Journal of Science and Technology in Civil Engineering, HUCE, 2024, 18 (2): 28–44

A NOVEL INTEGRATED SYSTEM UTILIZING A DEEP LEARNING APPROACH AND THE NIOSH LIFTING INDEX TO ENHANCE CONSTRUCTION SAFETY

Quangdung Tran[a,*], Nguyen Anh Duc[a]

[a] *Faculty of Building and Industrial Construction, Hanoi University of Civil Engineering, 55 Giai Phong road, Hai Ba Trung district, Hanoi, Vietnam*

Article history:
Received 05/6/2024, Revised 18/6/2024, Accepted 21/6/2024

Abstract

Musculoskeletal disorders are a prevalent issue in the construction industry, largely due to the physical demands of object-lifting tasks. Effective ergonomic assessment is crucial for preventing these injuries and enhancing workplace safety. This study introduces a novel integrated system that utilizes advanced computational models to assess the NIOSH Lifting Index in real-time, offering a significant improvement over traditional ergonomic assessment methods. The system combines pose estimation, object detection, cycle counting, and Long Short-Term Memory modeling to provide dynamic, real-time evaluations of lifting practices. The integrated approach allows for continuous monitoring and analysis of lifting tasks, providing immediate feedback that can be used to adjust working conditions proactively. This system was tested in a controlled environment, demonstrating high accuracy in predicting the lifting index and identifying ergonomic risks with impressive precision and recall metrics. The practical applications of this system in real-world settings suggest substantial benefits for improving safety standards and reducing the incidence of Musculoskeletal disorders on construction sites. The study also explores the challenges faced during the implementation of the system, including limitations related to pose estimation accuracy and the requirement for predefined object weights in detection processes. Finally, future research directions were also discussed.

Keywords: Musculoskeletal disorders; NIOSH's lifting equation; construction safety; real-time monitoring; ergonomic assessment.

https://doi.org/10.31814/stce.huce2024-18(2)-03 © 2024 Hanoi University of Civil Engineering (HUCE)

1. Introduction

Musculoskeletal disorders (MSDs), particularly low back pain (LBP), are among the most prevalent occupational health issues faced by workers in the construction industry. These disorders not only lead to significant health-related expenses but also impact productivity due to lost workdays and increased disability claims [1]. The physical demands and dynamic nature of construction work necessitate effective preventive measures and reliable methods for assessing ergonomic risks associated with manual object-lifting tasks.

The National Institute for Occupational Safety and Health (NIOSH) Lifting Equation has been extensively adopted as a tool for evaluating the risk factors contributing to LBP. This equation, known for its utility in calculating a Lifting Index (LI), has become a cornerstone in the ergonomic assessment of lifting tasks. The LI provides a quantitative measure of the physical demand of object-lifting activities, where higher values correlate with an increased risk of developing LBP. Studies consistently validate the relationship between a high LI and the occurrence of LBP, reinforcing the importance of this tool in occupational health assessments [2, 3].

*Corresponding author. *E-mail address:* dungtq@huce.edu.vn (Tran, Q.)

The evolution of ergonomic risk assessments, particularly those aimed at preventing MSDs in workplace settings, has seen significant advancements through the integration of technology [4]. Firstly, the automation of the NIOSH lifting equation parameters initially utilized depth camera technologies like the Microsoft Kinect [5], which despite its rapid and robust performance, suffered from accuracy limitations due to its frame-by-frame analysis [6]. Efforts to correct these inaccuracies involved complex error-correction models that, while improving estimates, highlighted the technology's limitations in practical applications. Secondly, the integration of wearable sensors was seen as a promising advancement, with studies demonstrating their utility in capturing real-time, biomechanical data that could potentially transform ergonomic risk assessments [7]. However, wearable sensors have faced deployment challenges. They can be intrusive, affecting workers' convenience and attitude towards their use. This is compounded by the difficulty of maintaining these devices in operational conditions without interfering with worker productivity or comfort [8]. Additionally, the use of fixed and calibrated cameras, although beneficial for controlled data collection, does not reflect the dynamic and often unpredictable conditions of actual construction sites [9]. In response to these above challenges, recent research has pivoted towards employing advanced computational techniques, including machine learning and computer vision, to overcome the limitations of traditional sensors and manual assessments. Machine learning models, particularly those utilizing decision trees, support vector machines, and neural networks, have shown potential in accurately classifying ergonomic risks from biomechanical data collected via less intrusive methods like IMUs placed on the body [10].

Moreover, the development of video-based AI systems for object-lifting risk assessment indicates a shift towards non-wearable, non-invasive methods that can provide accurate, real-time feedback without the drawbacks associated with sensor-based systems [11]. These systems leverage the power of deep learning and computer vision to analyze video data unobtrusively, offering a practical solution to the limitations posed by wearable and fixed-camera systems. The trajectory of ergonomic risk assessment technologies suggests a growing reliance on sophisticated algorithms and machine learning models that can adapt to the complexities of real-world environments.

In response to the inherent shortcomings of manual calculations of the Lifting Index (LI), such as susceptibility to errors, inability to provide real-time alerts, and challenges in continuously monitoring workers throughout their shifts, this present study incorporates advanced technological solutions. The study introduces an integrated system that leverages machine learning techniques to automate the calculation of the LI. By utilizing real-time data acquisition from cameras, the system can dynamically assess the risk levels associated with manual object-lifting tasks. This approach not only reduces the potential for human error in calculations but also enables immediate feedback to workers and supervisors, thereby enhancing the potential for preventive interventions. Additionally, the continuous data collection capability of this system allows for comprehensive monitoring throughout entire work shifts, addressing critical gaps in traditional manual methods. This methodology thus promises to significantly improve the precision, reliability, and practicality of ergonomic risk assessments in the construction industry.

2. Literature review

2.1. Introduction, history, and influences of NIOSH lifting guidelines

Construction workers are prone to musculoskeletal disorder, among which low back pain is among the most popular [12] and low back injuries account for the majority of the total cost because of sick day off, clinical cost and insurance cost [13, 14]. One of the most impactful reasons of low back pain is repetitive heavy-object lifting task [15], in which, biomechanical factor postures and exerted force influence the seriousness of the disorders [7, 16]. Therefore, one way to reduce the risk to low back injuries or musculoskeletal disorders in general is to analyze biomechanical factors and external load in assessing task and designing tasks. Roman-Liu (2014) synthesized a resourceful review of approaches to assess external load based on parts of the body and the nature of tasks, such as NIOSH, REBA – Rapid Entire Body Assessment…

NIOSH was initially introduced in 1981, updated in 2021 and has been a cornerstone in the ergonomic assessment of lifting tasks within workplace settings [17, 18]. This tool is designed to evaluate the physical demands of manual lifting tasks and their potential risk for causing low back pain (LBP), a prevalent issue among workers in various industries. The revised version, known as the Revised NIOSH Lifting Equation (RNLE), incorporates a broader range of variables including the weight of the load, the conditions under which the lifting occurs, and the posture of the lifter, to provide a more comprehensive risk assessment [2]. Despite its widespread use, the RNLE is not without its critics [2, 3]. Practical limitations such as the need for precise measurements and the complexity of calculations have been highlighted as barriers to its application, especially in dynamic and unpredictable field settings like construction sites [19]. Additionally, factors such as lifting frequency and the horizontal reach required by the task have been shown to significantly influence the recommended weight limits, often diverging from real-world capabilities and task demands [20, 21].

While the NIOSH Lifting Equation has played a pivotal role in enhancing our understanding and management of ergonomic risks associated with manual lifting, its application in modern industrial environments requires further adaptation and integration with new technologies. This evolution will enable more accurate, real-time assessments, contributing to safer workplace practices and the prevention of musculoskeletal disorders among workers. The ongoing research and development in this area reflect a concerted effort to align traditional ergonomic assessments with contemporary workplace needs and technological advancements.

2.2. Detailed guidelines of the RNLE

In building up the RNLE, some notions need to be defined (Fig. 1 depicts some of the notions):

- Lifting Task: the act of manually grasping an object of definable size and mass with two hands, and vertically moving the object without mechanical assistance.

- Load Weight (L): Weight of the object to be lifted.

- Horizontal Location (H): Hand distance from ankle midpoint, measured in inches or centimeters (at lift start and end).

- Vertical Location (V): Distance of the hands above the floor, measure at the origin and the destination of the lift.

- Vertical Travel Distance (D): Absolute value of the difference between the vertical heights at the destination and origin of the lift.

- Asymmetry Angle (A): Angular displacement from worker's mid-sagittal plane at lift start/end, in degrees (measured at lift origin and destination). Asymmetry angle is based on load position relative to mid-sagittal plane, not foot position or body twist extent.

- Neutral Body Position: the position of the body when the hands are directly in front of the body and there is minimal twisting at the legs, torso, or shoulders.

- Lifting Frequency (F): Average number of lifts per minute over a 15 minute period.

- Lifting Duration: Categorized into three tiers based on work-time and recovery-time distribution (work pattern). Classifications include short (1 hour), moderate (1–2 hours), or long (2–8 hours), determined by the work pattern.

- Coupling Classification: Hand-to-object coupling quality classification (e.g., handle, cut-out, or grip) as good, fair, or poor.

- Significant Control: situations demanding precise load placement at the lift's destination. Typically, this occurs when the worker needs to (1) re-grasp the load near the destination, (2) momentarily hold the object there, or (3) carefully position or guide the load.

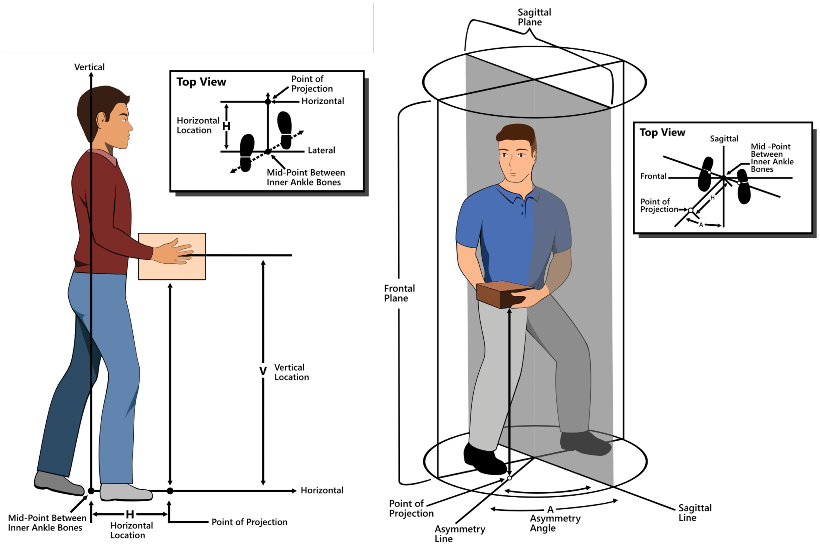


Figure 1. Graphic representation of the LI's elements [18]

The RNLE is defined as below:

$$RWL = LC \times HM \times VM \times DM \times AM \times FM \times CM$$

Then the Lifting Index is calculated as:

$$LI = \frac{Load\ Weight}{Recommended\ Weight\ Limit} = \frac{L}{RWL}$$

where LC is load constant – set at 23 kg, HM is horizontal multiplier = $25/H$ (metric) or $10/H$ (U.S. Customary), VM is vertical multiplier = $1 - (0.003|V - 75|)$ (metric) or $1 - (0.0075|V - 30|)$ (U.S. Customary), DM is distance multiplier = $0.82 + (4.5/D)$ (metric) or $0.82 + (1.8/D)$ (U.S. Customary), AM is asymmetric multiplier = $1 - (0.0032A)$ (both metric and U.S. Customary), FM is frequency Multiplier (taken from NIOSH table), CM is coupling multiplier (taken from NIOSH table)

After LI is calculated, the NIOSH recommend to interpret the result as follows:

- $LI < 1$: Indicates biomechanical stress below the recommended limit, suggesting low injury risk.

- $LI \approx 1$: Suggests stress nearing the limit; monitor and consider interventions.

- $LI > 1$: Indicates stress exceeds the limit, posing a high injury risk; immediate interventions needed. Furthermore, NIOSH elaborates the range of $1.0 < LI < 3.0$ is the medium risk and $LI > 3.0$ is the classified as high risk.

NIOSH offers tables for users to record worker measurements and provides mobile applications compatible with iOS and Android platforms. However, utilizing these tools requires manual measurement of each component and inputting the data into the forms.

3. Methodology

This section outlines the comprehensive methodology employed to enhance construction safety through the integration of deep learning techniques for pose estimation and LSTM modeling for real-time NIOSH Lifting Index estimation from video data. The methodology comprises two primary steps: Pose Estimation and LSTM Model Training. The entire process is summarized in Fig. 2.

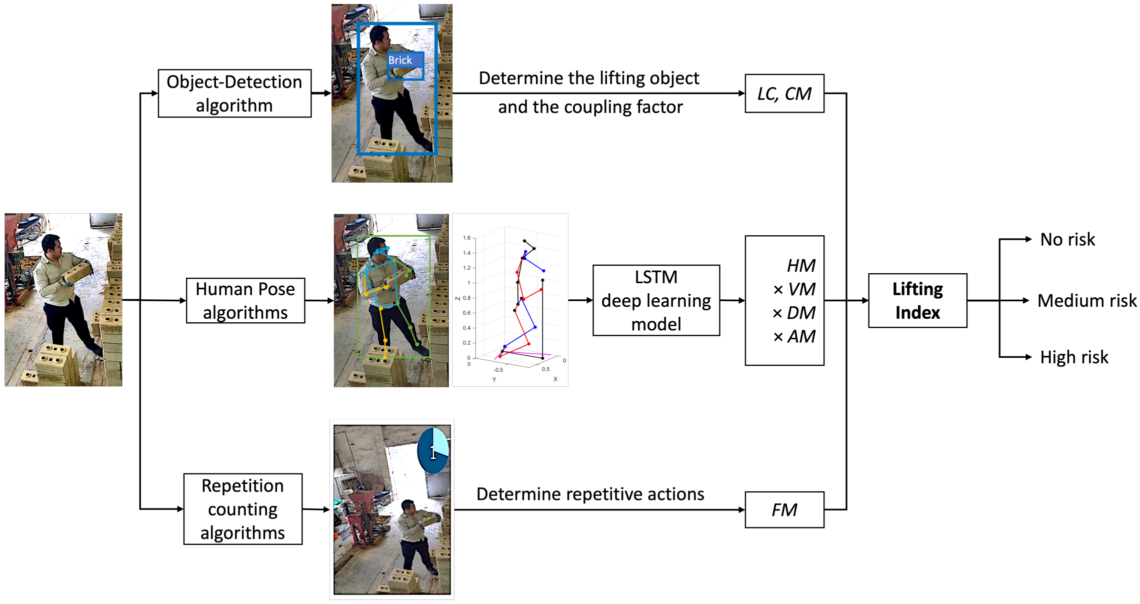


Figure 2. The planned process for automatic NIOSH's lifting index calculation

3.1. Pose Estimation

The initial step involves estimating the pose of workers utilizing sophisticated deep learning techniques. This process is subdivided into three sequential stages:

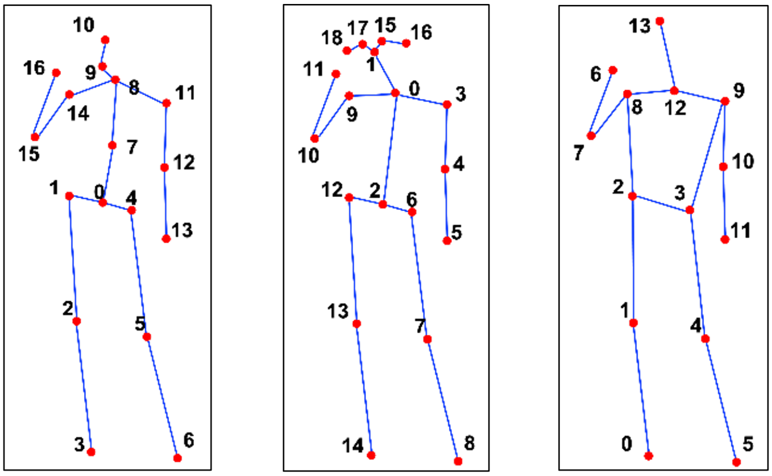
a. Human Detection and Skeleton Estimation

- RTMdet Model: The authors employ the RTMdet model [22] for human detection within video frames. This model, trained on the COCO-person dataset [23], has demonstrated exceptional performance. The COCO-person dataset, comprising 118,000 training images and 5,000 validation images, provides a robust foundation for detecting human instances.

- RTMpose Model: Subsequently, the RTMpose model [24] is applied to estimate the 2D skeletal poses of the detected human subjects. Trained on the Body8 dataset, which amalgamates data from various sources including AI Challengers, MS COCO, CrowPose, MPII, sub-JHMDB, Halpe, PoseTrack18, and OCHuman, RTMpose ensures accurate estimation of skeletal structures.

- VideoPose3D: The 2D skeletal poses are then transformed into their corresponding 3D, representations using the VideoPose3D model [25]. The dataset and the type of the skeletal posture is the Human3.6M, a popular dataset used for health safety research. This limitation in the number of points (17 points) helps to eliminate the smallest connections related to the details of the hands and feet, going deeper into the kinematic chain related to the wrist and ankle joints [26]. The Human3.6 dataset also contains more important information needed for health safety studies when compared to other datasets (Fig. 3).

Both skeletons from the CMU Panoptic and Campus and Shelf datasets do not have points that represent the curvature of the spine. While the data from the CMU Panoptic set has too many unnecessary facial points [27], the Campus and Shelf set has only one point, which does not represent the direction of head rotation [28]. The 3D Whole-Body datasets can also detect posture, however these data are large in size due to the need to reconstruct the entire detected body [29].



(a) Human3.6M (b) CMU Panoptic (c) Campus and Shelf

Figure 3. Comparison of different 3D skeletal models

b. Normalization of 3D Skeletal Postures

- Following the extraction of 3D skeletal postures from the video data, a normalization process is conducted. A conversion matrix is applied to standardize the orientation of the 3D skeletal postures, ensuring uniformity across varying camera perspectives.

3.2. LSTM Model Training

The subsequent step involves training an LSTM model [30] to estimate the NIOSH Lifting Index (LI) in real-time from the processed video data. This phase encompasses three concurrent processes:

a. LSTM Training

- LSTM Model: The LSTM model is trained to predict the value of $HM \times VM \times DM \times AM$ based on the 3D skeletal posture data obtained from the previous step. The model is trained on a meticulously curated dataset comprising 17-point 3 (x, y, z) 17,490 instances, ensuring a robust training regime. LSTM's inherent capability to capture temporal dependencies in sequential data renders it the ideal choice for this task.

b. Object Detection

- YOLOv8: Object detection, particularly for identifying the object being lifted (LC) and the coupling variable, is executed using the YOLOv8 algorithm [31]. Renowned for its efficiency and accuracy in object detection tasks, YOLOv8 ensures precise identification of lifting objects and coupling factors.

c. Cycle Counting

- Repnet Algorithm: The Repnet algorithm [32] is leveraged to automatically determine the cycle count of the lifting task (FM variable). Developed by Dwibedi et al. [33], Repnet's efficacy in accurately counting lifting cycles from video data has been extensively validated, making it an indispensable component of our methodology.

By integrating advanced deep learning techniques for precise pose estimation and LSTM modeling for real-time NIOSH Lifting Index estimation, our methodology offers a holistic approach to enhancing construction safety in lifting tasks.

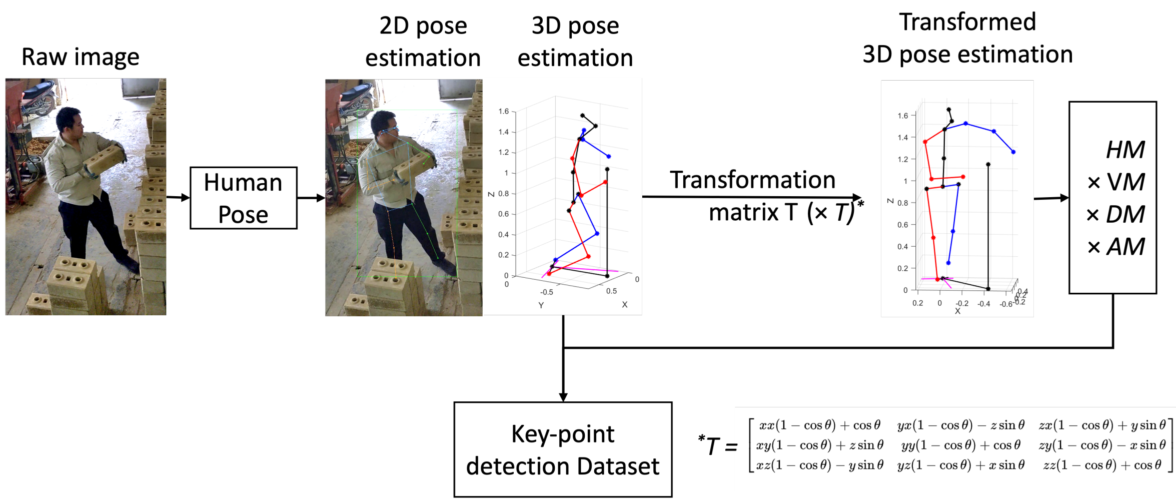
4. Results and discussions

4.1. The pose estimation results

The initial stage of the present methodology involved the use of the RTMpose model, which performed exceptionally well in generating 2D skeletal poses from video footage of construction workers engaged in various lifting tasks. These poses were accurate and detailed, providing a solid foundation for further analysis.

Upon obtaining the 2D skeletal poses, the VideoPose3D model was applied to convert these into 17-joint 3D skeletal poses. This transformation was based on logical geometrical relationships, such as ensuring the alignment of the pelvis center on the ground fell on the line between the two feet. To achieve this, 3D poses were straightened using transformation matrices, providing standardized and accurate 3D representations of the workers' postures.

With the 3D poses established, all coordinates of the 17 joints were accurately identified. This allowed for the automatic computation of critical components like the Horizontal Multiplier (HM), Vertical Multiplier (VM), Distance Multiplier (DM), and Asymmetric Multiplier (AM), essential for the lifting index calculations. Fig. 4 illustrates the entire process from capturing the raw footage to building the comprehensive 3D pose dataset.



$$T = \begin{bmatrix} xx(1-\cos\theta)+\cos\theta & yx(1-\cos\theta)-z\sin\theta & zx(1-\cos\theta)+y\sin\theta \\ xy(1-\cos\theta)+z\sin\theta & yy(1-\cos\theta)+\cos\theta & zy(1-\cos\theta)-x\sin\theta \\ xz(1-\cos\theta)-y\sin\theta & yz(1-\cos\theta)+x\sin\theta & zz(1-\cos\theta)+\cos\theta \end{bmatrix}$$

Figure 4. Entire process of building the 3D pose dataset

To evaluate the precision of the 3D poses and the calculated components of the lifting index, the authors classified worker postures into four distinct groups, presented in Fig. 5: standing and lifting (T01), lowering to pick up (T02), walking while holding an object (T03), and placing down the object (T04). The study then compared measurements such as height (H), vertical (V), distance (D), and angle (A) between manual calculations following NIOSH guidelines and those derived automatically from the 3D pose coordinates. Table 1 provides a detailed analysis of these comparisons, showing the percentage errors across different cases. In Table 1, the results demonstrate variability in the accuracy of the algorithmic calculations across different measures and tasks. For example, significant errors are observed in the angular measures (A), which show the highest discrepancies, particularly in tasks T01 and T04. Despite some errors, the overall performance of the algorithm is deemed satisfactory by experts, as all deviations fall within acceptable limits for practical applications.

Fig. 6 offers a visual representation of the transition from raw footage to 2D skeletal poses, and ultimately to refined 3D skeletal poses, illustrating the effectiveness and accuracy of the pose estimation process.

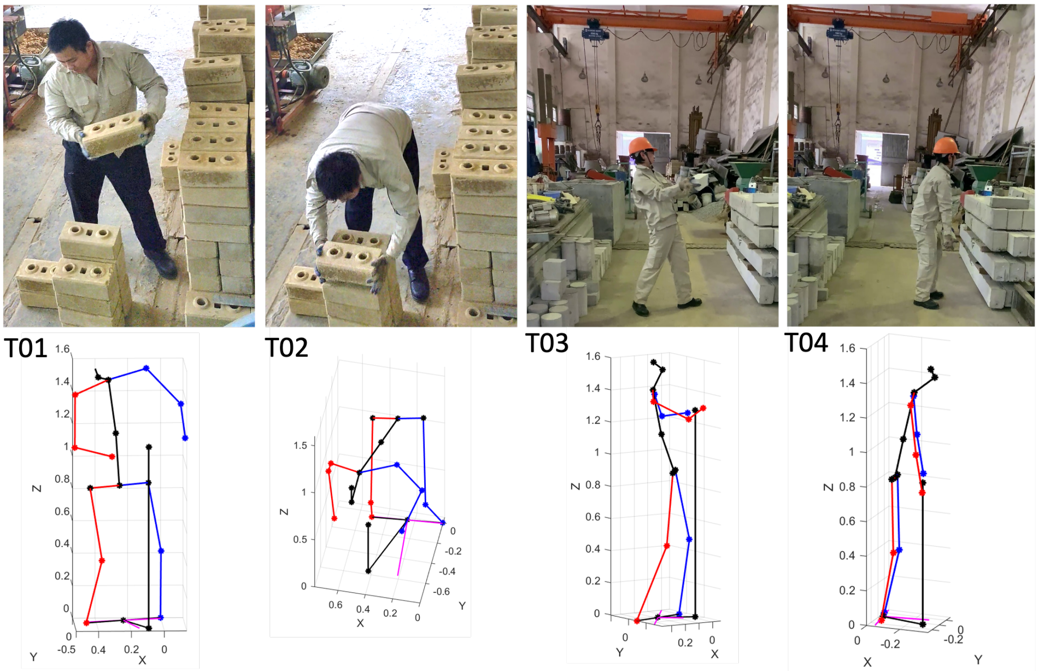


Figure 5. 3D pose generation from four posture groups

Table 1. Validation of the 3D pose dataset

Geometric measures										
Case	Measured	L (kg)	H (cm)	H err. %	V (cm)	V err %	D (cm)	D err. %	A (rad)	A err %
T01	NIOSH Algorithm	10	40 36.73	8.2%	110 111.5	1.3%	120 111.5	7.1%	5 9.22	84.4%
T02	NIOSH Algorithm	10	56 59.77	6.7%	50 49.32	1.4%	120 70.68	41.1%	15 15.67	4.5%
T03	NIOSH Algorithm	15	45 36.82	18.2%	125 128.3	2.7%	100 103.3	3.3%	15 12.84	14.4%
T04	NIOSH Algorithm	15	32 28.4	11.3%	78 81.86	4.9%	100 102.4	2.4%	25 19.88	20.5%
Lifting Equation components										
Case	Measured	L (kg)	HM	HM err %	VM	VM err %	DM	DM err %	AM	AM err %
T01	NIOSH Algorithm	10	0.63 0.68	7.9%	0.9 0.89	1.1%	0.83 0.86	3.6%	1 0.93	7.0%
T02	NIOSH Algorithm	10	0.45 0.42	6.7%	0.93 0.92	1.1%	0.85 0.88	3.5%	0.95 0.95	0.0%
T03	NIOSH Algorithm	15	0.56 0.68	21.4%	0.85 0.84	1.2%	0.82 0.86	4.9%	0.93 0.96	3.2%
T04	NIOSH Algorithm	15	0.78 0.88	12.8%	0.99 0.98	1.0%	0.88 0.81	8.0%	0.89 0.94	5.6%

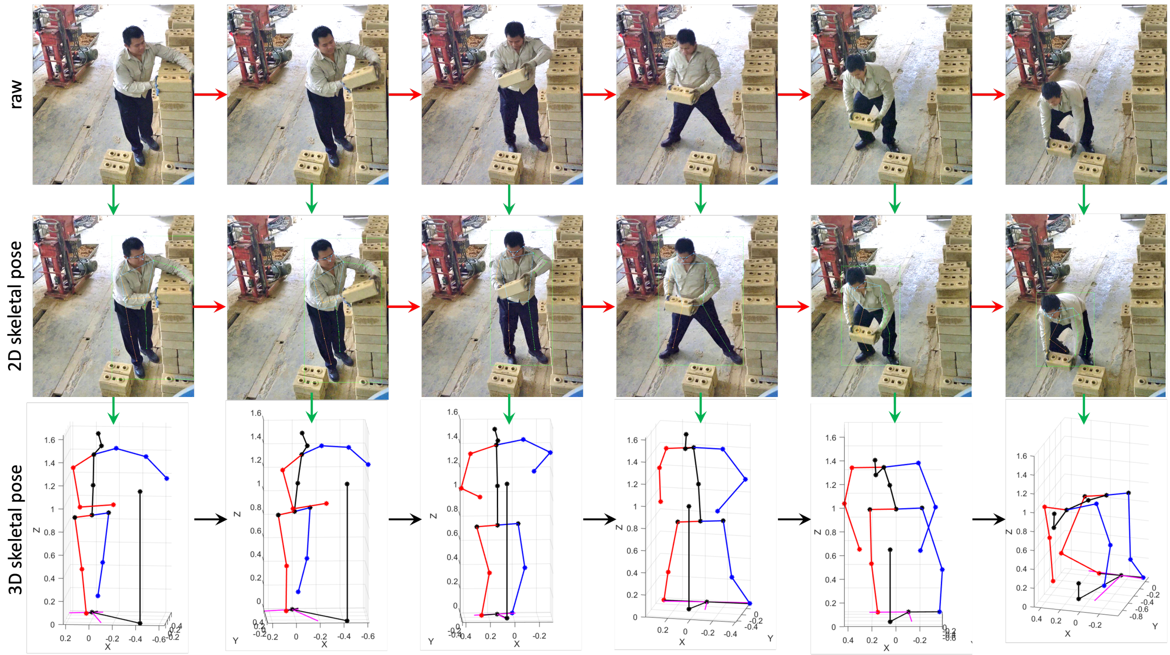


Figure 6. Transition process from raw footage to 2D and 3D skeletal poses

4.2. *The automated lifting index deep learning algorithm's (LSTM) performance*

a. Model Structure and Function

Fig. 7 visually illustrates the architecture of our LSTM model. The diagram traces the workflow from the initial input, which starts with a key-point detection dataset, progressing through a sequence input layer and an LSTM layer with multiple cells. This is followed by a fully connected layer leading to an output layer designed for regression. The output factors in multiple elements—HM, VM, DM, AM from the pose analysis, L and CM from the object detection model, and FM from the RepNet model—to compute the Lifting Index (LI). Finally, the output is classified into three risk categories: No risk, Medium risk, and High risk.

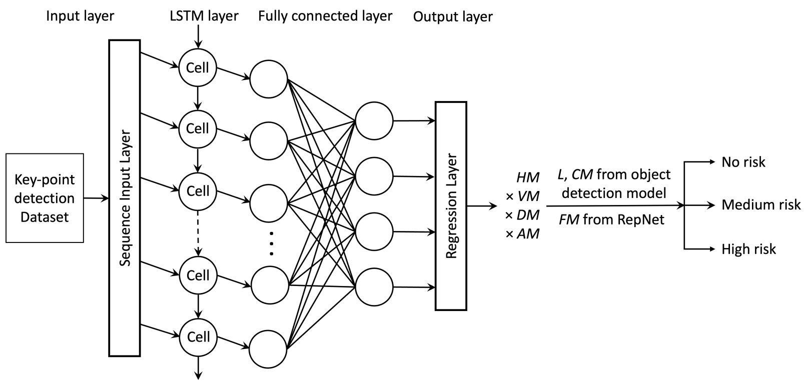


Figure 7. The Structure of the Constructed LSTM

b. Model's error and error analysis

Fig. 8 depicts RMSE and accuracy of the model in tandem. The charts show how quickly the model reached stable performance.

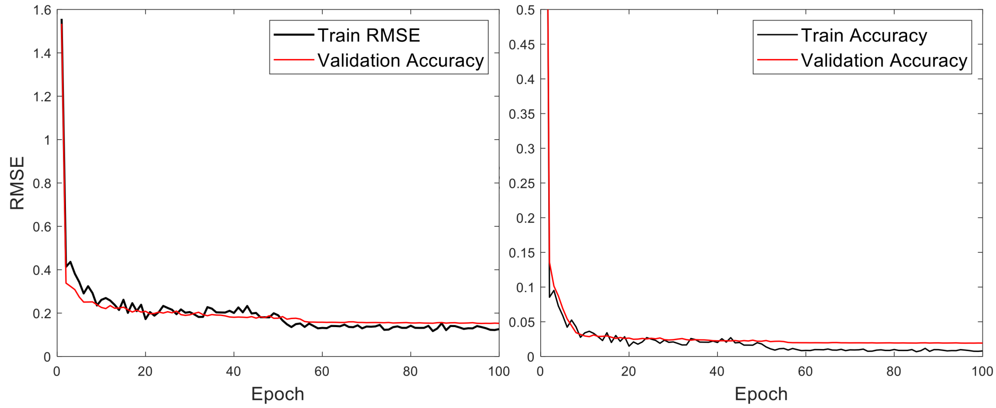
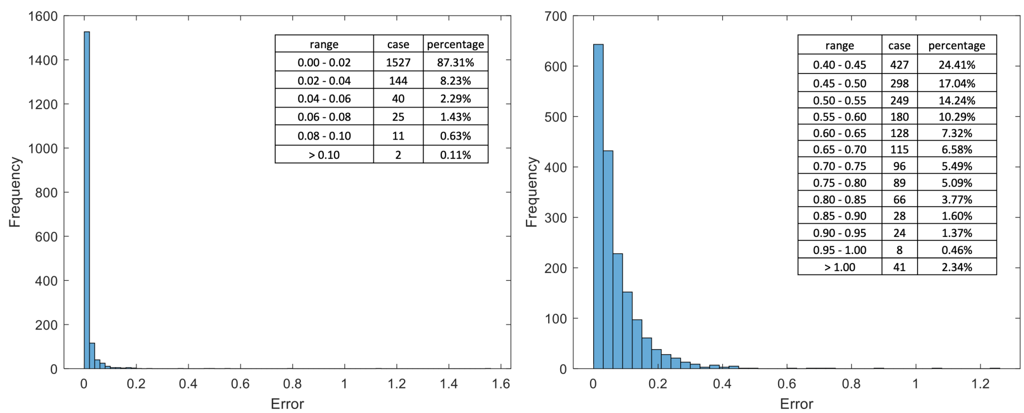


Figure 8. Training performance metrics

The RMSE and validation accuracy curves show initial rapid improvement, with RMSE dropping sharply to 0.4 and accuracy increasing significantly by the 5th epoch, then stabilizing by the 50th epoch. This indicates effective learning and a good fit, with the model extracting maximum information from the training data and generalizing well to validation data, thus avoiding overfitting. Both accuracy measures start at 0.5 and quickly improve, stabilizing by the 20th epoch. After the 50th epoch, the training accuracy remains consistently lower than validation accuracy, suggesting good model generalization possibly due to effective regularization or differences in data distribution between training and validation sets.

Fig. 9 presents histograms for RMSE and MAE, providing insight into the distribution of error values across all predictions made by the LSTM model.



a. RMSE histogram

b. MAE histogram

Figure 9. Error distribution

The majority of cases (87.31%) have an RMSE between 0.00 and 0.02, showing extremely high accuracy. Error rates gradually increase with the range, with very few cases (0.74%) having an RMSE greater than 0.06. A large proportion of the data has an MAE less than 0.75 (85.86%), with the largest percentage of cases (24.41%) between 0.40 and 0.45. The error rates are higher for MAE, with more variability in error magnitude compared to RMSE. These histograms highlight the model's precision and reliability, with most predictions exhibiting very low errors, indicative of strong performance.

Fig. 10 plots the actual versus predicted Lifting Index (LI). The data predominantly cluster around the diagonal line, indicating high accuracy of the model with minimal error deviation.

Most data points are concentrated between LI values of 0.5 and 1.75, with denser population extending up to 3.0. There are only a few outliers and four data points exceeding an LI of 3.0, denoting high-risk scenarios as defined by the NIOSH (LI < 1.0 - No risk; 1.0 < LI < 3.0 - Medium risk; LI > 3.0 - High risk).

Fig. 11 is a histogram that quantifies the distribution of predicted risk levels across different datasets. The histogram displays the distribution of predicted risk levels across training, validation, and test datasets reveals key insights into the LSTM model's performance. The training set, with 10,824 medium-risk cases, 2,753 no-risk, and 415 high-risk cases, illustrates a focus on medium-risk scenarios, which are also predominant in the validation (1,381 medium-risk) and test sets (1,362 medium-risk).

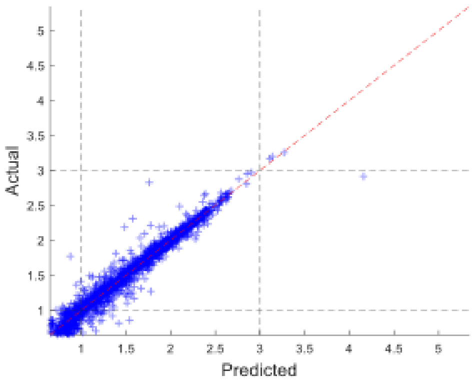


Figure 10. RMSE error analysis of prediction

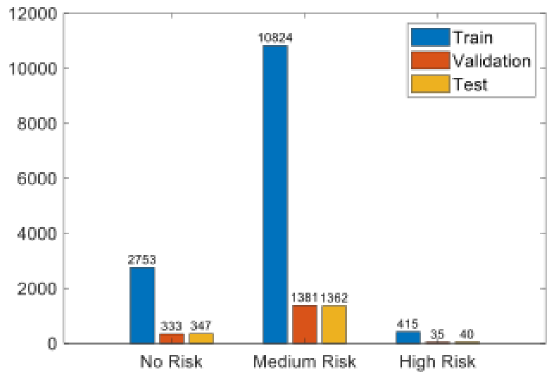


Figure 11. Histogram of the results by the LSTM Model

This distribution ensures the model is well-versed in handling the most common and varied lifting tasks, which are typically of medium risk. Fewer no-risk and high-risk cases across all sets ensure the model can accurately identify both less frequent high-risk situations, crucial for immediate safety interventions, and low-risk scenarios, important for overall risk management. The consistency in risk classification across different datasets demonstrates the model's robust capability to generalize well from training to real-world applications, making it a reliable tool for assessing lifting risks in construction and similar industries.

c. Model's Accuracy and Confusion Matrix

Fig. 12 showcases a confusion matrix of the model's predictions, with a part showing raw numbers and the other showing percentages. Part B, which is more detailed, reveals the true positive rates and misclassifications: for the 'No Risk' category, 17.15% were correctly predicted with a low false positive rate (2.69%); the 'Medium Risk' category showed a robust true positive rate of 75.19% with minor misclassification rates (2.40% classified as 'No Risk' and 0.29% as 'High Risk'); the 'High Risk' category had an impressively accurate prediction rate of 1.82% true positive, with a negligible misclassification rate from 'Medium Risk' (0.46%).

The LSTM model demonstrates strong predictive performance with a high degree of accuracy in classifying different risk levels of lifting tasks. The detailed analysis through error metrics, risk distribution, and a confusion matrix further validates the model's efficacy and reliability in real-time application environments aimed at enhancing workplace safety.

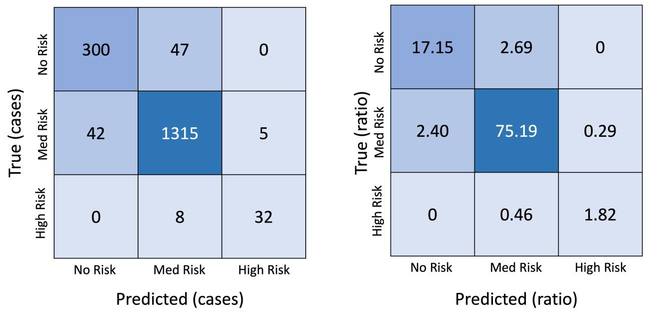


Figure 12. Confusion matrix of the results

4.3. The object detection and cycle counting results

This study selected five types of materials commonly carried on construction sites: bricks, cylinders, rectangles, and cubes. Specifically, brick samples were tested when carried singly or in pairs. The study used the Yolov8 model to detect these objects from a dataset compiled to record workers' carrying positions. Training was conducted on an Alienware 17R4 laptop with an NVIDIA GeForce GTX 1070, achieving a precision of 0.894 and recall of 0.754. Fig. 13 depicts some examples of successful detection of carried objects.

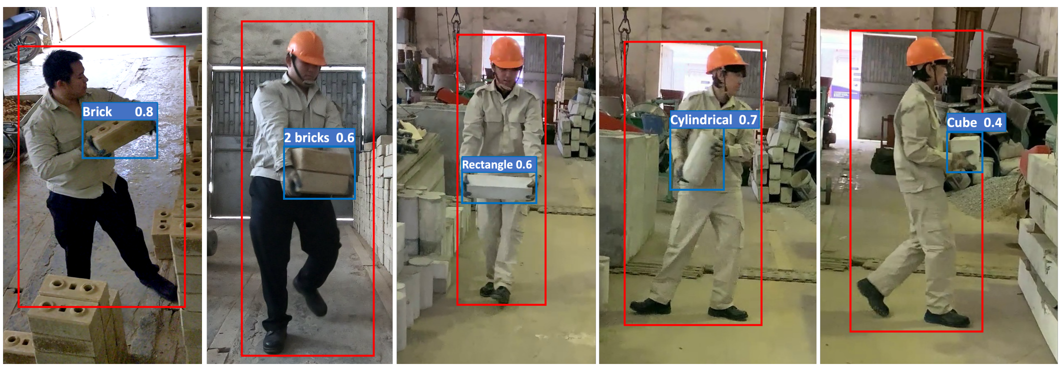


Figure 13. Carried object detected using Yolov8

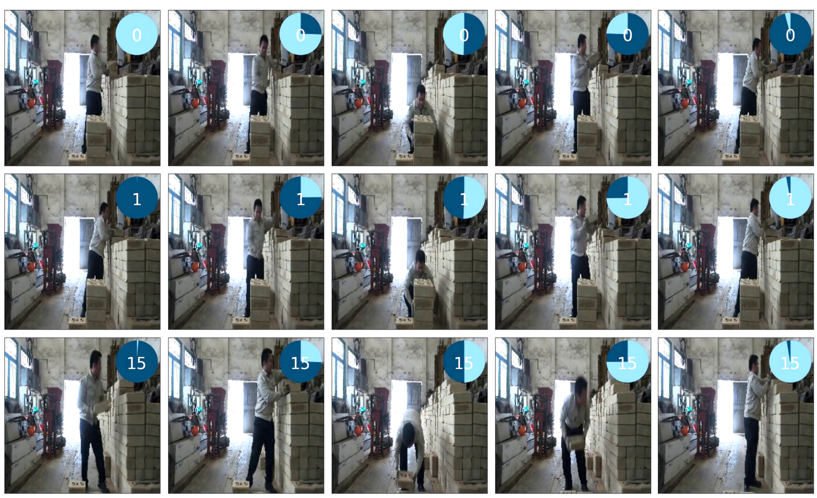


Figure 14. The cycle counting RepNet results

The selected model – RepNet – is a model designed to count repetitions and determine the cycle of repetitive actions in a video. It processes a video as a sequence of frames through three main components: a frame encoder, a temporal self-similarity matrix (TSM), and a cycle predictor. The encoder, structured like a ResNet model [34], extracts 2D features from each frame. The TSM, generated by comparing frames within the video, maps out the repetitive actions, allowing the model to analyze patterns over time. The cycle predictor then uses this matrix to estimate the length and periodicity of the action cycle, utilizing two fully connected layers of size 512 for each predictor. This architecture enables RepNet to effectively identify and quantify repetitive motions in video data. Fig. 14 illustrates an example that was extracted from the footage, in which the model successfully counted 15 cycles.

4.4. Integration of components

The individual components (i.e. pose estimation, LSTM model, object detection, cycle counting) are each complicated but advanced in doing their tasks. Though working on a modest computational device, the integrated system demonstrated high effectiveness and efficiency in real-time environments. The use of advanced algorithms allowed for accurate pose detection and object recognition, while the LSTM model effectively utilized the sequential data to make real-time predictions. The system's ability to process and analyze complex data on the fly significantly enhances its practical application in workplace safety, enabling proactive measures to mitigate injury risks. The integration not only ensures accuracy but also the adaptability of the system in various operational conditions, making it a robust tool for ergonomic assessment in dynamic settings.

4.5. Discussions

a. Practical implications

The study's results demonstrate the efficacy of an integrated approach in evaluating the ergonomic risks associated with object-lifting tasks on construction sites. The system combines pose estimation, object detection, cycle counting, and LSTM modeling to assess the NIOSH Lifting Index in real-time. This method offers a dynamic analysis of workers' interactions with various loads, providing insights that are crucial for improving workplace safety and ergonomics.

The integrated system successfully predicts the lifting index with high accuracy and efficiency, even on modest computational setups. Key findings reveal that real-time analysis and proactive risk assessments are feasible with the proposed model, which can significantly mitigate potential injuries by allowing immediate corrective actions. The application of such an advanced analytical tool in industry promises to enhance safety practices by providing continuous, real-time feedback. This system can be integrated into existing safety protocols, offering a more granular understanding of risk factors associated with specific lifting tasks and worker behaviors. Compared to traditional methods, which often rely on manual calculations of the NIOSH Lifting Index or use wearable sensors, our approach provides a more comprehensive and automated solution. While traditional methods provide static assessments, our system offers dynamic, real-time analysis, bridging the gap between observational studies and immediate application needs. Additionally, the primary advantage of the proposed approach lies in its ability to integrate multiple data sources into a cohesive analysis tool without the need for invasive sensors or extensive manual inputs. This not only ensures higher accuracy but also enhances user compliance and ease of use in field conditions.

The findings hold significant practical implications for various stakeholders in the construction industry, including site managers, safety officers, and workers themselves. By implementing this integrated system, stakeholders can achieve a higher level of situational awareness regarding ergonomic risks. This proactive approach not only enhances worker safety but also contributes to reducing the incidence of musculoskeletal disorders, which are prevalent in construction due to the physical nature of the work.To effectively implement this methodology in real-world settings, several steps are recommended:

1. Training and awareness: Train site managers and safety officers on how to interpret the system's outputs and integrate these insights into daily operational decisions.

2. Technology integration: Equip sites with the necessary hardware, such as cameras and computing devices, capable of supporting the system.

3. Continuous monitoring: Set up a continuous monitoring system that can provide real-time feedback to workers and site supervisors. This could involve setting up display screens at suitablelocations to alert workers when they are performing high-risk maneuvers.

4. Feedback mechanisms: Establish feedback mechanisms that allow workers to report about the system's effectiveness and usability, ensuring that the tool remains worker-centered and practical for daily use.

Moreover, the proposed system can be seamlessly integrated into existing safety protocols by:

1. Pre-shift assessments: Incorporating system checks into pre-shift safety briefings to assess risk levels and plan the day's work accordingly.

2. Real-time alerts: Utilizing the system's real-time analysis capabilities to provide immediate alerts to workers and supervisors if risky behaviors are detected. This could be linked directly to personal protective equipment (PPE) such as smart helmets or vests.

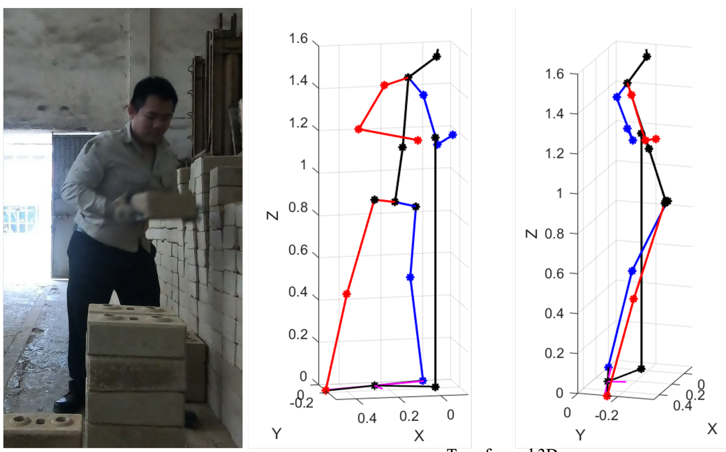
3. Training programs: Enhancing existing training programs with data from the system to educate workers about ergonomic risks and safe lifting practices. Use historical data collected by the system to highlight common risk patterns and teach corrective techniques.

4. Policy updates: Updating internal safety policies to include guidelines for the use of automated ergonomic assessment tools, specifying how data should be collected, analyzed, and acted upon.

By adopting these recommendations, construction sites can enhance their safety culture, making a tangible impact on reducing injuries and promoting a healthier workplace.

b. Limitations s and Future Research

One limitation encountered is the potential for error in pose estimation due to varying camera angles. Fig. 15 shows a case in which the 3D pose was built incorrectly due to the camera angles.



Raw image

Transformed 3D pose
(Errors due to camera angle defects in the 3D pose)

Figure 15. Incorrect 3D pose due to camera angle

Additionally, the effectiveness of object detection is contingent upon the reduction of carried objects to known weights. Future research should explore the enhancement of pose estimation accuracy and the expansion of object detection capabilities to include a broader range of object types and weights. Investigating the integration of environmental and contextual data could also enhance the model's applicability across diverse settings. It is also useful to conduct a long-term study to determine the level of improvement in worker health when applying this model to real production. Finally, the system's reliance on specific data conditions, such as camera angles and known object weights, presents limitations that need addressing to broaden its practical utility. Expanding the dataset to cover a broader array of objects and lifting scenarios will also enhance the system's generalizability and effectiveness across diverse construction environments.

5. Conclusions

This study introduces an integrated system that assesses the NIOSH Lifting Index in real-time, leveraging advanced computational models and real-time data analytics. The proposed system combines pose estimation, object detection, cycle counting, and LSTM modeling to dynamically assess lifting tasks on construction sites. This innovative approach offers a substantial improvement over traditional ergonomic assessment methods, which often rely on manual calculations or are constrained by the capabilities of wearable sensors. By employing cutting-edge algorithms and machine learning techniques, our methodology enables accurate, real-time evaluations of lifting practices, facilitating immediate and effective interventions to mitigate risk.

Several strategies are also recommended to incorporate this system into existing safety protocols. These include the use of real-time monitoring to provide immediate feedback on risky behaviors, integration of system outputs into daily safety briefings, and the enhancement of worker training programs with personalized, data-driven insights. Additionally, updating internal safety policies to include the use of such advanced technologies will ensure that safety practices remain robust and responsive to the evolving demands of construction work.

In conclusion, this study significantly advances construction safety by providing a sophisticated tool that not only evaluates ergonomic risks in real-time but also seamlessly integrates into construction site operations. This system not only supports the NIOSH guidelines but also pioneers a shift towards proactive safety management, ultimately fostering a safer and more efficient workplace. As this technology evolves, it will undoubtedly play a pivotal role in transforming occupational health standards, propelling the construction industry towards a future where preventive measures effectively mitigate the risks of lifting-related injuries.

References

[1] Tran, T. T., Taptagaporn, S., Dang, C. V., Kaewdok, T. (2023). Ergonomic risk factor-related musculoskeletal disorders among wood furniture workers in Binh Duong Province, Vietnam. *Asia-Pacific Journal of Science and Technology*, 28.

[2] Waters, T. R., Lu, M.-L., Piacitelli, L. A., Werren, D., Deddens, J. A. (2011). Efficacy of the Revised NIOSH Lifting Equation to Predict Risk of Low Back Pain Due to Manual Lifting: Expanded Cross-Sectional Analysis. *Journal of Occupational & Environmental Medicine*, 53(9):1061–1067.

[3] Lu, M.-L., Waters, T. R., Krieg, E., Werren, D. (2013). Efficacy of the Revised NIOSH Lifting Equation to Predict Risk of Low-Back Pain Associated With Manual Lifting: A One-Year Prospective Study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(1):73–85.

[4] Halim, I., Saptari, A., Abulais, M. F., Padmanathan, V. (2021). Integrated Measurement System of Postural Angle and Electromyography Signal for Manual Materials Handling Assessment. *The International Journal of Integrated Engineering*, 13(6).

[5] Lun, R., Zhao, W. (2015). A Survey of Applications and Human Motion Recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence*, 29(05):1555008.
[6] Spector, J. T., Lieblich, M., Bao, S., McQuade, K., Hughes, M. (2014). Automation of Workplace Lifting Hazard Assessment for Musculoskeletal Injury Prevention. *Annals of Occupational and Environmental Medicine*, 26(1).
[7] Wang, D., Dai, F., Ning, X. (2015). Risk Assessment of Work-Related Musculoskeletal Disorders in Construction: State-of-the-Art Review. *Journal of Construction Engineering and Management*, 141(6).
[8] Donisi, L., Cesarelli, G., Coccia, A., Panigazzi, M., Capodaglio, E. M., D'Addio, G. (2021). Work-Related Risk Assessment According to the Revised NIOSH Lifting Equation: A Preliminary Study Using a Wearable Inertial Sensor and Machine Learning. *Sensors*, 21(8):2593.
[9] Greene, R. L., Chen, G., Lu, M.-L., Hen Hu, Y., Radwin, R. G. (2021). Enhancing the Revised NIOSH Lifting Equation using ComputerVision. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1):467–471.
[10] Donisi, L., Cesarelli, G., Capodaglio, E., Panigazzi, M., Cesarelli, M., D'Addio, G. (2022). Machine Learning and Biosignals are able to discriminate biomechanical risk classes according to the Revised NIOSH Lifting Equation. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, IEEE.
[11] Zhou, G., Aggarwal, V., Yin, M., Yu, D. (2021). Video-based AI Decision Support System for Lifting Risk Assessment. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 275–282.
[12] Boschman, J. S., van der Molen, H. F., Sluiter, J. K., Frings-Dresen, M. H. W. (2012). Musculoskeletal disorders among construction workers: a one-year follow-up study. *BMC Musculoskeletal Disorders*, 13 (1).
[13] Wang, M.-J. J., Garg, A., Chang, Y.-C., Shih, Y.-C., Yeh, W.-Y., Lee, C.-L. (1998). The Relationship between Low Back Discomfort Ratings and the NIOSH Lifting Index. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):509–515.
[14] Bosch, T., de Looze, M. P., van Dieën, J. H. (2007). Development of fatigue and discomfort in the upper trapezius muscle during light manual work. *Ergonomics*, 50(2):161–177.
[15] Coenen, P., Gouttebarge, V., van der Burght, A. S. A. M., van Dieën, J. H., Frings-Dresen, M. H. W., van der Beek, A. J., Burdorf, A. (2014). The effect of lifting during work on low back pain: a health impact assessment based on a meta-analysis. *Occupational and Environmental Medicine*, 71(12):871–877.
[16] Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for MSD risk assessment. *Applied Ergonomics*, 45(3):420–427.
[17] Fox, R. R., Lu, M.-L., Occhipinti, E., Jaeger, M. (2019). Understanding outcome metrics of the revised NIOSH lifting equation. *Applied Ergonomics*, 81:102897.
[18] Waters, T. R., Putz-Anderson, V., Garg, A. (1994). *Applications manual for the revised NIOSH lifting equation*.
[19] Takala, E.-P., Pehkonen, I., Forsman, M., Hansson, G.-Å., Mathiassen, S. E., Neumann, W. P., Sjøgaard, G., Veiersted, K. B., Westgaard, R. H., Winkel, J. (2009). Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scandinavian Journal of Work, Environment & Health*, 36 (1):3–24.
[20] Genaidy, A. M., Karwowski, W., Christensen, D. M., Vogiatzis, C., Deraiseh, N., Prins, A. (1998). What is "heavy"? *Ergonomics*, 41(4):420–432.
[21] Elfeituri, F. E., Taboun, S. M. (2002). An Evaluation of the NIOSH Lifting Equation: A Psychophysical and Biomechanical Investigation. *International Journal of Occupational Safety and Ergonomics*, 8(2): 243–258.
[22] Liu, S., Zou, H., Huang, Y., Cao, X., He, S., Li, M., Zhang, Y. (2023). ERF-RTMDet: An Improved Small Object Detection Method in Remote Sensing Images. *Remote Sensing*, 15(23):5575.
[23] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Lecture Notes in Computer Science*, Springer Interna-

tional Publishing, 740–755.

[24] Yang, Z., Zeng, A., Yuan, C., Li, Y. (2023). Effective Whole-body Pose Estimation with Two-stages Distillation. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, IEEE.

[25] Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M. (2019). VideoPose3D. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[26] Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C. (2014). Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339.

[27] Joo, H., Simon, T., Li, X., Liu, H., Tan, L., Gui, L., Banerjee, S., Godisart, T., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., Sheikh, Y. (2019). Panoptic Studio: A Massively Multiview System for Social Interaction Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):190–204.

[28] Belagiannis, V., Amin, S., Andriluka, M., Schiele, B., Navab, N., Ilic, S. (2016). 3D Pictorial Structures Revisited: Multiple Human Pose Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):1929–1942.

[29] Lin, J., Zeng, A., Wang, H., Zhang, L., Li, Y. (2023). One-Stage 3D Whole-Body Mesh Recovery with Component Aware Transformer. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE.

[30] Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306.

[31] Glenn, J., Ayush, C., Jing, Q. (2023). *YOLOv8 by Ultralytics*.

[32] Zhang, X., Zhou, Q. (2023). RepNet: A Lightweight Human Pose Regression Network Based on Re-Parameterization. *Applied Sciences*, 13(16):9475.

[33] Dwibedi, D. (2020). *Google AI Blog: RepNet: Counting Repetitions in Videos*. Google AI Blog.

[34] He, F., Liu, T., Tao, D. (2020). Why ResNet Works? Residuals Generalize. *IEEE Transactions on Neural Networks and Learning Systems*, 31(12):5349–5362.