Global Academic Journal of Economics and Business

Available online at https://www.gajrc.com
DOI: https://doi.org/10.36348/gajeb.2026.v08i06.008

OPEN ACCESS



ISSN: 2706-9001 (P)
ISSN: 2707-2584 (O)

Review Article

AI-Driven Warehouse Robotics for Inventory Accuracy, Demand Forecasting and Automated Stock Replenishment in Saudi Arabia's Logistics Sector

Idrees Hafiz Syed Hussaini[1]*
[1]Independent Researcher

***Corresponding Author**
Idrees Hafiz Syed Hussaini
Independent Researcher

Article History
Received: 04.09.2026
Accepted: 27.10.2026
Published: 03.11.2026

Abstract: Warehouses increasingly depend on the interaction between physical stock movements and digital decisions. Yet robotic picking, inventory counting, demand prediction and replenishment are frequently evaluated as separate applications, obscuring the consequences of errors passed between them. This critical narrative review examines how artificial intelligence can connect these functions in Saudi Arabia's developing logistics network. Thirty DOI-identified publications from 2020–2025 were mapped across robotic fulfilment, inventory perception, forecasting, sequential inventory control and implementation governance. The synthesis differentiates findings demonstrated in operational settings from optimisation or simulation results, and proposes an integrated decision architecture linking verified stock states to probabilistic demand and constrained ordering. Evidence indicates that robotics can reduce travel and counting effort, while machine-learning forecasting and adaptive control offer decision support under variable demand; however, benefits cannot be assumed to transfer across warehouse layouts, product classes or service conditions. Inventory record integrity emerges as a prerequisite for dependable forecasting and automated orders, rather than a secondary outcome of automation. For Saudi organisations, deployment priorities include multilingual master-data governance, extreme-temperature resilience, interoperability, supplier lead-time visibility and workforce safety. The review proposes staged validation measures that combine record accuracy, fulfilment quality, stockout exposure, intervention rates and energy use. It concludes that the most defensible pathway is auditable, human-supervised automation in which perception, prediction and procurement are jointly evaluated.

Keywords: Autonomous Mobile Robots, Warehouse Inventory Accuracy, Computer Vision, Demand Forecasting, Automated Replenishment, Supply-Chain Resilience, Saudi Arabia.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

Saudi Arabia's ambition to strengthen its position as an international logistics hub places distribution-centre performance at the centre of industrial diversification. Warehouses serving retail, pharmaceuticals, industrial spare parts and cross-border commerce must reconcile fast order completion with accurate records and volatile replenishment requirements. A faster robot does not necessarily deliver better availability: if recorded

units differ from physical units, efficient movement can simply accelerate incorrect fulfilment. Similarly, a statistically impressive forecast is commercially unhelpful when it drives purchase orders against inaccurate receipts, unavailable supplier capacity or unrealistic lead times. The question is therefore systemic rather than technological.

Recent research has advanced autonomous mobile robot planning, shelf-to-person handling and machine-learning assistance in warehouse management [1–4]. Work on large robot fleets further shows why traffic coordination, rack assignment and human workflow are inseparable from throughput [5–13]. Parallel studies explore digital twins and computer vision for monitoring physical processes [14–18], while forecasting and inventory-control research addresses the timing and magnitude of future demand [19–25]. Taken together, these streams suggest a closed operational loop, but they rarely establish whether improvement at one stage survives uncertainty elsewhere. A counting model may report high image-classification accuracy without measuring the commercial cost of missed cases; an inventory policy may assume perfect stock observation even when its implementation depends on scanners and moving robots.

This review asks how AI-driven robotics can improve three linked outcomes: inventory accuracy, demand forecasting and automated stock replenishment in Saudi logistics facilities. Its aim is to construct an evidence-informed framework for integrating physical execution with planning, not to claim an untested national performance gain. The objectives are to distinguish the evidential maturity of major technologies; explain error propagation across stock, demand and purchasing decisions; identify implementable control safeguards; and establish a practical evaluation agenda. The argument develops from physical sensing to predictive decisions, and then to organisational feasibility. Saudi relevance is examined as a deployment context requiring local validation, rather than an assumption that foreign case-study results generalise automatically.

2. Review Design and Evidential Boundaries

A structured critical narrative approach was selected because the research question combines heterogeneous evidence: robotic scheduling models, empirical picking studies, computer-vision trials, forecast reviews and inventory-control experiments. These designs cannot be meaningfully pooled into a single effect estimate. The corpus comprises thirty published works dated 2020–2025, each checked for a matching digital object identifier, article title and publication record. Searches were informed by combinations of warehouse robotics, mobile fulfilment, inventory stocktaking, computer vision, demand forecasting, replenishment, reinforcement learning and supply-chain resilience. Reference relevance was judged against direct contribution to at least one component of the proposed operating loop.

The analytical procedure had four stages. First, studies were classified by decision layer: perception, execution, prediction, ordering or governance. Second, the reported research setting was distinguished as field evidence, experimental environment, simulation or conceptual synthesis. Third, mechanisms were compared across settings to identify shared prerequisites and incompatible assumptions. Fourth, these mechanisms were translated into evaluation propositions for Saudi warehouse operators. Reported percentage improvements are treated as study-specific findings, not pooled estimates or expected local returns. Table 1 presents the resulting cross-domain evidence map.

This is not an exhaustive or registered systematic review. No claim is made about the number of records screened, a universal absence of publication bias or a quantified overall treatment effect. Non-English evidence, unpublished operating data and supplier-specific integration projects may be underrepresented. These boundaries matter because simulation-rich robotics research and carefully selected pilots can exaggerate operational transferability. The contribution is thus conceptual integration and critical appraisal, with explicit separation of established observations from recommendations requiring future testing.

A claim-strength assessment was applied during synthesis. Field measurements were considered informative about feasibility but were examined for duration, baseline and confounding conditions; simulation findings were used to explain possible mechanisms rather than estimate realised savings; reviews were used to identify conceptual relationships and outstanding questions. No quantitative meta-analysis was attempted because outcomes, denominators and experimental conditions were not equivalent. This hierarchy helps prevent a simulation result for hundreds of identical robots from being treated as an expected performance level in facilities handling mixed pallets, manual exceptions and incomplete master data. The resulting framework is therefore explicitly conditional on operational verification.

Table 1: Evidence-to-decision map of technologies and limitations

Evidence domain	Potential contribution	Principal limitation / necessary verification
Mobile robots and routing [1–13]	Reduce travel; orchestrate shelves, pickers and fleets	Traffic congestion, ergonomics and end-to-end item confirmation
Vision, scans and digital twins [14–18]	Observe physical stock and simulate operating constraints	Occlusion, read failures, timestamp integrity and calibration
Forecasting [19–30]	Estimate item and lead-time demand	Censored sales, sparse histories and forecast bias
Replenishment optimisation [22–29]	Support reorder quantities under uncertainty	Assumed observability, supplier variability and override controls
Resilience research [26–28]	Design recoverable operations	Evidence on site-specific recovery and workforce readiness

3. Integrated Intelligence Architecture

An intelligent warehouse requires a trustworthy chain from observation to action. At the physical layer, robots, handheld scanners, fixed cameras and radio-frequency readers record item identity, location, handling events and exceptions. An event-processing layer associates observations with stock-keeping units, batches, bins and timestamps. Reconciliation checks then compare receipts, put-away confirmations, pick confirmations, dispatches and returns with the warehouse management system. Only after questionable events are resolved should a planning layer use the resulting balances as its starting inventory state.

Figure 1 expresses this relationship as a feedback architecture rather than a linear sequence of independent software products. Forecast services combine demand history, confirmed stockouts, promotion schedules and suitable exogenous information. Replenishment logic uses forecast distributions alongside lead times, service requirements, order constraints and physical receiving capacity. Approved purchase or transfer instructions return to warehouse execution; subsequent events update the next decision cycle. A digital twin can provide a sandbox for testing aisle layouts, robot availability and competing service policies, but its validity depends on calibration to actual handling events [14].

The design distinguishes state estimation from decision optimisation. When one camera misses an item, the system should record observation uncertainty rather than silently overwriting a verified count. When a forecast changes abruptly, the ordering service should disclose which inputs moved, quantify supplier exposure and obtain approval if the recommendation exceeds established thresholds. This separation permits component upgrades without destroying auditability. It also limits the risk that an end-to-end model hides uncertainty behind one opaque numerical recommendation.

Operational ownership must be explicit. Warehouse managers own location integrity and exception closure; planners own demand assumptions; procurement teams own supplier limits; and safety personnel own permissible robot actions. Integration should rely on unique event identifiers, synchronised clocks, access permissions and versioned rules. These fundamentals may appear less innovative than autonomous vehicles, yet the reviewed work on warehouse management and coordinated fleets indicates that system performance depends on interdependent decisions and stable interfaces [3-10].

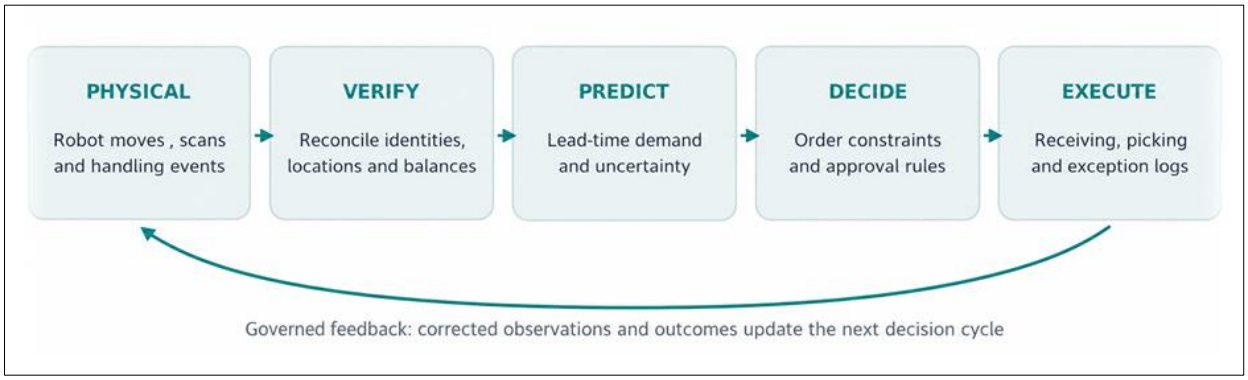


Figure 1: Proposed governed architecture linking physical events with forecast and purchasing decisions; conceptual synthesis by the author

4. Robotic Execution and Inventory Accuracy

Inventory accuracy is commonly expressed as agreement between recorded and physical quantities, but the metric requires greater precision. Exact-bin agreement, item-level count difference, batch traceability and monetary discrepancy capture different risks. A warehouse can achieve respectable aggregate accuracy while placing a critical component in the wrong aisle or assigning a pharmaceutical batch to an incorrect expiry date. For replenishment, the most consequential error is often falsely believing that stock is available. For fulfilment, the inverse error can trigger avoidable shortages or duplicate purchasing.

Robotic mobile fulfilment moves shelves or goods toward workers, reducing walking requirements, whereas autonomous mobile vehicles may carry picked loads between zones [1, 2]. The scheduling studies show that workstation assignment and robot sequencing must be considered together rather than optimised independently [6–8]. Multi-agent research extends this observation: traffic, routes, shelf movements and task allocation can create bottlenecks even when each individual vehicle follows an efficient path [5-13]. Consequently, robotics can improve inventory accuracy only when its handling events are verified and mapped correctly to system records. A recorded movement is not evidence that the expected item was physically lifted.

Computer vision, barcodes and radio-frequency identification address a different uncertainty: what is actually present at a location. Studies of automated stocktaking describe challenges involving occlusion, irregular packaging, lighting, equipment positioning and item heterogeneity [15]. An applied vision study reported improvements in count time and record accuracy in its particular warehouse, illustrating potential but not proving universal replicability [16]. Drone-based radio-frequency experiments likewise show that read performance depends on operating parameters and storage conditions [17]. A semi-automated barcode approach demonstrates why modest, targeted sensing can sometimes be more feasible than full fleet autonomy [18].

Practical reconciliation should retain three states: confirmed inventory, suspected discrepancy and quarantined exception. Duplicate scans, unreadable codes, damaged packaging and misaligned rack positions should create review tasks. Sampling must include slow-moving lines and unusual bins, not merely easy-to-recognise high-volume products. Cycle counts should be risk-weighted by discrepancy history, item value, safety significance and predicted shortage exposure. The business objective is not maximum scans per hour; it is a sufficiently accurate, timely and explainable stock estimate that supports reliable decisions.

Measurement uncertainty should be quantified separately from stock-record discrepancy. For example, a high-confidence barcode match may establish product identity without proving case quantity, whereas a radio-frequency read can indicate nearby tagged material without confirming its bin position. Combining modalities can reduce particular failure modes but also introduces calibration and duplicate-event risks. A practical acceptance test should vary shelf height, label angle, pallet wrap, worker proximity and network availability, then report sensitivity and false-positive rates by scenario. An exception workflow must preserve the raw observation so that disputed records can be reconstructed without assuming that a robot's first reading was correct.

5. Orchestration, Throughput and Human Factors

Robot deployment often begins with a productivity justification, yet system-wide throughput differs from robot utilisation. High utilisation may mean that vehicles spend too much time travelling, queueing for lifts or blocking shared intersections. Research on lifelong path finding shows how rolling-horizon coordination can address repeated assignments, while layout optimisation demonstrates that routing software cannot fully compensate for poor physical design [5-9]. Collaborative task scheduling and multi-agent path-planning studies similarly support joint consideration of tasks and routes [10, 11]. These findings argue against evaluating vehicle numbers as an isolated investment variable.

Human–robot coordination introduces another boundary. Workers must inspect exceptions, handle fragile products, address damaged packaging and respond when shelves arrive in an unexpected order. Field research on collaborative picking indicates context-specific time savings, but also shows that benefits vary with travel distance, pick density and the characteristics of handled goods [12]. Operational algorithms developed for large fulfilment systems provide useful principles for coordinated assignment, although their technology stack and transaction volumes may differ sharply from a regional Saudi warehouse [6]. Therefore, local pilot designs should include task difficulty and ergonomics alongside completed order lines.

Congestion prediction can support preventive routing when many vehicles compete for the same corridors [13]. However, adding predictors cannot replace hard collision-avoidance mechanisms, marked pedestrian zones or emergency-stop testing.

Supervisors need visible overrides, and the algorithm should communicate when it is deviating from the nominal route because of an obstruction. Throughput metrics should count successful, error-free fulfilment rather than raw tasks issued. An architecture that delivers faster but less accurate orders increases downstream corrective work and weakens the inventory state needed for future replenishment.

6. Demand Forecasting Under Operational Uncertainty

Forecasting supplies the anticipatory component of the warehouse intelligence loop. Relevant targets include item-level customer demand, expected picking workload, supplier arrivals and requirements over replenishment lead time. These are not interchangeable forecasts. A daily order-line forecast can support staffing, whereas purchasing requires an estimate of demand during a potentially variable supplier lead time. Reviews of artificial-intelligence forecasting methods emphasise that model selection depends on data volume, horizon and available explanatory variables [19-21]. The research on aggregation and hierarchical forecasting demonstrates why category-level signals and individual item trajectories should sometimes be reconciled rather than modelled in isolation [20, 30].

A disciplined baseline is essential. Seasonal naive forecasts, intermittent-demand methods and suitable statistical time-series models should be benchmarked before complex learning systems are adopted. Tree ensembles can use promotions, calendar effects and price changes, while recurrent or other neural architectures may suit sufficiently rich sequential data. Yet greater model capacity raises the risk of overfitting, unstable explanations and unnecessary computing requirements. Sparse spare parts, newly introduced products and rarely ordered industrial consumables may have insufficient histories to justify elaborate models. Forecast selection should be segmented by demand regularity and the economic consequences of errors, not by a single enterprise-wide preferred algorithm.

Warehouse observations also shape the training label. Recorded sales are censored when products are unavailable: zero sales may reflect a stockout rather than absent demand. A model trained without stockout flags can learn to forecast shortages as though they were normal customer behaviour. Returns, inter-warehouse transfers, promotional spikes and cancelled orders likewise require separate event definitions. Corrections to inventory records should not be confused with genuine market demand. These distinctions are particularly important for new logistics networks where legacy systems have inconsistent transaction histories.

Forecast quality should be evaluated with rolling-origin tests that preserve chronology and compare performance across product segments. Aggregate percentage error can conceal severe problems on low-volume items; weighted absolute error, forecast bias, pinball loss and interval coverage each reveal different weaknesses. Model outputs should include uncertainty bands for lead-time demand, not just single values. Most importantly, the evaluation must connect forecasts to decisions: a modest accuracy gain is valuable only if it reduces avoidable stockouts, excess inventory or expediting after ordering constraints are applied. A demand forecast is an input to service decisions, not a substitute for them.

Forecast governance also needs an explicit model-refresh policy. Frequent retraining may rapidly absorb data errors introduced by a new scanning system, while infrequent retraining may miss genuine demand shifts. Operators should maintain a frozen benchmark during each pilot and monitor forecast residuals by warehouse, product family, service class and forecast horizon. For high-consequence products, interval calibration deserves greater weight than a small improvement in average point error. Human override explanations should be captured as structured information; otherwise lessons about promotions, project demand or supplier rationing disappear into individual spreadsheets and cannot inform the next review cycle.

7. Automated Replenishment and Decision Safeguards

Replenishment converts estimated demand and verified stock into an action: order, transfer, defer or escalate. A basic inventory-position calculation combines on-hand units with outstanding receipts and subtracts committed or backordered quantities. The reorder decision should then consider uncertain lead-time demand, minimum order quantities, supplier calendars, shelf capacity, storage restrictions and targeted service levels. Safety stock cannot be inferred from demand volatility alone if lead times also change. For Saudi import-dependent items, port processing, inland routing and supplier concentration may be relevant scenario variables, but site-specific histories are required before numerical effects can be estimated.

Deep reinforcement learning has been studied for lost-sales, dual-sourcing and multi-echelon inventory environments [22]. Other contributions test reorder policies using continuous actions, bandit formulations and dynamic policies [23–25]. Distributional learning additionally investigates decisions sensitive to extreme adverse outcomes [29]. These results show methodological

possibilities rather than evidence that autonomous purchasing is always superior to traditional reorder-point policies. Each modelling study makes assumptions about state observability, supply behaviour, reward functions and stationarity. A warehouse that feeds stale counts into an adaptive policy may amplify the model's mistakes with high-frequency purchases.

A prudent operating model separates recommendation from commitment. Rules can approve routine replenishment within bounded quantities and supplier exposure, while exceptional decisions require planner review. Thresholds should be calibrated by stock criticality, budget risk and reversibility. An automated engine must refuse to reorder when inventory reconciliation is unresolved, when lead-time data are missing or when a proposed quantity exceeds storage constraints. Escalation is a designed feature, not a failure of autonomy. Procurement teams also need explainable reason codes: forecast increase, protection against variability, outstanding-order delay or stock correction.

Policies should be evaluated by the total cost of decisions, not purchase-order frequency or a model-specific reward score. Required outcomes include fill rate, cycle-service level, holding costs, expired or obsolete stock, supplier expedites and backorders. Multi-echelon trials must distinguish warehouse-level savings from costs shifted to suppliers or downstream locations. Robust control retains contingency suppliers and manual fallback procedures when data feeds, robotics or network connectivity fail. In this sense, intelligent replenishment means accountable decision support with calibrated autonomy.

A scenario-based simulator can test control policies before real purchasing authority is assigned. Scenarios should include incorrect on-hand balances, simultaneous late suppliers, sudden demand spikes, minimum-batch conflicts and receiving-dock congestion. The simulator must update its virtual inventory after every decision, otherwise predicted benefits may rely on impossible simultaneous receipts or unlimited warehouse capacity. Procurement review is especially important for products with long commitments, limited shelf life or difficult returns. The decision log should connect every recommended order to the observed stock state, forecast version and policy threshold that produced it, making post-event analysis practical.

8. The Accuracy–Forecast–Replenishment Feedback Loop

The main conceptual finding is that the three outcomes in the title interact through a feedback mechanism. Robotic handling creates stock transactions; identification systems determine whether those transactions represent the correct items; forecasting interprets historical demand relative to available stock; and replenishment decisions change future inbound volumes. If the stock ledger is wrong, a forecast may misinterpret a shortage and an ordering policy may compound the resulting imbalance. Conversely, more reliable scans can expose lost sales and make demand estimates more representative of actual customer interest.

Figure 2 locates these dependencies along two axes: decision speed and consequence. Perception decisions occur continuously and are often reversible through recounting, whereas purchase commitments operate over longer horizons and may incur substantial financial costs. This asymmetry justifies stricter approval controls as decisions move from local observation to supplier commitment. A cross-functional exception queue connects the layers, ensuring that unresolved discrepancies remain visible rather than being silently absorbed into forecast training.

The relationship also affects performance attribution. When a pilot reports fewer stockouts, the improvement might arise from better cycle counting, improved demand estimates, revised safety stock or simply higher working capital. Evaluation must therefore compare staged combinations against credible baselines. An informative design can introduce automated verification first, forecasting second and bounded ordering third, tracking whether each phase changes the downstream metric that motivated the investment. The relevant unit of analysis is a product–location–time series with associated event lineage, not an isolated algorithm accuracy score.

© 2026: Global Academic Journal's Research Consortium (GAJRC)

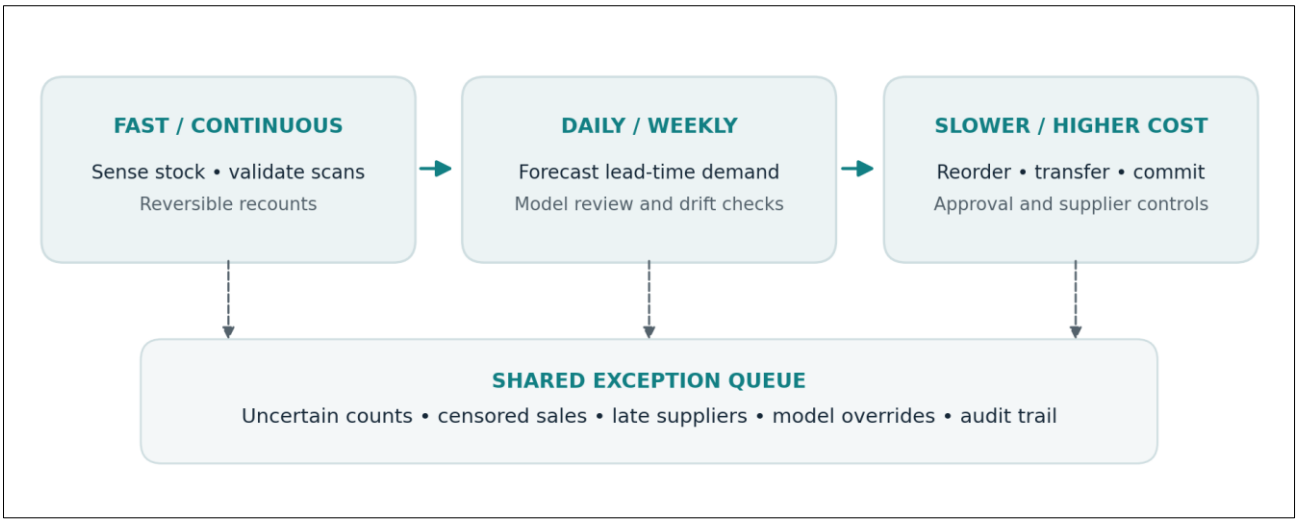


Figure 2: Decision horizons and risk-aware exception handling across the integrated warehouse control loop; conceptual synthesis by the author.

9. Saudi Arabia: Operating Conditions and Adoption Priorities

Saudi Arabia's national transport and logistics agenda seeks stronger logistics connectivity, modern operating systems and a wider network of logistics centres. Warehouse robotics can contribute to this ambition only where operators have economically viable use cases and dependable operational data. Facilities near ports, inland distribution corridors and major metropolitan markets may differ in volume, product mix, staffing arrangements and supplier lead times. A uniform automation blueprint would conceal these differences. Local business cases should therefore be built around throughput profiles, storage constraints and measurable error costs at each facility.

Environmental conditions influence both sensing and machinery. Dust, glare, uneven lighting and high ambient temperatures can affect camera performance, battery life and sensor reliability, particularly around doors, yards and mixed indoor–outdoor movement. These effects should be measured under actual shift conditions rather than presumed constant. Cold-chain warehouses face different constraints, including condensation, low-temperature battery operation and traceability of temperature-sensitive stock. Safety assessment must also account for loading bays, forklifts and pedestrians sharing congested transfer points.

Information architecture requires equally deliberate localisation. Item descriptions may appear in Arabic and English; suppliers may assign inconsistent packaging units; and customs or batch information can arrive through different systems. Master-data rules should preserve a single canonical item identity with explicit conversion factors. Event timestamps, accountabilities and procurement approvals need to be consistent across warehouse, enterprise resource planning and supplier platforms. The value of robotics diminishes when purchase orders are copied manually between disconnected software systems.

A sensible deployment sequence begins with stock-data remediation and baseline measurement. Next, test scanning and reconciliation in a bounded zone; follow with targeted robots on predictable routes; then establish demand forecasts that account for stockouts and lead times. Only after these components demonstrate stable operation should automated replenishment proceed with purchasing limits. Training must include robot supervisors, inventory controllers and buyers, rather than positioning advanced analytics as a replacement for operational knowledge. Worker consultation can reveal process exceptions that historical datasets overlook.

The strategic opportunity lies in resilient local capability. Open interfaces, documented maintenance arrangements, service-level agreements and staff training reduce dependence on one proprietary vendor. Local universities and technology partners can assist with evaluation and calibration; however, future Saudi-specific effectiveness claims require measured deployments. Alignment with the national logistics agenda is a plausible contribution pathway, not proof that an installation advances every strategic target.

A representative pilot portfolio might include a high-volume consumer-goods zone, a slower-moving industrial-spares area and a temperature-controlled storage location. Such contrast would expose whether a sensor or forecasting architecture depends on uniform cartons,

© 2026: Global Academic Journal's Research Consortium (GAJRC)

stable demand or uninterrupted temperature conditions. The pilot should span regular operations and at least one peak-demand period, with results split by location rather than reported as an overall average. For organisations handling imported products, the same protocol can track differences between supplier dispatch dates, customs release, inland transport and actual warehouse availability. This disaggregation supports better lead-time estimation without attributing every delay to the robotics installation.

10. Evaluation Framework and Implementation Economics

Table 2 presents a measurable evaluation design for phased deployment. Establish a sufficiently long pre-implementation baseline and keep its calculation rules fixed. Cycle-count agreement should be stratified by item value and inventory velocity; fulfilment errors should be attributed to picking, labelling and upstream record quality. Forecast studies require time-respecting holdouts and a stable benchmark, while replenishment studies should record total inventory position and exception overrides. Robot throughput alone is inadequate if stock accuracy or workplace safety deteriorates.

Economic assessment should include vehicles, charging, racking modification, software licences, network upgrades, integration, cyber controls, replacement parts and operational disruption during installation. Benefits may arise from avoided recounts, reduced travel, fewer expedited deliveries and lower obsolete inventory, but these savings can overlap. Double counting occurs when the same shorter task is booked simultaneously as a labour saving and increased capacity without demonstrating both outcomes. Scenario analysis should incorporate demand downturns, supplier delays and system outages.

Evidence on resilient supply chains emphasises visibility, responsiveness and preparedness [26–28]. Accordingly, resilience measures should cover time to restore manual operations, supplier substitution capacity and recovery of trusted stock records after an outage. Energy consumed per correct fulfilled order is more meaningful than vehicle electricity use alone because conveyors, charging and cooling systems contribute to total consumption. Where the technology displaces strenuous travel, worker fatigue and near misses should also be monitored. These outcomes make the investment appraisal operationally credible rather than narrowly algorithmic.

Table 2: Proposed pre/post deployment measures and minimum validation safeguards

Outcome	Operational metric	Validation safeguard
Inventory integrity	Exact-bin agreement; discrepancy value; unresolved exceptions	Blind recounts and stratified samples
Robotic fulfilment	Error-free lines/hour; queueing delay; safety incidents	Comparable shifts, task mix and incident logs
Demand forecasting	Bias; weighted error; interval coverage; stockout corrections	Rolling-origin testing against simple baselines
Replenishment	Fill rate; working capital; expedites; obsolete stock	Bounded orders, holdout locations and lead-time tracking
Resilience and resources	Recovery time; manual fallback; energy per correct order	Scenario drills and total-system metering

11. Governance, Limitations and Research Directions

Responsible automation requires technical assurance and organisational authority. Authentication, least-privilege access and segmented robot networks limit the consequences of compromised devices. Logs should retain who changed an inventory balance, which model generated an order and which person approved an exception. Forecast drift should trigger reassessment after promotions, supplier disruptions or shifts in customer demand. Safety overrides must remain operative without cloud connectivity. A system with impressive average results but no recovery procedure is poorly suited to critical inventory.

Three research gaps deserve priority. First, studies should quantify how perception error propagates into replenishment decisions, including false stock availability and misplaced batches. Second, comparisons should assess forecasting and control jointly under uncertain lead times and imperfect stock records, rather than giving ordering agents an unrealistically perfect state. Third, field trials should examine heterogeneous Saudi warehouse conditions and report performance across varying workforce arrangements, layouts and temperature regimes. Shared measurement definitions would make results more comparable without requiring disclosure of commercially sensitive transaction data.

The literature reviewed here mixes analytical, simulation and field designs and is unevenly distributed across those categories. Published innovations may be overrepresented relative to unsuccessful pilots, and results from large e-commerce fulfilment centres may not extrapolate to lower-volume industrial stores. The thirty-reference mapping was purposeful rather than exhaustive, so it should not be interpreted as a bibliometric census. The proposed integration model remains a testable proposition; it has not been experimentally validated in Saudi Arabia within this review. Future comparative studies should preregister endpoints, report uncertainty and distinguish gross efficiency from net organisational outcomes.

Cross-site replication should further separate installed technology effects from routine process discipline. Auditors can assign identical discrepancy definitions to manual and assisted areas, rotate sampling schedules and compare performance after initial learning has stabilised. Reported improvements should include confidence intervals, the duration of observation and an explanation of whether product assortment or customer demand changed during the pilot. A controlled pause in automated ordering can test whether safer manual procedures remain usable when interfaces are unavailable. Researchers should also disclose excluded product groups, intervention thresholds and the share of events requiring human adjudication. These details make results more informative for procurement committees considering investment across diverse warehouse facilities.

Future evaluations should report physical count provenance, document the age of inventory observations, and specify whether forecast accuracy is computed on requested demand or fulfilled orders. This prevents misleading attribution.

12. CONCLUSION

AI-driven warehouse robotics is most valuable when treated as an information-and-decision system rather than a collection of autonomous machines. The reviewed evidence supports targeted improvements in mobile fulfilment, stock sensing, forecasting and adaptive inventory decisions, but each result is conditional on data quality, operating environment and evaluation design. Physical stock verification is foundational because downstream prediction and purchasing inherit its errors. Demand models should account for censored sales and uncertain lead times, while replenishment systems require auditable limits and meaningful human intervention.

For Saudi logistics operators, the recommended pathway is phased deployment anchored in reliable master data, cross-system event traceability, safety engineering and locally measured business outcomes. Integrated tests should assess accurate fulfilment, stockout avoidance, working capital, recovery capability and energy intensity together. The central research contribution is a closed-loop framework that makes dependencies explicit and identifies where claims of benefit must be validated. A warehouse becomes intelligent not when every action is automated, but when verified observations produce appropriate decisions and every consequential action can be explained, challenged and improved.

REFERENCES

1. Fragapane G, de Koster R, Sgarbossa F, Strandhagen JO (2021). Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda. *European Journal of Operational Research, 294(2):405-426.* https://doi.org/10.1016/j.ejor.2021.01.019
2. Barros IRC, Nascimento TP (2021). Robotic Mobile Fulfillment Systems: A survey on recent developments and research opportunities. *Robotics and Autonomous Systems, 137:103729.* https://doi.org/10.1016/j.robot.2021.103729
3. Silva A, Coelho LC, Darvish M, Renaud J (2024). Manual and robotic storage and picking systems: a literature review. *INFOR, 62(3):313-343.* https://doi.org/10.1080/03155986.2024.2336280
4. de Assis RF, Faria AF, Thomasset-Laperrière V, Santa-Eulalia LA, Ouhimmou M, Ferreira WP (2024). Machine Learning in Warehouse Management: A Survey. *Procedia Computer Science, 232:2790-2799.* https://doi.org/10.1016/j.procs.2024.02.096
5. Li J, Tinka A, Kiesel S, Durham JW, Kumar TKS, Koenig S (2021). Lifelong Multi-Agent Path Finding in Large-Scale Warehouses. *Proceedings of the AAAI Conference on Artificial Intelligence, 35(13):11272-11281.* https://doi.org/10.1609/aaai.v35i13.17344
6. Allgor R, Cezik T, Chen D (2023). Algorithm for Robotic Picking in Amazon Fulfillment Centers Enables Humans and Robots to Work Together Effectively. *INFORMS Journal on Applied Analytics, 53(4):266-282.* https://doi.org/10.1287/inte.2022.1143
7. Wang Z, Sheu J-B, Teo C-P, Xue G (2022). Robot Scheduling for Mobile-Rack Warehouses: Human-Robot Coordinated Order Picking Systems. *Production and Operations Management, 31(1):98-116.* https://doi.org/10.1111/poms.13406
8. Zhang S, Zhuge D, Tan Z, Zhen L (2022). Order picking optimization in a robotic mobile

fulfillment system. *Expert Systems with Applications, 209:118338.* https://doi.org/10.1016/j.eswa.2022.118338
9. Zhang Y, Fontaine MC, Bhatt V, Nikolaidis S, Li J (2023). Multi-Robot Coordination and Layout Design for Automated Warehousing. *Proceedings of the International Joint Conference on Artificial Intelligence, 5503-5511.* https://doi.org/10.24963/ijcai.2023/611
10. Zhang H, Wu Y, Hu J, Wang Y (2023). Collaborative optimization of task scheduling and multi-agent path planning in automated warehouses. *Complex & Intelligent Systems, 9:5937-5948.* https://doi.org/10.1007/s40747-023-01023-5
11. Li B, Ma H (2023). Double-Deck Multi-Agent Pickup and Delivery: Multi-Robot Rearrangement in Large-Scale Warehouses. *IEEE Robotics and Automation Letters, 8(6):3701-3708.* https://doi.org/10.1109/LRA.2023.3272272
12. Koreis J, Loske D, Klumpp M (2025). Together, we travel: empirical insights on human-robot collaborative order picking for retail warehousing. *The International Journal of Logistics Management, 36(1):1-20.* https://doi.org/10.1108/IJLM-03-2023-0127
13. Yu G, Wolf MT (2023). Congestion Prediction for Large Fleets of Mobile Robots. *IEEE International Conference on Robotics and Automation, 7642-7649.* https://doi.org/10.1109/ICRA48891.2023.10161554
14. Rizqi ZU, Chou SY, Cahyo WN (2024). A simulation-based Digital Twin for smart warehouse: Towards standardization. *Decision Analytics Journal, 12:100509.* https://doi.org/10.1016/j.dajour.2024.100509
15. Daios A, Xanthopoulos A, Folinas D, Kostavelis I (2024). Towards automating stocktaking in warehouses: Challenges, trends, and reliable approaches. *Procedia Computer Science, 232:1437-1445.* https://doi.org/10.1016/j.procs.2024.01.142
16. Villegas-Ch W, Maldonado Navarro A, Sanchez-Viteri S (2024). Optimization of inventory management through computer vision and machine learning technologies. *Intelligent Systems with Applications, 24:200438.* https://doi.org/10.1016/j.iswa.2024.200438
17. Thomaidis NC, Zeimpekis V (2024). Investigation of operational parameters that affect the use of drones in goods' stock count process: Evidence from experimental results. *Journal of Industrial Engineering and Management, 17(1).* https://doi.org/10.3926/jiem.6446
18. Tong C (2025). An Efficient Intelligent Semi-Automated Warehouse Inventory Stocktaking System. *Operations and Supply Chain Management: An International Journal, 18(2):223-239.* https://doi.org/10.31387/oscm0610469
19. Mediavilla MA, Dietrich F, Palm D (2022). Review and analysis of artificial intelligence methods for demand forecasting in supply chain management. *Procedia CIRP, 107:1126-1131.* https://doi.org/10.1016/j.procir.2022.05.119
20. Babai MZ, Boylan JE, Rostami-Tabar B (2022). Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International Journal of Production Research, 60(1):324-348.* https://doi.org/10.1080/00207543.2021.2005268
21. Walter A, Ahsan K, Rahman S (2025). Application of artificial intelligence in demand planning for supply chains: a systematic literature review. *The International Journal of Logistics Management, 36(3):672-719.* https://doi.org/10.1108/IJLM-02-2024-0120
22. Gijsbrechts J, Boute RN, Van Mieghem JA, Zhang DJ (2022). Can Deep Reinforcement Learning Improve Inventory Management? Performance on Lost Sales, Dual-Sourcing, and Multi-Echelon Problems. *Manufacturing & Service Operations Management, 24(3):1349-1368.* https://doi.org/10.1287/msom.2021.1064
23. Geevers K, van Hezewijk L, Mes MRK (2024). Multi-echelon inventory optimization using deep reinforcement learning. *Central European Journal of Operations Research, 32:653-683.* https://doi.org/10.1007/s10100-023-00872-2
24. Preil D, Krapp M (2022). Bandit-based inventory optimisation: Reinforcement learning in multi-echelon supply chains. *International Journal of Production Economics, 252:108578.* https://doi.org/10.1016/j.ijpe.2022.108578
25. Hammler P, Riesterer N, Braun T (2023). Fully dynamic reorder policies with deep reinforcement learning for multi-echelon inventory management. *Informatik Spektrum, 46:240-251.* https://doi.org/10.1007/s00287-023-01556-6
26. Kassa A, Kitaw D, Stache U, Beshah B (2023). Artificial intelligence techniques for enhancing supply chain resilience: A systematic literature review, holistic framework, and future research. *Computers & Industrial Engineering, 186:109714.* https://doi.org/10.1016/j.cie.2023.109714
27. Zamani ED, Smyth C, Gupta S, Dennehy D (2022). Artificial intelligence and big data analytics for supply chain resilience: a systematic literature review. *Annals of Operations Research, online publication.* https://doi.org/10.1007/s10479-022-04983-y
28. Naz F, Kumar A, Majumdar A, Agrawal R (2022). Is artificial intelligence an enabler of supply chain resiliency post COVID-19? An exploratory state-of-the-art review for future research. *Operations*

Management Research, 15:378-398. https://doi.org/10.1007/s12063-021-00208-w

29. Wu G, de Carvalho Servia MA, Mowbray M (2023). Distributional reinforcement learning for inventory management in multi-echelon supply chains. *Digital Chemical Engineering, 6:100073.* https://doi.org/10.1016/j.dche.2022.100073
30. Mircetic D, Rostami-Tabar B, Nikolicic S, Maslaric M (2022). Forecasting hierarchical time series in supply chains: an empirical investigation. *International Journal of Production Research, 60(8):2514-2533.* https://doi.org/10.1080/00207543.2021.1896817.