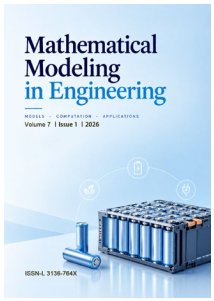
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 3, July, 2026



Cross-Modal Graphormer-CRF Prediction Model for Ramp Passage of Warehouse Robots

Aris Boukas[a,*], Dimitris Christopoulos[a]

[a] *School of Electrical and Computer Engineering, Technical University of Crete, Chania, 73100, Greece*

Received 21 June 2026; Revised 13 September 2026; Accepted 19 September 2026; Published 22 September 2026

Abstract

Warehouse robots frequently traverse short indoor ramps that connect receiving docks, sorting areas and storage floors. Ramp passage is not a routine navigation event because safety depends on slope, payload distribution, wheel slip, acceleration, battery condition, floor friction and route command. This paper proposes a Cross-Modal Graphormer-CRF prediction model for warehouse robot ramp passage. The model aligns inertial measurements, wheel encoder signals, motor current, depth images, load-cell readings and route-state data into passage windows. Cross-modal graph nodes are then constructed to represent ramp geometry, robot attitude, drive response, payload condition and sensing quality. A Graphormer encoder captures physical relations among these evidence groups, while a CRF decoding layer preserves the transition logic of approach, climb, boundary risk and passage failure. A simulated warehouse dataset of 3,210 ramp-passage windows is used for evaluation. The proposed model achieves 91.8% accuracy, 88.6% macro-F1 and 84.3% recall for boundary-risk windows, outperforming LSTM, Transformer and GCN baselines. Robustness tests with missing depth frames, load shift and encoder noise show that confidence decreases before the class label becomes unstable. These results indicate that structured cross-modal modelling can support reliable ramp-passage prediction and provide interpretable evidence for warehouse robot scheduling, route restriction and manual inspection.

Keywords: Warehouse Robot; Ramp Passage Prediction; Cross-Modal Fusion; Graphormer; Conditional Random Field; Risk Ranking

1. Introduction

Autonomous mobile robots are increasingly used in warehouses because they reduce repetitive transport work and improve storage efficiency [1]. Indoor logistics environments are no longer flat and uniform, because robots must pass dock ramps, temporary plates, lift transitions and sloped buffer zones [2]. Ramp passage changes wheel-ground contact, payload balance and braking distance, and these changes directly affect operational safety [3]. Recent studies on mobile robot slip estimation show that wheel slip can rapidly reduce tracking accuracy when the surface and load condition change together [4]. These observations indicate that ramp passage prediction should not rely only on route planning or a fixed slope threshold. A safe prediction system needs to understand the dynamic relationship among robot motion, ramp geometry, payload and local sensing quality.

Most warehouse navigation systems at present only have collision avoidance, path planning and task allocation. The above functions are necessary but not sufficient conditions for a robot to drive over a specific ramp under the given load and speed. The two robots are on the same road, but they are in different circumstances of danger; one has an uneven load, and the other has damaged drive wheels and temporary contamination of the ramp plate. Vision data can provide information about the shape and obstacle distribution of the ramp, and wheel encoders/motor current data

* Corresponding author: aris.bou@ece.tuc.gr

https://doi.org/10.68406/mme.2026.vol7iss3nm10:122-134

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

show drive response. Inertial data are pitch, roll and vibration information; load-cell data show payload position. The prediction model will use all the above sources. Research on sensor fusion for mobile robots has shown that single-channel observation is often unstable under terrain disturbance [5]. Ramp prediction is thus a cross-modal structured prediction problem.

A Cross-Modal Graphormer-CRF model for ramp passage prediction is presented in this paper. Graphormer is employed because the robot, ramp, load and drive-response variables have a relational structure and are not in a flat feature vector. Load or unevenness of the inclined plane causes changes in Pitch Angle and Slip Index. The relationship between motor current and passage risk is also different during acceleration and deceleration of the robot. Due to the ordered nature of the passage states, Conditional Random Field decoding is employed. A robot will generally move in the sequence of approach-to-climb, then to a stable pass or boundary risk, and finally to completion or failure. Similar structure prediction ideas have also been applied to sequential decision-making and activity recognition [6]. Therefore, the proposed model is in a relational form with state-transition constraints. First, an engineering prediction framework is constructed to connect various kinds of sensors and derive an intelligible path-selection decision. The model will not return only 'pass' or 'fail'. Return passage state, confidence, dominant evidence group and review priority. Suitable for the design of warehouse management; a low-confidence high-risk window may need to be re-measured, and a high-confidence severe-risk window may require route restriction. Data table construction and disturbance tests will also be conducted to evaluate the stability of the model in the presence of missing depth frames, abrupt payload shifts, and increased encoder noise. Therefore, the deployment of warehouses cannot be assumed to have ideal sensing conditions [7].

This study is positioned as a ramp-passage prediction method for warehouse robot operation rather than as a general navigation planner. Section 2 reviews ramp operation safety, multimodal sensing under slope and payload disturbance, and structured graph-sequence prediction. Section 3 presents feature alignment, cross-modal Graphormer encoding, CRF decoding and passage-risk scoring. Section 4 reports sample construction, model comparison, state-transition analysis, robustness testing, explanation output and engineering review priority. Section 5 summarizes the principal findings, deployment value, limitations and future validation requirements. The model is designed to determine whether the current robot, load and ramp conditions form a safe, uncertain or high-risk passage window before the robot reaches the end of the ramp.

2. Related Work

2.1. Warehouse Robot Ramp Operation and Safety Prediction

Safety prediction for warehouse robots has switched from rule-based collision detection to risk-aware navigation. The old system had a small slope and turning radius, and was not obstructed. The rules are easy to verify, but they are not suitable for short ramps with slopes, load transfer and changes in floor friction. Logistics robot research has found that, for high-efficiency operation of warehouses, both path quality and local motion stability are required [8]. The ramp-passage case is a special case, and the failure may occur even if the whole route is correct.

The State is affected by a temporary condition. The robot on the ramp will have a small increase in current, a pitch-angle jump, and a short encoder error. These signals are not always errors. However, when they occur together with a large payload offset and depth-frame uncertainty, they may be the result of boundary risk. Human-aware industrial robot safety research has shown that deployable navigation stacks need to integrate perception, prediction and control layers rather than using a final alarm for safety [9]. It supports the current focus on early ramp-state prediction.

Ramp operation also changes maintenance and scheduling decisions. If the robot continuously receives a low-confidence prediction for the same ramp, then inspect the surface and plate fixing or wheel condition of the ramp. If only that specific robot is likely to be in danger during the heavy-load operation, reduce the load or select a flatter route via the scheduling system. Therefore, the ramp prediction should return ranked evidence and confidence instead of a binary result.

2.2. Multimodal Sensor Fusion Under Slope and Payload Disturbance

Mobile Robot Sensing is naturally multi-modal. Encoders are for wheel displacement; Inertial Units are for body orientation; Current Sensors monitor drive-load; Depth cameras obtain ramp geometry; Load Cells measure weight distribution. The Signals have different Sampling Frequencies and Reliableities. Sequential sensor fusion for slip estimation has shown that proprioceptive signals can be combined to improve slip-state prediction in wheeled robots [10]. Ramps in warehouses should also have visual and proprioceptive cues; otherwise, only visual identification of wheel slip and motor response will be possible.

The first reason is that the modalities will not work for different reasons. The problems in the depth images may be caused by reflective ramp plates or occlusion of carried goods. Encoders can be affected by wheel slip. Motor Current

is affected by Battery Voltage and Acceleration. Load cells may be damaged due to impact from the robot driving over the ramp. Therefore, a quality mask can be used for cross-modal fusion to assign different weights to the various channels. Recently, studies on industrial monitoring have also focused on retaining data quality indicators in the prediction pipeline [11].

Payload Disturbance is relatively high. A robot with a centre of gravity and a robot with an off-centre load can be on the same slope but have different roll angles. An increase in the load will result in a rise of the motor current; although not immediately damaging, an eccentric load may cause lateral sliding at a relatively small slope. Therefore, the feature representation should be associated with slope, load, speed and slip evidence. A flat feature vector can store the variables, but it will not show their relationships. This motivates graph-based modelling.

2.3. Structured Prediction with Graph and Sequence Constraints

GNNs are used for data with relationships among the variables. At the ramp passage, the ramp node is associated with slope and surface conditions; the robot-state node is related to pitch and speed; the drive node is connected to encoders and motor-current responses; and the load node is connected to weight and centre of mass offset. Graph attention models can give different weights to these relations according to the current working window [12]. Graphormer extends the previous idea to add structural bias in attention for non-Euclidean data.

A structured prediction model is also needed for the ordered-process ramp-passage state. The model should not generate an unreasonable sequence of failures in which the robot cannot reach the ramp, or return to a stable entry after detecting a passage failure. Conditional Random Fields have long been employed to model label transitions in sequence prediction [13]. Although the neural model has learned the sequence pattern, explicit transition constraints still need to be added in the design of the state order, and the sample size is small. The two reasons for using Graphormer and CRF together now are: Graphormer first learns the cross-modal relations among slope, slip, load and sensing quality. Secondly, CRF decoding restricts the predicted state path and improves boundary recognition. The two are in harmony. Graphormer obtains the relevant relational evidence and CRF regularises the final state sequence.

3. Methodology

3.1. Ramp-State Feature Alignment

The input data include depth maps of the ramp area, inertial measurements, wheel encoder speeds, motor current, load-cell signals, route commands and battery status. Figure 1 shows the ramp-state feature alignment workflow. Robot motion data, visual ramp evidence and payload evidence are aligned by passage time, and the aligned window is divided into entry, climbing and exit periods. Quality masks are assigned to depth, inertial, encoder, current and load channels so that missing or noisy modalities can be down-weighted without discarding the whole passage window.

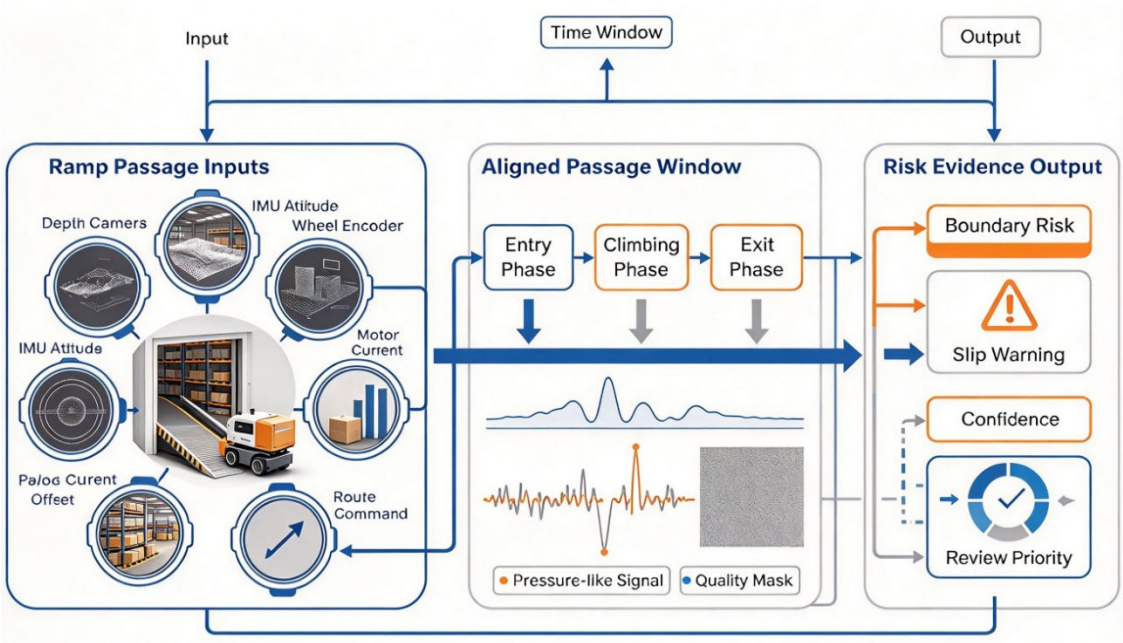


Figure 1. Ramp-State Feature Alignment Workflow

The Alignment Procedure is for warehouse operations, not for laboratory trajectories. The robot may halt before reaching the ramp; increase speed upon receiving a route command; or decrease speed due to the proximity of another robot. These events alter the time-position of the same physical ramp response. Therefore, a single fixed sample index for comparison cannot be used in this way. Align the passage window with the start of the pitch increase, the first ramp depth boundary, and the change in the motor-current gradient. Then, the three anchors are compared with the route command to confirm that the window shows a ramp passage instead of normal acceleration on a flat floor.

First, a ramp-response vector for each passage window must be generated. The vector components are Slope, Pitch, Speed, Slip, Load Offset, Sensing Quality indicators. It is small enough for online inference and has kept the necessary physical evidence of passage prediction.

$$x_t = [s_t, p_t, v_t, q_t, l_t, m_t] \tag{1}$$

In this expression, s_t denotes ramp slope, p_t denotes robot pitch, v_t denotes velocity, q_t denotes slip-related encoder response, l_t denotes payload offset and m_t denotes modality quality. The vector does not represent a final risk value. It is only the aligned observation state that will be transformed into graph nodes.

The slip-related encoder response is calculated from the difference between commanded wheel speed and observed displacement tendency. A high value does not always mean a failure, because short speed differences can appear during normal control adjustment. The payload-offset term is also interpreted cautiously. A load located away from the robot centre can increase roll response on a ramp, but it may still be safe when slope is low and speed is reduced.

$$m_t = [m_d, m_i, m_e, m_c, m_l] \tag{2}$$

A good mask shows that the depth image, inertial signal, encoder signal, current signal and load signal in the current window are all reliable. A mask is used to have the model reduce the weight of a modality when the measurement is unavailable or noisy, but not ignore that passage window for prediction entirely.

The aligned representation also has the passage phase label in the window. The beginning, rising and conclusion of the same number are all different. For instance, a higher starting pitch is expected; a sudden rise in pitch during a stable ascent may be caused by undulations of the ramp surface. Preserve the phase relationship to compare the same physical signal at different operating positions in the model.

3.2. Cross-Modal Graphormer Encoder

Figure 2 shows the Cross-Modal Graphormer-CRF model architecture. The left side represents ramp, robot-attitude, drive-response, payload and quality nodes. The middle part applies Graphormer attention with relation bias so that physically connected evidence groups can interact. The right side uses CRF decoding to generate an ordered passage-state sequence and review priority. This structure keeps ramp-passage prediction interpretable because each node corresponds to an operational evidence group rather than to an anonymous sensor column.

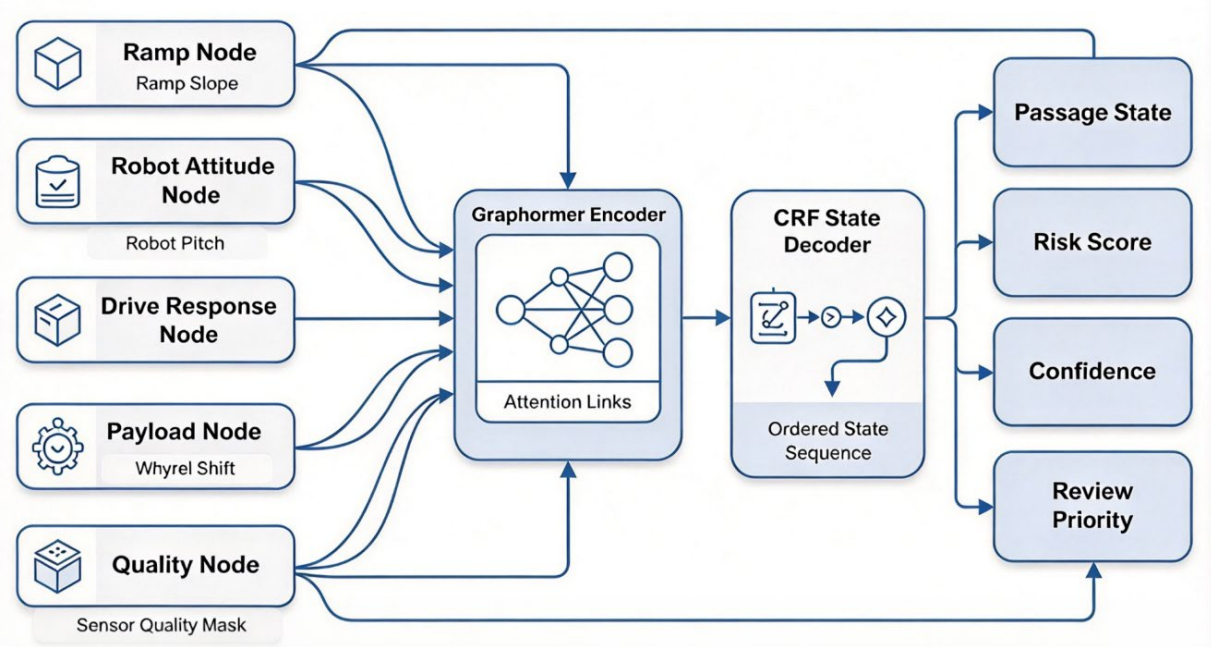


Figure 2. Cross-Modal Graphormer-CRF Model Architecture

The graph is organized according to evidence groups. A ramp node stores slope, surface continuity and depth-boundary reliability. A robot-attitude node stores pitch, roll and vertical acceleration. A drive-response node stores encoder deviation, motor-current residual and speed-control residual. A payload node stores weight, offset and load-bearing condition, while a quality node stores modality reliability. This node design avoids an excessively dense graph of raw sensor columns and allows the attention layer to compare physically meaningful relations.

$$h_i^0 = W_n z_i + b_n \tag{3}$$

The initial node embedding h_i^0 maps each evidence group into a common feature space. The same transformation is used for ramp, drive and payload nodes, but their structural relationship is preserved through relation bias in the attention layer.

The relation bias is not set as a fixed risk rule. It only tells the attention layer which pairs of evidence groups should be compared more directly. Slope and pitch are connected because they describe ramp geometry and robot body response. Motor current and encoder residual are connected because they describe drive effort and wheel-ground interaction. Payload offset and roll are connected because they jointly describe lateral stability.

$$a_{ij} = \mathrm{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d}} + r_{ij}\right) \tag{4}$$

The attention score a_{ij} includes a relation bias r_{ij}. This bias identifies whether two nodes are physically connected, such as slope with pitch, payload with roll, or motor current with slip. The relation term improves the ability of the model to preserve engineering structure.

$$h_i^{k+1} = \sum_j a_{ij} V_j \tag{5}$$

The updated node embedding h_i^{k+1} aggregates cross-modal evidence. When a robot enters a ramp with heavy load and rising slip index, the drive node and payload node can jointly influence the passage-state representation. After several encoder layers, the model obtains a context-aware representation for each evidence group and reduces false alarms caused by isolated signals.

3.3. CRF Decision Layer and Passage Risk Output

CRF decoding is employed to generate an ordered sequence of passage states. The layer receives emission scores from the Graphormer encoder and combines them with transition scores for entry, stable climbing, boundary risk and failed passage. This design prevents isolated or physically unreasonable state jumps, such as an immediate transition from entry to completed passage without climb evidence. The CRF layer therefore converts local cross-modal evidence into a state path that is consistent with warehouse ramp operation.

The set of states is based on warehouse operations. Based on the entry condition, the robot is near the ramp and has reduced its speed. A stable-climbing state is achieved when the robot is still on the incline, and the error has not increased substantially. A boundary-risk state is a situation where, during the movement of a robot, there is a high risk of wheel slip, excessive pitch response or load instability.

$$E(y_t) = W_o h_t + b_o \tag{6}$$

The emission score $E(y_t)$ describes the local support for each passage state at time t. It reflects the evidence learned by the Graphormer encoder but does not yet enforce transition order.

The emission score is useful for local diagnosis, but a single local score can be unstable in noisy sensor windows. A short wheel-speed disturbance may temporarily resemble boundary risk, and a missing depth frame may reduce the support for stable passage. The CRF layer therefore combines local evidence with the probability of state transition.

$$S(y) = \sum_t E(y_t) + \sum_t T_{y_{t-1}, y_t} \tag{7}$$

The sequence score $S(y)$ combines local evidence and transition scores. The transition matrix T encodes the engineering state order, so impossible or unlikely state jumps receive a lower score.

$$R_t = \mathrm{sigmoid}\left(w_r^T h_t + \alpha(1 - c_t)\right) \tag{8}$$

The passage risk R_t uses the decoded representation and confidence c_t. Lower confidence increases the risk score because uncertain sensing conditions require engineering caution. The output includes passage class, confidence, dominant evidence group and review priority.

Review priority is produced after the risk score rather than before it. A high-risk and high-confidence prediction is assigned to immediate restriction because the system has enough evidence to reroute or stop the robot. A high-risk but low-confidence prediction is assigned to remeasurement because poor sensing may explain the warning. A moderate-risk and stable-confidence prediction is assigned to scheduled monitoring, while a low-risk prediction is stored as an ordinary passage record. Confidence is therefore treated as evidence reliability, not as direct authorization for route closure.

4. Results and Discussion

4.1. Sample Construction and Ramp Condition Statistics

The dataset contains 3,210 passage windows of simulated warehouse routes with flat-entry, stable-ramp, heavy-load, wheel-slip, boundary-risk and failed-passage labels [14]. Figure 3 shows sample construction and operating distribution. Figure 3(a) presents label distribution, where stable and flat-entry windows are more frequent than boundary-risk and failed-passage windows. Figure 3(b) presents slope-load coverage and shows that high-load windows appear across several slope ranges. This construction is necessary because ramp risk is caused by the interaction among slope, payload, slip and pitch rather than by a single threshold. the dataset includes both normal transport windows and rare hazardous events.

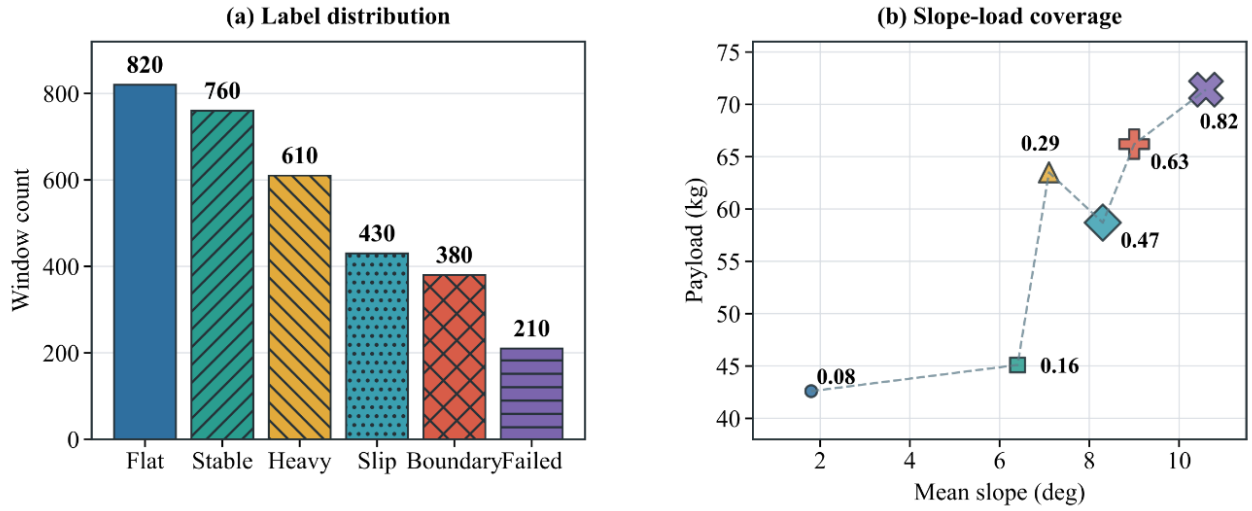


Figure 3. Ramp Passage Sample Construction. (a) Label distribution. (b) Slope-load coverage.

The Risk windows are close to the failed-passage windows in payload and pitch, but their slip indices are slightly lower. It is a problem that the model will not identify all the high-load ramps as failures. Research on the terrain response of mobile robots shows that payload and surface disturbance should be considered together with wheel-ground interaction [15]. As shown in the table, the stable ramp window will still be slightly reduced. Therefore, only the slope is not suitable for decision-making. The same table shows the reasons for a number comparison. The risk ratio is relatively stable for a long time, then gradually increases and finally declines; however, the speed of decline for payload and pitch is different. Therefore, the risk of a ramp is caused by multiple factors and not a single reason. A single threshold can be used to set a relatively conservative alarm rule, but scheduling decisions that would disrupt the closure of the route cannot be made.

Table 1. Numeric Statistics of Robot Ramp Passage Data

State	Windows	Mean Slope	Payload	Slip Index	Pitch Range	Risk Ratio
Flat Entry	820	1.8 deg	42.6 kg	0.031	1.2 deg	0.08
Stable Ramp	760	6.4 deg	45.1 kg	0.074	3.8 deg	0.16
Heavy Load	610	7.1 deg	63.5 kg	0.128	4.6 deg	0.29
Wheel Slip	430	8.3 deg	58.7 kg	0.214	6.9 deg	0.47
Boundary Risk	380	9.0 deg	66.2 kg	0.286	8.1 deg	0.63
Failed Passage	210	10.6 deg	71.4 kg	0.351	10.7 deg	0.82

The sample distribution is uneven because severe warehouse failures are infrequent during normal operation. Macro-F1 and boundary-risk recall are therefore used together with overall accuracy [16]. These metrics reduce the

possibility that the model appears accurate only because it predicts common safe conditions. The dataset also preserves rare states and difficult negative cases instead of treating them as noise. Robust sensor-fusion research similarly indicates that rare abnormal samples should be retained when missing modalities and sensor disturbances are expected during deployment [17].

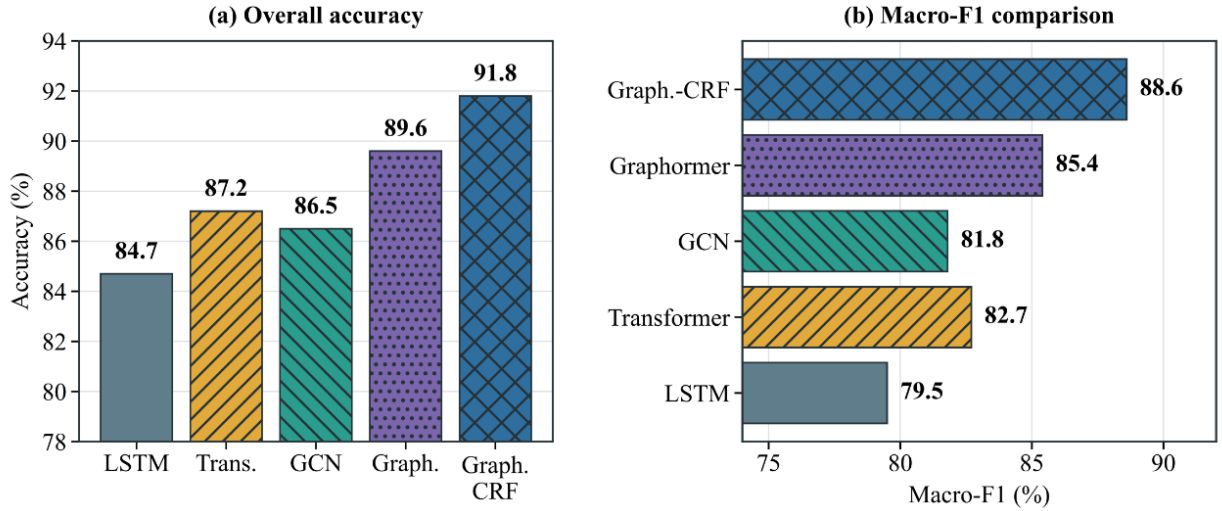


Figure 4. Comparative Prediction Results. (a) Overall accuracy. (b) Macro-F1 comparison.

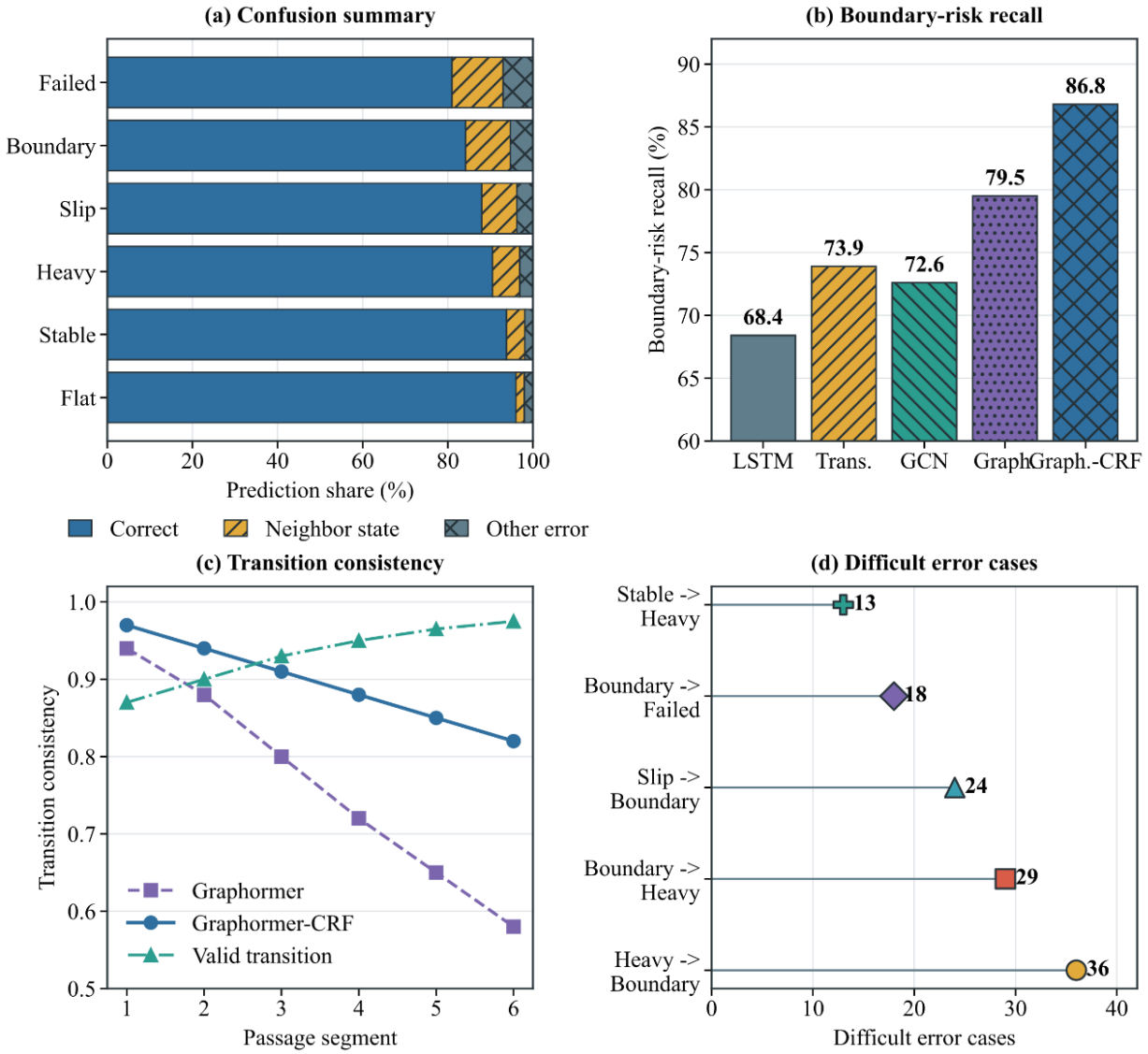


Figure 5. State-Transition and Error Boundary Analysis. (a) Confusion matrix. (b) Boundary-risk recall. (c) Transition consistency. (d) Difficult error cases.

4.2. Model Comparison and State-Transition Results

Figure 4 shows the prediction comparison. Figure 4(a) reports overall accuracy for the proposed model and the LSTM, Transformer, GCN and Graphormer-only baselines. Figure 4(b) reports macro-F1, where CRF decoding improves rare boundary-risk recognition. The results indicate that relational encoding is useful, but transition constraints further stabilize passage-state prediction [18]. Graphormer represents relations among load, slope and slip, while the CRF layer reduces implausible local labels that contradict the ordered ramp-passage process.

The model achieves 91.8% accuracy and 88.6% macro-F1. The LSTM baseline captures temporal dependence but does not explicitly model relations among load, slope and slip. The Transformer baseline captures long-distance dependencies but does not encode the physical structure of evidence groups. The GCN baseline represents graph relations but lacks structured sequential decoding. Graphormer-CRF combines relational attention with state-transition constraints, making it better suited to boundary-risk windows where medium slope, heavy load, reflective depth uncertainty and wheel response may occur together.

Figure 5 provides state-transition and error-boundary analysis. Figure 5(a) presents the confusion matrix, showing that flat-entry and stable-ramp states are well separated. Figure 5(b) reports boundary-risk recall, where Graphormer-CRF performs best. Figure 5(c) presents transition consistency and shows that CRF decoding reduces impossible jumps from start to end. Figure 5(d) presents difficult error cases, most of which occur near heavy-load ramp and boundary-risk states. These results are consistent with structured robot-state prediction research, where boundary labels are harder to acquire than stable labels [19].

Failure prediction should be treated as both a ranking problem and a classification problem. A boundary-risk window with low confidence should be re-observed, whereas a high-confidence failed-passage window may justify temporary route restriction. Therefore, the model output supports warehouse scheduling in addition to offline classification [20]. The review priority separates urgent restrictions from uncertain warnings, preserving throughput while reserving severe intervention for high-confidence risk windows. Reliability-aware multimodal monitoring also shows that review priority should be linked to confidence and evidence strength rather than only to the class label [21].

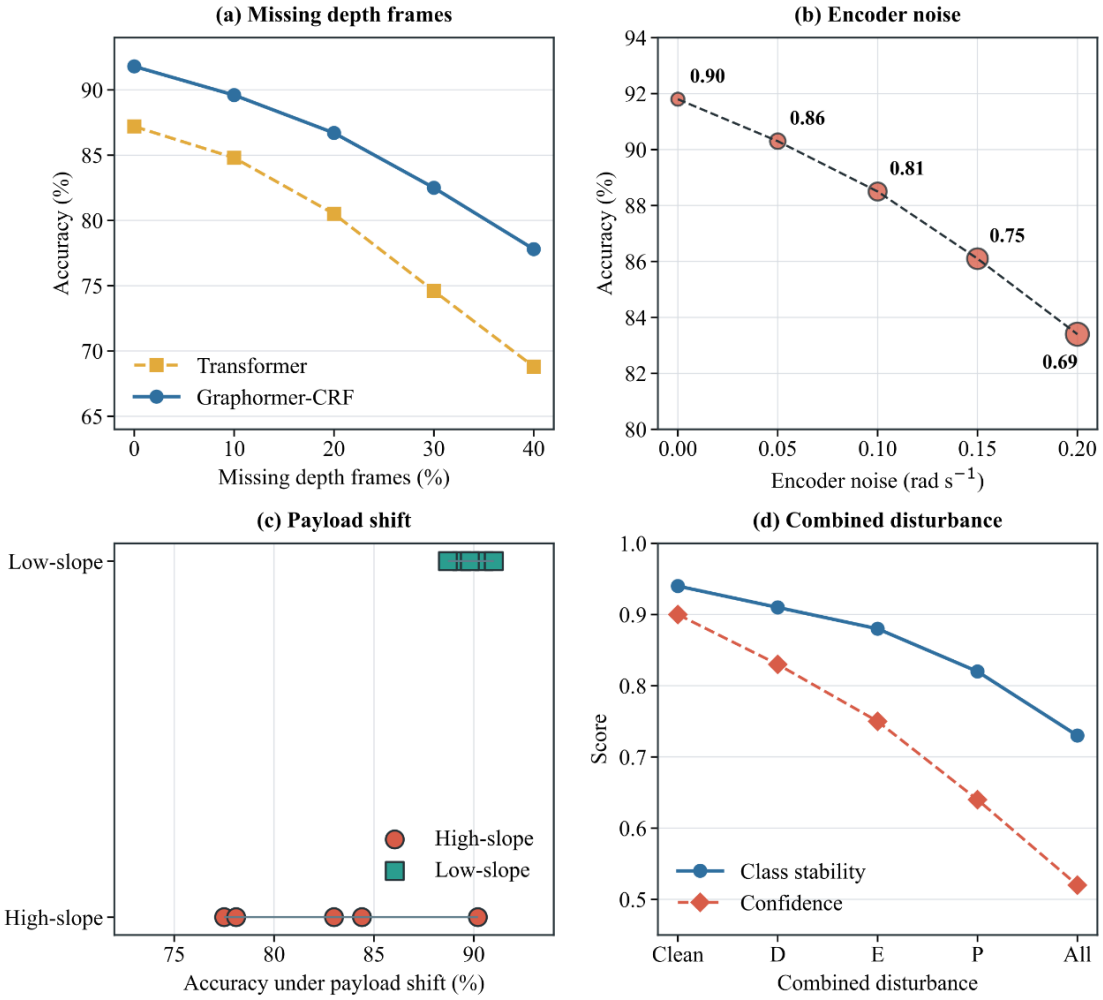


Figure 6. Disturbance Robustness Under Sensor Noise. (a) Missing depth frames. (b) Encoder noise. (c) Payload shift. (d) Combined disturbance.

4.3. Robustness and Engineering Review Output

Figure 6 shows disturbance robustness under sensor noise. Figure 6(a) presents missing depth-frame conditions, where the proposed model degrades more slowly than the Transformer baseline. Figure 6(b) presents encoder-noise disturbance, where the quality mask reduces overconfidence. Figure 6(c) presents payload-shift disturbance, showing that load evidence is important for high-slope windows. Figure 6(d) presents combined disturbance, where confidence drops before the predicted class changes. This behavior is desirable because an online system should alert operators when sensing conditions become unstable [22]. Edge AI research also supports explicit handling of missing and noisy sensor channels for stable local inference [23].

Based on the robustness analysis, the model does not weight all modalities uniformly. When depth maps are unavailable, the model relies more on IMU and encoder evidence; when encoder noise increases, pitch and load evidence provide context. This adaptive use of evidence is necessary for warehouse deployment because sensing quality changes with lighting, surface reflectance, wheel wear and load distribution [24]. Figure 7 shows cross-modal explanation results. Figure 7(a) presents modality contribution, where ramp geometry, pitch response and slip index are linked to boundary-risk prediction. Figure 7(b) presents review evidence structure, showing that load offset is dominant in heavy-load risk windows. The explanation output provides engineers with an inspection path rather than only a final label [25]. Traceable evidence is important for models connected to maintenance decisions [26].

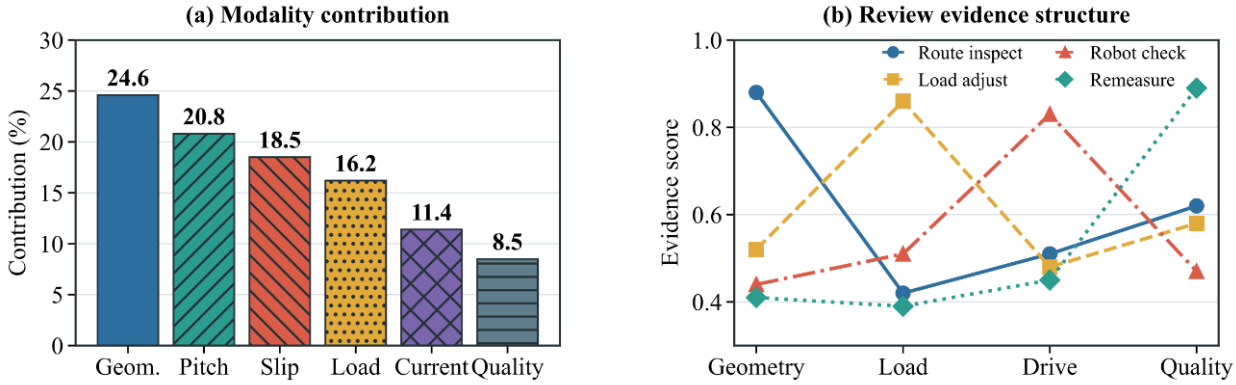


Figure 7. Cross-Modal Explanation Results. (a) Modality contribution. (b) Review evidence structure.

Explanation output is not only a visualization. It informs warehouse managers whether the risk is associated with ramp geometry, load offset or poor sensing performance. If geometry dominates, the route can be restricted; if load offset dominates, scheduling or loading position can be adjusted; if sensing quality dominates, remeasurement is more appropriate than immediate closure. Explainable maintenance research indicates that decision-support systems are more effective when they provide both evidence and confidence [27]. Figure 8 shows passage decision and review order. Figure 8(a) presents immediate restriction, scheduled monitoring, remeasurement and low-risk recording groups. Figure 8(b) shows the confidence-priority relation, Figure 8(c) shows route-level aggregation and Figure 8(d) shows robot-level aggregation. The output is therefore closer to warehouse operations than a binary pass-fail indicator [28].

The engineering value of this method is that it can be applied to task allocation. If the risk is small and the confidence is high, then a robot can be permitted to drive on the ramp. If the probability of boundary risk is relatively high, a flatter path can be used. If both the predicted risk and the uncertainty of sensing are high at that time, then suspend measurement. Given that the flow of goods in the warehouse changes constantly [29], a single, fixed threshold is impractical. The same decision structure can also support fleet-level management. If several robots are in danger on the same ramp, the facilities will be closed. If one robot is dangerous on different ramps, the car will be checked. If the predictions are frequently uncertain but not severe, change the sensing pipeline. The above differences are not suitable for a pass/fail assessment; instead, they can be considered risks, certainties and probabilities of occurrence. The output can be connected to the dispatch constraints. For example, a high-risk road may be temporarily closed to heavy-load trucks and only be used by empty-return robots. Therefore, the predicted results will meet the demand for productivity and pursue the goal of safe and efficient operation. Edge-computing research for predictive maintenance shows that deployable systems need to balance the speed of local decision-making with stable monitoring [30]. The module of route management will store all the decisions related to a particular robot, such as its ID, payload type, ramp number, and approval status. These records can enhance the transparency of the later threshold calibration by showing whether the alarm was caused by a route problem, a vehicle problem, or a sensing problem, and the results will also indicate the limitations of the current study. A dataset of artificial warehouse environments and controlled sensor disturbances was built. The warehouse is wet, the ramp plate is damaged, and the workers have to carry the goods to reach it. The following validation will add a long-term field log and maintenance record. Research on deployable robot navigation has shown that laboratory performance needs to be validated in real operating environments before wide-scale

application [31]. Therefore, the next step is to collect data on all the robot models, ramp materials, wheel-wear conditions and payload types. Record operator operations and any other changes or failures during use. Record can be used to update the CRF transition score and modify the review threshold for that warehouse site. Research on safe robot navigation has also shown that prediction, planning and control need to be done together in a dynamic operating environment [32]. Research on risk-averse task assignment has also suggested that, before deployment, route uncertainty should be taken into account in the assignment decision [33]. In the future, cross-modal graph learning for industrial logistics may incorporate cross-site validation and fleet-level adaptation [34]. The validation process will also compare the predicted review priority with the actual actions of the operator, and only when they are consistent can safe deployment proceed. Thus, the model can be used to make real-world scheduling and maintenance decisions more conveniently.

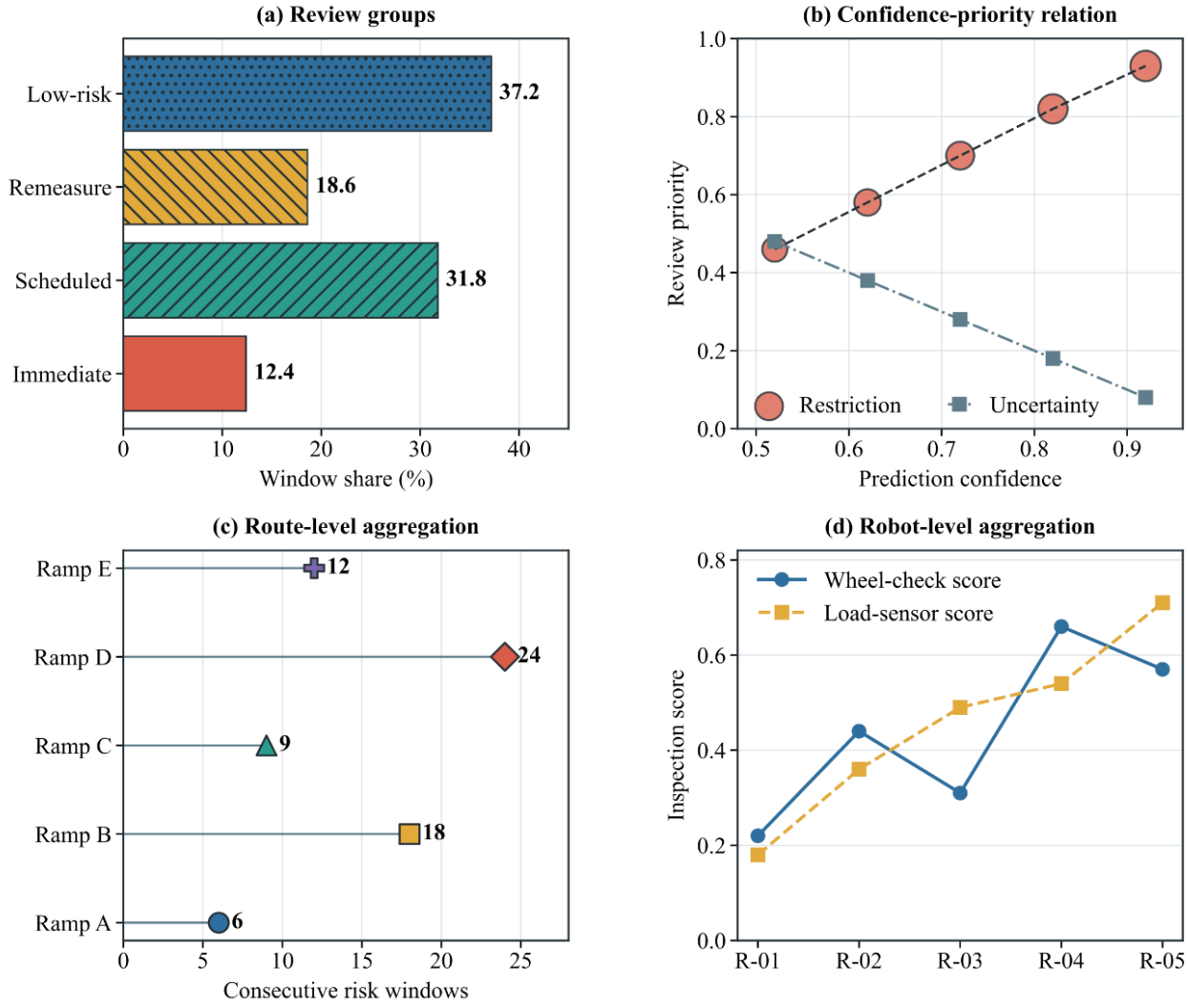


Figure 8. Passage Decision and Review Priority. (a) Review groups. (b) Confidence-priority relation. (c) Route-level aggregation. (d) Robot-level aggregation.

5. Conclusion

This study proposed a cross-modal Graphormer-CRF prediction model for warehouse robot ramp passage, and the overall conclusion is that ramp geometry, robot attitude, drive response, payload state and sensing quality should be evaluated jointly for reliable passage prediction. First, the simulated dataset shows that the proposed model achieves 91.8% accuracy, 88.6% macro-F1 and 84.3% recall for boundary-risk windows, demonstrating the value of cross-modal structured prediction for ramp-passage safety. Second, the comparative results show that relational Graphormer encoding improves evidence fusion, while CRF decoding reduces implausible state transitions and stabilizes boundary-risk recognition. Third, the main limitation is that the evaluation uses artificial warehouse conditions and controlled sensor disturbances; therefore, long-term warehouse logs, varied robot chassis, worn wheels, wet surfaces, temporary ramps and mixed traffic should be included in future validation.

The proposed output separates passage class, confidence, dominant evidence group and review priority. This structure supports immediate restriction for high-confidence severe risk, scheduled monitoring for stable moderate risk,

remeasurement for uncertain high-risk windows and low-risk recording for ordinary passage events. The confidence score is especially useful because it can decrease under missing depth frames, encoder noise or payload shift before the predicted class becomes unstable. The model therefore supports warehouse scheduling and ramp inspection while preserving evidence traceability for later audit.

Future work should incorporate long-term field logs from different robot chassis, ramp materials, wheel-wear conditions, payload types and human-robot mixed traffic scenarios. Deployment should also link prediction outputs with scheduling rules, maintenance records and ramp-infrastructure inspection. These extensions will support site-specific threshold calibration and help determine whether the model remains reliable under wet surfaces, damaged ramp plates, temporary ramps and changing warehouse traffic patterns.

6. Declarations

6.1. Author Contributions

Aris Boukas contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Dimitris Christopoulos contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

6.2. Funding and Acknowledgments

This research received no specific financial support from any funding agency.

6.3. Institutional Review Board Statement

Not applicable.

6.4. Informed Consent Statement

Not applicable.

6.5. Data Availability

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.7. Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

7. References

[1] Lackner, T., Hermann, J., Kuhn, C., & Palm, D. (2024). Review of autonomous mobile robots in intralogistics: State-of-the-art, limitations and research gaps. Procedia CIRP, 130, 930-935. https://doi.org/10.1016/j.procir.2024.10.187

[2] Aremu, M. B., Kabir, I. K., Ahmed, G., & El-Ferik, S. (2025). Autonomous mobile robot path planning techniques: A review: Classical and heuristic techniques. IEEE Access, 13, 117999-118022. https://doi.org/10.1109/ACCESS.2025.3579863

[3] Khan, M. A., Vasques, C. M. A., & Cavadas, A. M. S. (2026). A methodological survey of autonomous mobile robots and automated guided vehicles in industrial logistics. Encyclopedia, 6(9), 197. https://doi.org/10.3390/encyclopedia6090197

[4] Korayem, M. H., Safarbali, M., & Yousefi Lademakhi, N. (2024). Adaptive robust control with slipping parameters estimation based on intelligent learning for wheeled mobile robot. ISA Transactions, 147, 577-589. https://doi.org/10.1016/j.isatra.2024.02.008

[5] Dinh, N. V., Phan Thanh, D., Park, C. H., & Kim, G. W. (2024). Fusion consistency for industrial robot navigation: An integrated SLAM framework with multiple 2D LiDAR-visual-inertial sensors. Computers and Electrical Engineering, 120, 109607. https://doi.org/10.1016/j.compeleceng.2024.109607

[6] Geromel, M., & Cimiano, P. (2024). Sequence labeling via reinforcement learning with aggregate labels. Frontiers in Artificial Intelligence, 7, 1463164. https://doi.org/10.3389/frai.2024.1463164

[7] Aremu, M. B., Ahmed, G., Elferik, S., & Saif, A.-W. A. (2026). Autonomous mobile robot path planning techniques: A review: Metaheuristic and cognitive techniques. Robotics, 15(1), 23. https://doi.org/10.3390/robotics15010023

[8] Alirezazadeh, S., & Alexandre, L. A. (2025). A survey on task allocation and scheduling in robotic network systems. IEEE Internet of Things Journal, 12(2), 1484-1508. https://doi.org/10.1109/JIOT.2024.3491944

[9] Haney, J. M., & Liang, C. J. (2024). A literature review on safety perception and trust during human-robot interaction with autonomous mobile robots that apply to industrial environments. IISE Transactions on Occupational Ergonomics and Human Factors, 12(1-2), 6-27. https://doi.org/10.1080/24725838.2023.2283537

[10] Nourizadeh, P., Stevens McFadden, F. J., & Browne, W. N. (2024). In situ skid estimation for mobile robots in outdoor environments. Journal of Field Robotics, 41(1), 179-194. https://doi.org/10.1002/rob.22252

[11] Kang, J., Al Masry, Z., Varnier, C., Mosallam, A., & Zerhouni, N. (2024). A data quality management framework for equipment failure risk estimation: Application to the oil and gas industry. Engineering Applications of Artificial Intelligence, 136, 108834. https://doi.org/10.1016/j.engappai.2024.108834

[12] Chen, C., Wu, Y., Dai, Q., Zhou, H.-Y., Xu, M., Yang, S., Han, X., & Yu, Y. (2024). A survey on graph neural networks and graph transformers in computer vision: A task-oriented perspective. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 10297-10318. https://doi.org/10.1109/TPAMI.2024.3445463

[13] Corro, C., Lacroix, M., & Le Roux, J. (2025). Bregman conditional random fields: Sequence labeling with parallelizable inference algorithms. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics, 29557-29574. https://doi.org/10.18653/v1/2025.acl-long.1430

[14] Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. Artificial Intelligence Review, 57, 137. https://doi.org/10.1007/s10462-024-10759-6

[15] Wang, Y., Hong, M., Chai, H., Zhang, Y., Wang, G., Wu, C., & Guo, M. (2024). Research on foot contact state detection technology of wheel-legged robot. Sensors, 24(12), 3956. https://doi.org/10.3390/s24123956

[16] Gaudreault, J. G., & Branco, P. (2024). Empirical analysis of performance assessment for imbalanced classification. Machine Learning, 113, 5533-5575. https://doi.org/10.1007/s10994-023-06497-5

[17] Dai, R., Li, C., Yan, Y., Mo, L., Qin, K., & He, T. (2025). Unbiased missing-modality multimodal learning. Proceedings of the IEEE/CVF International Conference on Computer Vision, 24507-24517. https://doi.org/10.1109/ICCV51701.2025.02272

[18] Li, Z.-P., Ma, C., Zhang, Q.-H., Xing, M., Gong, D.-J., Wang, L.-H., & Huang, D.-S. (2026). Graph transformers for graph-level representation learning in the foundation model era: Taxonomy, generalization, trustworthiness, and benchmarks. Neurocomputing, 134969. https://doi.org/10.1016/j.neucom.2026.134969

[19] Zhang, Q., Wei, Y., Han, Z., Fu, H., Peng, X., Hu, Q., Deng, C., Xu, C., Wen, J., Hu, D., & Zhang, C. (2026). Multimodal fusion on low-quality data: A comprehensive survey. Information Fusion, 135, 104437. https://doi.org/10.1016/j.inffus.2026.104437

[20] Zhen, L., Tan, Z., de Koster, R., He, X., Wang, S., & Wang, H. (2025). Optimizing warehouse operations with autonomous mobile robots. Transportation Science, 59(5), 1130-1152. https://doi.org/10.1287/trsc.2024.0800

[21] Gao, Z., Jiang, X., Xu, X., Shen, F., Li, Y., & Shen, H. T. (2024). Embracing unimodal aleatoric uncertainty for robust multimodal fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 26876-26885. https://doi.org/10.1109/CVPR52733.2024.02538

[22] Zou, K., Lin, T., Han, Z., Wang, M., Yuan, X., Chen, H., Zhang, C., Shen, X., & Fu, H. (2024). Confidence-aware multi-modality learning for eye disease screening. Medical Image Analysis, 93, 103214. https://doi.org/10.1016/j.media.2024.103214

[23] Bala, A., Rashid, R. Z. J. A., Ismail, I., Oliva, D., Muhammad, N., Sait, S. M., Al-Utaibi, K. A., Amosa, T. I., & Memon, K. A. (2024). Artificial intelligence and edge computing for machine maintenance: Review. Artificial Intelligence Review, 57, 119. https://doi.org/10.1007/s10462-024-10748-9

[24] Lymperaiou, M., & Stamou, G. (2024). A survey on knowledge-enhanced multimodal learning. Artificial Intelligence Review, 57, 284. https://doi.org/10.1007/s10462-024-10825-z

[25] Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. IEEE Access, 12, 57574-57602. https://doi.org/10.1109/ACCESS.2024.3391130

[26] Sufi, F. (2025). Beyond the sensor: A systematic review of AI's role in next-generation machine health monitoring. Applied

Sciences, 15(19), 10494. https://doi.org/10.3390/app151910494

[27] Ferreira, R. S., & Guérin, J. (2025). Safety monitoring of machine learning perception functions: A survey. Computational Intelligence, 41(2), e70032. https://doi.org/10.1111/coin.70032

[28] Kang, Y., Wang, R., Qin, Z., Yang, P., & Yan, Y. (2025). Warehouses with heterogeneous robots' collaboration: Operational policies and performance analysis. International Journal of Production Research, 63(23), 8698-8726. https://doi.org/10.1080/00207543.2025.2513576

[29] Boysen, N., & de Koster, R. (2025). 50 years of warehousing research—An operations research perspective. European Journal of Operational Research, 320(3), 449-464. https://doi.org/10.1016/j.ejor.2024.03.026

[30] Jiao, G., Huang, M., Song, Y., Li, H., & Wang, X. (2026). Online joint optimization of order picking process in robotic mobile fulfillment systems. Omega, 138, 103374. https://doi.org/10.1016/j.omega.2025.103374

[31] Silva Araujo, S. D. C., Ong Michael, G. K., Deshpande, U. U., Deshpande, S., Avalappa, M. G., Amasi, Y., Patil, S., Bhat, S., & Karigoudar, S. (2025). ResNet-18 based multi-task visual inference and adaptive control for an edge-deployed autonomous robot. Frontiers in Robotics and AI, 12, 1680285. https://doi.org/10.3389/frobt.2025.1680285

[32] Farrell, S., Li, C., Yu, H., Yoshimitsu, R., Gao, S., & Christensen, H. I. (2025). Safe human robot navigation in warehouse scenario. Proceedings of the 2025 IEEE 21st International Conference on Automation Science and Engineering, 548-555. https://doi.org/10.1109/CASE58245.2025.11164151

[33] Ul Abideen, S. Z., Ullah, I., Bai, X., Zhang, W., Li, J., & Ge, S. S. (2026). Multi-robot task assignment: A survey of algorithmic approaches, challenges, and future directions. Computer Science Review, 61, 100968. https://doi.org/10.1016/j.cosrev.2026.100968

[34] Athira, K. A., Udayan, J. D., & Subramaniam, U. (2025). A systematic literature review on multi-robot task allocation. ACM Computing Surveys, 57(3), 68. https://doi.org/10.1145/3700591