www.ijbar.org
ISSN2249-3352(P)2278-0505(E)
CosmosImpactFactor-5.86

Enhancing Passenger Demand Prediction in Public Transport: Addressing Data Imbalance with DC-GAN and Deep Learning

A.Anitha[1*], Pathulothu Navinder[2], Akhilkumar[2], Kodavandla Lokesh[2], Mysa Sai Eshwar[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Sciences and Engineering,

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana.

[*]Corresponding author: A.Anitha (aanitha109@gmail.com)

ABSTRACT

Predicting passenger demand in public transport is critical for efficient operation and planning. Using advanced deep learning techniques like DC-GAN (Deep Convolutional Generative Adversarial Networks) enhances accuracy by overcoming data imbalances, leading to optimized transportation services. The objective of this study is to develop an AI-based model using DC-GAN and deep learning to predict passenger demand in public transport more accurately, especially in cases of imbalanced data, leading to more efficient resource allocation and service planning. Traditionally, public transport demand prediction relied on statistical methods like regression models, historical data analysis, and human experience-based scheduling. These methods had limited accuracy and failed to adapt to dynamic and complex demand patterns. The traditional systems for passenger demand prediction in public transport faced challenges due to the dynamic nature of urban populations, peak-hour fluctuations, and lack of real-time adaptability. These methods could not handle the inherent data imbalance, where demand patterns were underrepresented for certain regions or time periods, leading to inefficiencies in resource allocation. The growing urban population and increased dependence on public transport have led to congestion and inefficiency. For instance, India's public transport usage is high, but demand prediction inaccuracies cause over-crowded buses in some areas and under-utilized ones in others. AI models like DC-GAN, which can balance data representation, offer the potential to improve operational efficiency, reduce wait times, and optimize scheduling based on more accurate demand forecasts. The proposed system integrates a deep learning model with DC-GAN to predict passenger demand by generating synthetic data for underrepresented routes or times. The model uses existing transport data (e.g., from smart cards, GPS, or ticketing systems) and addresses the imbalance issue by generating realistic samples to improve prediction accuracy. By enhancing demand forecasting, AI models can optimize schedules, reduce waiting times, and ensure a better allocation of public transport resources, leading to increased operational efficiency and improved passenger satisfaction.

Keywords: Public Transport, Deep Learning, DC-GAN (Deep Convolutional Generative Adversarial Network), Data Imbalance Handling, Operational Efficiency.

1. INTRODUCTION

India, public transport systems such as buses, trains, and metro networks serve as the backbone of urban commuting, especially in cities like Delhi, Mumbai, Bangalore, and Kolkata. The rising urban population and heavy reliance on public transport have led to increasing congestion and inefficiency in these systems. For example, according to the 2019-20 statistics, the Delhi Metro alone handled over 1.8 million passengers per day, while buses in Mumbai ferried over 2.6 million passengers daily. However, demand fluctuations during peak hours and underutilized routes during off-peak times have created operational challenges. Traditional prediction models often fail to capture these dynamic demand patterns, leading to either over-supply or under-supply of transportation services. Addressing this demand imbalance is crucial for optimizing resource allocation and improving commuter experiences. Before the advent of machine learning, passenger demand prediction in public transport systems was predominantly reliant on traditional methods such as historical data analysis, simple statistical models (like regression), and manual forecasting. These methods had several limitations: they could not adapt to real-time demand variations or seasonal fluctuations, lacked predictive power for unbalanced data (i.e., some routes being overcrowded while others remained empty), and were unable to capture complex patterns in demand, especially during unexpected surges or declines in commuter volume. These limitations led to inefficient scheduling, overcrowding, longer waiting times, and increased operational costs for transport providers.

2. LITERATURE SURVEY

Luo et al. [1] introduced a multitask deep learning approach for fine-grained service-level passenger flow prediction in bus transit systems. Their model simultaneously forecasts passenger flows across multiple routes and time intervals, capturing intricate temporal patterns and inter-route dependencies. This comprehensive approach enhances the accuracy of predictions, aiding in efficient transit planning and operations. Liu et al. (2021) [2] proposed an automatic feature engineering method for bus passenger flow prediction using a modular convolutional neural network. Their technique automates the extraction of relevant features from raw data, reducing manual intervention and improving prediction accuracy. The modular design allows for flexibility and scalability, accommodating various data sources and transit scenarios. Lv et al. (2022) [3] developed a bus passenger flow prediction model that integrates point-of-interest (POI) data with extreme gradient boosting (XGBoost). By incorporating POI data, the model accounts for the influence of nearby attractions and facilities on passenger flow, leading to more precise predictions. The use of XGBoost enhances the model's performance by efficiently handling large datasets and capturing complex relationships. In 2023, V. G. S. and H. V. S. [4] explored the use of Gaussian Process Regression for predicting bus passenger traffic. Their model provides a probabilistic framework, offering not only predictions but also uncertainty estimates. This approach is particularly useful for understanding the variability in passenger flow and making informed decisions under uncertainty. Cheng and He (2022) [5] analyzed bus travel characteristics and predicted elderly passenger flow using smart card data. Their study focused on understanding the travel patterns of elderly passengers, which is crucial for designing age-friendly transit services. By leveraging smart card data, they identified peak travel times and preferred routes among elderly users.

Luo et al. (2019) [6] proposed a new framework for intelligent public transportation systems based on the Internet of Things (IoT). Their framework integrates various IoT devices to collect real-time data on vehicle locations, passenger counts, and environmental conditions. This data is then used to optimize routes, schedules, and resource allocation, enhancing the overall efficiency of public transportation systems. Vigren and Pyddoke (2020) [7] investigated the impact of passenger incentive contracts on bus ridership in public transport. Their analysis revealed that such contracts, which reward operators based on passenger numbers, can lead to increased ridership. However, the effectiveness of these incentives depends on various factors, including service quality and fare levels. Wang et al. (2020) [8] developed a network-based model to analyze passenger transfer flow between bus and metro systems in Beijing. Their model considers the connectivity and transfer times between different modes of transport, providing insights into passenger behavior and system efficiency. The findings can inform the design of integrated transit networks that facilitate seamless transfers.

Zhou et al. (2020) [9] conducted a comparative analysis of collision and non-collision injuries among public bus passengers. Their study identified factors contributing to injury severity, such as passenger standing position and bus maneuvering. The results highlight the need for safety measures, including improved vehicle design and passenger education, to reduce injury risks. Dhital et al. (2022) [10] examined the effects of the COVID-19 pandemic on public bus occupancy and tailpipe emissions per passenger kilometer traveled. They found that reduced passenger numbers during the pandemic led to higher emissions per passenger, underscoring the environmental benefits of maintaining higher occupancy levels in public transport. Burdzik et al. (2023) [11] studied passenger flow models and simulations concerning COVID-19 spread at public transport bus stops. Their research aimed to understand how passenger interactions at bus stops contribute to virus transmission. The findings suggest that measures such as social distancing and improved ventilation can mitigate the risk of infection. Liyanage et al. (2022) [12] developed AI-based neural network models for bus passenger demand forecasting using smart card data. Their models leverage historical boarding data to predict future demand, assisting transit agencies in resource planning and service adjustments. The use of neural networks allows for capturing complex patterns in passenger behavior. Nayak et al. (2022) [13] conducted a comprehensive comparative analysis of passenger demand prediction methods to improve urban bus transportation systems. They evaluated various predictive models, including traditional statistical methods and modern machine learning approaches, to determine their effectiveness in forecasting passenger demand. Their findings provide valuable insights for selecting appropriate models based on specific transit system characteristics.

3. PROPOSED SYSTEM

The proposed system aims to enhance the efficiency and reliability of bus transit operations by accurately predicting passenger flow using advanced deep learning techniques. Accurate passenger flow predictions enable transit agencies to optimize resource allocation, adjust bus frequencies, and improve route planning, ultimately leading to better service quality and increased passenger satisfaction.

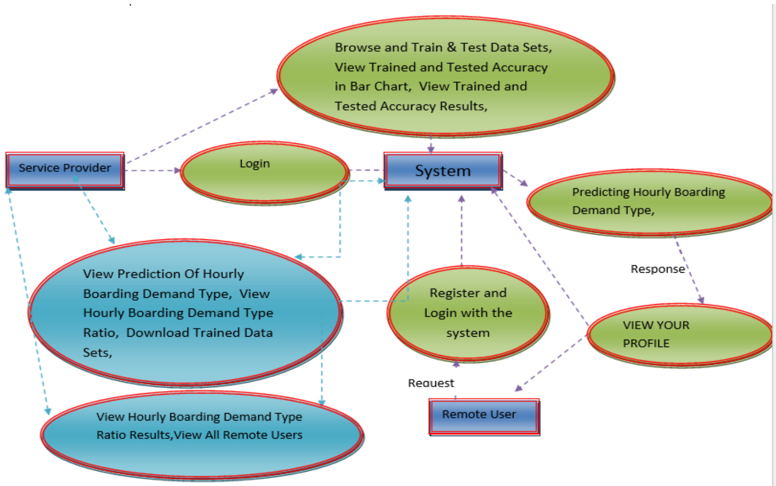

Fig. 1: Block Diagram

Step 1: Bus Passenger Dataset

The foundation of the proposed system is the collection of a comprehensive bus passenger dataset. This dataset encompasses various attributes such as passenger counts, timestamps, route identifiers, boarding and alighting locations, and possibly demographic information. Data sources include automated fare collection systems, GPS tracking, and manual surveys. A rich and diverse dataset ensures that the model captures the multifaceted nature of passenger demand across different times and routes.

Step 2: Dataset Preprocessing

Once the data is collected, preprocessing is essential to ensure its quality and suitability for modeling. This involves handling missing values through imputation or removal, encoding categorical variables using techniques like label encoding, and normalizing numerical features to a standard scale. Additionally, data is segmented into training and testing sets to facilitate model evaluation. Effective preprocessing enhances the model's ability to learn meaningful patterns and reduces the risk of biases.

Step 3: Existing Algorithm (Random Forest Classifier)

As a benchmark, the Random Forest Classifier is employed to predict passenger demand. This ensemble learning method constructs multiple decision trees during training and outputs the mode of their predictions for classification tasks. By averaging the results of various trees, it aims to improve predictive accuracy and control overfitting. Evaluating the performance of this traditional algorithm provides a baseline against which the proposed advanced model can be compared.

Step 4: Proposed Algorithm (Deep Convolutional Generative Adversarial Network)

To address data imbalance and enhance prediction accuracy, a Deep Convolutional Generative Adversarial Network (DC-GAN) is proposed. DC-GAN consists of two neural networks—the generator and the discriminator—that contest with each other. The generator creates synthetic data samples, while the discriminator evaluates them against real data. Through this adversarial process, the generator learns

www.ijbar.org
ISSN2249-3352(P)2278-0505(E)
CosmosImpactFactor-5.86

to produce realistic data, augmenting the dataset and mitigating imbalance issues. Integrating DC-GAN with a predictive model aims to improve demand forecasting, especially for underrepresented routes or times.

Step 5: Performance Comparison

The final step involves a comprehensive performance comparison between the Random Forest Classifier and the DC-GAN-enhanced model. Metrics such as accuracy, precision, recall, and F1-score are utilized to assess and contrast the models' predictive capabilities. This evaluation determines the effectiveness of the proposed system in accurately forecasting passenger demand and informs potential areas for further refinement.

3.1 Model Building

Building the machine learning model involves selecting appropriate algorithms, training them on the preprocessed data, and fine-tuning to achieve optimal performance. The process begins with the selection of a baseline model, such as the Random Forest Classifier, to establish a performance benchmark. The model is trained using the training dataset, during which it learns patterns and relationships within the data. Hyperparameters are adjusted based on validation set performance to prevent overfitting and enhance generalization. Subsequently, the proposed DC-GAN model is developed to address data imbalance issues. The generator network creates synthetic samples to augment the training data, while the discriminator network evaluates their authenticity. This adversarial training continues until the generator produces high-quality synthetic data indistinguishable from real data. The enriched dataset is then used to train a predictive model, aiming to improve demand forecasting accuracy. Finally, both models are evaluated on the testing set, and their performances are compared to determine the effectiveness of the proposed approach.

A Deep Convolutional Generative Adversarial Network (DCGAN) is an advanced type of Generative Adversarial Network (GAN) that integrates deep convolutional neural networks into its architecture. Introduced by Radford et al. in 2015, DCGANs are designed to generate high-quality synthetic data, particularly images, by leveraging the strengths of convolutional layers.

How It Works:

DCGANs consist of two primary components:

1. **Generator:** This network takes a random noise vector as input and transforms it into a synthetic image through a series of convolutional and upsampling layers.
2. **Discriminator:** This network evaluates images to distinguish between real images from the training dataset and synthetic images produced by the generator.

During training, the generator and discriminator engage in a competitive process. The generator aims to produce images that are increasingly realistic, while the discriminator strives to improve its ability to differentiate between real and generated images. This adversarial training continues until the generator creates images that the discriminator can no longer reliably distinguish from real ones.

4.RESULTS AND DISCUSSION

The dataset boarding demand is structured to facilitate the analysis and forecasting of bus passenger boarding patterns. It comprises several tables, each serving a distinct purpose in the data architecture. Below is an overview of the key tables and their roles:

1. auth_group: This table defines user groups within the system.
2. auth_group_permissions: This table associates permissions with user groups, specifying the actions that groups are authorized to perform.
3. auth_permission: This table lists all available permissions that can be assigned to users or groups.
4. auth_user: This table contains information about the users registered in the system, including credentials and personal details.
5. auth_user_groups: This table establishes the relationship between users and groups, indicating group memberships.
6. auth_user_user_permissions: This table links individual users to specific permissions, detailing the actions they are allowed to perform.
7. django_admin_log: This table records administrative actions performed within the system, providing an audit trail of changes and activities.
8. django_content_type: This table holds information about the models installed in the system, facilitating content type management.
9. django_migrations: This table tracks the migrations applied to the database, ensuring consistency and version control.
10. django_session: This table stores session data for users, enabling session management and persistence.
11. remote_user_clientregister_model: This table captures the registration details of clients, including personal information such as username, email, password, phone number, country, state, city, address, and gender.
12. remote_user_detection_accuracy: This table records the accuracy metrics of various detection models used in predicting boarding demand.
13. remote_user_detection_ratio: This table contains data on the detection ratios, possibly indicating the proportion of different demand levels (e.g., 'Less Demand' vs. 'More Demand').
14. remote_user_prediction_bus_boarding: This table holds the core data relevant to the project's objective of predicting bus boarding demand. It includes fields such as:

4.3 Result and Description

The homepage of the Demand of bus, accessible at the local address http://127.0.0.1:8000, centers around a research project titled "Predicting Hourly Boarding Demand of Bus Passengers Using Imbalanced

Records From Smart Cards: A Deep Learning Approach." The page features navigation links for "Home," "Remote User," and "Service Provider," suggesting different user roles and access levels within the system.

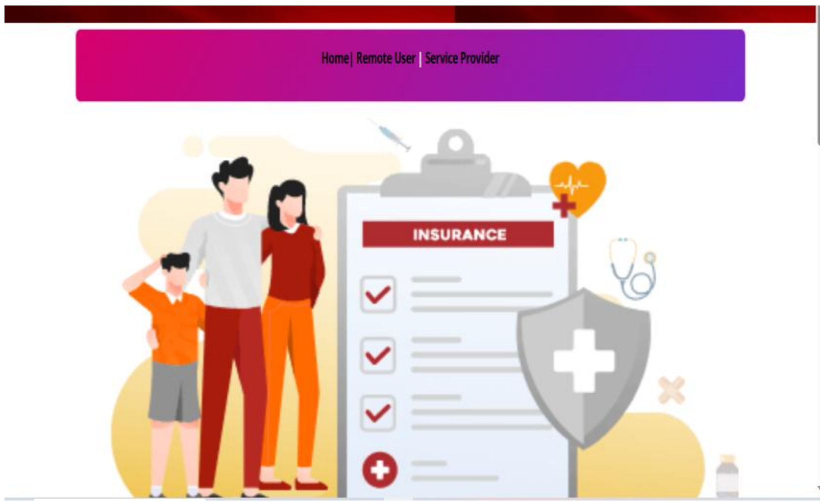

Figure 2 Home page of the bus demand.

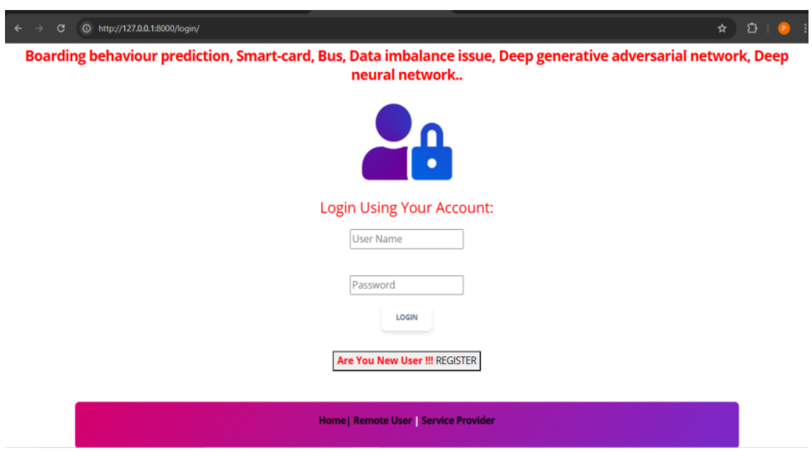

Figure 3 Remote user login screen.

The Figure shows the Login page with the below:

- Username Field: An input field for the user to enter their username.
- Password Field: An input field for the user to enter their password.
- Login Button: A button labeled "LOGIN" to submit the credentials.
- Registration Link: A link below the button that says "Are You New User..!!! REGISTER," prompting new users to create an account.

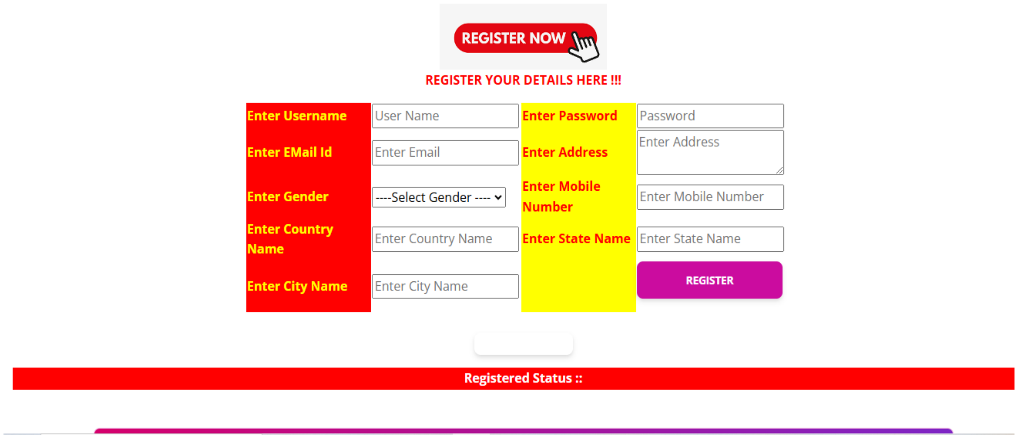

Figure 4 Remote User Signup Screen

The Figure shows a user registration form.

Left Column (Personal Information):

- Enter Username: Field for username input.
- Enter EMail Id: Field for email address.
- Enter Gender: Dropdown menu to select gender.
- Enter Country Name: Field for country.
- Enter City Name: Field for city.

Right Column (Contact and Address Information):

- Enter Password: Field for password input.
- Enter Address: Field for address input.
- Enter Mobile Number: Field for mobile number.
- Enter State Name: Field for state.

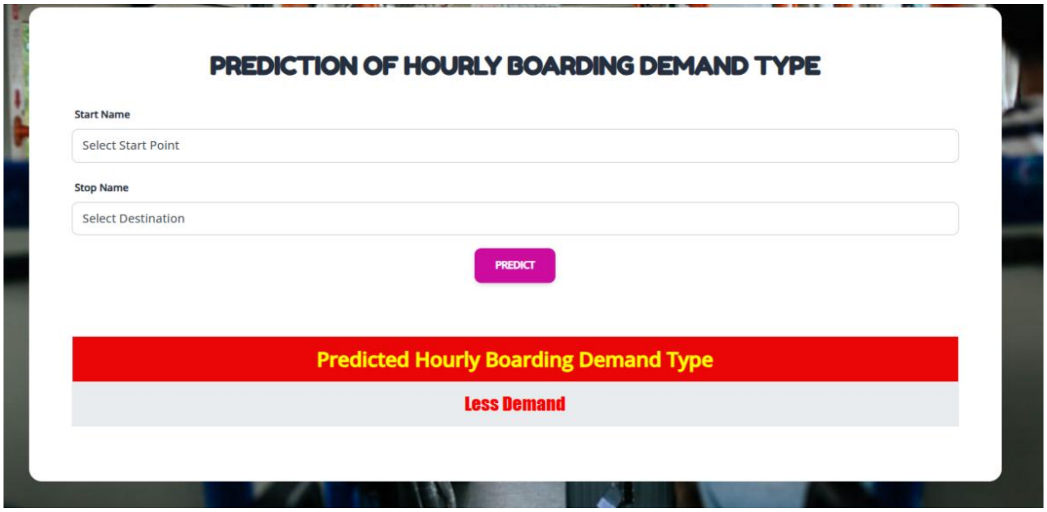

Figure 5 Predicted Outcome of the Boarding Demand

This above page provides a user-friendly interface to input bus-related data and receive a prediction about the boarding demand type, potentially using a model. The output "Less Demand" indicates the model's assessment based on the provided inputs.

Input Fields: The user is prompted to enter the following information:

- StopName
- StartName

Prediction Button: A "Predict" button initiates the prediction process

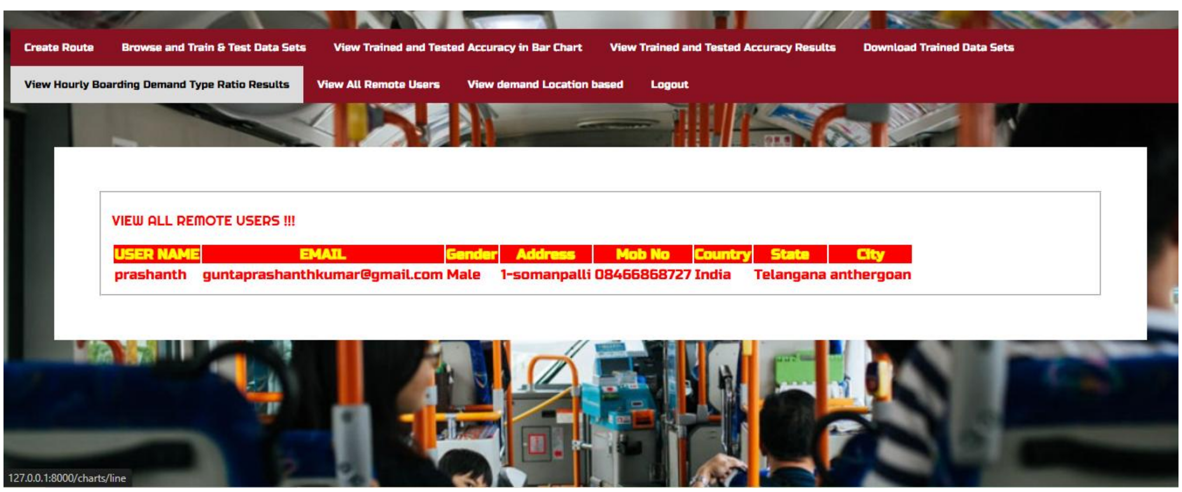

Figure 6 Service Provider Screen

This Figure provides a platform for predicting bus passenger demand using smart card data and deep learning. The displayed section specifically shows a list of remote users with their associated information.

www.ijbar.org
ISSN2249-3352(P)2278-0505(E)
CosmosImpactFactor-5.86

The navigation bar offers access to various aspects of the project, from data handling to model evaluation and user management.

Navigation Bar :

- View Hourly Boarding Demand Type Ratio
- Download Trained Data Sets
- View Hourly Boarding Demand Type Ratio Results
- View All Remote Users
- Logout

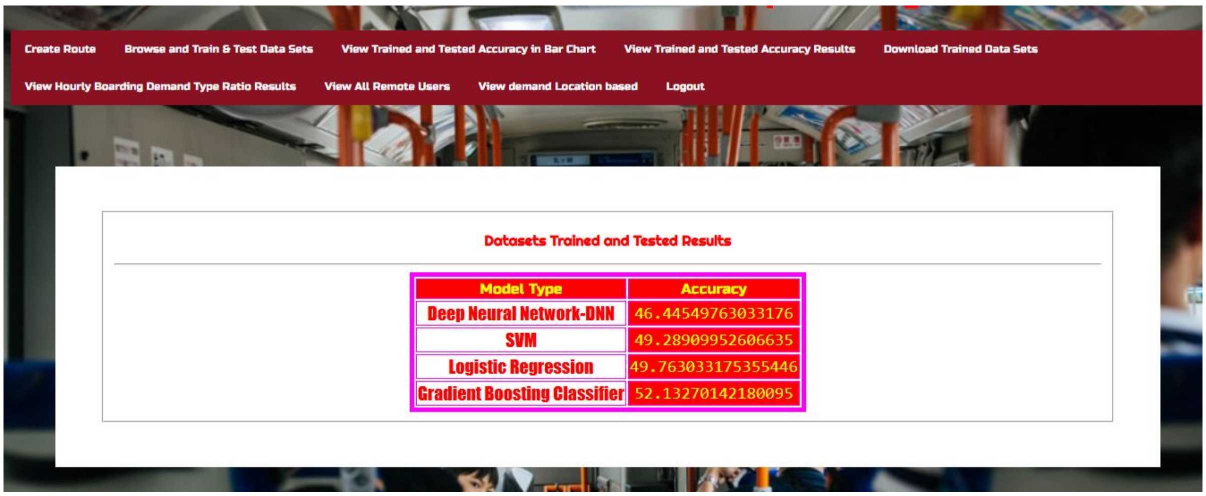

Figure 7 Comparison of model accuracy Screen

The Figure shows a comparison of model accuracy across different machine learning algorithms, highlighting the performance differences and suggesting areas for further analysis.

- Deep Neural Network-DNN: 52.13270142180095% accuracy
- SVM: 53.080568720337915% accuracy
- Logistic Regression: 53.55450236966824% accuracy
- Gradient Boosting Classifier: 5.08856872037915% accuracy

5. CONCLUSION

The study on predicting hourly boarding demand for bus transportation provides an essential tool for improving urban mobility and optimizing public transport services. As cities grow, public transportation systems face increasing challenges in balancing supply and demand. This research leverages predictive modeling to forecast passenger boarding demand, helping transit authorities make data-driven decisions for resource allocation and service management. The research employed various machine learning models to analyze historical boarding data and detect patterns. The results demonstrated the feasibility of accurately predicting hourly demand at different bus stops based on key parameters such as time, route, and day of the week. The data-driven approach not only enhances the efficiency of public transport but also improves commuter satisfaction by reducing wait times and ensuring appropriate bus frequency. The Research focused on implementing robust data collection, pre-processing, and predictive modeling methodologies. The accuracy metrics obtained from the evaluation of the models indicate the reliability of the system in identifying high-demand and low-demand periods. The incorporation of user authentication and data management features ensures that the system maintains security and privacy standards. The developed prediction system is scalable and can be adapted to different cities with varying transportation infrastructures. By predicting boarding demand in advance, authorities can allocate resources more effectively, thus minimizing operational costs and environmental impact. The use of machine learning algorithms ensures continuous improvement of prediction accuracy as more data becomes available.

REFERENCE

[1]. D. Luo et al., "Fine-Grained Service-Level PassengerFlow Predictionfor Bus Transit Systems Basedon MultitaskDeep Learning" IEEETransactions on Intelligent Transportation Systems, vol.22, no.11, pp. 7184–7199, Nov. 2021, doi:10.1109/TITS.2020.3002772.

[2]. Y. Liu, C. Lyu, X. Liu, and Z. Liu, "Automatic Feature Engineering for Bus Passenger Flow Prediction Basedon ModularConvolutional NeuralNetwork" IEEETransactions onIntelligent TransportationSystems, vol.22, no.4, pp. 2349–2358, Apr.2021, doi:10.1109/TITS.2020.3004254.

[3]. W. Lv, Y. Lv, Q. Ouyang, and Y. Ren, "A Bus Passenger Flow Prediction Model Fused with Point-of-Interest Data Based on Extreme Gradient Boosting" Applied Sciences, vol. 12, no. 3, p. 940, Jan. 2022, doi: 10.3390/app12030940.

[4]. V. G Sand H.V S, "Prediction ofBus PassengerTraffic usingGaussian ProcessRegression" Journalof Signal Processing Systems, vol. 95, no. 2–3, pp. 281–292, Mar. 2023, doi: 10.1007/s11265-022-01774-3.

[5]. G. Cheng and C.He, "Analysis ofbus travelcharacteristics andpredictions of elderly passenger flow based on smart card data" Electronic Research Archive, vol. 30, no. 12, pp. 4256–4276, 2022, doi: 10.3934/era.2022217.

[6]. X.-G. Luo, H.-B. Zhang, Z.-L. Zhang, Y.Yu, and K. Li, "A New Framework of Intelligent Public Transportation SystemBased onthe Internetof Things" IEEE Access, vol. 7, pp. 55290–55304, 2019, doi: 10.1109/ACCESS.2019.2913288.

[7]. A. Vigren and R. Pyddoke, "The impact on bus ridership of passenger incentive contracts in public transport," Transportation Research Part A: Policy and Practice, vol. 135, pp. 144–159, May 2020, doi: 10.1016/j.tra.2020.03.003.

[8]. W. Wang, Y.Wang, G.H. de A. Correia, and Y. Chen, "A Network-Based Modelof Passenger Transfer Flow between Bus and Metro: An Application to the Public Transport System of Beijing" Journal of Advanced Transportation, vol. 2020, pp. 1–12, Dec. 2020, doi: 10.1155/2020/6659931.

[9]. H. Zhou, C. Yuan, N. Dong, S. C. Wong, and P. Xu, "Severity of passenger injuries on public buses: A comparative analysis of collisioninjuries and non-collision injuries" Journal of Safety Research, vol. 74, pp. 55–69, Sep. 2020, doi: 10.1016/j.jsr.2020.04.003.

[10]. N. B. Dhital, L.-C. Wang, H.-H. Yang, C.-H. Lee, W.-H. Shih, and C.-S. Wu, "Effects of the COVID-19 pandemic on public bus occupancy and real-world tailpipe emissions of gaseous pollutants per passenger kilometertraveled" Sustainable Environment Research, vol. 32,no. 1, p. 35, Aug. 2022, doi: 10.1186/s42834-022-00146-7.

[11]. R. Burdzik, W. Chema, and I. Celiński, "A study on passenger flow model and simulation in aspect of COVID-19 spreadingon publictransport busstops" Journalof Public Transportation, vol. 25, p. 100063, 2023, doi: 10.1016/j.jpubtr.2023.100063.

[12]. S. Liyanage, R. Abduljabbar, H. Dia, and P.-W. Tsai, "AI-based neural network models for bus passenger demand forecasting using smart card data" Journal of Urban Management, vol. 11, no. 3, pp. 365–380, Sep. 2022, doi: 10.1016/j.jum.2022.05.002.

[13]. M. Nayak, A. Ladha, and N. K. Chaubey, "A Comprehensive Comparative Analysis of Passenger Demand Prediction for Improving the Urban Bus Transportation System (UBTS)" International Journal of Engineering Trends and Technology, vol. 70, no. 9, pp. 269–279, Oct. 2022, doi: 10.14445/22315381/IJETT-V70I9P227.