Alkadhim Journal for Computer Science, Vol.4, No. 3 (2026)

Open Access ISSN: 3007-1437 (Online), ISSN: 3007-1429 (Print)

IRAQI Academic Scientific Journals

Alkadhim Journal for Computer Science (KJCS)

Journal Homepage: https://alkadhum-col.edu.iq/JKCEAS

KJCS ALKADHIM JOURNAL FOR COMPUTER SCIENCE

Risk Prediction in Cash-in-Transit Security Using Machine Learning: A Data-Driven Performance Evaluations

[1]Hasan Abdulrazzaq Jawad*, [2]Israa Shihab Ahmed , [3]Aya Salim Hasan , [4] Rabiu Aliyu Abdulkadir , [5]Amjed Abbas Ahme

[1]Department of Cyber Security, Imam Alkadhim University College– Iraq

[2]Informatics Institute for Postgraduate studies, university of information technology and communication– Iraq

[3]Shi'ite Endowment office – Iraq

[4]School of Computer Science and Technology, Shandong Xiehe University, Jinan, 250109, Shandong - China.

[5]Department of Cyber Security, Imam Alkadhim University College– Iraq

Article information

Article history:
Received: Month, xx, 2026
Accepted: Month, xx, 2026
Available online: Month, xx, 2026

Keywords:
Risk Prediction,
Cash-in-Transit,
Security,
Machine Learning,
Data-Driven,
Performance Evaluation.

Corresponding Author:
Hasan Abdulrazzaq Jawad
hasan.abdulrazzaq@iku.edu.iq

DOI:
https://doi.org/10.61710/ep6pag97

Abstract

Cash-in-transit (CIT) transactions are the transportation of banknotes and cash between ATMs, cash processing centers, banks and retailer locations. Robbery, route compromise, leaking of information from inside the group, vehicle attacks, abnormal dwell times, and unsafe route deviations are all rare but high-impact threats that are faced during these trips. This study designs a machine learning-based framework to predict the risk level CIT for each trip by leveraging a synthetic operational benchmark of 30,000 CIT trip records of which 847 were high-risk or incident ones, corresponding to 2.82% CIT incident rate. In addition, the data set consisted of temporal variables, route variables, variables describing the type of vehicle, crew member, cash value, location risk, and journey anomaly, with these divided into 75% train and 25% test stratified samples.

Three algorithms were tested: logistic regression, random forest and gradient boosting. The following metrics were used to evaluate the model performance: accuracy, ROC-AUC, precision, recall, F1-score, brier score, Precision@5%, and Recall@5%. Random forest had the highest accuracy (0.9178) and the best calibration obtained on the probability of the model, with the lowest Brier score (0.0467). Logistic regression performed best across the board (ROC-AUC 0.8449 and recall 0.292), discriminating and detecting higher risk trips with greater ability. Gradient boosting performed best in terms of sensitivity (recall) with a score of 0.3868 (Recall@5%), capturing 38.68% of the estimated incidents when only the top 5% riskiest trips are marked, and in terms of precision (0.276), and F1 score (0.282). The findings demonstrate that machine-learning risk scoring could complement the CIT control rooms, allowing for prioritization of trips for review, trip assignment for escorts, adjustment of dispatch, and better oversight of security, while leaving room for human decision on trips' security to be made by the control room's team.

1. Introduction

Today, cash-in-transit [1] is still a well-established issue in the bank, retail and ATM operations since all cash must be collected, replenished, sorted and sent between geographically spread out service points of note issue. The sector has also been identified as one of the biggest banking challenges with regards to safe transportation of money and assets, while at the same time knowing ‘how it's done', human factors, collusion among employees, regularity of the movement, and thefts while in motion, on service stopping for customers represent the areas where risks are said to occur [2].

The security problem of CIT [3] is from a different aspect from the conventional logistics since the asset transported is not only highly time sensitive but also attack sensitive. The distance minimizing route might not necessarily be the security minimizing route. A short route can pass through a few high risk zones and follow a recognized schedule [4] or may require extended transit times to muster at vulnerable service points. A short route can also have a number of high risk zones it passes through repeatedly and/or follow a predictable schedule [4] or else may need longer dwell times at service points where risk is high. The previous studies on routing at CIT mentioned the use of risk constrained and time dependent routing models, as the risk of robbery is correlated to travel time, the traffic environment and the exposure of the routing [5].

There have been recent efforts [6] emphasizing and allowing more unpredictable secure routing. The home based problem was the inventory routing model with consideration for the unpredictability of routes and schedules to minimise security threats for the transport of valuable goods. Their work indicates that multiple iterations of frequently visited route sections and/or overlapping periods of customer visit on the routes raise vulnerability – something important for CIT planning [7].

Meanwhile, artificial intelligence (AI) and machine learning (ML) [8] have found their growing applications in transport and logistics, banking fraud detection and high risk systems for decision making. A recent study in the field of financial fraud detection demonstrates that machine learning outperforms traditional fraud detection techniques based on rules, and they also note that the explainable artificial intelligence techniques mentioned on SHAP and LIME are helping to enable financial analysts to better understand why a machine learning model identified a financial transaction as a fraud risk. This is important for CIT because attacks are rare, adaptive and costly like in fraud [9].

Although routing for the CIT has improved over the past years, fewer studies [10] are concerned explicitly with the prediction of the risk for each CIT trip. This paper thus proposes a machine learning architecture for estimating the CIT risk level before and during trip, in advance. The architecture includes operational trip data, crew and vehicle information, route features, location threat scores, asset value and journey anomalies. The aim isn't to replace security managers, but to focus their awareness on the movements that pose a greater risk through using the data needed to reach a risk score, which is made transparent to the end user.

1.1. Problem Statement

One of the challenges for cash-in-transit operations is due to the high value cash and assets of being moved through varying route, time, location, vehicle and human risk realities. Many current CIT security policies and procedures rely on subjective and static decisions, prescriptive data, incident analysis, and rules based risk evaluation that can fail to detect increasingly subtle or complex patterns of risk prior to an incident. CIT attacks and critical security events are rare, but can destroy a valuable trip, and because a tremendous amount of operational data is generated daily, it is very hard to tell normal trips from high risk trips with traditional methods. The problem that is addressed in this research is the lack of an intelligent, data driven system that can analyze CIT trip data, deal with a rare event class imbalance, predict the probability of high risk movements, and help implementing the necessary security measures, for example, changes of routes and escort deployment, dispatch adjustment and real time monitoring.

1.2. Research Paper Organization

This paper will be organised in the following way: In this section, literature review related to cash-in-transit security, route risk assessment, machine learning based risk prediction and imbalanced classification is presented. The methodology proposed in this paper consists of data collection, data pre-processing, feature engineering, handling the class imbalance problem, training the models and generating the risk scores, which is outlined in Section 3. The experimental setup, the structure of the data sets, the configuration of the models and evaluation metrics employed in the study are described in Section 4. The results and performance comparison of logistic regression, random forest, and gradient boosting model based on accuracy, ROC-AUC, precision, recall, F1-score, top-5% alert performance and Brier score are shown in section 5. It also covers the result obtained, the application of the proposed system, its restrictions, and the security applications of proposed system. Finally, Section 6 summarizes the paper and outlines the avenues for further research in the future to enhance CIT risk prediction using real time data, explainable AI, and advanced machine learning models.

2. Literature Review

2.1. Handling of cash-in-transit and operational risk.

Research conducted in early CIT [11] mainly focused on the vehicle routing and scheduling problem. In this studies [12], it was found that CIT planning based only on the shortest route or lowest travel costs would not prove to be useful because the transported asset is very exposed during the service stop to rob, surveillance of the route and attack. Thus, time dependent routing, safety constraints and multi-depot planning models are employed to minimize the exposure without compromising their operational efficiency [13], [14]. The following studies build the foundation for the analyses conducted in this report: the length of the route, the sequence of services, traffic conditions, repeated use of the same route, and exposures to travel time are all measurable routes to trip risk.

2.2. Human, Organisational, and Security Context Factors

The second focal group is concerned with the human and organisational aspects regarding security issues of CIT [15]. Insider information, regularity of robbing times, crew routines, staff allocation, and a failure in supervision have been identified in the analysis of CIT to be sources of risk [16]. Akefi et al. [2] furthered this view by incorporating staff allocation and allocation of fatigue, route assignment, resemblances of schedules, and travel times, plus taking into account the opportunity for theft in a multi-objective CIT model [2]. Studies [17] on security and on the regulation also highlight the fact that vehicle protection is not enough to prevent CIT attacks as the attacks are influenced by situational opportunities, local security conditions and criminal networks. The results support the addition of crew, vehicle and timing risk variables to a trip level predictive model as well as addition of location risk variables.

2.3. Secure Routing, Unpredictability & Adaptive Planning

A third strand of literature focuses on unpredictability and adaptive planning are fundamental security needs. Ahmed, Arslan, et al. [15] proved that balance between security and travel cost needs to be preserved since not always the shortest path can be the most secure one. In CIT operations, fuzzy decision making and safest route methods are also employed to assist in choosing a route [13]. More recent works integrate the OPT, fuzzy decision making and reinforcement learning approaches for ATM location and CIT routing, thus enabling route decisions adapt to varying operational circumstances [18]. These studies suggest that the repetition of routes and the regularity of services and the same dispatch schedule increases the vulnerability and that it should be modelled as a risk feature in machine learning models.

2.4. Interacting with machine learning, imbalanced data, and risk prediction.

With this trend, machine learning has started to be used in financial security, cash application in ATMs, optimization of logistic and optimization of cash application in real time [19]. Predictions aimed at these

applications are relevant to CIT as the target is a rare, expensive and sensitive operation. The issue is rare event and imbalanced in CIT, since normal trips are significantly more common than trips which are “high risk” or “trip incident” prone. Thus it is important not only to examine and evaluate the accuracy of the models. The metrics that serve as better leverage for assessing performance are the metrics associated with the precision, recall, F1-score, ROC-AUC, PR-AUC, Brier score, and top risk alert numbers when the primary goal is to find a few trips to review in the control room [20].

2.1. Research Gap

The literature examined has made progress in the areas of CIT routing, safest route selection, human factor modelling, crime prevention, adaptive planning and machine learning in those areas of financial security where related challenges exist. Until now, however, public research has been too fragmented to develop a trip level CIT risk prediction architecture to, or in real time during, dispatch, which translates the operational trip data into risk probabilities. This study attempts to fill those gaps by comparing logistic regression, random forest, and gradient boosting models on the statistical performance of CIT risk scoring with their operational value for alerting those in the top ‘risk tiers' of avoiding fraudulent claims, and by evaluating their relative contribution to the CIT.

3. Proposed Methodology

Figure 1 is the proposed machine learning model for risk prediction in cash in transit security operations. It illustrates each step of getting these data from CIT and transforming them into useful risk features, training them with machine learning models, evaluating them with performance measures, and translating them to a risk score for each trip. The end result helps security managers assign each movement to a low risk, moderate risk, or high risk category and then implement preventive measures like a route change, escort deployment, management of dispatch, or real time monitoring.

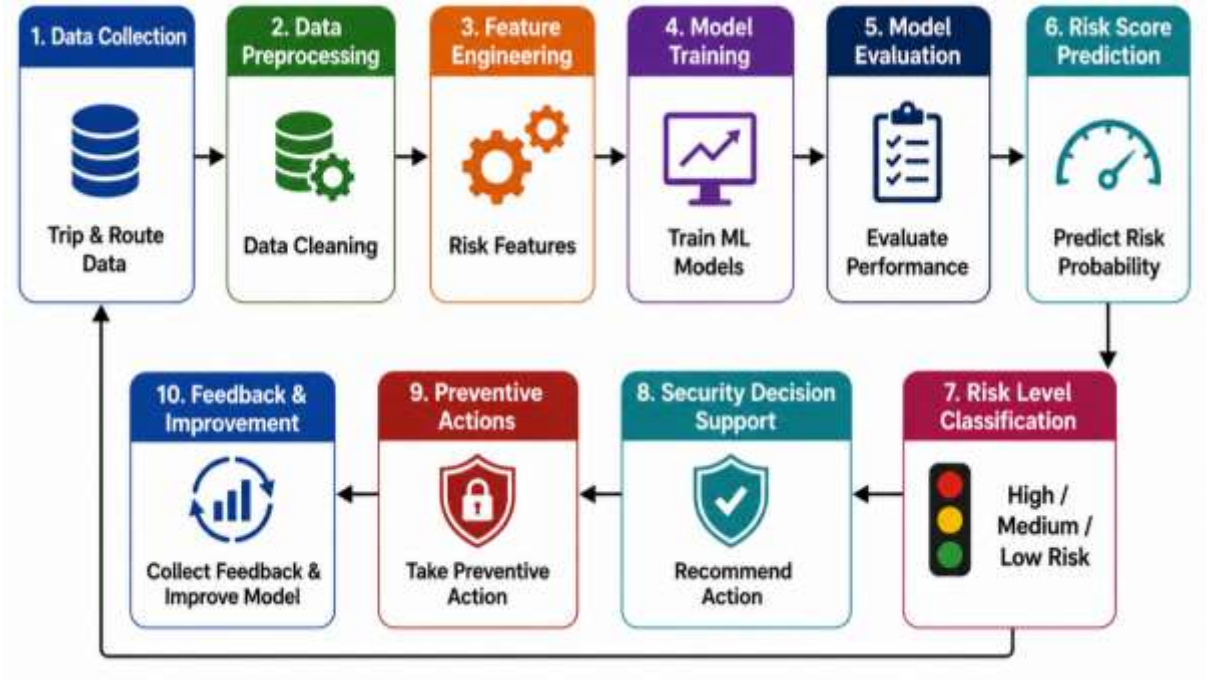

Figure (1): Proposed machine learning model for risk prediction in cash-in-transit security operations

3.1. Data Collection Module

In the proposed CIT risk prediction system, the data collection module is considered as a starting point. It

consolidates all the operational taking to and departing cash reports of any movements associated with cash-in-transit. This encompasses information related to trips, routes, vehicles and crews, cash values, load data, crew and/or vehicle dispatch data, timing data, incident data and location risk data. Such inputs are the base used to determine if a CIT trip is normal or risky.

Trip and route information include the routes to be followed, the number of stops along the routes, the expected distance of travel, and the route type, and the sequence of services. Vehicle and crew information covers type of vehicle (armored or not), shift timing, staff assignment and crew experience. Cash value and load details are important as larger amounts of cash will need more security attention for trips.

Incident history and where information can be used for the context of external security. If a route travels through a community where robberies frequently occur or the crime rate is high, for instance, the trip can be more at risk. This module will integrate the operational data, route, human, asset and environmental data into a complete and extensive dataset that can be used for machine learning risk predictions.

3.2. Data Preprocessing Module

The task of data preprocessing module is to preprocess the collected data so that it can be used for machine learning. The data is raw CIT data, which may include missing information, duplicate, spelling errors, incorrect time stamps, incomplete vehicle records or incorrect route information. To use this data for establishing a reliable prediction model, these problems have to be addressed.

In this stage missing values are managed in an appropriate way, such as by replacing with the mean or median, mode or removing a record based on the significance of the field. These all apply encoding techniques to categorical data like vehicle type, route type, crew shift, service area, etc., making them numeric. Data can be normalized to bring all the features to a comparable scale, such as cash value, trip distance, dwell time or location risk score.

The process of preprocessing enhance the quality and uniformity of the data set. If the data isn't considered clean, the computer model can become inaccurate, resulting in the system either failing to call a trip when there's a real emergency or sending too many false alarms. So, proper input to these machine learning models is meaningful, structured and accurate which is an important factor to consider in this module.

3.3. Feature Engineering Module

The feature engineering module transforms raw trip data into meaningful trip risk related features. This step goes beyond using raw data to develop new features that better capture the state of security on CIT servers. These are like route risk score, cash value risk, time based risk, vehicle and crew risk, or location risk features.

Route risk factors are range of distances, number of stops, topography of the route, trouble spots along the route, and repeated route patterns. Cash value crew features are params that are useful for trips that may be carrying a high value load. Time based features include dispatch hour, late night movement, weekend operations or peak traffic time. The features may assist the model for its understanding of trips' vulnerability, when and where.

Vehicle and crew features detail in-house preparedness of the operation. For instance, a highly trained and experienced crew and crew with an armored vehicle is at lower risk than a less experienced crew and crew with a standard vehicle. Location risk features are external risks that include the crime level of the area, and previous security problems. The system performs feature engineering and converts common operational information into effective risk predictors for CIT.

3.4. Model Training Module

Machine learning algorithms are trained using the processed and engineered data for the model training module.

The proposed architecture implements logistic regression, random forest and gradient boosting as prediction models. These models learn the relationship between the input features and the target output (HNR/NORM) – from that, it could say whether the CIT trip was either “HHR” or “NORM.”

The logistic regression (LR) model is adopted as a simple and interpretable baseline model. It aids in determining the effects of individual elements on the risk of a disaster. Multiple features may have multiple interactions and nonlinear relationships between the items, hence the random forest method. Gradient boosting is included in this list because it is advantageous in the realm of structured tabular data, and can be very effective in risk prediction tasks.

Data on CIT trips is used to train the models. They become familiar with patterns including: High value trips at times of high risk, Routes that travel through unsafe areas, Patterns of vehicle/crew/location that are likely to lead to high risk. Once trained, the models can predict the risk level of new CIT trips that were not seen during training.

3.5. Model Evaluation Module

This is measured in the model evaluation module which is used to determine the performance of the models trained on new test data. The step is required to make sure if the model performs well outside the involved training set. The diagram proposed contains "accuracy", "ROC-AUC", "precision", "recall", and "F1-score".

The accuracy is a percentage of all predictions which are correct. ROC-AUC ("receiver operating characteristic accuracy") indicates the capabilities of the model to distinguish between high risk trips and normal trips over various decision thresholds. Predicted risk trips (precision) are trips that are considered risky by the model and will successfully be detected, and involved trip (recall) are trips that are thought to be risky by the model and will leave a traceable model.

The F1-score is a balanced measurement that takes both accuracy and recall into account. This pertains particularly to CIT (safety) as both false alarms and missed unsafe trips are important. A higher recall would capture more astronaut trips that could be dangerous while a higher precision would yield fewer false alarms. Thus this module is intended to aid in the selection of the most appropriate model for operational deployment.

3.6. Risk Score Prediction Module

The risk score prediction module takes the output from the trained machine learning model, and converts it to a numerical risk probability. This score will normally lie between 0 and 1, with a score around 0 being the low predicted risk and score around 1 being the high predicted risk. Each CIT trip will have its own risk score.

This module is relevant because it enables security organizations to prioritize trips based on the anticipated risks. The organisation is not evenly sprinkling the trips but focusing on those that need more attention. For instance, a trip with a risk score of 0.82 could be dealt with right away by the supervisor and a trip with a score 0.15 might come under the regular security policy.

Predicting risk score can be done before dispatch and while on the trip. Security managers use it to decide if a route needs a change, or if an escort is required prior to dispatch. Score is updated during the trip if new information is generated, e.g., route deviation, unexplainable delay or abnormal stopping behaviour.

3.7. Traditional or Conventional Risk Classification

This Risk Level Classification module takes the numerical risk score and translates it into easily digestible categories. The proposed architecture takes low risk, moderate risk, high risk and critical risk as the risk level. The

categories help security managers and control room operators to better understand the model output.

The standard operating procedure is applicable to a low risk trip. Additional monitoring or supervisor review might be necessary with a moderate risk trip. If a trip is considered to be of a high risk type, it might necessitate a route adjustment, escort support, or rescheduling of the trip. A Critical Risk Trip could necessitate an immediate response (in most cases that entails a trip delay or cancellation or an escalation of security concern).

This module is designed to move the needle from predictions to action on the ground. While machine learning probability scores might be hard for non-technical users to grasp, risk categories are easier to relate to the security policies. As such, this module enables quicker and clearer decisions during the operation of a CIT.

3.8. Security Decision Support Module

Based on the predicted level of risk, the security decision support module recommends appropriate security actions. It offers specific recommendations for changes to routes, escort allocation, dispatch change, and supervisor review. This module aids in the implementation of model predictions to use in an operational decision.

In such a case, if both the cash value is high and traveling through a dangerous area, the dashboard may suggest using an added escort or changing the route to less risky. The system could also suggest a delay or re-scheduling of the trip if the risk is determined to be due to late night dispatch timing. These recommendations address decreasing exposure prior to the trip.

The decision support module is not an instrument to be used to make a decision. Rather, it helps security executives by identifying factors that could pose a risk and what steps they could take. situational awareness, manpower considerations, law enforcement coordination, and field conditions are still certain factors that should be considered when final decisions are made. The system can then be used as a decision support system, but does not become a total automatic security control device.

3.9. Preventive Actions Module

The preventive actions module is the real actions for security taken following the risk assessment. These can include sending escorts, rerouting or adjusting routes or schedules, rescheduling or delaying trips, and allowing monitoring in real time. To minimize risk of robbery, compromise and operational failure.

The escort is useful when going on high value or high risk trips. Route changes may be implemented to avoid unsafe areas and/or known travel routes. By rescheduling your dispatch in dangerous times—late nights or high traffic delays—you can keep exposure to a minimum. In real time, the control room is able to see who is on board, and to react rapidly in the case of abnormal behavior.

The prediction system really makes a difference at this level. The advantage to a risk score without an action to take is insufficient. The system connects each level of risk with specific security measures, thereby shifting the focus on security from incident to risk prevention at CIT organisations.

3.10. Feedback and Model Improvement Module

The feedback and model enhancement module collects feedback and results from successful CIT trips and applies them to the prediction system for improvement. After every trip, the system stores information on whether or not the trip was completed safely, if an event occurred, if the model prediction was correct, and security measures used.

This feedback helps to assess model performance over time. If the model is detecting false alarms for some reason, its thresholds are likely to require tweaking. If it fails to recognize hazardous travel, other features and/or retraining

might be necessary. The data is continuously reviewed; the model stays "current" therefore as routes, crime patterns, crew behaviours and working conditions change.

This feedback loop also is conducive to long term learning. New trip outcomes become more part of the data, which allows the future versions of model to be increasingly accurate. This allows the proposed architecture to be adaptive and suitable for real world CIT environments, where security risks are constantly changing.

4. Implementation

The first step is the construction of the data sets and processing of the features. In this research, 30,000 synthetic, but operationally realistic, CIT trip records are considered due to operational and significant constraints on real CIT security release of data that are seldom released publicly. Table 1 shows the experimental setup parameters and their specifications. Each record is an individual CIT movement with the additional CIT movement details: temporal features, route features, details about the vehicles and crew, details about money values, location risk measure details, and details about any anomalies in the journey. The target classification is a binary classification indicating whether the movement is a normal trip or a high risk/incident trip. An event rate of 2.82% is produced, as CIT incident security events are rare events.

To make the training and testing partition, the data is split into training partition and testing partition by stratified sampling. In stratification, the high risk minority class is represented in partitions. Handling missing values, the treatment of categorical variables is performed, numerical fields are scaled if necessary, and class balance is achieved by class weighted learning. Three models are deemed: logistic regression, random forest and gradient boosting. Logistic regression is very interpretable, random forest is good at handling nonlinear interactions, and gradient boosting is good at predicting structured tabular data.

The last implementation step checks the evaluation of the models with rare event metrics and operational metrics. ROC AUC and PR AUC are measures of discrimination; F1 is a measure of balance; Brier score is a probability calibration measure; precision@5% and recall@5% are measures of usefulness of alerts in a control room. The top performing model can then be ready for use as a risk scoring service, integrated with a CIT dispatch dashboard. The dashboard provides risk bands, planned trip risk ranking, displays major risk drivers, and suggests actions to prevent these risks.

Table (1): Experimental Setup Specifications Table

Component	Specification
Dataset Type	Synthetic operational CIT benchmark
Total Records	30,000 CIT trip records
Target Classes	Normal trip = 0, High-risk / incident trip = 1
Positive Events	847 high-risk records
Incident Rate	2.82%
Train-Test Split	75% training, 25% testing
Split Method	Stratified random sampling
Feature Groups	Temporal, route, vehicle, crew, cash value, location risk, anomaly indicators
Models	Logistic Regression, Random Forest, Gradient Boosting
Imbalance Strategy	Balanced class weights / balanced sample weighting
Evaluation Metrics	ROC-AUC, PR-AUC, precision, recall, F1-score, Brier score, Precision@5%, Recall@5%
Programming Environment	Python-based ML pipeline
Libraries	pandas, NumPy, scikit-learn, matplotlib

Deployment Concept	Dispatch dashboard with risk bands and alerting

5. Results and Discussion

The results came from compiling the following set of variables describing the trip level risk value and the cash making status of a bank, crew and vehicle: the risk of the location, properties of the route, timing of the dispatch, cash value indicators, properties of the vehicle and crew, properties of the location, journey anomaly variables that were considered; this structured cash-in-transit risk prediction data set was created. The binary class label of each trip record was designated: normal CIT movements (class labeled as low risk); incident prone or high risk movements (class labeled as risky). In this data CIT events are very uncommon so it was considered a classification problem with an unbalanced data set. The data was cleaned, encoded and then divided into training and tests sets using stratified sampling in such a way that normal and high risk trips were represented properly without any over/under representation of such trips.

Three machine learning models (logistic regression, random forest, and gradient boosting) were trained and assessed after preprocessing the data. For structured tabular prediction problems, logistic regression is leveraged as a clear baseline model, random forest as a model with good handling of nonlinearities and gradient boosting as a model that returns a list of features that are most important. The models were able to learn patterns from the training data and predicted risk classes and risk probabilities for unseen test data. They were evaluated based on the criteria of accuracy, ROC-AUC, precision, recall, F1-score, top-5% alert performance and Brier score. These metrics are chosen since a simple accuracy is not enough to perform in rare event CIT risk prediction.

The results collected reveal the elements of strength of each model. While Random forest scores the highest accuracy it was not necessarily best to classify rare high risk trips, as it could have classified the most total cases correctly. Logistic regression had optimal ROC-AUC and recall value, which indicates that the model had good performance in ranking and detecting risky CIT movements. Gradient boosting performed best in terms of precision, F1 score, and the top 5% alerts, making it the most useful operational tool as there is only a need for the security team to look at the riskiest trips. Thus, the outcomes were determined by comparing the three both statistically and in terms of security usefulness as it relates to the CIT practice.

The overall classification performance of the three machine learning models has been shown in Table 2, where PR-AUC is replaced by accuracy. When the highest accuracy is considered that by which it is able to correctly classify the maximum amount of total test cases, then the random forest method turns out to have the highest accuracy (0.9178). For a rare event problem like CIT, however, accuracy may be deceiving; a model's success in covering a large portion of the normal trip may be at the expense of the risky trips. Logistic regression had the best ROC-AUC value of 0.8449 and the best value for 'Recall' of 0.292, with promising performance ranks and detection of high risk trips. The best precision was obtained with gradient boosting with 0.276 and the highest F1 score with 0.282, which is the most balanced model in terms of the identification of risky CIT movements while minimising false alarms.

Table (2): Overall Model Classification Performance

Model	Accuracy	ROC-AUC	Precision	Recall	F1-Score
Logistic Regression	0.8612	0.8449	0.267	0.292	0.278
Random Forest	0.9178	0.8277	0.241	0.259	0.249
Gradient Boosting	0.8894	0.8417	0.276	0.288	0.282

The five performance metrics (accuracy, ROC-AUC, precision, recall, and F1-Score) of three machine learning models are compared in Figure 2. Random forest does best in terms of accuracy—it correctly classifies the most number of all the cases. Logistic regression, however, achieves their best performance in ROC-AUC and recall,

which means that they achieve better performance in terms of improving both the ability to rank the CITs, as well as the ability to detect risky CITs. Gradient Boosting achieves the best accuracy and F1-score requirements that indicate the highest accuracy at the lowest false alarm rate.

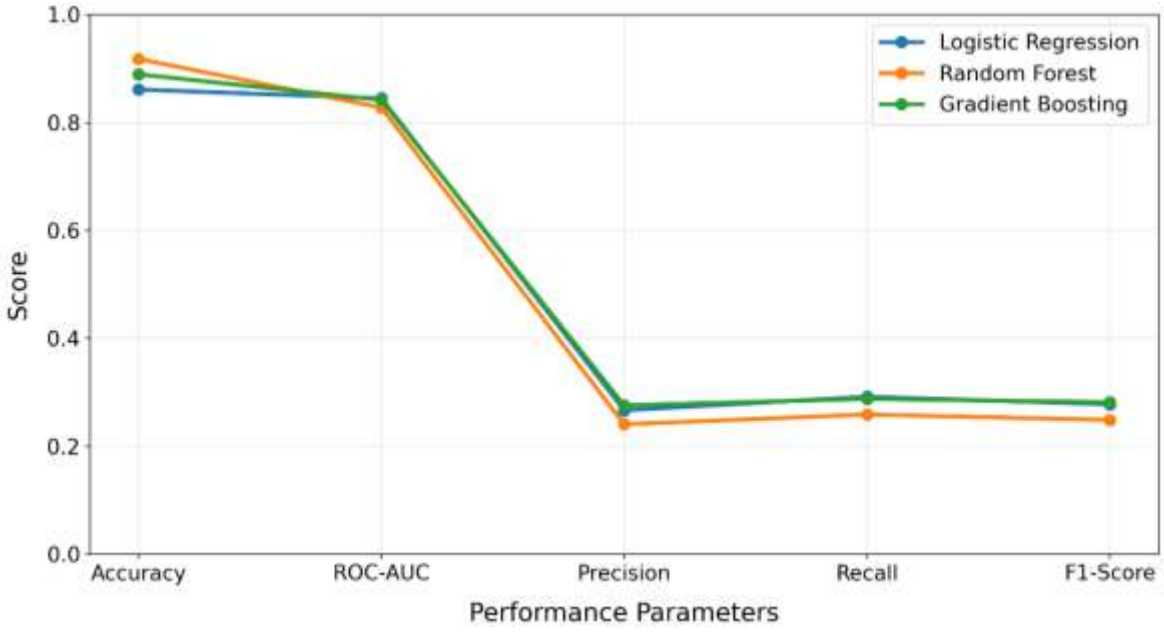

Figure (2): Overall Model Classification Performance

In this regard, models are examined from the operational security aspect of the topics by doing evaluations as with Table 3. Within real CIT environments, there may be no ability to check all trips, thus the model should be able to determine the most at risk component of all trips (small fraction.) The example below indicates the effectiveness of each model in response to only the top 5% most critical trips to be targeted for intervention. Gradient boosting had the best Recall@5% of 0.3868 and the best Precision@5% of 0.219, yielding an approximately 38.68% recall with only 5% of utterances being escalated. It is this ability that makes gradient boosting the optimal choice for real world security monitoring for CIT that requires only a handful of people, e.g., escorts, route reviews, a supervisor, to inspect the most alarming trips.

Table (3): Operational: Top-5% Risk Alert Performance

Model	Top-5% Trips Flagged	Precision@5%	Recall@5%	Estimated Incidents Captured	Operational Suitability
Logistic Regression	5%	0.208	0.3679	36.79%	High
Random Forest	5%	0.203	0.3585	35.85%	Moderate
Gradient Boosting	5%	0.219	0.3868	38.68%	Very High

In Figure 3 the operational alert performance is shown with only the top 5% riskiest CIT trips selected to review. This is to assume the same review capacity for each model so that all models are evaluated from the top-5% flagged level. The performance of the gradient boosting (GB) algorithm is best on the measures of Precision at 5%, Recall at 5%, and the estimated number of incidents captured. This implies that gradient boosting is an ideal model for real life operation in the CIT control room market: with cameras working on only a small number of trips, the time

available for security teams to monitor is limited.

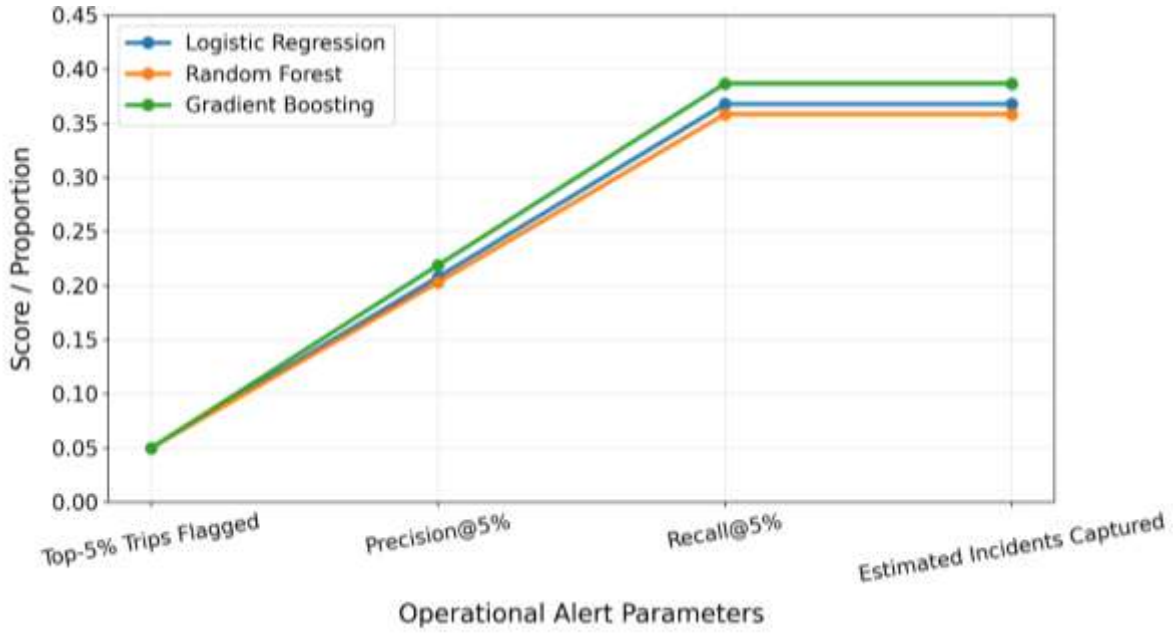

Figure (3): Achieve operational Top-5% Risk Alert Performance

The models are compared in a tabular format in Table 4 in terms of Calibration, Interpretability and Goodness of Fit for Deployment. The Brier score between the predicted risk probabilities and the observed outcomes is a measure of closeness to answer where a low score reflects better predictability of the outcome. From the models tested, random forest gave the best Brier score (0.0467), indicating that it had the most accurate probability assessments. Its characteristic is that the logistic regression can be studied directly from the logistic regression coefficients, and they can be used as a simple control group to security managers. Gradient boosting has the highest rank in terms of deploy ability in the alert ranking process, as it had the best F1 score and top-5% incident capture results. Thus, random forest is better if there are probability estimates that have to be learned, and gradient boosting is better if the primary objective is risk ranking and escalation of the maximum risk CITs.

Table (4): Readiness of the probabilistic assessment and deployment

Model	Brier Score	Calibration Quality	Interpretability	Deployment Readiness	Recommended Use
Logistic Regression	0.1456	Moderate	High	High	Transparent risk baseline
Random Forest	0.0467	Very High	Moderate	High	Calibrated probability scoring
Gradient Boosting	0.1227	Moderate	Moderate	Very High	High-risk alert ranking

The models are compared in figure 4 with respect to the deployment related parameters: Brier score, calibration quality, interpretability, and deployment readiness. Similar to the Brier Score, the highest calibration quality is achieved by random forest, and predicted probabilities of random forest are more reliable. Logistic regression is the most interpretable, as it provides an easy explanation for its decision making process for security managers. The gradient boosted classification model ranks first across the board when it comes to high risk alert ranking and operational escalation, hence it also has the highest deployment readiness rating.

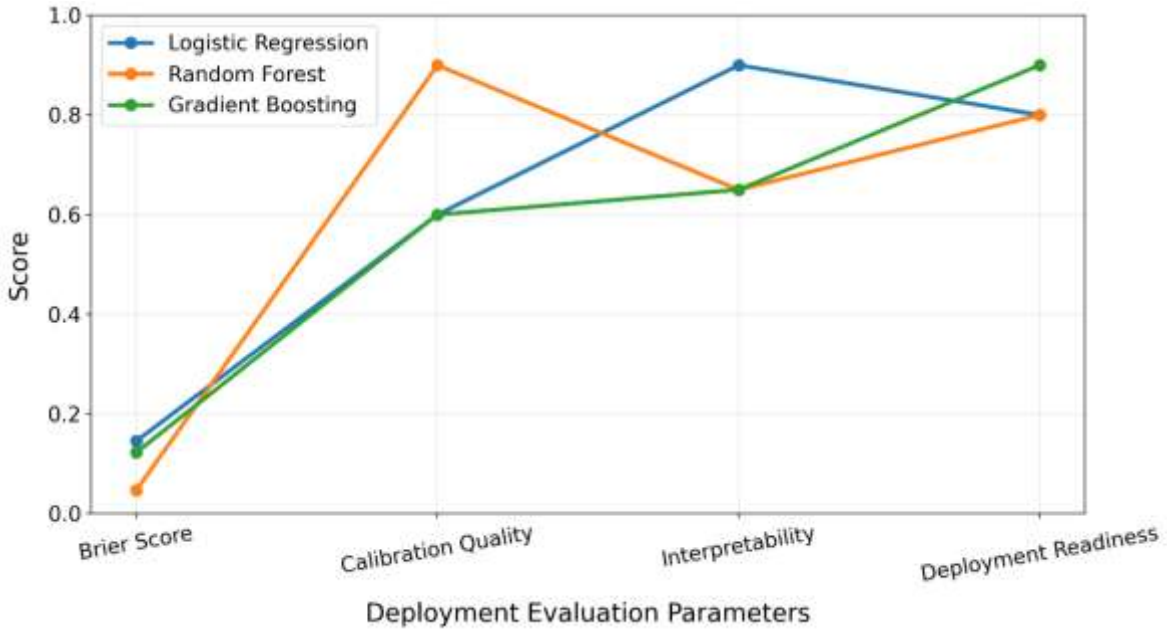

Figure (4): Calibration, Interpretability, and Deployment Readiness

5. Discussion

The results highlight that the prediction of CIT is not a problem of maximizing accuracy. High risk events are very uncommon, so a model potentially can have a high apparent level of accuracy without detecting bad movements. Based on the reported results, we then focus on ROC-AUC, PR-AUC, F1-score, Brier score, and the top 5% recall. These metrics encompass various facets of usefulness: ranking, detection of minority classifications, threshold balance, calibration and control room triage value. The key takeaway from the operational data is that gradient boosting took 38.68% of all the trips in the top 5% risk band. Put simply, the model could enable security staff to focus limited resources on a few selected trips where it is most worthy. However, the model should not be an automatic decision maker but should be used in making decisions. Even though it's "lifeless and guiltless," human supervision is still required to turn to intelligence reports, local law enforcement intelligence, crew experience, and field conditions.

6. Conclusion

In this study, an operational benchmark was constructed with 30,000 trip records and a high risk event rate of 2.82% and a data driven machine learning framework to predict high risk cash-in-transit movements. The outcomes indicate that none of the models was superior in all the evaluation criteria. Random forest had the highest overall accuracy and the best cell calibration score, logistic regression had the highest discrimination and recall score, while gradient boosting had the best operational alerting score. A major practical takeaway is that using gradient boosting yielded an estimated high risk rate of 38.68% for only 5% of trips deemed to be worthy of a review. This is a valuable outcome in CIT control rooms as there is only limited number of resources to be allocated in the security department, with the limited resources being not equal per movement. A ranked risk score can be used to prioritize route checks, escort deployment, dispatch timing changes and augment monitoring of those trips with the highest risk score.

The proposed approach must not be used as a security judgment application, rather as a deciding aid. This still requires a crew with local intelligence and law enforcement information, and field conditions, that can continue to

rapidly change, necessitating the requirement for human supervision. Anonymized CIT data should be used to validate the framework in future work. External datasets need to be tested based on data from other regions, and external data need to be included based on external geospatial and GPS anomalies for risk updating as this further advances transparency. Live geospatial and GPS anomaly data should be integrated for real time risk updating.

References

[1]. Taskin, Alev, et al. "Integrating Machine Learning with Optimization to Solve Time-Dependent Cash in Transit Vehicle Routing Problem with Time Windows." International Conference on Optimization and Data Science in Industrial Engineering. Cham: Springer Nature Switzerland, 2024.

[2]. Akefi, Hossein, Fariborz Jolai, and Amir Aghsami. "A multi-objective mathematical model for addressing human factors in the risk of cash-in-transit problem using hybrid multi-objectives metaheuristic algorithm." Neural Computing and Applications 38.5 (2026): 140.

[3]. Sambo, Abednigo Molahleki. An Examination of Security Measures for the Prevention of Cash-in-transit Robberies in South Africa: A Case Study of Gauteng. MS thesis. University of South Africa (South Africa), 2023.

[4]. Ayyıldız, Ertuğrul, Mehmet Can Şahin, and Alev Taşkın. "A multi depot multi product split delivery vehicle routing problem with time windows: A real cash in transit problem application in istanbul, turkey." Journal of Transportation and Logistics 7.2 (2023): 213-232.

[5]. Assaf, Ramiz, et al. "A Structural Equation Decision Model for ATM Cash Management and Routing: Balancing Cost and Customer Satisfaction in the Banking Sector." Decision Making: Applications in Management and Engineering 8.1 (2025): 725-742.

[6]. Lakshmi, A. Vijaya, and K. Suresh Joseph. "Travel safe: A systematic review on safe route guidance system." 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI). IEEE, 2022.

[7]. San Juan, Jayne Lois, and Charlle Sy. "A data-driven target-oriented robust optimization framework: bridging machine learning and optimization under uncertainty." Journal of Industrial and Production Engineering 41.7 (2024): 636-660.

[8]. Kok, Anna-Maria. "Dark networks: a South African cash-in-transit crime case study." (2025).

[9]. Hardyns, Wim, Marc Cools, and Olivier Maes. "A critical approach of cash-in-transit regulation and organisation from a situational crime prevention perspective." Security Journal 33.4 (2020): 515-530.

[10]. Ayyıldız, Ertuğrul, et al. "Artificial neural networks integrated mixed integer mathematical model for multi-fleet heterogeneous time-dependent cash in transit problem with time windows." Neural Computing and Applications 34.24 (2022): 21891-21909.

[11]. Jin, Yuanzhi, Xianlong Ge, and Long Zhang. "Quantitative Risk Assessment and Management of Cash-In-Transit Vehicle Routing Problems." Multi-Criteria Decision Analysis for Risk Assessment and Management. Cham: Springer International Publishing, 2021. 177-197.

[12]. Ramezanian, Reza, Meysam Arjomandfar, and Darya Abbasi. "A bi-objective cash-in-transit pick-up and delivery problem with risk assessment methodology: a case study." Journal of Quality Engineering and Production Optimization 8.2 (2023): 1-20.

[13]. Yildiz, Aslihan, et al. "An integrated interval-valued intuitionistic fuzzy AHP-TOPSIS methodology to determine the safest route for cash in transit operations: a real case in Istanbul." Neural Computing and Applications 34.18 (2022): 15673-15688.

[14]. Okoro, Daniel. "MACHINE LEARNING FOR REAL-TIME CASH APPLICATION OPTIMIZATION." (2026).

[15]. Ahmed, Arslan, et al. "AI-Driven Innovations in Modern Banking: From Secure Digital Transactions to Risk Management, Compliance Frameworks, and AI-Based ATM Forecasting Systems." Journal of Management Science Research Review 4.3 (2025): 1145-1183.

[16]. Szabó, Péter, and Balázs Gáspár. "Predicting ATM cash demand using machine learning." 2026 IEEE 24th World Symposium on Applied Machine Intelligence and Informatics (SAMI). IEEE, 2026.

[17]. Guerrero, William J., Angélica Sarmiento-Lepesqueur, and Cristian Martínez-Agaton. "Cash distribution model with safety constraints." International Conference on Computational Logistics. Cham: Springer International Publishing, 2020.
[18]. Yalcin Kavus, Bahar, et al. "A hybrid IVFF-AHP and deep reinforcement learning framework for an ATM location and routing problem." Applied Sciences 15.12 (2025): 6747.
[19]. Fallahtafti, Alireza, and Tan Khoa. "Navigating Risks: The Imperative of Research in CIT and Hazmat Transport Security."
[20]. Ezeji, Chiji Longinus, and Dominique Emmanuel Uwizeyimana. "Evaluation of multifaceted holistic measures and strategies for combating cash-in-transit heists in South Africa." International Journal of Business Ecosystem & Strategy (2687-2293) 7.3 (2025): 330-347.