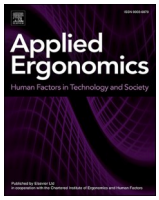
Applied Ergonomics 127 (2025) 104513

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo



Computer vision and tactile glove: A multimodal model in lifting task risk assessment

Haozhi Chen [a], Peiran Liu [a], Guoyang Zhou [a], Ming-Lun Lu [b], Denny Yu [a,*]

[a] Purdue University, West Lafayette, IN, USA
[b] National Institute for Occupational Safety and Health, Cincinnati, OH, USA

ARTICLE INFO

Keywords:
Computer vision
Tactile gloves
Lifting risk prediction

ABSTRACT

Work-related injuries from overexertion, particularly lifting, are a major concern in occupational safety. Traditional assessment tools, such as the Revised NIOSH Lifting Equation (RNLE), require significant training and practice for deployment. This study presents an approach that integrates tactile gloves with computer vision (CV) to enhance the assessment of lifting-related injury risks, addressing the limitations of existing single-modality methods. Thirty-one participants performed 2747 lifting tasks across three lifting risk categories (LI $<$ 1, 1 $\leq$ LI $\leq$ 2, LI $>$ 2). Features including hand pressure measured by tactile gloves during each lift and 3D body poses estimated using CV algorithms from video recordings were combined and used to develop prediction models. The Convolutional Neural Network (CNN) model achieved an overall accuracy of 89 % in predicting the three lifting risk categories. The results highlight the potential for a real-time, non-intrusive risk assessment tool to assist ergonomic practitioners in mitigating musculoskeletal injury risks in workplace environments.

1. Introduction

Work-related injuries continue to affect roughly 3 percent of the private industry workers in the United States (US) and cost hundreds of billions of dollars per year, including substantial losses in wages, productivity, medical expenses, and administrative costs (National Safety Council), (US Bureau of Labor Statistics, 2023). Among occupational injuries, overexertion, which encompasses lifting, remains the top contributing factor based on the Liberty Mutual Workplace Safety Index (LMWSI) (Marucci-Wellman et al., 2015). Repetitive or heavy lifting in poor postures, sometimes influenced by poorly designed tasks or workstations, was found to be a significant contributor to occupational injuries, particularly those in the low back region (Craig et al., 2003), (Van Dieen et al., 1999). Poor lifting postures often refer to postures that increase strain on the spine and reduce stability, such as excessive torso bending, excessive twisting on the torso, and holding the load too far (Burgess-limerick and Abernethy, 1997), (Marras et al., 1993), (Waters et al., 1993). Many tools have been developed for assessing the physical demands and risk factors associated with lifting tasks, such as the Revised NIOSH Lifting Equation (RNLE), the Liberty Mutual Manual Material (MMH) Handling Table, the Health Safety Executive (HSE) Manual Handling Assessment Charts (MAC), the ACGIH Threshold Limit Value (TLV) for Lifting, the Washington State Lifting Calculator, and the recently developed Lifting Fatigue Failure Tool (LiFFT) (Waters et al., 1993), (Potvin et al., 2021), (Gallagher et al., 2017), (American Conference of Governmental Industrial Hygienists (ACGIH), 2005), (Health and Safety Executive, 2019). A previous study investigating the risk assessment tools used by certified ergonomists in four English-speaking countries showed that RNLE was the most popular among all tools used in the workplace (Lowe et al., 2019). The RNLE was widely validated and used as the base for the ISO ergonomic standard 11228-Part 1 for Manual Lifting, Lowering, and Carrying (Fox et al., 2019), (Lu et al., 2016). The ISO 11228-1 has been adopted by many countries as their national ergonomic standards, such as Italy, Spain, and Netherlands (Lu et al., 2016). Other aforementioned risk assessments have not been validated extensively using workplace exposure or health data. As seen in the evidence in the literature, the RNLE brought a significant impact on ergonomics assessments by providing practical guidance on workplace design for reducing the risk of overexertion injuries associated with lifting tasks (Fox et al., 2019), (Waters et al., 2011), (Ranavolo et al., 2017). However, the involvement of trained analysts in using the RNLE presents some barriers to efficient risk assessment and mitigation. For example, taking measurements for RNLE requires significant training and practice (Waters et al., 1998) (Dempsey and Fathallah,

* Corresponding author. Purdue University, 315 N Grant St, West Lafayette, IN, 47907-2050, USA.
E-mail address: dennyyu@purdue.edu (D. Yu).

https://doi.org/10.1016/j.apergo.2025.104513
Received 9 October 2024; Received in revised form 20 March 2025; Accepted 21 March 2025
Available online 1 April 2025

1999). Modern jobs involving lifting often require workers to lift in many lifting conditions, which makes the RNLE analysis time-consuming and labor-intensive (Lu et al., 2015), (Callaghan et al., 2001). We opted to assess the automation of using RNLE with new technologies, such as wearable sensors, computer vision, and tactile pressure sensor gloves.

Wearable sensors have been proposed to overcome the challenges of collecting the data needed for the RNLE (Santopaolo et al., 2022), (Donisi et al., 2021), (Varrecchia et al., 2022). Donisi et al. (Donisi et al., 2021) proposed inertial measurement units (IMUs) on the lumbar region to assess RNLE with acceleration and angular velocity signals. Other studies have used IMUs for lifting action detection and assessing the risk for back pain (Thomas et al., 2022), (Pesenti et al., 2023), (Lu et al., 2020). Another study expanded on this by analyzing data from IMUs attached to 12 different body locations (Snyder et al., 2021). However, wearable sensor-based methods have limitations due to calibration challenges and restricted sensor coverage. These techniques depend on placing sensors on specific body parts, such as the back or waist, to capture motion data. As a result, they may offer an incomplete representation of full-body motion or provide an inaccurate view of the entire body. In addition, surface electromyography (sEMG) sensors have been widely used to study spinal loads and trunk moments during dynamic lifting exertions (Dolan et al., 2001), (Granata and Marras, 1995). More recent studies have used sEMG on various upper body regions to evaluate risk factors during the execution of lifting tasks (Ranavolo et al., 2018), (Brandt et al., 2018), (Varrecchia et al., 2018), (Antwi-Afari et al., 2017). Most mentioned studies focused on indirect measures such as trunk kinematics or muscle activities to relate them to RNLE outputs based on known task conditions; these approaches did not estimate key parameters such as hand location or other RNLE variables. Yet, the placement of body-worn sensors can be obtrusive and uncomfortable for workers, raising questions about their feasibility and acceptance in real-world applications.

The task of human body pose estimation has been a challenge in the field of computer vision, and this research area has been receiving a high level of interest in ergonomics with the advancement of computer vision (CV) algorithms in recent years (Toshev and Szegedy, 2013), (Iskakov et al., 2019), (Cao et al., 2018), (Zhu et al., 2022). Previous work has demonstrated the potential of computer vision in ergonomics or occupational safety field by extracting and analyzing different features including body motion, joint angles, L5/S1 moments, posture, trunk kinematics, and facial expressions (MassirisFernández et al., 2020), (Greene et al., 2022), (Wang et al., 2021a), (Zhou et al., 2022), (Wang et al., 2019), (Greene et al., 2021), (Wang et al., 2021b), (Greene et al., 2019), (Mehrizi et al., 2018), (Murugan et al., 2024). These methods have demonstrated promising results in predicting lifting-related injury risk factors, the lifting index (LI), and the rating of perceived exertion. Innovative approaches have been taken to improve the performance of these CV-based assessment methods. Specifically, Wang et al. introduced a 2D video-based lifting monitor that automatically extracts spatial and temporal factors for lifting risk assessment using ghosting effect for accurate detection of lifting instances (Wang et al., 2019). Zhou et al. integrated facial expression features with body motion and posture features and achieved an area under the curve (AUC) of 0.89 in classifying the LI (Zhou et al., 2022). However, two major gaps exist in CV for lifting risk assessments. First, the weight of the lifting object is one of the most critical contributors to injury risks, but weight can be difficult to assess using computer vision. Second, most studies on CV occur in the laboratory setting. The practicality of CV methods is restricted in occupational settings, where camera angles, occlusions, and multiple workers are common issues (Egeonu and Jia, 2024), (Debnath et al., 2022) and will affect the performance of CV-based assessment methods.

Existing techniques that assess lifting risk in ergonomics based on either sensor data or computer vision (CV) have several limitations. Sensor-based methods, despite their precision, can be intrusive due to the need for multiple sensors affixed to the worker's body. While pose estimation with CV techniques is useful for identifying hand locations, CV cannot directly measure the weight of the lifted object required for RNLE. A previous study that proposed CV methods for estimating hand loads using body kinematic information cannot be used for estimating the weight of the load for RNLE (Zhou et al., 2022). Hand loads are often referred to as dynamic loads acting on the hands during lifting for biomechanical analyses. We did not use the hand load information for a biomechanical analysis, rather, we explored a multi-modality method that measures both the pose and hand pressure factors for estimating both the lifting weight and body pose risk factors for calculating the lifting index. We developed an integrated approach using a tactile glove sensor that is non-intrusive (as gloves are already commonly used by workers (Zhou et al., 2023a)) and CV techniques that only require raw camera footage to estimate the NIOSH lifting equation-related parameters with existing 3D joint coordinates. Our study evaluates this computer vision and glove sensor informatics approach for determining lifting injury risk across various task and weight conditions with the long-term goal of providing an all-in-one sensor and computer vision tool for estimating lifting injury risks. We hypothesize that the accuracy of predicting the risk of injury (as measured by the RNLE Lifting Index) will be improved by integrating tactile gloves with computer vision than single-modality approaches of just CV or tactile gloves alone.

2. Method

2.1. Participants

This experiment was approved by the Purdue University Institutional Review Board (IRB). Thirty-one participants were recruited from the university population to perform the study. All participants had no professional experience in workplace lifting tasks, and no participant reported any musculoskeletal pain or discomfort at the time of the experiment. The sample contained 19 males and 12 females, and 80 % of the participants were right-handed.

2.2. Experimental design and procedure

Tactile gloves (PPS, Inc., Hawthorne, CA) were utilized in this study (Fig. 1). Sixty-five pressure sensors were embedded on the palmar side of each glove. Per the glove manufacturer, each sensor measured the pressure that was exerted on the palmar side of the hand with a scanning rate of 25–40 Hz, and pressure data was streamed continuously to a computer via Bluetooth connection. The gloves were calibrated using the manufacturer's software before each data recording session while participants were in a relaxed, non-working state. All lifting tasks were recorded by a GoPro camera (30hz) positioned on the left side of the participant when the participant faced the platform. The camera position was fixed across all task conditions.

Six different lifting tasks (Fig. 3) were performed in a controlled laboratory environment. Participants performed six tasks that involved moving a box with different assigned weights between four specific locations (Fig. 2). The platform's height level (location 2) was adjusted to three different levels (high [1.1m], middle [0.9m], low [0.7m]). Each task (1–6) was performed at every height level. The weights for each lifting task were determined using the Revised NIOSH Lifting Equation (RNLE). To manage participant risk levels, the Recommended Weight Limits (RWL) were estimated independently using the task conditions of the 6 tasks with variations in platform height (Location 2). For each platform height, the estimated RWLs were used to propose a range of lifting weights, roughly categorized the risk into the three LI ranges [LI < 1, $1 \leq$ LI ≤ 2, and LI > 2]. The exact distribution of weights was selected based on the availability of weight plate combinations and grouped for each weight/height combination to improve the robustness of the lifting weight prediction results, resulting in a total of nine combinations (e.g., a light weight at low/middle/high heights, a medium weight at low/middle/high heights, and a heavy weight at low/

Fig. 1. Tactile gloves from PPS Inc.

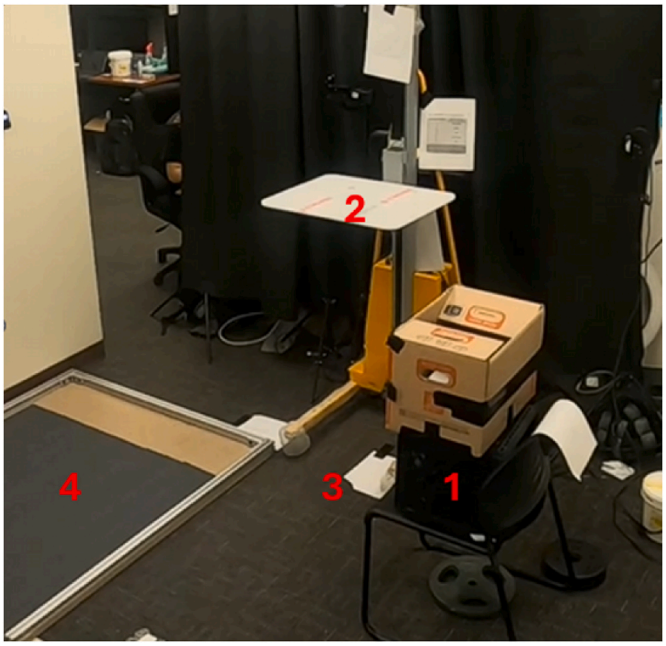


Fig. 2. Experimental setup showing the four specific locations (1, 2, 3, 4) used for the lifting tasks. Location 1 represents the desk as starting location, location 2 represents the height adjustable platform, location 3 represents the floor level under the platform, and location 4 represents the floor-level position.

middle/high heights). The weight was constant throughout the six tasks for a specific LI category. A Frequency-Independent Lifting Index approach (FM = 1) was used, and each lifting task was treated as an independent action, and a good coupling factor (CM = 1) was assumed for the boxes used in this study because they all had handles that ensured good hand-to-object contact. Specifically, the range of weights (Table 1) for light, medium, and heavy conditions corresponded to LI < 1, 1 ≤ LI ≤ 2, and LI > 2, respectively (Table 1).

Participants completed one action set (tasks 1–6) twice for each weight/height combination. After completing the action set twice for a specific weight, the platform's height was adjusted for the next set of tasks. After completing all action sets on the three height levels for a given weight, a new weight level (light, medium, or heavy) would be assigned for the participants to repeat the same process. The height order was controlled, and the order of task 1–6 within an action set was not randomized since the start/end location of each task depended on the previous task (the box ending location for Task 2 is the start location for Task 3). We prioritized randomizing weight assignment and order across participants in the study design to focus on the impact of how different weights at same task designs influenced model performance. A 10-s rest was given between each lifting action, a 2-min rest was given after the participants completed an action set (tasks 1–6), and a 10-min rest was given after the participants completed all required tasks at each weight level. Each participant completed 108 lifting/lowering actions in total. For safety purposes, no lifting task with a LI higher than three was given.

2.3. Data processing

2.3.1. 3D human pose estimation

The collected 2D videos were processed with MotionBert (Zhu et al., 2022), a current state-of-the-art 3D human pose estimation and body rebuild model based on transformer structure. The use of 3D estimates enables a more accurate and detailed representation of human movements, capturing depth and spatial relationships that are crucial for realistic and nuanced posture analysis. It was previously shown to outperform existing algorithms on the Human3.6M and 3DPW datasets (Zhu et al., 2022). With the video input, we used MotionBERT to create a skeleton representation of human motion comprising 17 distinct joint coordinates (see Appendix A1), each characterized by a three-dimensional vector with x, y, and z coordinates.

2.3.2. Pose related task variables

All recorded lifting videos were segmented into clips corresponding to each lift. The beginning of a clip was defined as the point when participants placed both hands on the load and initiated upward movement. The end of the clip was defined when the load reached the target height and remained steady, indicating that the lift had been completed and the load was stabilized. These clips were processed by MotionBert with pre-trained weights validated on the Human3.6M dataset, and the coordinate information for each joint at each frame was stored as output. To reduce the impact of frame-to-frame variations, the coordinates for each joint at the first and last five frames were averaged to estimate the starting and end posture for each lift. The Horizontal Location, Vertical Location, and Asymmetry Angle (defined in Table 2) were calculated for both the start and end of each lift, and the Vertical Travel Distance was also calculated.

Because MotionBert could only infer all joint coordinates in a scaled pixelated coordinate system (scaled with the resolution of the input video), we used the height of each participant to scale all features from the pixelated unit into the metric unit system. All task variables were then converted to metric units as following:

$$Task\ variable\ (cm) = Height\ (cm)^* \frac{Task\ variable\ (in\ pixel\ unit)}{Sum\ of\ joint\ segments\ connecting\ head\ to\ foot\ (in\ pixel\ unit)}$$

2.3.3. Glove feature

The pressure-sensing glove provides a direct and reliable measure of hand pressure and grip forces, which are key predictors of lifting weight during lifting tasks (Zhou et al., 2023b). The glove sensors were summed into the following hand regions: thumb, index, middle, ring, pinky, upper palm, and lower palm. These crafted features reduced the number of predictor variables and improved model interpretability by enabling analysis of the contribution of each hand region to the lifting risk assessment. Outputs exceeding the maximum sensor threshold (80psi) were removed before feature extraction. The data was denoised with a 1D Gaussian filter. A peak detection algorithm was used to detect lifting actions in the filtered glove data. For each peak, gradients within a 2-s window before and after the peak were calculated and sorted. The 2-s window was used to ensure the relevant motion would be captured since the average lifting action lasted 4 s. From the top five largest gradients before and after the peak, the one with the lowest corresponding force was selected to mark the start and end points of each lifting action (Zhou et al., 2023a). The glove data did not require

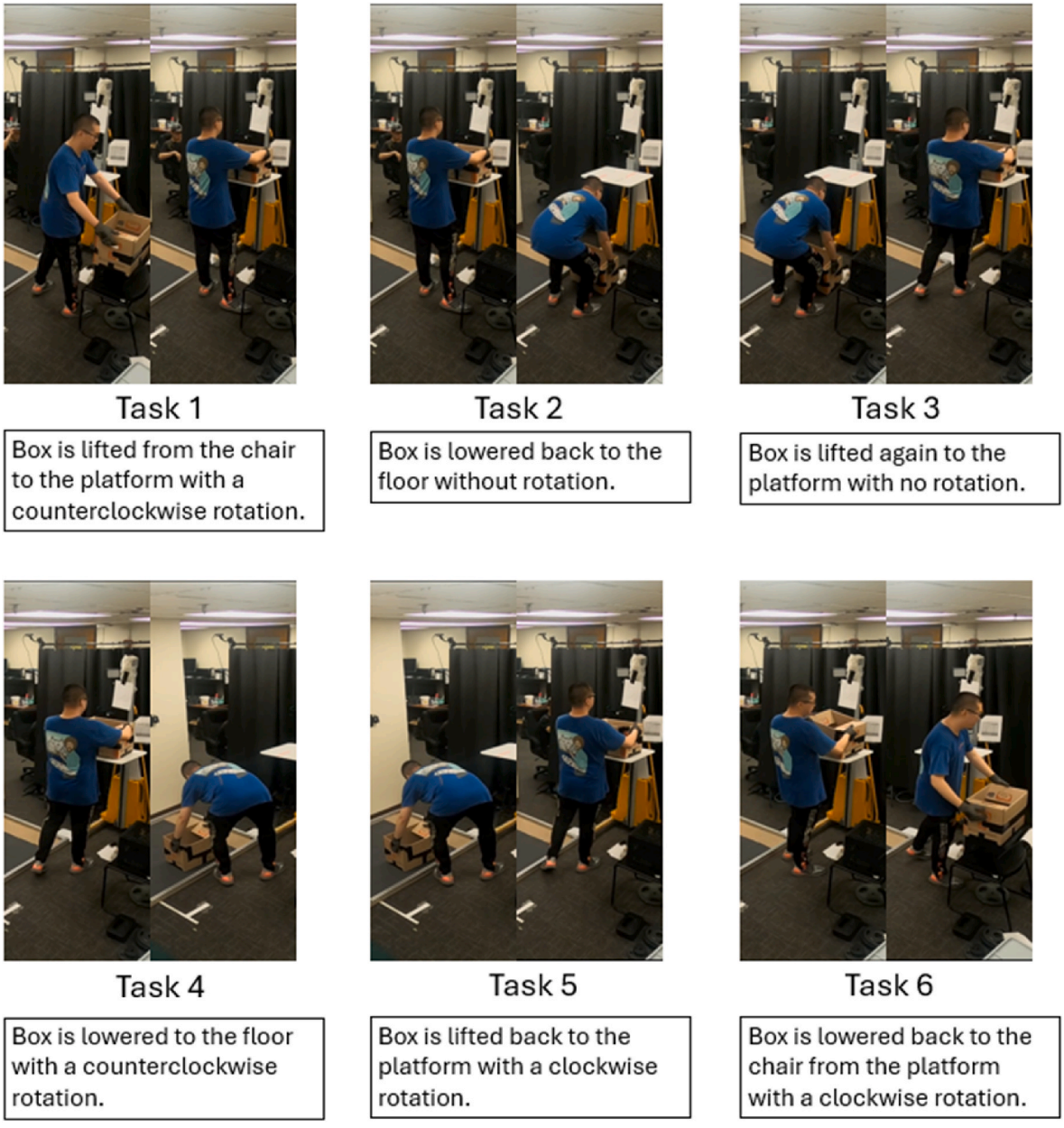


Fig. 3. Six lifting tasks that composed our action set were repeated twice for each weight/height combination with 108 lifting/lowering actions completed for each participant.

Table 1
Weight (kg) range for the studied lifting heights that are categorized into the three LI ranges [LI < 1, 1 ≤ LI ≤ 2, and LI > 2]).

Weight (kg)	Platform Height		
	Low	Middle	High
Light	3.4, 4.5	2.3, 3.4	1.1, 2.3
Medium	10.2, 11.3, 12.5	9.1, 10.2, 11.3, 12.5	7.9, 9.1, 10.2
Heavy	13.6, 14.7, 15.9, 17.0	13.6, 14.7, 15.9	13.6, 14.7

Table 2
Estimation method of pose-related task variables using the 17 joints from the pose estimation algorithm.

Features	Estimation Method
Horizontal Location	Displacement of mid-point between feet (foot joints) to mid-point between hands (wrist joints) projected to the ground (z-x plane).
Vertical Location	Vertical (y) displacement of the mid-point between both hands (wrist joints) from the mid-point between both feet (foot joints).
Asymmetry Angle	[modified from the definition used by NIOSH due to CV constraints] The angle formed between the line connecting the mid-point of the hands (wrist joints) and the center of the pelvis (asymmetric line) and the orthogonal line drawn from the line connecting the left and right hip (sagittal line) projected to the ground (z-x plane).
Lifting Distance	Absolute difference between the start and end vertical location.

synchronization with video recordings, and the average pressure of sensors throughout the lift for each region was calculated as input features for lifting index prediction.

2.3.4. Ground truth parameters and risk classification

From sections 2.3.2 and 2.3.3, a total of 4 pose-related features and 14 glove features were included in the final dataset (see appendix A2). The ground truth Frequency Independent Lifting Index (LI) for each lifting action was calculated based on the designed task variables. The vertical location was based on the 3 assigned shelf heights (0.7m, 0.9m, 1.1m) and the ground, the horizontal distance and asymmetric angle were maintained at 0.5m and 0°, respectively, and a coupling factor of 1 (good coupling) was assumed.

For the classification task, we have categorized the dataset into three distinct risk levels (an LI below 1 was categorized as 'low risk', an LI between 1 and 2 as 'moderate risk', and an LI greater than 2 as 'high risk'), serving as the ground truth labels.

2.4. Data analysis

2.4.1. Statistical machine learning model

Classification models, including Multinomial Logistic Regression, Support Vector Machine, and Random Forest, were used to process sensor data and predict risk levels. For all ML models, the independent variables are the pose features, the glove features defined in section 2.3.2 and 2.3.3, and the dependent/output is the LI. The predicted LI from these models was compared against the ground truth label of risk (LI) to evaluate the accuracy of each model in estimating ergonomic risk. All models were trained and validated using 5-fold cross-validation. The dataset was split into 80 % for training/validation and 20 % for testing. The performance of all the classification models was evaluated using accuracy, precision, recall, and F1 score. Precision measures how often a model correctly predicts a specific risk level out of all times it predicted that risk level, while recall measures how well the model identifies all actual instances of a risk level. Since our dataset was imbalanced, the Adaptive Synthetic Sampling Approach (ADASYN) (He et al., 2008) was applied to the training set to facilitate robust model training and evaluation. ADASYN has the advantage over the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) by helping adaptively shift decision boundaries based on the minority samples that are difficult to classify, and it avoids the overfitting issue faced by the random oversampling method that replicates existing samples.

2.4.2. Convolution neural network

A Convolutional Neural Network (CNN) model was also used for multi-class classification and regression tasks. The model was designed to process input data through one-dimensional convolutions. The CNN architecture comprises three convolutional layers, with each layer followed by a max-pooling operation to reduce dimensionality and capture the most salient features.

The CNN model was trained using a backpropagation algorithm with the Adam optimizer (Kingma and Ba, 2014), which adapts the learning rate for each parameter. We set the initial learning rate to 0.001 and trained the model over 200 epochs with a batch size of 32. To mitigate the risk of overfitting, we employed a dropout strategy in the fully connected layers and used a validation set to monitor the model's generalization performance.

A weighted-loss function was utilized to handle the imbalanced dataset issue and avoid noise. A higher weight was assigned to errors made on the minority class in this modified Cross-Entropy Loss function (see equation) to quantify the difference between the predicted probabilities and the actual class labels (Phan et al., 2020).

$$L(\theta) = -\frac{1}{N}\sum_{1=i}^{N}\sum_{c=1}^{M}\omega_c \bullet y_{ic} \bullet \log\left(p_{ic}(\theta)\right)$$

N is the number of observations, M is the number of classes, y_ic is a binary indicator of whether class c is the correct classification for observation i, $p\_ic(theta)$ is the predicted probability of observation i being of class c, and theta represents the model parameters. W_c represents the weight assigned to class c that is inversely proportional to the frequency of that class in the dataset. A complete workflow of the risk assessment method is shown in Fig. 4.

2.4.3. Evaluation metrics

The overall performance of each classification model was evaluated using accuracy, precision, recall, and F1-score.

2.4.4. Feature importance

The SHAP (Shapley Additive exPlanations) methodology quantifies the contribution of each individual feature to a prediction by employing a concept from a game theory known as the Shapley value (Hart et al., 1989). The Shapley value was adapted in machine learning to interpret the impact of features within a predictive model. The Deep SHAP explainer, an enhanced version of the DeepLift algorithm (Shrikumar et al., 2017) was used to explain the CNN model. Specifically, the SHAP value for a feature reflects the degree to which it influences the predictive outcome relative to a baseline, thereby explaining whether the existence of a feature increases or decreases the prediction made by the model.

3. Result

3.1. Model input dataset

A total of 3348 lifts were observed, but 576 lifts were removed due to missing video files, and 31 lifts were removed due to missing glove recordings. The final dataset encompassed 2747 lifting tasks. For each task, the ground truth LI (based on task variables) and algorithm-predicted LI (based on sensor measurements) were calculated at both the origin and destination of the lift, thus resulting in a final dataset with 5494 entries. For the classification task, there were 1784 lifting instances in the low-risk category, 2694 lifting instances in the moderate-risk category, and 1016 instances in the high-risk category.

We compared the task variables estimated from computer vision to

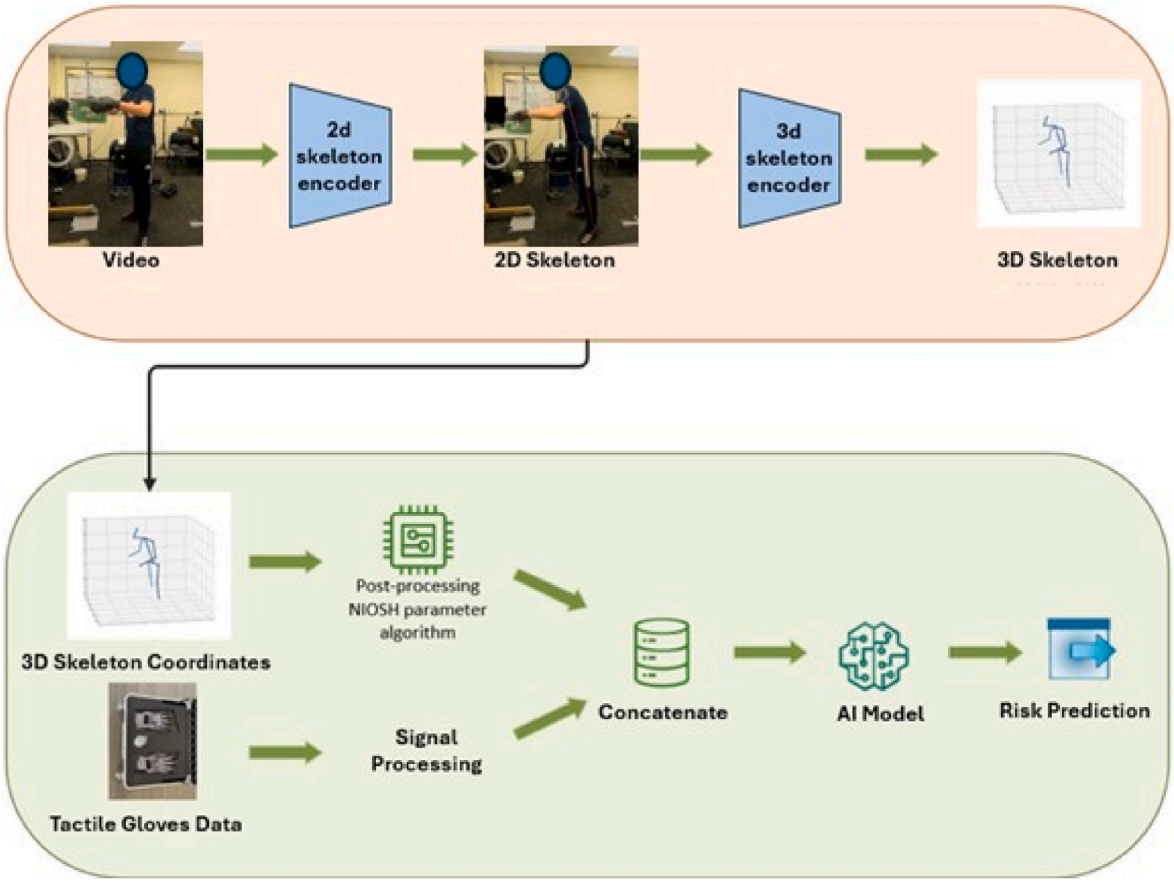


Fig. 4. Workflow of the proposed risk assessment method: the orange section represents the extraction of key body joint locations using MotionBert (Zhu et al., 2022), and the green section represents the estimation of lifting risks using the pose-related task variables and glove features. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Comparison (mean of the absolute difference between estimated values and the corresponding task condition) of estimated task variables using CV (the mean value for each variable) and values from experiment conditions.

	From 3D pose estimation	Experiment condition (including 10 cm offset for container)	Mean absolute difference
V: High_shelf	Mean: 116.41 cm (std:12.95)	120 cm	7.69 cm
V: Middle_shelf	Mean: 100.21 cm (std:9.06)	100 cm	6.48 cm
V: Low_shelf	Mean: 83.31 cm (std:9.33)	80 cm	7.28 cm
Horizontal Distance	Mean: 47.02 cm (std:17.67)	50 cm	13.99 cm
Average Difference			**8.86 cm**
Asymmetric Angle	Mean: 18.71° (std:15.80)	0°	18.71°

the values defined from the designed task conditions (Table 3). The result yielded an average difference of 8.86 cm by averaging the mean absolute differences (excluding asymmetric angle) and 18.71° (asymmetric angle) compared to the expected experimental values for each task condition.

3.2. Machine learning results

The results of the machine learning models are presented in Table 4.

Table 4
Performance (weighted across three classes) of all ML models in the classification task using overall accuracy, precision, recall, and F1 Score (* indicates the performance of the CNN model with the best overall results).

	Accuracy	Precision	Recall	F1 Score
Multinomial Logistic Regression	0.75	0.78	0.75	0.75
Support vector machine (SVM)	0.77	0.82	0.77	0.77
Random Forest	0.81	0.85	0.81	0.81
CNN				
Pose + Glove*	0.89*	0.88*	0.87*	0.88*
Pose-only	0.48	0.44	0.47	0.44
Glove-only	0.83	0.80	0.80	0.80

The CNN was the best-performing model and achieved a 0.89 overall accuracy score and a 0.88 precision score. The detailed classification performance of all tested models is presented in the appendix section (Appendix A3-5).

Using the best model (CNN), we further explored the contribution of sensor modalities to prediction performance. Comparing the performance of the three models (glove + pose, pose-only, glove-only) constructed with different modalities, the model with only the pose-related variables as inputs yielded performance with less than 50 % overall accuracy. A significant improvement in overall performance was achieved by using only the glove features as inputs. The multimodal pose + glove classification yielded an additional 0.08 in F1 score than the glove-only model. The primary improvements between the multimodal classification model and the glove-only model were observed in the moderate and high-risk classification tasks, with the corresponding F1 scores increased from 0.83 to 0.89 and 0.60 to 0.77, respectively (Appendix A3-

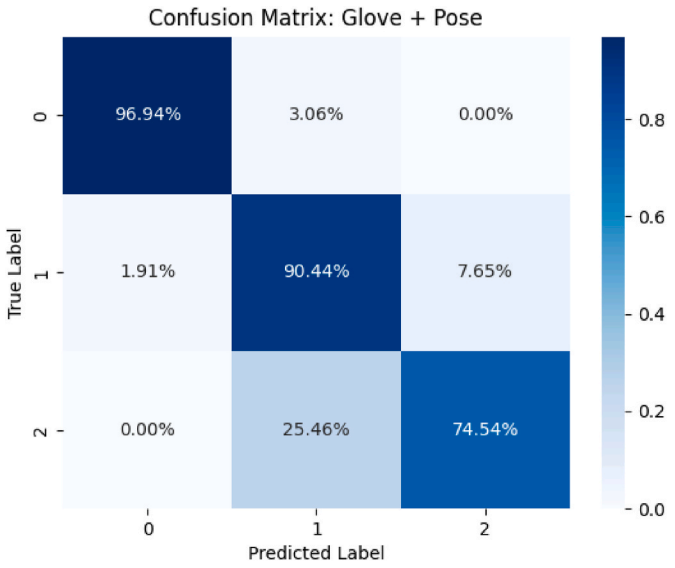


Fig. 5. Confusion matrix showing the recall score for classification of each class (0: LI < 1; 1: 1 ≤ LI ≤ 2; 2: LI > 2); the diagonal cells (from top left to bottom right) represent the percentages of correctly classified instances for each class.

5). The confusion matrix of the multimodal classification model shows a decreasing trend in prediction performance from low risk to high risk (Fig. 5).

3.3. Feature importance

The SHAP values for each feature of our classification model were generated to explain the feature contributions to the model performance (Fig. 6). The mean absolute SHAP value of the right hand was 9 % higher than that of the left hand, indicating the right hand was more important in the classification model. Fingers were more important than palms as the mean absolute SHAP value of the fingers was 14 % higher than the palms. Both the middle and index fingers from the left and right hand were the more significant contributors to our model.

D and V both had mean SHAP values close to 4, as shown in Fig. 6B, indicating that they were the most significant contributors to overall model performance among all the input features. Among all pose-related task variables, both the asymmetric angle and horizontal distance had a weaker contribution to the model compared to the vertical distance (V) and the lifting distance (D).

Fig. 7 illustrates the contribution of glove and pose-related variables based on the cumulative distribution of SHAP values across all three classes. In lower-risk tasks, glove features contributed significantly more to the model than pose-related variables. However, as the risk level increased, the importance of pose-related variables became more prominent in the model's predictions.

4. Discussion

This study demonstrates the feasibility of integrating tactile gloves with computer vision to assess lifting risks based on RNLE, achieving higher accuracy than single-modality approaches. Our integrated method has achieved risk assessments with an overall accuracy of nearly 90 %. This combination addresses some limitations of current single-modality approaches, which frequently overlook direct force measurements and concentrate on narrow areas of impact. The proposed system shows the potential for a non-intrusive solution for field ergonomic risk assessments for manual lifting tasks, enabling real-time data collection for RNLE and aiding in the identification of potential risks associated with two-handed lifting tasks.

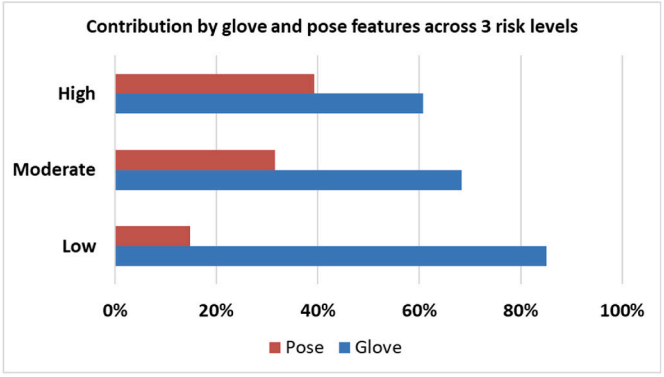


Fig. 7. Comparing the cumulative SHAP values between all glove and pose-related variables across the three risk levels.

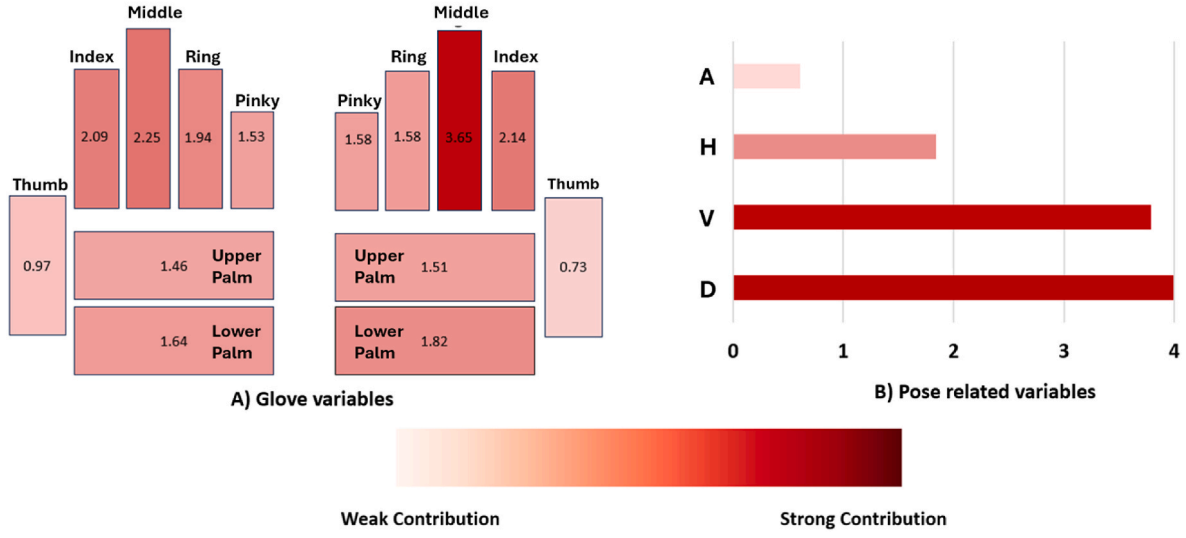


Fig. 6. Mean absolute SHAP value of all glove (7 regions) and pose-related variables (A: asymmetry angle, H: horizontal distance, V: vertical distance, D: lifting distance), the values are averaged across all three risk levels.

4.1. Extraction of pose-related variables

For this study, the participants were instructed to perform symmetric liftings (asymmetric angle = 0) with a fixed horizontal distance. However, relatively inconsistent estimation of foot positions was observed due to camera angles and occlusion of feet. The optimal camera placement for this study should be at a height approximately level with the subject's upper torso, perpendicular to the sagittal plane, with minimized occlusion and distortion for accurate body segment detection. Another issue arose from the fact that the sagittal line for asymmetric angle calculation is difficult to estimate perfectly from the joint-based skeleton, which is also a common challenge faced by the industrial practitioners during the implementation of the NIOSH lifting equation because the sagittal line is well-defined only when the body is in symmetrical postures (Dempsey and Fathallah, 1999). We used unconventional body landmarks (wrist joints, center of pelvis) as practical adaptations to the constraints of the pose estimation algorithm. We used the asymmetry line connecting the mid-point of the hands (wrist joints) and the center of the pelvis (the root joint) instead of the mid-point of the inner ankle bones (as given by the standard NIOSH definition) to estimate the asymmetric angle, and we used foot joints as a proxy for the ground level because the ground plane was not defined by the algorithm. This approach can introduce variability in vertical (~10 cm) and horizontal distance measurements (~17 cm) with a mean absolute error of ~8.86 cm, and it provides only an approximation of the true sagittal line and limits the model's ability to capture subtle variations in foot placement during lifting actions. This approach was established to accommodate limitations in the real-world application of computer vision-based pose estimation where precise anatomical landmarks may not always be detectable and ensure efficiency in real-world adaptation. This trade-off is balanced by the system's ability to automate risk classification efficiently based on the estimated pose information without manually measuring and inputting the task conditions.

Using non-standard body landmarks could introduce slight systematic shifts in vertical and horizontal distance estimations. Variation in estimated pose variables could lead to an under- or over-estimation of LI and potentially misclassify tasks near risk category thresholds into incorrect risk categories. Compared to the fixed task conditions, the estimated asymmetric angle (A) and horizontal location (H) exhibited considerable standard deviations, reflecting higher variability in those pose estimation outputs. In real-world applications, individuals rarely perform tasks in the exact manner specified by the workstation design (Attaianese and Duca, 2012). This can be attributed to multiple factors, including personal habits, comfort, or adaption to specific task demands, which often result in random movements and body twists (Falck and Rosenqvist, 2012). While the task design specified a zero-asymmetry angle, subtle twists occurred as participants anticipated moving toward a subsequent workstation positioned 90° away. These slight movements, which are not captured in the standard NIOSH asymmetry angle definition, introduce inaccuracies in the estimation. In practice, field data suggest variability in how the asymmetry angle is measured by different practitioners. Both the sagittal and asymmetry lines are imaginary and cannot be physically marked or drawn in the workplace, and feet movement during lifting tasks adds to the difficulty of capturing such static measurements (Dempsey and Fathallah, 1999). Moreover, participants wouldn't have identical lifting actions for every repetition, potentially due to fatigue or learning from the lifting tasks (Zhang et al., 2019). Consequently, sensor-based systems provide a more dynamic and precise assessment of body postures during lifting tasks, enhancing the NIOSH definitions by accounting for the variability caused by body movements.

A key challenge in biomechanic assessments is the requirement of complex sensor or camera calibration/setup to account for varying distances, angles, and environments. Despite known limitations of a single-camera setup, it remains more feasible than multiple cameras (Wang et al., 2021b), (Arvanitis et al., 2023) and wearable motion capture systems (Lu et al., 2020), (Bortolini et al., 2018) in an industry setting due to its adaptability and ease of use. The use of 3D pose estimation from a single camera coupled with height-based scaling offers versatility; standardizing measurements across individuals ensures that joint coordinates are represented consistently regardless of the relative position of the person to the camera. Height-based scaling also offers ease of implementation, height is a readily available and easily measurable parameter, making it convenient for scaling purposes. A limitation of our proposed method is the reliance on manual height measurement, which introduces an additional step that can lead to potential error, especially since the estimated skeleton only roughly reflects the person's actual height. Additionally, the use of a 3D pose estimation algorithm with uncalibrated cameras can affect the precision of joint coordinate mapping, especially in complex environments with occlusions and inconsistent camera angles. To obtain the most accurate pose information, the gold standard would still be to calibrate the camera or use a multi-camera setup to enhance accuracy. Furthermore, we showed that a single camera was accurate in extracting the proposed pose-related features for our model in our study environment, recognizing that laboratory-level precision is not always essential for all occupational safety applications (Greene et al., 2019). Although it demonstrates the potential of our camera and glove approach to risk assessment, we acknowledge the next step is to test the robustness of this approach against the known limitations of single camera studies.

4.2. Modeling results

We achieved the best classification result using the CNN model in comparison to other statistical ML models (Table 4). The CNN model had the best performance at correctly predicting the low-risk lifts, and the performance progressively dropped as the risk level got higher and decreased when predicting high risks. Several reasons could contribute to this behavior. First, approximately 25 % of the high-risk lifts were mistakenly classified as moderate risks. Specifically, the average LI difference between low and moderate was 1.22, while the average LI difference between moderate and high was 0.70. The fact that tactile gloves are the sole predictors for lifting weights might also be a contributing factor to reduced performance. The earlier regression model of weight prediction based on the same dataset but using the time-series glove sensor data reported significantly lower accuracy when estimating heavier weights (Zhou et al., 2023b), and lifting weight was the most significant contributor to increasing lifting risks in this study. The full CNN model with both glove- and pose-only showed marginal improvement over the glove-only model, likely due to the experiment's design, which focused on weight as the main risk factor (LI: 1, 2, 3), with limited variation in other variables like vertical and horizontal distance. This also explains why the pose-only model had reduced performance.

Current methods typically rely on the indirect estimation of lifting weight factors, utilizing tools like sEMG for muscle activity, IMUs for dynamic movement data, or facial expressions (Brandt et al., 2018), (Varrecchia et al., 2018), (Zhou et al., 2022). Our multimodal approach provided a direct assessment of lifting risk variables (both posture and lifting weight information) with precision scores for each class (Low, Moderate, High) as 0.97, 0.88, and 0.80. While the majority of previous efforts strived to achieve accurate estimation of lifting-risk-related factors such as joint kinematics, range of motions, or human postures using CV-based or IMU-based methods (Greene et al., 2022), (Mehrizi et al., 2018), (Conforti et al., 2020); methods that provided complete lifting risks assessment were still relying on human posture only and limited to risk prediction based on ACGIH lifting zones (Snyder et al., 2021) or Rapid Upper Limb Assessment (RULA) scores (MassirisFernández et al., 2020). From efforts that tried to include lifting weight as a significant contributor to risk assessment in lifting, previous studies have included kinematic pose-related variables such as acceleration/jerk or facial expressions to account for lifting weight and perceived effort in their lifting risk prediction models; and each has achieved either a 0.65

accuracy in binary weight classification task (Hlucny and Novak, 2020) or 0.83, 0.63, and 0.67 precision score in a three-class risk prediction task (Zhou et al., 2022). Using only the static pose-related variables as the predictors in our model to predict lifting risks has yielded a very poor performance, showing that pose-related variables alone were not enough to distinguish differences in lifting index without considering the associated load weight.

4.3. Feature importance

A possible explanation for the observation that the asymmetric angle and horizontal distance contributed significantly less than the rest of the pose-related variables is that they were designed (i.e., inputs into ground truth LI) to be constant values throughout the lifting experiment. The variations in asymmetric and horizontal measurements reflects how participants performed the task vs. what was expected through design. Since the ground truth was calculated on designed variables, this is a potential explanation for why asymmetric angle and horizontal distance were not important predictor features in our model.

Results suggest that glove variables tend to have a greater impact on prediction compared to pose-related variables, particularly for tasks classified as low-risk, as illustrated by the cumulative SHAP values in Fig. 7. For the classification of low LI categories, pose-related variables on average contribute significantly less to model performance compared to glove features. However, as the lifting risk rises, the increasing importance of pose-related variables may be attributed to changes in posture in response to varying weights. Although poses alone may not be strong indicators of lifting risk, their relevance grows when heavier weights and higher risk levels are involved. For example, lighter weight generally places less strain on the body, making it easier for workers to choose different poses without significantly increasing the risk of injury. However, for heavy lifting tasks, postures that minimize extended reach and movements for workers may be needed to mitigate strains on various body joints and reduce overall risk.

Similar to the earlier study using tactile gloves to predict lifting weight (Zhou et al., 2023b), our CNN prediction model gave more weight to the right hand than the left when assessing lifting risks, likely due to stronger grip strength in the dominant hand (80 % of all participants are right-handed) (Petersen et al., 1989), (Incel et al., 2002). Moreover, the model relied slightly more on the other four fingers (index, middle, ring, and pinky) than on the thumb or palm. Since the boxes used in the experiment had either fully cut-out or squared-edge handles, this observation can be explained as most load from the boxes was applied to the participants' index, middle, ring, and pinky fingers during lifting. However, it has been shown that hand and wrist positions could impact gripping strength (Shurrab et al., 2015). Thus, the inclusion of different types of handles and hand postures could make the model more robust toward the lifting weight component under different task conditions, bringing this initial risk prediction model closer to workplace applications.

4.4. Applications to practitioners

Our work demonstrated the feasibility of a non-intrusive method for predicting lifting risks using computer vision and tactile gloves, achieving high accuracy and addressing the limitations of single-modality approaches. Wearing gloves is common (Zhou et al., 2023a) and recommended (Zhao et al., 2021) for workplace lifting tasks. Compared to methods that require the placement of multiple IMU or sEMG sensors on the workers (Donisi et al., 2021), (Ranavolo et al., 2018), (Donisi et al., 2022), the non-intrusive nature of our method enhances its potential applicability and simplifies implementation in workplace settings.

In practical application, the proposed algorithm provides a potential tool for ergonomists and safety professionals in the field of manual material handling (MMH). The integration of computer vision with sensor gloves offers a novel approach to risk assessment. To implement this system, practitioners would equip workers with sensor gloves and place cameras to capture workers' movements without the need for specialized calibration except for noting worker height, thereby reducing setup complexity. Upon capturing the data, the algorithm processes the information and provides feedback on lifting risk levels that can be used for proactive interventions to prevent musculoskeletal injuries. While using standard cameras is low-cost and easy to implement, the current tactile gloves used in this study are an emerging technology with a higher cost, primarily designed for research purposes. The current research-grade gloves feature 65 sensors per glove, but reduction of sensors (e.g., our feature extraction reduced the 65 sensors into 7 hand regions) and scaling manufacturing may make the technology accessible to more workers once future research establishes the robustness of this approach in industry settings.

5. Limitations and future work

Although the all-in-one lifting risk assessment tool facilitated by the CV-glove multimodal CNN model achieved an overall F1 score of 0.88 in the lifting risk classification task under our experimental conditions, several limitations need to be addressed before this technique is ready for application in the workplace. First, future studies should include more varying lifting task conditions, including more weight/height combinations, horizontal distance, lifting velocity, and body twists, to make the model more robust and generalizable. Including different objects that would require different handling techniques would expose the model to a broader range of grip patterns and pressure distributions and improve the prediction power of the tactile gloves. The study design did not do full randomization of task, weight, and height, instead keeping task and height order consistent while randomizing the weight order to reduce the complexity of the experimental design. Potential learning/anticipation effects will need to be explored in larger-scale studies to explore potential confounding effects of task and height order. More weight and weight variation can ensure that the model performs well across a broader range of lifting conditions. Lifting velocity was not controlled in this study; variations in velocity may influence gripping force and glove pressure, potentially affecting pressure distribution across sensors. However, we expect our dataset of varying task conditions and over 31 subjects to capture some of these potential variations, and the resulting model was able to classify low/medium/high LI's. Incorporating lifting velocity measurements could provide additional insights into the dynamics of lifting tasks to enhance the robustness of the proposed model, as gripping force/pressure reading may vary depending on the hand velocity and handling technique. Second, there may be inaccuracies in the pose estimation algorithm. A pretrained MotionBert model trained and finetuned on the Human3.6M dataset (Ionescu et al., 2014) with an average mean per joint position error (MPJPE) of 37.5 mm (Zhu et al., 2022) was implemented, and the training set didn't contain actions representing manual material handling (MMH) tasks. A more accurate and robust pose estimation model for this pipeline could be finetuned by creating a human pose dataset with MMH tasks or augmenting existing 2D datasets to high-quality 3D pseudo-annotations with state-of-the-art techniques such as Exemplar Finetuning (Joo et al., 2021). A more refined pose estimation algorithm may enable a more precise calculation of the RNLE parameters by adhering to the standard NIOSH definitions, making the model more accurate and robust. Aligning more closely with standard definitions can facilitate broader adoption and validation of the model in ergonomic risk assessments. Despite utilizing a height-based scaling approach, developing a dedicated 3d pose estimation algorithm is recommended to focus on a wide range of lifting actions with calibrated cameras to generate precise metric calculations of 3d poses. Finally, many other architectures of neural networks exist with exceptional performance in handling different tasks and data structures. The current model's diminished performance in the high-risk category could

underestimate high-risk tasks and mistakenly predict them into moderate-risk tasks, the ideal model should prioritize genuinely high-risk tasks for maximized utility. Future studies can improve the model by adjusting the penalty for misclassifying a high-risk task and include more high-risk task data in the training (the current dataset is unbalanced with fewer high-risk tasks) for robustness and performance.

6. Conclusion

Our study demonstrated a novel multimodal method of predicting lifting risk based on RNLE measurements through tactile gloves and computer vision. The proposed method outperforms single-modality approaches by capturing comprehensive input data required for the RNLE and provides a robust, non-intrusive method for automated risk assessments in manual lifting tasks. Under experimental conditions, our current model achieves an overall F1 score of 0.88. This work has substantial implications for workplace safety, offering a practical and flexible tool for real-time risk assessment that could be easily implemented in various industrial settings.

CRediT authorship contribution statement

Haozhi Chen: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peiran Liu:** Writing – review & editing, Methodology, Formal analysis. **Guoyang Zhou:** Writing – review & editing, Data curation, Conceptualization. **Ming-Lun Lu:** Writing – review & editing, Conceptualization. **Denny Yu:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

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

Acknowledgment

This work was funded in part by a grant from NIOSH (#1R21OH012710). Findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the National Institute for Occupational Safety and Health (NIOSH), the Center for Disease Control and Prevention (CDC). The mention of any company or product does not constitute endorsement by NIOSH, CDC.

Appendix

A1. List of the 17 distinct joints estimated in this study with computer vision

Joint index	Joints
[1]	Root (Hip)
[2]	Right Hip
[3]	Right Knee
[4]	Right Foot
[5]	Left Hip
[6]	Left Knee
[7]	Left Foot
[8]	Spine
[9]	Thorax
[10]	Neck/Nose
[11]	Head
[12]	Left Shoulder
[13]	Left Elbow
[14]	Left Wrist
[15]	Right Shoulder
[16]	Right Elbow
[17]	Right Wrist

A2. List of features

Feature index	Feature name
[1]	right_thumb
[2]	right_index
[3]	right_middle
[4]	right_ring
[5]	right_pinky
[6]	right_upper_palm
[7]	right_lower_palm
[8]	left_thumb
[9]	left_index
[10]	left_middle
[11]	left_ring
[12]	left_pinky
[13]	left_upper_palm
[14]	left_low_palm
[15]	Vertical distance

(*continued on next page*)

(*continued*)

Feature index	Feature name
[16]	Horizontal distance
[17]	Lifting distance
[18]	Asymmetric angle

A3. Performance of classification models for low risk (LI < 1)

	Precision	Recall	F1 Score
Multinomial Logistic Regression	0.78	0.94	0.85
SVM	0.79	0.94	0.86
Random Forest	0.91	0.94	0.92
CNN			
Pose + Glove	0.97	0.97	0.97
Pose-only	0.39	0.26	0.31
Glove-only	0.99	0.97	0.98

A4. Performance of classification models for moderate risk (1 < LI < 2)

	Precision	Recall	F1 Score
Multinomial Logistic Regression	0.89	0.60	0.72
SVM	0.92	0.63	0.75
Random Forest	0.92	0.68	0.78
CNN			
Pose + Glove	0.88	0.90	0.89
Pose-only	0.59	0.60	0.59
Glove-only	0.82	0.84	0.83

A5. Performance of classification models for high risk (LI > 2)

	Precision	Recall	F1 Score
Multinomial Logistic Regression	0.54	0.78	0.64
SVM	0.55	0.85	0.67
Random Forest	0.55	0.91	0.68
CNN			
Pose + Glove	0.80	0.75	0.77
Pose-only	0.36	0.54	0.43
Glove-only	0.61	0.59	0.60

References

American Conference of Governmental Industrial Hygienists (ACGIH), 2005. Threshold Limit Values for Chemical Substances and Physical Agents & Biological Exposure Indices.

Antwi-Afari, M.F., Li, H., Edwards, D.J., Pärn, E.A., Seo, J., Wong, A.Y.L., 2017. Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers. Autom. Constr. 83, 41–47. https://doi.org/10.1016/j.autcon.2017.07.007.

Arvanitis, G., Piperigkos, N., Anagnostopoulos, C., Lalos, A.S., Moustakas, K., 2023. Real time enhancement of operator's ergonomics in physical human - robot collaboration scenarios using a multi-stereo camera system. In: Proceedings of the IEEE International Conference on Industrial Technology. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICIT58465.2023.10143035

Attaianese, E., Duca, G., 2012. Human factors and ergonomic principles in building design for life and work activities: an applied methodology. Theor. Issues Ergon. Sci. 13 (2), 187–202. https://doi.org/10.1080/1463922X.2010.504286.

Bortolini, M., Gamberi, M., Pilati, F., Regattieri, A., 2018. Automatic assessment of the ergonomic risk for manual manufacturing and assembly activities through optical motion capture technology. In: Procedia CIRP. Elsevier B.V., pp. 81–86. https://doi.org/10.1016/j.procir.2018.03.198

Brandt, M., et al., 2018. Accuracy of identification of low or high risk lifting during standardised lifting situations. Ergonomics 61 (5), 710–719. https://doi.org/10.1080/00140139.2017.1408857.

Burgess-limerick, R., Abernethy, B., 1997. Toward a quantitative definition of manual lifting postures. Hum. Factors 39 (1), 141–148.

Callaghan, J.P., Salewytsch, A.J., Andrews, D.M., 2001. An evaluation of predictive methods for estimating cumulative spinal loading. Ergonomics 44 (9), 825–837. https://doi.org/10.1080/00140130118541.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity fields [Online]. Available: http://arxiv.org/abs/1812.08008.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Conforti, I., Mileti, I., Del Prete, Z., Palermo, E., 2020. Measuring biomechanical risk in lifting load tasks through wearable system and machine-learning approach. Sensors (Switzerland) 20 (6). https://doi.org/10.3390/s20061557.

Craig, B.N., Congleton, J.J., Kerk, C.J., Amendol, A.A., Gaines, W.G., Jenkins, O.C., 2003. A prospective field study of the relationship of potential occupational risk factors with occupational injury/illness. Am. Ind. Hyg. Assoc. J. 64 (3), 376–387. https://doi.org/10.1080/15428110308984830.

Debnath, B., O'Brien, M., Yamaguchi, M., Behera, A., 2022. A review of computer vision-based approaches for physical rehabilitation and assessment. Multimed. Syst. 28 (1), 209–239. https://doi.org/10.1007/s00530-021-00815-4.

Dempsey, P.G., Fathallah, F.A., 1999. Application issues and theoretical concerns regarding the 1991 NIOSH equation asymmetry multiplier. Int. J. Ind. Ergon. 23, 181–191.

Dolan, P., et al., 2001. An EMG technique for measuring spinal loading during asymmetric lifting. Clin. BioMech. 16 (1), 17–24 [Online]. Available: www.elsevier.com/locate/clinbiomech.

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). https://doi.org/10.3390/s21082593.

Donisi, L., et al., 2022. A logistic regression model for biomechanical risk classification in lifting tasks. Diagnostics 12 (11). https://doi.org/10.3390/diagnostics12112624.
Egeonu, D., Jia, B., 2024. A systematic literature review of computer vision-based biomechanical models for physical workload estimation. Ergonomics 68 (2), 139–162. https://doi.org/10.1080/00140139.2024.2308705.
Falck, A.C., Rosenqvist, M., 2012. What are the obstacles and needs of proactive ergonomics measures at early product development stages? - An interview study in five Swedish companies. Int. J. Ind. Ergon. 42 (5), 406–415. https://doi.org/10.1016/j.ergon.2012.05.002.
Fox, R.R., Lu, M.L., Occhipinti, E., Jaeger, M., 2019. Understanding outcome metrics of the revised NIOSH lifting equation. Appl. Ergon. 81 (Nov). https://doi.org/10.1016/j.apergo.2019.102897.
Gallagher, S., Sesek, R.F., Schall, M.C., Huangfu, R., 2017. Development and validation of an easy-to-use risk assessment tool for cumulative low back loading: the Lifting Fatigue Failure Tool (LiFFT). Appl. Ergon. 63, 142–150. https://doi.org/10.1016/j.apergo.2017.04.016.
Granata, K.P., Marras, W.S., 1995. An EMG-assisted model of trunk loading during free-dynamic lifting. J. Biomech. 28 (11), 1309–1317.
Greene, R.L., et al., 2019. Predicting sagittal plane lifting postures from image bounding box dimensions. Hum. Factors 61 (1), 64–77. https://doi.org/10.1177/0018720818791367.
Greene, R.L., Chen, G., Lu, M.L., Hu, Y.H., Radwin, R.G., 2021. Enhancing the revised NIOSH lifting equation using computer vision. In: Proceedings of the Human Factors and Ergonomics Society. SAGE Publications Inc., pp. 467–471. https://doi.org/10.1177/1071181321651211
Greene, R.L., Hayden, M., Hu, Y.H., Radwin, R.G., 2022. Estimating trunk angle kinematics during lifting using a computationally efficient computer vision method. Hum. Factors 64 (3), 482–498.
Hart, S., 1989. In: Theory, Game, Eatwell, J., Milgate, M., Newman, P. (Eds.), Shapley Value. Palgrave Macmillan, London, pp. 210–216.
Health and Safety Executive, 2019. Manual handling assessment charts (the MAC tool) [Online]. Available: https://www.hse.gov.uk/pubns/indg383.pdf.
He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks. IEEE World Congress on Computational Intelligence), pp. 1322–1328.
Hlucny, S.D., Novak, D., 2020. Characterizing human box-lifting behavior using wearable inertial motion sensors. Sensors (Switzerland) 20 (8). https://doi.org/10.3390/s20082323.
Incel, N.A., Ceceli, E., Durukan, B., Erdem, R., Yorgancioglu, R., 2002. Grip strength: effect of hand dominance. Singap. Med. J. 43 (5), 234–237.
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 Trans. Pattern Anal. Mach. Intell. 36 (7), 1325–1339. https://doi.org/10.1109/TPAMI.2013.248.
Iskakov, K., Burkov, E., Lempitsky, V., Malkov, Y., 2019. Learnable triangulation of human pose [Online]. Available: http://arxiv.org/abs/1905.05754.
Joo, H., Neverova, N., Vedaldi, A., 2021. Exemplar fine-tuning for 3D human model fitting towards in-the-wild 3D human pose estimation. In: Proceedings - 2021 International Conference on 3D Vision, 3DV 2021. Institute of Electrical and Electronics Engineers Inc., pp. 42–52. https://doi.org/10.1109/3DV53792.2021.00015
Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization [Online]. Available: http://arxiv.org/abs/1412.6980.
Lowe, B.D., Dempsey, P.G., Jones, E.M., 2019. Ergonomics assessment methods used by ergonomics professionals. Appl. Ergon. 81. https://doi.org/10.1016/j.apergo.2019.102882.
Lu, M.L., Waters, T., Werren, D., 2015. Development of human posture simulation method for assessing posture angles and spinal loads. Hum. Factors Ergon. Manuf. 25 (1), 123–136. https://doi.org/10.1002/hfm.20534.
Lu, M.L., Putz-Anderson, V., Garg, A., Davis, K.G., 2016. Evaluation of the impact of the revised national Institute for occupational safety and health lifting equation. Hum. Factors 58 (5), 667–682. https://doi.org/10.1177/0018720815623894.
Lu, M.L., Barim, M.S., Feng, S., Hughes, G., Hayden, M., Werren, D., 2020. Development of a wearable IMU system for automatically assessing lifting risk factors. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, pp. 194–213. https://doi.org/10.1007/978-3-030-49904-4_15.
Marras, W.S., et al., 1993. The role of dynamic three-dimensional trunk motion in occupationally-related low back disorders: the effects of workplace factors, trunk position, and trunk motion characteristics on risk of injury. Spine 18, 617–628.
Marucci-Wellman, H.R., et al., 2015. The direct cost burden of 13 years of disabling workplace injuries in the U.S. (1998-2010): findings from the Liberty Mutual Workplace Safety Index. J. Saf. Res. 55, 53–62. https://doi.org/10.1016/j.jsr.2015.07.002.
MassirisFernández, M., Fernández, J.Á., Bajo, J.M., Delrieux, C.A., 2020. Ergonomic risk assessment based on computer vision and machine learning. Comput. Ind. Eng. 149. https://doi.org/10.1016/j.cie.2020.106816.
Mehrizi, R., Peng, X., Xu, X., Zhang, S., Metaxas, D., Li, K., 2018. A computer vision based method for 3D posture estimation of symmetrical lifting. J. Biomech. 69, 40–46. https://doi.org/10.1016/j.jbiomech.2018.01.012.
Murugan, A.S., et al., 2024. Optimising computer vision-based ergonomic assessments: sensitivity to camera position and monocular 3D pose model. Ergonomics 68 (1), 120–137. https://doi.org/10.1080/00140139.2024.2304578.
National Safety Council. Work Injury Costs [Online]. Available: https://injuryfacts.nsc.org/work/costs/work-injury-costs/#:~:text=Work%20Injury%20Costs%20and%20Time%20Lost&text=The%20total%20cost%20of%20work,administrative%20expenses%20of%20%2457.5%20billion. (Accessed 31 March 2024).
Pesenti, M., Invernizzi, G., Mazzella, J., Bocciolone, M., Pedrocchi, A., Gandolla, M., 2023. IMU-based human activity recognition and payload classification for low-back exoskeletons. Sci. Rep. 13 (1). https://doi.org/10.1038/s41598-023-28195-x.
Petersen, P., Petrick, M., Connor, H., Conklin, D., 1989. Grip strength and hand dominance: challenging the 10% rule. Am. J. Occup. Ther. 43 (7), 444–447. https://doi.org/10.5014/ajot.43.7.444.
Phan, H., Yamamoto, K., Phan, T.H., Yamamoto, K., 2020. Resolving class imbalance in object detection with weighted cross entropy losses [Online]. Available: https://arxiv.org/abs/2006.01413.
Potvin, J.R., Ciriello, V.M., Snook, S.H., Maynard, W.S., Brogmus, G.E., 2021. The Liberty Mutual manual materials handling (LM-MMH) equations. Ergonomics 64 (8), 955–970. https://doi.org/10.1080/00140139.2021.1891297.
Ranavolo, A., et al., 2017. Mechanical lifting energy consumption in work activities designed by means of the 'revised NIOSH lifting equation. Ind. Health 55 (5), 444–454.
Ranavolo, A., et al., 2018. Surface electromyography for risk assessment in work activities designed using the "revised NIOSH lifting equation". Int. J. Ind. Ergon. 68, 34–45. https://doi.org/10.1016/j.ergon.2018.06.003.
Santopaolo, A., et al., 2022. Biomechanical risk assessment of human lifting tasks via supervised classification of multiple sensor data. In: IEEE-RAS International Conference on Humanoid Robots. IEEE Computer Society, pp. 746–751. https://doi.org/10.1109/Humanoids53995.2022.10000147.
Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences [Online]. Available: https://arxiv.org/abs/1704.02685.
Shurrab, M., Mohanna, R., Shurrab, S., Mandahawi, N., 2015. Experimental design to evaluate the influence of anthropometric factors on the grip force and hand force exertion. Int. J. Ind. Ergon. 50, 9–16. https://doi.org/10.1016/j.ergon.2015.09.005.
Snyder, K., et al., 2021. A deep learning approach for lower back-pain risk prediction during manual lifting. PLoS One 16 (2). https://doi.org/10.1371/journal.pone.0247162.
Thomas, B., Lu, M.L., Jha, R., Bertrand, J., 2022. Machine learning for detection and risk assessment of lifting action. IEEE Trans. Hum. Mach. Syst. 52 (6), 1196–1204. https://doi.org/10.1109/THMS.2022.3212666.
Toshev, A., Szegedy, C., 2013. DeepPose: human pose estimation via deep neural networks [Online]. Available: https://arxiv.org/abs/1312.4659.
US Bureau of Labor Statistics, 2023. Employer Reported Workplace Injuries and Illnesses 2021-2022 [Online]. Available: (Accessed 30 March 2024).
Van Dieen, J.H., Hoozemans, M.J.M., Toussaint, H.M., 1999. Stoop or squat: a review of biomechanical studies on lifting technique. Clin. BioMech. 14, 685–696 [Online]. Available: www.elsevier.com/locate/clinbiomech.
Varrecchia, T., et al., 2018. Lifting activity assessment using surface electromyographic features and neural networks. Int. J. Ind. Ergon. 66, 1–9. https://doi.org/10.1016/j.ergon.2018.02.003.
Varrecchia, T., Conforto, S., De Nunzio, A.M., Draicchio, F., Falla, D., Ranavolo, A., 2022. Trunk muscle coactivation in people with and without low back pain during fatiguing frequency-dependent lifting activities. Sensors 22 (4). https://doi.org/10.3390/s22041417.
Wang, X., Hu, Y.H., Lu, M.L., Radwin, R.G., 2019. The accuracy of a 2D video-based lifting monitor. Ergonomics 62 (8), 1043–1054. https://doi.org/10.1080/00140139.2019.1618500.
Wang, H., Xie, Z., Lu, L., Li, L., Xu, X., 2021a. A computer-vision method to estimate joint angles and L5/S1 moments during lifting tasks through a single camera. J. Biomech. 129 (Dec). https://doi.org/10.1016/j.jbiomech.2021.110860.
Wang, X., Hu, Y.H., Lu, M.L.J., Radwin, R., 2021b. Load asymmetry angle estimation using multiple-view videos. IEEE Trans. Hum. Mach. Syst. 51 (6), 734–739. https://doi.org/10.1109/THMS.2021.3112962.
Waters, T.R., Putz-Anderson, V., Garg, A., Fine, L.J., 1993. Revised NIOSH equation for the design and evaluation of manual lifting tasks. Ergonomics 36 (7), 749–776. https://doi.org/10.1080/00140139308967940.
Waters, T.R., Baron, S.L., Kemmlert, K., 1998. Accuracy of measurements for the revised NIOSH lifting equation. Appl. Ergon. 29 (6), 433–438.
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. J. Occup. Environ. Med. 53 (9), 1061–1067. https://doi.org/10.1097/JOM.0b013e31822cfe5e.
Zhang, L., Diraneyya, M.M., Ryu, J.H., Haas, C.T., Abdel-Rahman, E.M., 2019. Jerk as an indicator of physical exertion and fatigue. Autom. Constr. 104, 120–128. https://doi.org/10.1016/j.autcon.2019.04.016.
Zhao, C., Li, K.W., Yi, C., 2021. Assessments of work gloves in terms of the strengths of hand grip, one-handed carrying, and leg lifting. Appl. Sci. (Switzerland) 11 (18). https://doi.org/10.3390/app11188294.
Zhou, G., Aggarwal, V., Yin, M., Yu, D., 2022. A computer vision approach for estimating lifting load contributors to injury risk. IEEE Trans. Hum. Mach. Syst. 52 (2), 207–219. https://doi.org/10.1109/THMS.2022.3148339.
Zhou, G., Lu, M.L., Yu, D., 2023a. Investigating gripping force during lifting tasks using a pressure sensing glove system. Appl. Ergon. 107 (Feb). https://doi.org/10.1016/j.apergo.2022.103917.
Zhou, G., Lu, M.L., Yu, D., 2023b. Tactile gloves predict load weight during lifting with deep neural networks. IEEE Sens. J. 23 (16), 18798–18809. https://doi.org/10.1109/JSEN.2023.3289670.
Zhu, W., Ma, X., Liu, Z., Liu, L., Wu, W., Wang, Y., 2022. MotionBERT: a unified perspective on learning human motion representations [Online]. Available: http://arxiv.org/abs/2210.06551.