Using Predictive Analytics in Logistics Industry to Associate Tolls and Damages to Customers

ABSTRACT

Predictive analytics has the potential to transform cost management and damage tracking in the logistics industry by integrating telematics, toll records, and automated damage assessment systems. This study focuses on designing a predictive system to streamline cost attribution, improve operational transparency, and reduce billing disputes for a trailer-as-a-service provider. The importance of this work lies in its ability to address inefficiencies caused by manual processes, data gaps, and a lack of real-time insights. Using machine learning models and Al-powered image recognition, we developed a framework that assigns costs to specific drivers, lanes, or trailers while automating damage detection. The findings indicate that the integration of predictive tools not only enhances operational efficiency but also builds customer trust through accurate billing and transparent processes.

Keywords: predictive modeling, computer vision, telematics, cost attribution, logistics, AI damage detection, billing transparency

INTRODUCTION

The global third-party logistics (3PL) market size is anticipated to grow from USD 1.08 trillion to USD 2.23 trillion in 10 years (The Brainy Insights, 2024). The market will experience rapid growth due to technological advancements in third-party logistics (3PL) during the forecast period (Yahoo Finance, 2024). One of the service third party logistics provide is transportation service, where the company own their trailer and distribute goods for the customer. For asset-based 3PL, fleet management could be very complicated. There are 3 types of trailer management -- traditional leasing, lease-to-own or direct ownership. However, direct owning trailer assets demand huge initial investment and maintenance cost, which makes capacity very non-flexible, alongside with potential loss on asset downtime. Some 3PL opt for traditional leasing to avoid the maintenance cost, but it failed to meet the flexibility on peak and low-capacity periods. As a result, traditional 3PL logistics business model face challenges on balancing capacity and efficiency.

Trailer as a Service (TaaS) provides a flexible solution for this challenge. Companies subscribe for capacity on the trailer, and TaaS takes care of maintenance. This solves the pain point of non-flexible, seasonal capacity and huge capital investment, and expand the opportunities to not only 3PL but brokers and carriers. While providing solutions on capacity, inaccurate billing and untracked operational cost of tolls and damages expenses become challenges for our client running the TaaS service. As the subscribers could utilize the trailer by assigning to different carriers, tracking tolls and damage expenses become challenging, given our client only has subscribers and asset list rather than the actual parties operate the trailers. This could lead to inaccurate billing and could further strained relationships with customers and carriers.

Predictive Model & Data Flow for Toll and Damage Attribution

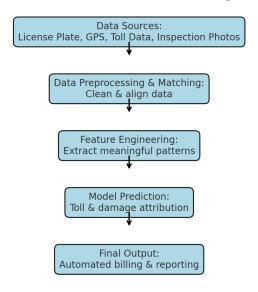


Figure 1. Predictive Model & Data Flow

In our study, the motivating business problem this paper focuses on is to build efficient workflow and model help attributing the tolls and damage to the appropriate party in a timely manner without relying on data provided from service subscribers.

Our research focuses on:

- 1. What approaches from predictive modeling along with data analytics systems should be adopted to enhance both accuracy and efficiency of toll attribution operations within TaaS systems?
- 2. Establish methods that enhance trailer damage monitoring together with attribution processes despite existing inspection irregularities and trailer exchange frequency.

The analysis will build predictive tools to automate toll expense allocation and damage inspection functions that support operational requirements of the business. Our research will review different techniques to pair toll payments with proper customers despite restricted data availability through the combination of license plate recordings timestamp records and GPS data.

An algorithm needs to be developed to process large datasets effectively throughout making correct predictions with limited available information. Our research investigates the application of computer vision technology for pre-trip and post-trip inspection photo comparison automation which helps detect possible manual inspection misses of vehicle damages. With integrating data from tolls agency and GPS data, our client will be able to automate the tolls attribution and bill the damage expense on the appropriate parties, which further reduced unnecessary overhead.

The following paper structure includes: Section 2 contains a literature review about transportation industry predictive models alongside damage detection techniques and toll attribution methods. Our proposed methodology structure comprises data preprocessing alongside feature engineering and model selection which appears in Section 3. Our experimental results with model descriptions appear in Section 4 following our method implementation description. The evaluation of our models' performance and practical applications and possible enhancement opportunities takes place in Section 5. Section 6 delivers the paper's summary including our study findings along with their effects on TaaS business and prospective research paths. The research we conduct addresses crucial operational challenges of the TaaS model with the goal to boost digital transformation within the transportation and logistics industry through efficient and transparent operations.

LITERATURE REVIEW

The use of transfer learning on damage detection on containers has been research widely, especially for MobileNetV2 model. This has been concluded by (Zixin Wang et al., 2021) and (Pavel Cimili, et al., 2022) that MobileNetV2 has advantages in multiple class of damage detection and it is useful in large scale container inspection scenarios. The former compared transfer and semi-supervised approaches by testing two separate models (for the lower and the upper part), and it was concluded that semi-supervised training could not outperform transfer learning model due to the complex structure of the trailer surface and its defects. They suggest that future improvement is needed for the classification of multiple damage classes.

The potential of MobileNetV2 model for multiple types of damage detection is still been researched. Zixin Wang et al., 2021 proposes a multitype damage detection model for containers based on MobileNetV2. They performed on-stie experiment on deploying model on the mobile terminal obtains images through the smartphone camera for real-time damage detection. They concluded that experiment results show that the multitype container damage detection model can give the corresponding damage types and prediction results. However, it is still necessary to quantify the degree of damage to the container according to the severity of the injury to the container and support the intelligent decision-making of container damage.

This provides the groundwork for our paper as we will focusing on attributing tolls cost based on geographical data and identifying damage classification. Our goal is integrating this information and match them to the appropriate party.

We summarize our findings in Tables 1 and Tables 2.

Study	Insights	Research Gap
(Sheng Xu et al., 2020)	They introduced a decision support platform based on shared logistic platform, which can be used to collect and integrate data in the process of time-driven activity-based costing.	
(Pavel Cimili, et al., 2022)	MobileNetV2 based on transfer learning is capable of damage detection if there is enough training data.	The model has the potential to applied on not just for binary classification but also for multiple class damage detection.

(Jiahao Chen et al., n.d.)	This paper proposes an improvement to the YOLOv5 model based on the Transformer self-attention mechanism for container damage detection, demonstrating superior performance compared to commonly used object detection algorithms.	Assessing the severity of multiple damaged areas in containers still requires enhancement.
(Zixin Wang et al., 2021)	This paper proposes a multitype damage detection model for containers based on MobileNetV2, which has excellent advantages in largescale container inspection scenarios.	A real-time monitoring system will needed to be developed based on port IP network cameras and integrated into the port management system. Also, it is necessary to quantify the severity of damage to the container.

Table 1: Key papers and identified research gaps

	Paper Aspect				
Study	Shared Logistics Platform	MobileNetV 2	Semi- Supervised Learning	Multiple Type damage classification	Model Enhancement
(Sheng Xu et al., 2020)	Yes				
(Pavel Cimili, et al., 2022)		Yes	Yes		
(Jiahao Chen et al., n.d.)				Yes	Yes
(Zixin Wang et al., 2021)		Yes		Yes	
Our Study		Yes	Yes		

Table 2: Relation of our study to other academic papers

DATA

Due to the sensitive nature of the data and confidentiality constraints imposed by the data provider, we did not use real-world raw data for our analysis. Instead, we employed large language models (LLMs) to generate synthetic datasets that closely mimic the structure and semantics of the original data. The synthetic dataset preserved the integrity of the three primary data categories—Asset Location, Toll Records, and Inspection Records—while ensuring that no personally identifiable or proprietary information was exposed. This allowed us to simulate real-world scenarios while adhering to strict data privacy requirements.

The dataset includes three main tables: Asset Location, Inspections, and Tolls.

- Asset Location: Tracks trailer information like unique identifiers (asset_vin), geospatial data (position), current motion status (asset_motion_status), and time stamps (reported_time, created time). It also includes provider data (telematic provider).
- Inspections: Contains details about inspections with variables such as inspection ID (id), trailer ID (vin), inspection type (inspection_type), and time stamps (start_time, end_time). It also stores metadata like user IDs (created by, updated by) and inspection data in JSON format.
- **Tolls**: Records toll event data, including date and time (Posted_Date, Invoice_Date), toll detection method (Read_Type), vehicle ID (Device_Plate_Id), toll charges (Toll_Charge), and plaza information (Entry_Plaza, Exit_Plaza). It also tracks disputes (Dispute_Status, Dispute_Reason) and account associations.

Variable	Туре	Description
asset_vin	Categorical	Unique identifier assigned to each trailer
position	Geospatial	Geospatial coordinates indicating the trailer's location
location	Text	Corresponding address derived from geospatial data
asset_motion_status	Categorical	Current motion status of the trailer
asset_name	Categorical	Name assigned to the trailer within the organization
organization_anon	Categorical	Anonymized name of the subscriber
telematic_provider	Categorical	Service provider responsible for geospatial and telematics data
reported_time	Time	Timestamp reflecting registering time
created_time	Time	Timestamp indicating when the row was created in the dataset

Table 3: Asset Location

Variable	Туре	Description
created_by	Categorical	User ID of the person who created the inspection record
updated_by	Categorical	User ID of the person who last updated the inspection record
created_time	Time	Timestamp when the inspection record was created
updated_time	Time	Timestamp when the inspection record was last updated
id	ID	Unique identifier for the inspection record
configuration_id	ID	Identifier for the trailer's configuration
data	JSON	Contains detailed inspection data in JSON format
vin	Categorical	Vehicle Identification Number (VIN) of the trailer
inspection_type	Categorical	Indicates whether the inspection is a Pre or Post trip
organization_id	Categorical	Identifier for the organization performing the inspection
document_id	ID	Unique identifier for associated documents
trailer_number	Categorical	Trailer's company-assigned number, mapped to asset_id
start_time	Time	Inspection start time
end_time	Time	Inspection end time
inspection_number	ID	Unique number assigned to the inspection

Table 4: Inspections

Variable	Туре	Description
Posted_Date	Time	Date the toll was incurred
Invoice_Date	Time	Date the toll was billed
Source	Text	Source system or method of toll detection (if applicable)
Read_Type	Categorical	Type of signal or method detected by the toll pass
Transponder_Status	Categorical	Status of the vehicle's transponder at the time of detection
Device_Plate_Id	Categorical	License plate number of the vehicle
Vehicle_Number	Categorical	Company-assigned vehicle ID; Corresponds to asset_id in Asset Location and trailer_number in Inspections
Agency	Categorical	Tolling agency responsible for processing the toll
Entry_Plaza	Text	Name or identifier of the plaza where the vehicle entered
Entry_Date	Time	Date and time the vehicle entered the toll plaza
Exit_Plaza	Text	Name or identifier of the plaza where the vehicle exited
Exit_Date	Time	Date and time the vehicle exited the toll plaza

Class	Categorical	Classification of the vehicle type
Miles	Float	Distance traveled between entry and exit plazas (if available)
Toll_Charge	Numeric	Monetary amount charged for the toll
Dispute_Status	Text	Status of any dispute raised (if applicable)
Dispute_Reason	Text	Reason for the toll dispute (if applicable)
Account	Categorical	Account to which the toll charge is assigned

Table 5: Tolls

METHODOLOGY

This study presents the development of automated frameworks for toll and damage attribution within the Trailer-as-a-Service (TaaS) model, integrating rule-based logic, Al-driven image analysis, and proposing predictive modeling enhancements to improve operational efficiency and accountability.

The toll attribution framework was designed to automate the identification and allocation of toll charges to the appropriate customer, thereby minimizing manual intervention and reducing billing disputes. The process leverages two primary data sources: daily transaction-level toll reports received from toll agencies, which include timestamps, vehicle identifiers, and toll amounts; and trailer subscription records detailing customer usage periods. Upon ingestion, an automation engine aligns toll transactions with customer subscriptions by correlating the trailer's Vehicle Identification Number (VIN) and timestamp data. This matching process is executed through a daily batch update to the internal database. Once attribution is complete, automated notifications summarizing toll charges are generated and dispatched to customers. This rule-based approach ensures timely, accurate, and auditable toll cost allocation.

For damage attribution, a scalable two-tiered framework was implemented to address the complexities of assigning responsibility for trailer damages. The first component involves Al-driven damage analysis, wherein inspection images are processed using a MobileNetV2-based convolutional neural network (CNN). This model classifies the type, location, and severity of damage, providing an objective validation layer, particularly valuable in cases of dispute or uncertainty. The second component employs deterministic, rule-based logic triggered upon receipt of inspection reports indicating damage. The system searches for prior inspection records to establish the timeframe during which the damage likely occurred. It then queries telematics data to identify customers who operated the trailer within this period. If a single customer is identified, responsibility is directly attributed; if multiple customers are detected, a list of potential responsible parties is generated. In instances where historical data is insufficient, cases are flagged for manual review. The outcome is a structured report consolidating customer usage data, telematics insights, and Al-generated damage assessments.

To further enhance attribution accuracy in complex scenarios, a predictive modeling approach is proposed. This involves integrating external environmental data sourced from real-time APIs, sensor-derived metrics such as vehicle overload indicators, and behavioral patterns including trip frequency, time-of-day usage, and historical driver behavior. These features would inform a machine learning model designed to predict the most likely customer responsible for operational incidents. A standard 70/30 traintest data partitioning strategy, combined with k-fold cross-validation, is recommended to ensure robust

model evaluation. Performance would be assessed using precision, recall, and F1-score metrics, given the critical importance of minimizing false attributions in operational contexts.

MODEL

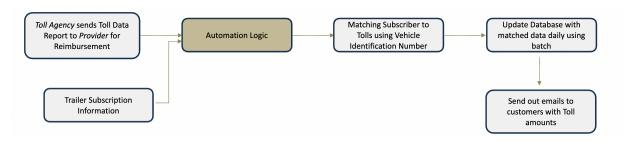


Figure 1. Toll Model Design

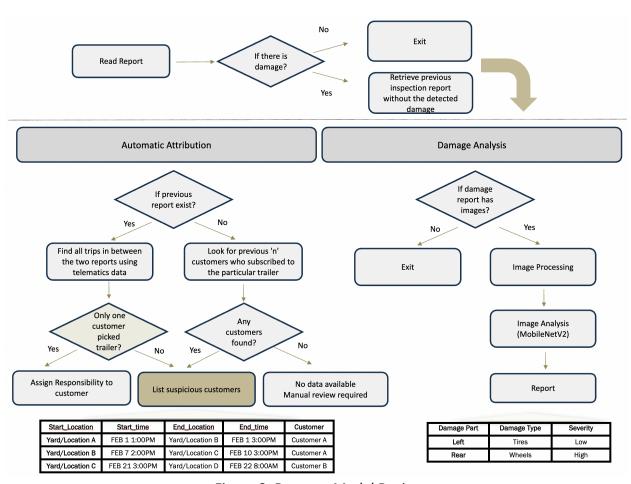


Figure 2. Damage Model Design

This project employs a hybrid methodology combining rule-based automation, Al-powered image recognition, and proposes the future integration of predictive modeling techniques to enhance decision-making processes within the TaaS framework.

The rule-based automation logic forms the foundation of both the toll and damage attribution systems. These deterministic engines utilize timestamp alignment, VIN matching, and subscription data to automate attribution tasks. The primary advantages of this approach include high transparency, ease of auditing, and rapid execution. However, its effectiveness is contingent upon the completeness and accuracy of input data, and it offers limited flexibility in ambiguous cases where deterministic rules cannot conclusively assign responsibility.

For image-based damage classification, the MobileNetV2 convolutional neural network was deployed due to its balance of computational efficiency and classification accuracy. MobileNetV2 is well-suited for industrial applications requiring scalable, real-time image processing. It effectively identifies damage type, location, and severity from inspection images. While the model performs well under standard conditions, it may exhibit limitations when processing highly complex or low-quality images compared to more resource-intensive CNN architectures. The model was fine-tuned using domain-specific datasets, with key hyperparameters such as image resolution, dropout rates, and learning rates optimized to improve performance within the operational environment. MobileNetV2 was selected for its proven capability in lightweight image recognition tasks and its seamless integration into automated workflows.

In addressing cases where rule-based systems are insufficient, the proposed predictive modeling component recommends the use of ensemble methods, such as Random Forest or Gradient Boosting Machines (GBM). These models are robust in handling heterogeneous data types and capturing non-linear relationships inherent in operational and behavioral datasets. Key features for this model would include environmental conditions, telematics data, and customer usage patterns. Hyperparameters such as the number of trees, maximum depth, and learning rate would be systematically tuned to optimize predictive performance. The adoption of predictive modeling provides a data-driven decision support mechanism, enhancing attribution accuracy by leveraging contextual insights where deterministic logic falls short.

EXPECTED RESULTS

In the logistics industry, approximately 10% of the total cost of each trip comes from tolls (1.5%), damages (8%), and maintenance in total (ATRI, 2024). While this may appear to be a relatively minor component, the operational inefficiencies tied to manually managing these costs can significantly impact the bottom line. Before the introduction of the automation process, the review of toll and damage reports has been a time-consuming and labor-intensive task requiring human work comparing images, verifying charges, and matching records to the appropriate trailers and clients. This process not only require a significant amount of time, but it is also prone to human error, leading to inaccurate attributions, delayed billing, and compromised compliance.

To better understand the scope of this inefficiency, we considered a scenario with approximately 1,000 trailers in active operation. Assuming each trailer completes an average of four trips per week and each trip generate one toll charge and requires one damage report comparison; there will be roughly 4,000 toll charges and 4000 times of comparison works per week. We assume that after each trip, the manual review would take an average of two minutes for tolls and damage report comparison, in total the manual review process takes around 133 hours of labor weekly. At a standard labor cost of \$15 per hour, the expense incurred solely for this manual task amounts to nearly \$2,000 per week or over \$100,000 annually. This

estimation didn't include the additional time and resource required for resolving disputes, which means that this process can be more time-consuming and costly than the estimation.

The introduction of an automated system has substantially transformed this process. By leveraging computer vision for image comparison in damage attribution and batch processing in toll attribution, the solution not only identifies and matches toll charges and damages with the correct driver and trailer but also flags discrepancies for further review. As a result, companies have realized a reduction in manual labor approximately equivalent to 2–3 full-time employees. In financial terms, this has translated into an annual cost savings of approximately \$500,000. Moreover, we estimate that automation can greatly enhanced the accuracy of attributions to 80%, significantly reducing the frequency of misassigned charges and improving the overall trustworthiness of the company, which is important in terms of maintaining good relationships with the customers.

Beyond saving costs, the implementation of automation brings strategic benefits that align with broader industry imperatives. Timely and accurate attribution is very important in the logistics sector, where payment cycles are closely tied to the accuracy of documentation and claim processing. Automated toll and damage identification enables invoicing to customers in a timely manner, expedites insurance claim submissions, and accelerates the dispute resolutions. These improvements collectively enhance cash flow management, improve working capital efficiency, and make financial planning more predictable for the company.

In sum, automating toll and damage attribution not only reduces costs and saves time but is also essential for improving the operational robustness of logistics firms. It supports more accurate financial reporting, enhance customer trust through faster settlements, and fosters a data-driven approach to fleet and cost management—benefits that extend far beyond the initial scope of finding the accurate clients and reducing manual work.

RECOMMENDATIONS

To facilitate a transition toward a data-driven decision-making framework, a two-phase approach is recommended, with emphasis placed on establishing a strong data infrastructure and enhancing the analytical understanding of collected data.

Establishment of Foundational Data Infrastructure

It is recommended that organizations prioritize the implementation of robust data pipelines and cloud-based storage systems. Automated and scalable ETL (Extract, Transform, Load) processes should be established using tools such as Airflow, Fivetran, or dbt, allowing for efficient ingestion and transformation of data. Data lineage and version histories should be tracked to ensure transparency and reproducibility.

Several foundational pillars should be addressed:

 Relevance and Representativeness: Data should be collected in alignment with clearly defined business and modeling objectives, ensuring that only pertinent and representative information is retained.

- Integrity and Quality: Continuous validation processes must be applied to ensure that data is
 accurate, consistent, and complete. Without such measures, the reliability of downstream
 analyses may be compromised.
- **Structure and Storage**: Standardized schemas and naming conventions should be adopted. Additionally, it is advised that data be stored using scalable, cloud-native platforms such as Google BigQuery or Snowflake, which offer high performance and flexible accessibility.

Enhancement of Data Understanding

Once foundational systems have been established, it is advised that analytical practices be developed to derive insights and create value from the data:

- Exploratory Analysis: Descriptive statistics and data visualizations should be conducted to identify
 patterns, anomalies, and potential outliers. This process allows trends to be observed and better
 understood prior to formal modeling.
- **Feature Engineering**: Raw data should be transformed into meaningful variables using domainspecific knowledge. When carefully constructed, such features are known to enhance the predictive power and interpretability of machine learning models.

By following these recommendations, organizations may be better positioned to progress from the mere collection of data to the strategic use of data as a driver of operational efficiency, innovation, and competitive advantage.

Advanced Modeling with Data

Once foundational infrastructure and analytical capabilities are in place, organizations should invest in advanced modeling techniques to automate processes, improve predictive accuracy, and reduce manual efforts. These models can unlock significant operational benefits by harnessing the full potential of telematics and sensor data.

For instance, automating toll updates and payment tracking by integrating toll, trip, and financial data into a real-time status engine—powered by rule-based logic and machine learning—can help save operational time and reduce billing errors. Similarly, damage-prone situations can be identified in advance by analyzing trip conditions, route data, and sensor inputs, enabling anomaly detection models to trigger early alerts and prevent breakdowns.

Predictive maintenance is another high-impact use case, where time-series modeling of sensor readings such as temperature and vibration can help forecast trailer failures and minimize downtime. Additionally, telematics data like acceleration, braking, and deviation patterns can be aggregated to profile driver behavior, enabling organizations to assign risk scores, improve safety, and incentivize responsible driving through clustering or supervised learning techniques.

By embedding such intelligent systems into daily operations, organizations not only reduce costs and increase uptime but also drive strategic value through automation, risk mitigation, and proactive decision-making.

CONCLUSIONS

The global logistics industry continues to undergo significant transformation, with the subscription-based trailer access model offering flexibility that traditional ownership or leasing models cannot provide. However, this innovative business approach introduces unique challenges in cost attribution and damage tracking due to multiple parties involved in the freight chain.

Our research has directly addressed these challenges by developing predictive frameworks for toll attribution and damage assessment that operate effectively even with limited data availability.

Our toll attribution framework successfully automates the process of matching toll records with the appropriate customers by leveraging telematics data and timestamp analysis. By implementing this system, logistics providers can significantly reduce manual processing time and human error, resulting in more accurate billing and improved customer trust. The damage attribution system, combining inspection report analysis with computer vision technology, creates a robust mechanism for determining when damage occurred and which party was responsible during that time window.

The integration of these systems yields substantial operational and financial benefits. As demonstrated in our analysis, before automation, toll and damage attribution required significant manual effort, with longer processing times and higher risk of human error. By implementing an automated process, logistics providers can reduce processing time by approximately 2-3 FTEs and cut costs by close to \$0.5 Million USD annually. Additionally, automation minimizes manual errors, enhancing the accuracy and reliability of the attribution process by up to 80%.

Beyond direct cost savings, these systems create strategic value through improved trust and transparency with customers. In the logistics industry where payment cycles are tight, timely and accurate attribution is critical. The automation of toll and damage identification ensures faster invoicing of carriers and accelerates insurance claims and settlement processes. Ultimately, this enhances cash flow management, improves operational efficiency, and strengthens compliance with industry standards.

While our study has yielded promising results, we acknowledge certain limitations. Data latency, the occasional need for manual review, and the scope of our methodologies focusing specifically on toll assignment and damage assessment without addressing all related processes such as dispute resolution present opportunities for future research. Further work could explore integrating real-time toll data processing, enhancing image recognition models to detect more subtle damages, and developing more sophisticated predictive algorithms to anticipate potential damages based on route patterns and carrier histories.

In conclusion, our research demonstrates that predictive analytics, when applied to logistics operations, can transform cost attribution and damage tracking from reactive, manual processes into proactive, automated systems. For subscription-based trailer access model providers, this represents not just an operational improvement but a strategic competitive advantage in an increasingly data-driven industry. As the logistics sector continues to evolve, we anticipate that the integrated use of telematics data, computer vision, and predictive modeling will become standard practice, enabling more transparent, efficient, and customer-focused operations.

REFERENCES

- The Brainy Insights. (2024, September 11). Third-party logistics market to reach USD 2,230.96 billion by 2033: Thriving e-commerce sector in emerging countries to propel growth. GlobeNewswire. Retrieved April 13, 2025, from https://www.globenewswire.com/news-release/2024/09/11/2944643/0/en/Third-Party-Logistics-Market-to-Reach-USD-2-230-96-Billion-by-2033-Thriving-E-Commerce-Sector-in-Emerging-Countries-to-Propel-Growth.html
- 2. Namazi, M. (2016). Time driven activity based costing: theory, applications and limitations. Interdisciplinary Journal of Management Studies (Formerly known as Iranian Journal of Management Studies), 9(3), 457-482.
- 3. Cimili, P., Voegl, J., Hirsch, P., & Gronalt, M. (2022, September). Automated damage detection of trailers at intermodal terminals using deep learning. In *Proceedings of the 24th International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation (HMS 2022), Rome, Italy* (pp. 19-21).
- 4. Chen, J., Dong, C., & Wan, Y. Enhancing Container Damage Detection with improved YOLOv5 Model: Integrating Swin Transformer.
- 5. Wang, Z., Gao, J., Zeng, Q., & Sun, Y. (2021). Multitype damage detection of container using CNN based on transfer learning. *Mathematical Problems in Engineering*, 2021(1), 5395494.
- 6. Xu, S., Niu, J., & Cai, X. (2020). Optimize Logistics cost model for shared logistics platform based on time-driven activity-based costing. In *Journal of Physics: Conference Series* (Vol. 1437, No. 1, p. 012115). IOP Publishing.
- 7. ATRI, An Analysis of the Operational Costs of Trucking: 2024 Update, https://truckingresearch.org/about-atri/atri-research/operational-costs-of-trucking/