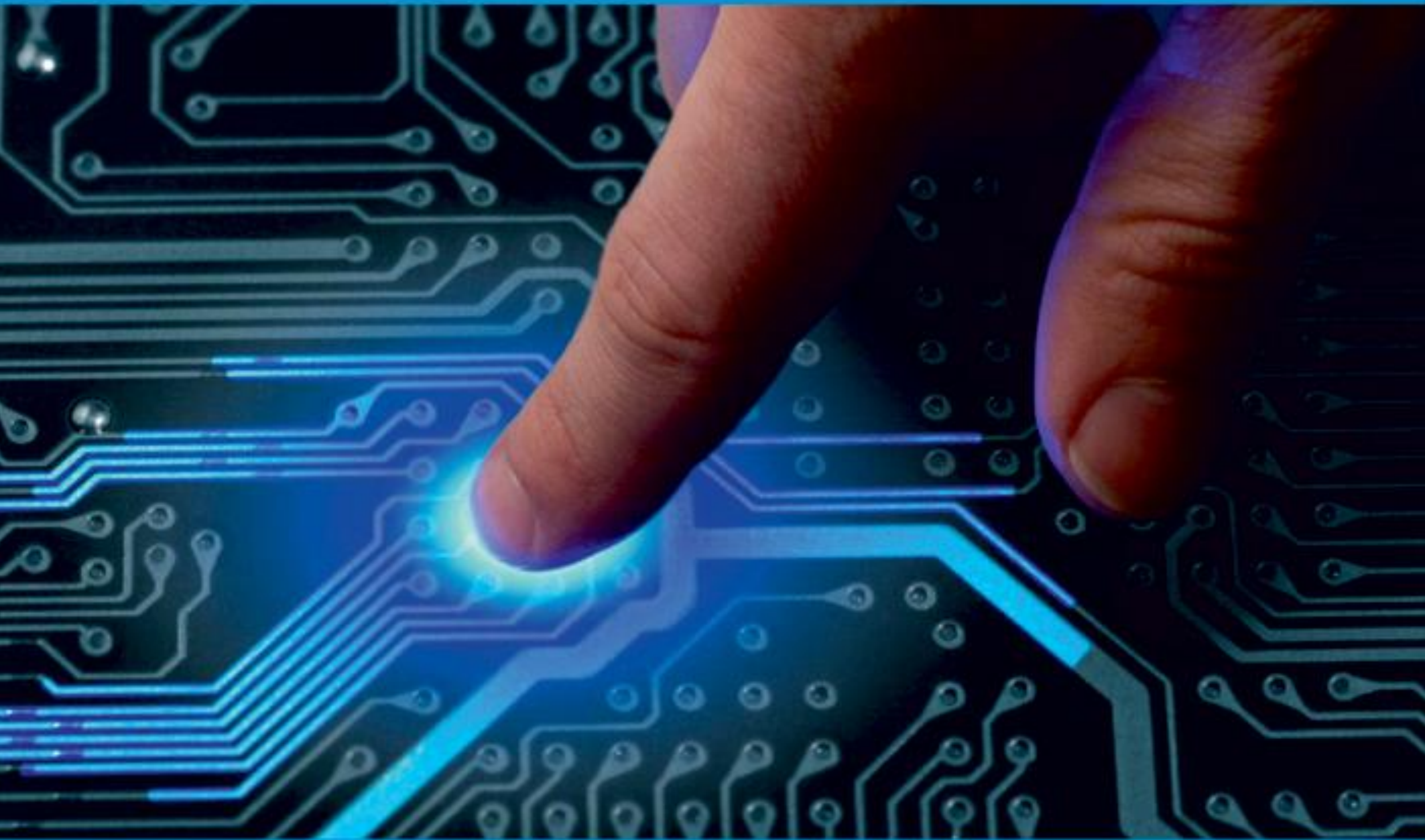
e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205245 |

Flight Delay Prediction Using Deep Learning

Dr. Archana P. Kale, Arpita Jagtap, Vicky Chavan, Bhavesh Kate, Kshitija Surywanshi

Department of Computer Engineering, MES Wadia College of Engineering, Pune, India

ABSTRACT: Deep learning techniques offer a powerful framework for flight delay prediction, leveraging their ability to extract complex patterns from diverse data sources, model temporal and spatial dependencies, and adapt to changing conditions over time. Predicting flight delays is challenging due to the myriad of factors involved, including weather conditions, runway availability, and air traffic management. Traditional models have struggled to comprehensively consider these factors, resulting in inaccuracies in delay predictions. To address this issue, the Random Deep Fusion (RDF) model is introduced, leveraging advanced deep learning techniques to revolutionize flight delay prediction. By combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Random Forests into a unified architecture, RDF achieves a new level of predictive accuracy. Notably, LSTM component achieves a remarkable training accuracy of 91.92%, marking a substantial improvement over previous models. With this significant enhancement, final RandomDeepFusion (RDF) Model achieves an accuracy of 88.9%. With CNN capturing spatial relationships, LSTM handling temporal dependencies, and Random Forests contributing to ensemble learning, RDF provides a comprehensive understanding of the complex patterns driving flight delays.

KEYWORDS: Flight Delay Prediction, Convolutional Neural Network (CNN), LSTM,Random Forest

I. INTRODUCTION

In the world of aviation, flight delays are a major problem. These delays have caused bad flight experiences for passengers and also resulted in financial losses of airline industry. According to research by the Air Transport Research Society, delays at airports cost the global economy an estimated 75.5 billion annually hence there is need of better management of these delays[1],[2].When transportations are interrupted, they give rise to delays, which create uncertainty regarding the expected arrival times for travellers. Flight delays not only disrupts the financial gains of airlines but also disrupts the passenger's travel plans. Particularly during holiday season when passenger volumes are high, the economic losses resulting from flight delays are even more significant. The world of aviation is a marvel of human ingenuity, allowing us to connect across vast distances. However, behind the scenes, it's a complicated process influenced by many things like weather, air traffic, and how airports work. Flight delays, even if they seem small, can cause big headaches. They mess up travel plans, test passengers' patience, and cost airlines and airports a lot of money. These delays cause problems not only for the people directly involved but also for the larger economy. In recent decades, the aviation industry has witnessed a growing interest in addressing the challenge of estimating and predicting flight delays. This critical issue has prompted numerous efforts to develop effective predictive models, leveraging the power of machine learning-based network architectures. Researchers have explored various approaches to comprehensively evaluate the performance of these architectures, encompassing a range of prediction tasks spanning both classification and regression domains.

One notable approach, introduced in [3][4] is the utilization of Random Forest methodology for flight delay prediction. This approach stands out by its integration of priority information derived from a wide array of weather and non-weather-related features. The various deep learning techniques like Multiscale Spatial-Temporal Adaptive Graph Convolutional Neural Network (MSTAGCN) is used to model the spatio-temporal features of airport networks, addressing the multi-airport scenario. It employs graph-structured inputs and incorporates a temporal convolutional block to capture the evolving dependencies within air traffic sequences[5]. [6] shows the use of ANN for handling nominal variables and enhance the relationship between input and output variables.

There are following challenges that cannot solve by single machine learning and deep learning technique : 1. Flight delays are influenced by a multitude of factors, including weather, airport operations, air traffic, and historical data. These factors create complex, multidimensional data that traditional models struggle to handle 2. The spatial layout of airports and their proximity to each other play a significant role in delays. Capturing these spatial dependencies effectively is a challenge. 3. Flight delays are influenced by time-related factors, such as the time of day, day of the week, and historical patterns. Understanding and modelling these temporal dependencies are essential for accurate predictions. The main contribution of our work is discussed as follows: The RandomDeepFusion model is designed to address these challenges by integrating advanced machine learning techniques. It combines Random Forest for

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 5, May 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1205245 |

robustness, CNN for spatial pattern recognition, and LSTM for capturing temporal dependencies. This fusion allows us to comprehensively analyse multidimensional data, consider spatial and temporal aspects, and make accurate predictions, ultimately contributing to more efficient airline operations and improved passenger experiences.

II. RELATED WORK

The authors tap into a wide spectrum of elements that might influence flight delays, adopting a more holistic strategy. Their work involves an in-depth comparison of various machine learning-cantered models, all ingeniously crafted to predict flight delays. This prediction activity comprises of several classification along with a regression task[3]. A model specifically designed for flight delays is con- structed using the random forest methodology. They develop an analytical method built on a clustering algorithm. The Random Forest framework encapsulates a probability sampling approach[4] The writer has formulated a method to predict flight delays leveraging the graph convolutional neural network technology. This deep learning structure inherently uses graph-shaped inputs and has been titled the Multiscale Spatial-Temporal Adaptive Graph Convolutional Neural Network (MSTAGCN). The program is adept at managing flight delay prediction complications. Flight delays have stunted the growth of civil aviation on a worldwide scale in recent times. Also, delay dispersion significantly contributes to flight interferences, commonplace at overly busy airports or those running close to their highest capacity[5].

The writer utilized the artificial neural network (ANN) approach, an instrument that holds potential for forecasting air travel delays. Not only is it able to handle nominal variables, it also provides meaningful insights. This suggested technique was put to the test for forecasting delays of incoming flights at JFK airport[6]. The writer suggests implementing two methods for predicting flight delays using deep convolution neural network that merges meteorological information. One method is called DCNN (Dual-channel Convolutional Neural Network), which is associated with the ResNet network structure. The second method is referred to as SE-Dense Net (Squeeze and Excitation Densely Connected Convolutional Network), which mixes the strengths of DenseNet and SENet.
These two deep convolutional neural networks, mentioned in this document, can potentially enhance prediction accuracies. The primary objective of this proposed study is to explore flight delays. Among all, the¬ random forest model showed top performance for predicting departure delay [7].The writer suggested a system for forecasting air travel delays by employing the random forest method and route- detection techniques. This system pinpoints the fastest air routes from one¬ specified location to another, taking advantage of the data offered by the user and making the most of complimentary or public APIs[8].

The writer introduces two techniques, namely random forest and logistic regression. These strategies are applied in a trifold manner. Initially, they formulate SP and ML algorithms. Following that, the generated results from step one is utilized to educate various supervised ML models. The last phase involves harnessing these insightful ML models to forecast existing delays[9]. The primary aim of the suggested project is to scrutinize flight holdups. The random forest has been identified as the most effective model to assess flight departures' lateness. In [8], the author introduced a scheme to forecast flight delays merging random forest coupled with a path-finding algorithm. This model directs towards the swiftest flights from one point to another, capitalizing on data provided by the client and harnessing freely available or public Application Programming Interfaces (APIs). In[10], the author has designed a blended approach that combines Random forest regression with the maximal information coefficient. This suggested RFR-MIC model displays commendable performance when measured against linear regression (LR), knearest neighbours (KNN), artificial neural network (ANN), and conventional Random Forest Regression. In[11] the writer has created a fresh graph-to-sequence learning structure incorporating the attention model for projecting multiple steps ahead of hourly leave and join delays throughout the full system. The AG2S-Net includes three critical elements: a graph convolutional neural network, a reciprocal LSTM neural network, and a sequence-to-sequence scheme which embraces an integral attention mechanism.

A framework was put forward in [12], grounded in deep learning theory. The scholars have introduced a method that hinges on stack denoising autoencoder, incorporated into the suggested model. Additionally, they utilized the Levenberg Marquart algorithm. Demonstrating remarkable potential, the proposed methodology has effectively navigated the complexities of large datasets and successfully identified critical factors influencing delays[13]. The researchers implemented Random Forest Regression and Maximal Information Coefficient, also known as RFR-MIC. They suggested that the RFR-MIC model delivers a more satisfactory performance when juxtaposed with traditional linear regression (LR), k-nearest neighbours (KNN), artificial neural network (ANN), and standard Random Forest Regression (RFR)[14]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205245 |

We've introduced a model [15], hailing as a part of the supervised machine learning concept of Naive Bayes Delays; this is a method focused on constructing a system. Departures are prone to shifting because of adverse weather scenarios, peak demands during holidays or seasons, the dictates of the airline companies, technical snags touching aspects like airport facilities, baggage management, mechanical apparatus, and the spillover of delays from earlier flights. The creators introduced their framework which encompasses three stages: (a) preparation of the data; (b) forecasting using a combined Convolutional Neural Network (c) The final step involves contrasting the framework's effectiveness with the proposed model through a review of the trial findings[16]. Researchers often put a great amount of energy into fine-tuning their models, aiming to achieve precise outcomes, but sometimes, they miss an important phase: comprehending the data and improving specific aspects, especially in a data mining project [17].In [18], these same researchers proposed a method based on visual elements. This approach consists of utilization of a snapshot from the airport's situational awareness map, encom- passing diverse aircraft trajectories and contextual attributes like weather conditions and airline schedules. They introduce an end-to-end deep learning framework known as TrajCNN, designed to extract and process both spatial and temporal information from the situational.

One remarkable research [19] on deep learning techniques presented an inclusive system, termed as Attention-based Bidirectional Long Short-Term Memory' (ATTBI-LSTM),specifically designed to predict flight delays. The Bidirectional LSTM model is powerful in capturing both spatial and temporal aspects of the flight network, along with incorporating weather conditions. The writer constructs an innovative geospatially weighted recurrent neural network model with the aim to improve flight time forecasts. This model envisions nationwide air traffic information garnered from an array of sources. These sources comprise of Automatic Dependent Surveillance - Broadcast (ADSB), Meteorological Aerodrome Reports (METAR), and records from airlines. [20]. The writers have put forward an inventive, structured system using machine learning for the prediction of delay times and durations of flights in sequence. They've incorporated different machine learning strategies blended with varying sampling methods [21].

III. PROPOSED METHODOLOGY

The challenges in the airline industry have prompted the integration of advanced machine learning techniques to address them effectively. One notable approach involves combining Random Forest for robustness, CNN for spatial pattern recognition, and LSTM for capturing temporal dependencies.

This fusion enables comprehensive analysis of multidimensional data, taking into account both spatial and temporal aspects. The outcome is more improved predictions, fostering more efficient airline operations and enhancing passenger experiences.

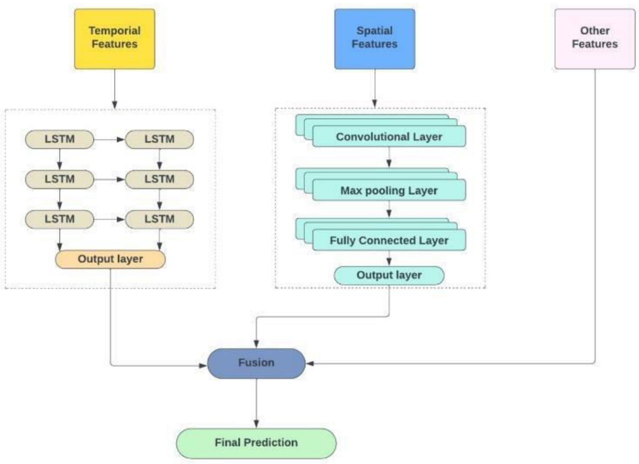


Fig. 1: *Proposed Methodology [RDF] Data Flow*

A. Random Deep Fusion Model

1. Convolutional Neural Network(CNN): The CNN component of the Random Deep Fusion model would continuously analyse spatial data, such as weather patterns and airport congestion, in real-time. By extracting spatial features from ongoing data streams, the CNN can identify spatial patterns and relationships as they unfold, allowing for timely

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205245 |

detection of factors that may contribute to flight delays. The CNN's ability to capture spatial patterns and relationships is crucial in understanding the spatial context of flight delay data. For instance, it can detect localized weather phenomena, such as storms or fog, that may affect airport operations. By continuously analysing spatial data, the CNN provides realtime insights into the spatial distribution of relevant factors, enabling proactive decision-making to mitigate potential delays. When combined with temporal insights from the LSTM and external features, the spatial features extracted by the CNN contribute to a comprehensive representation of the input data. This holistic approach allows the model to capture complex interactions between spatial, temporal, and external factors, resulting in more accurate and robust predictions of flight delays in real-time.

2. Long Short-Term Memory (LSTM): The LSTM component continuously analyses time-series data representing historical trends and patterns related to flight delays[fig.1]. By capturing long-term dependencies in sequential data, the LSTM can identify subtle temporal patterns and trends that may indicate the likelihood of future delays. This real-time analysis enables the model to adapt and respond quickly to changing conditions, enhancing its predictive capabilities. The LSTM's ability to capture temporal dependencies is crucial for understanding the temporal dynamics of flight delay data. For instance, it can identify recurring patterns in flight schedules, such as peak travel times or seasonal variations, which may impact delay probabilities. By continuously analysing timeseries data, the LSTM provides real-time insights into the evolving temporal context of flight operations, allowing for proactive decision-making and resource allocation. When combined with spatial features from the CNN and external factors, the temporal insights provided by the LSTM contribute to a comprehensive understanding of the factors influencing flight delays. By integrating temporal, spatial, and external features, the Random Deep Fusion model gains a holistic view of airport operations, enabling it to make more accurate and timely predictions of delay probabilities.

3. RandomDeepFusion model(RDF) : The Random Deep Fusion model operates by integrating spatial features from CNNs, temporal insights from LSTMs, and external factors such as weather conditions and flight schedules[fig.1.] Continuously analysing incoming data streams, the model dynamically adapts to capture evolving spatial and temporal patterns, providing timely and accurate predictions of flight delays. By combining these diverse sources of information, the model enables proactive decision-making to mitigate delays and optimize airport operations, ultimately enhancing passenger satisfaction and operational efficiency.

B. Dataset

The historical flight data provides by the Bureau of Transport Statistics consist mainly 5,00,000 flights per month which makes it 60,00,000 annually(https://www.bts.gov/).The weather data on airports is collected from National Climate Data Centre(https://www.ncdc.gov),which includes features like temperature, wind speed, humidity, etc

IV. EXPERIMENTAL RESULTS

In this paper, The Random Deep Fusion model integrates CNN for spatial feature extraction, LSTM for temporal feature extraction, and a Random Forest classifier for predicting flight delays.

A. Performance of LSTM model

Case	Accuracy	Loss
RDF before feature selection	63	-
Feature selection and sequence length=24	72.92	0.5807
Applied early stopping	82.05	0.4013
Batch Normalization and added dropout layer	91.91	0.200

TABLE I COMPARISON FOR LSTM MODEL

The TABLE I presents a comparative analysis of the performance of an LSTM (Long Short-Term Memory) model across different configurations aimed at enhancing its effectiveness in a specific task, likely classification or prediction. The first scenario, labeled "RDF before feature selection," represents the baseline model trained without any feature selection techniques applied beforehand. Despite achieving a moderate accuracy of 63%, no loss value is provided. In

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205245 |

contrast, the subsequent case, "Feature selection and sequence length=24," involved implementing feature selection methods prior to training, alongside setting the sequence length to 24. This adjustment led to a notable improvement in accuracy, reaching 72.92%, with a corresponding loss of 0.5807. Further refinement was achieved by employing early stopping, a technique aimed at preventing overfitting by halting training when performance on a validation dataset deteriorates. This intervention boosted accuracy to 82.05% and reduced the loss to 0.4013 compared to the previous configuration. Moreover, additional enhancements, such as batch normalization and the inclusion of a dropout layer, were introduced in the final scenario. These adjustments yielded a significant improvement in accuracy, achieving a high 91.91%, with a notably reduced loss of 0.200. Collectively, the table underscores the iterative process of model optimization, demonstrating how each adjustment contributes to enhancing the LSTM model's performance, as indicated by accuracy improvements and loss reductions across various configurations.

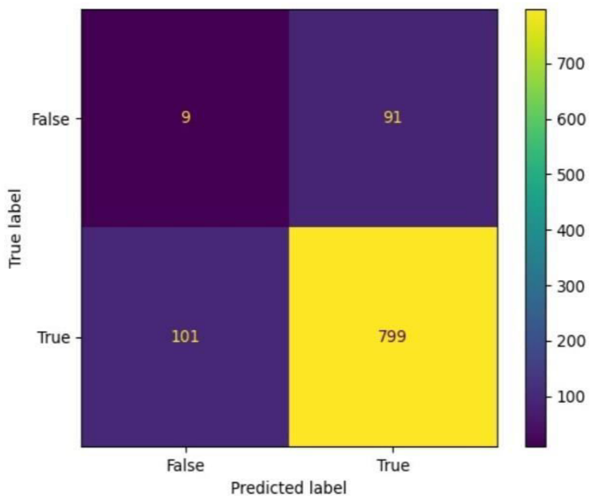


Fig. 2 *Confusion matrix of the RDF Model*

Model	Precision	Recall	F-Score	Accuracy
RDF	89.7	88.8	89.3	88.9

TABLE II TESTING ACCURACY FOR RDF MODEL

The table II presents the testing accuracy metrics for a Random Forest (RDF) model employed in the context of flight delay prediction. These metrics include Precision, Recall, F-Score, and Accuracy, each providing distinct insights into the model's performance. Precision, measuring the accuracy of positive predictions, achieved a rate of 89.7%, indicating that nearly 89.7% of the flights identified as delayed by the RDF model were indeed delayed. Meanwhile, Recall, representing the model's ability to correctly capture all relevant instances of delay, stood at 88.8%, suggesting that around 88.8% of all actual delayed flights were successfully identified. The F-Score, a harmonized measure of Precision and Recall, reached 89.3%, striking a balance between the two metrics. Finally, the model's overall Accuracy, reflecting the correctness of its predictions across both delayed and non-delayed flights, reached 88.9%, indicating that approximately 88.9% of the RDF model's predictions were accurate. Collectively, these metrics provide a comprehensive assessment of the RDF model's performance, showcasing its precision, recall, balanced FScore, and overall accuracy in predicting flight delays.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a comprehensive solution for improving flight delay prediction using deep learning techniques, incorporating both weather and non-weather features. Our proposed model Random Deep Fusion combines CNN, LSTM, and Random Forest algorithms to leverage the strengths of each component. Through the utilization of LSTM, we achieved a remarkable accuracy of each component. Through the utilization of LSTM, we achieved a remarkable accuracy of 91.92 % starting the effectiveness of LSTM in capturing temporal dependencies within the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205245 |

data. Subsequently, by integrating Random Forest into the RandomDeepFusion architecture, we attained a final accuracy of 88.9%, further enhancing predictive performance. The integration of CNN and LSTM in the RandomDeepFusion model enables effective feature extraction from sequential data, allowing for comprehensive analysis of weather and non- weather features. This combined approach harnesses the power of deep learning to provide accurate and reliable flight delay predictions. Our results highlight the efficacy of the RandomDeepFusion model in handling the complexities of flight delay prediction tasks, showcasing superior performance com- pared to traditional methods. By incorporating both weather and non-weather features, our approach provides a holistic understanding of the factors influencing flight delays.

In future, expanding RandomDeepFusion[RDF] to include real-time data and global coverage can improve flight delay prediction. This means making the model adaptable to changing conditions, ensuring it can handle large-scale data efficiently, and making its predictions easy to understand. By continuously improving based on real-world experience, the model can help make air travel more reliable and satisfying for passengers everywhere.

REFERENCES

1. Federal Aviation Administration Benefit Cost Analysis. https://www.faa.gov/ regulations policies/policy guidance/benefit cost/ Air Travel Consumer Report,U.S. Dept. Transp,Washington, DC, USA, 2012.
2. Guan Gui,Fan Liu, Daly,Jinlong Sun,Jie Yang ,Ziqi Zhou and Dongxu Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning", Transportation Research Part E: Logistics and Transportation Review,2019
3. Qiang Li , Ranzhe Jing, and Zhijie Sasha Dong, "Flight Delay Prediction With Priority Information of Weather and Non-Weather Features", IEEE Transactions on Intelligent Transportation Systems,2023.
4. Kaiquan Cai ,Yue Li , Yi-Ping Fang and Yanbo Zhu, "A Deep Learning Approach for Flight Delay Prediction Through TimeEvolving Graphs", IEEE Transactions on Intelligent Transportation Systems,2022.
5. Sina Khanmohammadi,Salih Tutun,Yunus Kucuk, "A New Multilevel Input Layer Artificial Neural Network for Predicting Flight Delays at JFK Airport", Procedia Computer Science,2016.
6. Jingyi Qu,Ting Zhao, Meng Ye,Jiayi Li,Chao Liu, "Flight Delay Pre- diction Using Deep Convolutional Neural Network Based on Fusion of Meteorological Data", ringer Science+Business Media, LLC, part of Springer Nature,2020.
7. Ravi Kothari , Riya Kakkar , Smita Agrawa,Parita Oza, Sudeep Tan- war,Bharat Jayaswal, Ravi Sharma , Gulshan Sharma and Pitshou Bokoro, of Best Machine Learning Model to Predict Delay in Passenger Airlines", Digital Object Identifier 10.1109/ACCESS.2023.3298979,2023
8. V´ıctor M. Tenorioa, Antonio G. Marquesa , Luis Cadarsob, "Signal Processing and Machine Learning for Air Traffic Delay Prediction", Research Procedia 58 (2021) 463–47,2021.
9. Zhen Guo, Bin Yu, Mengyan Hao, Wensi Wang, Yu Jian, Fang Zong"A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient", Aerospace Science and Technology,2021.
10. Jie Bao, Zhao Yang, Weili Zeng, "Graph to sequence learning with attention mechanism for network-wide multi-step-ahead flight delay prediction", Transportation Research Part C,2019.
11. Maryam Farshchian Yazdi , Seyed Reza Kamel , Seyyed Javad Mahdavi Chabok and Maryam Kheirabadi "Flight delay prediction based on deep learning and Levenberg Marquart algorithm", Yazdi et al. J Big Data (2020) 7:106.
12. Bin Yu, Zhen Guo ,Sobhan Asian, Huaizhu Wang, ,Gang Chen "," Flight delay prediction for commercial air transport: A deep learning approach", Transportation Research Part E,2019.
13. Jaehyun Yoo and Karl H. Johansson, "Learning communication delay patterns for remotely controlled UAV networks", ScienceDirect ,2017.
14. Mrs Yogita Borse, Dhruvin Jain ,Shreyash Shrama ,Viral vora, "Flight Delay Prediction System", International Journal of Engineering Research Technology C,2020.
15. Hesam Shafienya and Amelia C.Regan, "4D flight trajectory prediction using a hybrid Deep Learning prediction method based on ADS-B technology: A case study of Hartsfield–Jackson Atlanta International Airport (ATL), Transportation Research Part C,2022.
16. Nuno Fernandes,Sergio Moro,Carlos J.Costa,Manuela Aparicio, "Factors influencing charter flight departure delay", Research in Transportation Business Management.
17. Wei Shao, Arian Prabowo, Sichen Zhao ,Piotr Koniusz, Flora D. Salim, "Predicting flight delay with spatio-temporal trajectory convolutional network and airport situational awareness map", Neurocomputing,2022.
18. Maged Mamdouh,Mostafa Ezzat,Hesham Hefny, "Improving flight delays prediction by developing attention-based bidirectional LSTM network", Expert Systems With Applications,2024.
19. Xinting Zhu and Lishuai Li,"Flight time prediction for fuel loading decisions with a deep learning approach", Transportation Research Part C,2021.
20. Waqar Ahmed Khan,Hoi-Lam Ma,Sai-Ho Chung,Xin Wen, "Hierarchical integrated machine learning model for predicting flight departure delays and duration in series", Transportation Research Part C,2021.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details