Deep Learning Based Forecasting of Delay on Flights

Araştırma Makalesi/Research Article

Anıl AYAYDIN, M.Ali AKÇAYOL

Bilgisayar Mühendisliği, Gazi Üniversitesi, Ankara, Türkiye
anil.ayaydin1@gazi.edu.tr, akcayol@gazi.edu.tr
(Geliş/Received:20.01.2022; Kabul/Accepted:28.04.2022)
DOI: 10.17671/gazibtd.1060646

Abstract— In this study, three different methods from machine learning and deep learning have been implemented for preventing financial and moral losses that may occur as a result of delays in flights and to take necessary precautions by predicting the flight delay in advance, which are a serious problem in the aviation industry. Deep recurrent neural network (DRNN), long-short term memory (LSTM), and random forest (RF) have been extensively tested and compared employing a real data set covering 368 airports across the world with relevancy the success rate of forecasting of delay on flights. The experimental results showed that the LSTM model had a higher success rate of 96.50% at the recall level than the others.

Keywords— estimation, deep learning, machine learning, aviation.

Derin Öğrenme Tabanlı Havacılık Uçuş Verilerinde Gecikme Durumunun Tahmin Edilmesi

Özet— Bu çalışmada, havacılık endüstrisinde ciddi bir sorun teşkil eden uçuşlarda yaşanan gecikmeler sonucu oluşabilecek maddi-manevi kayıpları önlemek ve uçuş gecikmesinin önceden tahmin edilerek gerekli önlemlerin alınabilmesi amacıyla makine öğrenmesi ve derin öğrenmeden oluşan üç farklı yöntem uygulanmıştır. Deep recurrent neural networks (DRNN), long-short term memory (LSTM) ve random forest (RF) yöntemleri kapsamlı bir şekilde test edilmiş ve dünya genelinde 368 havalimanını kapsayan gerçek bir veri seti kullanılarak uçuşların gecikme durumu tahmin edilmiştir. Deneysel sonuçlar, LSTM modelinin %96.50 recall değeriyle diğer modellere göre daha yüksek başarı oranına sahip olduğunu göstermiştir.

Anahtar Kelimeler— tahminleme, derin öğrenme, makine öğrenmesi, havacılık.

1. INTRODUCTION

According to the statistical data published by the United States Bureau of Transportation, 14.69% of the flights that have taken place up to now in 2020 haven't been on time thanks to delays or cancellations. of those flights, 0.19% were diverted (the plane landed at a unique airport than the planned airport) [1]. Delays in scheduled flights cause serious financial losses for airline companies, which also ends up in a decrease in passenger satisfaction. Supported this, many countries within the world have made deterrent legal arrangements so as to compensate their citizens for the damages caused by these delays that are likely to be experienced, and that they impose monetary penal sanctions on the relevant airline companies. However, although these sanctions are applied, there are still frequent delays and cancellations in flights. Delay codes for departures within the scope of economic passenger flights are standardized by The International Air Transport Association (IATA) and also the main causes of flight delays are unsuitable climatic conditions, traffic density, airport capacity (number of passengers, runways and bridges, etc.), technical and mechanical problems, national aviation systems and security [2].

In this study, three different methods from the deep learning and the machine learning are implemented for forecasting of delay on flights, which constitute a significant problem within the aviation industry. So as to train the model, unlike similar studies which haven't just some airports, operations were applied on the info set covering 368 airports.

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.

Three different models were presented within the study and independent result data were obtained by using different input files on these models. It's aimed that the results obtained within the scope of the study also will be a source for other studies associated with the aviation sector to be made with deep learning architectures [3].

2. LITERATURE REVIEW

There are many studies within the literature on the forecasting of aviation flights. Flight cancellation, delay and diversion are well known as a critical performance indicator of flights within the commercial aviation industry. Approximately 87.50% of educational studies on estimating flight delays transpire within the 2000s and were published between 2007 and 2017. Most of those studies are meted out with the assistance of Machine Learning algorithms and data processing methods [3]. Researchers evaluate flight delays from different perspectives. These are listed as optimization of airport planning, airport capacity increase, flight cancellation, facility location and flight change.
In the study by Kang and Hansen [4], the consequences of on-time and early arrival date on schedule block time (SBT) adjustment decisions were investigated using data from 5 major United States-based airlines. Within the study, changes in SBT were modeled for flights that befell two years in an exceedingly row. Flight data from January 2008 to April 2014 were obtained from the Federal Aviation Administration (FAA) and Aviation System Performance Metrics (ASPM) database. Within the methods used section, a mixed logit model is employed to capture the heterogeneous preferences of every airway and also the possible correlation between alternative flights. On-time and early arrival features are generated for every alternative arrangement. Mixed logic models are estimated supported these features and also the selected block time setting.
In the study of Kenan, Jebali and Diabat [5], the integrated flight planning, fleet allocation and therefore the problem of the aircraft routing was formulated.

Delays are considered to form the model more realistic. The upkeep route was considered indirectly by ensuring that every aircraft's route ends at the identical destination from which it started. Within the main problem, the answer is reached with the assistance of a column-based formulation containing an outsized number of variables. The dataset includes 228 destinations and 45 different destinations. The flights were allotted with 59 aircraft with 5 different aircraft types. During this study, a two-stage stochastic programming model is developed for the matter of integrated flight planning, fleet allocation and aircraft routing under uncertainty. The importance of this model lies within the possibilities which will extend beyond it. This model has proven to be of high complexity and so an advertisement solver like CPLEX cannot resolve samples of the complexity. During this study, three column generation-based approaches were developed, and every one three proved superior to CPLEX by resolving large samples in 4 hours to an optimality gap of 1%. Then, the effect of a number of these obtained parameters on the airline company was also examined by performing a sensitivity analysis.
Yazdi, Dutta, and Steven's study [6] examined the links between the applying of luggage charges and delayed flights within the airline industry. Because the dataset, a panel dataset of 46 quarters was collected from the Timely Performance Database from the US Department of Transportation's Bureau of Transportation Statistics. the dataset continues from the third quarter of 2003 to the fourth quarter of 2014. The On-Time Performance database provides information about non-stop domestic flights of major commercial airlines, including the actual point, estimated point, actual point and estimated point. It also reports minute flight-level delays divided into five categories of causes: Carrier Delay, Air Delay, Security delay, Late Aircraft Delay, and National Air System Delay. Within the study, it's been investigated how Baggage Fees (BF) affect the delays with the assistance of formulations created by arranging them per the flight time. Eleven carriers, ten of which apply baggage fees, were studied. The results show that, on average, RF implementation leads to improved on-time performance as judged by direct flights and indirectly through ticket prices and market demand. The results also show that developments are tormented by the hub-airports on the route and therefore the classification of passengers as leisure or business. The analysis shows that the primary introduce implementation will cause more flight delays, but full implementation really improves and still improves late flights.

Kim et al. [7] attempted to predict departure and arrival delays using flight order and weather data from the proposed Recurrent Neural Networks, National Oceanic and Atmospheric Administration. The accuracy of the delayed flight forecast was measured at 91.81% for McCarran International Airport and 71.34% for Sky Harbor International Airport because of the difference in data volume.

Choi et al. [8] focused on the link between flight delays and weather. Weather data were collected from the National Oceanic and Atmospheric Administration. As a result, Random Forest algorithm, which may be a proposed ensemble learning method, predicted the arrival delay with 80.36% accuracy.

Belcastro et al. [9] estimated flight delays because of weather condition conditions using Random Forest via MapReduce. The subsequent results were obtained using weather data from the National Oceanic and Atmospheric Administration. At a 15-minute delay threshold, 74.2% accuracy and 71.8% recall, at a 60-minute threshold, 85.8% accuracy and 86.9% recall values were obtained.

Thiagarajan et al. [10] made a flight delay estimation using the weather data on the locations of the departure and arrival airports, obtained via the globe Weather Online API service, with the assistance of Gradient Boosting, a machine learning technique. They achieved 94.35%

accuracy in arrival delays and 86.48% accuracy in departure delays.

Prasad et al. [11] achieved a 78% success rate in classification with Decision Tree and 77% in classification with Regression within the scope of flight delay estimation using the identical data source as Thiagarajan.

Yu et al. [12], using the departure and arrival flight delay data of PEK airport between January 2017 and March 2018, the accuracy rate was 93% with DBN-SVR estimation method, 87% with k-NN, 87% with Support Vector Machine and eventually it had been calculated as 82% with Logistic Regression.

Manna et al. [13] adopted the statistical approach method and using the Gradient Boosted Decision Tree classifier, on the flight delay dataset of the 70 busiest airports belonging to the US Department of Transportation within the April-October 2013 period, obtained approximately 92% accuracy in arrival delays and approximately 94% accuracy in departure delays.

In this study, a supervised learning model has been developed by using deep learning architectures to predict cancellation, delay or diversion situations in flights, which are a very important problem within the aviation industry. The information on the flights that transpire within the scope of 368 airports were used. At this time, a more comprehensive dataset than other studies within the literature has been studied. Thus, the high diversity of knowledge on the premise of airports and data lines which will repeat one another are prevented.

The dataset used has been normalized by digitization method by removing the erroneous data. DRNN, LSTM and RF methods were applied to the model, respectively. At the stage of teaching the model to the system, unlike similar studies within the literature, operations were administrated on the dataset covering 368 airports, not just some airports. The other literature studies were listed in below.

Table 1. The other literature studies

Source	Target	Method	Parameters
[14]	Taxi departure time estimation	Queuing model	Airline, terminal, air, destination, queue size
[15]	Delay prediction	Probability model, reinforcement leaning	Seasonal trend, delay data
[16]	Taxi departure time estimation	Reinforcement learning	Wheel departure time, wheel landing time, seasonal average number of taxi entries and exits
[17]	Delay classification	Probability model	Type of flight, number of passengers, number of delayed flights, visibility, wind speed
[18]	Delay prediction	Queuing model	Planned flight time, delay time
[19]	Delay prediction	Adaptive network	Destination, arrival time, arrival delay, scheduled arrival time
[20]	Delay prediction	Random Forest	Hour, day, month, delay status, delay day type, previous day type, airport delays, departure-destination places
[21]	Taxi departure time estimation	Linear regression	Route induced and interaction induced factors
[22]	Delay classification	Probability model	Airport and airline types, distance between airports, days
[23]	Delay prediction	Hybrid model	Departure-arrival flight delays
[24]	Delay classification	Queuing model	Number of daily flights by airport, expected fares from each airport, flight route
[25]	Delay classification	Data-based model	Planned and actual departure-arrival times.
[26]	Delay classification	Linear regression	Weather, operation demand rate and airport capacity at scheduled departure time, aircraft waiting time at the airport, scheduled return time, cargo delay, scheduled arrival time
[27]	Delay classification	Departure-arrival flight delays model	Type of aircraft, route, reason for delay, off-peak time, season
[28]	Delay classification	Stochastic Simulation model	City, route, aircraft type, season

3. DEEP LEARNING

Artificial intelligence encompasses the strategy of machine learning, within which machines can learn by experience and acquire skills without human. Deep learning, on the opposite hand, is defined as a subset of machine learning where artificial neural networks (ANNs) and algorithms inspired by human intelligence learn from data. Just like learning process of individuals, the deep learning algorithm has a much better result with a bit improvement for every time to enhance the end result. Any problem that needs thought is realized by deep learning. It's been determined that 2.5 quintillion bytes of information are produced a day [29]. Thanks to requiring more data to boost the learn, data generation has led to a rise the capabilities of the deep learning models in recent years. Additionally, deep learning algorithms make the

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.

most of the more powerful processing capacity available today, together with the event of ANN technology [30]. Artificial intelligence (AI) algorithms need for deep learning without an oversized initial investment. Even when employing a diverse, unstructured and interconnected dataset, deep learning enables machines to unravel successfully complex problems.

The deeper learning algorithms are trained, the higher they perform [31].

Convolutional neural network (CNN), which could be a forward-looking neural network from machine learning techniques, may be a sort of multilayer perceptron. It absolutely was first introduced within the early 1980s by LeCun [32] with the LeNet-5 architecture. This architecture consists of an input layer, several convolutional layers, pooling and output layers. On line with the convolutional layer level, they play a job within the extraction of features by performing operations on the inputs received from the previous layers. While the primary convolutional layer provides the lowest feature, it provides higher-level feature extraction because the convolutional layers are added, that is, the upper the extent. The pooling layer is; it's accustomed simplify the outputs produced with convolutional layers. The output layer of the architecture; it is often connect to completely connect layers or two layers, like the sigmoid layer. Data within the input layer; Since it may be multimedia data, like sound, image, video, the researchers preferred it in many signal processing fields in recent years thanks to its high performance [33]. A technique to cut back the margin of error is implemented by employing a back-propagation algorithm that adjusts the training weights to be updated with margins of error throughout the training process of CNN architectures [34].

Autoencoders are one in all the popular models within the field of deep learning. Because the name suggests, it aims to automatically learn to convert any data into a code. It consists of two parts: Encoder and Decoder. These two parts are trained together as if they were one model during the training phase. After the training is over, these models are separated and accustomed compress the info and decompressed the compressed data. As an example, if the info is to be moved from one place to a different, less data will be transported by placing an Encoder on the sender and a Decoder on the receiver.

Since recursive neural networks (RNNs) models the behavior of dynamically changing systems through their hidden layers and are defined as a sort of ANN during which the output from the previous step is fed as input to the subsequent step, it stands out because the method that works best with the information set utilized in the study and has been preferred in practice for this reason.

3.1. Recurrent Neural Networks

Recursive neural networks (RNN) models the behavior of dynamically changing systems through their hidden layers. RNN is a type of neural network in which the output from the previous step is fed as input to the current step. In traditional neural networks, all inputs and outputs are

independent of each other, but when it is necessary to predict the next word of a sentence, previous words are required and therefore there is a need to remember previous words. As a result, RNN emerged and solved this problem with the help of a hidden layer [35]. Long Short-Term Memory Networks (LSTM), as a RNN application, can make faster and more accurate predictions than standard RNN. The general structure of the RNN and LSTM architectures is explained in the following sub-title of this section. Afterwards, the benefits of using these networks together are stated, and the use of RNN to deepen the architecture is also explained.

Given the input sequence, $\mathrm{x} = \left(\mathrm{x}_1, \mathrm{x}_2, .., \mathrm{x}_\mathrm{k}, .., \mathrm{x}_\mathrm{T}\right)$ (1) the RNN calculates the ordinal values of the hidden layers $h = \left(h_1, h_2, .., h_K, .. h_T\right)$ (2). As a result, the output sequence becomes $\mathrm{y} = \left(\mathrm{y}_1, \mathrm{y}_2, .., \mathrm{y}_K, .., \mathrm{y}_\mathrm{T}\right)$ (3). These calculations are iteratively repeated by applying the given equations from t=1 to T. Here, x_k, h_k and y_k can be any arbitrary value. Vectors are specified as input value, hidden layer and output value [36].

$$h_t = \varphi_h \left(W_{hh} h_{t-1} + W_{xh} x_t + b_h\right) \quad (4)$$

$$y_t = \varphi_0 \left(W_{hy} h_t + b_y\right) \quad (5)$$

W_{hh}, denotes the weight matrix for the transition between the hidden layers in the previous step and the current step, the weight matrix W_{xh} between the input and W_{hy} the hidden layer, b_h and b_y the weight matrix between the hidden layer φ_h and φ_0 the output, and corresponds to the bias value in each equation, and correspond to the activation functions between the hidden layers and the output. For activation, one of the logistic sigmoid, hyperbolic tangents or ReLU functions can be applied [37].

3.2. Long-Short Term Memory Neural Networks

The LSTM architecture outperforms the traditional RNN architecture in the long run by storing the memory cells in the standard RNN architecture φ_h and φ_0 the hidden layer information expressed with. In this study, Alex Graves et al. The LSTM architecture proposed by [38] was used. This single memory cell is repeated in each iteration of the model. It consists of input gate (i), forget gate (f), output gate (o), cell activation vectors (c) and all are the same size as the hidden layer (h).

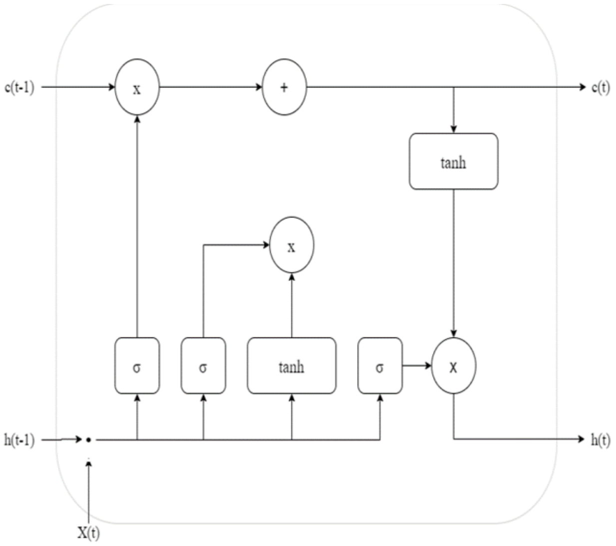

Figure 1. LSTM unit structure

The structure of the UKVH unit is shown as an example. In the UKVH unit structure shown in the Figure 1; as input, X(t) takes the current input value, h(t-1) takes the previous hidden state, and c(t-1) takes the previous memory state. As output; h(t) produces the current hidden state and c(t) the current memory state.

The following equations represent the calculations in the model.

$$i_t = \varphi\left(W_{xi}x_t + W_{ht}h_t - 1 + W_{ci}c_t - 1 + b_i\right) \ (1)$$

$$f_t = \varphi\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \ (2)$$

$$c_t = f_t c_t - 1 + i_t tanh\ \left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \ (3)$$

$$o_t = \varphi\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \ (4)$$

$$h_t = o_t tanh\ \left(c_t\right) \ (5)$$

φ symbol represents the logistic sigmoid function.

3.3. Deep Recurrent Neural Networks

When the recent studies [39], [40] are examined, it is seen that the deep and hierarchical model gives more efficient and accurate results. Based on this hypothesis, it was decided to design the model with deep learning architecture for flight cancellation and delay-steering prediction. For this purpose, four different LSA models have been created and these are expressed as input to hidden layer, hidden layer to output, hidden layer to hidden layer and hidden layer stack.

In this research, LSA was applied from the hidden layer to the output layer in the function of the input to the hidden layer direction. The related equation is given below:

$$h_t^{(l)} = f_h^{(l)}(h_t^{(l-1)}, h_{t-1}^{(l)} = \varphi_h(W_l h_{t-1}^{(l)} + U_l h_t^{(l-1)} \ (1)$$

Here $h_t^{(l)}$ represents the hidden layer at the level. It is used instead of $h_t^{(l-1)}$ when l=1. From l=1, the hidden layers all levels are recalculated.

4. ANALYSIS OF AVIATION FLIGHT DATA AND FUTURE PREDICTION

4.1.Dataset

In the dataset, there are domestic flights within the borders of our country between the years 2010-2020 and international origin flights between the flight points. Within the scope of the dataset, data with the amount of 62597 rows were obtained after the data cleaning phase, which was dispensed with the assistance of a console application developed separately with Python. Accordingly, by applying cross validation, 80% training and 20% test data. The dataset obtained during this study, flights were classified in line with delay/not delay status.

In the next sub-heading, recursive neural networks, which may be a deep learning architecture and determined as a way during this project, are going to be mentioned.

Table 2. Features of the flight dataset

Feature	Value
Class	Anonym
Category	Aviation and flight data
Subcategory	Flight Dataset
Data Owner	General Directorate of State Airports Authority
Description	Aviation Flight Dataset which has 62597 row data
Format	CSV
Keywords	Flight, Airport, Cancelled, Delayed, Diversion
Contact	https://www.dhmi.gov.tr
Row Count	62597
Update Ti	May 2020
Size	4.6 MB

4.1.1. Features of Dataset

The following 10 features were obtained from the raw data obtained. While the values of the used attributes are included as text in the raw data set, the values of all

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.

attribute fields are classified and categorized. These attributes and their categories are listed below:

YEAR: It is the year of the flight.

MONTH: It is the month of the flight.

DAY: It is the day of the flight.

TARIFE_HOUR: It is the time of the flight.

IS_DELAYED_OR_NOT: It is the cancellation or delay/diverting status of the flight (Not delay: 0, Delay: 1).

DESTINATION: The destination of the flight. The digitized categories are presented in the appendices.

ORIGIN: It is the departure point of the flight. The digitized categories are presented in the appendices.

DELAY_TIME: Indicates how much time spends for latency as hours of delay. The table shown as below.

Table 3. Time of delay column key value table

DELAY_TIME	
Key	**Value**
0	Empty
1	Between 0-1 Hours
2	Between 1-2 Hours
3	Between 2-4 Hours
4	4 Hours and Over

DELAY_REASON: The reason for the delay. The table shown as below.

Table 4. Reason of delay column key value table

DELAY_REASON	
Key	Value
0	Empty
1	From his arrival
2	Sourced from the Airport Operator
3	Meteorological
4	Company
5	Technical
6	Patient
7	Wind
8	Rain
9	Team Standby
10	Operational
11	Fuel Critic
12	Fault
13	Passenger Sickness
14	Team Sickness

DELAY_DESCRIPTION: It contains extra information about the delay, if any. The table has shown as below.

Table 5. Description of delay column key value table

DELAY_DESCRIPTION	
Key	Value
0	Empty
1	Runway Density
2	Maintenance / Repair
3	Anti-Icing Application
4	Icing
5	Snow
6	Fog
7	Wind
8	Rain
9	Team Standby
10	Operational
11	Fuel Critic
12	Fault
13	Passenger Sickness
14	Team Sickness

4.2. Forecasting of Aviation Flight Data with Deep Learning

Hyperparameters of the models have shown in Table 6 .

Table 6. Hyperparameters of the models

Hyperparameter	**DRNN**	**LSTM**	**RF**
Layer (Unit) Structure	128-64-64-1	64-32-1	-
Number of Dense	3	3	-
Activation Function	ReLU, Sigmoid	Sigmoid	-
Loss Function	MSE	MSE	-
Optimizer	Adam	RMSProp	-
Number of Estimators	-	-	13

4.2.1. Deep Recurrent Neural Network Method

In the scope of the study, DRNN has been used as a supervised learning model. The architecture used works iteratively as mentioned within the previous titles, and also the output data of the previous step affect the output of the following steps. Since the information sets employed in this study is compatible with the model in question, DRNN architecture has been the well-liked method of analysis. Since single-layer artificial neural network (ANN) models are insufficient for complex models, multi-layer feedback ANN, that is, the architecture where the outputs of the layers are fed back to the previous layers, is preferred. Within the model built as a result of this, there are multiple hidden layers that form the idea of deep learning architectures additionally of the input and output units. There's no analytical method for determining the number of layers and neurons within the hidden layers, and also the methods suggested by Karsoliya were used because the upper limit [41].

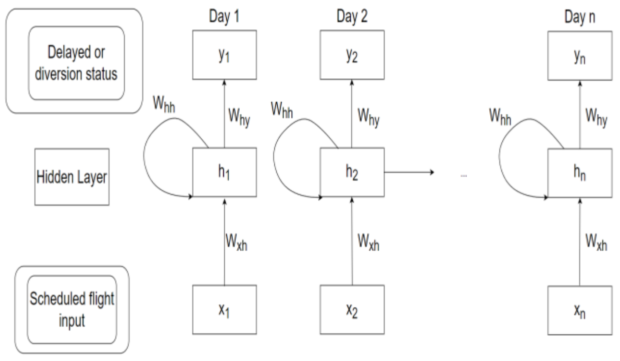

Figure 2. The DRNN model

In the diagram above corresponds to the load matrix for the transition between the previous step and therefore the hidden layers on the present step, , the load matrix between the input and hidden layers, , the burden matrix between the hidden layer and therefore the output. There are a variety of hidden layers between the planned flight input and output layers. Since the results of the previous step are iterative to affect the subsequent one, it's an acceptable model for the aviation flight data set utilized in this study.

4.2.2. Long Short-Term Memory Model

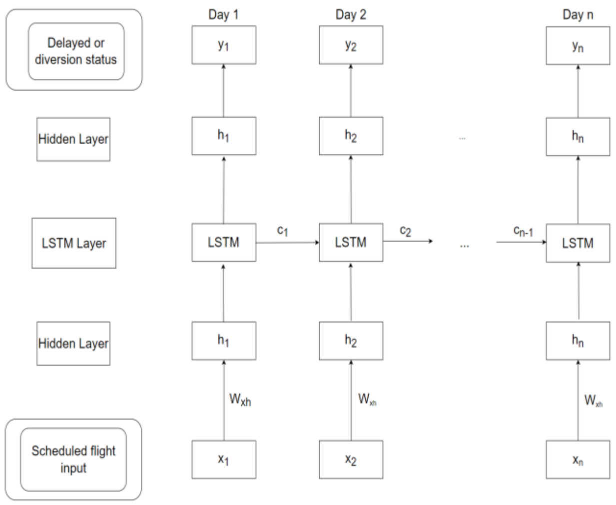

Figure 3. LSTM model

Here, and within the LSTM model. As n are several hidden layers and LSTM layers between the planned flight input and output layers. Refers to the memory value of this step and corresponds to the memory value of the previous step. Since the memory data of the previous step are kept within the LSTM layer, this model developed to be implemented on the aviation flight data set increases the performance.

4.2.3. Random Forest Model

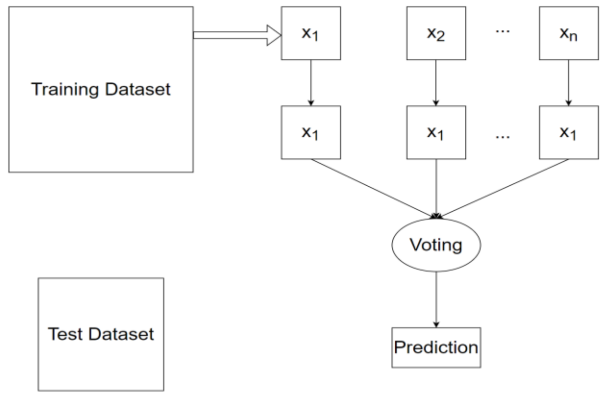

Figure 4. RF model

As described above, within the flight prediction model with RF, the simplest prediction result's obtained by comparing the planned flight inputs from the flight training data set with the previous one, taking the higher value of voting and similarly planning to the worth through this comparison with the worth of n.

Although there are different approaches within the literature as a training method, the DRNN model was compiled using Jupyter Notebook to see the foremost appropriate and effective training function, and it had been observed that the LSTM has better results compared to other tests, and this model was also preferred in experimental studies.

In the model built with the RF algorithm, the date, time, departure and arrival points were determined and estimation was made with 111 input parameters on a single row of data and the help of a way developed. Within the DRNN model, additionally to creating the estimation process using the identical parameters, the system was trained at the determined iteration value by creating a multiple hidden layer artificial neural network model and also the correct classification values were measured. Within the LSTM model, unlike the DRNN model, the results of the previous step are stored with the assistance of a memory cell and therefore the results are obtained. within the light of these data, the DRNN architecture has progressed numerically. However, no distinctive difference may well be detected in terms of accuracy rate between DRNN and LSTM recursive deep learning architectures.

Although the RF machine learning algorithm gives better leads to terms of correct classification, deep learning models are the main focus of the study. During this context, the LSTM model incorporates a higher accuracy value

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.

Compared to the DRNN model and can set an example in terms of getting used in future studies concerning the aviation industry.

5. THE EXPERIMENTAL RESULTS

Approximately 75% of the dataset used consists of delayed/redirected flights, while 25% consists of canceled flights. Accordingly, the success rate in classification was approximately 75% when the DRNN architecture, which was modeled as cancellation or delay/redirect by applying cross validation with 80% training and 20% test data, was trained. Dropout values are added to every hidden layer to scale back the memorization that may occur with each iterative step.

Using the DRNN deep learning architecture, the model was built up as 128-64-64-1 (input-hidden layer 1, hidden layer 2-output) neurons in the layers. The model was trained on 37557 samples, validated on 25020 samples. It has been trying to prevent overfitting by using dropout in the middle layers. As the activation function, the ReLU function is used, which gives good results for supervised iterative models in the input and hidden layers. The activation function in the output layer was determined as sigmoid and the best result was tried to be obtained. While building the model, *MSE* was chosen as the loss method, optimizer *MAN*, and *accuracy* as the metric value. When the established model is trained to have an epoch value of *10*, Figure 5. The Training and Test Accuracy Graph on the top and the Training and Test Error Graph in Figure 6 were obtained.

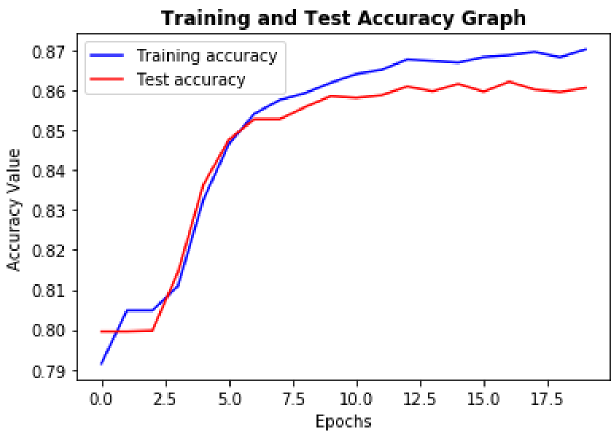

Figure 5. DRNN training and test accuracy graph

When the Training and Test Accuracy Graph in Figure 5 is examined, the training accuracy increased slightly after the first iteration, and the test accuracy tended to remain constant from the first step. When the model parameters are changed or a middleware is added, no significant change has been detected in the result, and the difference between training and test accuracy is very low. This demonstrates that the graph shows a consistent result.

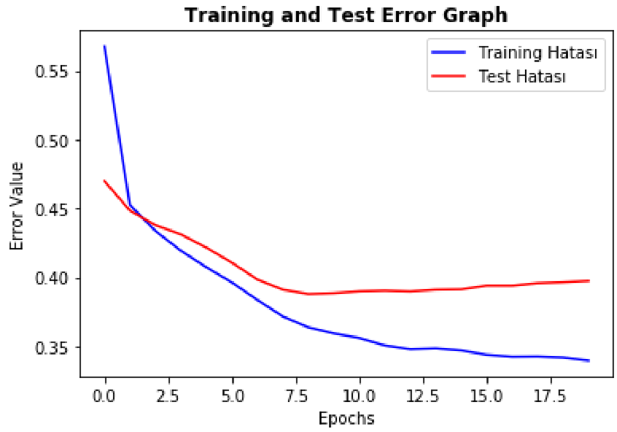

Figure 6. DRNN training and test error graph

When the Training and Test Accuracy Graph in Figure 6 is examined, the training error curve started to decline from the first step, and while the test error curve was stable, it tended to increase in between. This is due to the high batch size value. When this value is lowered, the stability increases, but the CPU and memory usage goes up to 99% on a machine with a 4-core CPU and 16 GB of memory. The model can be retrained on machines with high CPU-GPU-Memory values to obtain a more stable graphics. In this context, the training time of the model may be longer, but the amount of fluctuation in the error curve to be obtained will be less.

In the LSTM model, unlike the DRNN model, the information from the previous time can be transferred to the next with the help of the memory cell. Differently, scaling was performed between 0-1 values due to the dependency on the data scale. However, it was then converted back to real scale by inverse transformation. As another different element, the time step value has been defined and the size of the input data has been adjusted. By using the dropout parameter, as in the DRNN model, excessive memorization of the system is prevented. The 64-32-1 neuron layer structure was applied to the model. The model was trained on 37557 samples, validated on 25020 samples. Sigmoid is again preferred as the activation function in the output layer. Figure 7 shows the training and the test accuracy graph, and Figure 8 has the training and a test error graph.

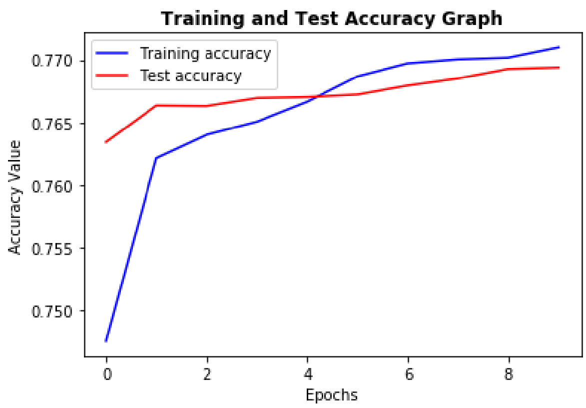

Figure 7. LSTM training and test accuracy graph

When the training and test accuracy values in Figure 7 compare with the DRNN model, it is observed that there is not much difference. The reason why the accuracy value was lower in the LSTM model was due to the fact that the amount of training data was reduced to 60% in order to obtain a stable result. When the same amount of training data is used, the LSTM model is ahead in terms of accuracy with a very small margin. As a deep learning model, in the light of these results, it is planned to use the LSTM model in future studies, since it can store the data from the previous iteration through the memory cell. On the other hand, there are some fluctuations in the test accuracy curve that do not have a great effect on the result. As in the previous model, this problem can be avoided by reducing the batch size and re-doing the training and testing processes on machines with higher equipment.

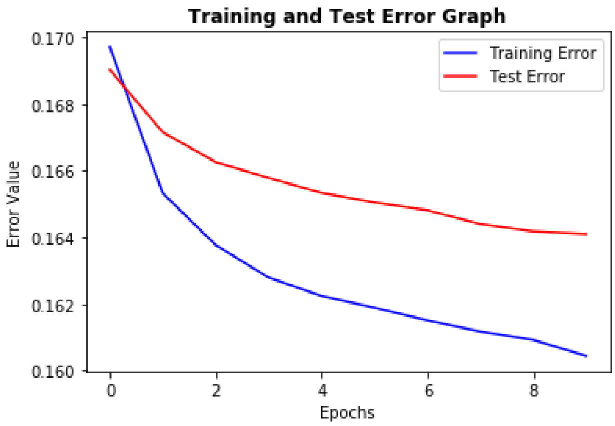

Figure 8. LSTM training and test error graph

In the training and the test error graph shown in Figure 8, the training error curve showed a sharp downward trend until the third step; however, after this step, the decline continued at a slower pace. The test error curve decreased until the second step and after the third step, the decrease slowed down and approached to be stable. As a result, when we look at the graph, it has been determined that there is a very small difference between the training and test error at the level of about 0.4%.

Since it can store the knowledge of the hidden layers within the memory within the long run, the success rate of roughly 79% was achieved in classification by applying the LSTM, which outperforms the DRNN.

Before the appliance of deep learning architectures, which is that the main subject of the study, with RF, a machine learning algorithm during which classification processes are applied by training the features of the architecture, the classification was made and also the result data were obtained. In line with this, DRNN was more successful than the RF method with an accurate classification rate of about 86.06%, although it had been not a giant difference. Meanwhile experimental results show that the recall value has the best success rate when LSTM is used as 96.50%. This difference is thought to be due to the use of different metrics while optimizing the model. For example, while the LSTM model was trained, RMSProp was used as the activation function as the sigmoid optimizer function, while ReLU was used as the activation in the DRNN model and the Adam function was used as the optimizer.

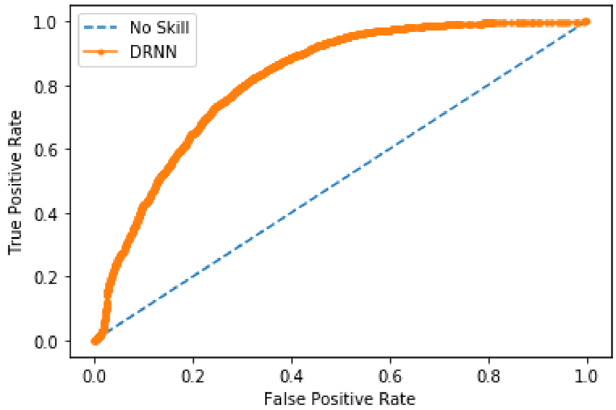

Figure 9. DRNN model ROC-AUC curve

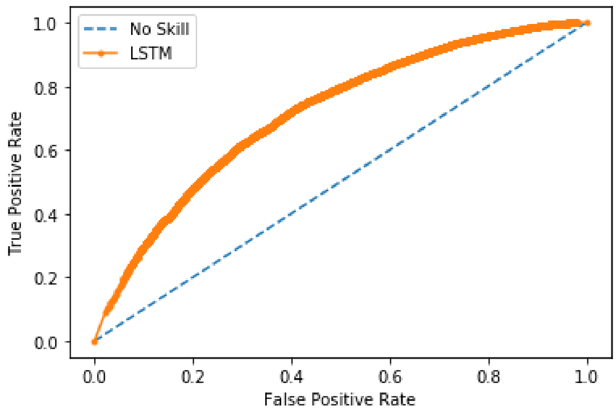

Figure 10. LSTM model ROC-AUC curve

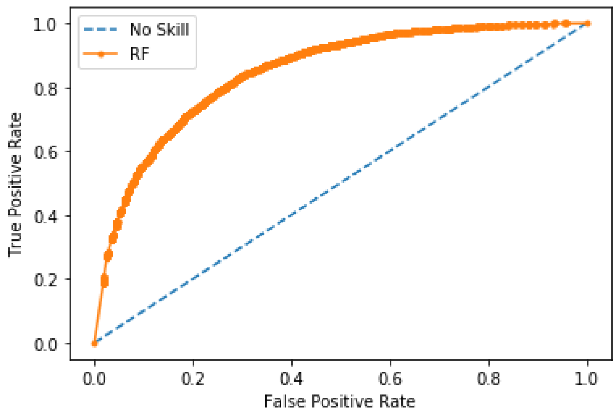

Figure 11. RF model ROC-AUC curve

According to the ROC AUC graphs, DRNN is considered as the most successful classifier model. When evaluated as a deep learning model, a more successful result was obtained as a result of the different number of layers and optimizing the model using different hyper parameters

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.

compared to the LSTM model. In addition to, the AUC value being in the range of 0.5-1 indicates that the model performs a successful classification.

Among the models applied to support deep learning, the LSTM is that the architecture with the very best numerical value in terms of correct classification and performance percentage. As a result of the experiments on all models, the leads to Table 7 were obtained.

Table 7. Values calculated by means of different models according to the parameters

Scores	**RF**	**DRNN**	**LSTM**
Accuracy (%)	82.21	86.06	76.96
Recall (%)	96.21	96.37	96.50
Precision (%)	96.18	87.47	77.96
F-score (%)	96.20	91.70	86.30
ROC-AUC	90.11	81.87	71.68

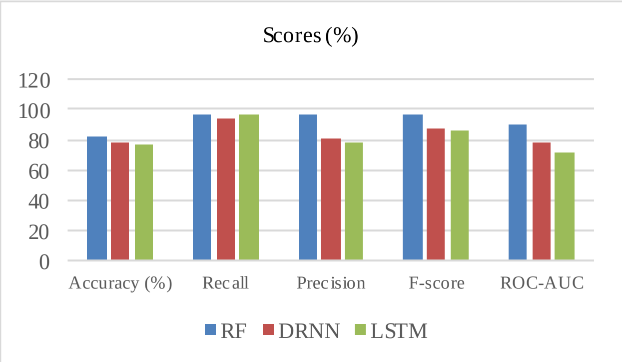

Figure 12. The graph of calculated performance metrics for different models

As shown in Figure 5, RF stands out as the most successful model, although there is no big difference when looking at the accuracy metric. Looking at the Recall metric, the most successful algorithm in classifying as flight delay was LSTM. Another metric is a precision, that is, the most successful algorithm for the correct classification of flight is RF. Within the scope of the F-score value, which is the harmonic mean of Recall and Precision values, the most successful algorithm was again RF. The most successful model in terms of the ROC-AUC metric, which expresses the classification capacity of the model is RF. The higher this value, the more successful the model is that. In general perspective, the value differences between the metrics are caused from the differences of the parameters used in the build phase of the models.

6. CONCLUSIONS

In this study, it's aimed to make sure the feasibility of operations like classification and estimation by creating a system within which deep learning architectures are applied by making use of the flight data obtained, within the scope of reducing the negative effects of cancellation, delay and diversion events within the aviation ecosystem, where our country has become one in all the few centers within the world in recent years. DRNN, LSTM and RF are applied for the forecasting of cancellation, delay and diversion on flights. a true data set covering 368 airports across the globe is used for training and testing of the models. The experimental results show that the recall value has the best success rate when LSTM is used as 96.50%.

ACKNOWLEDGEMENT

We would prefer to express our gratitude and relation to the General Directorate of State Airports Authority and its esteemed managers, who are the authority of the airport management sector in Turkey, while creating the flight data set within the scope of this study for allowing the utilization of the general public flight data and providing financial backing for participation to the conference.

REFERENCES

[1] Internet: Bureau of Transportation Official, Bureau of Transportation Statistics, https://www.bts.gov/, 15.11.2021

[2] L. Belcastro, F. Marozzo, D. Talia, and P. Trunfio, "Using Scalable Data Mining for Predicting Flight Delays", *ACM Trans. Intel. syst. Technol.*, 1, 5-20, 2016. [3]A. Sternberg, J. D. A Soares, D. Carvalho, E. Ogasawara, "A Review on Flight Delay Prediction.", *arXiv: Computers and Society*, 2017.

[3] A. Sternberg, J. D. A Soares, D. Carvalho, E. Ogasawara, "A Review on Flight Delay Prediction.", *arXiv: Computers and Society*, 2017.

[4] L. Kang, M. Hansen, "Behavioral analysis of airline scheduled block time adjustment", *Transp. pic. Part E: Log. Transp. Rev.* 103, 56–68, 2017.

[5] N. Kenan, A. Jebali, A. Diabat, "The integrated aircraft routing problem with optional flights and delay considerations", *Transp. pic. Part E: Log. Transp. Rev.*, 118, 355–375, 2018.

[6] A. A. Yazdi, P. Dutta, A. B. Steven. "Airline baggage fees and flight delays: a floor wax and dessert topping?", *Transp. pic. Part E: Log. Transp. Rev.*, 104, 83–96, 2017.

[7] YJ. Kim, S. Choi, S. Briceno, D. Mavris, "A deep learning approach to flight delay prediction.", **2016 IEEE/AIAA 35th Digital Avionics Systems Conference**, USA, 1-6, 2016.

[8] S. Choi, YJ. Kim, S. Briceno, D. Mavris, "Prediction of weather induced airline delays based on machine learning algorithms", **2016 IEEE/AIAA 35th digital avionics systems conference (DASC)**; USA, 1-6, 2016.

[9] L. Belcastro, F. Marozzo, D. Talia, P. Trunfio, "Using scalable data mining for predicting flight delays.", *ACM Trans Intel SystTechnol*, 8, 1–20, 2016.

[10] B. Thiagarajan, L. Srinivasan, A. V. Sharma, D. Sreekanthan, V. Vijayaraghavan, "A machine learning approach for prediction of on time performance of flights.", **In: Proceeding of IEEE/AIAA 35th digital avionics systems conference**, USA, 6, 2016.

[11] U. S. Prasad, P. A. Chauhan, L. S. Asha., "Data mining & predictive analysis on airlines performance.", **Int J Pure Appl Math.2018;118:1–12. AIAA 35th digital avionics systems conference (DASC)**, Sacramento, CA, USA, 6, 1–6, 2016.

[12] B. Yu, Z. Guo, S. Asian, H. Wang, G. Chen, "Flight delay prediction for commercial air transport: A deep learning approach", *Transportation Research Part E*, 125, 203 – 221, 2019.

[13] S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, S. Barman, "A Statistical Approach to Predict Flight Delay Using Gradient Boosted Decision Tree", **International Conference on Computational Intelligence in Data Science**, Chennai, India, 1-5, 2017.

[14] H. Idrıs, J. Clarke, R. Bhuva, L. Kang, "Queuing Model for Taxi-Out Time Estimation", *ATC Quarterly*,10(1), 1-22, 2002.

[15] Y. Tu, M. Ball, W. Jank, "Estimating Flight Departure Delay Distributions – A Statistical Approach with Long-Term Trend and Short-Term Pattern", *Journal of the American Statistical Association*, 103, 112-125, 2008.

[16] P. Balakrishna, R. Ganesan, L. Sherry, "Accuracy of Reinforcement Learning Algorithms for Predicting Aircraft Taxi-Out Times: Case-Study of Tampa Bay Departures", *Transportation Research Part C: Emerging Technologies*, 18, 950-962, 2012.

[17] R. Wesonga, F. Nabugoomu, P. Jehopio, "Parameterized Framework for the Analysis of Probabilities of Aircraft Delay at an Airport", *Journal of Air Transport Management*, 23, 1-4, 2012.

[18] M. Arikan, V. Deshpande, M. Sohoni, "Building Reliable Air-Travel Infrastructure Using Stochastic Models of Airline Networks", *Operations Research*, 61 (1), 45-64, 2013.

[19] S. Khanmohammadi, C. Chou, H. Lewis, D. Elias, (2014), "A Systems Approach for Scheduling Aircraft Landings in JFK Airport", **IEEE International Conference on Fuzzy Systems (IEEE 2014)**, Beijing, China, 11, 1578-1585, 2014.

[20] J. Rebollo, H. Balakrishnan, "Characterization and Prediction of Air Traffic Delays,", *Transportation Research Part C: Emerging Technologies*, 44, 231-241, 2014.

[21] O. Lordan, J. Sallan, M. Arroyo, Forecasting of Taxi Times: The Case of Barcelona-El Prat Airport, *Journal of Air Transport Management*, 56, 118-122, 2016.

[22] P. Rodriguez, P. Sanchez, E. Gomez, Modelling the Asymmetric Probabilistic Delay of Aircraft Arrival, *Journal of Air Transport Management*, 62, 90-98, 2017.

[23] P. Baumgarten, R. Malina, A. Lange, The Impact of Hubbing Concentration on Flight Delays within Airline Networks: An Empirical Analysis of the US Domestic Market. *Transportation Research Part E: Logistics and Transportation Review*, 66, 103-114, 2014.

[24] N. Pyrgiotis, K.M. Malone, A. Odoni, Modelling Delay Propagation within an Airport Network, Transportation Research Part C, Emerging Technologies, 27, 60-75, 2013

[25] P. Fleurquin, J.J. Ramasco, V.M. Eguiluz, Systemic Delay Propagation in the US Airport Network. *Nature Scientific Reports*, 3, 1-6, 2013.

[26] N. Xu, L. Sherry, K. Laskey, Multi-Factor Model for Predicting Delays at U.S. Airports. *Transportation Research Record: Journal of the Transportation Research Board*, 2052, 62-71, 2008.

[27] J.T. Wong, S.C. Tsai, A Survival Model for Flight Delay Propagation, *Journal of Air Transport Management*, 23, 5-11, 2012

[28] B. Schellekens, Stochastic Simulation of Delay Propagation: Improving Schedule Stability at Kenya Airways. *Master Thesis, Delft University of Technology*,2011**.**

[29] Internet: B. Marr, What Is Deep Learning AI? A Simple Guide With 8 Practical Examples, https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8- practical-examples/#3be4ab4c8d4b, 15.11.2021.

[30] K. Funahashi, Y. Nakamura, "Approximation of dynamical systems by continuous time recurrent neural networks", *Neural networks,* 6, 801–806, 1993.

[31] Internet: A. Amidi, Recurrent Neural Networks Cheatsheet, https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks, 15.11.2021.

[32] Y. Lecun, B.E. Boser, J.S. Denker, D. Henderson, R.E. Howard, W.E. Hubbard, LD. Jackel, "Handwritten digit recognition with a backpropagation network", *Advances in Neural Information Processing Systems,* 2, 396 –404, 1990.

[33] M. Valenti, S. Squartini, A. Diment, G. Parascandolo, T. Virtanen, "A convolutional neural network approach for acoustic scene classification", *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, Alaska, 2017.

[34] H. Eghbal-Zadeh, B. Lehner, M. Dorfer, G. Widmer, "A hybrid approach using binaural I-vectors and deep convolutional neural networks", **Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)**, Budapest, Hungary, 2016.

[35] Internet: A Review of Recurrent Neuarl Networks: LSTM Cells and Network Architectures, https://www.mitpressjournals.org/doi/full/10.1162/neco_a_01199, 15.11.2021.

[36] Internet: Rectified Linear Units (ReLU) in Deep Learning, https://www.kaggle.com/dansbecker/rectified-linear-units-ReLU-in-deep-learning, 15.11.2021.

[37] Internet: What is Deep Learning? https://machinelearningmastery.com/what-is-deep-learning/, 15.11.2021.

[38] K. Funahashi, Y. Nakamura, Approximation of dynamical systems by continuous time recurrent neural networks, *Neural networks*, 6, 801–806, 1993.

[39] A. Graves, A. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, Acoustics, **IEEE International Conference on. IEEE**, 6645–6649, 2013.

[40] R. Pascanu, C. Gulcehre, K. Cho, Y. Bengio, How to construct deep recurrent neural networks, *arXiv preprint arXiv*, 1312, 6026, 2013.

[41] S. Karsoliya, "Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture", *International Journal of Engineering Trends and Technology,* 6, 714–717, 2012.

It is the extended version of the paper "Deep Learning Based Forecasting of Cancellation, Delay and Orientation on Flights" accepted at the International Conference on Informatics and Computer Science 2021.