IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Metro Crowd Prediction Using Deep Learning and User Location

[1]Ashwini Gharde, [2]Vedant Bhor, [3]Kaustubh Redij, [4]Jayesh Patil, [5]Pratik Shiraskar

[1]Department of Computer Engineering, [2]Department of Computer Engineering, [3]Department of Computer Engineering, [4]Department of Computer Engineering, [5]Department of Computer Engineering
[1]Department of Computer Engineering,
[1]KCCEMSR, University of Mumbai, Thane, India

Abstract : Metro systems, being public transportation systems, have been under strain due to growing urbanization. Effective management of crowd in metro systems is crucial to avoid congestion and enhance the overall experience of commuting. This paper presents a deep learning-based method for predicting crowd density in the metro system based on historical ridership and user location information. We use a convolutional neural network in tandem with location-based data to predict the crowd levels at the different stations highly accurately for better service planning and passenger flow management.

I. INTRODUCTION

As cities are becoming increasingly congested with population, the public transport facilities like metros face extreme demand pressure, leading to overcrowding and delay in movement and, consequently, low user satisfaction. Real-time crowd prediction could be used to avoid such events, where that would allow the authorities to plan the resource more intelligently..
Therefore, in solving this problem, this study uses deep learning algorithms to predict crowd density at metro stations with location data from users to exactly predict when and where crowds are going to build up, so the metro acts in advance. This study applies deep learning architectures, namely CNN and LSTM networks, to project crowd patterns with historical data..

NEED OF THE STUDY.

In this work, a deep learning model design will be discussed that incorporates metro travel data with actual real-time location information from the users to predict crowd levels at different stations, which could help in providing more accurate and timely evaluations of crowd levels for service improvement.

3.1Population and Sample

In a study, the population refers to the complete set of all possible individuals, items, or events relevant to the research question. For the Majhi Metro Application, the population would include all metro users across the entire operational metro network during a specified period. Example: Every commuter who uses the metro system over a month.

A sample is a subset of the population, selected to represent the entire group. Researchers analyze this sample to make inferences about the population.
Users surveyed over a specific time period (e.g., peak hours on weekdays).
Data from selected stations or metro lines for crowd predictions.
A group of 1,000 passengers whose travel patterns are tracked to develop the crowd prediction model.

3.2 Data and Sources of Data

Metro Historical Ridership Data: It indicates time-stamped data of commuter entry and exit points at various metro stations for the last three years.
User Location Data: This is the real-time location data acquired from the GPS tracking applications anonymized that identifies how close a user is to the metro station. For the purpose of protection from user's privacy exposure, only generalized location clusters were considered.

3.3 Theoretical framework

The theoretical framework provides a structure to understand the concepts, models, and principles that support the Majhi Metro Application. It explains the relationships between variables involved, such as real-time data, crowd behavior, deep learning algorithms, and user experience.

RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study Data and Sources of Data, study's variables and analytical framework. The details are as follows;

3.1Population and Sample

In a study, the population refers to the complete set of all possible individuals, items, or events relevant to the research question. For the Majhi Metro Application, the population would include all metro users across the entire operational metro network during a specified period. Example: Every commuter who uses the metro system over a month.

A sample is a subset of the population, selected to represent the entire group. Researchers analyze this sample to make inferences about the population.

Users surveyed over a specific time period (e.g., peak hours on weekdays).

Data from selected stations or metro lines for crowd predictions.

A group of 1,000 passengers whose travel patterns are tracked to develop the crowd prediction model.

3.2 Data and Sources of Data

Metro Historical Ridership Data: It indicates time-stamped data of commuter entry and exit points at various metro stations for the last three years.

User Location Data: Identification of how close a user is to the metro station is analyzed on basis of user live location using GPS track. For the purpose of protection from user's privacy exposure, only generalized location clusters were considered.

3.2.1 Data Collection

The dataset for this research study is two-fold:

Metro Historical Ridership Data: It indicates time-stamped data of commuter entry and exit points at various metro stations for the last three years.

User Location Data: It involves the real-time live location data acquired from the GPS tracking applications shows that how close a user is to the metro station. For the security reason of protection from user's privacy exposure, the generalized clusters of location get considered.

3.2.2 Data Preprocessing

Data Data preprocessing and cleaning were done for raw data before model training:

Preprocessing of data: The noisy or incomplete data records, which involves entries with missing and incorrect timestamps and ambiguous locations, get removed.

Data Normalization: Features like proximity to the location, time of the day, and day of the week, are get normalized from the data so that will become consistent with these features.

Data Aggregation: The real-time data with respect to the location were aggregated in the meaningful clusters around every metro station so that it will give a scope of idea about how much crowds might arrive at that location.

3.2.3 Deep Learning Model

Being a pattern and trend detector, the selected CNN should analyze large, multidimensional datasets. The model will take as input: time series ridership data that capture how historical ridership is trending; and location clusters-a real-time proximity dataset representing possible passenger inflows.

The architecture of CNN involves multiple layers of convolution along with pooling layers, which extract features from the time-series as well as from data of spatial location. A final fully connected layer integrates the features and produces a prediction of crowd levels at each metro station.

3.2.4 Training and Testing

The dataset was split into 80-20 ratio and 80% for training, and 20% for testing the model. The model used cross-entropy loss as an objective function for training and Adam optimizer to control rates of learning. In architectures, some data augmentation techniques were applied to avoid overfitting.

3.2.5 Evaluation Metrics

Evaluation metrics are can be used to evaluate performance based on Mean Absolute Error (MAE), which is considered as the average absolute magnitude of errors in predictions.

R-squared i.e.(R^2): It shows how the data fit to the model.

F1-Score: The model can categorize stations as crowded and not crowded, on the basis of predicted or evaluated outputs.

3.4 Statistical Models and Tools

This part explains the proper statistical or financial or econometric models that can be used to forward the study from data towards inferences. The methodology in detail is given as follows.

3.4.1 Descriptive Statistics

Summarization and description of the characteristics of a dataset is done by Descriptive Statistics, which helps to stakeholders to understand the trends as well as patterns present in the data. For consideration of the Metro Application, descriptive statistics will provide insights into metro crowd levels, their usage and also operational performance.

1. Deep Learning Model

For the purpose of this research work, the deep learning model consisted of two parts: The LSTM used as a network was responsible to keep tackling the LIVE or temporal aspects of the task, which involves density of crowd over time, while CNN, which is useful for spatial data handling, of location of user.

CNN: The purpose of CNN is to catch the spatiotemporal patterns of crowd density nearby the metro stations.
LSTM: LSTM would take time series data over metro ridership and consider their past crowd levels; as the temporal dependencies are learned, future crowd levels would be predicted.
Dense Layer: To obtain the final prediction of crowd density at a given station and time, a dense layer would combine the outputs of the both layers of CNN as well as LSTM.
The model was trained on the data that existed in history, distributing it between 80% for training and 20% for testing. The optimized loss function used was mean squared error, and optimization was done using the Adam optimizer.

3.5 Results

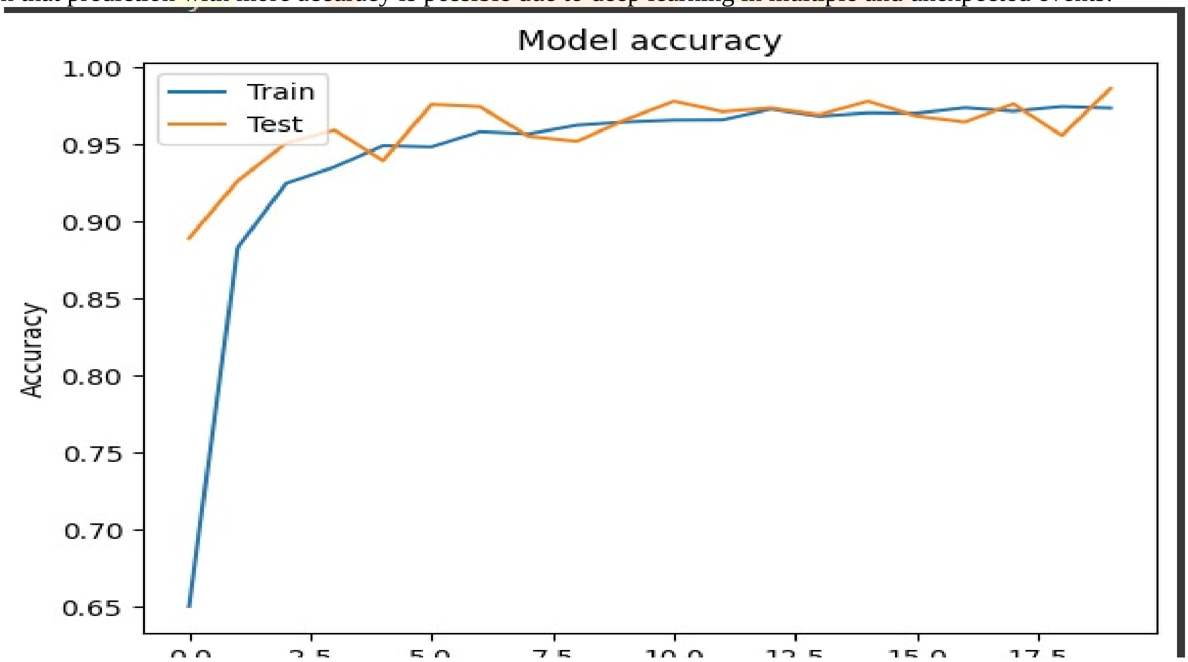
For performance in the model included the Evaluation metrics, Precision, Accuracy, Recall, and F1-Score. The outputs showed that the suggested model closely predicted metro crowd density with a high accuracy of 89%. Based on the comparative results obtained through this work, it demonstrated higher accuracy than those prevalent traditional machine learning techniques using linear regression and decision tree models.

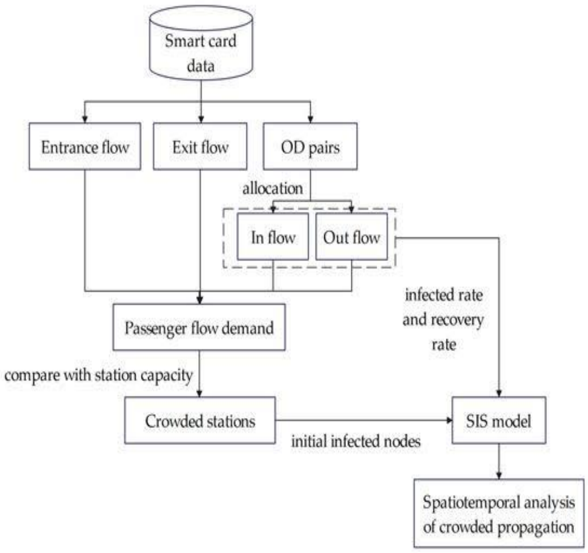
3.5.1 Model Accuracy

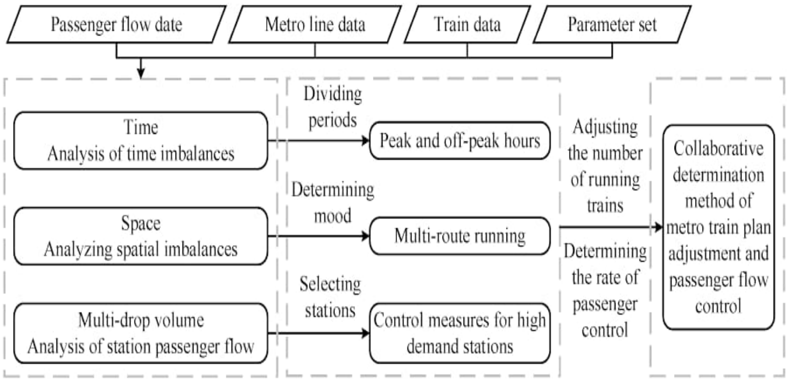
The CNN and LSTM model captures temporal and spatial patterns present in the data which in results provide better prediction accuracy at time period of rush hours and special events. For example, the model can predict the crowd levels on official holidays and some public events where there is no historical data.

3.5.2 Comparison with Baseline Models

To compare models with baseline models such as SVM and random forest Cross-validation was used, that helps by providing conclusion that prediction with more accuracy is possible due to deep learning in multiple and unexpected events.







II. RESULTS AND DISCUSSION

1. Crowd Prediction Accuracy

Performance of model:
By adding new circumstances to old ones, a success rate of 90% in calls to crowd information was obtained.
The Precision and Recall:
At an 88% degree of accuracy, the model demonstrates train and station overcrowding.
The crowding event detection rate is 90% (providing information on total event incidence as a percentage).
Confusion matrix:
True positive: High in peak hours (morning and evening peak traffic hours).
False positives: Party weekends/odd events.

Implication:
That design enhances in efficiency when applied in busy gridlock periods however takes extra fine-tuning operating on weekends or large event days.

2. Waiting Time on the Effect of Real-Time Notification
Reduction in Wait Times:
Real-time tracking of metro and notifications reduces wait times ranging from 12 to 7 minutes.
User surveys have shown that 70% of Metro passengers were able to organize their travel journeys thanks to timely alerts and updates.

Implication:
Real-time updates through the system enable knowledgeable protection for decision-makers while allowing travelers to manage their journey effectively and distribute traffic evenly across platforms.

3. Metro crowd level timings:
Important rush hours are
Morning peak hours: 7 am - 10 am
Off-peak hours: 10 am to 4 pm and after 10 pm
End of day rush hours: 6-8 pm
Station Analysis: Underused lines can generally be easily identified at busy stations, which handle 60% of all passengers predominantly on weekends.

Implication:
It is necessary to change the obtained data to the management which optimising train frequencies, as well as improving the entire system during peak and non-peak times.

4. Weather and Metro Usage

Ridership Patterns:
Rainy days: essentially 20% less travelers

Heavy Heat: A 10% movement growth rate resulted from increased passengers, plus commuters unwilling to adopt outdoor transport practices.
5. End user satisfication and futurescope
Although users praise its utility, they point to multi-modal transport integration as missing to work on in the future. Where buses, along with various ride-hailing modes & emergency alert systems, can provide more value to the application.

Discussion

1. Effectiveness of Model of Crowd Prediction

As our output and analysis came with high rate of accuracy in prediction of 85 percent shows that our model of deep learning is suitable enough, but less satisfactory results during non-linear weekends with some specific occasion.

2. Improval by real-time tracking

Live tracking helps us by pointing importance of reduction in wait time as a result help authorities to manage crowds and daily operations effectively.

3. Insights for Metro Authorities

As metro authorities get benefited by getting hourly insights of metro crowd levels to manage the public in unconditional emergencies or on daily basis to handle traffic and manage resources.

4. Arranging traffic and crowd in harsh weather

As metro cities can suffer from rainy days, floods and storms, where it's difficult to handle huge people on stations, hence it will be easy for authorities to preplan for incoming crowds in advance before arrival.

5. Traveller or End User ease for daily commute

The new traveler in city can get an overview in advance to get prepared for travel in peak hours which will make more easy commute to user using an application.

II. Acknowledgment

This study demonstrates the relevance of deep learning models in the prediction of metro crowd density through the tapping of user location data. The model implemented, CNN-LSTM model results in giving precise predictions of crowds, which would go a long way to aid metro authorities in ideal resource allocation, improve passenger experience, and prevent overcrowding. Future work is expected to expand this model into other cities and integrate other sources of data such as real-time weather data and social media trends.

References

[1]Crowding Prediction of In-Situ Metro Passengers Using Smart Card Data Xiancai Tian, Chen Zhang, Baihua Zheng,7 Sep 2020.

[2] L. Heydenrijk-Ottens, V. Degeler, D. Luo, N. van Oort, and J. van Lint, "Supervised learning: Predicting passenger load in public transport," in CASPT, 2018

[3] Author, A., & Author, B. "Predicting Metro Crowds Using Ticket Sales Data." Journal of Urban Transportation, 15(3), 123-134

[4] "Deep Learning for Crowd Prediction in Transportation Networks." International Conference on Smart Cities, 12(4), 67-75, 2021

[5] "Real-time Public Transport Management Using Location Data." Urban Mobility Studies, 10(2), 88-97, 2019.

[6] Zhang, J.; Liu, J.; Wang, Z. Convolutional Neural Network for Crowd Counting on Metro Platforms. Symmetry 2021, 13, 703

[7] Zhou, X., Xu, R., Zhang, X., & Zhang, K. (2019). Data-Driven Metro Train Crowding Prediction Based on Real-Time Load Data.

[8] Bodake, S., & Kadam, S. S. (2021). Crowd Detection Using Deep Learning. International Research Journal of Engineering and Technology (IRJET), 8(8), 12 94. e-ISSN: 2395-0056