Case Studies on Transport Policy 12 (2023) 101009

Contents lists available at ScienceDirect

Case Studies on Transport Policy

journal homepage: www.elsevier.com/locate/cstp

Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach

Juan Gerardo Muros Anguita [a], Oscar Díaz Olariaga [b,*]

[a] Escuela Internacional de Postgrado, Universidad de Granada, Granada, España
[b] Facultad de Ingeniería Civil, Universidad Santo Tomás, Bogotá, Colombia

ARTICLE INFO

Keywords:
Air cargo demand
Airport
Air transport public policy
Machine learning
Deep learning
Convolutional/recurrent neural networks

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), taking into account 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).

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

* Corresponding author.
E-mail addresses: e.juangerardo@go.ugr.es (J. Gerardo Muros Anguita), oscardiazolariaga@usta.edu.co (O. Díaz Olariaga).

https://doi.org/10.1016/j.cstp.2023.101009
Received 27 May 2022; Received in revised form 28 July 2022; Accepted 16 April 2023
Available online 20 April 2023

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 back-propagation 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) propose 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), aautoregressionintegrated 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) carry 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. Rodríguez 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 1990 s. 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

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

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

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.

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 $\boldsymbol{X}$ and, 3-dimensional tensors for the searched outputs $\boldsymbol{Y}$, 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 *fmodel 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 ($\boldsymbol{X}$ inputs), as shown in Equation (1).

$$\hat{y}_{j,t+n\_outs}, \hat{y}_{j,t+n\_outs-1}, \cdots, \hat{y}_{j,t} = f_{modelforecast}\left(x_{i,t-1}, x_{i,t-2}, \cdots, x_{i,t-n_{lags}}\right) \quad (1)$$

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 $t + n\_outs$ indicates the future prediction steps of the time t, therefore, the prediction Y consists of a vector of *n_outs* time components. The input variable $\boldsymbol{X}$ 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 $\boldsymbol{X}$ Input and $\boldsymbol{Y}$ Output couples necessary for supervised learning, can always be formed.

Also, the temporal data must be structured as 5D $\boldsymbol{X}$ and 3D dimensional $\boldsymbol{Y}$ tensors to adapt them to the input and output layers of the ConvLSTM2D neural model respectively (Ketkar and Moolayil, 2021). The $\boldsymbol{X}$ and $\boldsymbol{Y}$ 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 Fig. 1.

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 Fig. 2.

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.

Fig. 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 $\boldsymbol{X}$ 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 $\boldsymbol{Y}$ 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 Fig. 3, it is observed that the inputs of the $\boldsymbol{X}$ 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 $\boldsymbol{Y}$ 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.

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 || ||, where n is the number of samples to obtain the mean.

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

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

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

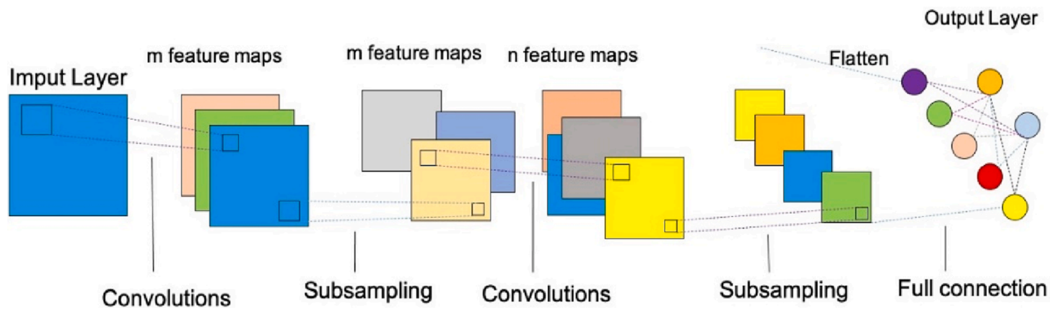

Fig. 1. General diagram to produce feature maps made by convolutional layers. .
Source: Hu et al., 2019

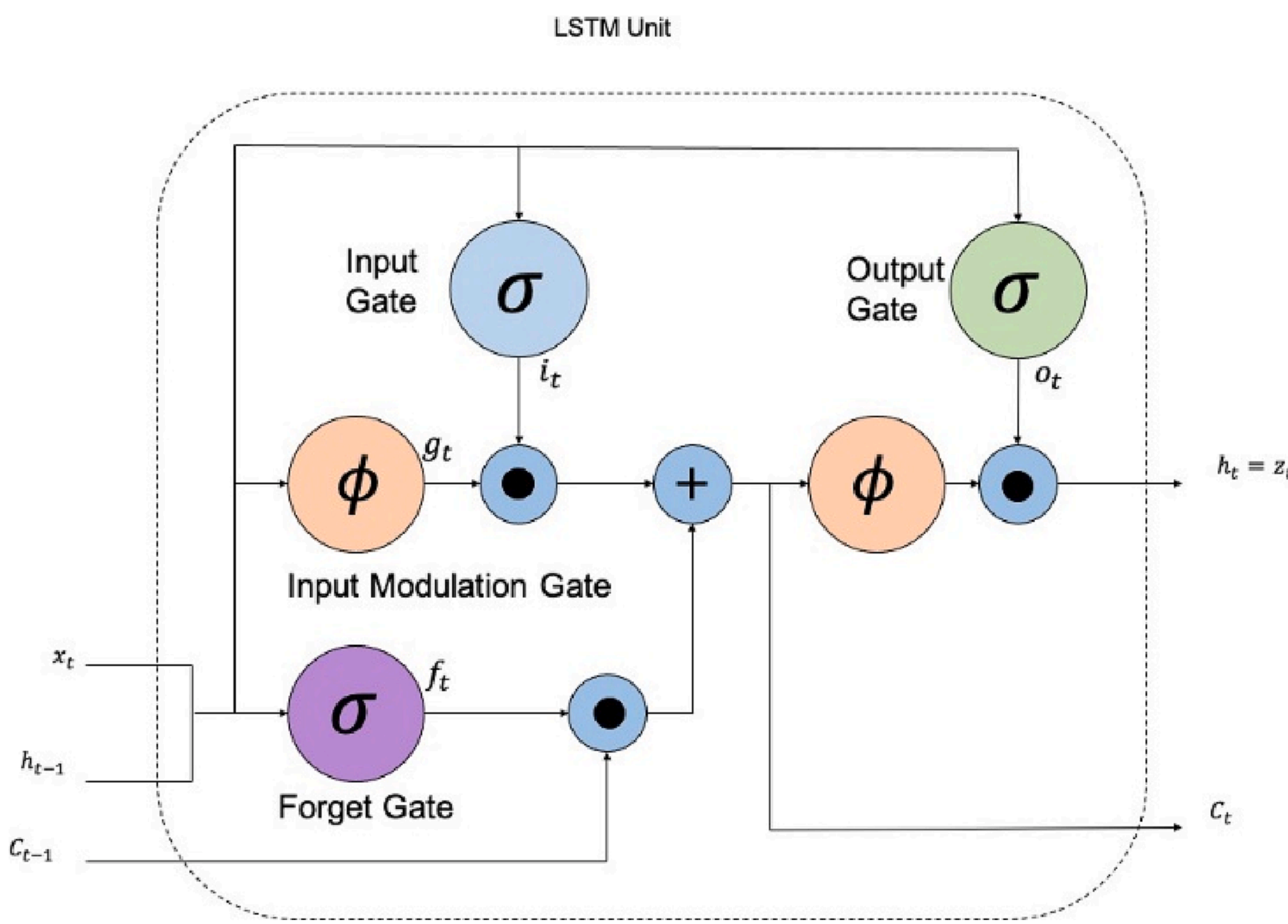

Fig. 2. Schematic of an LSTM long-short-term memory cell or unit. .
Source: Donahue et al., 2015

The training of the model consists of introducing the training tensor ***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 $\widehat{Y}(x)$, with which to compare the tensor of the expected output ***Y***, which defines the cost function MSE. Using theBackpropagation (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 ***X*** input and ***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 ***X*** input and ***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.

Figs. 4 and 5 show the change of the MSE during successive training

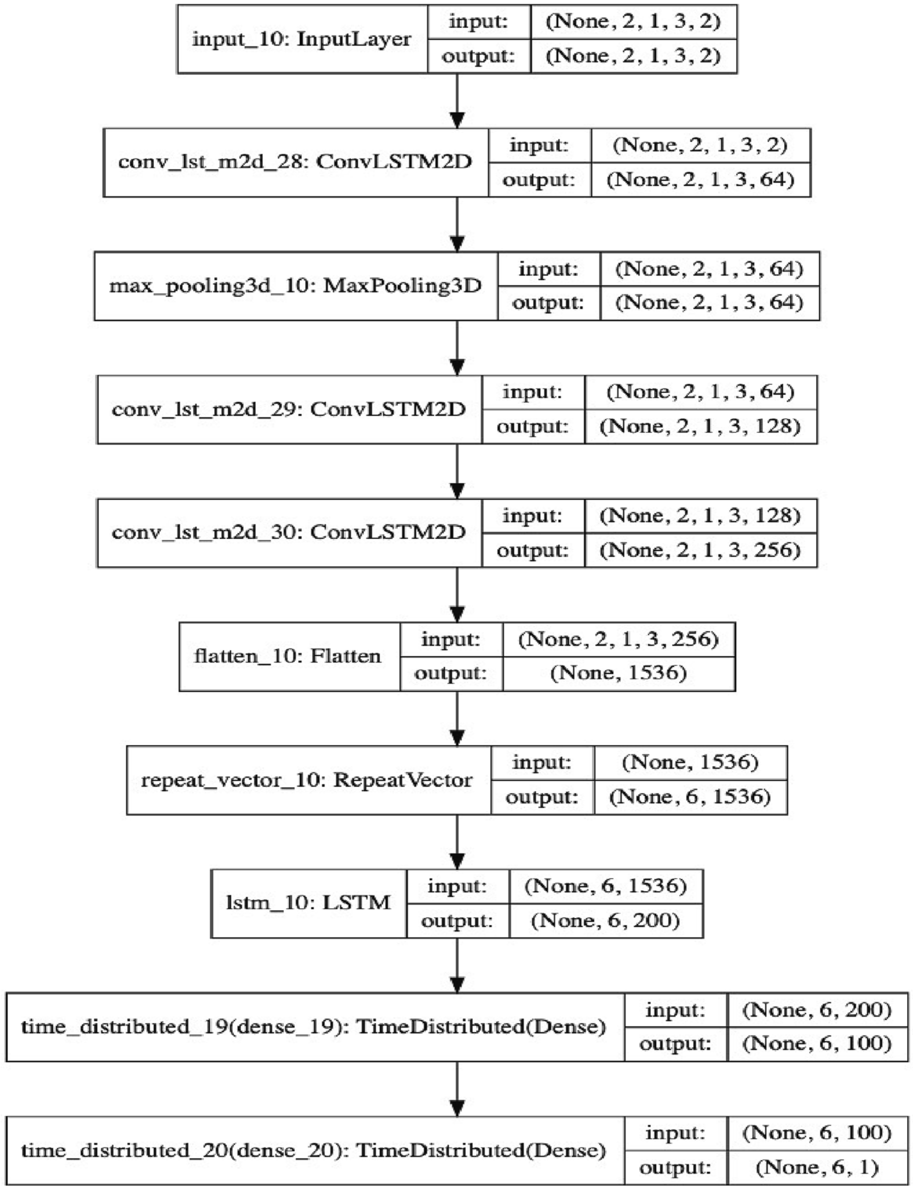

Fig. 3. Architecture implemented in the research, consisting of three blocks of ConLSTM2D layers. Source: authors.

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.

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.

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

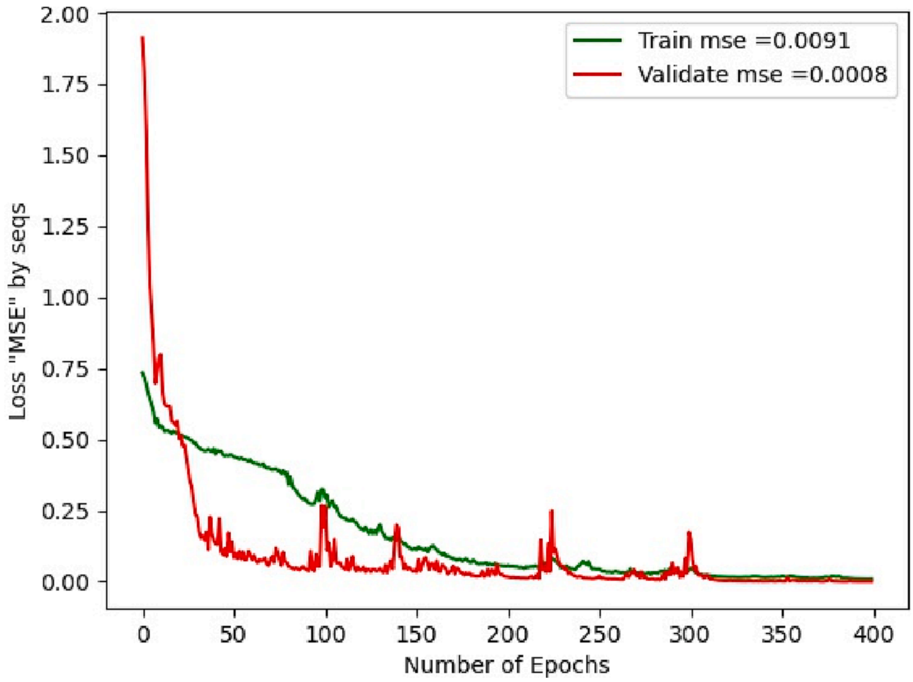

Fig. 4. Evolution of the MSE cost function during training vs number of epochs (or cycles) for both training and test samples. Source: authors.

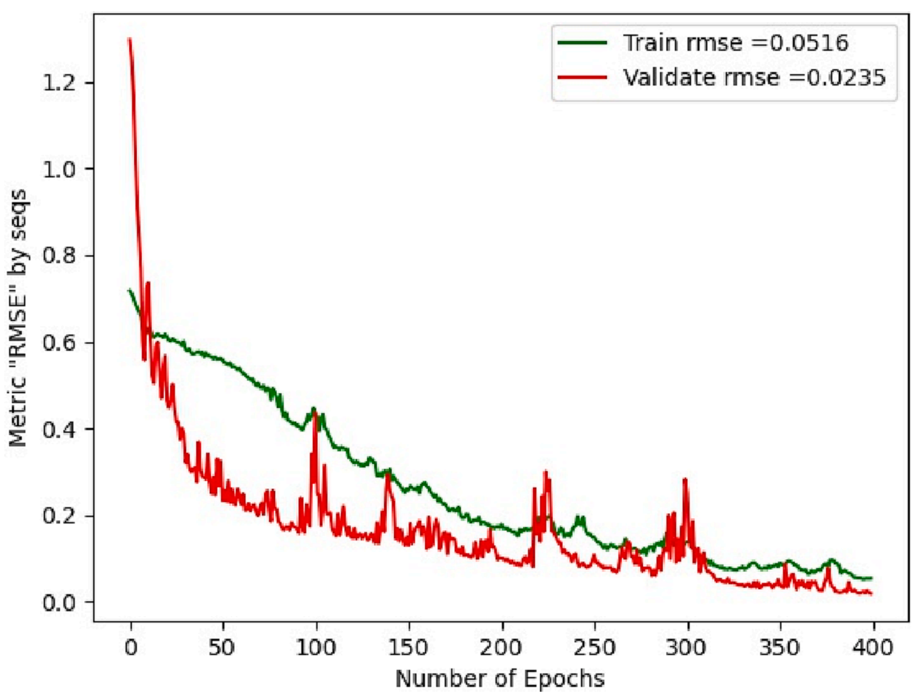

Fig. 5. Evolution of the RMSE metric during training vs cycle repetition for both training and test samples. Source: authors.

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 $\boldsymbol{X}$ inputs and 3D output $\boldsymbol{Y}$, 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

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.

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

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.

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.

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

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

To conclude this analysis, Fig. 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.

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.

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 $\boldsymbol{X}$ inputs and the 3D output $\boldsymbol{Y}$ 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

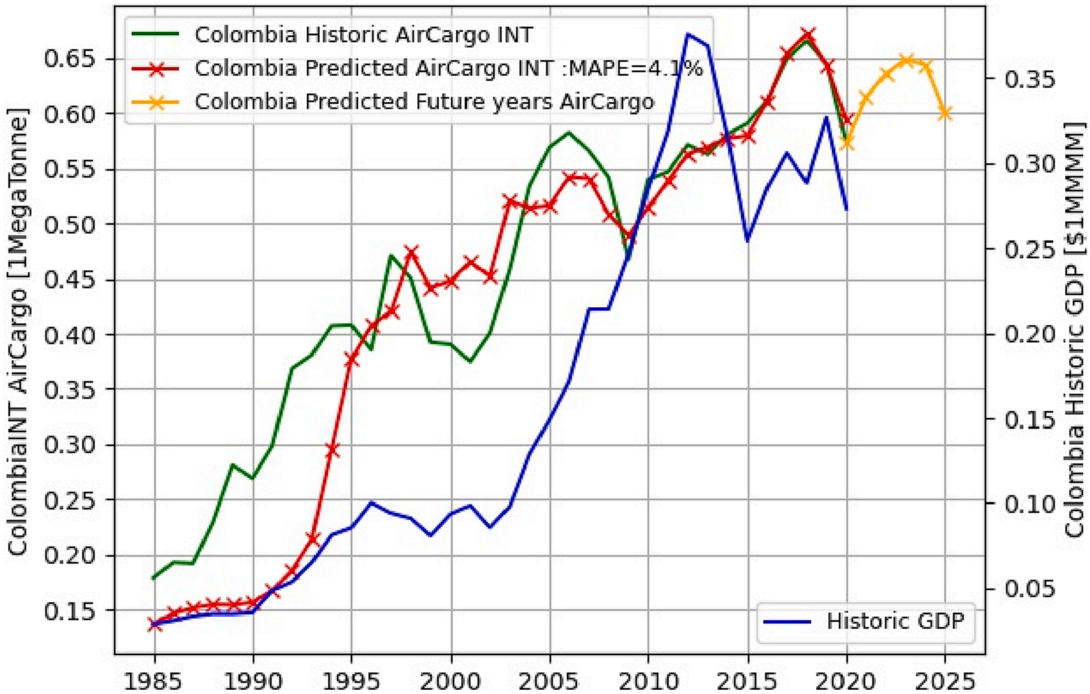

Fig. 6. Forecast of international air cargo demand (in million tons) as a function of the own historic evolution and the history of GDP (in USx10^{12}$).

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).

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

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

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.

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 concening international air cargo demand predictions.

Finally, Fig. 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

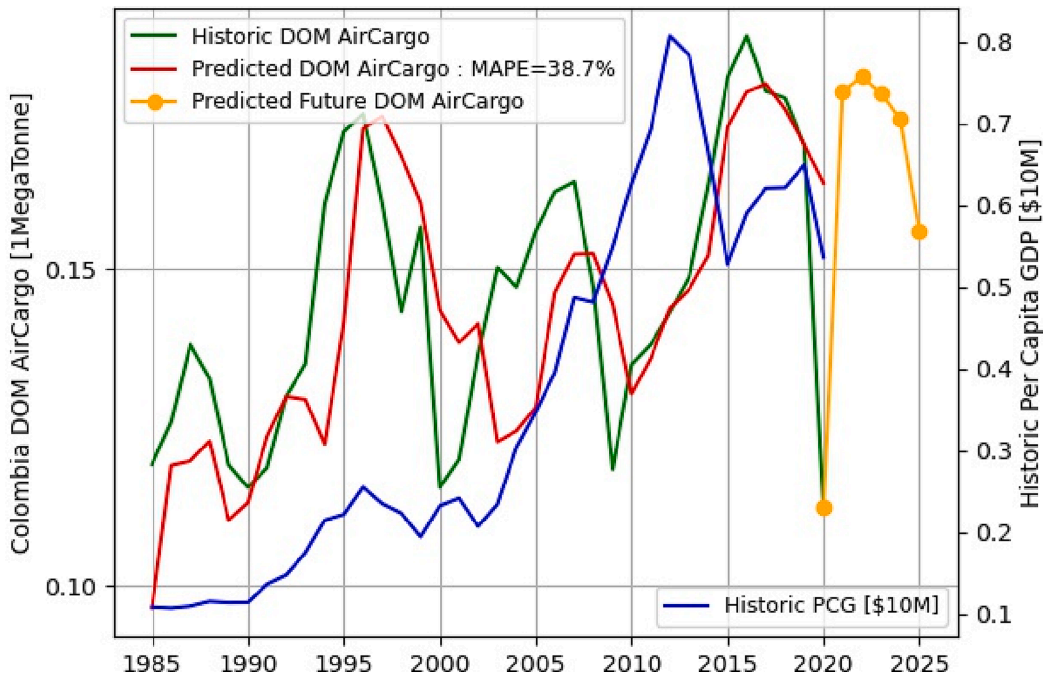

Fig. 7. Forecast of domestic air cargo demand (DOM) (in million tons) as a function of the historical series of the variable itself and the historical evolution of PCG (in US$x10^4).

how this curve is learning from less to more, adjusting better in the last years of the historical series).

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 1990 s, 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 Anguita 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 statical 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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aci, 2016. Traffic Forecast. ACI (Airports Council International), Montreal.
Aci, 2021. Annual World Airport Traffic Report. ACI (Airports Council International), Montreal.
Aerocivil (2022). 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. Springer International, Cham.
Airbus (2016). *Global Market Forecast 2016–2035*. Blagnac: Airbus.
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*. 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. https://doi.org/10.1016/j.tranpol.2020.04.004.
Ashford, N., Mumayiz, S., Wright, P., 2011. Airport Engineering. John Wiley & Sons, New Jersey.
Ashford, N., Stanton, M., Moore, C.A., Coutu, P., Beasley, J., 2013. Airport Operations. McGraw-Hill, New York.
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. https://doi.org/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. https://doi.org/10.1155/2012/431512.
Bianchi, F., Maiorino, E., Kampffmeyer, M., Rizzi, A., Jenssen, R., 2017. Recurrent Neural Networks for Short-Term Load Forecasting. Springer, Cham.
Blokdyk, G., 2017. Recurrent neural network. The Art of Service, Brendale (Australia).
Boeing,, 2016. World Air Cargo Forecast 2016–2035. Boeing, Seattle.
Boeing,, 2020. World Air Cargo Forecast 2020–2039. Boeing, Seattle.
Calin, O., 2020. Deep Learning Architectures. A Mathematical Approach. Springer, Cham.
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. https://doi.org/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. https://doi.org/10.1007/s11135-010-9342-8.
de Neufville, R., Odoni, A., 2013. Airport Systems, Planning, Design, and Management. McGraw-Hill, New York.
Díaz Olariaga, 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.
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/rsp3.1248326.
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.
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., Pulido, L., 2019. Measurement of airport efficiency. The case of Colombia. Transport and Telecommunication 20 (1), 40–51.
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.
Dingari, M., Reddy, M., Sumalatha, V., 2019. Air Traffic Forecasting Using Artificial Neural Networks. International Journal of Scientific & Technology Research 8 (10), 556–559.
DANE, 2022. https://www.dane.gov.co/.
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.jjimei.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. MIT Press, Cambridge (MA).

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.

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. Routledge, New York.

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. McGraw-Hill, New York.

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. https://doi.org/10.1016/j.jtrangeo.2010.09.001.

Iata, 2021a. World Air Transport Statistics 2021. IATA, Geneva.

Iata, 2021b. Air Cargo Market Analysis. IATA, Geneva.

Icao, 1987. Master Planning. Part 1. Doc 9184. ICAO, Montreal.

Icao, 2006. Manual on Air Traffic Forecasting. Doc 8991. ICAO, Montreal.

Icao, 2020. Presentation of 2019 Air Transport Statistical Results. ICAO, Montreal https://www.icao.int/annual-report-2019/Documents/ARC_2019_Air%20Transport%20Statistics_es.pdf.

Jadc, 2021. Worldwide Market Forecast 2020–2040. Japan Aircraft Development Corporation, Tokyo.

Janic, M., 2009. Airport analysis, planning and design: demand, capacity and congestion. Nova Science Publishers, New York.

Janic,, 2021. System Analysis and Modelling in Air Transport. CRC Press, Boca Raton.

Kazda, A., Caves, R., 2015. Airport design and operations. Bingley, Emerald.

Ketkar, N., Moolayil, J., 2021. Deep Learning with Python. Springer, New York.

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. https://doi.org/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. https://doi.org/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. https://doi.org/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., Piboonrungroj, 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.

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. Wiley, New York.

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

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

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*. Medellín, Colombia, pp. 277–294. https://doi.org/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. Springer, Cham.

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. https://doi.org/10.3846/aviation.2020.12273.

Sewak, M., Karim, R., Pujari, P., 2018. Practical Convolutional Neural Networks. Packt Publishing, Birmingham.

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. https://doi.org/10.1016/j.simpat.2012.05.012.

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

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.

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. Transportation Research Board, Washington DC.

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. McGraw-Hill, New York.

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. https://doi.org/10.1016/j.neucom.2020.07.061.