

AI-Powered Smart Ticketing and Bus Management System for MSRTC: A Review

Unde Manasi Ramdas¹, Bhosale Sachin Bajirao², Banger Abhishek Popat³, Khatri Anand A.⁴

Department of Computer Engineering / Jaihind College of Engineering / SPPU, Pune, India

Abstract- The Maharashtra State Road Transport Corporation (MSRTC) runs one of India's largest state bus networks and has recently added digital ticketing machines, UPI payments, RFID smart cards and GPS-based vehicle tracking. This paper reviews selected work on electronic ticketing, real-time bus tracking, bus arrival time prediction, passenger demand forecasting and AI-based fleet management, and relates it to MSRTC's operating context. The reviewed studies use GPS traces, smart-card records and neural-network or ensemble models. Reported results are encouraging, but a direct comparison is difficult because the studies use different cities, datasets and metrics, and almost none are tested on Indian state-transport conditions. The review highlights gaps in data availability, evaluation practice, rural connectivity and integration with existing ticketing systems, and proposes a conceptual system workflow for future work.

Keywords— Smart ticketing, MSRTC, bus management, arrival time prediction, passenger demand forecasting, vehicle tracking, machine learning, intelligent transport systems.

I. INTRODUCTION

Public buses carry a large share of daily travel in Maharashtra. According to a December 2023 press report, MSRTC operated a fleet of 15,945 buses serving roughly 53 lakh commuters a day [1]. At this scale, small inefficiencies in ticketing, scheduling and dispatch affect millions of passengers and a large amount of revenue.

Traditional operations rely on fixed timetables, cash handling and manual supervision. Digital tools are changing this: electronic ticket issuing machines (ETIM), UPI and QR payments, smart cards and GPS tracking now generate large volumes of data.

Machine learning can use this data to predict arrival times, forecast passenger demand and flag operational problems, while a ticketing platform can give passengers live information.

This review discusses selected research and reported deployments according to their objectives, data sources, techniques, results and limitations. It separates four tasks that are often mixed together: digital ticketing, vehicle tracking, arrival time prediction, and demand forecasting.

II. REVIEW METHODOLOGY

This is a narrative review of selected sources, not a systematic survey. Sources were located through web searches on MSRTC ticketing and tracking, bus arrival time prediction, and smart-card-based passenger demand forecasting. MSRTC deployment information comes from press reports and vendor pages, which are less rigorous than peer-reviewed papers and are treated as such. Each source was examined for its purpose, data, method, results and limitations. Four questions guided the analysis: (1) What data and AI methods are used? (2) What outputs do they provide? (3) What evidence supports the reported results? (4) What problems remain for an MSRTC-type system?

III. LITERATURE REVIEW

In December 2023, MSRTC announced that digital payment was available on all its buses, with 34,000 digital machines supplied to conductors. The machines were provided by EbixCash under a five-year contract that also covers maintenance, and let passengers pay by scanning a payment gateway app [1]. The vendor describes its ETIM platform as supporting debit and credit cards, NFC, UPI, wallets and QR payments, along with RFID smart cards, and

lists an online ticketing system and about 40 lakh smart cards issued for MSRTC [2]. These are deployment descriptions from a press report and a vendor page, not evaluations. They show that the ticketing data infrastructure already exists, which is useful because every transaction can become training data. They do not report fare-collection accuracy, revenue-leakage reduction, transaction failure rates or passenger satisfaction. Any AI layer should be measured against such baselines.

Reports describe MSRTC's vehicle tracking system (VTS) as covering around 16,000 buses, including 850 in the Pune division, and enabling passengers to view bus locations and arrival times; a companion mobile app with seat availability was described as under development, and connectivity problems in rural areas, ghats and tunnels were acknowledged as a challenge [3]. A July 2025 report states that GPS installation would proceed in phases, mentions plans for over 10,000 buses with modern VTS, and says the government plans more than 2,000 smart buses with AI features and automatic passenger counting [4]. An earlier report describes a first rollout on 956 buses on Mumbai, Pune and Nashik routes, with passenger information displays at 23 locations [5]. The reports differ in fleet figures and dates, so the exact coverage of VTS should be checked against official MSRTC sources before citing numbers in a final paper. More importantly, none of the reports describes prediction accuracy, how arrival estimates are computed, or how tracking gaps in poor-coverage areas are handled. Tracking tells passengers where a bus is; it does not by itself predict when it will arrive.

Rashvand et al. [6] built a fully connected neural network to predict the time for a bus to reach its next stop, using New York City bus data covering 232 lines and about 2.13 million records. Inputs included one-hot encoded line, distance to the stop, day type, rush-hour status, stop index and a distance-based flag. They report a validation RMSE of about 35.7 seconds against an average schedule deviation of 491 seconds in the data, and show that the network scaled to many lines where an SVR baseline became untrainable beyond about 30 lines. The authors note that passenger demand and weather were not available and could improve results.

Liu et al. [7] propose an evaluation framework for several arrival time models (delay-based, k-nearest-neighbour, kernel regression, additive model and LSTM) on raw GPS data from Dublin, including a preprocessing method using KD-tree nearest-point search. Their emphasis is on whether models are practical to deploy, not only accurate. As summarised in [6], Sun et al. [8] combined clustering and Kalman filters on Nashville bus data and reduced prediction error against the agency's baseline, and Noor et al. [9] used SVR on city-bus data from Petaling Jaya, finding that adding weather made little difference. Ensemble-learning approaches for bus travel time have also been reported [10], [11]. Results are encouraging but come from structured urban networks with regular stops and GPS coverage. MSRTC operates long intercity and rural routes, ghat sections and mixed traffic, where dwell time and road conditions vary more. A low RMSE on one city's data does not transfer automatically. Studies should also report error by route and time of day, not only an average, since [6] itself shows large differences between lines.

Liyanage et al. [12] trained LSTM and BiLSTM models on smart-card ticketing data from 18 Melbourne bus routes and 1,781 stops, reporting that BiLSTM outperformed the other models tested with over 90% accuracy. A study on predicting hourly boarding demand from smart-card records [13] addresses class imbalance, since boarding at a given stop and hour is rare compared with non-boarding, and uses a deep generative adversarial network to create balanced training data. This literature is the closest match to an AI ticketing system, because ticket transactions directly feed demand models. However, smart-card data covers only card users; MSRTC passengers also pay by cash, UPI or concession passes, so demand estimates built from one payment channel may be biased. "Accuracy over 90%" also needs a clear definition and a comparison with simple baselines such as last week's same-hour demand.

Beyond prediction, AI can support decisions. A multi-agent reinforcement learning approach to bus holding control has been proposed to reduce irregular spacing between buses [14], and the transit

decision-support work summarised in [6] uses historical delay patterns to help planners adjust schedules. Control methods need reliable real-time data and operator trust, and they are harder to test safely than prediction models. For a student project, recommendations such as "consider adding a bus on

route X at this hour" are a safer first step than automatic control.

IV. COMPARATIVE ANALYSIS

Table 1. Comparative Analysis of Existing Work on Smart Ticketing and Bus Management

Study	Input / approach	Main output	Key limitation
MSRTC digital ticketing [1], [2]	ETIM, UPI/QR, RFID cards	Cashless ticketing	Deployment report; no accuracy or revenue metrics
MSRTC VTS and AI buses [3]–[5]	GPS tracking, passenger counting (planned)	Live location, ETA display	Figures vary; prediction method not described
Rashvand et al. [6]	NYC bus records, fully connected NN	Next-stop arrival time	Urban data; no demand or weather inputs
Liu et al. [7]	Dublin GPS data, several models	Model evaluation framework	Preliminary results; non-Indian context
Sun et al. [8]; Noor et al. [9]	Clustering, Kalman filter; SVR	Arrival and delay prediction	Local datasets; cited via [6]
Liyanage et al. [12]	Melbourne smart-card data, BiLSTM	Stop-level passenger demand	Card users only
Smart-card boarding study [13]	Smart-card records, deep GAN	Hourly boarding demand	Focus on imbalance, not deployment
Chen et al. [14]	Multi-agent reinforcement learning	Bus holding control	Needs reliable real-time data

V. RESEARCH GAPS

- Data availability. No public MSRTC ticketing or GPS dataset was found in the reviewed sources. Student projects must use public datasets from other cities or simulated data and say so clearly.
- Different tasks treated together. Arrival time prediction, demand forecasting, anomaly detection and scheduling should be formulated and evaluated as separate tasks.
- Evaluation practice. Averages such as RMSE or accuracy hide poor performance on particular routes, peak hours or rare events. Per-route errors, MAE/MAPE and simple baselines should be reported.
- Data splitting. Random splits of time-ordered data can leak information from neighbouring timestamps. Chronological or route-wise splits give a more honest estimate.
- Indian operating conditions. Rural coverage, ghat sections, festival peaks and long-distance routes are not represented in the urban datasets reviewed; MSRTC itself acknowledges connectivity gaps [3].
- Mixed payment channels. Cash, UPI, smart-card and pass travel produce different data, so demand models trained on one channel may be biased.
- Dwell time and passenger counts. Both [6] and [4] point to passenger counting as a missing input that affects dwell time and crowding.
- Privacy and integration. Smart-card data is personal, and new software must work with existing vendor-run ETIM and tracking systems rather than replace them.
- Operational metrics. False alerts, prediction lead time, offline behaviour and conductor workload should be evaluated, not only model accuracy.

VI. PROPOSED IMPROVEMENTS

- Define the task first: for example, next-stop arrival time, hourly boarding demand by route, or detection of unusual ticket patterns.
- Document the dataset: source, size, collection interval, features, missing values and how labels were produced. If simulated, describe how.
- Start with baselines: compare a timetable or historical-average baseline with Random Forest, LightGBM and, if justified, an LSTM.
- Use proper testing: split by time or by route; fit preprocessing on training data only.
- Report suitable metrics: RMSE, MAE and MAPE for regression; precision, recall and F1 for anomaly detection; results broken down by route and hour.
- Handle poor connectivity: design the conductor device to store transactions offline and sync later, and test behaviour when GPS data is missing.
- Protect passenger data: anonymise card identifiers and limit what the dashboard exposes.
- Test each component separately: ticketing flow, tracking accuracy, prediction model and dashboard should each have their own tests.

VII. PROPOSED SYSTEM WORKFLOW

The workflow below is a conceptual architecture derived from the reviewed literature.

It has not been implemented or validated.

- Data Collection: Ticket transactions (route, stop, time, payment mode), GPS positions and, if available, passenger counts.
- Data Preprocessing: Remove duplicates and invalid GPS points, fill or flag gaps, and match trips to stop sequences.
- Feature Engineering: Time of day, day type, festival or weekend flag, distance to next stop, recent dwell times, route and segment identifiers.
- Passenger Application: Booking, live bus location, estimated arrival time and seat availability.

- AI Modules: (a) arrival time prediction, (b) short-term demand forecasting per route and hour, (c) anomaly detection on ticketing and trip data.
- Depot Dashboard: Delay alerts, predicted crowding and suggestions for extra or reduced trips, for human decision.
- Evaluation and Feedback: Compare predictions with actual arrivals and ticket counts, track errors per route, and retrain periodically.

VIII. CONCLUSION

MSRTC has already built much of the digital foundation needed for AI-assisted operations, including cashless ticketing machines, smart cards and vehicle tracking [1]–[5]. Research elsewhere shows that arrival time and passenger demand can be predicted from GPS and smart-card data [6], [7], [12], but the reported results come from other cities and are not directly comparable with each other or with MSRTC conditions. For future student projects, the most useful steps are to choose one clear task, use a documented dataset, test with time-based or route-based splits, report route-level metrics against simple baselines, and judge the system by measured results rather than expected improvements.

REFERENCES

1. Express News Service, "MSRTC goes digital: Now, pay for your tickets through UPI," Mumbai, Dec. 11, 2023 (press coverage hosted at ebixcash.com).
2. EbixCash, "Bus exchange solutions," company web page, ebixcash.com/ebixcash-business/bus-exchange-solutions/. [Accessed: Oct. 2026]
3. "MSRTC introduces vehicle tracking system for passengers across Maharashtra," AbhiBus blog, abhibus.com/blog/msrtc-introduces-vehicle-tracking-system/.
4. Staff Reporter, "MSRTC to install GPS in its buses," The Hitavada, Jul. 1, 2025.
5. "Maharashtra state transport buses get vehicle trackers and info system," Together We Can Mumbai, togethervcan.in/?p=5401. [Verify date]
6. N. Rashvand, S. S. Hosseini, M. Azarbayjani, and H. Tabkhi, "Real-time bus arrival prediction: A

- deep learning approach for enhanced urban mobility," arXiv:2303.15495, v3, Mar. 2024.
7. D. Liu et al., "BusTime: Which is the right prediction model for my bus arrival time?" arXiv preprint, Mar. 2020.
 8. F. Sun, Y. Pan, J. White, and A. Dubey, "Real-time and predictive analytics for smart public transportation decision support system," in Proc. IEEE Int. Conf. Smart Computing (SMARTCOMP), 2016, pp. 1–8.
 9. R. M. Noor et al., "Predict arrival time by using machine learning algorithm to promote utilization of urban smart bus," 2020. [Verify venue and volume]
 10. G. Zhong, T. Yin, L. Li, J. Zhang, H. Zhang, and B. Ran, "Bus travel time prediction based on ensemble learning methods," IEEE Intell. Transp. Syst. Mag., vol. 14, no. 2, pp. 174–189, 2020.
 11. H. Huang, L. Huang, R. Song, F. Jiao, and T. Ai, "Bus single-trip time prediction based on ensemble learning," Comput. Intell. Neurosci., 2022.
 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," J. Urban Manag., vol. 11, no. 3, pp. 365–380, 2022, doi: 10.1016/j.jum.2022.05.002.
 13. "Predicting hourly boarding demand of bus passengers using imbalanced records from smart-cards: A deep learning approach," 2023, White Rose Research Online, eprint 191453. [Add author list and venue from the source]
 14. C. Chen, W. Chen, and Z. Chen, "A multi-agent reinforcement learning approach for bus holding control strategies," Adv. Transp. Stud., 2015 (as cited in [6]).