Physical Ergonomics & Human Factors, Vol. 223, 2026, 20–27
https://doi.org/10.54941/ahfe1007792

Automated Ergonomic Problem and Solution Identification From Videos With a Knowledge-Retrieving Large Multimodal Model

Gunwoo Yong and SangHyun Lee

Dept. of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, 48109, USA

ABSTRACT

Workers across diverse industries experience a high prevalence of work-related musculoskeletal disorders (WMSDs), the leading cause of non-fatal injuries. To mitigate WMSDs, it is crucial for ergonomic experts to identify ergonomic problems and solutions. However, implementing such manual identification across diverse workplaces is challenging because it is time-consuming and resource-intensive, highlighting the need for accessible tools for on-site personnel. Recent advances in large multimodal models (LMMs) have demonstrated their potential to identify ergonomic problems and solutions, given their strong scene understanding capabilities. However, LMMs often generate plausible but incorrect information, known as hallucination, particularly when handling long-context video inputs. This limitation is critical because workers without ergonomic expertise cannot verify the correctness of identified problems and solutions, leading to ineffective or even harmful interventions. To address this, we aim to automatically identify ergonomic problems and solutions from videos, supported by guideline-based evidence, by applying an ergonomic knowledge-retrieving LMM. We developed an ergonomic knowledge retrieval pipeline that enables the LMM to retrieve ergonomic knowledge from a knowledge graph and ground its predictions accordingly. To evaluate the correctness of identification and the relevance of the retrieved knowledge, we used accuracy and context precision as evaluation metrics. Evaluation on 25 real-world workplace videos yielded an accuracy and context precision of 0.80, outperforming a state-of-the-art LMM. Our results highlight the importance of integrating ergonomic knowledge into LMMs in identifying ergonomic problems and solutions. Our knowledge-retrieving LMM automates ergonomic problem and solution identification grounded in verified knowledge, helping reduce WMSDs through easier, broader adoption.

Keywords: Ergonomic problem and solution identification, Large multimodal model, Knowledge retrieval

INTRODUCTION AND RELATED WORKS

Workers across diverse industries (e.g., health care, manufacturing, and construction) experience a higher incidence of work-related musculoskeletal disorders (WMSDs) than any other category of non-fatal injury. For the 2023–2024 period, the highest incidence of days away from work, job restriction,

Received February 7, 2026; Revised April 1, 2026; Accepted April 19, 2026; Available online July 20, 2026

or transfer (DART) cases in the U.S. was attributed to overexertion, repetitive motion, and bodily conditions, which are the primary drivers of WMSDs (BLS, 2026). To mitigate the prevalence of WMSDs, federal agencies such as the Occupational Safety and Health Administration (OSHA) recommend that ergonomic experts monitor workers and workplaces to identify ergonomic problems and their solutions (OSHA, 2020). However, implementing such manual identification is challenging because it requires significant time from experts, resulting in limited coverage of workers and workplaces and often yields inconsistent results depending on an individual's level of expertise (Li et al., 2020). This limitation highlights the need for accessible tools that can be used directly by on-site personnel, such as workers and managers.

In this context, large multimodal models (LMMs) emerge as an accessible tool when integrated with ubiquitous smartphone cameras. This is because LMMs can discern relationships between workers and their workplaces from imagery data, leveraging their robust scene understanding capabilities (Fan et al., 2024). However, LMMs often produce hallucinations, which are plausible but incorrect outputs, especially when processing longer contextual inputs, such as videos, and when performing domain-specific tasks (Huang et al., 2025; Li et al., 2025). Since workers and managers often lack the ergonomic expertise to distinguish correct outputs from hallucinated ones, such hallucinations may be accepted without verification, resulting in ineffective ergonomic solutions; therefore, reducing hallucinations is essential.

To address this limitation, we aim to identify ergonomic problems and their solutions from videos using guideline-based evidence by applying an ergonomic knowledge-retrieving LMM. Our LMM contributes to automating ergonomic problem and solution identification grounded in verified knowledge, offering an accessible tool to help reduce WMSDs in workplaces.

DEVELOPMENT OF ERGONOMIC KNOWLEDGE-RETRIEVING LMM

To enable the identification of ergonomic problems and solutions grounded in verified ergonomic knowledge from workplace videos, we proposed an ergonomic knowledge-retrieving LMM. We first built a knowledge base that contains ergonomic knowledge using publicly available ergonomic guidelines (e.g., "Simple Solutions" from the National Institute for Occupational Safety and Health (Albers and Estill, 2007)). Afterward, to guide an LMM to leverage the ergonomic knowledge stored in our knowledge base, we developed an ergonomic knowledge retrieval pipeline. Our pipeline is designed to extract the knowledge necessary for ergonomic problem and solution identification from workplace videos and corresponding worker motion.

Figure 1 presents an overview of the ergonomic problem and solution identification process of the proposed ergonomic knowledge-retrieving LMM. A workplace video is first processed to obtain 3D human motion data through pose estimation. The video and the extracted motion data are then provided as inputs to the knowledge retrieval pipeline, which retrieves relevant ergonomic knowledge from the knowledge base. Based on the retrieved ergonomic knowledge, the final LMM identifies the corresponding ergonomic problem and its solution. Details of the knowledge base and the knowledge retrieval pipeline are described in the following paragraphs.

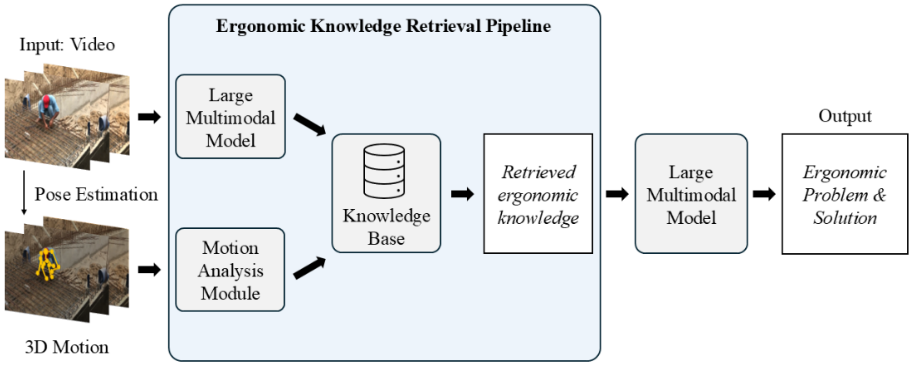


Figure 1: Ergonomic problem and solution identification process of the proposed ergonomic knowledge-retrieving LMM (youtube.com/@surveyingengineering designinfo).

As the knowledge base, we selected a knowledge graph, as it can show structured relationships among the information required for ergonomic problem and solution identification (Peng et al., 2023), thereby facilitating ergonomic knowledge retrieval. Given that the essential information for ergonomic problem and solution identification includes tasks, postures, ergonomic risk factors, symptoms, and controls (Albers and Estill, 2007; Torma-Krajewski et al., 2009), our knowledge graph is constructed around these elements. Figure 2 illustrates the template of the proposed knowledge graph, along with an example subgraph.

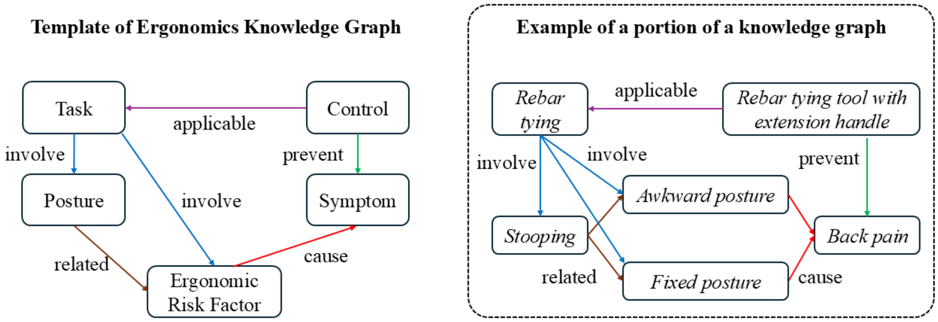


Figure 2: Template of the ergonomic knowledge graph and an example.

Our ergonomic knowledge retrieval pipeline consists of three components: 1) an LMM; 2) a motion analysis module; and 3) the knowledge graph. The LMM is employed to identify the task being performed and ergonomic risk factors unrelated to worker postures (e.g., the use of vibration tools) from videos. Posture-related ergonomic risk factors are not identified by LMMs because they have inherent limitations in precisely analyzing human body postures (Li et al., 2024). Accordingly, posture information and posture-related risk factors are identified using a dedicated motion analysis module. Our motion analysis module takes 3D human motion data obtained from 3D pose estimation as input, and determines whether awkward postures are repetitive or fixed over extended periods. Using this module, we can focus on identifying severe, repetitive, or fixed, awkward postures to support accurate ergonomic problem and solution identification. Specifically, for each video frame, we compute body joint angles associated with major awkward postures commonly observed in industrial settings (e.g., stooping, squatting, kneeling, arm raising, and neck bending) from 3D human motion data. Afterward, we compute a self-similarity matrix from the time-series patterns of joint angles to determine whether specific awkward postures are repetitive or fixed. The pose-based self-similarity matrix can capture repetitive or fixed posture patterns by measuring pairwise pose similarity across time (Chen and Yu, 2024). Using this representation, we identify: (1) repetitive postures by checking whether a specific posture exhibits a periodic pattern occurring at least four times within one minute; and (2) fixed postures by checking whether highly similar postures are sustained over a one-minute duration. These criteria are defined based on an established ergonomic risk assessment tool, specifically the Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000). Finally, based on the identified task, posture, and ergonomic risk factors, we retrieve the corresponding symptom and control measure from the knowledge graph by leveraging the encoded relationships among these elements. The relationships among all retrieved elements constitute ergonomic knowledge, which is then injected into the final LMM to enable the identification of ergonomic problems and solutions based on verified ergonomic knowledge. For the backbone LMM, we selected Gemini-3-Pro (Google DeepMind, 2025) for its superior performance in processing video data.

TESTING

Testing Dataset Preparation

Our testing objective was to evaluate whether our knowledge-retrieving LMM can identify ergonomic problems and solutions from videos grounded in ergonomic knowledge. To this end, we first collected a set of workplace videos that show guideline-based ergonomic problems to assess the potential of our LMM in real-world settings. Specifically, we collected 25 workplace videos from the web (e.g., YouTube). Each video depicts a single ergonomic problem and has a duration of at least one minute, as sufficient temporal context is required to identify repetitive and fixed postures. The videos cover a diverse range of tasks (e.g., overhead drilling and brick laying) and include various ergonomic risk factors, such as awkward posture, fixed posture, repetition, excessive force application, localized pressure, and vibration. The videos were used exclusively for testing to isolate and assess performance improvements attributable solely to ergonomic knowledge retrieval, without the confounding effects of additional training.

Evaluation Metric

To evaluate both whether the LMM correctly identifies ergonomic problems and solutions and whether its outputs are supported by retrieved ergonomic knowledge, we used two evaluation metrics. Accuracy was used to assess correct identification, while context precision assessed whether the retrieved knowledge was accurate and relevant to the inputs (Es et al., 2023). Both metrics were manually evaluated by two reviewers, who independently assessed each output using publicly available ergonomic guidelines as the reference. The inter-reviewer agreement ratio was 0.93, indicating a high level of agreement between the two assessments. To ensure a conservative evaluation, when discrepancies occurred, the lower of the two scores was adopted as the final score. To examine whether ergonomic knowledge retrieval improves ergonomic problem and solution identification beyond what can be achieved with pretrained knowledge alone, we compared our LMM with its backbone model, Gemini-3-Pro, using accuracy. Context precision was not compared, as the backbone model cannot perform knowledge retrieval.

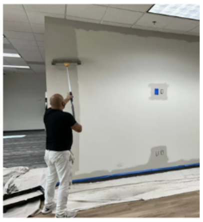
RESULTS AND DISCUSSION

Table 1 reports the accuracy and context precision of our ergonomic knowledge-retrieving LMM, along with the accuracy of the baseline model, Gemini-3-Pro. The reason for selecting Gemini-3-Pro as the baseline is that it is a leading state-of-the-art LMM capable of processing video inputs with strong generalization performance. Our ergonomic knowledge-retrieving LMM achieved an accuracy of 0.80, substantially outperforming the baseline model, which scored 0.56. This performance gap indicates that incorporating external ergonomic knowledge significantly improves the LMM's ability to correctly identify ergonomic problems and solutions, compared to relying solely on pretrained knowledge. Furthermore, our LMM's accuracy represents a high level of performance relative to the consistency observed in judgments of ergonomic experts, given that the average agreement on identified ergonomic problem-solution pairs between experts is approximately 64.93% (Jones et al., 1999). In addition to accuracy, our LMM achieved a context precision of 0.80. The most widely used search platform, Google, reported an overall user satisfaction score of 0.81, which reflects users' perceived relevance and usefulness of retrieved search results (ACSI, 2024). In this context, a precision of 0.80 indicates that our LMM can retrieve ergonomic knowledge with a level of relevance comparable to what humans commonly experience.

Table 1: The accuracy and context precision of the proposed ergonomic knowledge-retrieving LMM, and the accuracy of the baseline model.

Models	Accuracy	Context Precision
Ergonomic knowledge-retrieving LMM	0.80	0.80
Baseline (Gemini-3-Pro)	0.56	-

Figure 3 presents examples of successful ergonomic problem and solution identification by our LMM using retrieved ergonomic knowledge, in contrast to the baseline model's failure case. Note that the photographs shown in the figure are not from the actual test videos but are copyright-free images included solely for illustrative purposes. Our LMM correctly identified ergonomic problems and solutions in alignment with ergonomic guidelines by leveraging retrieved ergonomic knowledge, whereas the baseline model identified incorrect problems and solutions. For example, in the drywall finishing task, a substantial portion of the video shows the worker with arms raised while applying pushing force. Our LMM correctly identifies the corresponding ergonomic problem and its solution, whereas the baseline model incorrectly focuses on back bending as the dominant ergonomic risk factor. As a result, it proposes an ineffective ergonomic solution that fails to address the primary problem and may introduce additional symptoms (e.g., knee pain).



Correctly Identified Ergonomic Problem and Solution by Our LMM
Retrieved Knowledge: *Drywall finishing involves awkward posture (arms raised), repetition, and excessive force for pushing. Repetitive arms raised posture with excessive force for pushing causes shoulder pain. Using a pneumatic drywall finishing system prevents excessive force for pushing and shoulder pain.*
Problem: *The worker is performing drywall finishing with his arms raised in an awkward posture while applying force repetitively, resulting in shoulder pain.*
Solution: *Use a pneumatic drywall finishing system.*

Incorrectly Identified Ergonomic Problem and Solution by Baseline (Gemini 3 Pro)
Problem: *The worker is bending his back to apply cement to the wall, resulting in back pain.*
Solution: *The worker should squat to keep the back straight.*

Correctly Identified Ergonomic Problem and Solution by Our LMM
Retrieved Knowledge: *Using hand-held powered tools involves vibration. Vibration causes hand pain. Using reduced vibration tools with anti-vibration gloves prevents vibration. Using reduced vibration tools with anti-vibration gloves prevents hand pain. Using reduced vibration tools with anti-vibration gloves is applicable to when using hand-held powered tools.*
Problem: *The worker is using a hand-held powered tool, exposing his hands to vibration, resulting in hand pain.*
Solution: *Use reduced vibration tools with anti-vibration gloves.*

Incorrectly Identified Ergonomic Problem and Solution by Baseline (Gemini 3 Pro)
Problem: *The worker is continuously walking in circles to guide the plate compactor, which is a repetitive and monotonous task that can lead to fatigue and leg pain.*
Solution: *Drive a stake into the center of the work area and tether the plate compactor to it using a rope, allowing the machine to compact the gravel automatically without manual guidance.*

Figure 3: Correctly identified ergonomic problems and solutions by our LMM and incorrectly identified ergonomic problems and solutions by the baseline model (Gemini-3-Pro) (pxhere.com; finefinishescustompainting.com).

CONCLUSION

To automatically identify ergonomic problems and solutions with verified knowledge, we present an ergonomic knowledge-retrieving LMM. To this end, we develop an ergonomic knowledge retrieval pipeline that enables the LMM to retrieve relevant ergonomic knowledge from videos using a knowledge graph constructed based on ergonomic guidelines. Tested on 25 real-world workplace videos, our LMM achieved an accuracy of 0.80 and a context precision of 0.80, outperforming a state-of-the-art LMM that cannot retrieve knowledge from videos. Our results demonstrate that enabling the retrieval of ergonomic knowledge effectively reduces hallucinations and improves the accuracy of LMMs in identifying ergonomic problems and solutions. The main contribution of our ergonomic knowledge-retrieving LMM is the automation of ergonomic problem and solution identification grounded in verified knowledge, facilitating broader adoption across diverse workplaces to help protect workers from WMSDs.

REFERENCES

ACSI (2024) "ACSI Search and Social Media Study 2024." American Customer Satisfaction Index.

Albers, J.T. and Estill, C.F. (2007) *Simple Solutions: Ergonomics for Construction Workers*. 2007–122. NIOSH (The National Institute for Occupational Safety and Health).

BLS (2026) *Employer-Reported Workplace Injuries and Illnesses - 2023–2024*. USDL-2609191. U.S. Bureau of Labor Statistics.

Chen, X. and Yu, Y. (2024) "Automatic repetitive action counting for construction worker ergonomic assessment," *Automation in Construction*, 167, p. 105726. Available at: https://doi.org/10.1016/j.autcon.2024.105726.

Es, S., James, J., Espinosa-Anke, L. and Schockaert, S. (2023) "RAGAS: Automated Evaluation of Retrieval Augmented Generation." arXiv. Available at: https://doi.org/10.48550/arXiv.2309.15217.

Fan, C., Mei, Q., Wang, X. and Li, X. (2024) "ErgoChat: a Visual Query System for the Ergonomic Risk Assessment of Construction Workers." arXiv. Available at: https://doi.org/10.48550/arXiv.2412.19954.

Google DeepMind (2025) *Gemini 3 Pro*, *Google DeepMind*. Available at: https://deepmind.google/models/gemini/pro/ (Accessed: February 3, 2026).

Hignett, S. and McAtamney, L. (2000) "Rapid Entire Body Assessment (REBA)," *Applied Ergonomics*, 31(2), pp. 201–205. Available at: https://doi.org/10.1016/S0003-6870(99)00039-3.

Huang, L. Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., et al. (2025) "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Transactions on Information Systems,* 43(2), pp. 1–55. Available at: https://doi.org/10.1145/3703155.

Jones, J.A.R., Cockcroft, A. and Richardson, B. (1999) "The Ability of Non-ergonomists in the Health Care Setting to Make Manual Handling Risk Assessments and Implement Changes," *Applied Ergonomics*, 30(2), pp. 159–166. Available at: https://doi.org/10.1016/S0003-6870(97)00018-5.

Li, J., Wu, M., Jin, Z., Chen, H., Ji, J., Sun, X., Cao, L., and Ji, R. (2025) "MIHBench: Benchmarking and Mitigating Multi-Image Hallucinations in Multimodal Large Language Models." arXiv. Available at: https://doi.org/10.48550/ARXIV.2508.00726.

Li, L., Martin, T. and Xu, X. (2020) "A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders," *Applied Ergonomics*, 87, p. 103138. Available at: https://doi.org/10.1016/j.apergo.2020.103138.

Li, Y., Hou, R., Chang, H., Shan, S. and Chen, X. (2024) "UniPose: A Unified Multimodal Framework for Human Pose Comprehension, Generation and Editing." arXiv. Available at: https://doi.org/10.48550/arXiv.2411.16781.

OSHA (2020) *Ergonomics - Identify Problems*, *Occupational Safety and Health Administration*. Available at: https://www.osha.gov/ergonomics/identify-problems (Accessed: May 22, 2023).

Peng, C., Xia, F., Naseriparsa, M. and Osborne, F. (2023) "Knowledge Graphs: Opportunities and Challenges," *Artificial Intelligence Review*, pp. 1–32. Available at: https://doi.org/10.1007/s10462-023-10465-9.

Torma-Krajewski, J., Steiner, L.J. and Burgess-Limerick, R. (2009) *Ergonomics processes: implementation guide and tools for the mining industry.* 2009–107. NIOSH: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. Available at: https://doi.org/10.26616/NIOSHPUB2009107.