National Safety Council

MSD Solutions Lab

an nsc program

Research to Solutions:

Findings from the MSD Solutions Lab Grant Program 2023 - 2024

Introduction

Musculoskeletal disorders (MSDs) affect a quarter of the global population and represent the most common category of workplace injury. Often referred to as ergonomic injuries, MSDs impact the muscles, nerves, tendons, joints, cartilage, and spinal discs, and typically result from common workplace risk factors such as forceful exertions, awkward or static postures, and repetitive movements. When the work environment or job tasks cause, worsen, or extend these conditions, they are classified as work-related MSDs. According to the Liberty Mutual Workplace Safety Index, MSDs cost employers billions of dollars annually in lost productivity, worker compensation, absenteeism, presenteeism, turnover, and recruitment challenges. These disorders are also the leading causes of disability, involuntary retirement, and limitations to gainful employment. Moreover, low wage workers and communities of color disproportionately occupy jobs with greater MSD risk where they often have little control over their work environment. In addition to being a good business investment, MSD risk reduction efforts contribute to a more equitable workplace.

To address this critical issue, the MSD Solutions Lab was established in 2021 as a strategic initiative with the goal of engaging key stakeholders, conducting research, identifying new technology, innovating solutions, and scaling the results so all workplaces can benefit. The lab operates under four pillars:

- **Engage:** Meaningfully engage employers, workers, researchers, and innovators.
- **Research:** Conduct impactful, practical research, analyze data, and disseminate insights across industries.
- **Solve:** Identify, pilot, scale, and promote unique solutions.
- **Amplify:** Create a global effort to engage operations and safety leaders across all industries.

The lab has many activities across all four pillars, including the Research to Solutions (R2S) and MSD Solutions Pilot grants, which are key aspects of the Solve pillar. These grant programs are designed to increase collaboration among academic institutions, businesses, industries, and emerging technology providers to mitigate MSDs. The goal of the R2S grant program is to develop, evaluate, and/or disseminate effective solutions for MSDs, focusing on occupational injury risk reduction. The pilot grant program also aims to develop solutions for preventing MSDs by matching organizations with innovative technology providers to trial emerging technologies in real-life applications. The pilot grants are discussed in more detail in the accompanying **pilot grant findings report**.

MSD Solutions Lab Research to Solutions Grant Program Overview

For the 2023 – 2024 R2S grant program, up to $75,000 was awarded to projects focusing on specific priority research areas:

- **Emerging technologies for risk assessment and mitigation:** Examples include, but are not limited to, leveraging computer vision, machine learning, natural language processing, smart sensors, exoskeletons and exosuits, robotics (including collaborative and service), augmented/virtual and mixed reality, digital twins, and automation in addressing issues of musculoskeletal health and disorders.
- **Legacy MSD high-risk jobs or tasks:** Examples include, but are not limited to, solutions to jobs or tasks known to have high MSD risk but for which there is insufficient evidence regarding methods to sufficiently mitigate the risk. These can come from any industry sector, including health care and social assistance; retail trade; manufacturing; transportation and warehousing; and construction.
- **Future of work:** Examples include, but are not limited to, the role of non-traditional work (e.g., hybrid, work-from-anywhere, remote, and gig work), demographic shifts (aging, sex/gender, racial inequalities, obesity, and social determinants of musculoskeletal health and disorders), and the role of COVID-19 on musculoskeletal health and disorders.
- **MSD management systems:** Examples include, but are not limited to, systems that incorporate or could be integrated with MSD solution strategies, emphasize the hierarchy of controls to "design out" hazards and risk using Prevention through Design and integrate ergonomics with process optimization through MODAPTS®, MTM and Lean Six Sigma.
- **Total worker wellbeing:** Examples include, but are not limited to, integrating occupational safety and health interventions and worksite health promotion programs addressing musculoskeletal health and disorders.

Faculty members at academic institutions, graduate students, post-doctoral fellows, and organizations pursuing unique solutions in one or more priority areas were invited to submit proposals. A review committee consisting of internal National Safety Council personnel and volunteers from the MSD Solutions Lab advisory council evaluated each proposal and scored them based on the following criteria:

- **Scientific Importance/Relevance to Priority Area:** Does the proposal address an important scientific, technical, or practical question? Will the potential findings substantially add to understanding the priority area investigated?
- **Significance of Research/Field Knowledge:** Is the project original and innovative? Will the proposal work develop, test, and evaluate a new methodology or solution? Will the proposed work demonstrate a strong understanding of the area of inquiry and the underlying scientific issues? Will the proposed work make a clear case for how the research fits into a larger context of the issue under consideration for the study?
- **Approach:** Are the variables and controls clearly defined for the study design, if relevant? Are correct quantitative/qualitative measures utilized for evaluating potential research outcomes? Are proper data/statistical analyses described, if relevant?
- **Impact of Work:** Does the proposal state the strategies for MSD prevention and solution development clearly so they can be easily implemented? Can claims of the uniqueness of the proposal or additions to the existing solutions be justified? Can this solution be transferable to similar industries?
- **Dissemination of Findings and Budget Logistics:** Does the proposal document sharing findings through virtual MSD Solutions Lab symposiums or presentations at NSC events, research papers, and knowledge transfer documentation? Is the budget justified and itemized appropriately?

Upon completion of the grant, grant recipients were expected to present their findings at the NSC Safety Congress & Expo.

Brief Description of Grant Projects

For the 2023 – 2024 R2S grant program, there were four grant winners:

- **Rutgers University (New Brunswick, NJ, USA):** Awarded $75,000 to implement an automated image captioning system, designed to help employers better identify ergonomic risk factors and real-time solutions.
- **Virginia Tech (Blacksburg, VA, USA):** Awarded $51,000 to implement inexpensive, camera-based marker-less sensors along with machine-learning models to assess worker physical exposures and MSD risks more efficiently, accurately and comprehensively.
- **Iowa State University (Ames, IA, USA):** Awarded $61,000 to develop a predictive model and artificial intelligence (AI)-based ergonomics app for risk assessment and mitigation that enables employers in high-risk industries to understand shoulder MSD risk in different scenarios impacting their workers – with and without an exoskeleton – to make more informed decisions about injury mitigation.
- **University of Waterloo (Waterloo, ON, Canada):** Awarded $37,000 to generate evidence-based guidance on computer vision-based MSD risk assessment technology in the workplace, so employers can better determine the optimal approach and timing for integrating computer vision-based MSD risk assessment tools into their ergonomics programs.

The following sections present an overview of each R2S project and provide more detailed discussion of the problems being solved, project aims, study methods, accomplishments, lessons learned, and publications resulting from the project. This report serves as a resource for safety professionals, offering invaluable insights into the challenges and benefits of emerging technologies for mitigating MSDs. Furthermore, the report can be used to better understand how to leverage these innovative solutions, ultimately improving worker wellbeing.

Research Study 1: Automated Ergonomic Problem and Solution Identification by Explainable Image Captioning – Rutgers University

What's the Problem?

Accurate, reliable ergonomic programs are necessary to maintain a safe working environment and prevent MSDs, and typically involve detecting ergonomic problems, assessing risk levels, and identifying solutions. Historically, ergonomic evaluation programs have relied on manual observation, requiring an on-site ergonomic expert and significant resources. Manual ergonomic processes can also be subjective and experience-dependent, vulnerable to human error and inconsistency. These factors have motivated the development of automated approaches to ergonomic evaluation, including the adoption of AI to conduct risk assessments. While assessment of risk level is automated in some cases, manual analysis is still often required for identifying ergonomic problems and corresponding solutions. Automating the identification of ergonomic problems and solutions has the potential to increase access to and allow broad application of ergonomic programs without the high cost and manpower, thereby mitigating ergonomic risks.

Project Aims

The goal of this research was to automate steps in the ergonomic process, specifically identifying ergonomic problems and solutions. By leveraging image captioning technology and applying techniques to current AI models, the research team sought to improve the accuracy of ergonomic problem and solution identification and to visualize the AI models' reasoning in predictions. They also aimed to develop prototype software to demonstrate how an end-user could use the technology.

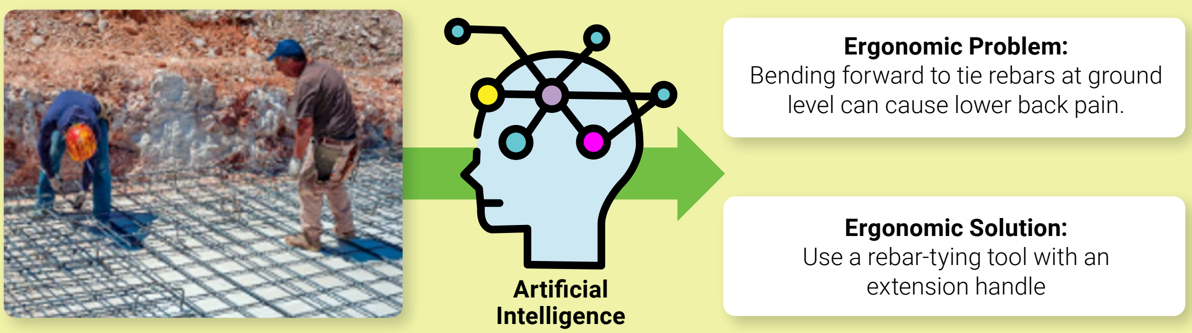


Figure 1. *An image is processed with AI to produce a visual explanation of the ergonomic problem and solution. Image source: © peuceta - stock.adobe.com*

Explanation of Technology and Research Methodology

The research team identified and evaluated the performance of relevant existing image captioning models to understand the causes of incorrect and missing predictions. They hypothesized three techniques to improve the models' accuracy while embedding explainability to enable visual evaluation of the ergonomic problem and solution prediction process. The three techniques are — visualize attention and cohesiveness, visualize pose-awareness, and embed knowledge (Figure 2).

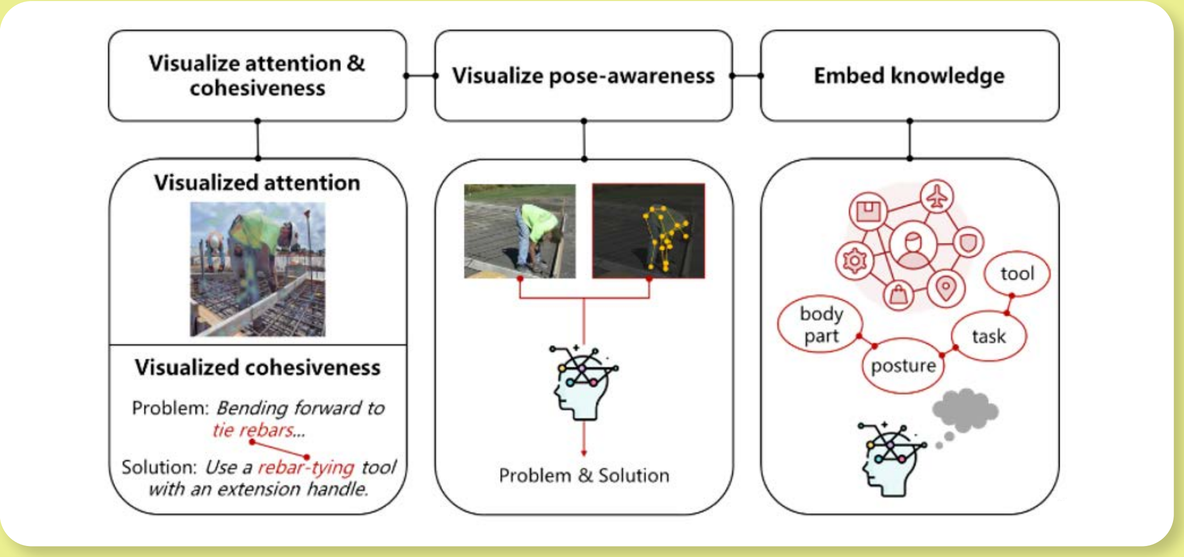


Figure 2. *Chart illustrating techniques to improve ergonomic problem and solution prediction accuracy and visualize explanations. Image source: © tcpalm.com/@Weit;* ***youtube.com/@MikeDayConcrete***

Visualize attention refers to visualized focus of the AI model on a specific part of the image. The AI model superimposes a heatmap to highlight its attention on the image so developers and users can evaluate if the identified ergonomic problem is predicted from a critical and related area on the image. Cohesiveness refers to whether the solution corresponds to the identified problem, and can be visualized in a graph (Figure 3).

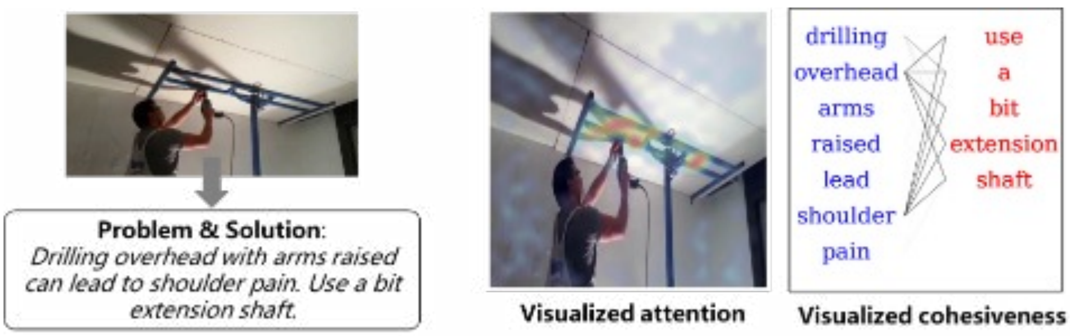


Figure 3. *An image is processed with AI to produce a two-sentence visual explanation of the ergonomic problem and solution. A heatmap visualizes attention on the relevant part of the image and a bipartite graph (image with blue/red text) visualizes the cohesiveness between keywords of the explanation. Image source: ©* ***youtube.com/@andrespenaflor4589***

Visualize pose-awareness refers to the identification and mapping of key body points (Figure 2). Embed knowledge refers to the process of enforcing reasonable association between key elements (e.g., body part, posture, task, and tool) that form the ergonomic problem and solution. Embedding knowledge enables the model to make connections between key elements to finalize predictions.

To develop and train the models, the research team collected workplace images from the internet and with the support of ergonomists, manually annotated each image with a two-sentence caption describing an ergonomic problem and its corresponding solution. To help the model adapt to diverse wording when applied, the training dataset was automatically diversified using a large language model (i.e., ChatGPT) prompted to paraphrase the caption in four different ways, creating a total of five different but synonymous captions for each image. The resulting dataset consisted of approximately 10,000 images and captions and was organized into categories related to ergonomic problems and solutions.

The three techniques were then applied to augment the image captioning models. Specifically, the bootstrapping language-image pre-training (BLIP) model was selected as the backbone due to its performance in image captioning tasks. The accuracy of the developed models was evaluated by ergonomic professionals with new workplace images collected from the internet. Additionally, to evaluate the image captioning models augmented by each technique, the bilingual evaluation understudy (BLEU) metric automatically assessed the quality of machine-generated captions by measuring their similarity to human-generated captions on a scale from 0 to 1.

The three techniques can be implemented simultaneously (Figure 4); however, for the purposes of the study, each of the three techniques was applied to existing models one at a time and tested independently to show their respective impact on accuracy.

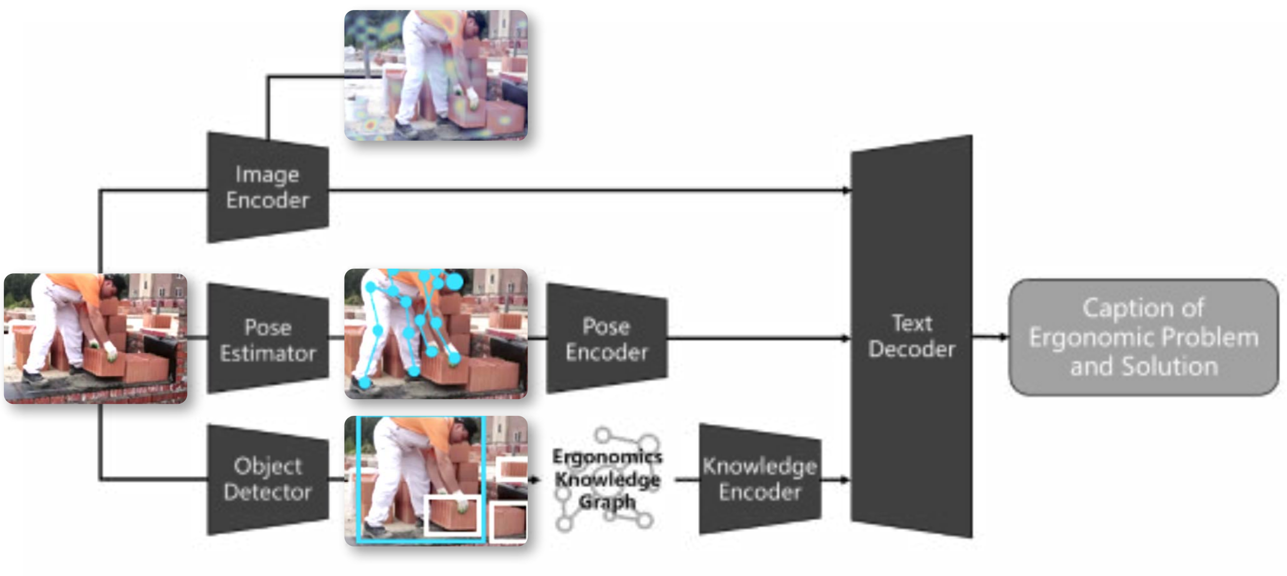
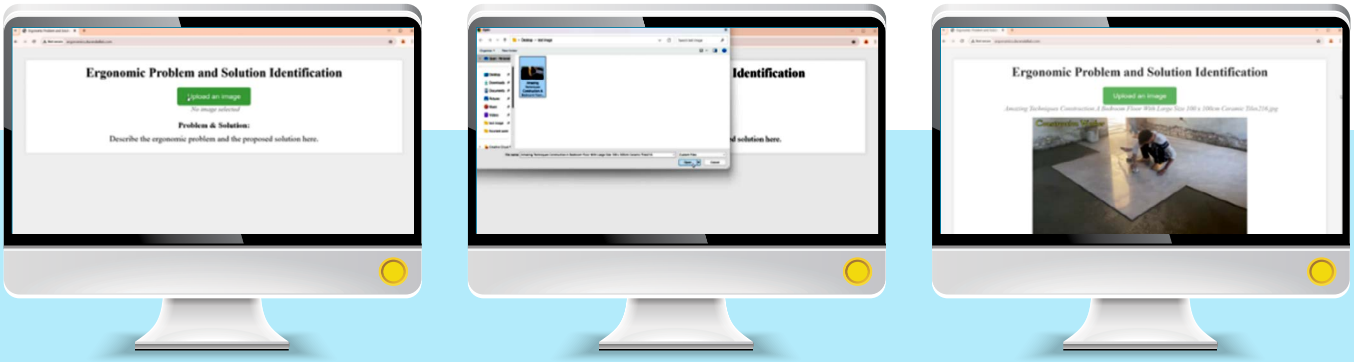


Figure 4. *Framework illustrating how the three techniques can be used to augment an image captioning model simultaneously to produce an ergonomic problem and solution with explanations. Image source: ©* https://www.youtube.com/watch?v=Ak9I3kyxI28

Additionally, a simple web application was developed that allows users to upload an image, process that image by clicking a button, and review the generated two-sentence text of the identified ergonomic problem and solution (Figure 5). The web application is not currently usable for the public. The developed AI model is hosted on a cloud server provided by Amazon Web Services (AWS).



Open the website | **Upload an image** | **Review the result**

Figure 5. *Example showing how to upload an image to the prototype website and the generated result.*

Results and Lessons Learned

The research team developed techniques that can flexibly be applied to existing image captioning models of different architectures. Tests measuring the similarity between ergonomic problems and solutions predicted by the AI models and those that were manually annotated demonstrated that applying each technique independently led to improvements in accuracy (Table 1). Additionally, the feature that enables models to visually explain their reasoning increases transparency and trustworthiness because it allows practitioners to qualitatively evaluate the intermediate information.

Table 1. *Results comparing *Bilingual Evaluation Understudy (BLEU) scores of the baseline image captioning models and the image captioning models with each technique applied.*

Comparative result of visualizing attention and cohesiveness:

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4
CNN-LSTM (baseline 1)	0.799	0.698	0.611	0.558
BLIP (baseline 2)	0.884	0.834	0.794	0.761
New Model (X-BLIPBART)	**0.942**	**0.886**	**0.842**	**0.796**

Comparative result of visualizing pose-awareness:

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4
Baseline (InstructBLIP)	0.8773	0.8270	0.7722	0.7121
New Model (PALIP)	**0.9551**	**0.9308**	**0.8907**	**0.8406**

Comparative result of embedding knowledge:

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4
Baseline (BLIP)	0.6892	0.5890	0.5054	0.4061
New Model	**0.7833**	**0.7011**	**0.6407**	**0.5810**

Key
**The bilingual evaluation understudy (BLEU) metric automatically assesses the quality of machine-generated captions by measuring their similarity to human-generated captions on a scale from 0 to 1.*
CNN-LSTM *- Convolutional Neural Network - Long Short-Term Memory,* ***BLIP*** *- Bootstrapping Language-Image Pre-Training,* ***BART*** *- Bidirectional and Auto-Regressive Transformer,* ***PALIP*** *- Pose-Aware Language Image Pre-training*

Publications and Presentations

- Yong, G., Liu, M., & Lee, S. (2024). Explainable image captioning to identify ergonomic problems and solutions for construction workers. *Journal of Computing in Civil Engineering, 38*(4), 04024022.
- Yong, G.,Liu, M., andLee, S. (2024).Automated Captioning for Ergonomic Problem and Solution Identification in Construction Using a Vision-Language Model and Caption Augmentation.ASCE Construction Institute & Construction Research Congress (CI & CRC) Joint Conference. Des Moines, Iowa.
- Yong, G., Liu, M., and Lee, S. (2024). Vision-based Ergonomic Problem and Solution Identification with Pose-Aware Image Captioning. ASCE International Conference on Computing in Civil Engineering (I3CE 2024). Pittsburgh, Pennsylvania.
- Yong, G., Liu. M., and Lee, S. (2024). Ergonomic Problem and Solution Identification by Applying Image Captioning with Embedded Ergonomic Knowledge. 2024 AHFE International Conference on Human Factors in Design, Engineering, and Computing. Honolulu, Hawaii.

Research Study 2: Evaluation of Markerless Motion Capture to Assess Biomechanical Exposures during Material Handling Tasks – Virginia Tech

What's the Problem?

Measuring and monitoring worker exposures to MSD risk factors in the workplace is critical to understanding occupational exposure limits and designing interventions to reduce risk. While several tools exist for assessing physical exposures in the workplace, they often have limitations in accuracy and scope, especially when capturing dynamic movements, and can be resource-intensive to implement. Technologies such as wearable sensors have addressed some of these limitations by providing precise measurements of risk exposures. However, challenges have limited widespread adoption, such as discomfort from using body-worn sensors for an extended period, incompatibility with personal protective equipment, and the need for user compliance (e.g., proper usage and charging). Developing solutions that can automatically and continuously monitor physical exposures during manual material handling tasks—with minimal reliance on user compliance—could improve the ability to estimate, understand, and control the biomechanical demands that workers face when performing their jobs.

Project Aims

This research project aimed to develop and test methods for automatically classifying manual tasks and estimating associated biomechanical demands using markerless motion capture data, with an emphasis on improving the accuracy and efficiency of occupational physical exposure assessments. The project had three specific aims:

1. Assess the performance of a markerless motion capture system, together with a machine learning algorithm, to continuously classify diverse manual material handling tasks that require complex, dynamic behaviors.
2. Investigate the feasibility of alternative measures such as markerless motion capture and/or in-sole pressure systems to predict the load level, specifically the 3D dynamic hand force, for workers completing tasks.
3. Evaluate whether the kinematic data from aim one and predicted hand force from aim two are sufficient for biomechanical analysis.

Explanation of Technology and Research Methodology

Markerless cameras have ambient sensors that can non-intrusively identify anatomical landmarks (key points on a person's body) and monitor worker postures and motions with vision data alone. To evaluate the capability of markerless motion capture in producing sufficient kinematic data for biomechanical analysis, a variety of manual material handling tasks were simulated by participants recruited from local student and community populations. Whole body kinematics were collected using three markerless cameras (Microsoft Azure Kinect, WA, USA) that captured information from one-handed and two-handed tasks including symmetric and asymmetric box lifting, pushing/pulling, cart pushing, and overhead lifting (Figure 6). For validation purposes, kinematic and in-sole pressure data were also collected with wearable sensors (Noraxon Ultium Motion, AZ, USA). Hand forces were measured and recorded during the simulated tasks from two triaxial load cells (AMTI MC3A-6, MA, USA) attached at the lateral faces near the top of the box, and from load cells on the cart. Participants completed the entire sequence of tasks several times under varying conditions, with differences in box mass, box dimension, grip type, starting position, and lift origin.

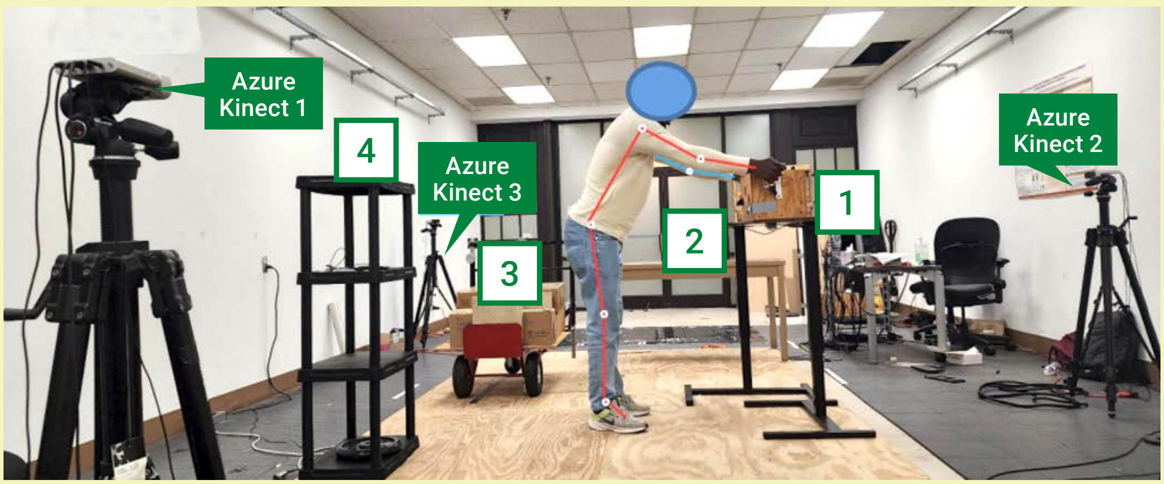


Figure 6. *Image demonstrating the simulated tasks and pin-pointed anatomical landmarks on a person's body. Note that the green boxes indicate the location of the three Azure Kinect systems. The white numbered rectangles indicate the positions of symmetric and asymmetric box lifting (1), pushing/pulling (2), cart pushing (3), and overhead lifting (4).*

Kinematic data collected from the markerless motion capture system were used as inputs to three recurrent neural network models trained to automatically classify tasks. The methods used to develop these models are described in detail in an **article** authored by the study investigators. Performance of the neural network-based models was evaluated based on their accuracy in correctly classifying tasks.

To examine the feasibility of alternative measures to predict 3D dynamic hand forces, kinematic data from the markerless motion capture system and ground reaction forces measured by the in-sole pressure system were used as inputs to three machine learning models trained to estimate external hand force. Model performance was evaluated by comparing predicted hand forces to the actual hand forces measured by load cells.

Additional work evaluated whether the data obtained from markerless motion capture and in-sole pressure measurement systems are sufficient to accurately estimate biomechanical demands. For this, a full-body musculoskeletal model was developed and the estimates derived from markerless motion capture data inputs were compared to estimates derived from gold standard inputs (kinematic data from Noraxon inertial measurement units and the measured hand force).

Results and Lessons Learned

From the first aim, kinematic data from the markerless motion capture system and machine learning algorithms yielded a mean precision of 85 to 97% in classifying diverse material handling tasks and task conditions (such as lift height and hand width). Classification performance (i.e., mean accuracy, precision, recall and F1-score) varied depending on the material handling task, with misclassification most frequent for tasks with comparable kinematics (e.g., box carriage vs. box pulling).

Findings from the second aim indicated that the machine learning models with markerless motion captured kinematics and in-sole pressure system data as inputs performed reasonably in estimating dynamic hand forces. Performance differed between material handling tasks, input feature sets, and box mass conditions, with tasks involving pushing and pulling being the most difficult for the models to estimate accurately. Using data from both markerless motion capture and in-sole pressure systems led to better estimates than using data from only the in-sole pressure system, suggesting that markerless motion capture could increase the accuracy of hand force estimations and predictions.

In the third aim, the musculoskeletal model using kinematic data and predicted hand force data derived from aims one and two yielded biomechanical demand estimates similar to those using gold standard inputs. The magnitude of differences between the estimates varied between manual material handling tasks and mass conditions. There also appeared to be differences between sexes regarding the performance of markerless motion capture and the machine learning models to classify manual tasks and estimate biomechanical demands.

This study demonstrates the potential of using markerless motion capture to automatically and continuously classify several common occupational tasks, and to estimate the associated biomechanical demands on workers. The methods developed in this research could enable stakeholders to rapidly assess an individual worker's exposure to physical demands during diverse manual tasks. The findings also support further investigation, especially to better understand how biological sex may influence the use of markerless motion systems for physical exposure assessments as differences related to sex were found in all three aims. The research team also recommends future research on markerless systems in varied environments and among diverse anthropometric characteristics. Particularly for the latter, markerless systems may struggle to accurately identify anatomical landmarks in individuals with obesity or physical disabilities that alter body kinematics.

Publications

- Ojelade, A., Rajabi, M.S., Kim, S., & Nussbaum, M.A. (2024). A Data-Driven Approach to Classifying Manual Material Handling Tasks Using Markerless Motion Capture and Recurrent Neural Networks. SSRN Electronic Journal.

Research Study 3: Addressing Shoulder Fatigue and Injury Risk: Developing a Predictive Model and AI-based Application for Ergonomics Assessment – Iowa State University

What's the Problem?

Shoulder injuries are a prevalent source of MSDs among workers in the manufacturing, construction, retail and warehousing industries, where work involving repetitive and/or sustained elevated arm postures are common. Guidelines exist to define the maximum level of exertion a worker can endure without experiencing upper limb muscle fatigue; however, their practical application is limited because implementing these guidelines requires direct measurement of muscle activity, typically using body-worn electromyography sensors. AI-based video motion capture that can provide accurate time-series motion profiles of workers is a novel approach to ergonomic assessments that can help overcome this limitation. Ergonomic applications that integrate video motion capture data with models predicting upper limb fatigue can enhance the ability to quickly and accurately assess the risk of upper limb MSDs. These tools can also help identify cases where ergonomic interventions, for example, using an exoskeleton, may be beneficial.

Project Aims

This project aimed to advance the understanding of the relationship between elevated arm work and muscle fatigue, and to develop an AI-based application that objectively quantifies shoulder MSD risk. The following were specific objectives of the project:

1. Develop a predictive model for shoulder MSD risk that maps arm posture to muscle activation, allowing for practical use of a widely referenced standard that predicts muscle fatigue above/below an acceptable threshold value.
2. Validate AI-based video motion capture as an approach to ergonomic assessment and the muscle fatigue predictive model from the first objective in lab and workplace settings.
3. Improve the ability to identify jobs with high shoulder MSD risk and predict whether exoskeleton adoption will be useful.

Explanation of Technology and Research Methodology

The American Conference of Governmental Industrial Hygienists (ACGIH) Upper Limb Localized Fatigue Threshold Limit Value (TLV) is a logarithmic curve that shows the relationship between muscle exertion and muscle fatigue. It is a widely referenced assessment that defines the maximum level of repetitive hand activity or muscle exertion that a worker can be exposed to without experiencing significant localized fatigue in their upper limbs. The researchers developed a predictive model that determines if the upper limb fatigue TLV is exceeded during a job task and integrated it into a prototype mobile application they developed called Shoulder Injury Risk Assessment (SIRA).

Cell phone videos of workers completing job tasks can be uploaded into the SIRA app, where the data are processed using AI, and TLV assessment calculations are performed automatically within the app (Figure 7). Figure 8 provides an example of output from the SIRA app that predicts fatigue in the right shoulder. The TLV and methods to develop SIRA are described in more detail in this **article** authored by the study investigators. The application is not currently available to the public.

3D SIRA Assessment Methodology: Schematic

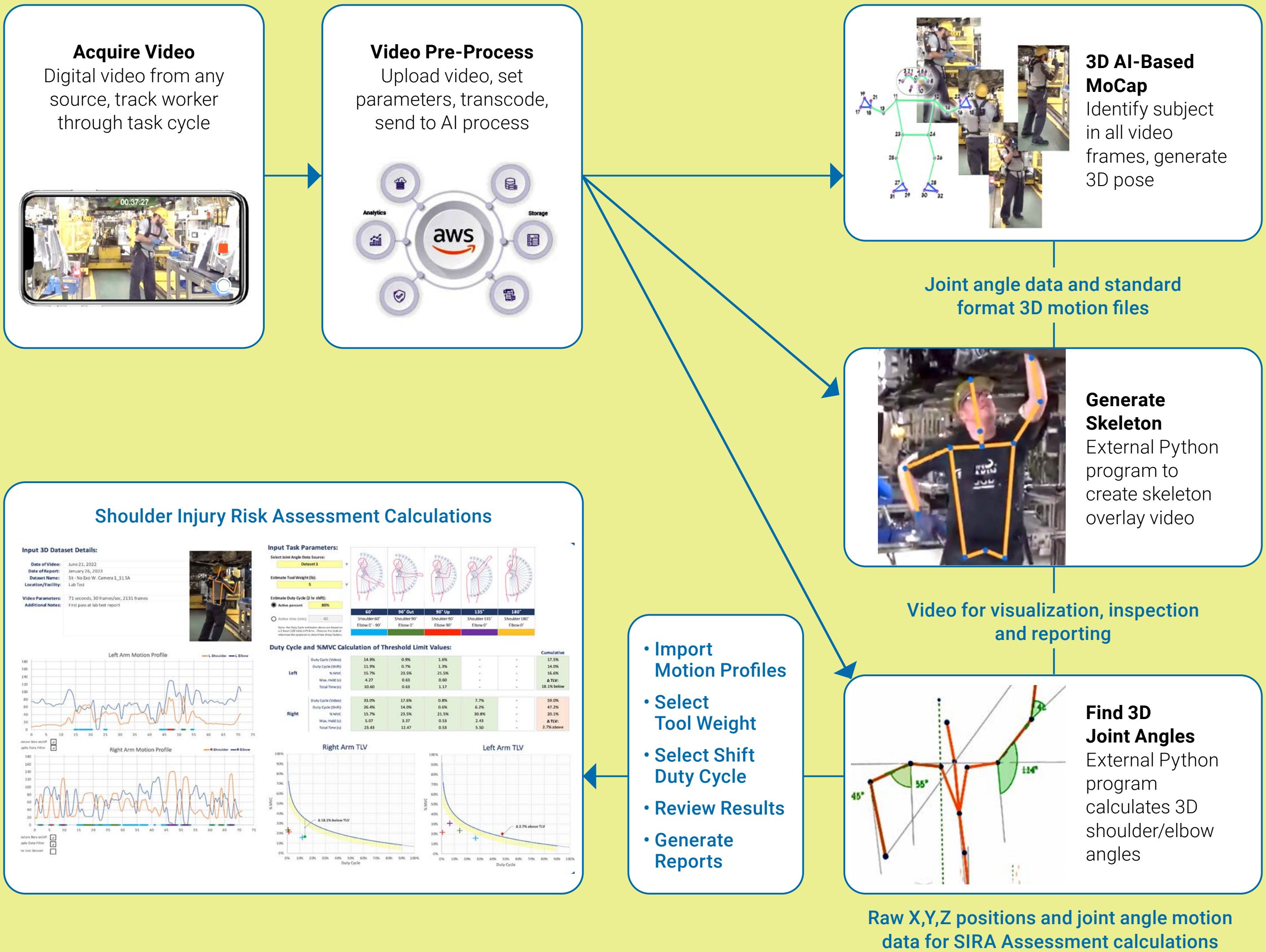


Figure 7. *Schematic illustrating the five phases of AI video processing and SIRA: Acquire the video, upload the video, generate tracking skeletons, calculate 3D angles and perform TLV assessment calculations. Image Source:* [*https://www.assp.org/docs/default-source/psj-articles/f1butleryounggillette_1224.pdf?sfvrsn=1c287a46_0*](https://www.assp.org/docs/default-source/psj-articles/f1butleryounggillette_1224.pdf?sfvrsn=1c287a46_0)

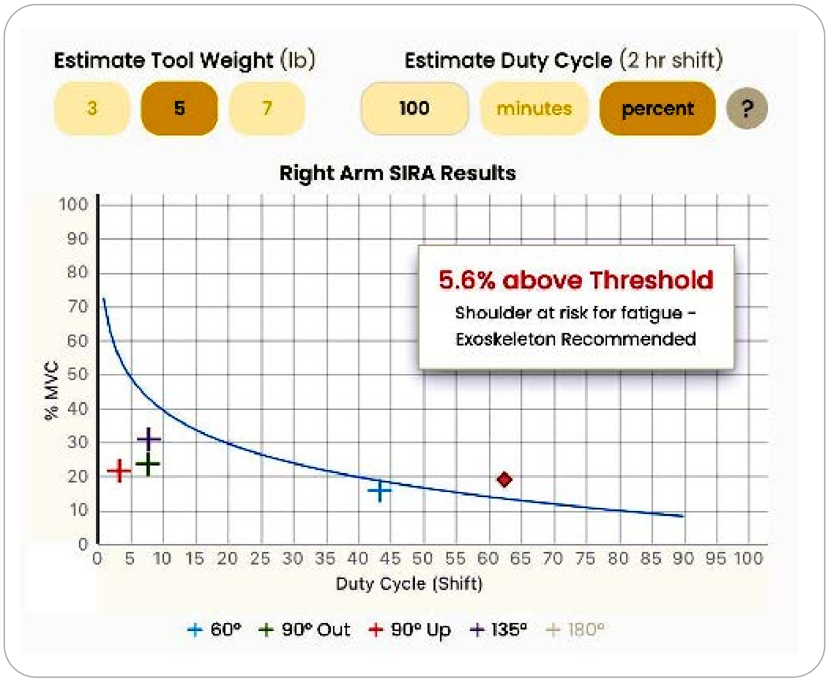


Figure 8. *Example output from the SIRA application for a task involving the use of a 5-lb tool and 100% duty cycle. The crosses represent points on the ACGIH graph for each arm posture and the red diamond represents the cumulative effect of the arm postures relative to the fatigue TLV (blue curve). The cumulative effect of all postures shows that fatigue is predicted in the right shoulder. Image Source:* ***https://www.assp.org/docs/default-source/psj-articles/f1butleryounggillette_1224.pdf?sfvrsn=1c287a46_0***

To test the predictive model and ability of AI-based motion capture, data were collected from workers completing tasks involving elevated shoulder postures at a construction worksite and in a lab setting. For the construction worksite, researchers collaborated with Weitz Construction (Des Moines, IA) to complete assessments among 22 participants at a Microsoft data center worksite. Data were collected while they completed their jobs with and without an exoskeleton, including shoulder muscle activity, perception of exertion/fatigue, and exoskeleton usability. Videos from two cell phones were uploaded to the SIRA app.

For the lab assessments, a protocol was designed to test the accuracy and validity of the AI-based motion capture and SIRA app predictions. Twenty participants performed simulated standing and seated assembly tasks, with and without an exoskeleton, and the data collected included shoulder muscle activity, multi-camera video motion capture, perception of exertion/fatigue, and exoskeleton usability. Videos from two cell phones were uploaded to the SIRA app. The shoulder angles calculated from AI-processed cell phone video data were compared to the angles calculated from lab-based multi-camera motion capture data to test for accuracy. The fatigue predictions from the SIRA app were compared to muscle activity measured by surface electromyography data as a validation measure.

Results and Lessons Learned

When it came to the AI video motion capture, there were challenges collecting suitable data due to obstructed views and the presence of multiple workers in the frame. Figure 9 shows an example of preliminary accuracy testing, comparing the detected shoulder elevation from the lab-based multi-camera motion capture and from the motion capture from the two smartphones.

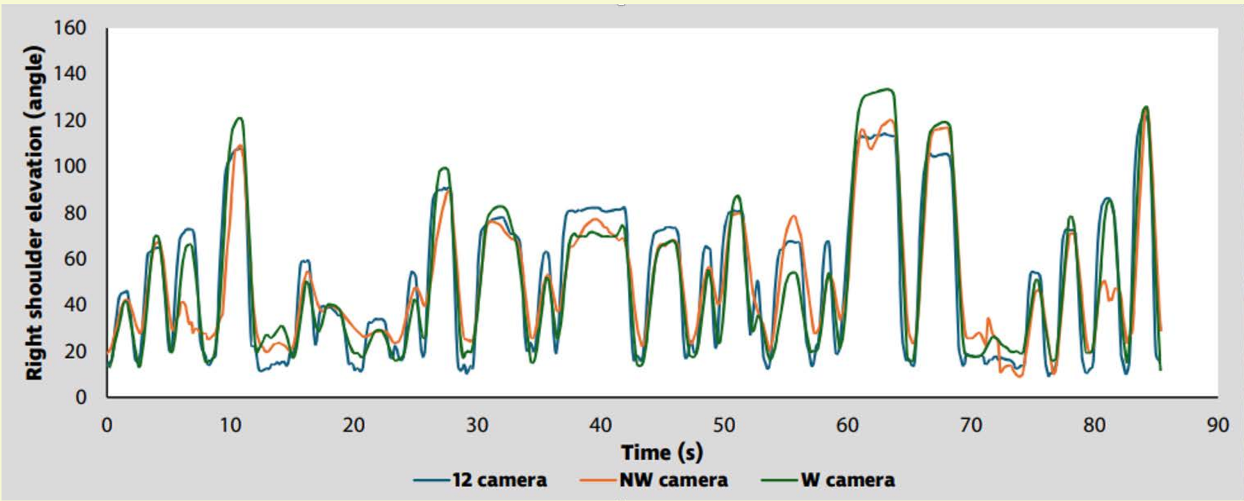


Figure 9. *Example of accuracy testing comparing shoulder elevation angles from the 12-camera lab motion analysis system and motion capture from two smartphone locations. Image Source:* ***https://www.assp.org/docs/default-source/psj-articles/f1butleryounggillette_1224.pdf?sfvrsn=1c287a46_0***

To date, the researchers are finalizing accuracy and validity findings for the study. Specifically, the research team is conducting the following data analyses:

- **SIRA measurement accuracy:** comparing AI-based motion capture to lab motion capture
- **SIRA prediction accuracy:** comparing predicted muscle activation to electromyographic sensor data
- **SIRA validation:** compare SIRA app to electromyographic fatigue prediction using the ACGIH TLV

When evaluating exoskeleton usage in both the construction worksite and lab setting, exoskeleton usage was associated with reduced muscle activity. However, perceptions of how useful exoskeletons are and their effect on perceived exertion varied, with more participants in the lab setting finding the exoskeleton useful to perform tasks and to reduce perceived exertion of the shoulder, elbow, and back. Further research should continue exploring the utility of exoskeletons across industries and tasks.

Overall, the findings of this study indicate that exoskeletons can reduce muscle activity when matched to job tasks that require elevated arm posture, but they need to be trialed in various settings to determine barriers such as confined spaces. When it comes to AI-based motion capture and the prototype SIRA application, the research team is optimistic that the technology can significantly increase capacity to conduct accurate shoulder injury risk assessments and could even be applied to standard ergonomic assessments or more complex biomechanical assessments. Further research is needed to improve the accuracy of AI-based motion capture, especially in handling obstructed views and multiple workers during video capture. Developing applications like SIRA also requires addressing programming maintenance, video storage, and AI processing costs.

Publications

- Butler T, Young M, & Gillette JC. (2024). Artificial intelligence (AI) for injury prevention. *Professional Safety, 69*(12): 24-30.

Research Study 4: Generating evidence-based guidance for implementing computer-vision-based technologies for MSD risk assessment – University of Waterloo

What's the Problem?

Computer-vision-based technologies provide a novel approach to capture inputs necessary for MSD risk assessments in the workplace. They differ from traditional methods that rely on manual observation, instead using AI to automatically and objectively analyze images or videos of work tasks. With the continued development of technologies that leverage AI for MSD risk assessments, safety professionals and organizations need guidance on how best to utilize these tools. To successfully implement the technology, workplaces must evaluate whether different assessment tools and models produce consistent results, determine which tools best suit specific tasks and environments, and establish standardized processes for conducting AI-based MSD risk assessments.

Project Aims

This study sought to fill a research gap by generating evidence to help workplaces understand when and where to make use of computer-vision-based MSD risk assessment tools to optimize MSD prevention efforts. The specific objectives of the project were to investigate whether 1) the angle of the video recording, 2) the pace of movement completing a task, or 3) the computer vision approach influences the accuracy of pose-estimations (postures that are estimated from the identification of key joints and angles on the body) when using computer-vision-based technologies to conduct MSD risk assessments.

Explanation of Technology and Research Methodology

Computer vision systems are designed to identify key body joints from image or video data to create a digital "skeleton", or 3D model, of the individual completing a task. By tracking the key body joints, the system can calculate joint angles, estimate postures, and compare them to predefined risk thresholds to assess the risk of MSDs.

To generate evidence on the practical application of computer-vision-based MSD risk assessment, 40 participants were recruited to complete six simulated work tasks (above shoulder work, cutting/trimming, packaging, sagittal lifting, palletizing, cart push/pull; Figure 10). To assess the influence of the video recording angle, each participant's body motion was captured as they performed each task using eight synchronized 2D video cameras, oriented in ~45° increments around the participant. Participants completed the tasks at both fast and slow speeds, with the faster pace determined by production line video data from an industry partner or by the speed the task could be completed with minimal errors.

Video data from all eight cameras were processed together to produce the ground-truth reference (obtained from the joint angular outputs from Theia Markerless). To assess whether using a particular computer vision approach impacts the accuracy of the MSD risk assessment, video data from each of the eight cameras was also processed individually through 4 to 5 different computer vision-based MSD risk assessment models (Figure 11). Trunk inclination and shoulder elevation angle estimates based on the individual camera angles and different computer vision approaches were compared to the ground-truth reference.

Above Shoulder Work

Cutting/Trimming

Packaging

Sagittal Lifting

Palletizing

Cart Push/Pull

Figure 10. *Images of the simulated work tasks completed for the study.*

Single-camera Pose Estimation Models

Multi-camera Model
from Theia Markerless

Figure 11. *On the left, video from a single camera angle of a participant completing a task is processed using five different models to produce single-camera 3D poses of skeletal segments and joints. On the right, video data from all eight cameras together are processed using Theia Markerless to produce the ground-truth, reference 3D poses of skeletal segments and joints.*

Results and Lessons Learned

Statistical analysis for this project is ongoing, with the research team continuing to analyze single-camera video data and compare it to the ground-truth reference data. The following are key findings so far.

Angle of Video Recording Influence: In general, across the computer-vision-based models, videos captured perpendicular to the plane of motion (i.e., from the side for forward arm raising or forward trunk bending) of the participant produced pose-estimations with less joint angle differences from the ground-truth data than videos captured from angles that were parallel or off-axis from the primary plane of motion (i.e., from the front or back when trying to observe forward arm raising or forward trunk bending). Pose-estimation was also more accurate when the video was captured from the same side of the body region of interest, for example capturing video of the left shoulder from the left side of the body (Figure 12).

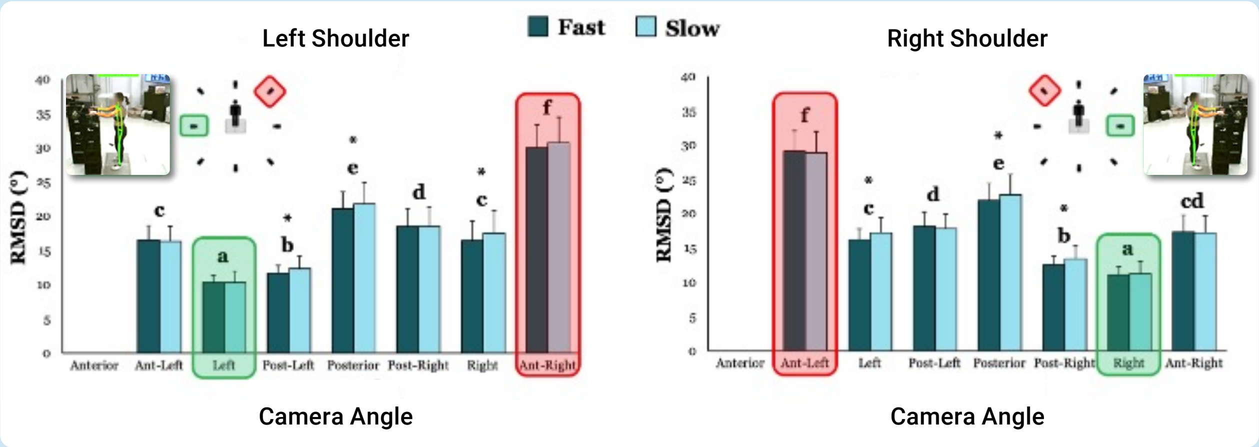


Figure 12. *Mean (+SD) root mean square differences (RMSD), of left and right shoulder elevation angles, in degrees (°), across three computer vision approaches during the above shoulder work task for each camera angle at both fast and slow movement paces. The green and red overlayed boxes indicate the camera angles with the lowest and highest RMSD, respectively, highlighting increased accuracy when capturing videos from the same side of the body region of interest. Inset images show the camera view that produced the lowest RMSD differences for the left and right side, respectively. Different letters above the bars indicate significant differences between camera angles, while the asterisks indicate significant differences between movement speed ($p < 0.05$).*

Movement Pace Influence: Initial observations indicate that when videos are captured from the ideal angle (i.e., perpendicular to and on the same side as the primary motion and limb of interest), movement pace did not influence outputs from single-camera pose-estimation models. However, when videos were captured from less ideal angles, movement pace had a modest effect (pacing had a smaller effect than angle) (Figure 12). It is important to note that more ballistic movements which were not captured in this study may still cause image blurring, potentially impacting the accuracy of the pose-estimation.

Computer Vision Approach Influence: While the angle of capture is the primary driver of differences between single-camera data and ground-truth, the computer vision approach can influence outputs. First, when capturing from the ideal angle (i.e., perpendicular to and on the same side as the primary motion and limb of interest), differences between ground-truth and single-camera computer vision were 5 to 15° depending on the joint and angle of rotation. This difference from ground-truth is likely explained by differences in how each “skeleton” or 3D model is defined between each single-camera system and the ground-truth model, in addition to how each model defines 0°, or a neutral posture. When comparing between single-camera systems (e.g., computer vision approaches) where data were recorded from the ideal angle, inter-system differences of 0 to 5° were observed. However, differences between single-camera systems vary by a greater margin when video is recorded from non-ideal angles. This is likely explained by how each underlying computer vision model was developed. When viewed from the ideal angle, the camera can “see” the joints of interest and computer vision can identify those joints of interest well. However, at non-ideal angles, where the camera cannot “see” all joints of interest, the computer vision must make a prediction. Preliminary data indicate that when video is recorded from non-optimal angles, the different computer vision approaches predict differently. As data analysis continues, the research team now has detailed documentation to describe the various ways that angles are calculated, which can help confirm whether differences are due to accuracy issues or different choices in how to identify joints and calculate angles, and how those different choices may influence their usability in given situations.

Publications

The study investigators have several articles and resources in preparation.

Key Takeaways

The four research projects focused on various aspects of emerging technology to address MSD risk, investigating solutions to automate parts of the risk assessment process and increase the capacity and readiness of workplaces to utilize technology to prevent MSDs. Specifically, the projects addressed the use of AI-based tools to detect MSD risks, assess the risk level, and identify solutions. They also provided practical guidance on how to implement these technologies. The Rutgers University project focused on reducing reliance on manual observation for identifying ergonomic problems and solutions. They achieved this by developing techniques to augment existing image captioning models, enabling them to visually explain the reasoning behind problem identification and solution prediction. Their results demonstrated that simple images of individuals performing tasks could serve as inputs for these models, quickly generating ergonomic solutions that are both efficient and easy to evaluate for accuracy.

The Virginia Tech and Iowa State University projects focused on risk assessment, investigating the use of AI-based motion capture as an approach to collecting data inputs for quantifying MSD risk. The Virginia Tech project focused on markerless cameras as a source of data inputs to quantify biomechanical demands of a task. Their findings indicated that the markerless motion capture data is adequate to classify a variety of manual material handling tasks and, when combined with in-sole pressure system data, produces accurate estimates of hand force and the biomechanical demands associated with a task. Data analysis for these projects is ongoing, but the initial findings underscore the need for further research into how factors such as sex, as well as specific work tasks and conditions, may influence the effectiveness of AI-based motion capture. Additionally, combining computer-vision-based technology with other methods, such as non-invasive wearable sensors, may be a promising approach.

The Iowa State University project concentrated specifically on shoulder MSDs and assessed the use of cell phone video data as an input for muscle fatigue predictive models. They developed a predictive model for shoulder Injury Risk Assessment and the SIRA mobile application that enables MSD risk assessment by simply uploading a cell phone video to the app. While they faced challenges with the video footage when it came to blocked camera views and multiple workers being in frame, the prototype SIRA application demonstrates how AI-based motion capture and predictive models can be applied for simple, low resource MSD risk assessment.

Continued development and refinement of these novel technologies is needed, along with practical guidance for using them. This was the goal of the University of Waterloo project—to generate and document evidence for the optimal use of computer-vision-based technologies for MSD risk assessments. While their analyses are ongoing, initial results provided insights into the impact of video angle and movement pace on pose-estimation accuracy. The team faced challenges when it came to comparing different computer vision approaches, however they extensively documented the various ways that angles and postures are calculated which will help users understand what the differences mean and how to interpret them. Upon finalization of data analysis, the researchers plan to develop a best practice guide to help safety professionals understand when and how to use these tools.

Conclusion

The 2023 - 2024 Research to Solutions grant program was an opportunity for organizations to advance research on MSD solutions in the workplace. The four awarded projects generated evidence that highlights significant developments and further research opportunities in the utility of emerging technology for MSD risk reduction. The MSD Solutions Lab is excited to continue the R2S grant program and its goal to inspire collaboration among academic institutions, businesses and industries to uncover promising, scalable, and transferable solutions that mitigate injury risk across sectors. More information about the grant programs can be found **here**.

Authors

Marley Reynoso, MPH and Paige DeBaylo, PhD

Funding

This research was supported by Amazon.com, Inc. Its contents and results reflect the authors' contributions and are solely their responsibility, and do not necessarily represent the views of Amazon, or the official position of NSC.