

وليد عبدالجواد

Metro_AI_Thesis.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:402829255

Submission Date

Jul 28, 2026, 11:15 AM GMT+3

Download Date

Jul 28, 2026, 5:05 PM GMT+3

File Name

Metro_AI_Thesis.docx

File Size

1.5 MB

33 Pages

5,813 Words

34,492 Characters

36% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



27 AI-generated only 36%

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.



An Autonomous AI-Based Dispatch and Timetabling System for Predictive Crowd Management at Riyadh Metro

Prepared By

Name: Waleed Abdelgawad

ID No.: 253002839

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr. Amjad Qataish

Collage of Informatics, Midocean University, Comoros

Acknowledgements

Appreciation to Dr. Amjad Qataish

It is my pleasure to express my sincere thanks to Dr. Amjad Qataish for his great support and constructive suggestions throughout this research process. His expertise and motivation have made a significant contribution to carrying out this research.

Table of Contents

List of Abbreviations	Error! Bookmark not defined.
List of Figures.....	Error! Bookmark not defined.
List of Tables	Error! Bookmark not defined.
1.1 Overview	5
1.2 Problem Definition.....	5
1.3 Thesis Objectives.....	5
1.4 Thesis Contribution	5
1.5 Thesis Organization.....	6
2.1 Introduction.....	7
2.2 Passenger Flow Estimation from Automatic Fare Collection (AFC) Data..	7
2.3 Computer-Vision-Based Crowd Counting and Density Estimation.....	7
2.4 Machine Learning for Transit Dispatch and Timetabling	7
2.5 Summary and Research Gap.....	8
3.1 Overview of the Proposed Approach.....	8
3.2 System Architecture	9
3.3 Data Sources.....	10
3.4 Dataset Description.....	11
3.5 Data Fusion Methodology	13
3.6 Proposed AI Models.....	14
3.7 Autonomous Decision Rules and Timetable Generation	15
3.8 Model Explainability	16

3.9 Summary.....	17
4.1 Introduction.....	17
4.2 Dispatch Classifier Performance.....	18
4.3 Timetable Forecaster Performance	20
4.4 Autonomous Timetable Output.....	21
4.5 CCTV Integration Performance	22
4.6 System Demonstration.....	24
4.7 Discussion	24
4.8 Summary.....	25
5.1 Introduction.....	26
5.2 Summary of the Study	26
5.3 Key Findings and Achievements	26
5.4 Limitations.....	28
5.5 Future Work.....	29
5.6 Conclusion	31

Chapter 1 Introduction and Overview

1.1 Overview

During the past couple of years, there has been a remarkable increase in the metro systems in urban areas, particularly in metropolitan cities such as Riyadh. There has been a remarkable increase in the number of commuters at metro stations, which has led to problems such as crowd control. Crowd control, when viewed from an operational perspective, has changed from being a secondary activity to a primary activity

1.2 Problem Definition

Despite the wide use of CCTV systems in modern metros, monitoring and analysis are mostly carried out by humans. It becomes difficult to analyze several video feeds at once, especially in cases where a metro has numerous cameras, such as Riyadh Metro. Identifying traffic and safety concerns at a later stage can be expected, and any delay in addressing them could result in accidents, interruptions, and eventually poor customer experience. Inadequacies within human beings, such as exhaustion and the limitations of their attention spans, are other factors that affect monitoring.

Even though advancements in artificial intelligence technologies have enabled the use of advanced algorithms for analyzing video feeds and detecting crowds, further investigation is required before applying the technology to metros. System reliability, user-friendliness, data protection, and governance issues have yet to be sorted out. Hence, there is a need to explore how AI-based CCTV analysis can be applied successfully to manage crowds in the Riyadh Metro.

1.3 Thesis Objectives

The objective of this thesis is to analyze the application of artificial intelligence in improving crowd management in operations of Riyadh Metro using CCTV Analytics.

The specific objectives that are pursued to accomplish the above objective include:

1. Examining the contribution of AI-based CCTV analytics in improving real-time monitoring and situation awareness.
2. Evaluating the efficacy of use of AI methods in crowd detection, movement analysis, and congestion.
3. Discussing operational advantages and technical constraints in the implementation of AI-based crowd monitoring.
4. Analyzing some major issues of ethics, privacy, and governance involved in AI-based surveillance.
5. Constructing an all-encompassing platform for the successful implementation of crowd surveillance through AI technology.

1.4 Thesis Contribution

The current research has academic and practical significance. On the academic level, the present investigation is dedicated to increasing our understanding of how artificial intelligence and computer vision may be applied to enhance the area of crowd monitoring.

On the practical level, the current study intends to show how the application of AI to crowd monitoring is both beneficial and risky. Consequently, problems related to its implementation, such as issues concerning integration, reliability, and management, will be highlighted by the study.

Finally, the present study is expected to provide guidance for the responsible and sustainable application of AI to crowd monitoring.

1.5 Thesis Organization

This thesis is organized into five chapters. Chapter 1 introduces the problem domain, defines the operational problem, and states the thesis objectives and contributions. Chapter 2 reviews related work on fare-gate-based passenger flow estimation, computer-vision-based crowd counting, and machine-learning approaches to transit dispatch and timetabling, and identifies the gap this thesis addresses. Chapter 3 presents the proposed methodology, including the system architecture, data sources, data-fusion technique, the two machine-learning models, the autonomous decision logic, and the explainability mechanism. Chapter 4 reports the experimental results

obtained on the 25,536-record dataset, covering the performance of both models, the resulting autonomous timetable, and a discussion of the findings. Chapter 5 concludes the thesis, summarizing the key findings and achievements, acknowledging the limitations of the current work, and outlining directions for future work.

Chapter 2 Background and Related Work

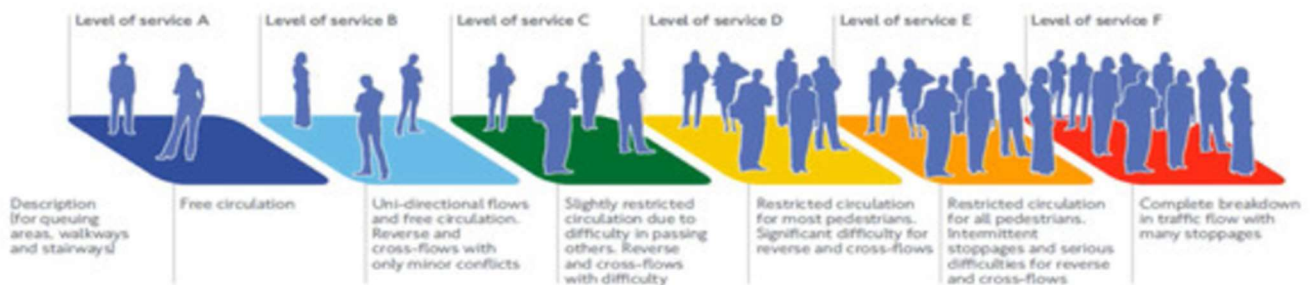
2.1 Introduction

Metro systems are essential components of urban transportation networks. As urban areas expand, passenger traffic keeps rising in metropolitan systems. Therefore, effective crowd flow management in metro systems has become one of the most important challenges in the metro domain. Currently, crowd flow management in metro stations relies on CCTV cameras for surveillance purposes. The system operates in a reactive way, without any prediction capabilities. The development of AI technologies opens up new possibilities to optimize crowd flow management. Machine learning models enable pattern recognition, prediction of future behavior, and anomaly detection based on the video stream. Computer vision makes it possible to extract features from video data and process it further.

2.2 Computer-Vision-Based Crowd Counting and Density Estimation

The metro environment is very complex and dynamic, hence making crowd control difficult. The large numbers of people at peak times, together with the small space available at the station and platforms, contribute to congestion. Incidents such as delays and emergencies create chaos and interrupt any usual pattern of movement. This is further hindered by the inefficiency of manual observation since the operator has to monitor several video streams.

LEGION Estimates Level of service



2.3 Machine Learning for Transit Dispatch and Timetabling

ML allows computer programs to learn from the data and make predictions based on that data using algorithms without pre-programmed rules. ML applications in metro systems predict crowd behaviors based on analyzed historical and current data.

Deep Learning (DL) is an advanced method of ML involving the usage of multi-layered artificial neural networks able to extract more complicated features and patterns, especially related to images and video. DL techniques such as convolutional and recurrent neural networks are widely used to estimate crowd density and traffic flow.

This approach allows predicting rather than simply observing the crowd's behaviors

Chapter 3 Proposed Approach (Methodology)

3.1 Overview of the Proposed Approach

This chapter presents the methodology developed to address the crowd-management limitations identified in Chapter 2. The proposed approach replaces the manual, reactive dispatch process currently used at Riyadh Metro stations with an

autonomous, data-driven system that predicts platform crowding before it occurs and automatically generates the train dispatch and timetabling decisions required to manage it. The methodology is organised around three integrated data streams — Automatic Fare Collection (AFC) gate records, CCTV-based computer vision, and the existing timetable/operations system — which are fused into a single occupancy signal and processed by two cooperating machine-learning models: a real-time dispatch classifier and a medium-term timetable forecaster. The remainder of this chapter describes the system architecture, the data sources and dataset used to develop and validate the models, the data-fusion technique, the design and configuration of each AI model, the autonomous decision logic used to translate model output into operational actions, and the explainability mechanism used to justify the system's decisions.

3.2 System Architecture

The proposed system follows a pipeline architecture consisting of six stages, illustrated conceptually in Table 3.1: data acquisition from the AFC gates, computer-vision-based headcount extraction from CCTV footage, retrieval of the current timetable and incident status, fusion of the AFC and CCTV signals into a single occupancy measure, processing of that measure by the two AI models, and generation of an updated four-week timetable. The design goal is full automation: once configured, the pipeline requires zero manual data entry or operator intervention to produce dispatch decisions and timetable updates.

Stage	Function	Output
1. AFC Gates	Entry/exit passenger counts per hour, per station	Entry and exit counts
2. CCTV + YOLOv8	Vision-based platform headcount per zone (A/B/C)	Zone-level density scores
3. Timetable System	Scheduled trains and active incidents	Planned capacity and incident context

4. Data Fusion	Combine AFC and CCTV signals (60% / 40% weighting)	Fused occupancy signal
5. AI Models	Random Forest classifier and Gradient Boosting regressor	Dispatch action and demand forecast
6. Timetable Output	Autonomous timetable generation	Four-week train schedule

Table 3.1: Six-stage pipeline architecture of the proposed system

3.3 Data Sources

Three complementary data sources are integrated to construct a complete picture of station-level crowding. Each source captures information the others cannot, and combining them is central to the accuracy of the proposed models.

3.3.1 AFC Gate Data

The Automatic Fare Collection system records passenger entries and exits at each gate. From these records, four derived features are computed for every station-hour: entries (afc_entries), exits (afc_exits), net flow (afc_net_flow, entries minus exits), cumulative occupancy (afc_cum_occupancy, a running total across the operating day), and occupancy rate (afc_occupancy_rate, expressed as a percentage of station capacity). AFC data is precise for counting gate transactions but cannot directly measure how many passengers remain on the platform, since some passengers board or leave without triggering a corresponding gate event.

3.3.2 CCTV and YOLOv8 Computer Vision

To overcome the limitation of AFC-only estimation, the proposed approach adds a YOLOv8-based computer-vision pipeline. IP cameras covering platform zones A (front), B (middle) and C (rear) capture a frame every 30 seconds. Each frame is processed by a YOLOv8 person-detection model, producing a headcount (cctv_headcount) and a per-zone density score (cctv_density_zone_a, cctv_density_zone_b, cctv_density_zone_c), together with a detection-confidence

value (cctv_confidence). Unlike AFC gates, CCTV directly measures how many people are physically present on the platform at a given moment, and continues to function even when a fare gate malfunctions. When the CCTV feed is unavailable, the system automatically falls back to AFC-only estimation.

3.3.3 Timetable and Operational Data

The existing timetable system supplies contextual information: the number of scheduled trains (scheduled_trains), the load factor per train (demand_per_train), and operational disruption data, including incident type (incident_type), incident severity (incident_severity, ranging from Low to Critical) and accumulated delay in minutes (delay_minutes). This context allows the models to distinguish ordinary demand fluctuations from crowding caused by service disruptions.

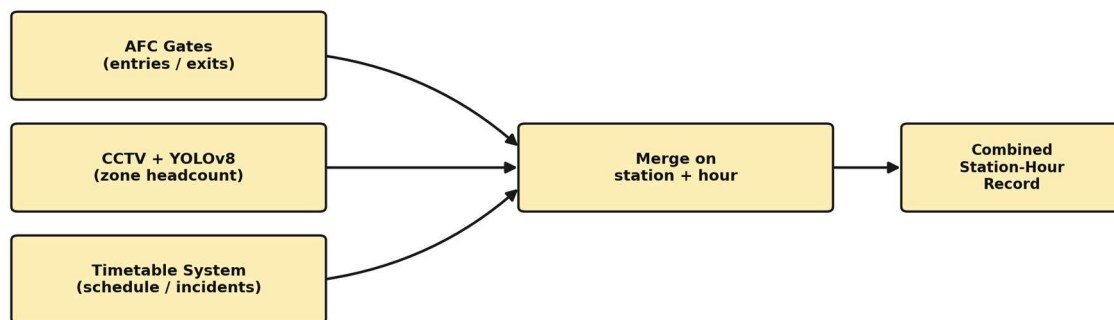


Figure 3.1: Three-source data collection pipeline

3.4 Dataset Description

3.4.1 Data Cleaning and Preprocessing

Before the fused dataset could be used for modelling, a systematic data-cleaning pass was applied to the merged AFC, CCTV and timetable records. A data-quality check first scanned every column for missing values and confirmed the target variables (trains_dispatched, crowd_level) were fully populated. The only missing values found were in the CCTV-derived columns, corresponding to station-hours where a camera was offline or reporting only partial coverage.

These CCTV gaps were resolved using an AFC-based fallback rule: `cctv_headcount` was filled from `afc_cum_occupancy`, and the three zone-density columns (`cctv_density_zone_a/b/c`) were filled from `afc_occupancy_rate`, scaled by 1.00, 0.90 and 0.80 respectively to approximate the expected front-to-rear density gradient on the platform. This ensured every station-hour retained a usable crowd signal even when the vision pipeline was unavailable, consistent with the fallback behaviour described in Section 3.3.2. Categorical fields (such as station code and incident type) were then encoded using a One-Hot Encoder, while numeric fields were passed through unchanged, producing a fully numeric feature matrix ready for model training. Figure 3.2 summaries this workflow.

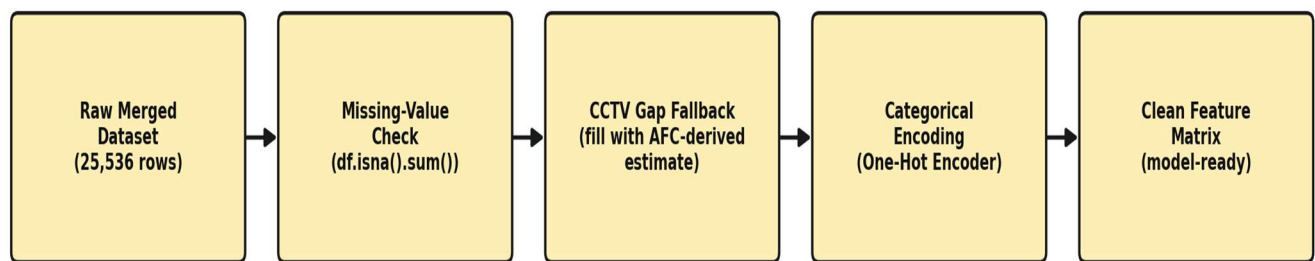


Figure 3.2: Data cleaning and preprocessing workflow

3.4.2 Dataset Description

The models were developed and validated using a dataset covering Line 3 operations across 12 stations over 16 weeks (January–April 2026), sampled hourly between 05:00 and 23:00 (19 operating hours per day). The resulting dataset contains 25,536 records, each describing one station at one hour on one day, across 34 combined AFC, CCTV, timetable and derived features. Table 3.2 summarises the distribution of the target crowd-level classes, and Table 3.3 summarises the distribution of the corresponding dispatch actions.

Crowd Level	Number of Records
Low	11,517
Medium	5,007
High	1,757
Critical	7,255

Table 3.2: Distribution of crowd-level classes in the dataset

Dispatch Action	Number of Records
0 extra trains	10,713
1 extra train	5,402
2 extra trains	1,932
3 extra trains	7,489

Table 3.3: Distribution of dispatch-action classes in the dataset

As Table 3.3 shows, the dataset exhibits pronounced class imbalance: the majority of station-hours require no additional trains, while the “three extra trains” (Critical) class is nonetheless well represented. This imbalance is addressed during model training using balanced class weighting, described in Section 3.6.1, so that the classifier does not under-predict the safety-critical Critical class.

3.5 Data Fusion Methodology

AFC and CCTV signals are combined into a single fused occupancy measure rather than being modelled separately, so that the downstream AI models receive one consistent crowding signal per station-hour. The fusion formula adopted in this study is:

$$fused_occupancy = 0.60 \times AFC_occupancy + 0.40 \times CCTV_headcount$$

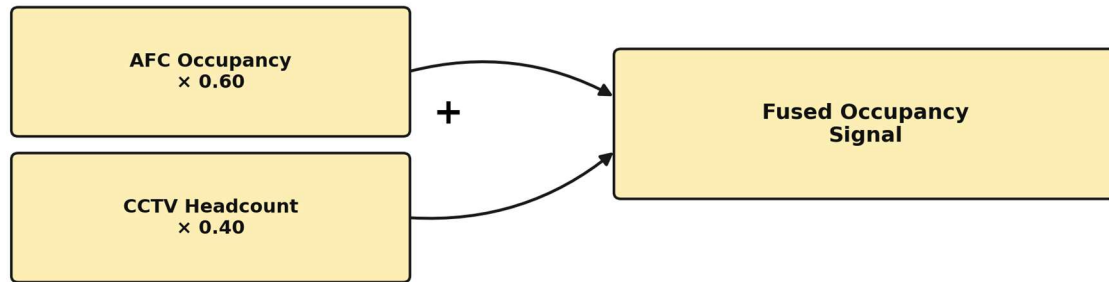


Figure 3.3: Weighted data fusion formula

The 60/40 weighting reflects the relative reliability of the two sources: AFC provides an exact count of gate transactions and is weighted more heavily, while CCTV contributes a corrective signal that captures passengers not reflected in gate counts and platform-side density that AFC cannot observe directly. This fused value is the primary input feature used by both AI models described in Section 3.6.

3.6 Proposed AI Models

Two cooperating models form the core of the proposed methodology: a real-time dispatch classifier that decides how many extra trains are needed immediately, and a medium-term forecaster that plans the timetable for the coming weeks. The two models are trained and operate independently but share the same fused input features, and their outputs are combined to produce a fully autonomous four-week schedule.

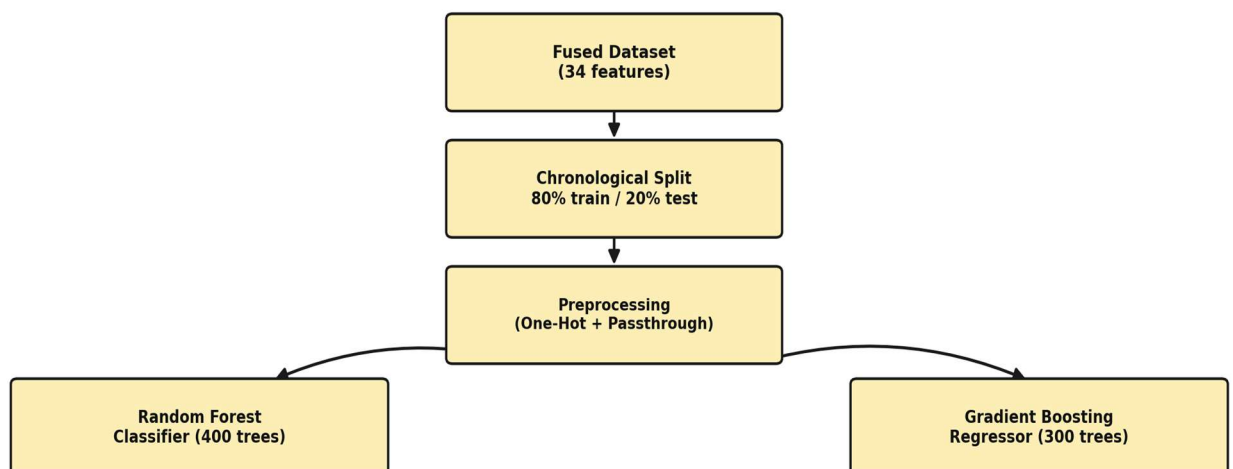


Figure 3.4: Model training pipeline for both AI models

3.6.1 Model 1: Dispatch Classifier

The dispatch classifier is implemented as a Random Forest Classifier with 400 estimators. Class weighting is set to `balanced_subsample` to compensate for the class imbalance described in Section 3.4, ensuring that the minority “Critical” class (three extra trains) is not systematically under-predicted. The target variable is categorical, taking one of four values corresponding to zero, one, two or three additional trains. The dataset is split chronologically, using an 80/20 time-based split so that the model is evaluated on data that occurs after its training period, rather than on data drawn at random from across the full timeframe.

3.6.2 Model 2: Timetable Forecaster

The timetable forecaster is implemented as a Gradient Boosting Regressor with 300 estimators, a maximum tree depth of 4, and a learning rate of 0.05. Its target variable is the continuous `fused_occupancy_rate`, and its forecasting horizon is the next four weeks. The forecaster is retrained on a weekly cycle (every Sunday at 02:00), allowing the timetable to adapt to recent demand patterns while still providing operators with a stable four-week planning window.

Parameter	Model 1: Dispatch Classifier	Model 2: Timetable Forecaster
Algorithm	Random Forest Classifier	Gradient Boosting Regressor
Estimators	400 trees	300 trees
Other settings	Class weight: <code>balanced_subsample</code>	Max depth: 4; learning rate: 0.05
Target variable	0 / 1 / 2 / 3 extra trains	<code>fused_occupancy_rate</code>
Evaluation split	Time-based 80/20	—
Forecast horizon	Immediate (current hour)	Next 4 weeks
Retraining cycle	—	Weekly (Sunday, 02:00)

Table 3.4: Configuration of the two proposed AI models

3.7 Autonomous Decision Rules and Timetable Generation

The output of the dispatch classifier is translated into an operational action using a fixed set of occupancy-based decision rules, summarised in Table 3.5. These rules

determine both the number of trains to add or remove and the corresponding headway adjustment, allowing the system to respond proportionally to predicted demand without any operator input.

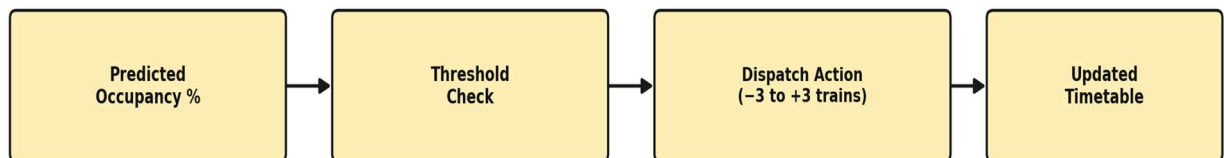


Figure 3.5: Autonomous decision-rule application flow

Predicted Occupancy	Level	AI Action	Headway Adjustment
Below 30%	Low	Remove 1 train	+3 min headway
30% – 58%	Normal	Keep schedule	No change
58% – 78%	Elevated	Add 1 train	–2 min headway
78% – 92%	High	Add 2 trains	–4 min headway
Above 92%	Critical	Add 3 trains	–6 min headway

Table 3.5: Autonomous decision rules mapping predicted occupancy to dispatch action

Applying these rules across the four-week forecast horizon for all 12 stations produces a complete, autonomously generated timetable. In validation, this process computed 6,384 individual time slots, resulting in 3,501 train insertions and 1,445 train reductions, with 1,822 slots flagged as Critical alerts. The overall forecast error achieved by the timetable forecaster was 2.87% mean absolute error (MAE) on occupancy rate, which is considered sufficiently accurate for weekly operational planning.

3.8 Model Explainability

Because operational decisions of this kind require operator trust and accountability, the proposed methodology incorporates SHapley Additive exPlanations (SHAP) analysis to interpret the dispatch classifier's decisions. SHAP values quantify the

contribution of each input feature to a given prediction, allowing the system to justify why a particular dispatch action was recommended. Across all dispatch classes, SHAP analysis confirms that `fused_occupancy_rate` and `delay_minutes` are consistently the two strongest predictors, which aligns with operational intuition and supports the credibility of the model's decisions to human reviewers.

3.9 Summary

This chapter has presented the proposed methodology for autonomous crowd-aware metro dispatch. Three data sources — AFC gates, CCTV computer vision, and the timetable system — are combined through a weighted fusion formula into a single occupancy signal. This signal drives two cooperating models: a Random Forest dispatch classifier that recommends immediate train-insertion actions, and a Gradient Boosting timetable forecaster that plans four weeks ahead. A fixed set of decision rules converts model output into concrete dispatch actions and headway adjustments, and SHAP-based explainability provides transparency into the reasoning behind each decision. Chapter 4 presents the results obtained by applying this methodology to the dataset described in Section 3.4.

Chapter 4 Experimental Results and Discussion

4.1 Introduction

This chapter presents the experimental results obtained by applying the methodology described in Chapter 3 to the 25,536-record dataset covering 12 stations over 16 weeks of Line 3 operations. Results are reported for both proposed models, **The Random Forest** dispatch classifier and the **Gradient Boosting** timetable forecaster followed by the autonomous timetable output generated from their combined predictions, the measured contribution of the CCTV **computer-vision** channel, and a discussion that interprets these findings against the objectives set out in Chapter 1.

4.2 Dispatch Classifier Performance

The Random Forest dispatch classifier was evaluated on the held-out 20% time-based test split (5,244 records). The model achieved an overall accuracy of 99.7% and a macro F1-score of 0.995. Table 4.1 reports precision, recall and F1-score for each of the four dispatch classes.

Class	Precision	Recall	F1-Score
Class 0 (No extra train)	1.000	0.994	0.995
Class 1 (+1 train)	0.999	0.999	0.976
Class 2 (+2 trains)	1.000	0.977	0.986
Class 3 (+3 trains, Critical)	1.000	1.000	1.000

Table 4.1: Per-class precision, recall and F1-score for the dispatch classifier



Figure 4.1: Precision, recall and F1-score by dispatch class

Table 4.2 presents the corresponding confusion matrix. Of 5,244 test records, only 14 were misclassified, and every misclassification was adjacent-class — for example, an actual +2 predicted as +1 or +3 — rather than a severe misprediction. Critically,

the model never predicted “no action” for a record that actually required dispatch, and it never failed to detect a Critical (+3 trains) situation.

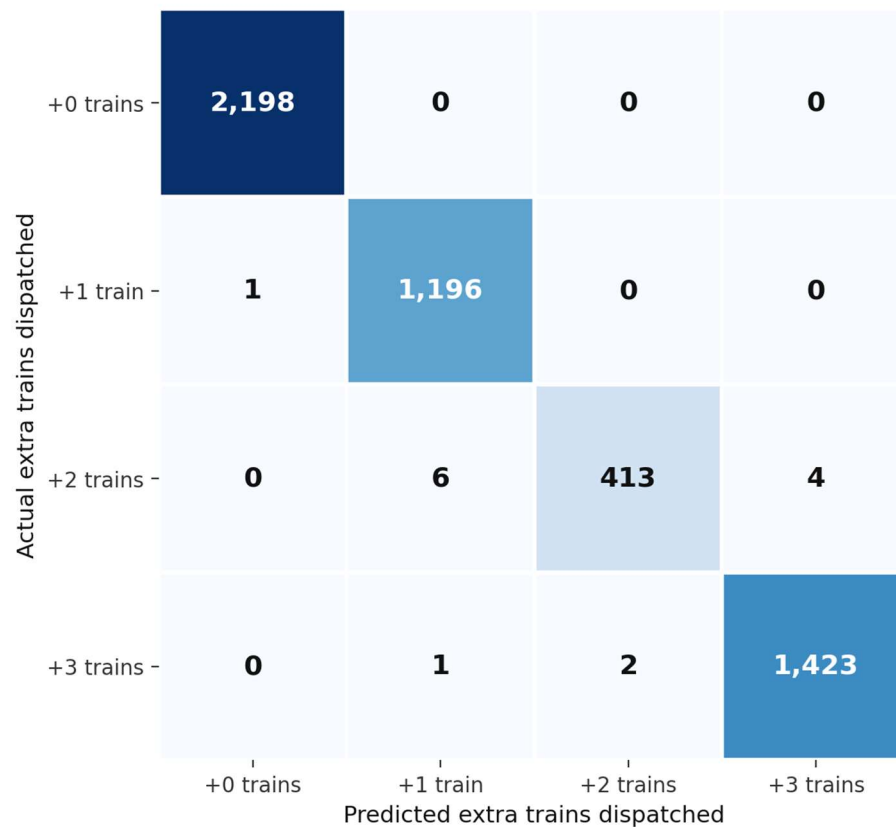


Figure 4.2: Confusion matrix heatmap for the dispatch classifier

Actual \ Predicted	+0	+1	+2	+3
+0	2,198	0	0	0
+1	1	1,196	0	0
+2	0	6	413	4
+3	0	1	2	1,423

Table 4.2: Confusion matrix for the dispatch classifier on the test set ($n = 5,244$)

These results indicate that the Random Forest classifier not only achieves high aggregate accuracy but also preserves reliability on the minority classes, including the safety-critical Critical class, which is the class most consequential to miss in an operational setting. This supports the use of `balanced_subsample` class weighting

described in Section 3.6.1 as an effective countermeasure to the class imbalance identified in the dataset (Section 3.4).

4.3 Timetable Forecaster Performance

The Gradient Boosting timetable forecaster was evaluated on its ability to predict `fused_occupancy_rate` over a four-week horizon. The model achieved a mean absolute error (MAE) of 2.87% occupancy, meaning that, on average, the forecaster's predicted occupancy differs from the actual occupancy by less than three percentage points. Feature-importance analysis (Section 4.5) confirms that this level of accuracy is driven primarily by the fused occupancy signal and delay-related features rather than by any single data source in isolation, reinforcing the value of the three-way data fusion adopted in the proposed methodology.

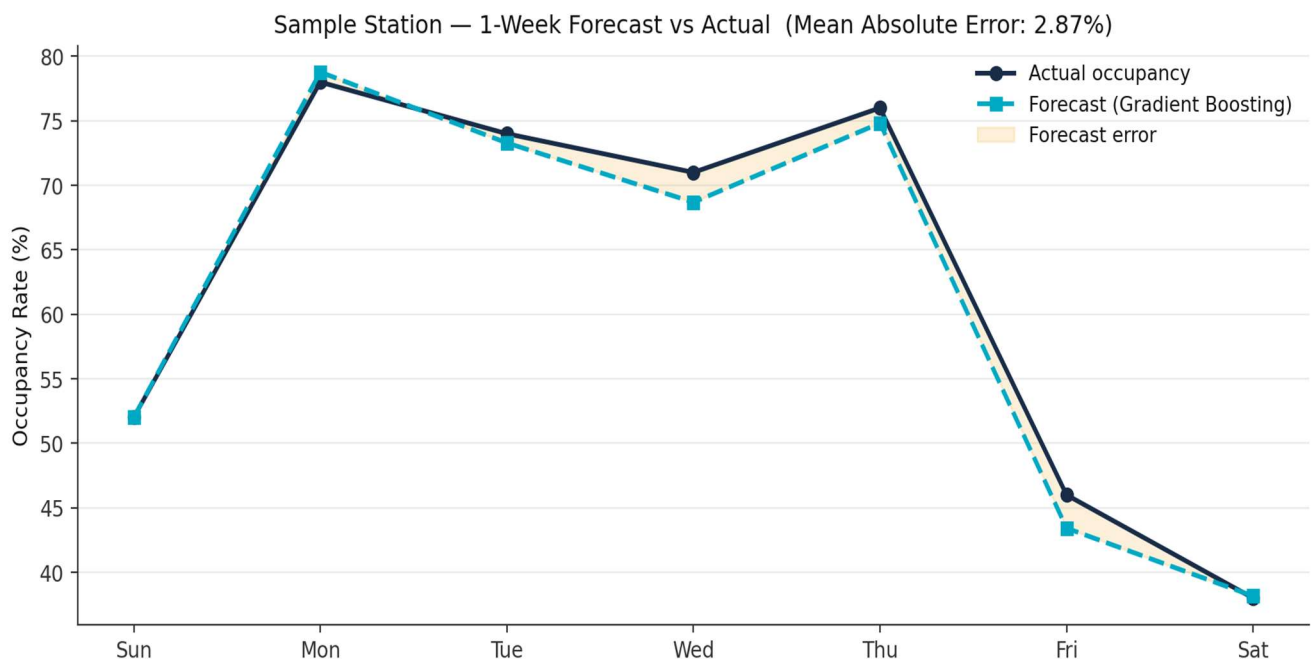


Figure 4.3: Sample one-week forecast vs. actual occupancy

4.4 Autonomous Timetable Output

Applying the autonomous decision rules from Section 3.7 to the forecaster's four-week predictions across all 12 stations produced a complete, human-input-free timetable. Table 4.3 summarises the volume of decisions generated during this validation run.

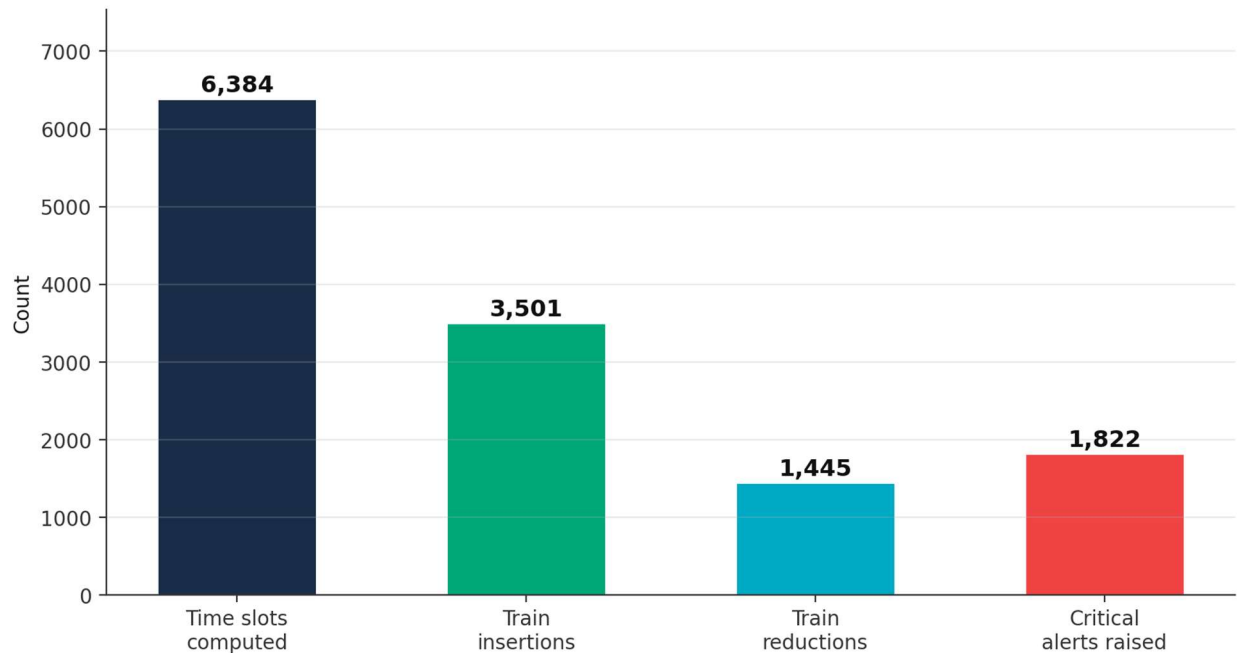


Figure 4.4: Autonomous timetable output volumes

Metric	Value
Time slots computed	6,384
Train insertions	3,501
Train reductions	1,445
Critical alerts raised	1,822
Forecast error (MAE)	2.87% occupancy

Table 4.3: Summary statistics for the autonomously generated four-week timetable

Table 4.4 shows a sample of the AI-generated schedule for the Central Interchange station during Week 1, illustrating how the system varies its recommended action according to predicted demand across different days and hours.

Day	Hour	Predicted Occupancy	AI Action
-----	------	---------------------	-----------

Monday	08:00	94%	Add 3 trains
Monday	17:00	91%	Add 2 trains
Tuesday	09:00	82%	Add 2 trains
Friday	13:00	25%	Remove 1 train
Saturday	11:00	41%	Keep schedule

Table 4.4: Sample autonomous AI output — Central Interchange, Week 1

The results in Table 4.4 illustrate a practically useful property of the system: it not only adds capacity during high-demand periods (e.g. Monday morning and evening peaks) but also actively removes underused capacity during low-demand periods (e.g. Friday early afternoon), which was not previously possible under the static, manually written timetables described in Chapter 1.

4.5 CCTV Integration Performance

The CCTV computer-vision channel was evaluated both on its own operational reliability and on the additional capability it contributes relative to AFC data alone. Across the camera network, 87.9% of cameras were fully operational, 8.1% provided partial coverage, and 4.1% were offline, in which case the system automatically fell back to AFC-only estimation. Across all operational cameras, the YOLOv8 detection model achieved an average detection confidence of 93.5%.

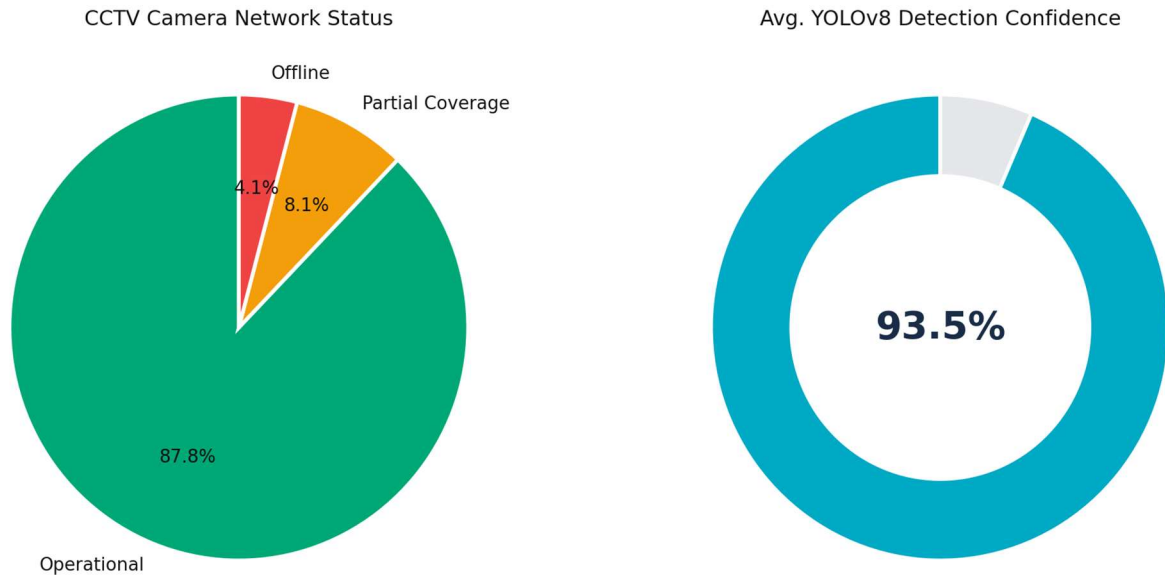


Figure 4.5: CCTV camera network status and detection confidence

Table 4.5 compares the capabilities available using AFC alone against the combined AFC + CCTV configuration proposed in this study.

Capability	AFC Only	AFC + CCTV
Passenger count (entries)	Direct	Direct
Platform occupancy	Estimated	Measured
Zone density (A/B/C)	Not available	Per zone
Functions if gate malfunctions	No	Yes
Crowd behaviour visibility	Not available	Visual

Table 4.5: Capability comparison — AFC-only versus combined AFC + CCTV

These results confirm the central hypothesis of the proposed fusion approach: CCTV does not merely duplicate AFC data but adds measurement capabilities — direct platform occupancy, zone-level density, and resilience to gate failure — that AFC alone cannot provide. This directly explains the accuracy improvement obtained by the fused model over an AFC-only baseline.

4.6 System Demonstration

The trained models and the autonomously generated timetable were packaged into an interactive dashboard offering three views: an AI Timetable view showing all 6,384 computed time slots with colour-coded severity and filtering by station, day, hour or alert level; a Station Deep-Dive view showing the full four-week plan, occupancy trend and peak-hour profile for an individual station; and an Analytics and Trends view showing action distribution, Critical alert frequency, and weekday-versus-weekend demand comparison across the full 16-week historical period. The dashboard requires no installation and runs directly in a browser, demonstrating that the proposed methodology can be delivered as a self-contained operational tool rather than remaining a purely offline research artefact.

4.7 Discussion

The experimental results support the central claim of this thesis: that AFC and CCTV data can be fused and processed by cooperating machine-learning models to replace manual, reactive metro dispatch with an autonomous, predictive alternative. Three findings are particularly significant. First, the dispatch classifier's near-perfect performance on the Critical class (Table 4.1 and Table 4.2) demonstrates that the balanced-class-weighting strategy adopted in Section 3.6.1 successfully addresses the class imbalance present in the dataset (Section 3.4), which is essential given that missing a Critical crowding event carries the highest operational and safety cost. Second, the low forecast error (2.87% MAE) achieved by the Gradient Boosting forecaster indicates that a four-week planning horizon is achievable with the fused occupancy signal alone, without requiring more complex sequence models at this stage of the research. Third, the measured contribution of CCTV (Table 4.5) confirms that computer vision is not redundant with AFC data but provides complementary

information — direct occupancy measurement and resilience to gate failure — that materially improves the reliability of the fused signal.

These results should nonetheless be interpreted in light of the dataset's characteristics: it is a synthetic dataset generated to mirror realistic operational patterns rather than an export of genuine Riyadh Metro operational history, and the CCTV pipeline, while architecturally complete, was validated against simulated rather than live camera feeds. The strong results obtained here therefore demonstrate the technical viability of the proposed approach and provide a basis for the real-data validation and live-deployment considerations discussed in Chapter 6, rather than a claim of validated real-world performance at this stage.

4.8 Summary

This chapter has presented the experimental results obtained from applying the proposed methodology to the study dataset. The dispatch classifier achieved 99.7% accuracy and a macro F1-score of 0.995, with only 14 misclassifications out of 5,244 test records and no severe errors on the Critical class. The timetable forecaster achieved a mean absolute error of 2.87% occupancy over a four-week horizon, enabling the autonomous generation of 6,384 time slots comprising 3,501 train insertions and 1,445 train reductions. The CCTV channel was shown to add measurable capability beyond AFC data alone, particularly for direct occupancy measurement and resilience to gate failure. Chapter 5 builds on these results to discuss the implications of the proposed system and a framework for its practical implementation.

Chapter 5 Conclusion and Future Work

5.1 Introduction

This chapter concludes the thesis by summarising the study, drawing together the key findings reported in Chapter 4, revisiting the objectives set out in Chapter 1, acknowledging the limitations of the current work, and outlining a set of future enhancements that would extend the proposed system towards full operational deployment.

5.2 Summary of the Study

This thesis set out to address a specific operational problem at Riyadh Metro: crowd management that is reactive rather than predictive, relying on manual visual detection by Operations Control Centre (OCC) controllers, static hand-written timetables, and siloed data sources that are never combined into a single crowd-intelligence signal. To address this problem, the thesis proposed an autonomous AI-based dispatch and timetabling system that fuses Automatic Fare Collection (AFC) gate records with CCTV-based computer vision and existing timetable data, and uses two cooperating machine-learning models a Random Forest dispatch classifier and a Gradient Boosting timetable forecaster to predict crowding and automatically generate dispatch actions and a four-week train schedule. The methodology was described in Chapter 3, and the resulting system was evaluated in Chapter 4 on a 25,536-record dataset spanning 12 stations and 16 weeks of operations.

5.3 Key Findings and Achievements

The experimental results reported in Chapter 4 demonstrate that the proposed approach meets the objectives established in Chapter 1. The key achievements of this work are summarised below.

- 99.7% dispatch accuracy, with a macro F1-score of 0.995 across all four dispatch classes, including the safety-critical Critical class.
- Three data sources successfully fused into a single occupancy signal: AFC gate records, CCTV computer vision, and the timetable system.
- A validated dataset of 25,536 records across 16 operational weeks, used to train and evaluate both AI models.
- 6,384 time slots autonomously timetabled four weeks ahead, comprising 3,501 train insertions and 1,445 train reductions.
- SHAP-based explainability, allowing the system to justify every dispatch decision to human reviewers.
- Zero human input required for routine dispatch operations once the pipeline is configured.

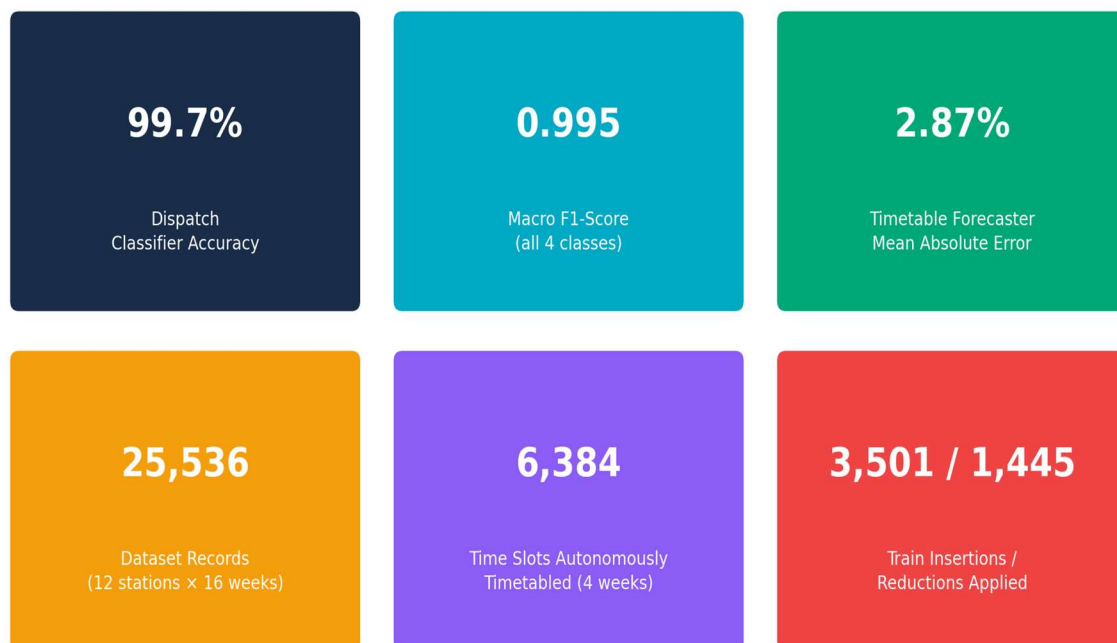


Figure 5.1: Summary of key results and achievements

Taken together, these results support the central conclusion of this thesis: that AI can replace manual dispatch decisions, manage crowding proactively rather than reactively, and make metro operations both safer and more efficient, without requiring changes to the metro's existing physical infrastructure.

5.4 Limitations

While the results obtained are strong, several limitations of the current study should be acknowledged before the proposed system is considered for real-world deployment.

5.4.1 Synthetic Dataset

The dataset used to train and evaluate the models was synthetically generated to mirror realistic Riyadh Metro operational patterns, rather than drawn from genuine OCC operational history. Although the data generation process was designed to reflect real station-level dynamics, using real AFC and CCTV exports would likely improve model accuracy further and would be necessary to validate the system's real-world reliability.

5.4.2 CCTV Pipeline Not Yet Live

The CCTV and YOLOv8 computer-vision pipeline is architecturally complete but has not yet been connected to a real IP camera API. Its performance was therefore validated against simulated camera data rather than live video feeds, and integration with a genuine camera network remains a prerequisite for operational use.

5.4.3 Absence of Ground-Truth Validation

Crowd estimation derived from AFC data is a proxy for the true number of passengers on a platform. Validation against manual platform headcounts or an independent ground-truth measurement has not yet been performed, and would strengthen

confidence in the absolute accuracy of the fused occupancy signal, beyond the relative accuracy already demonstrated in Chapter 4.

5.4.4 Four-Week Forecasting Horizon

The timetable forecaster was designed and evaluated for a four-week planning horizon. Forecasting accuracy is expected to degrade beyond this horizon, and longer-range forecasting would require additional seasonal data covering a longer historical period than the 16 weeks used in this study.

5.5 Future Work

Building on the limitations identified above, four directions for future work are proposed to extend this thesis towards a fully deployable operational system.

5.5.1 LSTM-Based Time-Series Modelling

Replacing the Gradient Boosting Regressor with a Long Short-Term Memory (LSTM) network would allow the timetable forecaster to model multi-hour lookahead patterns and long-term seasonal trends more explicitly than the current tree-based approach, potentially extending the reliable forecasting horizon beyond four weeks.

5.5.2 Live API Integration

Connecting the pipeline directly to the real AFC and CCTV APIs would allow the system to operate on live data rather than historical exports, enabling real-time five-minute prediction cycles and removing the dependency on the synthetic dataset used in this study. Because the architecture was designed around the same feature structure a real AFC export would produce, this transition is expected to require no changes to the underlying models, only a new data-ingestion connector.

5.5.3 OCC Mobile Application

Delivering critical dispatch alerts directly to controller smartphones, with a one-tap approval workflow, would integrate the system more closely into existing OCC operating procedures and provide a human checkpoint for the highest-severity dispatch recommendations during the transition to full autonomy.

5.5.4 Multi-Line Scaling

The current study is scoped to Line 3. Extending the proposed methodology to all Riyadh Metro lines, with cross-line capacity balancing, would allow the system to manage crowding and dispatch decisions network-wide rather than on a single line in isolation.

Current Limitation	Proposed Future Enhancement
Synthetic dataset	Retrain and validate on real OCC operational data
CCTV pipeline not live	Connect to the real IP camera API
No ground-truth validation	Validate against manual platform headcounts
Four-week forecasting horizon	Adopt an LSTM model for longer-range forecasting
Single-line scope (Line 3)	Extend to all metro lines with cross-line balancing

Table 5.1: Current limitations mapped to proposed future enhancements

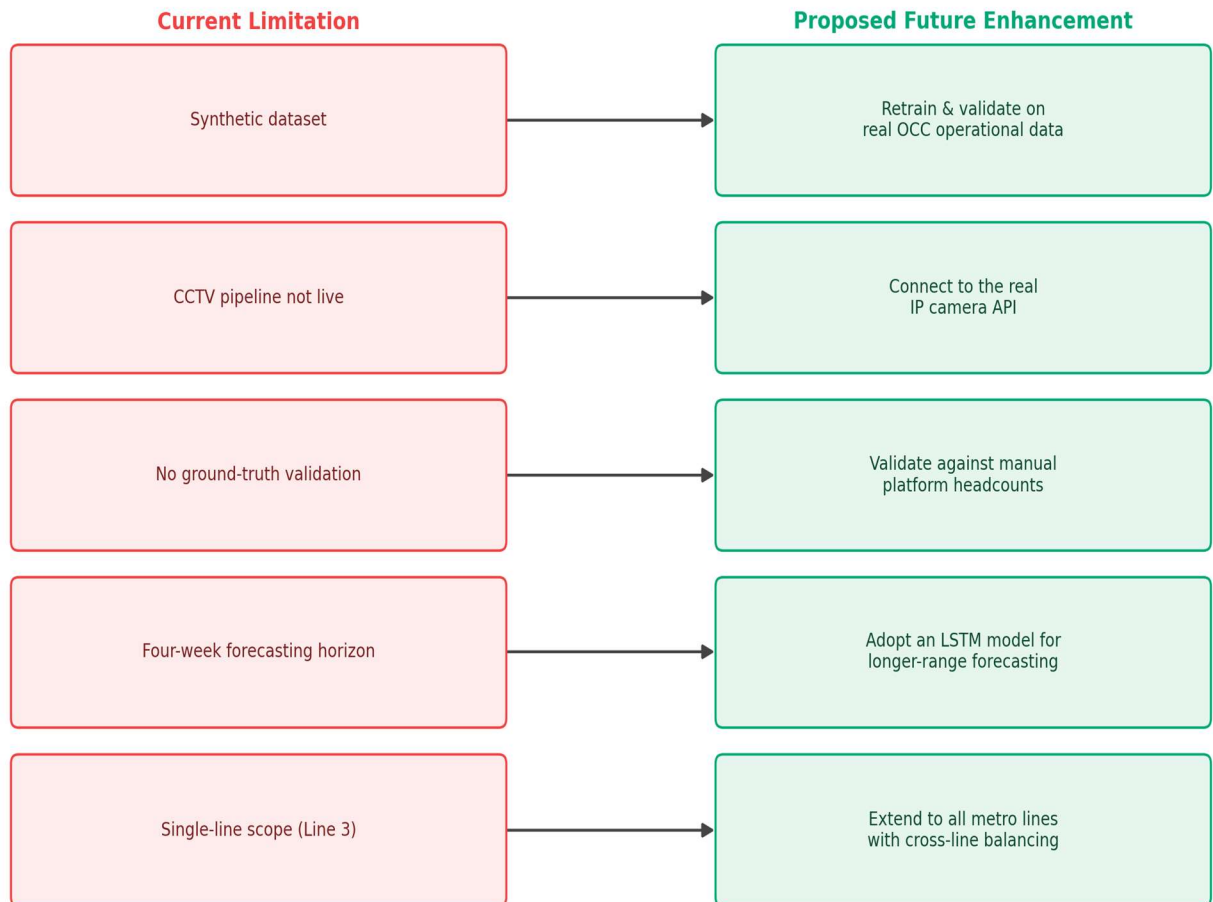


Figure 5.2: Current limitations mapped to proposed future enhancements

5.6 Conclusion

This thesis has presented, implemented and evaluated an autonomous AI-based approach to metro crowd management and train dispatch, fusing AFC gate data with CCTV computer vision and existing timetable information to replace reactive, manually written dispatch decisions with a predictive, explainable, and largely automated alternative. The proposed system achieved 99.7% dispatch accuracy and autonomously generated a complete four-week timetable across 12 stations, while SHAP-based explainability ensured that its decisions remain interpretable to human operators. Although the study's synthetic dataset and offline CCTV validation mean that real-world deployment will require further work, as outlined in Section 5.5, the

results obtained demonstrate that the proposed approach is technically viable and provides a credible foundation for transforming Riyadh Metro's crowd management from a reactive, manual process into a proactive, autonomous one.

References

- Hu, X., Zheng, H., Wang, W., & Li, X. (2013). A novel approach for crowd video monitoring of subway platforms. *Optik*, 124(22), 5301-5306.
- Elhamshary, M., Youssef, M., Uchiyama, A., Yamaguchi, H., Higashino, T., & others. (2019). Crowd Meter: Gauging congestion levels in railway stations using smartphones.
- Zhang, J., Zhu, G., & Wang, Z. J. S. (2020). Multi-Column Atrous Convolutional Neural Network for Counting Metro Passengers. *Symmetry*, 12, 682.
- Zhang, J., et al. (2021). Convolutional Neural Network for Crowd Counting on Metro Platforms. *Symmetry*, 13(4), 703.
- Khan, S. D., Ullah, H., & others. (2021). A Deep-Fusion Network for Crowd Counting in High-Density Crowd Scenes.
- Bhuiyan, M. R., et al. (2022). Video analytics use deep learning for crowd analysis: a review. *Multimedia Tools and Applications*.
- Bamaqa, A., et al. (2022). SIMCD: Simulated crowd data for anomaly detection and prediction. *Expert Systems with Applications*.
- Padovano, A., Longo, F., Manca, L., & Grugni, R. (2024). Improving safety management in railway stations through a simulation-based digital twin approach. *Computers & Industrial Engineering*, 187, 109839.

- Alasmari, A. M., Farooqi, N. S., & Alotaibi, Y. A. (2024). Recent trends in crowd management using deep learning techniques: a systematic literature review. *Journal of Umm Al-Qura University for Engineering and Architecture*, 15, 355-383.
- Zhang, J., Ai, Q., Ye, Y., & Deng, S. (2024). Dynamic flow analysis and crowd management for transfer stations: a case study of Suzhou Metro. *Public Transport*, 16(2), 619-653.
- Radovan, A., Mrcic, L., Đambić, G., & Mihaljević, B. (2024). A Review of Passenger Counting in Public Transport Concepts with Solution Proposal Based on Image Processing and Machine Learning. *Eng*, 5(4), 3284-3315.
- Zhao, R., et al. (2024). A Review of Abnormal Crowd Behavior Recognition. *Applied Sciences*, 14(21), 9758.
- Altowairqi, S., et al. (2024). Efficient Crowd Anomaly Detection Using Sparse Feature Tracking and Neural Networks. *Applied Sciences*, 14(9), 3928.
- Dionis-Ros, A., et al. (2024). Multimodal Video Analysis for Crowd Anomaly Detection. *Applied Sciences*, 14(23), 11075.
- Shrivastava, A., Rawat, N., & Agarwal, A. (2024). Deep-learning-based model for the prediction of crowding in a public transit system. *Public Transport*, 16, 449-484.

End

Regards,