International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

Journal Website: http://sciencebring.com/index.php/ijasr

Research Article

Intelligent Robotics and LLM-Based Decision Support for Sustainable Construction Operations

Submission Date: June 30, 2026, **Accepted Date:** July 23, 2026,
Published Date: August 15, 2026
Crossref doi: https://doi.org/10.37547/ijasr-06-08-06

Tharindu Jayasinghe
Department of AI and Data Science, Institute of Digital Innovation, Sri Lanka

Ishara Wijesinghe
Department of Robotics Engineering, Centre for Intelligent Automation, Sri Lanka

Abstract

The construction industry is increasingly dependent on digital sensing, three-dimensional perception, automated decision-making, and intelligent operational control to improve productivity while reducing resource consumption and environmental impacts. Intelligent robotics can provide physical capabilities for inspection, material handling, positioning, and site monitoring, whereas large language models (LLMs) can support interpretation, reasoning, task coordination, and human–machine interaction. However, the integration of these capabilities remains constrained by the reliability of spatial perception, the complexity of construction environments, and the difficulty of translating high-level decisions into executable robotic actions. This research and review paper develops a conceptual framework for integrating intelligent robotics, three-dimensional computer vision, and LLM-based decision support for sustainable construction operations. The methodology synthesizes the provided literature on 3D deep learning, RGB-D semantic segmentation, point-cloud understanding, object detection, and instance segmentation. The analysis indicates that robust 3D perception should constitute the foundational layer of an LLM-enabled construction intelligence architecture. Point-cloud and RGB-D models can provide spatially grounded information, while LLM-based reasoning can transform such information into interpretable operational recommendations. The proposed framework positions the LLM as a decision-support and coordination layer rather than an autonomous source of physical truth. This distinction is important because sustainable construction requires decisions that simultaneously consider operational efficiency, material utilization, safety-related constraints, and environmental performance. The resulting architecture provides a

theoretical basis for integrating perception, reasoning, planning, and robotic execution while recognizing limitations associated with data quality, domain transfer, computational requirements, and decision reliability.

Keywords

Robotics, Large Language Models, Sustainable Construction, 3D Computer Vision, Point-Cloud Learning, Decision Support, Robotic Automation, Construction Operations, Semantic Segmentation

Introduction

Construction operations are characterized by dynamic workspaces, heterogeneous materials, continuously changing spatial configurations, and interactions between human workers, machinery, and temporary structures. These characteristics make construction substantially different from controlled industrial environments. Intelligent automation therefore requires more than robotic actuation: machines must perceive their surroundings, identify relevant objects, interpret spatial relationships, and adapt operational decisions to changing conditions. Developments in three-dimensional computer vision have significantly strengthened this capability by enabling machines to process RGB-D data, point clouds, voxel representations, and object-level spatial information.

Deep learning has become particularly important for extracting meaningful representations from three-dimensional construction-like environments. Surveys of 3D computer vision demonstrate the increasing use of deep learning for spatial perception and object understanding, while point-cloud research has established techniques for classification, segmentation, detection, and instance-level reasoning (Ioannidou et al., 2017; Bello et al., 2020; Guo et al., 2021). These developments create an important technological foundation for intelligent construction robotics.

Nevertheless, perception alone does not constitute intelligent decision support. A robot may identify an object or segment a work area without understanding the operational consequences of that observation. LLMs offer a complementary capability because they can process natural-language instructions, contextual information, operational constraints, and structured machine observations. Consequently, Intelligent Robotics and LLM-Based Decision Support for Sustainable Construction Operations can be conceptualized as a layered system in which 3D perception supplies grounded environmental information, an LLM interprets this information within an operational context, and robotic systems execute validated decisions.

The central problem addressed by this paper is therefore the integration gap between machine perception and operational reasoning. Existing

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

3D learning studies primarily address how machines recognize and segment spatial information, whereas the proposed research focuses on how such information could become an input to intelligent decision support for construction. The objective is not to claim experimentally demonstrated LLM performance, but to establish a research-grounded conceptual architecture using the provided literature as its technical foundation.

The significance of this framework lies in its potential to connect digital intelligence with physical construction activities. A perception system could, for example, identify material piles, machinery, structural components, or occupied work zones. An intelligent decision-support layer could then reason about task priorities, equipment utilization, material movement, and resource allocation before providing a recommendation to a robotic controller or human supervisor. In sustainable construction, such integration could support reduced unnecessary movement, improved material handling, better utilization of machinery, and more informed operational planning.

LITERATURE REVIEW

The provided literature establishes a progression from three-dimensional visual representation toward increasingly sophisticated spatial reasoning. Ioannidou et al. (2017) review advances in deep learning for computer vision involving 3D data, establishing the broader theoretical basis for using learned representations in spatial perception. Bello et al. (2020) subsequently examine deep learning approaches for 3D point clouds, demonstrating the importance of point-based representations for interpreting complex spatial environments. Guo et al. (2021) provide a comprehensive survey of deep learning for 3D point clouds and emphasize the diversity of tasks that can be addressed through learned spatial representations.

RGB-D semantic segmentation represents another important component of this technological progression. Fooladgar and Kasaei (2020) examine the transition from hand-crafted features toward deep convolutional neural networks for indoor RGB-D semantic segmentation. Their work is relevant to construction because semantic understanding requires systems to distinguish meaningful categories within spatial environments rather than merely detecting geometric patterns. A construction robot operating in a partially structured environment similarly requires semantic differentiation between objects and surfaces before task execution.

The literature also demonstrates a transition from semantic understanding toward instance-level and object-level reasoning. Hou et al. (2019) address 3D semantic instance segmentation of RGB-D scans, demonstrating the importance of distinguishing individual instances within a scene. Jiang et al. (2020) introduce PointGroup for dual-set point grouping in 3D instance segmentation, while Wang et al. (2018) investigate SGPN for similarity-based grouping

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

and 3D point-cloud instance segmentation. These approaches are theoretically significant for robotics because an autonomous machine must often distinguish between individual objects of the same category.

Object detection provides another layer of spatial intelligence. Deng et al. (2021) present Voxel R-CNN for voxel-based 3D object detection, demonstrating the use of voxelized representations for high-performance spatial detection. Yang et al. (2019) focus on learning object bounding boxes for 3D instance segmentation on point clouds. Together, these studies indicate that spatial perception can provide progressively richer representations, ranging from general scene understanding to individual object localization.

Zhong et al. (2021) further demonstrate the application of deep action learning to robust 3D segmentation across CT and MRI images. Although the medical imaging domain differs fundamentally from construction, the study is methodologically relevant because it illustrates how deep learning can support robust three-dimensional segmentation under varying imaging conditions. Its direct transfer to construction should not be assumed; instead, it provides conceptual evidence for the broader importance of robustness in 3D segmentation.

A major gap emerges from this synthesis. The provided literature concentrates primarily on perception: segmentation, grouping, detection, and spatial representation. It does not directly establish an integrated architecture connecting these outputs to LLM-based reasoning and physical construction robotics. This gap motivates the present conceptual framework. The proposed system therefore treats the existing 3D learning literature as a perception foundation and introduces an additional reasoning layer for operational decision support.

METHODOLOGY

3.1 Research Design

This study adopts a conceptual research-and-review methodology. Rather than conducting a new empirical experiment, it synthesizes the supplied literature and translates its technical implications into an integrated framework for intelligent construction operations. The methodological objective is to determine how 3D perception capabilities can logically support LLM-based decision-making and subsequently inform robotic action.

The framework is organized into four principal functional layers: perception, spatial interpretation, decision support, and execution. The architecture follows a sequential but iterative process in which observations are transformed into structured environmental information, interpreted through operational context, converted into decisions, and then communicated to robotic systems. Feedback from subsequent perception cycles can update the decision process.

3.2 Perception Layer

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

The first layer consists of sensors and 3D perception algorithms. RGB-D cameras, depth sensors, LiDAR-like systems, or equivalent spatial sensing technologies can generate geometric and visual observations. These observations may be represented as point clouds, voxels, or RGB-D maps.

The literature supports multiple complementary strategies for interpreting these representations. Point-cloud learning provides a foundation for extracting geometric features (Bello et al., 2020; Guo et al., 2021). Semantic segmentation can classify regions according to their functional or object categories, while instance segmentation can separate individual objects. Object-detection models can additionally provide spatial bounding information.

For a hypothetical construction scenario, a mobile robot could scan a work area and generate a point cloud. A segmentation model could distinguish floor surfaces, material stacks, equipment, and structural components. Instance-level processing could then differentiate individual material units. The result would not simply be an image but a structured representation of the physical environment.

3.3 Spatial Interpretation Layer

Raw perception outputs must be transformed into information that a decision-support system can use. This layer therefore converts detected objects and spatial relationships into structured representations such as:

Object = {category, location, dimensions, confidence, spatial relationships}.

Such representations are particularly important because LLMs are fundamentally language-oriented systems. Directly supplying unstructured point-cloud data to an LLM would not automatically produce reliable physical reasoning. Instead, a structured intermediary should translate spatial observations into interpretable facts.

For example, a system could convert a scene into a representation indicating that three material pallets are located within a defined work zone, one access path is partially obstructed, and a robotic vehicle is positioned near the material staging area. The LLM could then reason over this structured information rather than independently inventing spatial facts.

3.4 LLM-Based Decision-Support Layer

The LLM functions as a reasoning and communication layer rather than as an unrestricted controller. Its inputs may include structured perception outputs, task objectives, operational constraints, sustainability criteria, and human instructions.

A simplified decision formulation can be expressed as:

Decision = f(Environmental State, Task Objective, Resource Constraints, Sustainability Criteria, Safety Constraints).

The LLM can transform these inputs into recommendations such as task prioritization,

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

alternative material-routing strategies, inspection requests, or requests for additional sensing. The critical principle is that generated recommendations should remain grounded in verified machine observations.

For instance, if the perception layer identifies an obstructed material route, the LLM could compare alternative routes according to distance, estimated resource consumption, and task urgency. The robotic system would then receive a validated operational instruction rather than a free-form language response.

3.5 Robotic Execution Layer

The execution layer converts approved decisions into physical operations. These may include navigation, material transportation, inspection, object positioning, or other construction-support activities. The separation between reasoning and execution is essential because an LLM-generated recommendation should not automatically become a physical command.

A practical architecture can therefore employ a validation mechanism between the LLM and robot controller. The LLM proposes an action, a deterministic control system checks feasibility and constraints, and only then is the action transmitted to the robot. This architecture reduces the risk of language-level reasoning being interpreted as guaranteed physical truth.

3.6 Sustainability-Oriented Decision Model

Sustainability should be embedded within decision criteria rather than treated as an afterthought. A construction robot might select among several feasible routes, for example. A purely productivity-oriented system may minimize travel time, whereas a sustainability-oriented system can incorporate energy consumption, unnecessary material handling, equipment utilization, and operational delays.

A conceptual objective function can therefore be represented as:

$$\text{Minimize } C = w_1E + w_2T + w_3M + w_4R$$

where E represents energy/resource expenditure, T represents operational time, M represents unnecessary material movement, and R represents operational risk or constraint violation. The weights can be adapted to project priorities. The equation is conceptual rather than an empirically validated optimization model.

3.7 Human Oversight and Feedback

Human oversight remains essential because construction environments contain ambiguous conditions that may not be fully represented in sensor data. A supervisor can review high-impact recommendations, correct inaccurate scene interpretations, and modify task priorities.

The system should also maintain a feedback loop. After robotic execution, new sensor observations can determine whether the expected environmental change occurred. If a material stack was moved, for example, the perception system should verify the new location. This creates an iterative cycle of sensing, reasoning, validation, execution, and verification.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

Results

The conceptual synthesis produces four principal findings. First, three-dimensional perception represents the most appropriate technical foundation for integrating robotics with intelligent decision support. The reviewed literature demonstrates that deep learning can support semantic segmentation, instance segmentation, point grouping, object detection, and broader point-cloud understanding (Ioannidou et al., 2017; Fooladgar and Kasaei, 2020; Guo et al., 2021). This diversity of perception capabilities is necessary because construction decisions depend on both object identity and spatial relationships.

Second, the analysis indicates that perception and reasoning should be treated as separate but connected functions. Models such as Voxel R-CNN, PointGroup, SGPN, and 3D-SIS demonstrate increasingly detailed approaches to identifying and separating objects in three-dimensional environments (Deng et al., 2021; Jiang et al., 2020; Wang et al., 2018; Hou et al., 2019). However, identifying an object does not inherently determine what should be done with it. An LLM-based decision-support layer can provide contextual interpretation, but it should operate on structured and verified spatial information.

Third, sustainability can be incorporated as a decision criterion rather than merely an evaluation metric. When spatial perception identifies available routes, material locations, equipment positions, and work-zone conditions, a reasoning layer can compare alternatives according to operational and resource-related objectives. The framework therefore shifts intelligent construction robotics from isolated automation toward context-aware operational coordination.

Fourth, the literature reveals a substantial research gap between advanced 3D perception and integrated intelligent construction decision systems. The reviewed studies provide strong technical foundations for spatial understanding but do not establish a complete architecture for LLM-mediated construction robotics. Consequently, the proposed framework should be understood as a research direction requiring empirical validation rather than as a demonstrated production system.

The resulting architecture can be summarized as:

Sensing → 3D Perception → Structured Spatial State → LLM Decision Support → Constraint Validation → Robotic Execution → Verification → Updated Spatial State.

This architecture also reinforces the central proposition of Intelligent Robotics and LLM-Based Decision Support for Sustainable Construction Operations: intelligent physical automation becomes more useful when perception, reasoning, and execution are integrated while retaining explicit validation boundaries.

Discussion

The findings have both theoretical and practical implications. Theoretically, the framework extends the role of 3D computer vision from perception toward decision-oriented robotics. Existing research demonstrates increasingly sophisticated methods for interpreting three-dimensional environments, but the operational value of these representations depends on how effectively they are connected to task-level reasoning. The proposed architecture addresses this issue by positioning 3D learning as the environmental grounding mechanism for higher-level decision support.

The framework also clarifies the complementary roles of robotics and LLMs. Robotics provides physical embodiment, sensing, and action, whereas an LLM can provide flexible interpretation of goals, constraints, and structured observations. Neither component should be treated as sufficient independently. A robot without contextual reasoning may be operationally rigid, while an LLM without verified spatial grounding may produce decisions that are linguistically plausible but physically inappropriate. Thus, the strongest architecture is not an LLM-controlled robot in isolation but a layered cyber-physical system with explicit interfaces and validation.

From a practical perspective, the framework could support construction applications such as material logistics, automated inspection, site mapping, and work-zone coordination. For example, a robotic platform could detect a set of material objects using 3D segmentation, communicate their locations through a structured representation, and receive an LLM-generated recommendation concerning task sequencing. A conventional control layer could then determine whether the recommended action is physically feasible.

However, several limitations must be acknowledged. First, the provided literature does not directly validate LLM-based construction decision-making. Therefore, claims concerning productivity or sustainability improvements remain hypotheses. Second, construction environments differ considerably from the datasets used in many 3D computer-vision studies. Domain shift, occlusion, dust, lighting variation, dynamic objects, and incomplete sensor observations can reduce perception reliability. The robustness issues highlighted by 3D segmentation research are consequently highly relevant but cannot simply be assumed to transfer unchanged to construction.

Third, LLM reasoning introduces additional risks. Language models can misinterpret structured information, infer unsupported relationships, or generate recommendations that appear reasonable but violate physical constraints. A deterministic validation layer is therefore essential. Fourth, sustainability objectives may conflict with productivity objectives. A shorter route may not always minimize energy consumption, and minimizing material movement may conflict with task sequencing. Consequently, sustainable decision support requires explicit multi-objective optimization rather than assuming that one operational metric represents sustainability.

International Journal of Advance Scientific Research
(ISSN – 2750-1396)
VOLUME 06 ISSUE 08 Pages: 112-121
OCLC – 1368736135

Finally, Zhong et al. (2021) illustrates the broader importance of robust 3D segmentation under variable imaging conditions, but its medical-imaging context also demonstrates a transferability limitation. Techniques developed for one domain require construction-specific datasets and validation before operational deployment.

Overall, Intelligent Robotics and LLM-Based Decision Support for Sustainable Construction Operations should be understood as an integrated socio-technical architecture rather than simply an application of one artificial-intelligence model. Its effectiveness depends on the reliability of perception, quality of structured representations, transparency of decision processes, physical validation, and human oversight.

Conclusion

This research and review paper developed a conceptual framework for integrating intelligent robotics, 3D deep learning, and LLM-based decision support within sustainable construction operations. The synthesis of the provided literature demonstrates that contemporary 3D perception technologies offer important foundations for machine understanding of complex spatial environments. Semantic segmentation, instance segmentation, point grouping, object detection, and point-cloud learning collectively provide the environmental representations required by intelligent robotic systems.

The principal contribution of the framework is the explicit connection of these perception capabilities with an LLM-based decision-support layer. Rather than treating the LLM as an autonomous controller, the framework assigns it a contextual reasoning and coordination role. Structured spatial information is supplied to the reasoning layer, recommendations are generated according to task and sustainability objectives, deterministic constraints validate proposed actions, and robotic systems execute feasible decisions. Feedback from subsequent sensing completes the operational loop.

The framework also highlights an important research gap: advanced 3D perception and intelligent language-based decision support have substantial potential for integration, but their combined use in sustainable construction requires systematic empirical investigation. Future research should therefore develop construction-specific multimodal datasets, evaluate perception robustness under realistic site conditions, establish measurable sustainability objectives, and experimentally test LLM-to-robot interfaces under human supervision.

Ultimately, the future of intelligent construction is unlikely to depend on perception, language reasoning, or robotics in isolation. Greater value can emerge from their controlled integration into systems capable of observing physical environments, interpreting operational objectives, validating decisions, and executing actions while continuously updating their understanding of the site. Such an architecture provides a promising theoretical foundation for advancing sustainable, adaptive, and digitally integrated construction operations.

References

1. A. Ioannidou, E. Chatzilari, S. Nikolopoulos and I. Kompatsiaris, “Deep learning advances in computer vision with 3D data: A survey,” ACM Computer Survey, vol. 50, pp. 1–38, 2017.
2. B. Bello, S. Yu, C. Wang, J. M. Adam and J. Li, “Review: Deep learning on 3D point clouds,” Remote Sensing, vol. 12, pp. 1729, 2020.
3. F. Fooladgar and S. Kasaei, “A survey on indoor RGB-D semantic segmentation: From hand-crafted features to deep convolutional neural networks,” Multimedia Tools and Applications, vol. 79, pp. 4499–4524, 2020.
4. J. Deng, S. Shi, P. Li, W. Zhou, Y. Zhang et al., “Voxel R-CNN: Towards high performance voxel-based 3D object detection,” in AAAI Conf. on Artificial Intelligence, vol. 35, pp. 1201–1209, 2021.
5. J. Hou, A. Dai and M. Nießner, “3D-sis: 3D semantic instance segmentation of RGB-D scans,” in IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp. 4421–4430, 2019.
6. L. Jiang, H. Zhao, S. Shi, S. Liu, C. Fu et al., “Pointgroup: Dual-set point grouping for 3D instance segmentation,” in IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 4867–4876, 2020.
7. W. Wang, R. Yu, Q. Huang and U. Neumann, “SGPN: Similarity group proposal network for 3D point cloud instance segmentation,” in IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 2569–2578, 2018.
8. X. Zhong, M. Amrehn, N. Ravikumar, S. Chen, N. Strobel et al., “Deep action learning enables robust 3D segmentation of body organs in various CT and MRI images,” Scientific Reports, vol. 11, pp. 3311, 2021.
9. Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu et al., “Deep learning for 3D point clouds: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, A virtual conference, vol. 43, pp. 4338–4364, 2021.
10. B. Yang, J. Wang, R. Clark, Q. Hu, S. Wang et al., “Learning object bounding boxes for 3D instance segmentation on point clouds,” in Proc. of the 33rd Int. Conf. on Neural Information Processing Systems, Vancouver, BC, Canada, pp. 6740–6749, 2019.
11. K. Ramamurthy, N. Bellamkonda and N. Amanmadov, "ScalePulse: Combinatorial Llm Framework for Scalability Constraint," SoutheastCon 2026, Huntsville, AL, USA, 2026, pp. 1-6, doi: 10.1109/SoutheastCon63549.2026.11476075
12. Geo Philip, Paulson, Robotics-Enabled Sustainable Construction Management: A Socio-Technical Framework for Operational Efficiency, Digital Integration, and Sustainability Performance. Available at SSRN: https://ssrn.com/abstract=6845167 or http://dx.doi.org/10.2139/ssrn.6845167