Predicting Weight and Strenuousness from High-Speed Videos of Subjects Attempting Lift

Joseph Judge
judgejc@clarkson.edu
Clarkson University
Potsdam, NY, USA

Priyo Ranjan Kundu Prosun
prosunp@clarkson.edu
Clarkson University
Potsdam, NY, USA

Owen Talmage
otalmage@delsys.com
Delsys
Natick, MA, USA

Austin Dykeman
dykemaa@clarkson.edu
Clarkson University
Potsdam, NY, USA

Sean Banerjee
sbanerje@clarkson.edu
Clarkson University
Potsdam, NY, USA

Natasha Kholgade Banerjee
nbanerje@clarkson.edu
Clarkson University
Potsdam, NY, USA

ABSTRACT

Intelligent understanding of human motions during lifting minimizes overexertion injuries by offering continuous monitoring and early intervention for people attempting heavy lifts at home or work. Standardized lift assessment methods such as the revised National Institute for Occupational Safety and Health lift equation require physical therapy expertise and lack subject perceptions of strenuousness in attempting lift tasks. We provide one of the first approaches to perform non-intrusive automated detection of strenuousness, weight lifted, and subject's knowledge of weight using joint motions of subjects from multi-view high-speed color videos as subjects lift varying weights, with and without prior weight knowledge. We show average accuracies of 81.32% and 77.23% by using convolutional neural networks to automatically detect low versus high weight and strenuousness. Our work informs monitoring technologies for individuals engaged in heavy lifting in manufacturing, warehousing, nursing, and emergency response.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; • **Computing methodologies** → **Neural networks**.

KEYWORDS

healthcare artificial intelligence, lift analysis, deep networks

ACM Reference Format:
Joseph Judge, Priyo Ranjan Kundu Prosun, Owen Talmage, Austin Dykeman, Sean Banerjee, and Natasha Kholgade Banerjee. 2022. Predicting Weight and Strenuousness from High-Speed Videos of Subjects Attempting Lift. In *ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE' 22), November 17–19, 2022, Washington, DC, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3551455.3559602

CHASE' 22, November 17–19, 2022, Washington, DC, USA

ACM ISBN 978-1-4503-9476-5/22/11…$15.00
https://doi.org/10.1145/3551455.3559602

1 INTRODUCTION

A large number of activities and occupations involve potentially injurious interactions with objects. Pervasive artificial intelligence (AI) that is human-aware, yet non-intrusive and seamless, plays an important role in ensuring safety during strenuous interactions while allowing people to conduct activities with minimal impact on performance and efficiency. In 2019, 9.8% of all nonfatal occupational injuries and illnesses as reported to the U.S. Bureau of Labor Statistics, and 12.5% of transportation and warehousing injuries, were classified as 'overexertion in lifting or lowering', ranking fourth among all 20 categories of injury [13]. Workers sustaining these injuries were out of work for a median of 13 days [14]. Nursing assistants [6], blue-collar workers in warehousing and manufacturing [15], and first responders such as firefighters and emergency personnel [2, 3, 7] routinely cite back and musculoskeletal injuries in performing lifting of people or heavy objects as part of the requirements of the job. Musculoskeletal injuries reduce the lifespan of workers in these occupations [12], which is especially of concern in rural areas, where aging populations rely on the longevity of work lifespans for sustenance and support of community members. Women and older workers are more likely to report lower back pain [23], and be impacted in reduction of work lifespan and inclusivity due to time lost from work [9]. Pervasive AI in home and work environments that continuously monitors people for strenuous activity such as heavy lifts can improve quality of life by performing early detection of injury potential, demonstrating concern through sympathetic human-computer interfaces, and signalling for assistance.

Assessment of lifting activities to date has been largely manual, conducted by physical therapy experts in controlled environments. Workplace assessments are usually sporadic. The most common form of lift assessment occurs by leveraging the revised National Institute for Occupational Safety & Health (NIOSH) lift equation (RNLE) [22], that provides recommended weight limit (RWL) using measurements of a series of parameters related to object weight, joint positions, frequency of lifts, duration of lift, and hand-to-object coupling quality, as subjects attempt a lift in a controlled setting. The lift index (LI), computed as a ratio of the actual weight carried to the RWL, provides an estimate of injury risk. Manual collection of data related to the lift equation is time-intensive, requires physical therapy expertise for complex factors such as the quality of

coupling, and is unlikely to occur frequently, if at all, in unstructured environments on the job. While automated approaches exist to estimate the parameters of the RNLE using wearables [1, 5, 17–20], not all parameters of the RNLE may be relevant for every form of lift that subjects may encounter on the job, e.g., performing a single-handed lift or encountering a hazard during a lift, such as a slip or fall. Importantly, the RNLE fails to provide knowledge of perception of the lift task from the perspective of the *human*, i.e., how strenuous did the person perceive the task to be or whether they had prior knowledge of the object's weight. Subjects' perceptions of weight are likely to impact how they lift weights. Intelligent monitoring systems are more likely to be successful at offering assistance if they are cognizant subjects' perceptions of lift.

In this work, we perform non-intrusive analyses of human lift to facilitate automated prediction of task parameters such as prior weight knowledge, task-based weight recognition, and strenuousness by using data captured with high-speed color cameras as part of a human subjects' study with $N = 23$ subjects. Each subject attempts multiple randomized 2-handed lifting of identical cartons of 4 different weights, with and without being informed of the weight prior to lift. We track subject posture using OpenPose [4] to garner an understanding of subject-to-subject variabilities in adaptation of human posture to varying weights during lift. We obtain average accuracies of 81.32% and 77.23% in performing low versus high weight and strenuousness detection, and 58.01% in performing 4-class weight prediction from body posture.

2 RELATED WORK

2.1 Automated Assessment of Lift Injury Risk

Much automated lift assessment focuses on automating computation of RWL and LI using wearable or RGB-D sensors. Spector et al. [18] use the Microsoft Kinect RGB-D sensor to automatically compute the input parameters of the RNLE from skeletons estimated from Kinect data in order to circumvent manual measurement of the parameters. While horizontal, vertical, and distance measurements are fairly straightforward to acquire from the skeleton, their work does not clarify how hand-to-object coupling quality is derived, which has a complex breakdown according to the RNLE application manual [22]. Given the complexity of directly estimating the RNLE input parameters, several approaches instead use machine learning to directly predict RWL and LI using input from wearable sensors or motion capture. Ground truth values of RWL and LI are computed by manual measurement of the RNLE input parameters, or using motion capture [19, 20] with the coupling parameter set to 'good'. Varrecchia et al. [20] use surface electromyography (sEMG) sensors installed on the trunk for 20 subjects to estimate the LI using neural networks. They improve accuracy in a later study [19] by using motion capture data for the same subjects to estimate energy parameters fed as additional input to the networks. Wang et al. [21] use motion history images of a subject's silhouette to detect lifting, and hand and feet location, and to use them in RWL prediction.

The work of Snyder et al. [17] uses deep convolutional neural networks (CNNs) to classify risk of injury in terms of the LI as low, medium, and high from accelerometer and gyroscope data using a NIOSH dataset [1] of 10 subjects fitted with 6 wearable inertial measurement units (IMUs) for whom risk is pre-calculated using the RNLE as part of the ground truth. Donisi et al. [5] provide a similar approach to detect LI by evaluating a variety of off-the-shelf machine learning (ML) algorithms using data collected in-house from a single IMU installed on the lower back for seven subjects. While these approaches provide an estimate of risk, they do not factor in the subject's perceptions of the task. Additionally, for those approaches that use automated estimation of RNLE input parameters [18–20], the hand-to-object coupling quality is typically manually set to a single value, whereas for real-world tasks, the coupling factor has a far more complex dependence on object size, regularity of object geometry, presence of cutouts or handles, location of the object to facilitate grasp, and quality of grasp [22].

Figure 1: Subject 20 performing a lift from the west, north, and east views from the Point Grey cameras.

2.2 Automated Object Weight Prediction

Very few approaches exist for automated weight prediction upon lift. Palinko et al. [16] provide classifiers that enable a robot to estimate the weight of a tabletop object using data from the robot's force and torque sensors, as the robot attempts lift. Their work cannot be directly ported to at-a-distance assessment of humans performing lift. While force sensors may be installed on the object, propagation to everyday objects proves unscalable.

Lastrico et al. [11] use deep learning to estimate carefulness and weight from motion capture and optical flow data of users handing over four transparent glasses to a robot. Two glasses are weighted down with coins and screws. One weighted and non-weighted glass is filled with water to the brim inducing subjects to move these glasses with care. The authors find high success at detecting carefulness, but limited success with weight prediction. In their work, users have visual access to the glass contents. Our work predicts perceptual and physical parameters of lift with opaque containers where subjects do not have access to the box contents, and lack knowledge of the weight in one lift session.

3 DATA COLLECTION

3.1 Recording Environment

We performed data collection using our in-house capture environment consisting of a circular space 22 feet in diameter. We used three FLIR Point Grey BlackFly S cameras, each of which captures 8-bit color with a spatial resolution of 1440×1080 and a maximum

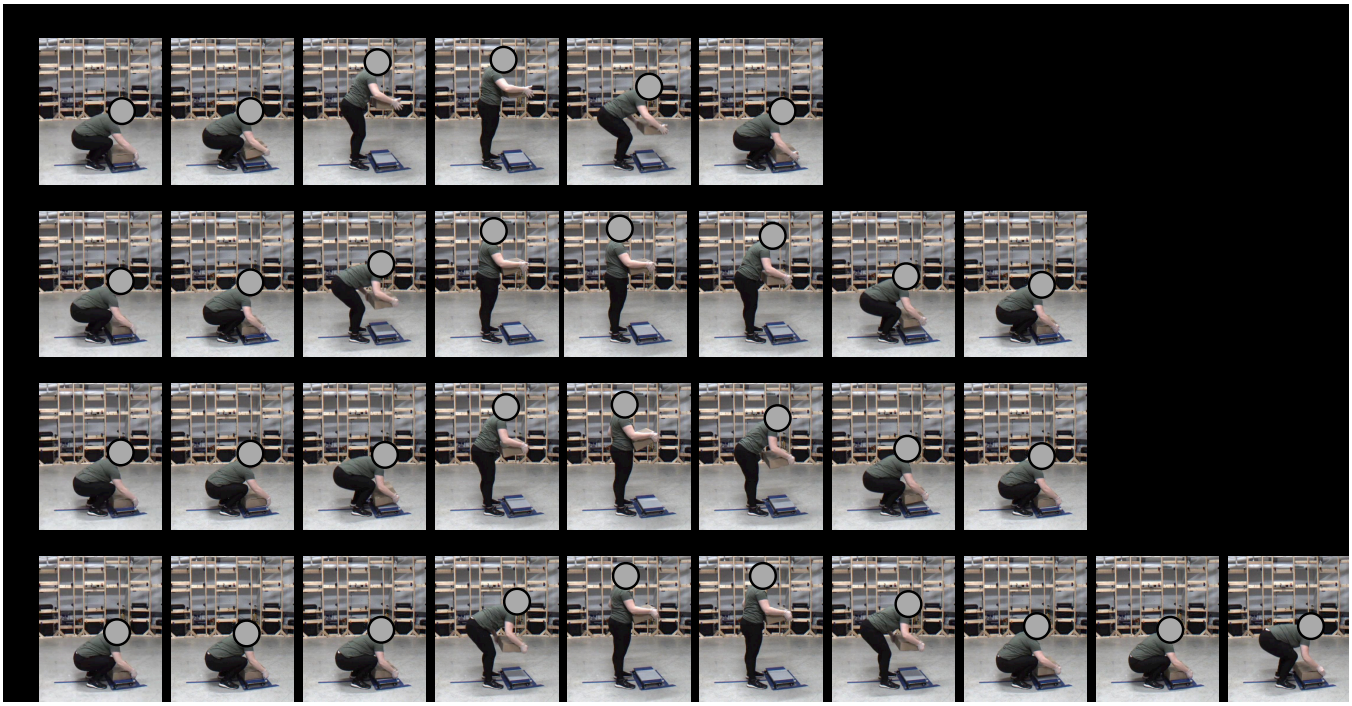

Figure 2: Full lifts for Subject 24 attempting weight lift with 0 lb, 15 lb, 30 lb, and 45 lb weights without weight knowledge. 0 second indicates first hand touch marking start of action. Last visible frame for the filmstrip indicates first hand release marking end of action. Black boxes indicate no frame data due to end of interaction for that weight class.

temporal resolution of 226 frames per second (FPS). We lowered the capture temporal resolution to 70 FPS as the Point Grey records at low exposure for higher frame rates. The cameras were arranged along cardinal directions of north, east and west representing the front, right, and left views of a subject respectively. The high speed of 70 FPS in contrast to conventional 30 FPS camera rates enables the Point Grey cameras to acquire fine-grained information on time taken for initial lift-off from the ground, accelerations as the person lifts the object up, and detail on struggle during lift performance. The three cameras were each connected to a corresponding custom-built Intel Core i5-10600K powered computer, with 32GB of RAM, and 8TB of SSD storage. The three computers were controlled from a fourth computer that acts as a central control. We used an in-house C# command line application for data capture and extraction. We used an overhead green light strobe programmed to flash twice for synchronization across multiple computers. We synchronized all cameras to the second falling edge of the light strobe by detecting peaks in absolute differences between adjacent frames for static patches on the ground. Figure 1 shows example high-speed color images as a subject performs a lifting action.

3.2 Recruitment of Subjects

We performed recruitment through an email to students, faculty, and staff at Clarkson University, United States. The study was approved by the Institutional Review Board (IRB) under protocol number 21-03, and all methods were performed in accordance with these guidelines. All research personnel on the study went through the CITI Human Subjects Research training. We recruited 28 subjects. Subjects visited the capture environment on two days to perform lifts with objects of varying weights. On the control or 'non-blind' day, we informed each subject of the weight they were lifting at each weight presentation. On the experimental or 'blind' day, subjects did not receive this information, to capture interaction without prior knowledge of the weight. Each capture day took 90 minutes per subject. Non-blind and blind day assignments were randomized across subjects. We obtained informed consent from all subjects and provided a $25 monetary incentive upon completion. Subjects completed a demographics questionnaire on age, gender, height, weight, physical fitness regimen, and handedness.

3.3 Study Method

Upon providing informed consent, each subject was told that they would interact with four weights, 0 lb, 15 lb, 30 lb, and 45 lb, with each weight being presented to them 10 times, for a total of 40 weight presentations. We randomized the presentation orders to each subject, and varied randomizations across subjects. We presented weights as dumbbells distributed in bags put in a single corrugated cardboard box of dimensions 18"×12"×4" on a dolly. On the blind day, we asked subjects to look away when the weight dolly was moved, and we moved the dolly at an even pressure and rate to prevent subjects being primed with weight knowledge through visual, auditory, or temporal cues during rolling motion. We locked the casters after bringing the dolly to the center of the space. On the non-blind day, we informed the subject of the weight. We informed

the subject to attempt a box lift, hold it for around 2 seconds or as comfortable, and lower the box when done. No other instructions were provided to capture natural lift. After lift performance, we requested the subject to rate strenuousness from 1 to 5 on the Likert scale. For both days, we requested subjects to respond 'Yes' or 'No' to whether they needed assistance during lift. Except subject 30, no subject reported need for assistance. Subject 30 reported need for assistance for 8 45 lb weight presentations on the non-blind day and all 10 45 lb weight presentations on the blind day. On the blind day, we request subjects to provide a weight guess after lift completion. We performed a post-capture validation check over our entire data to remove spurious captures. We eliminated data for 5 out of the 28 subjects due to issues with capture quality, retaining data for 23 subjects. For each lift, we manually identified the start time as the frame corresponding to first hand touch, and the end time as the frame corresponding to first hand release. At 2 days per subject, 40 captures per subject on each day, and 3 viewpoints per capture, we recorded a total of 5,520 high-speed videos.

Figure 2 shows filmstrips for a subject attempting lift with the four weight classes studied in this work. The time instant '0 seconds' corresponds to the manually marked location of the start of the interaction based on first hand touch on the box. All filmstrips end at the manually marked first hand release after lowering. As shown by the figure, the entire lift and lowering process takes longest for the highest weight class of 45 lb at 12.67 seconds, and least time for the lowest weight class of 0 lb at 7.67 seconds. While the time taken for the intermediate classes of 15 lb and 30 lb is similar, with the subject reaching standing pose by around 6.13 seconds, we observe differences in positioning of individual joints prior to the lift which provide clues as to the weight lifted. For instance, during the upward phase of the lift, at 3.07 seconds, the subject's hips are higher for the 15 lb weight than for the 30 lb weight, indicating that the subject's lower body moves upward faster for the lower weight. Since we note lift completion as the time of first hand release, for the 45 lb weight, the subject spends more time in the lowermost posture and retains contact with the weight while beginning return to standing. We notice this for several subjects, presumably attributed to need for carefully lowering and stabilizing the box.

4 POSTURE-BASED PREDICTION

Subjects' posture over time as they interact with objects has the potential to inform how heavy the weight may actually be and how heavy they may perceive the weight by identifying signs of struggle, delay in lifting, or differences in lift rate. We represent subject posture by extracting skeletons from the Point Grey images captured from each viewpoint using the OpenPose [4] library. Figure 3 shows OpenPose skeletons aligned to frames from the east, north, and west viewpoints of a subject. We analyze predictive ability of the posture by using 1D convolutional neural networks (CNNs) on time series data corresponding to the motion of key joint locations during the initial phase of the lift, i.e., during hold and attempt to perform lift-off, in order to assess potential for early intervention. We define the initial phase as the first 200 frames of the Point Grey data, i.e., the first 2.857 seconds starting from point of first hand touch manually marked as discussed in Section 3. Our use of 1D CNNs is motivated by recent observations of their success for time series [8, 10], with higher accuracies demonstrated over traditional time series architectures, e.g., recurrent neural networks [10].

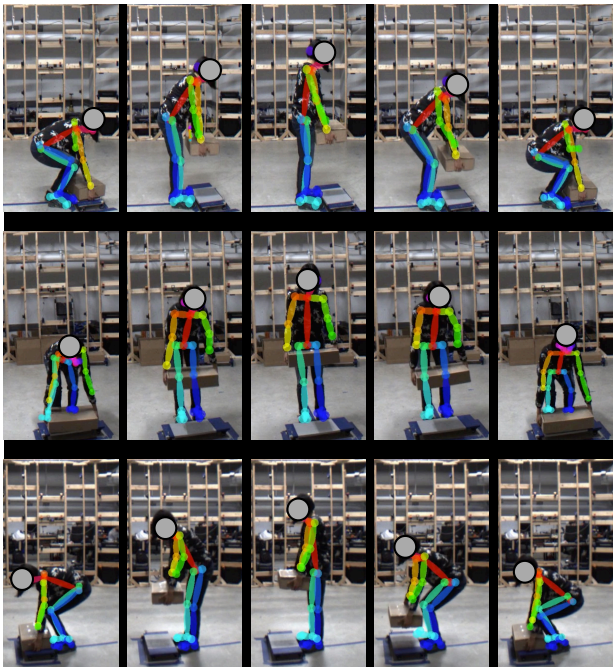

Figure 3: OpenPose skeleton aligned to Subject 30 for east, north, and west viewpoints.

We use the ankle, knee, hip, shoulder, and wrist joints on the right side of the subject as key joints from the east viewpoint. We choose the same joints on the left side of the subject from the west viewpoint. We select the torso and left and right shoulder locations from the north viewpoint that captures the front of the subject. With 6 joints each from the east and west viewpoints, and 3 from the north viewpoint, we have 15 joint locations across all views.

Figure 4 shows filmstrips of Subject 24 during the first 3 seconds of lift for the four weight classes. The figure reaffirms the observations in Figure 2 that lift-off from the ground occurs sooner, and upward movement occurs at a faster rate with lower weights. Subtle differences are seen in the 30 lb and 45 lb filmstrips, where toward the end of the 3 second period, the hip and knees start moving upward for the 30 lb weight, whereas they remain closer to the ground for the 45 lb weight. Figure 5 shows filmstrips of 0 lb and 45 lb lifts for Subject 24 on blind and non-blind days. For the 0 lb weight, we notice that the slope of the motion trajectory for the right wrist is sharper for the blind lift than for the non-blind lift. As shown by the inset, the hand location changes from a 'safe' under-package positioning to a positioning at the side of the package. These features suggest that the subject may have been prepared for a higher weight class, and then been surprised by the empty package, resulting in sudden faster upward movement and hand posture adjustment. For the non-blind lift, the lower slope and non-changing hand posture suggest controlled motion, i.e., that the subject may have been prepared *a priori* for the 0 lb weight. For the 45 lb weight, lift-off from the ground occurs later for the

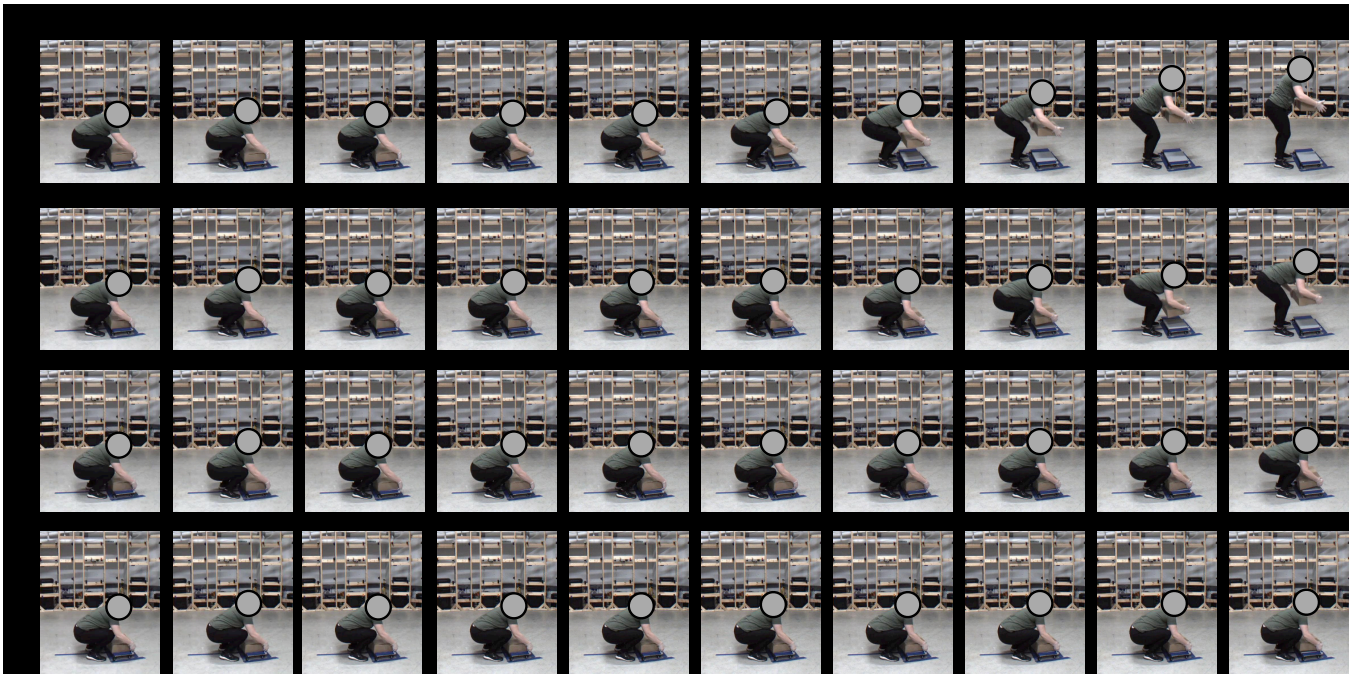

Figure 4: First 3 seconds of lift for Subject 24 for 0 lb, 15 lb, 30 lb, and 45 lb weights without weight knowledge.

Figure 5: First few seconds of lift for Subject 24 for 0 lb and 45 lb weights with and without weight knowledge (red line shows manually marked trajectory of wrist).

blind lift than for the non-blind lift, indicating that the subject may have been prepared for an intermediate weight during the blind lift, and may have needed to adjust posture and timing for the higher weight. We expect the 1D CNNs to learn to distinguish trajectory

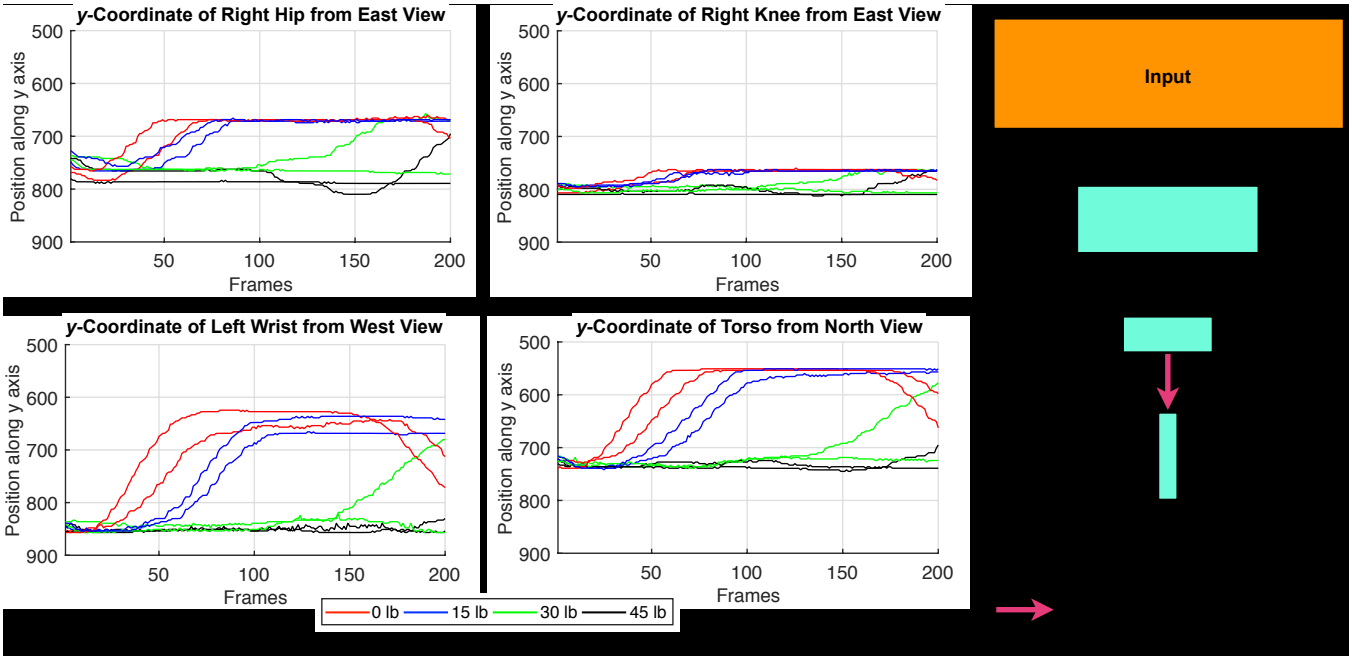

Figure 6: Plots of *y*-coordinates for a selection of joints extracted from each camera viewpoint for the 4 weight classes. (b) 1D CNN for joints-based prediction of weight, strenuousness, and weight knowledge.

features pertaining to differences in lift-off and upward motion across weights and blind/non-blind days.

Figure 6(a) shows trajectories for the *y*-coordinates of the right hip and knee, left wrist, and torso taken from the east, west and north views for the first 200 frames. As the figure demonstrates, upward motion along the *y*-axis occurs earlier for lighter weights, and later for heavier weights, to the extent that the wrist and torso show minimal motion for the 45 lb weight. During lift, the subject performs a slight dip with their hip before lifting a heavier weight, presumably in order to gain balance during weight lifting. Times to lift for the 0 and 15 lb weights are close to each other, though wrist and torso rises occur faster for the 0 lb weight. While the figure shows *y*-coordinates, in this work, we use both the *x*- and *y*-coordinates. Prior to feeding the trajectories to the network, we center each trajectory about the trajectory mean in order to remove the effect of location and physical differences such as height. With 2 coordinates per joint for 30 joints and 200 frames of data, each sample to the 1D CNN is fed as a matrix of size 30×200.

Figure 6(b) shows our 1D CNN architecture. The 30×200 input is subjected to 1D convolution with size 3 filters, followed by batch normalization, application of a rectified linear unit (ReLU) activation, and size 2 max pooling to yield a feature map of size 8×100. The feature map is similarly subjected to convolution, batch normalization, activation, and max pooling to generate a 4×50 feature map which is then restructured into a dense fully connected layer corresponding to the number of classes. We use the 1D CNN to predict the following outputs.

(1) **4-Class Weight:** We use joint data from the skeletons to assess potential for prediction of the 4 weight classes, i.e., 0 lb, 15 lb, 30 lb, and 45 lb.
(2) **Binary Strenuousness:** Strenuousness ratings from the subjects demonstrate an imbalance with several subjects largely providing low ratings. To mitigate the imbalance, we predict strenuousness as a binary output, i.e., 'Low' if the ratings are 1 or 2, and 'High' if the ratings are 3 to 5.
(3) **3-Class Weight:** We find that some subjects perceive the high weights, i.e., 30 lb and 45 lb to be similar. The 30 lb was perceived as being 45 lb 27 times, while the 45 lb weight was perceived as being 30 lb 7 times. For instance, Subject 21 mis-perceived the 30 lb to be 45 lb 4 out of 10 times. Since finer distinction between weights may be challenging, we perform coarse predictions by classifying weights into 3 classes, as 0 lb in the first class, 15 lb in the second class, and 30 lb and 45 lb in the third class.
(4) **Binary Weight:** For the same reason as (3), we also predict weight as a binary output similar to strenuousness with low representing weights of 0 lb or 15 lb, and high representing the 30 lb and 45 lb weights.
(5) **Subject's Prior Knowledge of Weight:** We train 1D CNNs that predict whether the subject had prior knowledge of weight as a binary output, i.e., 'No' or 'Yes'. Data for no knowledge comes from the blind capture, while data for the presence of knowledge comes from the non-blind capture.

We perform two types of predictions, one where the 1D CNN is agnostic of the subject and one where the 1D CNN has prior awareness of the subject built in. The user-agnostic 1D CNN is trained using data from users that are not present in the test set, and enables creating generalizable predictors that may be propagated to novel subjects. The user-aware 1D CNN is trained by using data from the subjects present in the test set, with care taken to ensure mutual exclusivity of test and training data. User-aware 1D CNNs

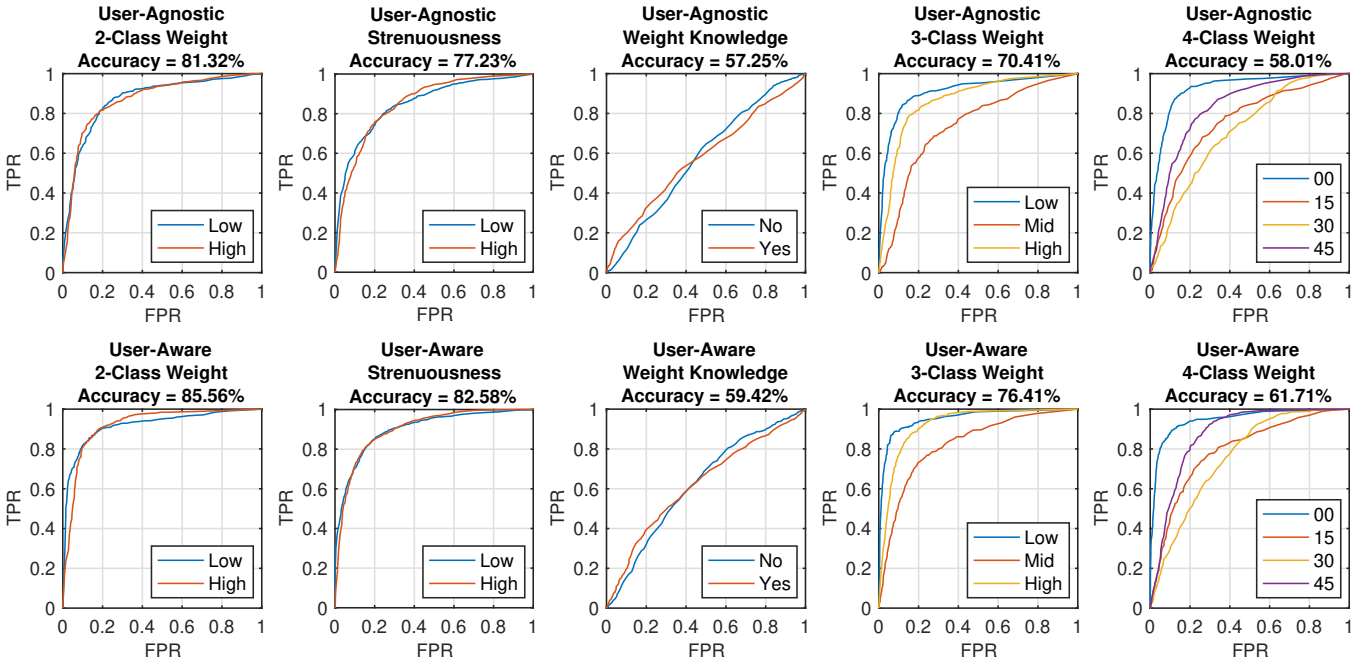

Figure 7: ROC curves for user-agnostic and user-aware prediction of weight classes, strenuousness, and awareness of weight.

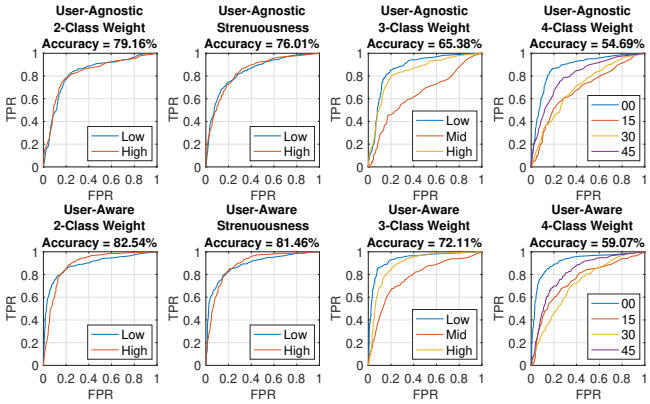

Figure 8: ROC curves for user-agnostic and user-aware prediction of weight classes, strenuousness, and awareness of weight using blind data only.

have value in environments where an AI agent observes a set of individuals over time, e.g., in a manufacturing facility or in an older adults' home, and the agent can be provided pre-deployment data about the individual through a set of trial exercises. Within each type of 1D CNN, we conduct three forms of prediction—one that uses data from captured solely during the blind study i.e. without subject's knowledge of weight, one that uses data captured solely during the non-blind study, and one that combines the data from the two studies in training and in test. For the first two forms, we do not provide prediction of subject's weight knowledge.

We generate results by performing n-fold cross validation, with $n = 4$ folds for user-agnostic CNNs and $n = 2$ folds for user-aware CNNs. With 23 users, we have 6 test users in 3 folds for the user-agnostic CNNs and 5 in the fourth fold. The user-aware CNNs have 12 subjects in one fold and 11 in the other fold. Subject assignments to folds are done at random. We perform ablation testing through exhaustive search over batch sizes of 4, 8, and 16, learning rates of $1e{-}3$ and $1e{-}4$ and 50, 75, and 150 epochs. We report results with maximum accuracy from ablation testing. We obtain accuracy by summarizing the number of times the highest output from the network for an output class matches the class label. We also obtain the receiver operating characteristic (ROC) curve for each output class by treating its class probabilities as positive and the sum of remaining classes probabilities as negative, and obtaining true positive rates (TPRs) and false positive rates (FPRs) for varying thresholds of separation.

5 RESULTS

Figures 7, 8, and 9 summarize ROC curves for each output and class when using combined blind and non-blind or mixed data, blind data only, and non-blind data only. Titles to each plot provide the maximum overall accuracy. In both user-agnostic and user-aware CNNs, highest prediction accuracy is provided by 2-class weight prediction at 81.32% and 85.56% respectively using mixed data. The next highest accuracy is reported by strenuousness detection at 77.23% and 82.58% for user-agnostic and user-aware CNNs using mixed data. The prediction rate for strenuousness is lower than binary weight, with maximum difference at 4.09%. Slightly lower strenuousness detection compared to weight prediction may be attributed to the subjectivity of ratings. With 3-class and 4-class weights, the performance drops to 70.41% and 58.01% for user-agnostic CNNs and 76.41% and 61.71% for user-aware CNNs using mixed data.

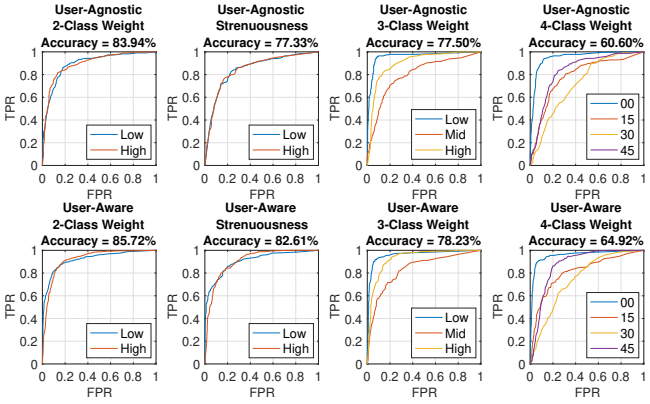

Figure 9: ROC curves for user-agnostic and user-aware prediction of weight classes, strenuousness, and awareness of weight using non-blind data only.

From the ROC curves in Figures 7, 8, and 9, we see that the area under the curve is higher for 0 lb and 45 lb, indicating that elements of delay during lift may help discern if a person is lifting something very light or very heavy. The accuracies for weight knowledge prediction of 57.25% for user-agnostic and 59.42% for user-aware are close to chance, indicating that separation based on joint positions is difficult when using all weight classes together. Results with training using blind and non-blind data separately demonstrates that non-blind data shows higher accuracies than mixed for weight and strenuousness prediction, while blind data shows lower accuracies. This may be due to postural adjustment in response to surprise when the actual weight does not match up with an expected pre-lift weight as, e.g., in Figure 5.

For the purpose of analyzing need for intervention, it is of interest to understand the trade-off between consistent intervention at the risk of inconvenience to the subject or reduction in subject independence versus non-intervention at the risk of safety. Assuming that intervention is needed at a higher weight class and a higher strenuousness level, we observe that at an FPR of 0.3, the TPR for prediction of weight classes that are 30 lb and higher is 0.8669 for 'High' in 2-class weight prediction, 0.8746 for 'High' in 3-class weight prediction, and 0.8246 for 45 lb in 4-class weight when using user-agnostic classifiers with mixed data. TPR at FPR of 0.3 is 0.8291 for strenuousness prediction using mixed data. The high value of TPR for 2- and 3-class prediction suggests that with some compromise on convenience emphasis on safety can be increased.

Figures 10, 11, and 12 show confusion matrices for the various prediction experiments. We observe that in 2-class weight prediction, accuracies are higher for recognizing the lower class consisting of 0 lb and 15 lb weights, than for the higher class, though the order is flipped for user-aware blind prediction. Higher accuracies are also observed for prediction of lower strenuousness. Highest accuracies are observed in the 3-class and 4-class confusion matrices for the 0 lb weight, followed by the highest weight class, i.e., the 30 lb/45 lb class in the 3-class prediction and the 45 lb class in the 4-class prediction. Confusion matrices for 3-class weight prediction demonstrates the obvious expectation of the intermediate (15 lb) class showing low accuracy. Confounding is higher on blind days than on non-blind days, and tends to occur with the highest class with 30 lb and 45 lb weight. Interestingly, when the graininess is increased to 4 classes, the 15 lb weight confounds more with the lower weight of 0 lb than with higher weights.

Tables 1 and 2 show per-subject per-class accuracies for user-agnostic and user-aware classification using mixed, blind, and non-blind data respectively for 2-class and 4-class predictions. We omit the 3-class predictions in the interest of space. 3-class per-subject summaries mirror those of 4-class predictions. 'NaN' denotes cases where no data was available for the subject, e.g., if the subject consistently provided low ratings for every lift. For most subjects, the lowest weight class shows highest accuracy, similar to the trends observed in overall summaries from the ROC curves and confusion matrices. Results for using blind and non-blind data show similar trends, with lower values for blind and higher for non-blind. In binary weight classification, Subjects 16, 21, and 31 show amongst the highest accuracies. In strenuousness, we see highest average accuracies for Subjects 24 and 26. Subjects 14 and 29 perform worst for 'High' and 'Low' detection respectively for user-agnostic CNN. Figure 13 shows a comparison of the trajectories of the $y$coordinates for the hip and wrist of two 45 lb lifts of Subject 14 compared to the hip and wrist y-coordinates of 45 lb lifts for well-performing subjects 31 and 16. We observe that Subject 14 starts their 45 lb lift early. Subject 16 ranks both lifts as 3, indicating that the strenuousness may be intermediate rather than high. The figure also shows trajectories for two 0 lb lifts of subject 29 compared to the 0 lb lifts of Subjects 31 and 16. Subject 29 appears to lift their hips and wrists later than the other two subjects. Subject 29 also performs a gentler raise, which appears to be more characteristic of heavy objects for most subjects. Our classifier uses the general trend of most subjects lifting heavier objects later to incorrectly predict low weight for ground-truth heavy weights of Subject 14, and high weight for ground-truth lighter weights of Subject 29. In user-aware detection, strenuousness values are generally higher. Trends of user-aware 2-class weight detection are similar to user-agnostic detection. Multi-class weight predictions per user reflect findings in the confusion matrices in Figures 10 to 12 where most subjects show confounding of higher weights.

6 DISCUSSION

In this work, we have presented the first approach to understand subject perceptions while lifting objects with the goal of informing ubiquitous monitoring systems on non-verbal cues related to subject experiences with lift. Unlike prior work that has focused on tabletop objects, our work focuses on large weight categories in containers of volumes typical in lift and carry operations. We demonstrate results of predicting weight, strenuousness, and prior knowledge of weight from parameters such as joint locations at the start of the lift. Our work analyzes two-handed lifts which tend to be common for standardized rectilinear containers. We are interested in expanding the scope of objects carried to include diverse geometries, aspect ratios, mass distributions, and flexibilities lifted from varying heights and with various hand configurations.

While current accuracies are low, especially for fine-grained weight prediction, continued progress on this work is essential due

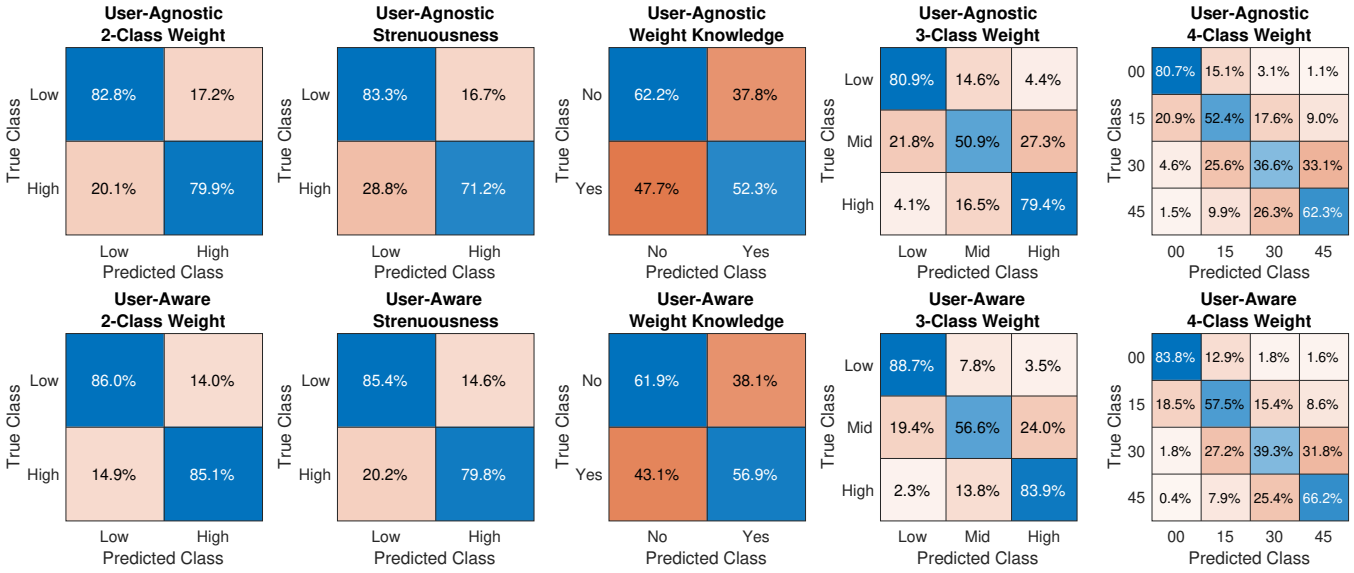

Figure 10: Confusion matrices to predict weight, strenuousness, and weight knowledge.

ID	2-Class Weight		2-Class Stren.		Weight Knowl.		4-Class Weight				2-Class Weight		2-Class Stren.		4-Class Weight				2-Class Weight		2-Class Stren.		4-Class Weight			
	Lo	Hi	Lo	Hi	No	Yes	0	15	30	45	Lo	Hi	Lo	Hi	0	15	30	45	Lo	Hi	Lo	Hi	0	15	30	45
	Mixed										Blind								Non-Blind							
6	0.95	0.73	0.98	0.53	0.43	0.80	1.00	0.25	0.35	1.00	0.90	0.90	0.95	0.70	1.00	0.30	0.20	0.60	0.95	1.00	0.95	0.50	1.00	0.10	0.60	0.80
7	0.84	0.80	0.95	0.60	0.50	0.87	0.82	0.45	0.35	0.60	0.83	0.90	0.89	0.74	0.75	0.30	0.40	0.80	1.00	0.55	0.91	0.69	1.00	0.60	0.20	0.30
8	0.98	0.58	0.86	NaN	0.95	0.20	0.95	0.90	0.25	0.10	1.00	0.60	0.98	NaN	1.00	0.30	0.10	0.10	0.90	1.00	0.88	NaN	0.90	0.90	0.60	0.40
10	0.63	0.98	0.48	0.95	0.85	0.70	0.50	0.40	0.00	0.95	0.50	0.90	0.55	1.00	0.00	0.00	0.00	1.00	0.80	0.90	0.45	0.85	0.60	0.70	0.10	0.70
11	0.80	0.92	0.90	0.71	0.58	0.38	0.80	0.55	0.75	0.63	0.85	0.90	0.90	1.00	0.90	0.80	0.70	0.20	0.85	0.79	0.90	0.37	0.90	0.70	0.60	0.44
12	0.83	1.00	1.00	0.37	0.68	0.63	0.90	0.65	0.55	0.80	0.90	0.90	1.00	0.09	1.00	0.50	0.60	0.40	0.80	1.00	0.80	0.80	0.90	0.40	0.30	0.70
14	0.98	0.18	0.98	0.22	0.82	0.40	1.00	0.00	0.00	0.10	1.00	0.05	1.00	0.39	1.00	0.00	0.00	0.00	0.95	0.20	1.00	0.14	1.00	0.20	0.00	0.10
15	0.93	0.53	0.71	0.67	0.58	0.50	0.70	0.80	0.40	0.15	1.00	0.15	0.82	1.00	0.10	1.00	0.30	0.10	0.95	0.70	0.66	0.50	0.80	0.60	0.60	0.50
16	0.93	0.97	0.97	0.90	0.18	0.85	0.80	0.90	0.89	0.25	1.00	0.79	1.00	0.68	0.90	0.70	0.33	0.30	0.85	0.95	1.00	0.86	0.70	0.80	0.67	0.90
17	0.93	0.60	0.92	0.30	0.88	0.43	0.90	0.45	0.25	0.20	0.85	0.50	0.83	0.40	0.90	0.00	0.20	0.40	1.00	0.45	0.93	0.30	1.00	0.80	0.20	0.00
18	0.63	1.00	0.98	0.44	0.55	0.41	0.45	0.00	0.60	0.89	0.70	0.95	0.95	0.75	0.90	0.80	0.80	1.00	0.35	1.00	1.00	0.53	0.60	0.20	0.10	1.00
19	0.97	0.38	0.95	0.50	0.60	0.36	1.00	0.21	0.05	0.10	1.00	0.15	1.00	0.33	0.90	0.00	0.00	0.00	0.95	0.55	0.95	0.37	1.00	0.89	0.10	0.40
20	0.70	0.93	0.60	0.96	0.83	0.23	0.85	0.65	0.25	0.95	0.65	0.90	0.50	0.79	0.90	0.40	0.00	0.90	0.95	1.00	0.66	0.91	1.00	1.00	0.10	0.80
21	1.00	0.90	0.98	0.74	0.70	0.53	1.00	0.55	0.60	0.85	1.00	1.00	1.00	0.84	1.00	0.90	0.80	1.00	1.00	0.80	1.00	0.65	1.00	0.40	0.30	0.70
22	0.68	0.95	0.71	0.90	0.37	0.55	1.00	0.84	0.35	0.85	0.83	0.70	0.68	0.81	0.78	0.56	0.40	0.90	0.70	1.00	0.85	0.71	0.90	0.70	0.60	1.00
24	0.87	0.95	0.95	0.86	0.93	0.33	0.58	0.70	0.85	0.75	0.80	1.00	0.72	0.91	0.90	0.70	0.80	0.70	1.00	0.80	0.97	0.80	0.89	1.00	0.60	1.00
25	0.98	0.63	0.91	0.76	0.25	0.80	1.00	0.40	0.25	0.25	1.00	0.30	1.00	0.33	1.00	0.20	0.10	0.30	0.95	0.75	0.92	0.81	1.00	0.50	0.10	0.40
26	0.89	0.56	0.80	1.00	0.33	0.82	0.60	0.89	0.37	0.35	0.90	0.89	0.63	1.00	0.50	1.00	0.67	0.40	0.83	0.90	0.89	NaN	0.80	0.63	0.50	0.30
28	0.79	0.95	1.00	0.88	0.75	0.30	0.84	0.60	0.50	0.94	0.70	1.00	1.00	0.73	0.90	0.80	0.60	1.00	0.89	1.00	1.00	0.89	0.78	0.70	0.80	0.75
29	0.26	0.97	0.76	0.94	0.75	0.47	0.26	0.00	0.00	0.85	0.20	0.95	0.73	0.90	0.30	0.00	0.10	0.50	0.37	0.74	0.86	0.88	0.44	0.00	0.00	0.90
30	0.78	0.97	0.52	0.88	0.43	0.67	0.85	0.60	0.32	1.00	0.65	1.00	0.55	1.00	1.00	0.40	0.40	1.00	0.95	0.95	0.68	1.00	0.80	0.70	0.00	0.70
31	0.92	1.00	0.65	0.94	0.79	0.14	1.00	0.56	0.32	1.00	0.94	1.00	0.50	0.85	0.78	0.78	0.00	0.90	0.89	1.00	0.84	1.00	0.89	0.89	0.11	1.00
32	0.80	0.93	0.81	0.82	0.60	0.65	0.75	0.75	0.20	0.80	0.80	0.95	0.91	0.83	0.60	0.60	0.30	1.00	0.90	0.85	1.00	0.60	0.80	0.80	0.10	0.80

Table 1: Per-subject accuracies for user-agnostic predictions ('Lo'='Low', 'Hi'='High', 'Stren.'='Strenuousness', 'Knowl.'='Knowledge'). We exclude 3-class weight summaries as their results are similar to 4-class weight summaries.

to its impact in mitigating injury in workplace and home environments. As part of future work, we are interested in adding more joints, e.g., the head and the elbow, and in evaluating prediction with various viewpoint/joint combinations to determine the relative

ID	2-Class Weight		2-Class Stren.		Weight Knowl.		4-Class Weight				2-Class Weight		2-Class Stren.		4-Class Weight				2-Class Weight		2-Class Stren.		4-Class Weight			
	Lo	Hi	Lo	Hi	No	Yes	0	15	30	45	Lo	Hi	Lo	Hi	0	15	30	45	Lo	Hi	Lo	Hi	0	15	30	45
	Mixed										Blind								Non-Blind							
6	0.93	0.73	0.98	0.80	0.80	0.48	1.00	0.30	0.55	0.90	0.90	0.65	0.90	0.85	1.00	0.40	0.10	0.40	0.95	0.90	0.95	0.75	0.90	0.30	0.80	0.80
7	0.89	0.75	0.93	0.83	0.47	0.82	0.76	0.60	0.30	0.65	0.89	0.90	0.74	1.00	0.75	0.60	0.50	0.80	0.89	0.70	0.91	0.81	1.00	0.50	0.30	0.70
8	0.95	0.75	0.94	NaN	0.70	0.60	0.95	0.50	0.45	0.10	1.00	0.80	0.93	NaN	1.00	0.70	0.40	0.30	0.90	0.90	0.90	NaN	0.80	0.70	0.60	0.60
10	0.63	0.90	0.68	0.95	0.90	0.78	0.40	0.30	0.05	0.95	0.55	0.95	0.65	1.00	0.30	0.30	0.00	0.90	0.85	0.95	0.65	0.90	0.70	0.60	0.00	0.60
11	1.00	0.85	0.95	0.76	0.68	0.56	1.00	0.65	0.25	0.37	1.00	0.95	0.95	0.68	1.00	1.00	0.80	0.50	0.95	0.58	0.75	0.63	1.00	0.60	0.40	0.67
12	0.88	0.90	0.95	0.58	0.65	0.60	1.00	0.60	0.60	0.65	0.85	0.85	1.00	0.52	0.90	0.50	0.90	0.60	0.85	0.95	0.90	0.95	0.70	0.60	0.40	0.90
14	0.98	0.67	0.98	0.31	0.67	0.45	1.00	0.35	0.21	0.45	1.00	0.79	1.00	0.33	1.00	0.20	0.11	0.10	0.95	0.45	0.96	0.43	1.00	0.40	0.30	0.30
15	0.98	0.63	0.79	0.33	0.73	0.60	0.70	0.75	0.30	0.35	0.80	0.60	0.97	1.00	0.40	0.40	0.40	0.10	0.90	0.75	0.71	0.50	0.90	0.60	0.70	0.70
16	0.98	0.95	1.00	0.85	0.72	0.74	0.85	0.85	0.67	0.90	0.90	0.84	1.00	0.79	0.70	0.90	0.67	0.60	0.85	1.00	0.94	0.86	0.90	0.90	0.67	1.00
17	0.90	0.83	0.80	0.60	0.75	0.58	0.85	0.65	0.45	0.30	0.80	0.60	0.73	0.60	0.80	0.30	0.20	0.30	1.00	0.75	0.90	0.80	1.00	0.80	0.60	0.50
18	0.90	0.95	0.85	0.87	0.68	0.31	0.90	0.80	0.85	0.89	0.80	1.00	0.95	0.65	0.90	0.90	0.50	1.00	0.70	0.95	0.75	0.68	0.80	0.40	0.40	0.89
19	0.85	0.68	0.90	0.58	0.65	0.54	0.95	0.53	0.15	0.55	0.95	0.60	0.84	0.71	0.50	0.10	0.10	0.40	0.84	0.75	0.85	0.79	0.90	0.78	0.30	0.50
20	0.83	0.95	0.67	0.92	0.53	0.43	0.85	0.45	0.25	0.80	0.55	0.80	0.58	0.79	0.80	0.30	0.20	0.70	0.65	1.00	0.59	0.91	0.90	0.50	0.20	0.90
21	0.93	0.88	0.98	0.79	0.48	0.73	0.90	0.70	0.60	0.80	1.00	1.00	1.00	0.79	1.00	1.00	0.80	0.90	0.95	0.75	1.00	0.80	0.90	0.60	0.50	0.60
22	0.92	0.93	0.83	0.97	0.42	0.70	0.84	0.79	0.50	0.90	0.83	0.90	0.91	0.88	1.00	0.44	0.60	0.90	0.95	1.00	0.81	0.86	0.90	0.80	0.60	0.70
24	0.79	0.93	0.86	1.00	0.75	0.46	0.68	0.75	0.70	0.75	0.80	0.95	0.76	0.91	0.80	0.60	0.80	0.80	1.00	0.85	0.90	1.00	0.78	1.00	0.70	0.80
25	0.98	0.75	0.89	0.62	0.48	0.60	1.00	0.75	0.20	0.45	1.00	0.70	0.95	0.56	1.00	0.30	0.50	0.40	0.90	0.85	0.92	0.56	0.90	0.80	0.40	0.40
26	0.92	0.92	0.89	1.00	0.49	0.61	0.85	0.67	0.26	0.15	0.95	0.84	0.87	1.00	0.60	1.00	0.67	0.70	0.89	0.80	0.89	NaN	1.00	0.50	0.50	0.40
28	0.77	0.89	0.92	0.95	0.55	0.32	0.84	0.75	0.55	1.00	0.70	0.95	1.00	1.00	0.90	0.70	0.40	1.00	0.95	1.00	1.00	0.89	1.00	0.90	0.70	1.00
29	0.36	0.92	0.60	0.96	0.53	0.55	0.26	0.00	0.05	0.75	0.20	0.85	0.55	0.93	0.20	0.10	0.00	0.70	0.47	0.89	0.57	1.00	0.44	0.00	0.00	1.00
30	0.70	0.95	0.63	0.92	0.48	0.67	0.75	0.30	0.58	0.90	0.60	0.90	0.72	0.91	0.80	0.40	0.20	0.70	0.80	0.95	0.68	0.93	0.90	0.40	0.11	0.80
31	0.92	0.97	0.72	0.91	0.63	0.38	0.94	0.67	0.11	0.95	0.89	0.80	0.56	0.85	0.78	0.44	0.30	0.90	0.89	1.00	0.80	1.00	0.89	0.44	0.11	0.90
32	0.83	0.95	0.90	0.82	0.53	0.58	0.95	0.55	0.40	0.75	0.85	0.95	0.82	0.78	0.90	0.60	0.30	1.00	0.80	0.85	0.85	0.85	0.80	0.60	0.20	0.80

Table 2: Per-subject accuracies for user-aware predictions (conventions are same as in Table 1).

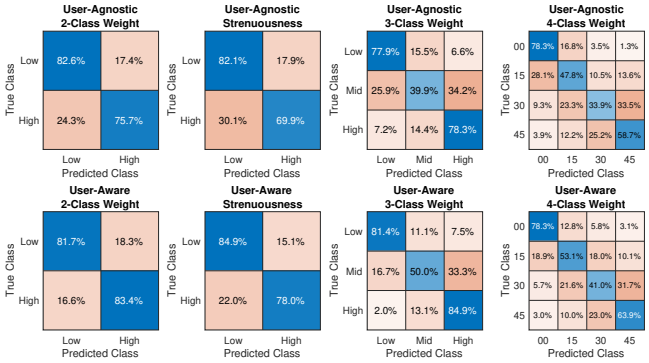

Figure 11: Confusion matrices to predict weight and strenuousness weight knowledge using blind lifts.

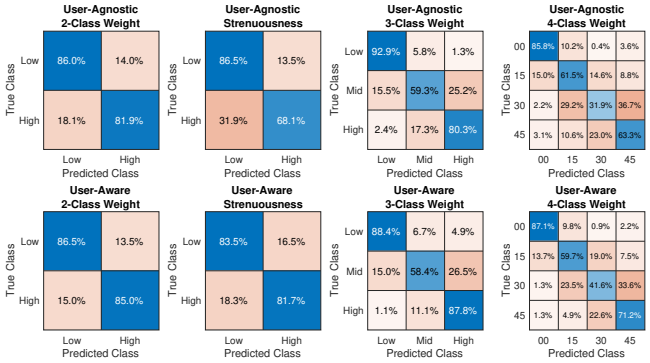

Figure 12: Confusion matrices to predict weight and strenuousness using non-blind lifts.

importance of the left versus right sides of the body, the anterior versus posterior viewpoints, the dorsal versus ventral viewpoints, and discriminative abilities of individual joints or joint groups, and differences versus correlations across joints.

The differences in prediction accuracies for outputs such as weight and strenuousness using blind data and non-blind data alone suggest that there is a potential separation between the blind and non-blind data, indicating that it should be possible to leverage the data collected in this study to perform prediction of weight knowledge. However, our classifiers demonstrate weight knowledge prediction near chance. One reason for the near-chance performance may be that the weight knowledge prediction networks combine the data from multiple weight classes so that the signal representing actual weight or strenuousness, e.g., large-scale changes in joint locations during lift, may overpower the subtle signals related to weight knowledge. Struggle may occur due to lack of perception

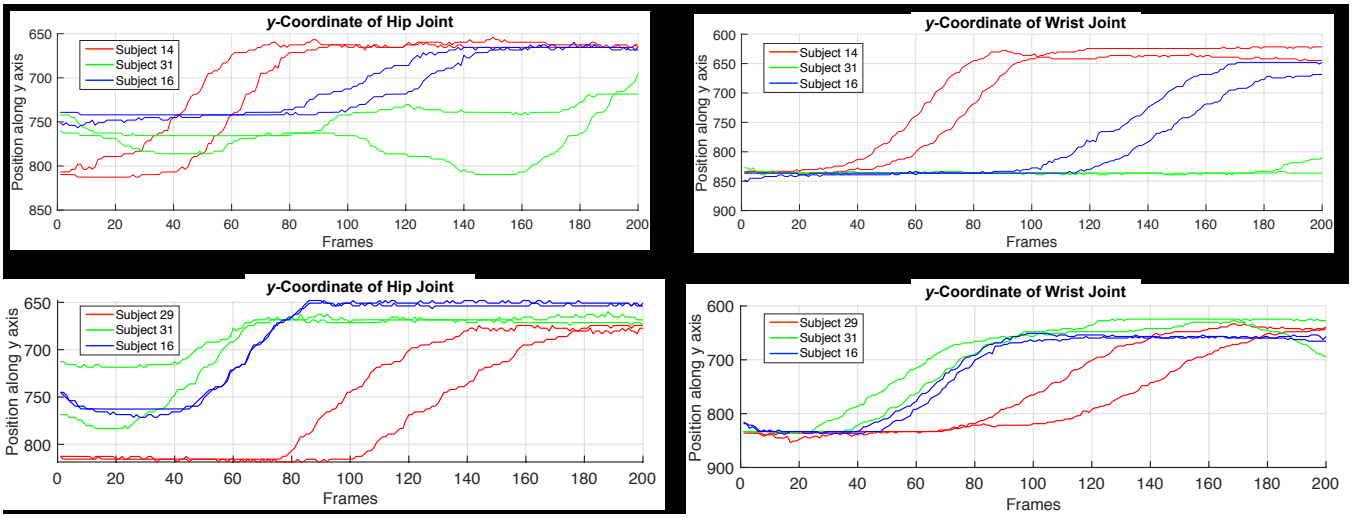

Figure 13: Top: Plots of the *y*-coordinates of hip and wrist for Subject 14 demonstrate earlier lift of 45 lb weight in comparison to well-performing subjects. Bottom: *y*-coordinate hip and wrist plots demonstrate later gentler-paced lifting of 0 lb weight for Subject 29 in comparison to well-performing subjects.

without knowledge or due to mis-perception of weight with knowledge, inducing the signals to appear similar. For future work, we are interested in performing weight knowledge detection within each weight class, rather than by combining all weight classes, as the variations seen in Figure 5 show the potential of using trajectories within each class for weight knowledge prediction.

For the user-aware CNNs, we make the choice of training the CNN with all subjects' data in order to enable the CNN to acquire sufficient information to learn network weights. Using all subjects' data also enables the CNN to be applicable for a wide range of subjects in an organization. Higher accuracy may be achievable by having CNNs be individual-specific over time, which may be applicable for monitoring in an environment with few people, e.g., the home of an older adult. For fine-grained weight prediction, there may be a benefit of performing regression rather than classification, especially if regression provides outputs that are within a weight uncertainty that cannot be perceived by the subject. Accuracies may also be improved if 2D information from multiple viewpoints is fused to generate a single 3D skeleton to yield a reduction in data dimensionality without loss of information, making network training more well-conditioned. While current convention for 3D body-posture analysis is to use depth cameras due to their ability to provide single-camera 3D information acquisition and their generalizability, in this work, we choose to use the Point Grey BlackFly S RGB cameras as we were analyzing the benefit of higher recording speeds on detecting weight and strenuousness. Consumer depth cameras are currently unavailable for temporal resolutions higher than 30 FPS. As part of future work, we are interested in evaluating joint combinations to assess single-camera generalizability, given that some joints will be occluded from some viewpoints during lift. We are also interested in evaluating prediction with reduced frame rates to enable proliferation to lower frame-rate depth cameras and to handle drop in acquisition rates under situations such as altered lighting conditions or simultaneous acquisition and detection.

Important considerations for future work also include performing evaluations from the subject's perspective, e.g., obtaining weight guesses from the subject, and analyzing trends between subject-reported weight and strenuousness, and actual weight. Future work should also garner information on subject preference for continuous intervention while compromising independence versus passive monitoring at the risk of safety. A comprehensive understanding of lift necessitates connecting subject perceptions with knowledge from subject matter experts. In ongoing research, we are working with a physical therapist to assign standardized lift assessment ratings to our subjects' lift performances and relate the assessments to weight carried and subjects' perceptions of lift. We plan to conduct a full-scale study on relationship between lift parameters and subject demographics such as age, gender, and physical fitness level.

Our work evaluates lift perceptions on a per-lift basis, however, work in this domain will have the highest impact if intelligent monitoring systems perform cumulative monitoring over various timescales, ranging from short timescales of a single lift/lowering action, medium scales of a few hours where a subject performs several lift-and-carry operations as part of their work routine, and longer scales over weeks, months, and years. Continuous strenuousness prediction over a single lift/lower action via sliding windows can prove beneficial, e.g., in controlled physical therapy or exercise routines where the therapist/trainer may have the subject perform the lifting activity on their own, and offer assistance at the top of the lift to remove the lowering component. Continuous prediction is also of benefit to assess fatigue during repetitive lift performance, as occurs in warehouse environments. Prediction of long-term parameters such as fatigue may be facilitated by using alternative modalities, e.g., surface electromyography sensors to measure muscle activation during lift and thermal imagers to record temperature changes during the lifting process. Mechanisms of investigation for longer timescales should include interviews related to number of injuries on the job, provision of recommendations to improve

posture based on expert knowledge, and longitudinal studies on continuity in following expert advice. Future work should also leverage multimodal sensing technologies to analyze the relationship between human performance of tasks and perceptions that range beyond the physical properties of the task to concerns about job security in the event of malperformance, and approaches to mitigate such concerns through empathetic feedback for optimal safety and productivity.

ACKNOWLEDGMENTS

This work was funded by National Science Foundation grant IIS-2026559.

REFERENCES

[1] Menekse S Barim, Ming-Lun Lu, Shuo Feng, Grant Hughes, Marie Hayden, and Dwight Werren. 2019. Accuracy of an algorithm using motion data of five wearable IMU sensors for estimating lifting duration and lifting risk factors. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 63. SAGE Publications, Sage, Los Angeles, CA, USA, 1105–1111.

[2] Lee D Cady, David P Bischoff, Eugene R O'Connell, Phillip C Thomas, and John H Allan. 1979. Strength and fitness and subsequent back injuries in firefighters. *Journal of occupational medicine.: official publication of the Industrial Medical Association* 21, 4 (1979), 269–272.

[3] Richard Campbell. 2018. US firefighter injuries on the fireground, 2010–2014. *Fire technology* 54, 2 (2018), 461–477.

[4] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 172–186.

[5] Leandro Donisi, Giuseppe Cesarelli, Armando Coccia, Monica Panigazzi, Edda Maria Capodaglio, and Giovanni D'Addio. 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 (2021), 2593.

[6] Laura P D'Arcy, Yasuko Sasai, and Sally C Stearns. 2012. Do assistive devices, training, and workload affect injury incidence? Prevention efforts by nursing homes and back injuries among nursing assistants. *Journal of advanced nursing* 68, 4 (2012), 836–845.

[7] Ben Evarts and Gary P Stein. 2019. *US fire department profile 2017*. National Fire Protection Association, Quincy, MA, USA.

[8] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery* 33, 4 (2019), 917–963.

[9] Phyllis King, Wendy Huddleston, and Amy R Darragh. 2009. Work-related musculoskeletal disorders and injuries: differences among older and younger occupational and physical therapists. *Journal of occupational rehabilitation* 19, 3 (2009), 274–283.

[10] Yuichiro Koyama, Tyler Vuong, Stefan Uhlich, and Bhiksha Raj. 2020. Exploring the best loss function for DNN-based low-latency speech enhancement with temporal convolutional networks.

[11] Linda Lastrico, Alessandro Carfi, Alessia Vignolo, and Alessandra Sciutti. 2021. Careful with that! Observation of human movements to estimate objects properties. In *Human-Friendly Robotics 2020: 13th International Workshop*. Springer, Springer, Berlin, Germany, 127–141.

[12] K Ngan, S Drebit, S Siow, Seungdo Yu, D Keen, and H Alamgir. 2010. Risks and causes of musculoskeletal injuries among health care workers. *Occupational medicine* 60, 5 (2010), 389–394.

[13] United States Bureau of Labor Statistics. 2019. TABLE R4. Number of nonfatal occupational injuries and illnesses involving days away from work by industry and selected events or exposures leading to injury or illness, private industry, 2019. https://stats.bls.gov/iif/oshwc/osh/case/cd_r4_2019.htm

[14] United States Bureau of Labor Statistics. 2019. TABLE R70. Number of nonfatal occupational injuries and illnesses involving days away from work by event or exposure leading to injury or illness and number of days away from work, and median number of days away from work, private industry, 2019. https://www.bls.gov/iif/oshwc/osh/case/cd_r70_2019.htm

[15] Rosimeire Simprini Padula, Maria Luiza Caires Comper, Emily H Sparer, and Jack T Dennerlein. 2017. Job rotation designed to prevent musculoskeletal disorders and control risk in manufacturing industries: A systematic review. *Applied ergonomics* 58 (2017), 386–397.

[16] Oskar Palinko, Alessandra Sciutti, Francesco Rea, and Giulio Sandini. 2014. Weight-aware robot motion planning for lift-to-pass action. In *Proceedings of the second international conference on Human-agent interaction*. ACM, New York, NY, USA, 193–196.

[17] Kristian Snyder, Brennan Thomas, Ming-Lun Lu, Rashmi Jha, Menekse S Barim, Marie Hayden, and Dwight Werren. 2021. A deep learning approach for lower back-pain risk prediction during manual lifting. *Plos one* 16, 2 (2021), e0247162.

[18] June T Spector, Max Lieblich, Stephen Bao, Kevin McQuade, and Margaret Hughes. 2014. Automation of workplace lifting hazard assessment for musculoskeletal injury prevention. *Annals of occupational and environmental medicine* 26, 1 (2014), 1–8.

[19] Tiwana Varrecchia, Cristiano De Marchis, Francesco Draicchio, Maurizio Schmid, Silvia Conforto, and Alberto Ranavolo. 2020. Lifting activity assessment using kinematic features and neural networks. *Applied Sciences* 10, 6 (2020), 1989.

[20] Tiwana Varrecchia, Cristiano De Marchis, Martina Rinaldi, Francesco Draicchio, Mariano Serrao, Maurizio Schmid, Silvia Conforto, and Alberto Ranavolo. 2018. Lifting activity assessment using surface electromyographic features and neural networks. *International Journal of Industrial Ergonomics* 66 (2018), 1–9.

[21] Xuan Wang, Yu Hen Hu, Ming-Lun Lu, and Robert G Radwin. 2019. The accuracy of a 2D video-based lifting monitor. *Ergonomics* 62, 8 (2019), 1043–1054.

[22] Thomas R Waters, Vern Putz-Anderson, and Arun Garg. 1994. Applications manual for the revised NIOSH lifting equation.

[23] Haiou Yang, Scott Haldeman, Ming-Lun Lu, and Dean Baker. 2016. Low back pain prevalence and related workplace psychosocial risk factors: a study using data from the 2010 National Health Interview Survey. *Journal of manipulative and physiological therapeutics* 39, 7 (2016), 459–472.