Evaluation of a Markerless Motion Capture System to Assess Physical Exposures During Manual Material Handling Tasks

Aanuoluwapo Ezekiel Ojelade

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Industrial and Systems Engineering

Maury A. Nussbaum, Chair
Sunwook Kim
Deborah E. Dickerson
Frederick Paige
Kwok-Leung Tsui

9th February, 2024
Blacksburg, Virginia

Keywords: classification; machine learning; prediction; biomechanical modeling

Evaluation of a Markerless Motion Capture System to Assess Physical Exposures During Manual Material Handling Tasks

Aanuoluwapo Ezekiel Ojelade

ABSTRACT

Manual material handling (MMH) tasks are associated with the development of work-related musculoskeletal disorders (WMSDs). Minimizing the frequency and intensity of handling objects is an ideal solution, yet MMH remains an integral part of many industry sectors, including manufacturing, construction, warehousing, and distribution. Physical exposure assessment can help identify high-risk tasks, guide the development and evaluation of ergonomic interventions, and contribute to understanding exposure-risk relationships. Physical exposure can be evaluated using self-assessment, observational methods, and direct measurements. Nevertheless, implementing these methods *in situ* can be challenging, time consuming, expensive, and infeasible or inaccurate in many cases. Thus, there is a critical need to improve physical exposure assessments to protect workers and save costs.

This dissertation assessed the accuracy of a markerless motion capture system (MMC) to quantify physical exposures during MMH tasks using three studies. Specifically, the first study investigated the performance of an MMC system, together with machine learning algorithms, for classifying diverse MMH tasks during a simulated complex job. In the second study, the feasibility of predicting dynamic hand forces was determined, using alternative measures, such as kinematics from MMC and/or in-sole pressure systems, coupled with a machine learning algorithm. Finally, in the third study, we systematically evaluated MMC for assessing biomechanical demands, by comparing outputs from a full-body musculoskeletal model driven by kinematic and kinetics from gold standard input and estimates derived from the MMC and in-sole pressure measurement system.

Overall, the findings of these studies demonstrated the potential of using MMC to classify several common occupational tasks and to estimate the associated biomechanical demands for a given worker (automatically and with minimal physical contact). Additionally, the methods developed here can help stakeholders rapidly assess an individual worker’s exposure to physical demands during diverse tasks.

Evaluation of a Markerless Motion Capture System to Assess Physical Exposures During Manual Material Handling Tasks

Aanuoluwapo Ezekiel Ojelade

GENERAL AUDIENCE ABSTRACT

Manual material handling (MMH) tasks expose workers to known risk factors for work-related musculoskeletal disorders (WMSDs) such as back and shoulder pain. Accurately quantifying workplace exposures to these risk factors is an essential aspect of identifying high-risk working conditions and for developing/evaluating workplace interventions to reduce WMSD risks. Current physical exposure assessment tools are labor-intensive, offer crude measures, and have limited application due to costs or feasibility. Using markerless motion capture (MMC) systems in the workplace could enable full or partial automation for the collection of critical measures such as the tasks a worker performs, the hand forces involved, and their biomechanical demands. New approaches are needed, though, since such automation is challenging due to variations in the type of input data required for different physical exposure assessments. In this dissertation, our goal was to assess the accuracy of MMC as a tool to quantify physical exposures during MMH tasks. In support of our goal, three studies were completed.

In the first study, we investigated the accuracy of using data from MMC together with machine learning algorithms to classify diverse MMH tasks, and distinguish among different task conditions. Our results emphasized that classification performance was satisfactory, though it differed between feature sets, MMH tasks, and between males and females. The second study explored combining MMC and IPM data with machine learning algorithms to predict hand forces during MMH tasks. Our results were encouraging overall, but predictions were less accurate in pushing and pulling tasks. In the third study, we evaluated an approach for estimating biomechanical demands on data obtained from MMC and in-sole pressure measurement systems. We compared estimates from a musculoskeletal model driven by kinematics from a whole-body inertial measurement unit and kinetics from direct measures of hand loads, and kinematics from MMC. Our findings support using MMC and kinetics from predicted hand forces as input for estimating biomechanical demands.

Overall, findings from these studies show that MMC can automatically classify common occupational tasks, predict dynamic hand forces, and estimate biomechanical demands with minimal physical contact. This new approach could allow stakeholders to assess worker's exposure and the efficiency of ergonomic interventions.

Acknowledgements

I would like to express my sincere gratitude to everyone who contributed to the completion of this dissertation.

I am deeply thankful to my advisor, Dr. Maury A. Nussbaum, for his unwavering support, guidance, and invaluable insights throughout the research process. His expertise and encouragement played a pivotal role in shaping the direction of this work. Besides knowledge and experience, he trained me with empathy and patience.

I extend my appreciation to the members of my dissertation committee, Drs. Sunwook Kim, Frederick Paige, Deborah Dickerson, and Kwok Tsui, for their valuable feedback and constructive criticism. Their expertise greatly enriched the quality of this research.

I am grateful to my lab mates at Occupational Ergonomics and Biomechanics Laboratories for their collaboration and meaningful discussions that enhanced the depth and breadth of my work.

Special thanks to the National Safety Council for their financial support, which made this research possible.

To the best friend I can ask for – Salewa – thank you for teaching me how to take meaningful breaks and checking on me when working during those late hours.

I want to acknowledge my friends for their understanding, encouragement, and unwavering belief in my abilities. Indeed, a conducive atmosphere begotten the sharpest mind. Also, their emotional support sustained me during challenging times.

To my family: Mom and dad, this degree belongs to you! Your words and admonitions always strengthen me when I feel like giving up. They are fuel to my spirit. I love you guys to the moon and back.

Last but not least, I express my gratitude to all the participants who generously contributed their time and insights to this study. This dissertation would not have been possible without the collective support and encouragement from all these individuals. Thank you for being part of this academic journey.

Table of Contents

List of Tables

List of Figures

1. Chapter 1

1.1 Introduction

Work-related musculoskeletal disorders (WMSDs) remain an important occupational health problem. In 2020, these disorders accounted for over one-third of workplace injuries, resulting in a median of ~13 lost workdays per incident (U.S. Bureau of Labor Statistics, 2021) and a substantial financial impact on workers and stakeholders. The manufacturing and warehousing sectors, as examples, paid about $2.24 and $1.50 billion, respectively, in 2020 to compensate for the direct costs of overexertion and other bodily injuries (Liberty Mutual Insurance, 2023). WMSDs are most common in the back, accounting for approximately ~30% of all cases (U.S. Bureau of Labor Statistics, 2021). Notably, epidemiological studies have attributed the high burden of low-back WMSDs to physical risk factors that include forceful exertions (Eriksen et al., 2004; Punnett and Wegman, 2004; Thomsen et al., 2007), repetitive lifting and bending (Andersen et al., 2003), non-neutral postures (Bernard and Putz-Anderson, 1997), and a rapid work pace (Harkness et al., 2004). To quantify work-related exposures to these risk factors, researchers have developed several occupational exposure assessment methods. Yet, implementing these methods on a job site can be challenging, such as due to differences in outcome measures, diagnostic criteria, and difficulty in interpreting assessment results (da Costa and Vieira, 2010; Garg and Kapellusch, 2009). Thus, there is a critical need to improve occupational exposure assessments to protect workers and save costs.

Occupational exposure assessment is the process of estimating the intensity, duration, and frequency of a worker's exposure to physical risk factors. Such assessment is essential for determining WMSD risk, designing and evaluating interventions, and understanding occupational exposure limits (i.e., dose-response relationships; Figure 1.1). The levels of such

exposures can be assessed in multiple ways, most commonly in terms of biomechanical loads, postures, and/or muscle activities in different body regions, either acutely or cumulatively (e.g., Bruno Garza et al., 2014; Coenen et al., 2012; Dolan and Adams, 1993; Dolan et al., 1999; Gallagher et al., 2017; Garg and Kapellusch, 2016; Jäger and Luttmann, 1989; Karhu et al., 1977; Kumar, 1990; Marras et al., 1993; Norman et al., 1998; Ringheim et al., 2015). As an example of the latter, Norman et al. (1998) identified peak spine loads, hand loads, and trunk kinematics as predictors of low back pain. Similarly, Marras et al. (1993) found that low-back moment, lift rate, lateral trunk velocity, sagittal trunk angle, and trunk twisting velocity could discriminate among job tasks with different historical risks of developing low back pain.

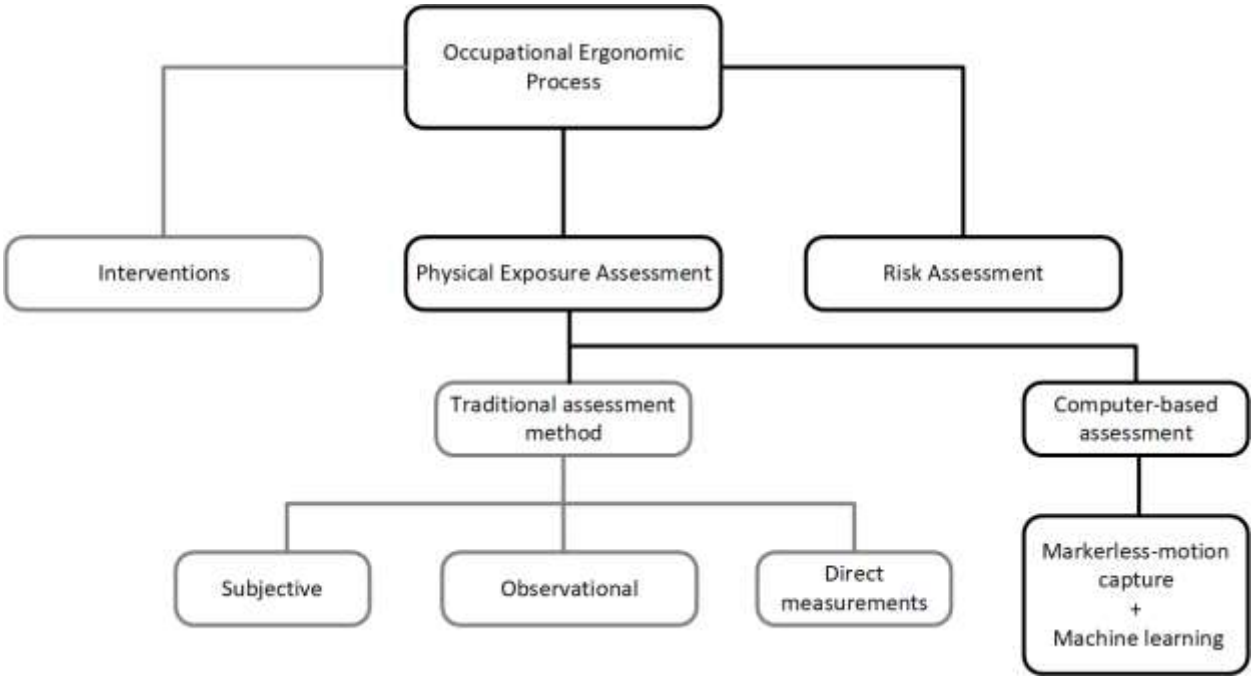

Figure 1.1. Overview of the typical occupational ergonomic process. Solid black lines and boxes indicate areas of focus in the current dissertation.

Such studies have demonstrated an association between physical exposures and WMSDs. However, a major challenge is the lack of methods for obtaining exposure metrics efficiently, thereby posing difficulties when identifying potential WMSD risks in diverse occupational

settings. In addition, some existing methods yield only crude exposure measures, which makes it challenging to obtain detailed estimates of risk or to infer the presence of a dose-response relationship. Therefore, there is a need for more detailed physical exposures that can be obtained more efficiently to promote a better understanding of risk factors for WMSDs and to aid in developing and evaluating interventions.

Current occupational exposure assessment approaches can be categorized broadly into three types: subjective (i.e., self-report), observational, and direct measurement (David, 2005a). The former two approaches are simple, quick, and cheap to use in the field, and can roughly identify body parts at most risk. However, major challenges are inherent to these approaches. For example, self-reported exposures and observational methods can be imprecise and unreliable due to large inter-rater variability, substantial time may be needed to perform detailed assessments, and can require extensive training (Burdorf and van der Beek, 1999; David, 2005a). Further, worker literacy levels might confound interpretation or comprehension of questions (Spielholz et al., 2001), and expert evaluations can be biased by inter- and intra-observer variability (David, 2005b).

A direct measurement approach typically involves the use of wearable sensors (e.g., accelerometers, inertial measurement units [IMUs], electromyography, and in-sole pressure sensors). Often in laboratory settings, these wearable sensors are attached to the body (skin or clothes) to record human parameters for ergonomic assessment. For example, IMUs were used in one study to quantify the physical exposure during simulated manual material handling tasks (Kim and Nussbaum, 2013). In another, kinematics data obtained from optical sensors were combined with electromyography measurements to estimate joint and muscle loadings during different activities simulated in the laboratory (Lloyd and Besier, 2003). These and many similar

studies (e.g., Antony and Keir, 2010; Antwi-Afari et al., 2018; Itoh et al., 2004; Kim and Nussbaum, 2014; Sommerich et al., 2000) indicate that direct measurement methods provide objective information to support physical exposure assessment. Wearable sensors though, can be intrusive, cause discomfort, or be inconvenient or infeasible in practice (Antwi-Afari et al., 2018; Golabchi et al., 2016), and their long-term use relies on the user's compliance with wearing and charging them. There are also limitations to specific sensors. Examples of such limitations for IMUs include cumulative error in estimating orientation, which can caused by gyroscope drift and magnetic disturbances in the presence of ferromagnetic materials that are often present in industrial environments (Robert-Lachaine et al., 2017).

In recent years, various emerging technologies, in conjunction with machine learning, have provided new alternatives to perform exposure assessments – specifically computer-based assessments. Of particular relevance here are advancements in automated data collection and analysis, hardware sensors, and computer vision applications, which have enabled the accurate identification of human postures (MassirisFernández et al., 2020). Both plain (RGB) and depth (RGB-D) cameras – referred to here as markerless cameras – are *ambient sensors* that offer an approach to quantifying occupational physical exposures without the need for any on-body sensors. As example applications, markerless cameras have been used to extract 3D human pose (Liu et al., 2016), classify worker actions and the equipment used to perform those actions (Koppula et al., 2013; Li et al., 2020), and estimate risk scores using ergonomic risk assessment tools (Humadi et al., 2021; Parsa et al., 2019). These and similar reports (Scano et al., 2020; Xu et al., 2017; Yan et al., 2017; Zhang et al., 2018) have demonstrated that the combination of machine learning with a markerless camera can quantify occupational exposures continuously

and efficiently (and, in some cases, quantify WMSD risk by providing input to existing risk assessment tools).

However, reported applications of markerless cameras in the workplace are still limited. Specifically, there are important drawbacks associated with technology and data processing, coupled with a lack of thorough validation. Regarding the former, some markerless cameras have a narrow field of view (Dutta, 2012; Spector et al., 2014), low resolution (Khoshelham and Elberink, 2012), an inability to parse data from sensors efficiently (Yu et al., 2019), and limited usability in outdoor conditions (Spector et al., 2014). Regarding validation, reported research has been limited to classifying fairly simple tasks with high inter-task variability (Escorcia et al., 2012; Khosrowpour et al., 2014; Yang et al., 2016), which are relatively "easy" classification problems. New approaches are required to more fully harness the potential of markerless cameras for occupational exposure assessments. Note that *depth cameras* are of interest here, because plain cameras require a large setup space and longer calibration time, and they are more sensitive to background color, texture, and clothing (Regazzoni et al., 2014).

1.2 Research Needs

As noted, markerless camera technology could enable continuous monitoring and quantification of occupational physical exposures. This dissertation was performed to help realize such potential, with an overall objective to develop and evaluate new approaches for quantifying physical exposures in the workplace using markerless camera technology. In support of this research needs, three studies were completed:

Development of Automated Occupational Task Classifiers

Many studies have focused on the accuracy of markerless cameras in tracking and classifying simple, non-occluded worker postures, but little has been done to generalize such

findings to more complex working postures. For example, a deep neural network was used in one recent study to estimate 3D postures from two markerless cameras (Mehrizi et al., 2019b). While this approach demonstrated sufficient accuracy to justify the use of markerless cameras as a risk assessment tool in industry, a major limitation was that the 3D posture estimation yielded substantially larger errors during complex lifting conditions. Further, Tang and Golparvar-Fard (2021) showed that body occlusion greatly reduced the accuracy and sensitivity of models in predicting risk severity levels. Similar limitations were reported by others (Albert et al., 2020; Xu et al., 2015), and emphasize that new methods are needed to determine if markerless cameras are effective for use in industries where workers assume complex and/or varying postures.

Predicting Dynamic Hand Forces during MMH Tasks

Existing evidence suggests that data from markerless cameras could be used to classify hand loads categorically and detect excessive loads handled by workers. Such information could be useful as a first step in overcoming the problem of collecting information on physical loads in the workplace (Lee et al., 2020; Yang et al., 2020). For example, Lee et al. (2020) predicted different physical load levels (i.e., no, standard, or excessive load) using three markerless cameras and deep learning. In another study, kinematic features generated from IMU data were used as predictors when automatically classifying load levels as no, or 50%/75% of maximum acceptable weight of carry (Lim and D'Souza, 2019). This latter study supports the potential for using data from markerless cameras to classify loads being handled by workers. On the other hand, classifying loads only categorically has limited utility in risk assessment, because biomechanical models and some risk assessment tools require more precise information to compute risk scores. For example, the NIOSH lifting equation requires the specific load weight being handled to assess a lifting task.

Estimating Biomechanical Demands during MMH Tasks

Data extracted from markerless cameras have been used to estimate risk assessment scores. For example, RULA scores were automatically predicted in one study from digital video using computer vision techniques (MassirisFernández et al., 2020). While this and similar studies (Hu et al.; Li et al., 2020; Mehrizi et al., 2019a; Tang and Golparvar-Fard, 2021) show the potential of markerless camera to provide accurate ergonomic assessments, none of the proposed methods consider the effect of dynamic forces and moments in complex linkage systems. Joint loadings, however, can be substantially affected by dynamics, especially at the spine (Bush-joseph et al., 2007). Therefore, existing methods are not applicable and/or have not been tested with realistic dynamic working conditions and are likely underestimate biomechanical exposures (De Looze et al., 1994; Freivalds et al., 1984; Garg et al., 1982; Leskinen et al., 1983; Wood and Hayes, 1974).

The current work is innovative in three main ways. First, it uses deep learning algorithms to automatically classify diverse occupational tasks. While several studies have evaluated the performance of markerless camera for posture estimation, the effect of complex kinematics (e.g., asymmetric postures and motions) and high intra- and inter-task variability have received little attention, despite the complexity of the kinematics required during diverse occupational tasks. Using data from markerless cameras together with machine learning algorithms is an innovative approach that could help identify specific MMH tasks and distinguish among different task conditions. Second, this work is innovative in developing methods to predict 3D dynamic hand forces involved when manipulating objects. While some research has addressed classifying load levels, knowing the actual external loads is critical for exposure assessment (and potential biomechanical loading). Third, the current work is the first, to our knowledge, to quantify the

ability of a markerless camera in estimating biomechanical load when dynamic aspects are considered.

1.3 Impacts

Detailed and valid physical exposure assessments in the workplace are needed to help reduce the prevalence of WMSDs and better understand exposure-response relationships. Markerless cameras have clear potential to provide detailed and descriptive information on diverse physical and exposures in the workplace. The current work showed that markerless cameras, in conjunction with machine learning, could assist ergonomists in estimating physical exposures and biomechanical demands using non-intrusive approaches. As such, this work helps stakeholders rapidly assess an individual worker's physical exposure for realistic occupational tasks and thereby aid in job redesign. Further, this work could improve the standardization of risk assessments performed in various workplaces.

1.4 References

Albert, J.A., Owolabi, V., Gebel, A., Brahms, C.M., Granacher, U., Arnrich, B., 2020. Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. Sensors 20, 5104.

Andersen, J.H., Kaergaard, A., Mikkelsen, S., Jensen, U.F., Frost, P., Bonde, J.P., Fallentin, N., Thomsen, J.F., 2003. Risk factors in the onset of neck/shoulder pain in a prospective study of workers in industrial and service companies. Occup Environ Med 60, 649-654.

Antony, N.T., Keir, P.J., 2010. Effects of posture, movement and hand load on shoulder muscle activity. Journal of Electromyography and Kinesiology 20, 191-198.

Antwi-Afari, M.F., Li, H., Yu, Y., Kong, L., 2018. Wearable insole pressure system for automated detection and classification of awkward working postures in construction workers. Automation in Construction 96, 433-441.

Bernard, B.P., Putz-Anderson, V., 1997. Musculoskeletal disorders and workplace factors; a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back.

Bruno Garza, J.L., Eijckelhof, B.H.W., Huysmans, M.A., Johnson, P.W., van Dieen, J.H., Catalano, P.J., Katz, J.N., van der Beek, A.J., Dennerlein, J.T., 2014. Prediction of trapezius muscle activity and shoulder, head, neck, and torso postures during computer use: results of a field study.

Burdorf, A., van der Beek, A., 1999. Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. Scand J Work Environ Health 25 Suppl 4, 25-30.

Bush-joseph, C., Schipplein, O., J Andersson, G.B., Andriacchi, T.P., 2007. Influence of dynamic factors on the lumbar spine moment in lifting.

Coenen, P., Kingma, I., Boot, C.R.L., Bongers, P.M., Van Dieën, J.H., 2012. The contribution of load magnitude and number of load cycles to cumulative low-back load estimations: A study based on in-vitro compression data. Clinical Biomechanics 27, 1083-1086.

da Costa, B.R., Vieira, E.R., 2010. Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. Am J Ind Med 53, 285-323.

David, G.C., 2005a. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occup Med (Lond) 55, 190-199.

David, G.C., 2005b. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occupational medicine 55, 190-199.

De Looze, M.d., Kingma, I., Thunnissen, W., Van Wijk, M., Toussaint, H., 1994. The evaluation of a practical biomechanical model estimating lumbar moments in occupational activities. Ergonomics 37, 1495-1502.

Dolan, P., Adams, M.A., 1993. The relationship between EMG activity and extensor moment generation in the erector spinae muscles during bending and lifting activities. Journal of Biomechanics 26, 513-522.

Dolan, P., Kingma, I., van Dieen, J., de Looze, M.P., Toussaint, H.M., Baten, C.T.M., Adams, M.A., 1999. Dynamic Forces Acting on the Lumbar Spine During Manual Handling: Can They be Estimated Using Electromyographic Techniques Alone? Spine 24, 698-703.

Dutta, T., 2012. Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. Appl Ergon 43, 645-649.

Eriksen, W., Bruusgaard, D., Knardahl, S., 2004. Work factors as predictors of intense or disabling low back pain; a prospective study of nurses' aides. Occupational and Environmental Medicine 61, 398-404.
Escorcia, V., Dávila, M.A., Golparvar-Fard, M., Niebles, J.C., 2012. Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras, Construction Research Congress 2012: Construction Challenges in a Flat World, pp. 879-888.
Freivalds, A., Chaffin, D.B., Garg, A., Lee, K.S., 1984. A dynamic biomechanical evaluation of lifting maximum acceptable loads. Journal of Biomechanics 17, 251-262.
Gallagher, S., Sesek, R.F., Schall, M.C., Huangfu, R., 2017. Development and validation of an easy-to-use risk assessment tool for cumulative low back loading: The Lifting Fatigue Failure Tool (LiFFT). Applied Ergonomics 63, 142-150.
Garg, A., Chaffin, D.B., Freivalds, A., 1982. BIOMECHANICAL STRESSES FROM MANUAL LOAD LIFTING: A STATIC VS DYNAMIC EVALUATION. IIE Transactions (Institute of Industrial Engineers) 14, 272-281.
Garg, A., Kapellusch, J.M., 2009. Applications of biomechanics for prevention of work-related musculoskeletal disorders. Ergonomics 52, 36-59.
Garg, A., Kapellusch, J.M., 2016. The Cumulative Lifting Index (CULI) for the Revised NIOSH Lifting Equation: Quantifying Risk for Workers With Job Rotation. Human factors 58, 683-694.
Golabchi, A., Han, S., Fayek, A.R., 2016. A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations. Canadian Journal of Civil Engineering 43, 294-303.
Harkness, E.F., Macfarlane, G.J., Nahit, E., Silman, A.J., McBeth, J., 2004. Mechanical injury and psychosocial factors in the work place predict the onset of widespread body pain - A two-year prospective study among cohorts of newly employed workers. Arthritis Rheum-Us 50, 1655-1664.
Hu, B., Kim, C., Ning, X., Xu, X., Using a deep learning network to recognise low back pain in static standing.
Humadi, A., Nazarahari, M., Ahmad, R., Rouhani, H., 2021. In-field instrumented ergonomic risk assessment: Inertial measurement units versus Kinect V2. International Journal of Industrial Ergonomics 84, 103147.
Itoh, Y., Akataki, K., Mita, K., Watakabe, M., Itoh, K., 2004. Time-frequency analysis of mechanomyogram during sustained contractions with muscle fatigue. Systems and Computers in Japan 35, 26-36.
Jäger, M., Luttmann, A., 1989. Biomechanical analysis and assessment of lumbar stress during load lifting using a dynamic 19-segment human model. Ergonomics 32, 93-112.
Karhu, O., Kansi, P., Kuorinka, I., 1977. Correcting working postures in industry: A practical method for analysis. Applied Ergonomics 8, 199-201.
Khoshelham, K., Elberink, S.O., 2012. Accuracy and resolution of Kinect depth data for indoor mapping applications. Sensors (Basel) 12, 1437-1454.
Khosrowpour, A., Niebles, J.C., Golparvar-Fard, M., 2014. Vision-based workface assessment using depth images for activity analysis of interior construction operations. Automation in Construction 48, 74-87.

Kim, S., Nussbaum, M.A., 2013. Performance evaluation of a wearable inertial motion capture system for capturing physical exposures during manual material handling tasks. Ergonomics 56, 314-326.
Kim, S., Nussbaum, M.A., 2014. An evaluation of classification algorithms for manual material handling tasks based on data obtained using wearable technologies. Ergonomics 57, 1040-1051.
Koppula, H.S., Gupta, R., Saxena, A., 2013. Learning human activities and object affordances from RGB-D videos. Int J Robot Res 32, 951-970.
Kumar, S., 1990. Cumulative Load as a Risk Factor for Back Pain. Spine 15, 1311-1316.
Lee, H., Yang, K., Kim, N., Ahn, C.R., 2020. Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. Automation in Construction 120, 103390.
Leskinen, T., Stålhammar, H., Kuorinka, I., Troup, J., 1983. A dynamic analysis of spinal compression with different lifting techniques. Ergonomics 26, 595-604.
Li, L., Martin, T., Xu, X., 2020. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. Applied Ergonomics 87, 103138-103138.
Liberty Mutual Insurance, 2023. 2023 Workplace Safety Index: The Top 10 Causes of Disabling Injuries.
Lim, S., D'Souza, C., 2019. Statistical prediction of load carriage mode and magnitude from inertial sensor derived gait kinematics. Applied ergonomics 76, 1-11.
Liu, M., Han, S., Lee, S., 2016. Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis. Construction Innovation 16, 348-367.
Lloyd, D.G., Besier, T.F., 2003. An EMG-driven musculoskeletal model to estimate muscle forces and knee joint moments in vivo. J Biomech 36, 765-776.
Marras, W.S., Lavender, S.A., Leurgans, S.E., Rajulu, S.L., Allread, S.W.G., Fathallah, F.A., Ferguson, S.A., 1993. The role of dynamic three-dimensional trunk motion in occupationally-related. Spine 18, 617-628.
MassirisFernández, M., Fernández, J.Á., Bajo, J.M., Delrieux, C.A., 2020. Ergonomic risk assessment based on computer vision and machine learning. Computers & Industrial Engineering 149, 106816.
Mehrizi, R., Peng, X., Metaxas, D.N., Xu, X., Zhang, S., Li, K., 2019a. Predicting 3-D lower back joint load in lifting: A deep pose estimation approach. IEEE Transactions on Human-Machine Systems 49, 85-94.
Mehrizi, R., Peng, X., Xu, X., Zhang, S., Li, K., 2019b. A Deep Neural Network-based method for estimation of 3D lifting motions. J Biomech 84, 87-93.
Norman, R., Wells, R., Neumann, P., Frank, J., Shannon, H., Kerr, M., Beaton, D.E., Bombardier, C., Ferrier, S., Hogg-Johnson, S., Mondloch, M., Peloso, P., Smith, J., Stansfeld, S.A., Tarasuk, V., Andrews, D.M., Dobbyn, M., Edmonstone, M.A., Ingelman, J.P., Jeans, B., McRobbie, H., Moore, A., Mylett, J., Outerbridge, G., Woo, H., 1998. A comparison of peak vs cumulative physical work exposure risk factors for the reporting of low back pain in the automotive industry. Clinical Biomechanics 13, 561-573.
Parsa, B., Samani, E.U., Hendrix, R., Devine, C., Singh, S.M., Devasia, S., Banerjeee, A.G., 2019. Toward Ergonomic Risk Prediction via Segmentation of Indoor Object

Manipulation Actions Using Spatiotemporal Convolutional Networks. Ieee Robot Autom Let 4, 3153-3160.
Punnett, L., Wegman, D.H., 2004. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. J Electromyogr Kinesiol 14, 13-23.
Regazzoni, D., De Vecchi, G., Rizzi, C., 2014. RGB cams vs RGB-D sensors: Low cost motion capture technologies performances and limitations. Journal of Manufacturing Systems 33, 719-728.
Ringheim, I., Austein, H., Indahl, A., Roeleveld, K., 2015. Postural strategy and trunk muscle activation during prolonged standing in chronic low back pain patients. Gait and Posture.
Robert-Lachaine, X., Mecheri, H., Larue, C., Plamondon, A., 2017. Effect of local magnetic field disturbances on inertial measurement units accuracy. Applied ergonomics 63, 123-132.
Scano, A., Mira, R.M., Cerveri, P., Tosatti, L.M., Sacco, M., 2020. Analysis of Upper-Limb and Trunk Kinematic Variability: Accuracy and Reliability of an RGB-D Sensor. Multimodal Technolog 4.
Sommerich, C.M., Joines, S.M., Hermans, V., Moon, S.D., 2000. Use of surface electromyography to estimate neck muscle activity. Journal of Electromyography and Kinesiology 10, 377-398.
Spector, J.T., Lieblich, M., Bao, S., McQuade, K., Hughes, M., 2014. Automation of Workplace Lifting Hazard Assessment for Musculoskeletal Injury Prevention.
Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., Kaufman, J., 2001. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. Ergonomics 44, 588-613.
Tang, S., Golparvar-Fard, M., 2021. Machine Learning-Based Risk Analysis for Construction Worker Safety from Ubiquitous Site Photos and Videos. Journal of Computing in Civil Engineering 35, 04021020.
Thomsen, J.F., Mikkelsen, S., Andersen, J.H., Fallentin, N., Loft, I.P., Frost, P., Kaergaard, A., Bonde, J.P., Overgaard, E., 2007. Risk factors for hand-wrist disorders in repetitive work. Occup Environ Med 64, 527-533.
U.S. Bureau of Labor Statistics, 2021. Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work. bls.gov, Washington, DC.
Wood, G., Hayes, K., 1974. A kinetic model of intervertebral stress during lifting. British Journal of Sports Medicine 8, 74-79.
Xu, X., McGorry, R.W., Chou, L.-S., Lin, J.-h., Chang, C.-c., 2015. Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking. Gait & posture 42, 145-151.
Xu, X., Robertson, M., Chen, K.B., Lin, J.H., McGorry, R.W., 2017. Using the Microsoft Kinect to assess 3-D shoulder kinematics during computer use. Appl Ergon 65, 418-423.
Yan, X.Z., Li, H., Li, A.R., Zhang, H., 2017. Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. Automation in Construction 74, 2-11.
Yang, J., Shi, Z.K., Wu, Z.Y., 2016. Vision-based action recognition of construction workers using dense trajectories. Adv Eng Inform 30, 327-336.
Yang, K., Ahn, C.R., Kim, H., 2020. Deep learning-based classification of work-related physical load levels in construction. Advanced Engineering Informatics 45, 101104-101104.

Yu, Y., Yang, X., Li, H., Luo, X., Guo, H., Fang, Q., 2019. Joint-Level Vision-Based Ergonomic Assessment Tool for Construction Workers. Journal of Construction Engineering and Management 145, 04019025.
Zhang, H., Yan, X., Li, H., 2018. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. Automation in Construction 94, 1-10.

2. Chapter 2. Classifying Manual Material Handling Tasks from Markerless Motion Capture: A Data-Driven Approach Using Recurrent Neural Networks

2.1 Abstract

Effective physical exposure assessment methods are needed in the workplace to quantify musculoskeletal disorder risks and the association between exposure and risks. While several tools are available, they are often limited in scope and can be resource-intensive. We investigated the use of kinematic data from a markerless motion capture system, with various machine-learning algorithms and input data features, to classify among eight distinct material handling tasks and specific task conditions (i.e., hand configuration and initial lifting height). Overall, using the kinematic data led to satisfactory results (e.g., mean precision of 85 – 97%) in classifying the tasks and the task conditions. Our results, though, also emphasize that classification performance differed between feature sets, tasks, and between males and females. Nonetheless, use of markerless motion capture appears to have clear potential for ergonomic exposure assessment

2.2 Introduction

Work-related musculoskeletal disorders (WMSDs) continue to be an important health concern. In the United States, WMSDs led to a median of 14 days away from work (U.S. Bureau of Labor Statistics, 2021) and involved substantial direct costs (about $2.24 billion) to employers (Liberty Mutual Insurance, 2023). Several risk factors have been associated with the development of WMSDs, including forceful exertions, repetition, and non-neutral postures. Exposures to such risk factors are particularly common during manual material handling (MMH) tasks, including lifting/lowering, pushing/pulling, holding and carrying (Andersen et al., 2003; Bernard & Putz-Anderson, 1997; da Costa & Vieira, 2010). Some tasks entail variable or complex exposures, however, due to variability in work styles or task demands, making it challenging to identify the specific ergonomic risks (Burdorf & van Riel, 1996; Dempsey, 1999). Effective assessment of physical exposures, though, is often critical to developing targeted interventions, and more generally for quantifying associations between exposures and risks (Marras et al., 2009; Plantard et al., 2017; Waters et al., 2007). A fundamental aspect of physical exposure assessment involves distinguishing the specific MMH tasks performed (aka, task classification); doing is an important step in identifying high-risk work conditions (Li & Buckle, 1999), especially since different tasks present distinct physical exposures and WMSD risks.

To classify MMH tasks, physical exposure assessment tools are needed that are accurate and compatible with the work environment. Physical exposures can be assessed using self-assessment, human observation, and direct measurement (David, 2005; Li & Buckle, 1999). Self-assessment and human observation approaches are quick, straightforward, and require no advanced technologies. However, these approaches can be compromised by individual biases or sub-optimal workplace conditions (e.g., occlusions and limited viewpoints) (Pedersen et al.,

2016; Plantard et al., 2017). Alternatively, direct measurements often involve attaching sensors to objects or directly to a worker's body, such as to obtain postural or force data (David, 2005). Direct measurements generally yield rich, precise data, but can be limited by cost, the time needed to perform assessments, sensor discomfort, and potential influences on work behaviors (Antwi-Afari et al., 2018; Golabchi et al., 2016; Nath et al., 2017). Overall, major limitations of direct measurement methods are that continuous measurements are both labor intensive (Rezagholi et al., 2012) and resource-demanding (Wells et al., 1994). Therefore, there would be clear value in automating data collection and analysis to address the challenges posed by existing physical exposure assessment methods.

Recently, advancements in hardware sensors and machine learning have provided new alternatives to perform occupational physical exposure assessments – specifically computer-based assessments (MassirisFernández et al., 2020). Of relevance here are *markerless motion capture* (MMC) systems, both plain and depth cameras. Existing work demonstrates the feasibility of a computer-based assessment approach for continuous motion tracking and quantifying physical exposures, without requiring on-body sensors (Plantard et al., 2017; Zhang et al., 2018). Two example applications emphasize such feasibility. In one, data extracted from a MMC system were used to automatically detect activities such as walking and specific tasks during construction drywall installation (Khosrowpour et al., 2014). In another, MMC data facilitated classifying different construction worker actions when laying bricks, transporting rebar, and making formwork (Yang et al., 2016).

However, reported applications of MMC have three important limitations. First, it is unclear if existing task classifiers can be applied when the tasks of interest include relatively complex motions. In both studies noted above (Khosrowpour et al., 2014; Yang et al., 2016),

reasonable classification accuracy (~75%) was found for simpler construction tasks, such as shoveling and transporting, but this accuracy was substantially lower for more tasks, such as bolting and plastering. Second, occupational task classification performance often depends on either the specific classification algorithm used or the features used as input variables to the algorithm. Several task classification algorithms have been explored – such as Support Vector Machines, K-Nearest Neighbors, Decision Trees, and Neural Networks (e.g., Escorcia et al., 2012; Park et al., 2016; Song et al., 2010; Yu et al., 2019; Zhan et al., 2012; Zhang & Tian, 2012) – and the performance for a given task varies based on the specific algorithm and input variables employed. Third, the tasks included in earlier reports have often been generic, lacking adequate representation of occupational tasks generally and MMH tasks specifically.

Our purpose in this study was thus to investigate the performance of an MMC system, together with a machine learning algorithm, for classifying diverse MMH tasks during a simulated complex job. Specifically, we explored the relative performance of using different machine learning algorithms as an ergonomic exposure assessment approach for identifying specific MMH tasks and distinguishing among different task conditions (e.g., initial lifting height). Several machine learning algorithms were tested, since no single model was expected to be best suited for MMH task classification (Jozefowicz et al., 2015). We also evaluated the effects of using various input variables (i.e., feature sets, based on kinematic data). Our study was exploratory in nature, seeking results that could inform future assessments of physical exposures using MMC.

2.3 Methods

MMH tasks were simulated in a controlled, laboratory setting, and these tasks were then classified using body kinematics obtained from an MMC system. Diverse tasks were simulated,

representative of physically-demanding activities in several occupational sectors (e.g., lifting, carrying, pushing). Several task classifiers were explored using different machine learning methods, and the performance of these classifiers were examined using common metrics.

Participants

A convenience sample of 36 young (14 females) participants completed the study and were recruited from the university and local community. Respective means (SD) of age, body mass, and stature were 27 (4.41) years, 77.5 (12.23) kg, and 176.4 (6.9) cm for males; and 27 (5.24) years, 68 (12.34) kg, and 170.1 (7.25) cm for females. All participants reported being physically active (i.e., exercising at least twice per week) and having no musculoskeletal disorders within the past year. The research reported herein complied with the tenets of the Declaration of Helsinki, and the study protocol was approved by the Institutional Review Board at Virginia Tech. Informed consent was obtained from all participants prior to any data collection.

Task Simulations

Eight MMH tasks were simulated in the laboratory, and these tasks involved some variations of manual box lifting, carrying, pushing, pulling, and reaching. The tasks simulated here are similar to those used in an earlier study that evaluated the efficacy of several classification models for MMH tasks (Kim and Nussbaum (2014). Specifically, six within the current set of tasks were also included in the noted earlier study. Distinct here, however, was the inclusion of cart pushing and the use of different hand configurations, box masses, lift origins, and starting positions. The specific MMH tasks that were simulated are described below (see also Figure 2.1). A single wood box was used in the study (width = 26.0 cm; depth = 41 cm; and height = 23.5 cm).

- Task 1: symmetric box lifting from two different origins (floor and individual knee height) to an individual hip height. Hip height was defined as the vertical distance from the floor to the greater trochanter.
- Task 2: asymmetric box lifting, from a table placed in front of the participant to another positioned 90º to the left of the participant. The two tables were adjusted to individual hip height.
- Task 3: box carriage from one table to another (i.e., lifting from hip height, carrying, and lowering to a height of 0.74m). Participants carried the box over a distance of 2.4 m, selected as the 50th percentile of the carrying distance of the U.S. workforce (Ciriello et al., 1999).
- Task 4: box pushing over a distance of ~0.7 m, with table height = 0.74 m.
- Task 5: box pulling toward the body over a distance of ~0.7m, with the table height fixed at 0.74 m.
- Task 6: cart pushing at individual waist height over a distance of ~1.8m. Cart mass was fixed at 86 kg, including the box, representing common cart loads (Hoozemans et al., 2004). Cart pushing was completed only using two hands.
- Task 7: overhead lifting from cart height to individual overhead height, with the box lifted from a height of 0.56 m (height of the cart). Overhead height was defined from individual anthropometric measures as the distance between the lateral epicondyle and the floor when the shoulder was flexed at 80º. This anthropometric measure was selected since working repeatedly with arm flexion or abduction beyond 80º has been associated with shoulder disorders (Bernard & Putz-Anderson, 1997).

- Task 8: lowering to different origins (floor and individual knee height) from overhead height. Participants carried the box over a distance of ~1.3m.

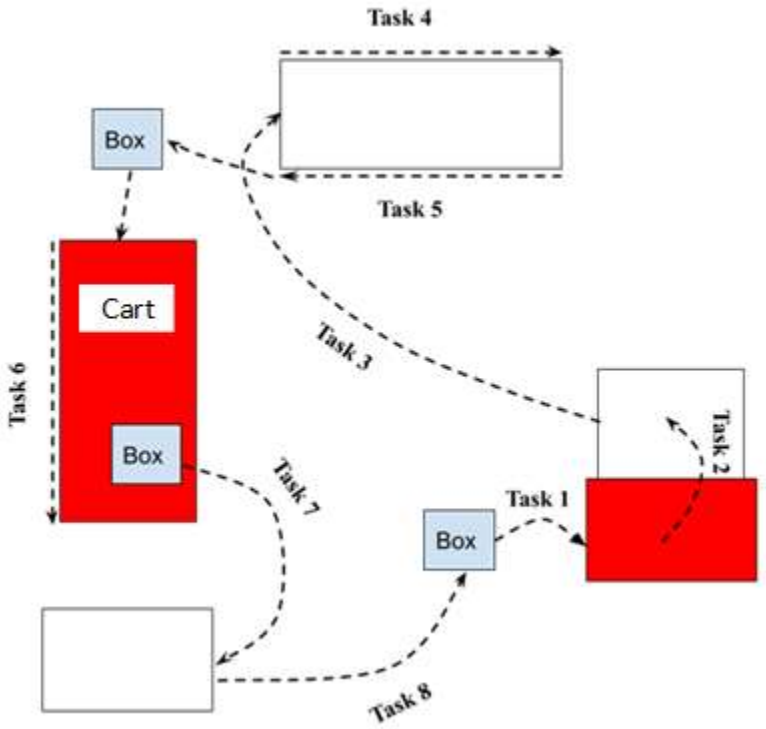

Figure 2.1. A top-view schematic of the simulated tasks. Dotted lines indicate the movement path, and red rectangles indicate two alternative starting positions.

A total of 36 experimental scenarios were completed for the simulated MMH tasks, involving all possible combinations of 36 task conditions – all combinations of three levels of *Hand Configuration,* three levels of *Box Mass,* two levels of *Lift Origin,* and two levels of *Start Position. Hand Configuration* had three levels: broad, narrow, and one-hand (Figure 2.2). Both the broad and narrow hand configurations involved using both hands, for which the handles were separate by 52 and 33 cm, respectively. In the one-hand configuration, the box handle was positioned in the middle of the box. Multiple hand configurations were used here since different hand widths and one- vs. two-handed lifting methods impose different biomechanical demands

on the lower back (Garg et al., 1982; Gary et al., 1996; Marras & Davis, 1998). Three levels of *Box Mass* – 6, 9, and 12 kg – were used for the broad and narrow hand configuration, while masses of 5, 7, and 9 kg were used for one-hand hand configuration. Different box masses were used for lifting tasks involving one vs. two hands, based on results from pilot testing (i.e., to ensure that most participants could complete all tasks). Specific masses used here were roughly within the 8th and 22nd percentiles of masses lifted by the U.S. workforce (Ciriello et al., 1999). Two levels of *Lift Origin* – floor and individual knee height – were included to impose different physical exposures during lifting. Finally, two levels of *Starting Position* were used to impose more task variability. Specifically, one starting position was set at Task 1 and another at Task 6 (Figure 1).

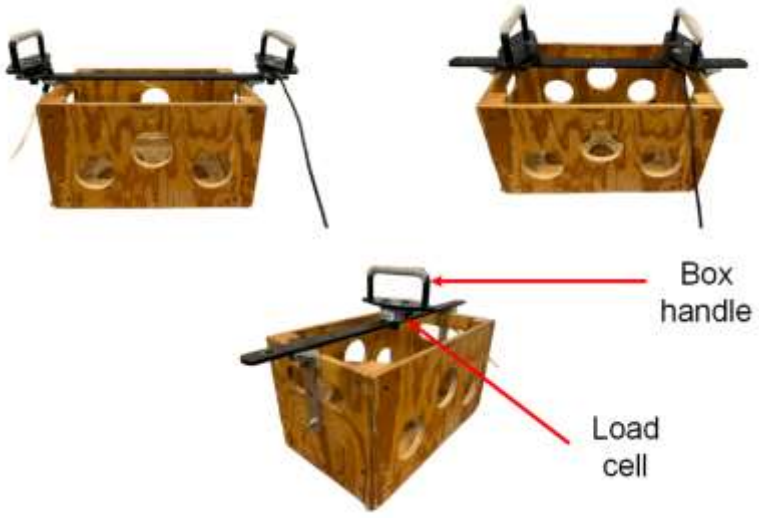

Figure 2.2. Illustration of the three levels of hand configurations with the broad, narrow and one-hand hand configurations shown on the left, middle, and right, respectively.

Experimental Procedures

Participants completed one experimental session (~3 hrs.), which consisted of training and experimental phases. During the training phase, participants were introduced to the MMH

tasks and the different task conditions, then practiced the simulated tasks. They were asked to perform all tasks using their own comfortable work strategies and speed, while assuming they were working in an industrial environment.

In the experimental phase, participants completed multiple trials of the MMH tasks, one trial in each of the 36 task conditions. A study *trial* involved completing all of the eight MMH tasks sequentially, with a given box mass, hand configuration, lift origin, and from a given starting position. The presentation order of *Hand Configuration* was counterbalanced using 3 × 3 balanced Latin Squares. Within a given *Hand Configuration*, the presentation order of *Box Mass* was also counterbalanced using 3 × 3 balanced Latin Squares, whereas the presentation orders of *Starting Position* and *Lift Origin* were alternated across participants. To mitigate physical fatigue, a minimum of four minutes of rest was given between each *Hand Position* condition.

Instrumentation

Whole-body kinematics were monitored at 30 Hz using three markerless camera systems (Azure Kinect™ [AKT], Microsoft Corporation, Seattle WA, USA). These systems were positioned ~1.74 m from the edge of the work area, a configuration that was determined to be effective during pilot testing and that aimed to optimize the coverage of the narrow camera field of view (Figure 2.13 Appendix A). The three AKTs were time-synchronized using a 3.5 mm audio cable connected in a daisy-chain configuration, where one of the AKTs was designated as the "master" device, with the remaining two as subordinate devices. An iPi Recorder (iPi Soft®; www.ipisoft.com) was used for sampling from the AKTs. At the start of each trial, participants completed a calibration step by standing in a "T-pose" while facing one of the markerless cameras, which involved them standing upright, with their arms abducted to the horizontal, and their feet together and pointing forward. This step was needed to provide a clear and consistent

starting point for tracking body joints. Note that the AKT is the latest depth camera from Microsoft, and it has clear improvements such as a wider field of view, better resolution, and a global shutter that allows for improved performance in sunlight (Microsft Corporation, 2022).

Data Processing and Feature Selection

Task classification can be viewed as multi-staged processing, which includes data collection and processing, feature selection, data labeling, classification stage, and model evaluation (Figure 2.3). The latter stages are each discussed in detail subsequently.

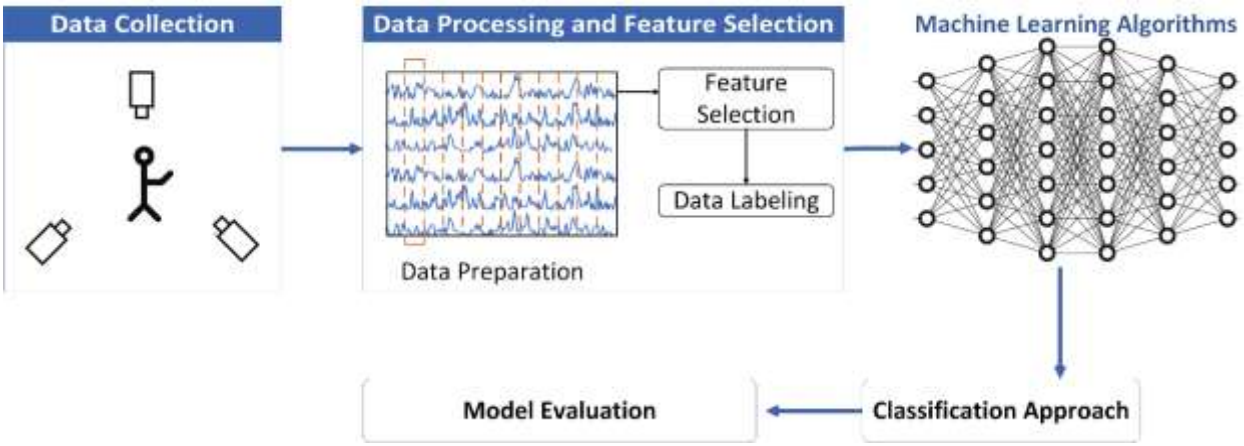

Figure 2.3. Overview of the task classification process.

Data Processing

Data recorded from the markerless camera systems were processed using iPi Motion Capture (MOCAP) Studio. Using this software, 3D body motions were tracked, and key body joints were identified from the video recorded using the iPi Recorder. Seventeen body joints were tracked and stored for further processing – pelvis, lower spine, middle spine, upper spine, neck, and bilateral acromioclavicular, shoulder, elbow, hip, knee, and talus. Tri-axial positions and quaternion joint rotations were extracted for each of the body joints using iPi Biomechanical add-on software. Joint kinematics were low-pass filtered (6 Hz cutoff; 4^{th} order Butterworth;

bidirectional) to remove sensor noise and other artifacts, with the cutoff frequency determined using residual analysis (Winter, 2009). Filtered joint kinematics were then normalized, using the Min-Max-Scaler function (Pedregosa et al., 2011), since machine learning models are sensitive to unscaled data (Zhou et al., 2022). Here, we used the function to transform both training and validation data to a (0, 1) range. All other offline data processing was completed using Python (vs. 3.10.11; https://www.python.org).

Feature Selection

A final processing step involved identifying *features* in the kinematics data. Features are independent measurable properties or characteristics of the data that serve as input for machine learning models. In their basic form, features can be in the form of raw data. However, in some instances, the most informative features are selected while discarding irrelevant or redundant ones using feature selection algorithms (Markovitch & Rosenstein, 2002). Therefore, two types of features were considered here: raw and "informative" features. First, the processed joint kinematics were included as input for model training described below. Specifically, there were 119 raw features (RF), consisting of the tri-axial and quaternion rotations of the 17 body joints. Second, high-dimensional features may lead to longer classification processes and overfitting. Reducing dimensionality, by selecting the most informative features, can simplify machine learning models and help improve model performance and interpretability (Chen et al., 2017). Dimensionality reduction is especially valuable in the context of MMC systems, since occlusions from the body or the environment can lead to data loss and inaccurate pose estimations (Plantard et al., 2017). A filter-based method for feature selection – Minimal-Redundancy-Maximum-Relevancy (mRMR) – was used to select subsets of features from among the entire kinematic feature pool (Peng et al., 2005). The primary goal of this approach is to find a feature subset that

minimizes redundancy between features, while maximizing their relevance to the target variable, and more details on this method have been reported elsewhere (Peng et al., 2005). Since there is no consensus on the number of features that could yield the best model performance, we arbitrarily tested the top 60 (TOP-60) and top 80 (TOP-80) features as input for each machine learning model. Listings of TOP-60 and TOP-80 for each classification category (see section 2.6.3) are provided in Tables 2.6 – 2.11 Appendix A.

Data Labeling

Raw, TOP-60, and TOP-80 features were labeled manually based on the specific tasks performed, using the recorded RBG-D data (ground truth). In each MMH task that involved manipulating a box, the task began when participants touched the box handle and ended when they removed their hands from the box. In the case of cart pushing (Task 6), the task started when participants touched the cart handles and concluded when their hands returned to their sides. Data corresponding to when participants were idle between each task was removed and the remaining data were concatenated (see example in Figure 2.15 Appendix A).

Classification Approach

Machine Learning Algorithms

Recurrent neural network (RNN) architectures were used for task classification based on kinematic measures. RNNs are a variant of an artificial deep learning network, designed to process and recognize patterns in sequential data (Levin, 1990). Unlike other neural networks and machine learning models that process cross-sectional data, RNNs maintain contextual relationships between time series data segments. Three RNN models were included and compared for classifying MMH tasks in this study: 1) Bidirectional Long Short-Term Memory (Bi-LSTM); 2) Gated Recurrent Units (GRU); and 3) Bidirectional Gated Recurrent Units (BGRU). Bi-LSTM is a variant of LSTM that is created by combining two independent RNNs, specifically backward and forward LSTM models (S. Hochreiter & J. J. N. c. Schmidhuber, 1997). Bi-LSTM models resolve the vanishing gradient problem prevalent in RNNs – which refers to the situation where the gradients of the loss function become extremely small, causing the network weights to update very slowly (S. Hochreiter & J. Schmidhuber, 1997; Jozefowicz et al., 2015) – by using a gating mechanism. A Bi-LSTM has three gates – input, forget, and output – that modify Bi-LSTM cell states. This type model increases task classification performance, since it combines input from the past to the future, a forward LSTM, and from the future to the past, a backward LSTM (Graves et al., 2013; Yang et al., 2020; Zhou et al., 2022). A GRU model is a simplified version of the LSTM model; thus, a GRU has fewer parameters, enabling it to "understand" long-term dependencies more easily. Finally, a BGRU is a GRU model that is enhanced by a two-layer structure. This structure provides the model output layer with complete contextual information of the input data at every instance. In brief, the input data are passed through feedforward and backward GRU networks, and the outputs of the two layers are

connected in the same output later. Some studies have shown that BGRU models are suitable for classification problems. For examples, BGRU models were used for human identification (Lynn et al., 2019) and for dialog intent classification (Wang et al., 2020).

Model Architecture

The model architecture consisted of an input layer, an RNN model layer (i.e., Bi-LSTM, GRU, and BGRU), a masking layer, dropout functions, dense layers, and an output layer (Figure 2.4). All implementations were done using Python (3.10.11; https://www.python.org).

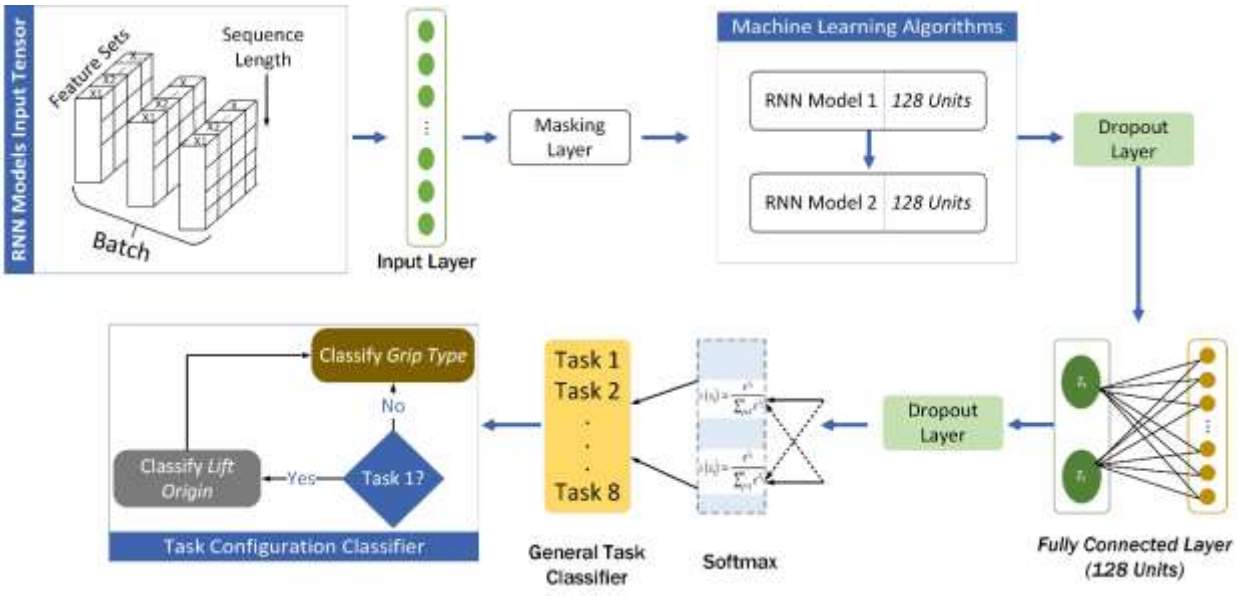

Figure 2.4. Overall architecture of recurrent neural network models for classification.

The input layer of each RNN model was designed to take a multidimensional matrix that is determined based on the sequence length, feature dimension, and batch size. Sequence refers to the ordered set on input data points processed sequentially over time steps, which was fixed at 30 here, corresponding to 1 second of data at the 30 Hz sampling rate. This sequence length was selected to account for tasks that have short duration, and it has been used in previous human activity recognition studies to improve recognition of short duration tasks (Bulling et al., 2014; Capela et al., 2015). Since the duration of each task varied, the sequence length was zero-padded

(Figure 2.14 Appendix A)to obtain a consistent sequence length (e.g., see (Dwarampudi & Reddy, 2019). A masking layer was added to each of the RNN models to mitigate padding effects. Specifically, this layer skips any sequences that have the special masking value. Feature dimension is the number of features, based on the dimensions of the features generated earlier. Raw, TOP-60, and TOP-80 features were 119-, 60- and 80-dimensional vectors, respectively. Finally, the batch size was set to 64, which was determined as the optimal size during initial testing. The input layer was connected to each of the RNN model architectures, consisting of a dropout function, a dense layer, an optimizer, a loss function, and an output classification layer. Our output classification layer used a many-to-one architecture, wherein a single output is produced from more complex input data.

Classification decisions were made *sequence-to-sequence* (i.e., one for each second of data). Each input sequence was classified independently, with the entire sequence being considered at every time step (Figure 2.5). In this context, a time step corresponds to each discrete unit of time within a sequential input data sequence, and for the present study this discrete time unit was 1 second. Using time steps enables RNNs to capture temporal dependencies and facilitate the modeling of dynamic input sequences. Time steps also enable the RNN models to maintain contextual relationships between past and present information, making RNNs well-suited for tasks involving time series data.

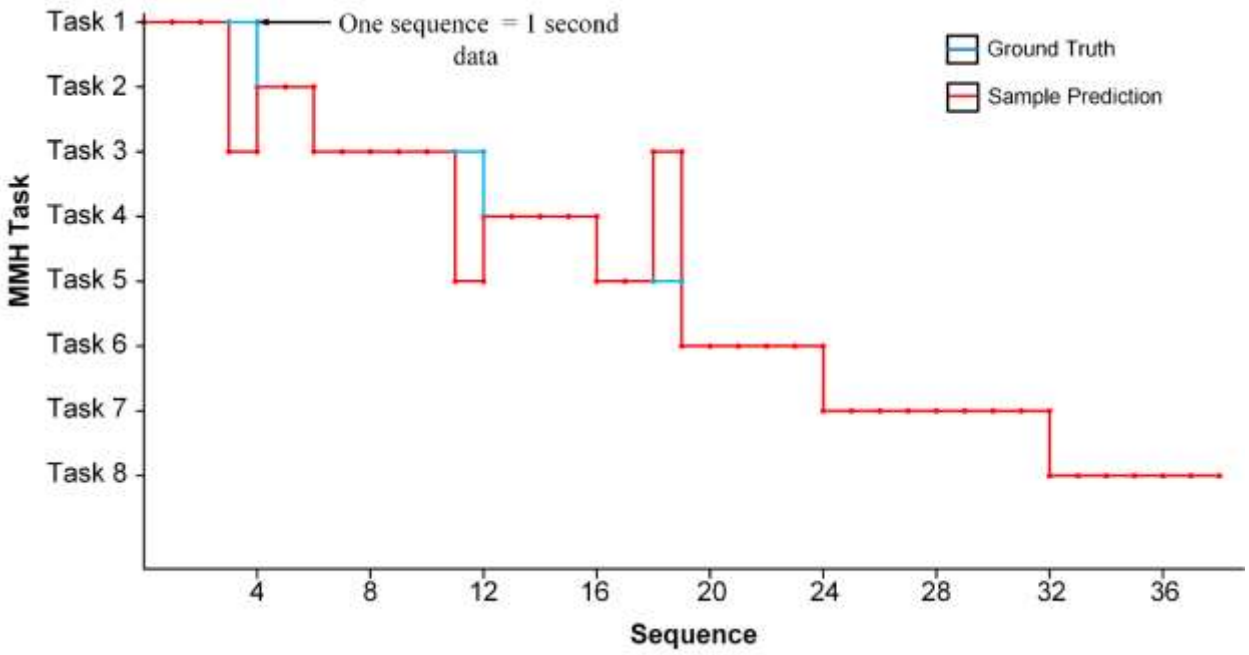

Figure 2.5. An example of continuous classification using a sequence-to-sequence approach. A Bi-LSTM model and Raw feature set were used in this example. Each dot represents sequence, blue lines indicate ground truth, and red lines indicate output from the model.

Classification Scheme

Two categories of RNN models were developed – *general task classifiers* and *task configuration classifiers.* General task classifiers were trained and evaluated using the extracted features to classify the eight simulated MMH tasks. Task configuration classifiers, in contrast, classified the different configurations within relevant simulated tasks (i.e., Lift Origins and/or Hand Configuration). We developed two categories of RNN models because developing a single classifier (e.g., multi-stage classifier) across all tasks and task configurations could result in convergence difficulties.

Model Training, Validation, Hyperparameters, and Evaluation

For each category of RNN model, experimental data were used as input to the models. We used a leave-one-subject-out approach to train and validate the RNN models; this approach is a special case of cross-validation, in which each subject is considered as a "fold". Thus, data from 35 participants were used for training, while data from one participant was used for validation, then this process was repeated 36 times (i.e., 36-fold cross-validation). While the leave-one-subject-out method has been used in past work (e.g., Kim & Nussbaum, 2014; Porta et al., 2021), it often results in high variance in accuracy since participants can perform the same tasks in different ways (Jordao et al., 2018). This method replicates real-world training and testing, though, wherein the model is trained offline using known subject data and then tested on an unseen subject (Jordao et al., 2018).

RNN models have several hyperparameters that are used to control the learning process and model complexity (Probst et al., 2019). Hyperparameter values were determined here using an empirical tuning process involving manual adjustments to each parameter and subsequent evaluation of the resulting model performance. Specific values tested and the final values adopted are presented in Table 2.5 Appendix A. Performance of the RNN models was assessed using four common metrics – accuracy, Precision, recall, and F1-score – which are defined in Appendix B.

Statistical Analyses

To account for the differing number of independent variables potentially affecting each performance metric, a two-stage analyses of variance of variance (ANOVA) approach was used. First, separate two-way repeated-measures ANOVAs were used to assess the effects of *Feature set* and *RNN model* on accuracy. The latter factor had two levels to represent general task

classifiers vs. task configuration classifiers. These models were used initially since accuracy could only be computed across all of the simulated MMH tasks. Second, separate three-way repeated-measures ANOVAs were used for outcomes of precision, recall, and F1-score. For general task classification, the independent variables were *Feature set, RNN model,* and *MMH task.* Two sets of models were used for the task configuration classifiers, with different independent variables*:* a) *Feature set, RNN models, and Hand Configuration*; and b) *Feature set, RNN models, and Lift Origin*. For each ANOVA model, biological sex (*Sex*) was included as a blocking effect. Significant interaction effects were explored using simple-effects testing and Tukey's HSD *post hoc* paired comparisons procedure. All statistical analyses were performed with JMP Pro 16 (SAS, Cary, NC) using the restricted maximum likelihood (REML) method. Parametric model assumptions were verified, and statistical significance was determined when $p < 0.05$. Summary data are reported as least-square means (with 95% confidence intervals) from statistical model fits.

2.4 Results

ANOVA results are summarized in Tables 2.1 – 2.4 Appendix A, and Figures 2.21 – 2.23 Appendix A provides confusion matrices for each category of RNN model. Sample confusion matrices are shown below (Figures 2.6, 2.8 and 2.10) using results representing the best classification performance. There were significant main or interactive effects of *RNN model* and *Feature set* for all classification performance metrics. More detailed results are provided below.

MMH Task Classification Performance

Accuracy: Mean accuracy was significantly affected by *RNN model* and *Feature set*. Significantly smaller accuracy was obtained using the GRU model (~91%), vs. using either the BGRU and Bi-LSTM models, though the magnitude of the difference was quite small (i.e., ~91

vs. ~92%, respectively). Using the TOP-60 feature set led to significantly poorer mean accuracy compared to the TOP-80 and RF feature sets, though again the difference was rather small (~90 vs. ~93%, respectively). A sample confusion matrix using a Bi-LSTM model and the TOP-80 feature set is shown in Figure 2.6; in this case, the model performed well on most tasks, with an overall accuracy of ~93%.

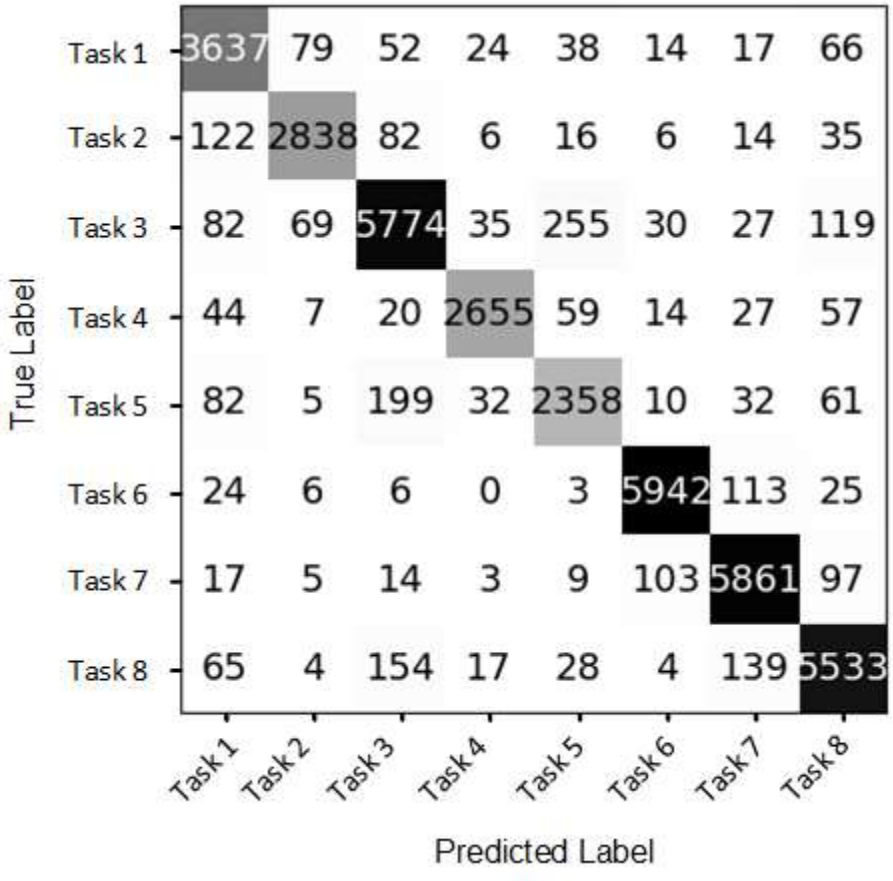

Figure 2.6. Overall confusion matrix from using a Bi-LSTM model using the TOP-80 feature set, representing the best performance when classifying the MMH tasks. For this and other confusion matrices, cells on the main diagonal indicate correct classifications, and lighter shades indicate classes with larger misclassifications.

Precision: *RNN model*, *MMH task*, and *Feature set* main effects, and *RNN model* × *MMH task*, *Feature set* × *MMH task*, *Sex* × *MMH task* interaction effects, were all significant (Table 2.2 Appendix A). Precision was comparable between the different RNN models for most of the MMH tasks and at relatively high levels (i.e., ~87-97%; Figure 2.7). However, using the GRU model led to precision that was up to ~3% poorer compared to the Bi-LSTM and BGRU models. All simple effects were significant, except for the effect of *RNN model* in Tasks 2 ($p = 0.37$), 4 ($p = 0.29$), 6 ($p = 0.37$), and 7 ($p = 0.36$) conditions. Using the TOP-60 feature set consistently led to precision that was 2-8% worse than when using the TOP-80 and RF feature sets, depending on the specific MMH task (Figure 2.7). All simple effects were significant, except for the effect of *Feature set* in Tasks 3 ($p = 0.87$), 6 ($p = 0.46$), and 7 ($p = 0.36$) conditions. Precision in some tasks was ~4% smaller among males, though there were no significant paired differences between sexes for any of the tasks (Figure 2.20 Appendix A). Simple effects were only significant for the effects of *MMH task* among males ($p = <0.0001$) and females ($p = <0.0001$).

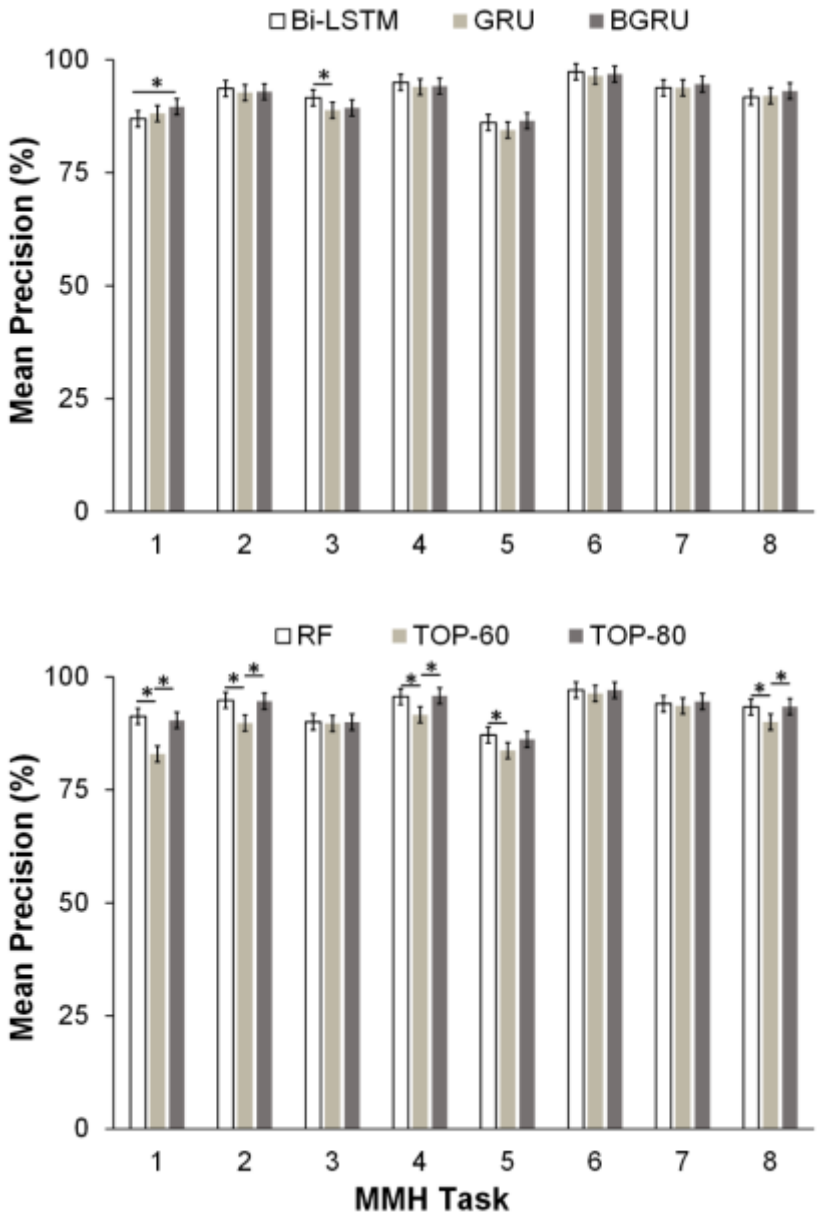

Figure 2.7. Interaction effects of *RNN Model* × *MMH task* (top) and *Feature set* × *MMH task* (bottom) on classification precision. For this and other figures below, error bars indicate 95% confidence intervals, and the symbol * indicates a significant difference between pairs of means.

Recall and F1-score: *RNN model*, *MMH task*, and *Feature set* main effects, and *Feature set* × *MMH task* and *Gender* × *MMH task* interaction effects, were all significant for both recall and F1-score (Figures 2.16 and 2.17 Appendix A). One consistent observation was that using the

TOP-60 features led to significantly reduced recall and F1-scores compared to TOP-80 and RF, by up to 7% depending on the specific MMH task. All simple effects were significant, except for the effect of *Feature set* in Tasks 6 ($p = 0.51$), 7 ($p = 0.20$) conditions. There were no significant paired differences between sexes or between MMH tasks. Similar to precision, though, these metrics were somewhat poorer among males, by ~1–3% depending on the specific MMH task. Simple effects were only significant for the effects of *MMH task* among males ($p = <0.0001$) and females ($p = <0.0001$).

Classifying Hand Configuration

Accuracy: There was a significant main effect of *Feature set* on accuracy, which was larger when using TOP-80 and RF feature sets (TOP-80 and RF = 94%) vs. TOP-60 (81%). An example confusion matrix illustrating a hand configuration classifier output based on Bi-LSTM and TOP-80 feature set is shown in Figure 2.8; in this case, the model had worse performance in classifying Broad and Narrow hand configurations.

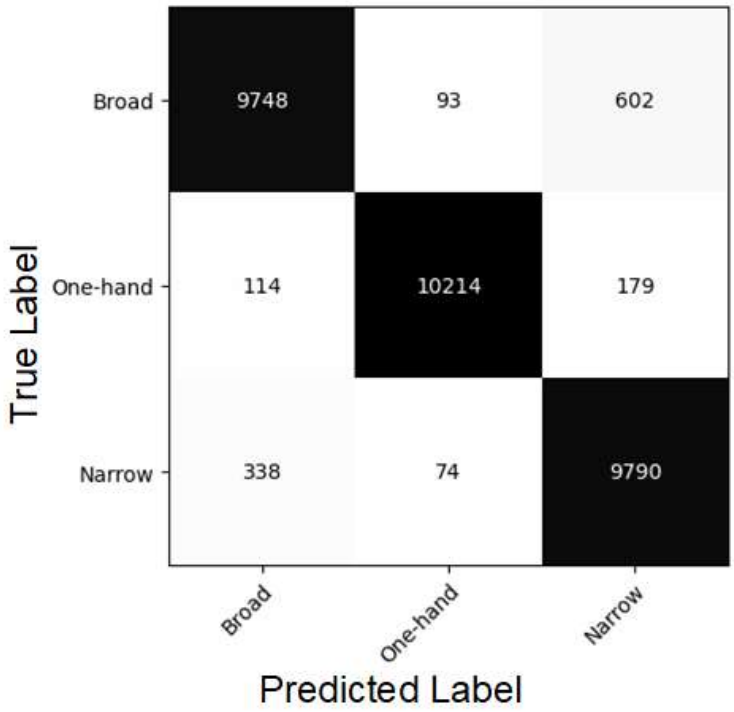

Figure 2.8. Overall confusion matrix for Bi-LSTM model using the TOP-80 feature set, representing the best performance when classifying hand configuration.

Precision: There were significant *RNN model*, *Feature set*, *Hand Configuration* main effects, as well as *Feature set* × *Hand Configuration,* and *Hand Configuration* × *Sex* interaction effects. Using GRU model led to significantly smaller precision, compared to using the BGRU and Bi-LSTM models, yet the magnitude of the difference was small (GRU = 89%, BGRU= 90%, Bi-LSTM = 91%). Across hand configurations, using the TOP-60 feature sets led to significantly less precision (by ~18%), compared to using the TOP-80 and RF feature sets, except for the one-hand configuration (Figure 9). All simple effects were significant except for the effects of *Feature set* in one hand ($p = 0.089$) condition. For a given *Sex*, the narrow hand configuration led to significantly less precision (by up to 11%), compared to either the broad or one-hand configurations (Figure 9). Simple effects were only significant for the effect of *Hand Configuration* among males ($p = <0.0001$) and females ($p = <0.0001$).

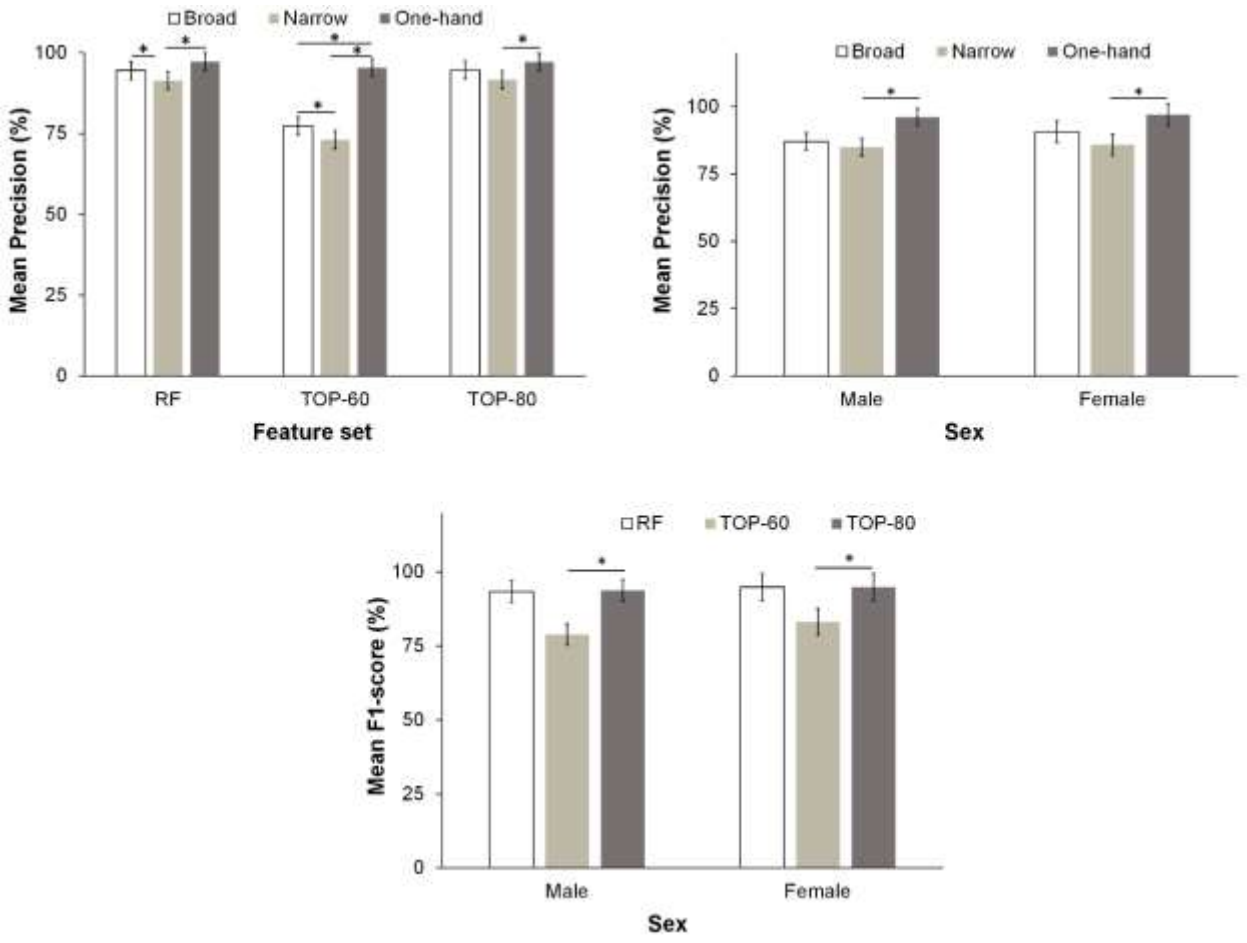

Figure 2.9. Significant interaction effects of *Feature set* × *Hand Configuration* on precision (left), *Hand Configuration* × *Sex* on precision (middle), and *Feature set* × *Sex* on F1-score (bottom).

Recall: There were significant *RNN model*, *Feature set*, *Hand Configuration* main effects and *Feature set* × *Hand Configuration*, *Hand Configuration* × *Gender*, and *Feature set* × *Gender* × *Hand Configuration* interaction effects on recall. Compared to the GRU model, using a Bi-LSTM led to significantly better recall (GRU = 88.9%, Bi-LSTM = 90.1%). Using the TOP-60 feature set led to significantly poorer recall in all three hand configurations (by 7-20%), compared to TOP-80 and RF (Figure 2.18 Appendix A). All simple effects were significant, with

differences between *Hand Configuration* significant for both male (p = <0.0001) and female (p = <0.0001) and differences between *Sex* were significant for all *Hand Configuration* (p = <0.0001).

F1-score: *RNN model*, *Feature set*, and *Hand Configuration* main effects, and *Feature set* × *Hand Configuration*, and *Feature set* × *Sex* interaction effects, were significant. Compared to the GRU model, using the Bi-LSTM model led to significantly larger F1-score up to 2%. Similar to recall, using the TOP-80 feature set led to significantly higher F1-score (93-97%), across all hand configurations vs. RF and TOP-60 feature sets (Figure 2.19 Appendix A). All simple effects were significant, with difference between *Hand Configuration* significant for all *Feature set* (p = 0.0002), and differences between *Feature set* were significant for all *Hand Configuration* (p = 0.0015). For a given sex, using TOP-80 feature set led to significantly larger F1-score by ~12-15% compared to TOP-60 feature set (Figure 2.9). Simple effects were only significant for the effects of *Hand Configuration* among males (p = <0.0001) and females (p < 0.0001).

Classifying Lift Origin

Accuracy: There were significant *RNN model*, and *Feature set* main effects on accuracy. Using GRU model led to significantly poorer accuracy (~80%) vs. using Bi-LSTM (~83%) and BGRU models (~84%). Using the TOP-60 feature set led to significantly smaller accuracy (81%), compared to RF (83%) feature set. An example confusion matrix illustrating a lift origin classifier output based on BGRU and RF feature set is shown in Figure 2.10. In this case, the model performs well in classifying both lift origins with an overall accuracy of ~84%.

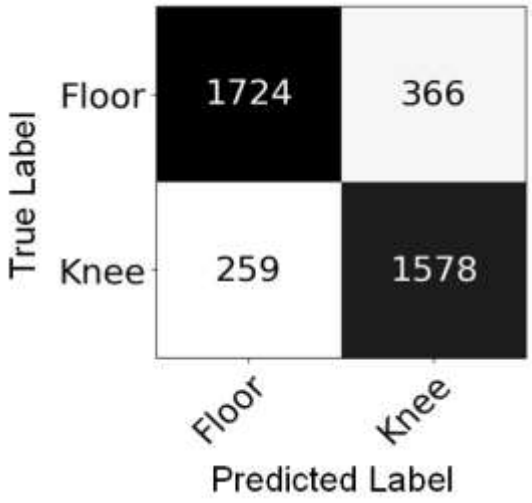

Figure 2.10. Overall confusion matrix for BGRU model using the RF feature set, representing the best performance when classifying *Lift Origin*.

Precision and Recall: There were significant main and interaction effects of *RNN model*, *Lift Origin*, and *Gender* × *Lift Origin.* Using the GRU model led to significantly poorer precision (81%) vs. using the Bi-LSTM (84%) and BGRU (85%). Similarly, using GRU model led to significantly smaller recall (80%) vs using the Bi-LSTM (83%) and BGRU (84%). Males exhibited smaller precision (up to 5%) when lifting from the knee origin compared to the floor origin. Simple effects were only significant for the effect of *Lift Origin* among males (p = <0.0001) and females (p = 0.039). Also, recall was significantly smaller among males (up to 5%) when performing lifts from the floor origin compared to the knee origin (Figure 2.11). Simple

effects were only significant for the effect of *Lift Origin* among males (p = <0.0001).

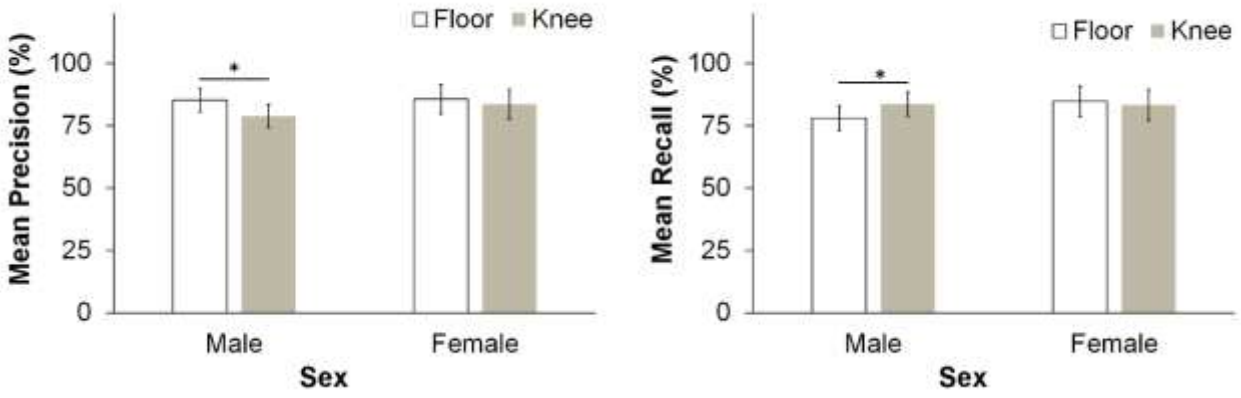

Figure 2.11. Significant *Sex* × *Lift Origin* interaction effects on precision (left) and recall (right) for classifying *Lift Origin*.

F1-score: *RNN model*, *Feature set*, *Lift Origin* main effects and *Sex* × *RNN model* interaction effect were significant for F1-score. Using RF and TOP-80 feature sets led to significantly larger F1-score (RF = 83%, TOP-80 = 82%), compared to TOP-60 feature set (80%). Lifting from the knee origin condition led to significantly poorer F1-score (81%) compared to lifting from the floor origin (83%). Using the BGRU model led to significantly smaller F1-score by 5% for males and 3% for females compared to the GRU model (Figure 2.12). Simple effects were only significant for the effect of *Lift Origin* among males (p = <0.0001) and females (p = 0.0032).

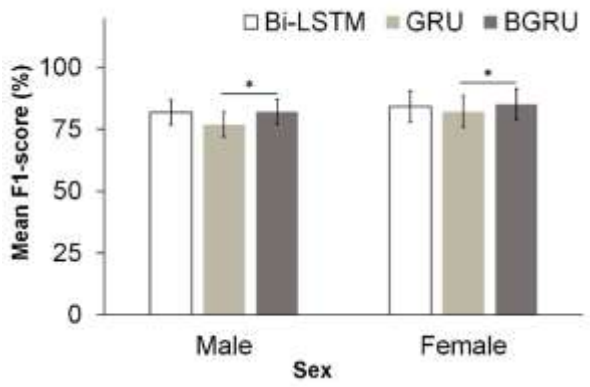

Figure 2.12. Significant *Sex* × *RNN model* interaction effects on F1-score for classifying *Lift Origin*.

2.5 Discussion

Combining MMC and RNN models could contribute to efficient occupational exposure assessment, specifically by helping to automate aspects of such assessments and by using ambient vs. wearable sensors. Using data from an MMC system, our goal here was to investigate the use of different RNN models and feature sets to classify diverse MMH tasks and specific task conditions. Across the MMH task types and feature sets, mean precision, recall, and F1-score values were, in our opinion, good to excellent, with each metric on the order of 85 – 97%. Performance was quite high for the more specific goal of classifying hand configuration, with mean precision, recall, and F1-scores of up to 96-97%. Performance of classifying lift origin, in contrast, varied depending on the feature sets used and between males and females. The following discussion addresses these effects in more detail, including the differential effects of the three models, features, dependencies on MMH task and task conditions, and differences related to sex.

Effects of Machine Learning Algorithms and Feature sets on MMH Task Classification

MMH task classification performance varied depending on the MMH task and feature set. Compared to the TOP-80 feature set, using the TOP-60 feature set substantially reduced mean precision and recall, by up to 7% depending on the specific MMH task. Recall that we used the mRMR algorithm to select the top 60 and 80 features from among the raw features. Upon inspection (Tables 2.6 – 2.8 Appendix A), the TOP-60 features consisted of tri-axial positions and quaternion joint rotations of the left forearm and shoulder, and the bilateral thigh, shin, and foot. TOP-80 features comprised tri-axial positions and quaternion joint rotations of the right forearm and shoulder, along with tri-axial information on the left clavicle and quaternion joint rotations of the right clavicle. Given these differences, arm dominance might explain why the TOP-80 feature set outperformed the TOP-60 (aside from simply having additional input data). An individual's dominant arm plays a crucial role in determining trajectory direction and speed, while the nondominant arm is critical for accurate positioning (Wang & Sainburg, 2007). Thus, including right-arm kinematics may have improved machine learning algorithms, improving classification performance.

Across the simulated MMH tasks, mean precision in task classification varied between 87 and 97%, but depended on the specific RNN model and MMH task (Figure 2.7). Performance of machine learning models is often found to depend on the specific model and task type (Barazandeh et al., 2017; Gong et al., 2011; Khosrowpour et al., 2014; Luo et al., 2018; Yang et al., 2016). In such studies, a range of construction and MMH tasks were simulated, including drilling, sawing, and lifting, and reported results showed that machine learning algorithms often misclassify MMH tasks, especially when two tasks involve similar body kinematics (e.g., drilling vs. sawing, lifting from knee vs. floor levels). We similarly found that machine learning

algorithms misclassified tasks with similar kinematics. Tasks involving pulling (Task 5) and carrying (Task 3) were the most "confused" tasks, irrespective of the machine learning algorithm (Figure 2.21). Notably, most of the current misclassifications occurred at the completion of Task 5 and Task 3 (Figure 2.5). After reviewing all labeled videos, we suspect that the most important reason for such misclassification is the similarity in body kinematics between Task 5 and Task 3, especially at the end of these two tasks.

Another possible reason for misclassification is that Task 5 had a class imbalance, which occurs when the instances of one class outnumber the instances of other classes (Guo et al., 2008). In some cases, precision and recall for a minority class may be adversely affected, causing the model to incorrectly classify minority class instances (Ali et al., 2013). In our dataset, Task 5 had the fewest sequences (N=2,779), while Task 3 had the most sequences (N=6,391). Thus, our machine learning algorithms could have developed a bias toward predicting the majority class. Generating synthetic data using heuristic oversampling to increase Task 5 occurrences could improve future classification performance. More generally, heuristic oversampling techniques, such as the synthetic minority over-sampling technique (Chawla et al., 2002), should be explored to improve the performance of machine learning algorithms when used for task classification.

Classifying Detailed Aspects of MMH Tasks

Precision of hand configuration classifiers (i.e., precision, recall, and F1-score) were substantially poorer for broad and narrow hand configurations (broad = 88.8%; narrow = 85.4%;), compared to one hand configuration (precision = 96.5%). The smaller precision of hand configuration classifiers in broad and narrow hand configurations was not surprising due to the similarity in kinematics, which differed from the one-hand configuration. Similar motion pattern in broad and narrow hand configurations could have led to redundant or correlated features,

potentially affecting the ability of the models to generalize to new information during evaluation (Kim & Nussbaum, 2014).

Compared to Bi-LSTM and BGRU models, the GRU model demonstrated substantially lower performance on both hand configuration and lift origin classifications. Notably, using the GRU model reduced mean precision by up to 3% and 4% for hand configuration and lift origin classifications, respectively. Despite its computational efficiency due to simplified gates, GRU model process information only in the forward direction (i.e., past to future). Capturing contextual information in both forward and backward directions could improve model performance for scenarios in which understanding the entire sequence is crucial for making accurate classifications (Alawneh et al., 2020). We suspect that Bi-LSTM and BGRU models outperformed the GRU model due to their ability to process contextual information in bilateral directions.

Differences in Classifying MMH Task and Hand Configuration with Biological Sex

There were differences related to sex in our study. For example, precision in some MMH tasks was 4% smaller among males, though not significant. Also, mean recall of task configurations was generally smaller (up to 5%) when classifying lift origins among males. Frankly, it is unclear to us why performance was smaller among males. A responsible machine learning algorithm should not be biased towards one or another group of people (Arrieta et al., 2020). Machine learning algorithms, though, can harbor hidden biases that emerge when they are used in the real world. These so-called latent biases, often inherited from the data used to train and validate machine learning algorithms, pose a critical challenge to fairness and equity. For example, though not a direct comparison, effects of sex-bias in training data led to an accuracy

gap in emotion recognition between male and female test sets using data from RGB camera (Domnich & Anbarjafari, 2021).

One reason for sex-related differences in our result could have been that larger kinematic variability among males might have contributed to the relatively poorer ability to classify lift origin and some of the MMH tasks. Visual inspection of our video data showed that males and females employed different postures, especially when performing Tasks 1, 5 and 8. Earlier studies have also shown males and females exhibited different kinematic when performing MMH tasks (Lindbeck & Kjellberg, 2001; Plamondon et al., 2014). For example, Lindbeck and Kjellberg (2001) found that males exhibited larger kinematic variability (up to 20° trunk and knee flexion) when performing MMH tasks.

Limitations

Several limitations of this study should be mentioned. First, participants were relatively young (i.e., 18 – 39 years old). Therefore, caution should be taken in generalizing the results to other populations, such as older individuals or those with musculoskeletal disorders, and future work is needed with larger and more diverse samples. Second, feature selection methods can substantially affect classification performance (Preece et al., 2009). We used a filter-based approach to select subsets of features from among raw kinematic features. While our approach led to high performance (up to 97%) in classifying MMH tasks, the method used here – mRMR – did not consider the temporal (time series) nature of the data. Future work should consider using feature selection methods that preserve temporal information. Third, we used machine learning algorithms that consist of non-linear structures, making them highly non-transparent in arriving at their decisions. Understanding the process or reasoning behind predictions is crucial – a concept known as Explainable AI. Integrating Explainable AI into machine learning can facilitate

verification of predictions, systematic identification of potential flaws and biases, and the understanding of the underlying decision-making processes of machine learning algorithms (Samek et al., 2017). Fourth, we standardized some conditions such as participant clothing, lighting source and brightness, and camera positioning, which might have improved the ability to track whole-body kinematics using MMC. Thus, future work should explore the effects of varying these conditions.

Our tasks included several components commonly seen in the workplace that could lead to environmental occlusions, and three MMC systems were used to monitor whole-body kinematics to reduce the impact of such occlusions. Kotsifaki et al. (2018) found that increasing the number of cameras could enhance tracking of whole-body kinematics. Nevertheless, there will likely be practical constraints on the number of cameras that is feasible in practice, such as due to workflow and space restrictions. Therefore, future work should explore the possibility of using a single MMC or different configurations of two MMCs for physical exposure assessment. The MMC systems employed here used depth sensing through the 3D time-of-flight principle (Microsoft Corporation, 2022). Our machine learning algorithms were trained on body kinematics extracted from RGB-D cameras. Recently, though, video surveillance cameras (plain cameras) are becoming more widely used in occupational settings (e.g., manufacturing industry) to enhance worker safety and to track productivity (Cocca et al., 2016; Kostal et al., 2022; Xu et al., 2015). We suggest potential value in investigating the feasibility of using plain cameras as an alternative approach to tracking body kinematics, which could be more efficient and less costly.

Practical Applications of Our Findings

For practical application, we suggest that the GRU model and TOP-80 feature set could be used as the base approach for MMH task classification using MMC. Using the GRU model

generally resulted in less accuracy (by 1%) compared to the Bi-LSTM and BGRU, though this difference seems of limited practical relevance. Further, training any of the RNN models with the TOP-80 feature set yielded a mean accuracy comparable to that using the RF features and roughly 2% better than when using the TOP-60 feature set. Notably, though, the mean epoch training time (the number of times that the model works through the entire training dataset) using the GRU model with the TOP-80 feature set was 19-35% less than when using the other models. This faster training time likely stemmed from the fact that the GRU model has a simplified architecture with two gates and fewer parameters, making it more computationally efficient for large datasets (Khandelwal et al., 2016). Moreover, using a streamlined feature set (vs. the RF feature set) helps promotes a more interpretable model and contributes to improved generalization performance, by reducing the likelihood of capturing noise or irrelevant patterns in the data (Chen et al., 2017; Xue et al., 2015).

Using RNN-based models and MMC is a novel approach to quantifying physical exposure assessment; as such, some comments about using this approach are warranted. One advantage is that using MMC provides a non-intrusive method for obtaining whole-body kinematics, and classification performance appears comparable to earlier results reported using inertial measurement units (IMUs). For example, earlier work that used Bi-LSTM and IMUs to classify MMH tasks reported a mean precision of 92% (Porta et al., 2021), while we achieved a mean precision of 91-92%. Moreover, one limitation of earlier work is that pushing and pulling tasks were often misclassified when using wearable sensors, by up to 5% (Kim & Nussbaum, 2014; Mokhlespour Esfahani, 2018; Porta et al., 2021). It is important to distinguish between pulling and pushing, since they impose different loads on the low back (Hoozemans et al., 2004).

Our approach was able to distinguish between pulling and pushing tasks with ambient sensing (i.e., MMC), with reasonable precision and recall (Figure 5).

Using a sequence-to-sequence classification approach enhances performance of RNN model by capturing temporal dependencies and enabling RNN models to incorporate contextual information from the entire input sequence, contrasting with a sample-by-sample method (Yin et al., 2017).

2.6 Conclusions

Effectively measuring and monitoring physical exposures in the workplace is critical to assessing and controlling the risk of WMSDs. While several tools are available, these tools are often limited in accuracy and scope and their use can be quite resource-intensive. We evaluated the use of ambient sensors here, specifically MMC, together with machine-learning models, to classify among eight diverse MMH tasks and several specific task conditions. Using the MMC data, task classification performance was quite good using several classification models and input feature sets: mean precision, recall, and F1-score were 85 – 97% across the MMH tasks. Performance for classifying hand configuration was quite high, with mean precision, recall, and F1-scores of up to 97%. In contrast, performance in classifying lift origin varied substantially depending on feature sets and between males and females. Overall, our findings indicate that the proposed approach has the potential to quantify MMH tasks exposures, offering a balance of simplicity and non-intrusiveness in exposure assessments. Future work will be needed, though, to assess the ability of the proposed method among diverse workers and in real working conditions.

2.7 References

Alawneh, L., Mohsen, B., Al-Zinati, M., Shatnawi, A., & Al-Ayyoub, M. (2020). A comparison of unidirectional and bidirectional lstm networks for human activity recognition. 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops),

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2013). Classification with class imbalance problem. *Int. J. Advance Soft Compu. Appl*, *5*(3), 176-204.

Andersen, J. H., Kaergaard, A., Mikkelsen, S., Jensen, U. F., Frost, P., Bonde, J. P., Fallentin, N., & Thomsen, J. F. (2003). Risk factors in the onset of neck/shoulder pain in a prospective study of workers in industrial and service companies. *Occup Environ Med*, *60*(9), 649-654.

Antwi-Afari, M. F., Li, H., Yu, Y., & Kong, L. (2018). Wearable insole pressure system for automated detection and classification of awkward working postures in construction workers. *Automation in Construction*, *96*, 433-441.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115.

Barazandeh, B., Bastani, K., Rafieisakhaei, M., Kim, S., Kong, Z., & Nussbaum, M. A. (2017). Robust sparse representation-based classification using online sensor data for monitoring manual material handling tasks. *IEEE Transactions on Automation Science and Engineering*, *15*(4), 1573-1584.

Bernard, B. P., & Putz-Anderson, V. (1997). Musculoskeletal disorders and workplace factors; a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back.

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, *46*(3), 1-33.

Burdorf, A., & van Riel, M. (1996). Design of strategies to assess lumbar posture during work. *International Journal of Industrial Ergonomics*, *18*(4), 239-249.

Capela, N. A., Lemaire, E. D., & Baddour, N. (2015). Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. *PLOS ONE*, *10*(4), e0124414.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Chen, Q., Zhang, M., & Xue, B. (2017). Feature selection to improve generalization of genetic programming for high-dimensional symbolic regression. *IEEE Transactions on Evolutionary Computation*, *21*(5), 792-806.

Ciriello, V. M., Snook, S. H., Hashemi, L., & Cotnam, J. (1999). Distributions of manual materials handling task parameters. *International Journal of Industrial Ergonomics*, *24*(4), 379-388.

Cocca, P., Marciano, F., & Alberti, M. (2016). Video surveillance systems to enhance occupational safety: A case study. *Safety Science*, *84*, 140-148.

da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *Am J Ind Med*, *53*(3), 285-323.

David, G. C. (2005). Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup Med (Lond)*, *55*(3), 190-199.
Dempsey, P. G. (1999). Utilizing criteria for assessing multiple-task manual materials handling jobs. *International Journal of Industrial Ergonomics*, *24*(4), 405-416.
Domnich, A., & Anbarjafari, G. (2021). Responsible AI: Gender bias assessment in emotion recognition. *arXiv preprint arXiv:2103.11436*.
Dwarampudi, M., & Reddy, N. (2019). Effects of padding on LSTMs and CNNs. *arXiv preprint arXiv:1903.07288*.
Escorcia, V., Dávila, M. A., Golparvar-Fard, M., & Niebles, J. C. (2012). Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras. Construction Research Congress 2012: Construction Challenges in a Flat World,
Garg, A., Chaffin, D. B., & Freivalds, A. (1982). Biomechanical stresses from manual load lifting: a static vs dynamic evaluation. *IIE transactions*, *14*(4), 272-281.
Gary, A. W., Marras, W. S., & PARNIANPouR, M. (1996). Trunk kinematics of one-handed lifting, and the effects of asymmetry and load weight. *Ergonomics*, *39*(2), 322-334.
Golabchi, A., Han, S., & Fayek, A. R. (2016). A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations. *Canadian Journal of Civil Engineering*, *43*(4), 294-303.
Gong, J., Caldas, C. H., & Gordon, C. (2011). Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models. *Advanced Engineering Informatics*, *25*(4), 771-782.
Graves, A., Jaitly, N., & Mohamed, A. (2013, 8-12 Dec. 2013). Hybrid speech recognition with Deep Bidirectional LSTM. 2013 IEEE Workshop on Automatic Speech Recognition and Understanding,
Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the class imbalance problem. 2008 Fourth international conference on natural computation,
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*, *9*(8), 1735-1780.
Hochreiter, S., & Schmidhuber, J. J. N. c. (1997). Long short-term memory. *9*(8), 1735-1780.
Hoozemans, M. J., Kuijer, P. P. F., Kingma, I., Van Dieën, J. H., De Vries, W. H., Van Der Woude, L. H., Veeger, D. J., Van Der Beek, A. J., & Frings-Dresen, M. H. (2004). Mechanical loading of the low back and shoulders during pushing and pulling activities. *Ergonomics*, *47*(1), 1-18.
Jordao, A., Nazare Jr, A. C., Sena, J., & Schwartz, W. R. (2018). Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art. *arXiv preprint arXiv:1806.05226*.
Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An empirical exploration of recurrent network architectures. International conference on machine learning,
Khandelwal, S., Lecouteux, B., & Besacier, L. (2016). *Comparing GRU and LSTM for automatic speech recognition* LIG].
Khosrowpour, A., Niebles, J. C., & Golparvar-Fard, M. (2014). Vision-based workface assessment using depth images for activity analysis of interior construction operations. *Automation in Construction*, *48*, 74-87.

Kim, S., & Nussbaum, M. A. (2014). An evaluation of classification algorithms for manual material handling tasks based on data obtained using wearable technologies. *Ergonomics*, *57*(7), 1040-1051.
Kostal, P., Prajova, V., Vaclav, S., & Stan, S.-D. (2022). An Overview of the Practical Use of the CCTV System in a Simple Assembly in a Flexible Manufacturing System. *Applied System Innovation*, *5*(3), 52.
Kotsifaki, A., Whiteley, R., & Hansen, C. (2018). Dual Kinect v2 system can capture lower limb kinematics reasonably well in a clinical setting: concurrent validity of a dual camera markerless motion capture system in professional football players. *BMJ Open Sport Exerc Med*, *4*(1), e000441.
Levin, E. (1990). A recurrent neural network: Limitations and training. *Neural networks*, *3*(6), 641-650.
Li, G., & Buckle, P. (1999). Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics*, *42*(5), 674-695.
Liberty Mutual Insurance. (2023). *2023 Workplace Safety Index: The Top 10 Causes of Disabling Injuries*. https://business.libertymutual.com/insights/2023-workplace-safety-index/
Lindbeck, L., & Kjellberg, K. (2001). Gender differences in lifting technique. *Ergonomics*, *44*(2), 202-214.
Luo, H., Xiong, C., Fang, W., Love, P. E., Zhang, B., & Ouyang, X. (2018). Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Automation in Construction*, *94*, 282-289.
Lynn, H. M., Pan, S. B., & Kim, P. (2019). A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access*, *7*, 145395-145405.
Markovitch, S., & Rosenstein, D. (2002). Feature generation using general constructor functions. *Machine Learning*, *49*, 59-98.
Marras, W. S., Cutlip, R. G., Burt, S. E., & Waters, T. R. (2009). National occupational research agenda (NORA) future directions in occupational musculoskeletal disorder health research. *Applied Ergonomics*, *40*(1), 15-22.
Marras, W. S., & Davis, K. G. (1998). Spine loading during asymmetric lifting using one versus two hands. *Ergonomics*, *41*(6), 817-834.
MassirisFernández, M., Fernández, J. Á., Bajo, J. M., & Delrieux, C. A. (2020). Ergonomic risk assessment based on computer vision and machine learning. *Computers & Industrial Engineering*, *149*, 106816.
Microsft Corporation. (2022). *Azure Kinect body tracking joints*. https://learn.microsoft.com/en-us/azure/kinect-dk/about-azure-kinect-dk
Microsoft Corporation. (2022). *Azure Kinect body tracking joints*. Retrieved 10/23/2023 from https://docs.microsoft.com/en-gb/azure/Kinect-dk/body-joints
Mokhlespour Esfahani, M. I. (2018). *Development and Assessment of Smart Textile Systems for Human Activity Classification* Virginia Tech].
Nath, N. D., Akhavian, R., & Behzadan, A. H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Appl Ergon*, *62*, 107-117.
Park, S., Park, J., Al-Masni, M. A., Al-Antari, M. A., Uddin, M. Z., & Kim, T.-S. (2016). A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services. *Procedia Computer Science*, *100*, 78-84.

Pedersen, S. J., Kitic, C. M., Bird, M.-L., Mainsbridge, C. P., & Cooley, P. D. (2016). Is self-reporting workplace activity worthwhile? Validity and reliability of occupational sitting and physical activity questionnaire in desk-based workers. *BMC Public Health*, *16*, 1-6.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226-1238.

Plamondon, A., Lariviere, C., Denis, D., St-Vincent, M., Delisle, A., & Group, I. M. R. (2014). Sex differences in lifting strategies during a repetitive palletizing task. *Applied Ergonomics*, *45*(6), 1558-1569.

Plantard, P., Shum, H. P. H., Le Pierres, A. S., & Multon, F. (2017). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*, *65*, 562-569.

Porta, M., Kim, S., Pau, M., & Nussbaum, M. A. (2021). Classifying diverse manual material handling tasks using a single wearable sensor. *Applied Ergonomics*, *93*, 103386.

Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors—a review of classification techniques. *Physiological measurement*, *30*(4), R1.

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, *9*(3), e1301.

Rezagholi, M., Mathiassen, S. E., & Liv, P. (2012). Cost efficiency comparison of four video-based techniques for assessing upper arm postures. *Ergonomics*, *55*(3), 350-360.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*.

Song, B., Kamal, A. T., Soto, C., Ding, C., Farrell, J. A., & Roy-Chowdhury, A. K. (2010). Tracking and activity recognition through consensus in distributed camera networks. *IEEE Trans Image Process*, *19*(10), 2564-2579.

U.S. Bureau of Labor Statistics. (2021). *Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work*. bls.gov. Retrieved May 7 from https://www.bls.gov/data/home.htm

Wang, J., & Sainburg, R. L. (2007). The dominant and nondominant arms are specialized for stabilizing different features of task performance. *Experimental Brain Research*, *178*, 565-570.

Wang, Y., Huang, J., He, T., & Tu, X. (2020). Dialogue intent classification with character-CNN-BGRU networks. *Multimedia Tools and Applications*, *79*(7), 4553-4572.

Waters, T. R., Dick, R. B., Davis-Barkley, J., & Krieg, E. F. (2007). A cross-sectional study of risk factors for musculoskeletal symptoms in the workplace using data from the General Social Survey (GSS). *Journal of Occupational and Environmental Medicine*, 172-184.

Wells, R., Moore, A., Potvin, J., & Norman, R. (1994). Assessment of risk factors for development of work-related musculoskeletal disorders (RSI). *Appl Ergon*, *25*(3), 157-164.

Winter, D. A. (2009). Biomechanics and motor control of human movement. In (Third ed., pp. 49-50). John wiley & sons.

Xu, X., McGorry, R. W., Chou, L.-S., Lin, J.-h., & Chang, C.-c. (2015). Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking. *Gait & posture*, *42*(2), 145-151.

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, *20*(4), 606-626.

Yang, J., Shi, Z., & Wu, Z. (2016). Vision-based action recognition of construction workers using dense trajectories. *Advanced Engineering Informatics*, *30*(3), 327-336.

Yang, K., Ahn, C. R., & Kim, H. (2020). Deep learning-based classification of work-related physical load levels in construction. *Advanced Engineering Informatics*, *45*, 101104-101104.

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. *arXiv preprint arXiv:1702.01923*.

Yu, Y., Yang, X., Li, H., Luo, X., Guo, H., & Fang, Q. (2019). Joint-Level Vision-Based Ergonomic Assessment Tool for Construction Workers. *Journal of Construction Engineering and Management*, *145*(5), 04019025.

Zhan, K., Ramos, F., & Faux, S. (2012, 5-7 Dec. 2012). Activity recognition from a wearable camera. 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV),

Zhang, C., & Tian, Y. (2012). RGB-D camera-based daily living activity recognition. *Journal of Computer Vision and Image Processing*, *2*(4), 12.

Zhang, H., Yan, X., & Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, *94*, 1-10.

Zhou, G., Aggarwal, V., Yin, M., & Yu, D. (2022). A Computer Vision Approach for Estimating Lifting Load Contributors to Injury Risk. *IEEE Transactions on Human-Machine Systems*.

Appendix A – ANOVA Results and Supplemental Figures

Table 2.1. Overview of ANOVA results assessing the effect of *RNN model*, *Feature set*, and *Sex* on the accuracy metric for classifying MMH task, hand configuration, and lift origin. Entries are F values (p values), and significant effects are highlighted in bold font.

Effect	MMH Task Classification	Hand Configuration Classification	Lift Origin Classification
RNN Model (M)	**9.80 (<.0001)**	2.59 (0.0769)	**20.86 (<.0001)**
Feature (F)	**112.69 (<.0001)**	**144.61 (<.0001)**	**5.61 (0.0041)**
Gender (G)	0.95 (0.3357)	0.70 (0.4099)	0.84 (0.3664)
F × M	0.41 (0.8002)	0.07 (0.9915)	0.74 (0.5684)
G × M	1.26 (0.2856)	0.16 (0.8501)	1.97 (0.1408)
G × F	0.04 (0.9575)	1.76 (0.1732)	0.13 (0.8810)
F × M × G	0.52 (0.7243)	0.18 (0.9509)	0.30 (0.8765)

Table 2.2. Summary of ANOVA results for the effects of *RNN model*, *MMH task*, *Feature set*, and *Sex* on **MMH task classification performance metrics**: precision, recall, and F1-score. Entries are F values (p values), and significant effects are highlighted in bold font.

Effect	Precision	Recall	F1-score
RNN Model (M)	**6.95(0.001)**	**5.16(0.0058)**	**11.60(<.0001)**
Feature (F)	**110.24(<.0001)**	**67.71(<.0001)**	**156.90(<.0001)**
MMH Tasks (T)	**167.67(<.0001)**	**124.01(<.0001)**	**239.55(<.0001)**
Gender (G)	0.82(0.3707)	1.162(0.2885)	1.05(0.3129)
F × M	0.45(0.7723)	0.12(0.9752)	0.37(0.8334)
M × T	**2.93(0.0002)**	1.42(0.1329)	0.96(0.4882)
F × T	**9.55(<.0001)**	**4.07(<.0001)**	**8.11(<.0001)**
G × M	1.05(0.3488)	0.76(0.4678)	1.74(0.1754)
G × F	0.12(0.889)	0.01(0.9961)	0.03(0.971)
G × T	**8.91(<.0001)**	**3.79(0.0004)**	**4.66(<.0001)**
M × F × T	0.69(0.8845)	0.23(1)	0.29(0.9999)
M × F × G	0.33(0.8612)	0.51(0.7294)	0.86(0.4866)
M × T × G	0.77(0.7083)	0.36(0.9843)	0.52(0.9233)
F × T × G	1.30(0.1977)	0.41(0.9714)	0.90(0.5547)

Table 2.3. Summary of ANOVA results for the effects of *RNN model*, *Hand Configuration*, *Feature set*, and *Sex* on **hand configuration classification performance metrics**: precision, recall, and F1-score. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

Effect	Precision	Recall	F1-score
RNN Model (M)	**7.24(0.0008)**	**3.79(0.023)**	**5.61(0.0038)**
Feature (F)	**333.46(<.0001)**	**207.43(<.0001)**	**322.51(<.0001)**
Hand Configuration (H)	**209.50(<.0001)**	**85.00(<.0001)**	**137.65(<.0001)**
Gender (G)	0.56(0.46)	0.68(0.4138)	0.69(0.4128)
F × M	0.07(0.9904)	0.08(0.9881)	0.25(0.9083)
M × H	0.10(0.984)	0.83(0.5038)	0.58(0.6744)
F × H	**60.85(<.0001)**	**28.78(<.0001)**	**45.96(<.0001)**
G × M	0.54(0.5829)	0.25(0.7799)	0.36(0.6998)
G × F	2.33(0.0975)	2.24(0.1068)	**3.93(0.0199)**
G × H	**3.96(0.0194)**	**8.13(0.0003)**	1.50(0.224)
M × F × H	0.17(0.9943)	0.46(0.8843)	0.23(0.9853)
M × F × G	0.25(0.9073)	0.25(0.9092)	0.55(0.7029)
M × H × G	0.07(0.99)	0.06(0.9925)	0.17(0.9561)
F × H × G	0.30(0.8779)	**3.36(0.0096)**	1.42(0.2262)
M × F × H × G	0.35(0.946)	0.15(0.9966)	0.06(0.9999)

Table 2.4. Summary of ANOVA results for the effects of *RNN model*, *Lift Origin*, *Feature set*, and *Sex* on **lift origin classification performance metrics**: precision, recall, and F1-score. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

Effect	Precision	Recall	F1-score
RNN Model (M)	**15.78(<.0001)**	**7.59(0.0006)**	**29.19(<.0001)**
Feature (F)	2.85(0.0589)	1.96(0.142)	**9.63(<.0001)**
Lift Origin (L)	**45.29(<.0001)**	**5.19(0.0231)**	**5.47(0.0197)**
Gender (G)	0.49(0.4879)	0.72(0.4035)	0.80(0.3753)
F × M	0.84(0.4981)	0.26(0.9025)	0.86(0.4852)
M × L	0.66(0.5177)	0.62(0.5382)	0.11(0.8918)
F × L	0.16(0.8564)	0.28(0.7597)	0.66(0.5156)
G × M	1.27(0.282)	0.67(0.5125)	**3.22(0.0405)**
G × F	0.10(0.9072)	0.07(0.9372)	0.18(0.8386)
G × L	**12.21(0.0005)**	**14.79(0.0001)**	2.81(0.0943)
M × F × L	0.07(0.991)	0.10(0.9811)	0.04(0.9961)
M × F × G	0.46(0.7645)	0.11(0.9787)	0.35(0.8407)
M × L × G	0.40(0.6713)	0.58(0.5609)	0.09(0.9135)
F × L × G	0.85(0.4283)	0.37(0.6918)	0.02(0.9747)
M × F × L × G	0.39(0.8139)	0.33(0.8605)	0.29(0.8879)

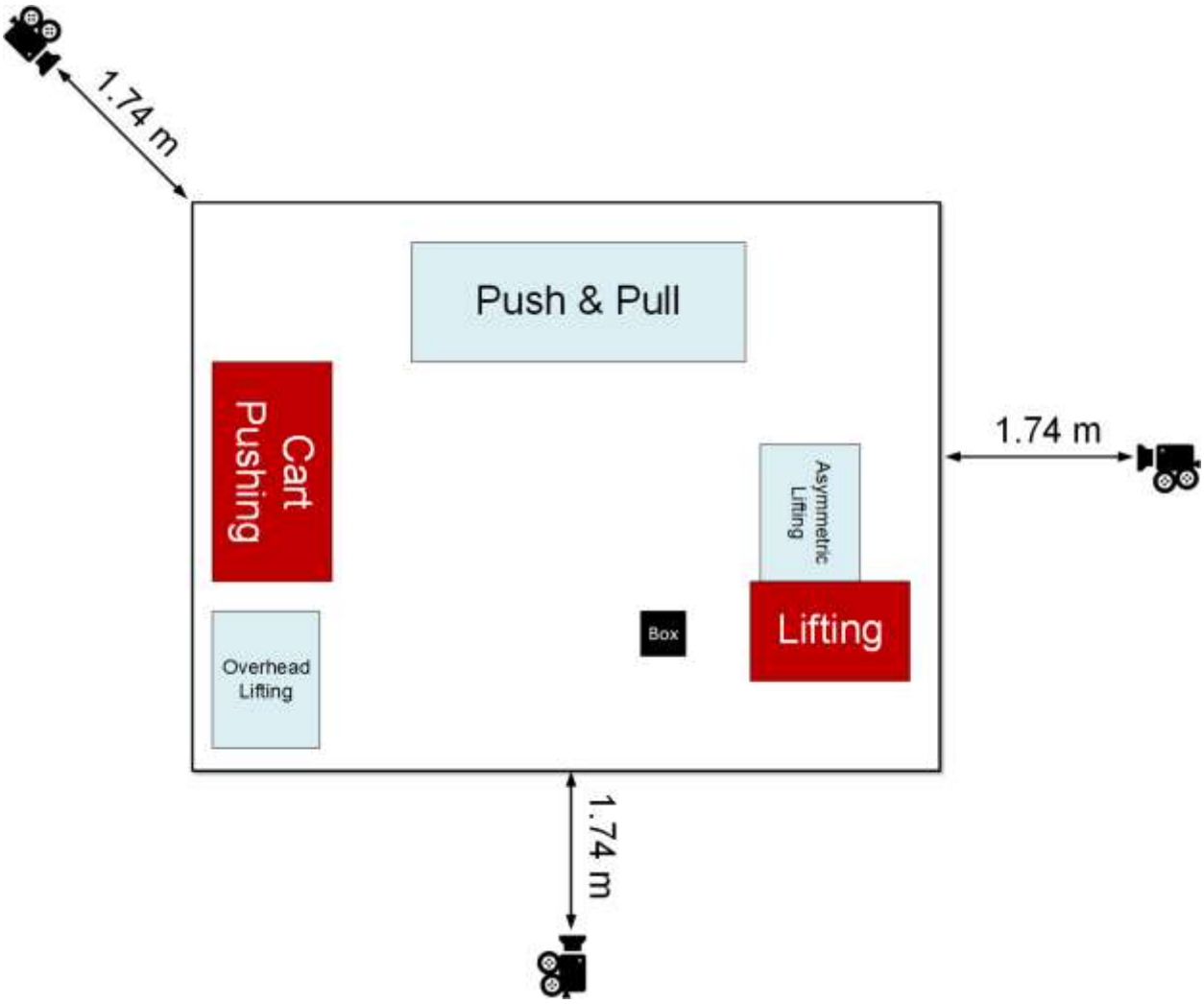

Figure 2.13. Work area configuration and camera layout

Table 2.5. Summary of Tested and Selected Hyperparameters for RNN Models

Hyperparameters	**Tested Parameters**	**Optimal Parameters**
RNN model layer	32, 64, 128, 256	128
Number of dense layers	32, 64, 128, 256	128
Training rate	1e-7 – 1e-4	1e-5
Optimizer	RMSProp, Adam, AdaDelta	Adaptive moment estimation method (Adam)
Dropout rate	0.1, 0.2, 0.5, 1	0.2
Decay rate	1e-10 – 1e-5	1e-8

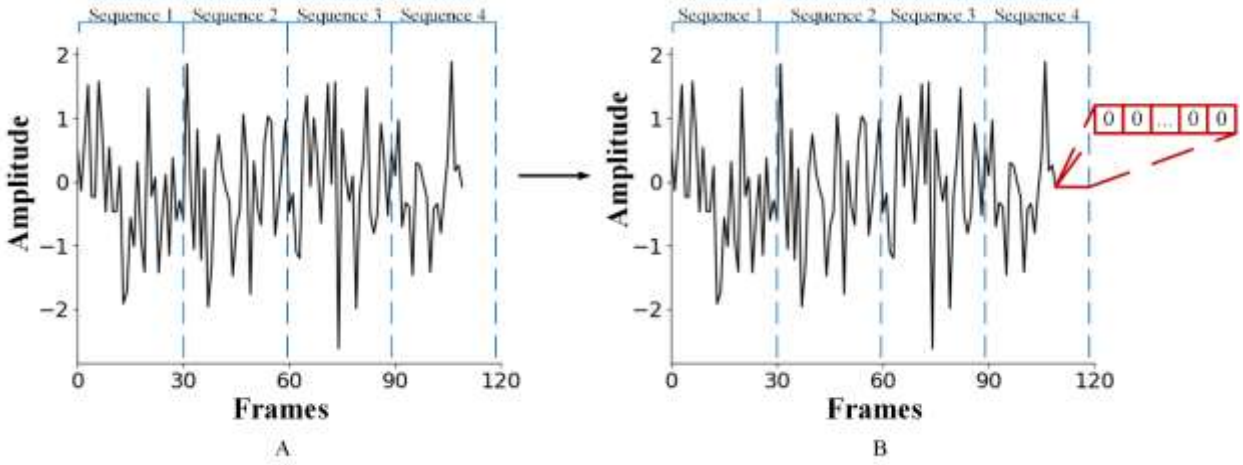

Figure 2.14. Creating a data sequence and zero-padding an incomplete sequence. Left: Sequences 1 - 3, each of length 30, were generated from the signal. Right: incomplete sequence four is zero-padded to achieve a length of 30.

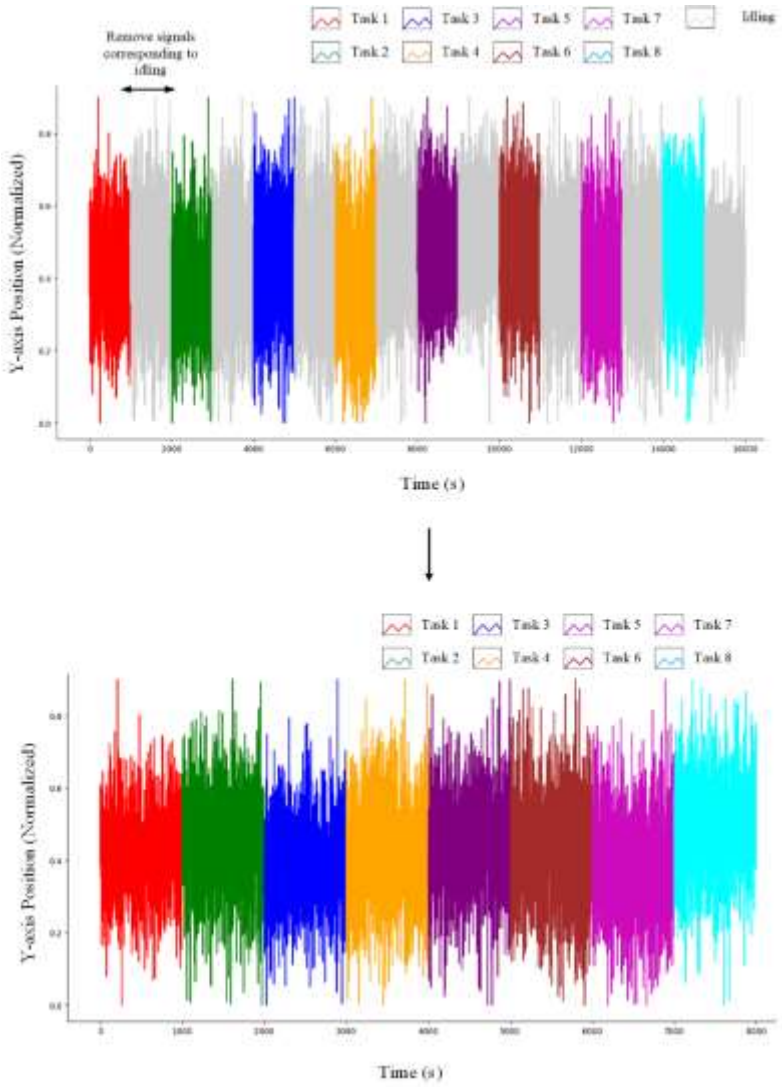

Figure 2.15. Steps to removing signals corresponding to "idling" when performing simulated tasks. Of note, signals in grey represent the idling signals, while signals in other colors represent different MMH tasks.

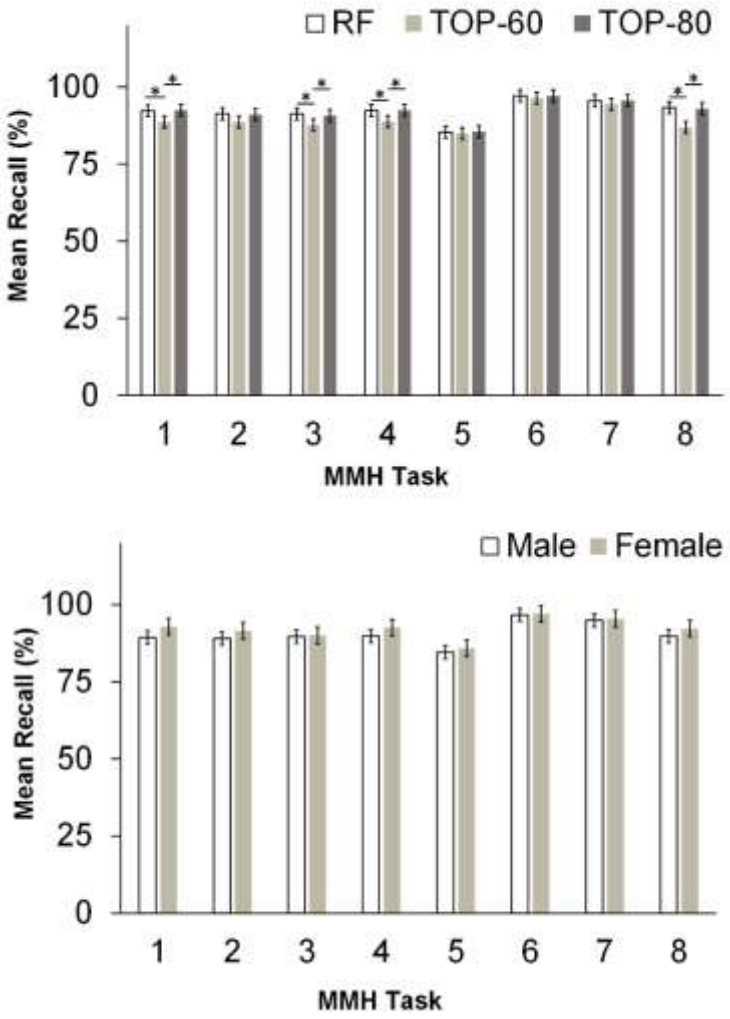

Figure 2.16. Significant *Feature set* × *MMH task* (top) and *Sex* × *MMH task* (bottom) interaction effects on mean recall for classifying MMH task.

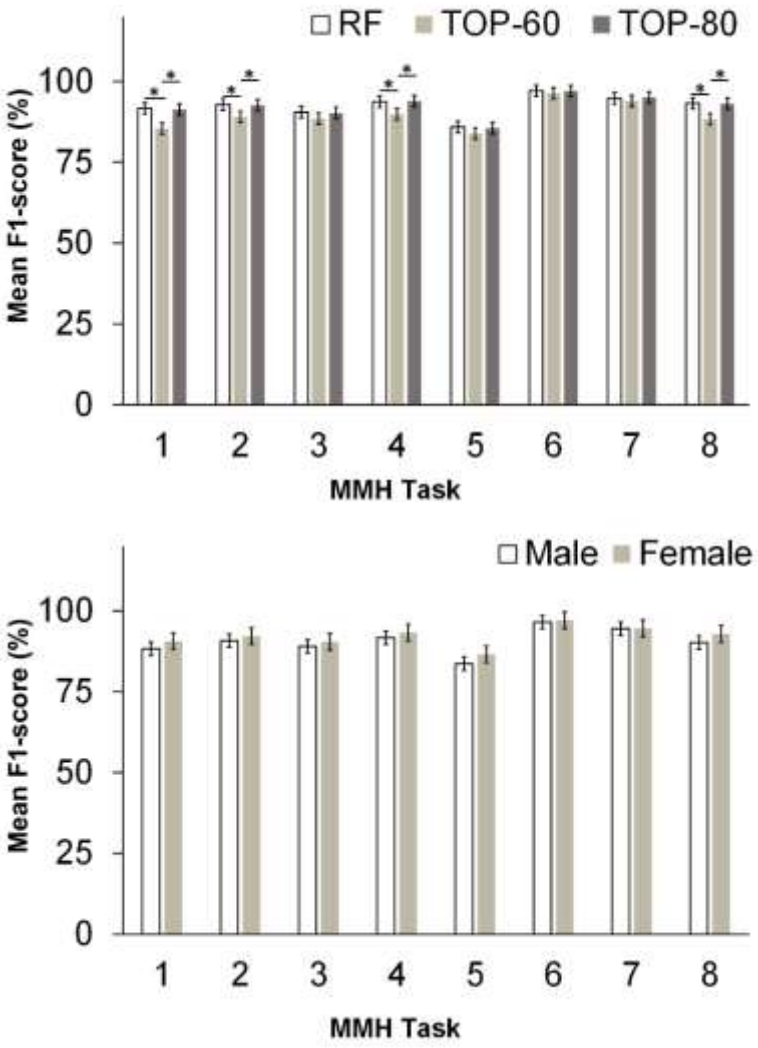

Figure 2.17. Significant *Feature set* × *MMH task* (top) and *Sex* × *MMH task* (bottom) interaction effects on mean recall for classifying MMH task.

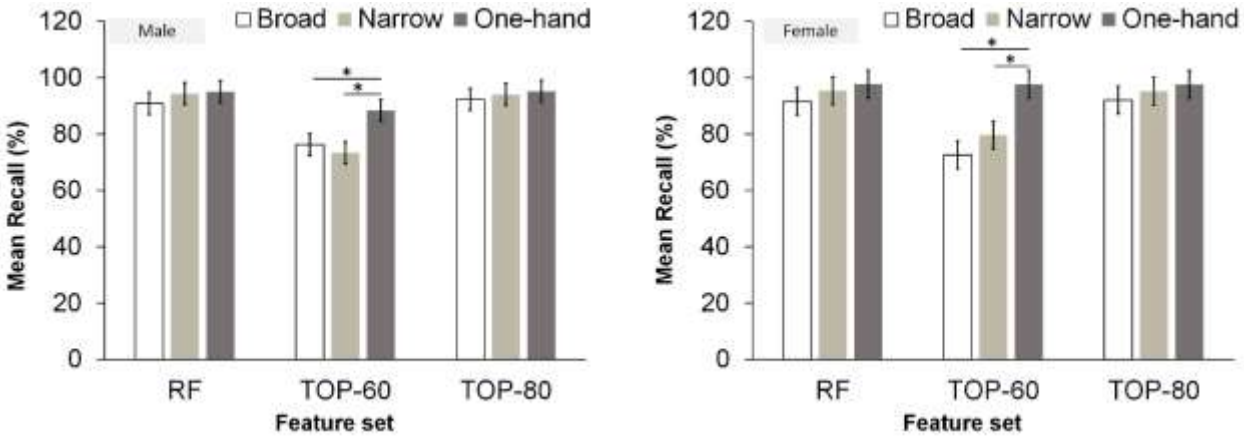

Figure 2.18. Significant *Feature set* × *Sex* × *Hand Configuration* interaction effects on mean recall for classifying hand configuration.

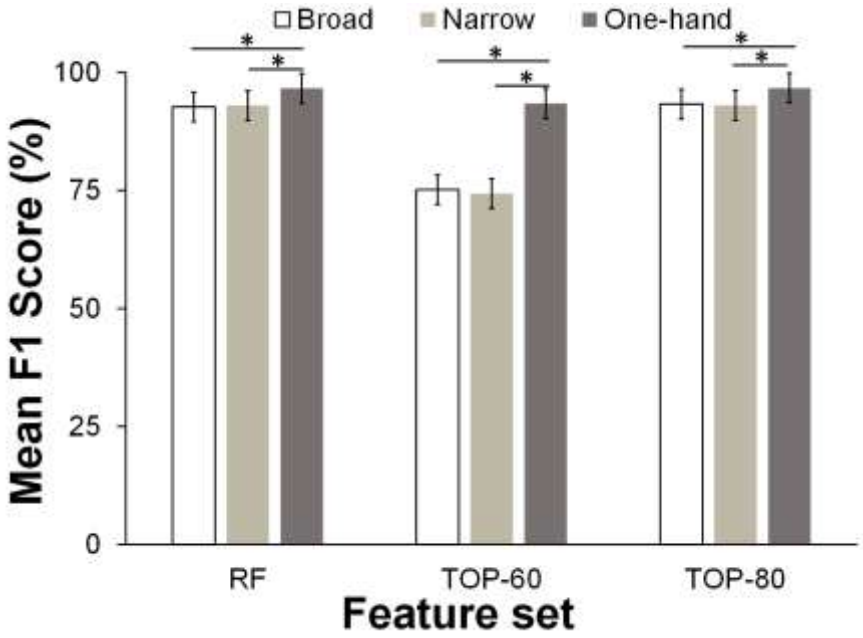

Figure 2.19. Significant *Feature set* × *Hand Configuration* interaction effects on mean F1-score for classifying hand configuration.

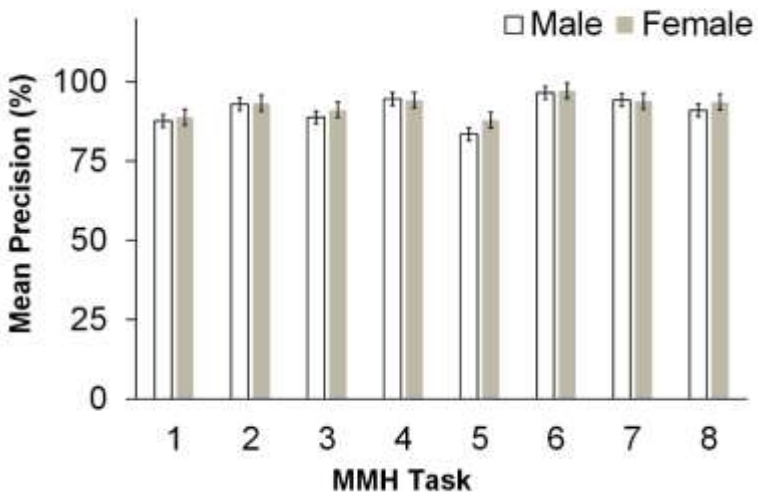

Figure 2.20. Significant *Sex* × *MMH task* interaction effects on mean precision for classifying MMH tasks.

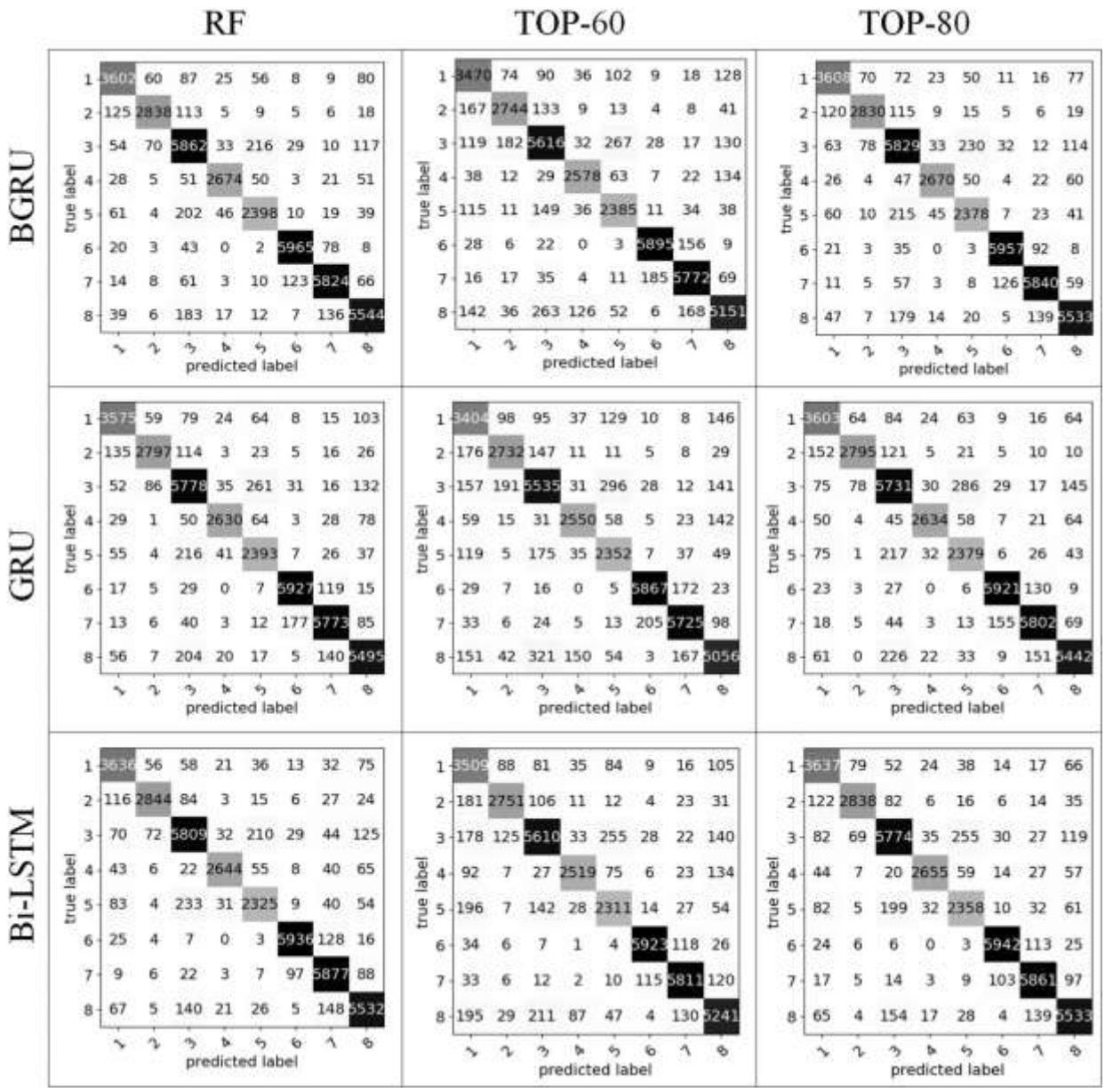

Figure 2.21. Overall confusion matrices for several feature sets. Confusion matrices A, B, and C represent RNN models' performance classifying MMH tasks using RF, TOP-60 and TOP-80 feature sets, respectively. Confusion matrices D, E, and F represent the classification performance of the Bi-LSTM, GRU, and BGRU models when classifying MMH tasks respectively.

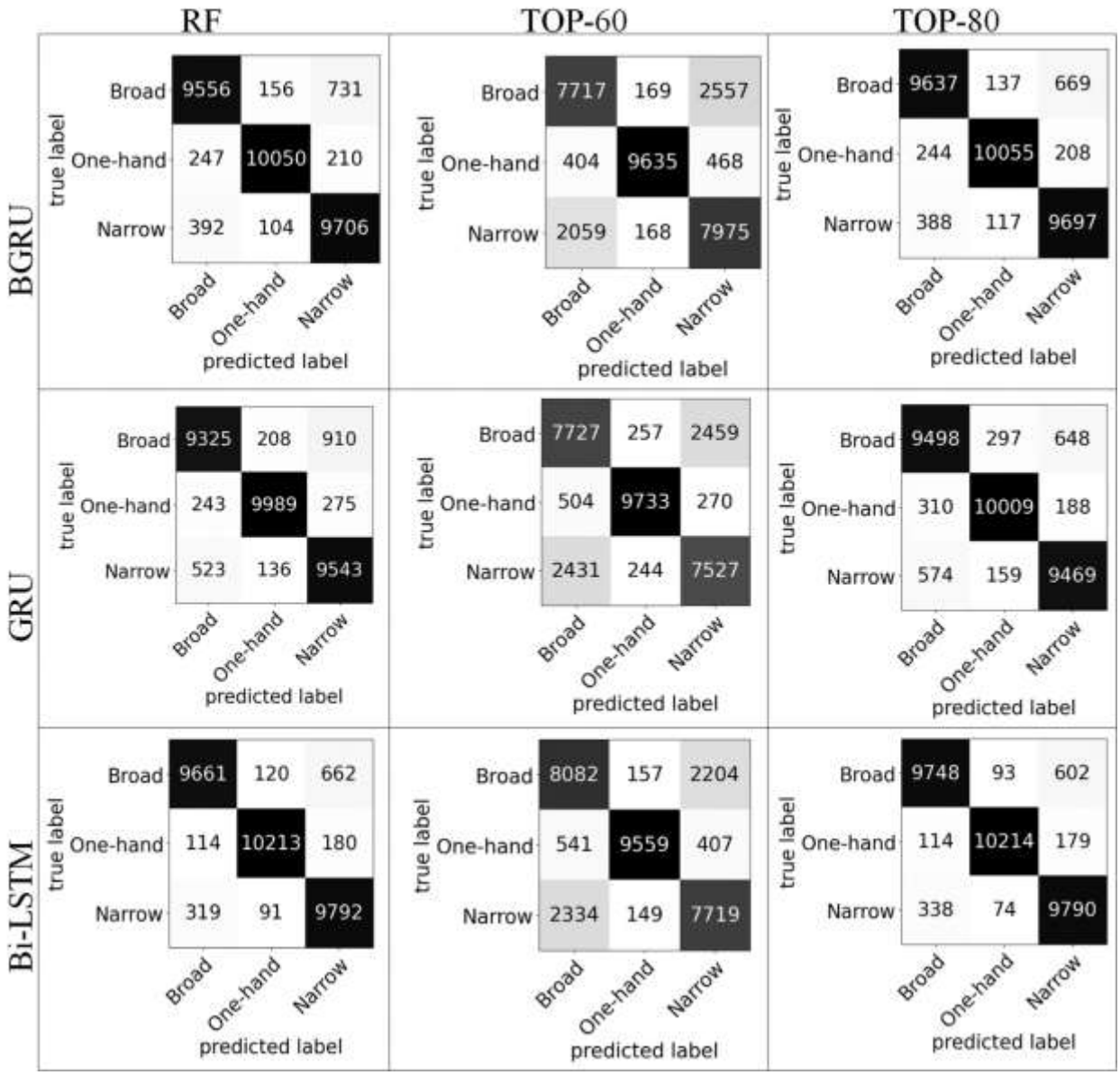

Figure 2.22. Overall confusion matrices for several feature sets. Confusion matrices A, B, and C represent RNN models' performance classifying hand configuration using RF, TOP-60 and TOP-80 feature sets, respectively. Confusion matrices D, E, and F represent the classification performance of the Bi-LSTM, GRU, and BGRU models when classifying hand configuration respectively.

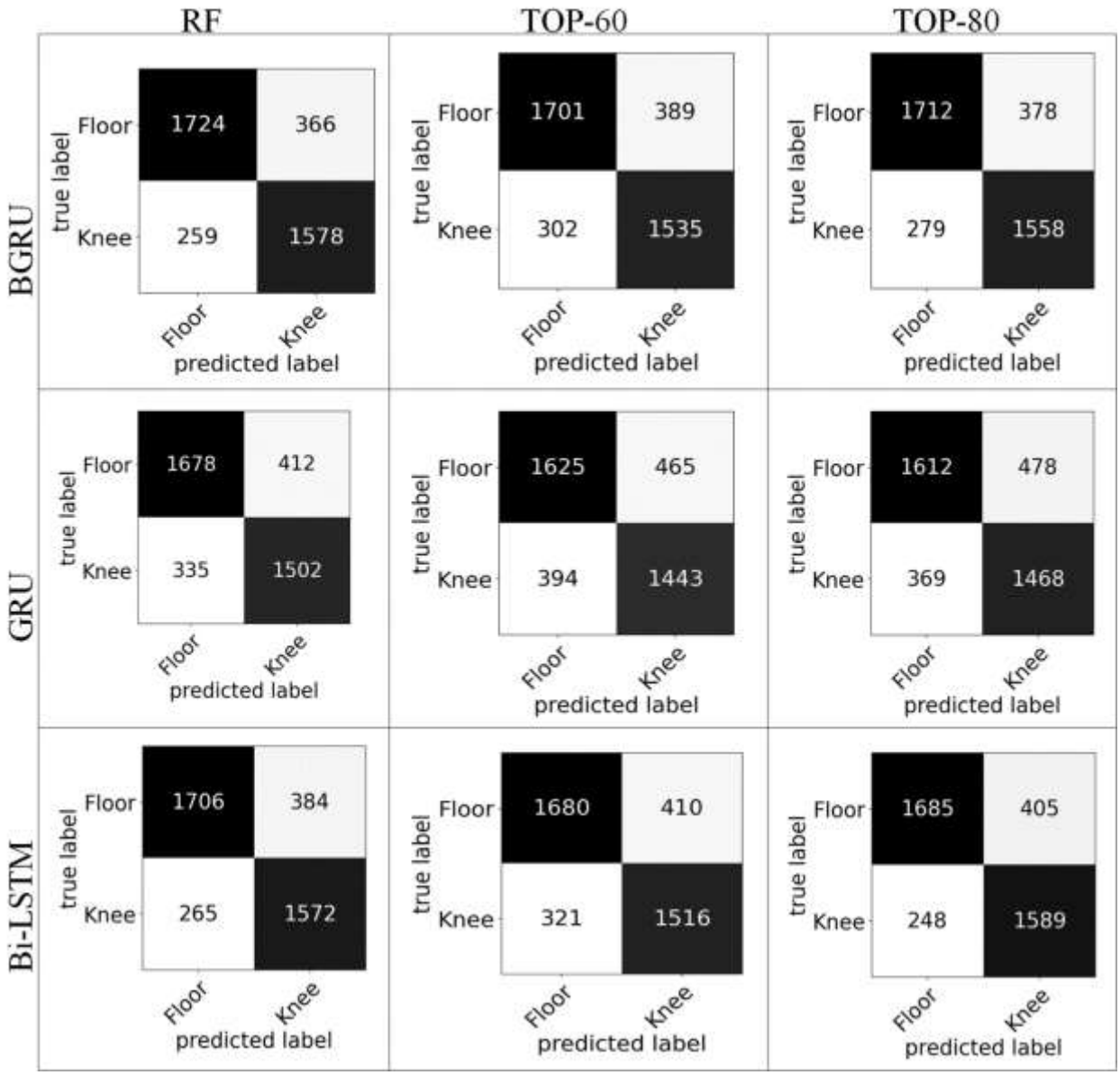

Figure 2.23. Overall confusion matrices for RNN model and features when classifying lift origin.

Table 2.6. Compilation of feature subsets obtained through the Minimal-Redundancy-Maximum-Relevancy (mRMR) method from the complete pool of kinematic features. List of the TOP-60 feature sets selected using the mRMR method for the purpose of classifying MMH tasks.

Joint Rotations	Joint Positions
V_LFoot	z_RShoulder
Z_LFoot	z_LFoot
Y_LFoot	y_LFoot
X_LFoot	x_LFoot
V_LShin	z_LShin
Z_LShin	y_LShin
Y_LShin	x_LShin
X_LShin	z_LThigh
V_LThigh	y_LThigh
Z_LThigh	x_LThigh
Y_LThigh	z_RFoot
X_LThigh	y_RFoot
V_RFoot	x_RFoot
Z_RFoot	z_RShin
Y_RFoot	y_RShin
X_RFoot	x_RShin
V_RShin	z_RThigh
Z_RShin	y_RThigh
Y_RShin	x_RThigh
X_RShin	z_LForearm
V_RThigh	y_LForearm
Z_RThigh	x_LForearm
Y_RThigh	z_LShoulder
X_RThigh	y_LShoulder
V_LForearm	x_LShoulder
Z_LForearm	
Y_LForearm	
X_LForearm	
V_LShoulder	
Z_LShoulder	
Y_LShoulder	
X_LShoulder	
V_LClavicle	
Z_LClavicle	
Y_LClavicle	

Table 2.7. List of the TOP-80 feature sets selected using the mRMR method for the purpose of classifying MMH tasks.

Joint Rotations	Joint Positions
V_LFoot	z_RShoulder
Z_LFoot	z_LFoot
Y_LFoot	y_LFoot
X_LFoot	x_LFoot
V_LShin	z_LShin
Z_LShin	y_LShin
Y_LShin	x_LShin
X_LShin	z_LThigh
V_LThigh	y_LThigh
Z_LThigh	x_LThigh
Y_LThigh	z_RFoot
X_LThigh	y_RFoot
V_RFoot	x_RFoot
Z_RFoot	z_RShin
Y_RFoot	y_RShin
X_RFoot	x_RShin
V_RShin	z_RThigh
Z_RShin	y_RThigh
Y_RShin	x_RThigh
X_RShin	z_LForearm
V_RThigh	y_LForearm
Z_RThigh	x_LForearm
Y_RThigh	z_LShoulder
X_RThigh	y_LShoulder
V_LForearm	x_LShoulder
Z_LForearm	z_LClavicle
Y_LForearm	y_LClavicle
X_LForearm	x_LClavicle
V_LShoulder	z_RForearm
Z_LShoulder	y_RForearm
Y_LShoulder	x_RForearm
X_LShoulder	y_RShoulder
V_LClavicle	x_RShoulder
Z_LClavicle	
Y_LClavicle	
X_LClavicle	
V_RForearm	
Z_RForearm	
Y_RForearm	
X_RForearm	
V_RShoulder	
Z_RShoulder	
Y_RShoulder	
X_RShoulder	
V_RClavicle	
Z_RClavicle	
Y_RClavicle	

Table 2.8. List of the TOP-60 feature sets selected using the mRMR method for the purpose of classifying hand configuration.

Joint Rotations	Joint Positions
Y_LShoulder	z_LFoot
V_LFoot	y_LFoot
Z_LFoot	x_LFoot
Y_LFoot	z_LShin
X_LFoot	y_LShin
V_LShin	x_LShin
Z_LShin	z_LThigh
Y_LShin	y_LThigh
X_LShin	x_LThigh
V_LThigh	z_RFoot
Z_LThigh	y_RFoot
Y_LThigh	x_RFoot
X_LThigh	z_RShin
V_RFoot	y_RShin
Z_RFoot	x_RShin
Y_RFoot	z_RThigh
X_RFoot	y_RThigh
V_RShin	x_RThigh
Z_RShin	z_LForearm
Y_RShin	y_LForearm
X_RShin	x_LForearm
V_RThigh	z_LShoulder
Z_RThigh	y_LShoulder
Y_RThigh	x_LShoulder
X_RThigh	
V_LForearm	
Z_LForearm	
Y_LForearm	
X_LForearm	
V_LShoulder	
Z_LShoulder	
X_LShoulder	
V_LClavicle	
Z_LClavicle	
Y_LClavicle	
X_LClavicle	

Table 2.9. List of the TOP-80 feature sets selected using the mRMR method for the purpose of classifying hand configuration.

Joint Rotations	Joint Positions
Y_LShoulder	z_LFoot
V_LFoot	y_LFoot
Z_LFoot	x_LFoot
Y_LFoot	z_LShin
X_LFoot	y_LShin

V_LShin	x_LShin
Z_LShin	z_LThigh
Y_LShin	y_LThigh
X_LShin	x_LThigh
V_LThigh	z_RFoot
Z_LThigh	y_RFoot
Y_LThigh	x_RFoot
X_LThigh	z_RShin
V_RFoot	y_RShin
Z_RFoot	x_RShin
Y_RFoot	z_RThigh
X_RFoot	y_RThigh
V_RShin	x_RThigh
Z_RShin	z_LForearm
Y_RShin	y_LForearm
X_RShin	x_LForearm
V_RThigh	z_LShoulder
Z_RThigh	y_LShoulder
Y_RThigh	x_LShoulder
X_RThigh	z_LClavicle
V_LForearm	y_LClavicle
Z_LForearm	x_LClavicle
Y_LForearm	z_RForearm
X_LForearm	y_RForearm
V_LShoulder	x_RForearm
Z_LShoulder	z_RShoulder
X_LShoulder	y_RShoulder
V_LClavicle	x_RShoulder
Z_LClavicle	
Y_LClavicle	
X_LClavicle	
V_RForearm	
Z_RForearm	
Y_RForearm	
X_RForearm	
V_RShoulder	
Z_RShoulder	
Y_RShoulder	
X_RShoulder	
V_RClavicle	
Z_RClavicle	
Y_RClavicle	

Table 2.10. List of the TOP-60 feature sets selected using the mRMR method for the purpose of classifying lift origin.

Joint Rotations	Joint Positions
V_LFoot	x_RShin
Z_LFoot	z_LFoot
Y_LFoot	y_LFoot
X_LFoot	x_LFoot

V_LShin	z_LShin
Z_LShin	y_LShin
Y_LShin	x_LShin
X_LShin	z_LThigh
V_LThigh	y_LThigh
Z_LThigh	x_LThigh
Y_LThigh	z_RFoot
X_LThigh	y_RFoot
V_RFoot	x_RFoot
Z_RFoot	z_RShin
Y_RFoot	y_RShin
X_RFoot	z_RThigh
V_RShin	y_RThigh
Z_RShin	x_RThigh
Y_RShin	z_LForearm
X_RShin	y_LForearm
V_RThigh	x_LForearm
Z_RThigh	z_LShoulder
Y_RThigh	y_LShoulder
X_RThigh	x_LShoulder
V_LForearm	
Z_LForearm	
Y_LForearm	
X_LForearm	
V_LShoulder	
Z_LShoulder	
Y_LShoulder	
X_LShoulder	
V_LClavicle	
Z_LClavicle	
Y_LClavicle	
X_LClavicle	

Table 2.11. List of the TOP-80 feature sets selected using the mRMR method for the purpose of classifying lift origin.

Joint Rotations	Joint Positions
V_LFoot	x_RShin
Z_LFoot	z_LFoot
Y_LFoot	y_LFoot
X_LFoot	x_LFoot
V_LShin	z_LShin
Z_LShin	y_LShin
Y_LShin	x_LShin
X_LShin	z_LThigh
V_LThigh	y_LThigh
Z_LThigh	x_LThigh
Y_LThigh	z_RFoot
X_LThigh	y_RFoot
V_RFoot	x_RFoot
Z_RFoot	z_RShin

Y_RFoot	y_RShin
X_RFoot	z_RThigh
V_RShin	y_RThigh
Z_RShin	x_RThigh
Y_RShin	z_LForearm
X_RShin	y_LForearm
V_RThigh	x_LForearm
Z_RThigh	z_LShoulder
Y_RThigh	y_LShoulder
X_RThigh	x_LShoulder
V_LForearm	z_LClavicle
Z_LForearm	y_LClavicle
Y_LForearm	x_LClavicle
X_LForearm	z_RForearm
V_LShoulder	y_RForearm
Z_LShoulder	x_RForearm
Y_LShoulder	z_RShoulder
X_LShoulder	y_RShoulder
V_LClavicle	x_RShoulder
Z_LClavicle	
Y_LClavicle	
X_LClavicle	
V_RForearm	
Z_RForearm	
Y_RForearm	
X_RForearm	
V_RShoulder	
Z_RShoulder	
Y_RShoulder	
X_RShoulder	
V_RClavicle	
Z_RClavicle	
Y_RClavicle	

Appendix B – Model Evaluation Metrics

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1-score = 2\,x\,\frac{Precision\ x\ Recall}{Precision+Recall}$$

3. Chapter 3. Predicting 3D Dynamic Hand Forces from Markerless Motion Capture using Attention-based Recurrent Neural Networks

3.1 Abstract

Accurately assessing three-dimensional, dynamic hand forces for physical exposure is challenging due to the need for specialized equipment. We explored the combination of MMC and IPM data with machine learning to predict hand forces during one-handed and two-handed MMH tasks. Our results were encouraging overall, but predictions were less accurate in pushing and pulling tasks. Also, the use of feature set from in-sole pressure system increased root mean square error up to 60% in most conditions. In summary, our approach demonstrates potential for predicting dynamic external hand forces non-intrusively. Future work is needed to enhance model performance in push and pull task conditions.

3.2 Introduction

Work-related musculoskeletal disorders (WMSDs) continue to be a major health problem in labor-intensive industries, accounting for a median of 14 days away from work (U.S. Bureau of Labor Statistics, 2021). WMSDs are most common in the back, contributing to nearly a third of all cases (U.S. Bureau of Labor Statistics, 2021). Workers engaged in manual material handling (MMH) tasks are at an increased risk, in particular lifting/lowering, pushing/pulling, or holding (Bernard & Putz-Anderson, 1997; da Costa & Vieira, 2010; Yang et al., 2016). However, MMH tasks are an intrinsic part of many industry sectors (e.g., manufacturing, construction, warehousing, and distribution), and it remains quite challenging to eliminate many MMH tasks. As part of the traditional ergonomics assessment and control process, a physical exposure assessment (aka, physical demands analysis) can help identify high-risk tasks, guide the development and assessment of ergonomic interventions, and contribute to understanding exposure-risk relationships (Mathiassen & Winkel, 1991; Spielholz et al., 2001; Wells et al., 1997; Winkel & Mathiassen, 1994). Physical exposure can be measured using self-reports, observational methods, and direct measurements (Burdorf & van der Beek, 1999; Koppelaar & Wells, 2005; Wells et al., 2004). While these assessments often emphasize postural exposures, hand forces – the loads acting on the hand due to exertions or activities occurring during MMH tasks – have also been identified as one of the key factors contributing to WMSD risks (Koppelaar & Wells, 2005; Wells et al., 2004).

Although self-reports and observational methods can assess various occupational risk factors using questionnaires, checklists, and discrete and continuous scales, applications in estimating hand forces are limited, since these methods yield relatively crude physical exposure scores. Such scores can be useful in screening or initial analyses, but they have limited use in

detailed risk assessment (Kumar, 1993; Spielholz et al., 2001; Wiktorin et al., 1996; Wiktorin et al., 1993). Direct measurements of hand forces are feasible, such as using force gauges and load cells (Burdorf, 1995; van der Beek & Frings-Dresen, 1998). Yet, measuring hand force *in situ* is typically quite difficult, since direct measurements require specialized equipment and expertise (Koppelaar & Wells, 2005; Wang et al., 2023) and can substantially interfere with the work being done.

Recent technological innovations offer the possibility to automatically quantify hand forces in the workplace and with minimal worker interference. These innovations include the development of low-cost sensors – especially inertial measurement units (IMUs) and *markerless motion capture* (MMC) systems – and fast, effective machine learning algorithms. In fact, there is existing work that demonstrates the feasibility of estimating hand forces from other measures. As an example, data extracted from four IMUs were used to automatically classify among three different hand loads manipulated during various load carriage modes (Lim, 2024; Lim & D'Souza, 2019). Similarly, data from three IMUs were used to classify among three levels of load mass during load carriage (Lee et al., 2020). However, IMUs have potential drawbacks, such as wearer discomfort, influences on work behaviors, challenges related to battery drainage, and interference from magnetic sources. IMU systems are also still fairly expensive. On the other hand, MMC offers a non-intrusive approach to obtaining body kinematics to quantify physical exposure. As an example, using two MMCs for assessing upper extremity joint angles yielded small mean absolute errors (MAE < 5°) compared to a marker-based camera system (Steinebach et al., 2020).

However, these and similar approaches still have three important limitations with respect to estimating hand loads. First, handled loads were only classified categorically in earlier reports,

either as: no, standard, or excessive loads (Lee et al., 2020); as no load, or 50%/75% of maximum acceptable weight of carry (Lim & D'Souza, 2019); or as no, small, and high loads (Yang et al., 2020). While such load classification has clear use in risk assessment, biomechanical models and some modern risk assessment tools require more precise information. For example, advanced biomechanical modeling software requires specific hand forces to perform top-down inverse dynamics analysis and to estimate internal (reaction) forces. Second, only classifying loads does not capture the effects of dynamics, and thus could substantially underestimate physical loads and risks (Garg et al., 1982). Continuous sampling of hand forces, however, could capture dynamic effects. Third, earlier studies did not consider 3D hand forces, though these are needed for comprehensive biomechanical analyses, given that asymmetric loading in MMH is an important WMSD risk factor (Arjmand et al., 2010; Gallagher et al., 1994).

The main goal of this study was to investigate the feasibility of estimating external hand forces using other measures, such as kinematics from MMC and/or in-sole pressure systems, and the secondary goal was to assess the effectiveness of various machine learning models in estimating external hand forces. Using kinematic measures as predictors was expected to help estimate hand forces, because such measures provide information about body trajectory, velocity, and acceleration when manipulating objects (Hlucny & Novak, 2020; Lee, 2015). For example, lower extremity velocity and angular motion differ across different load magnitudes, and participants tend to shorten the acceleration period when lifting heavier objects (Lee, 2015). Machine learning algorithms were used here due to the anticipated complex, high-dimensional, and non-linear relationships between predictors and hand forces. Yet, various machine learning algorithms were evaluated because no single algorithm was expected to be suited for predicting

external hand forces (Jozefowicz et al., 2015). Results from this work may enable more efficient and effective assessments of physical exposures, by establishing new procedures to predict external hand forces from other, more accessible measures.

3.3 Methods

Experimental Design and Data Collection

The current analysis was based on data obtained from a study described in Chapter 2. In that study, participants (14 females and 22 males) completed simulated MMH tasks involving 36 task conditions. Specific MMH tasks simulated were: symmetric box lifting from two different origins (floor and individual knee height); asymmetric box lifting; box carriage from one table to another; box pushing and pulling; cart pushing; overhead lifting from cart height; and lowering to different origins (floor and individual knee height). All tasks were performed using three levels of *Hand Configuration,* three levels of *Mass,* two levels of *Lift Origin,* and two levels of *Start Position*. A study trial involved completing all tasks sequentially with a given mass, hand configuration, lift origin, and from a given starting position. More detailed descriptions are provided in Chapter 2.

Whole-body kinematics were available from the MMC system at 30 Hz (see Chapter 2). Tri-axial hand forces were obtained from load cells (Michigan Scientific Corp. TR3D-A-1K, Charlevoix, MI, USA) attached near the top of the box (Figure 3.1). Hand forces were sampled at 200 Hz from the load cells using a custom LabView (NI, Austin, TX, USA) program. Further, ground reaction forces were collected using an insole pressure measurement (IPM) system (Ultium Insoles, Noraxon, Scottsdale, AZ, USA). Each IPM system has an 8-zone pressure sensor, and the data were available at 250 Hz. At the start of each trial, participants completed a process to calibrate the IPM system, which involved standing on one foot at a time. Data from

the load cells and IPM were synchronized using an analog output signal generated from LabView, whereas the MMC and IPM systems were synchronized using an identified movement at the start of each trial (e.g., raising a foot).

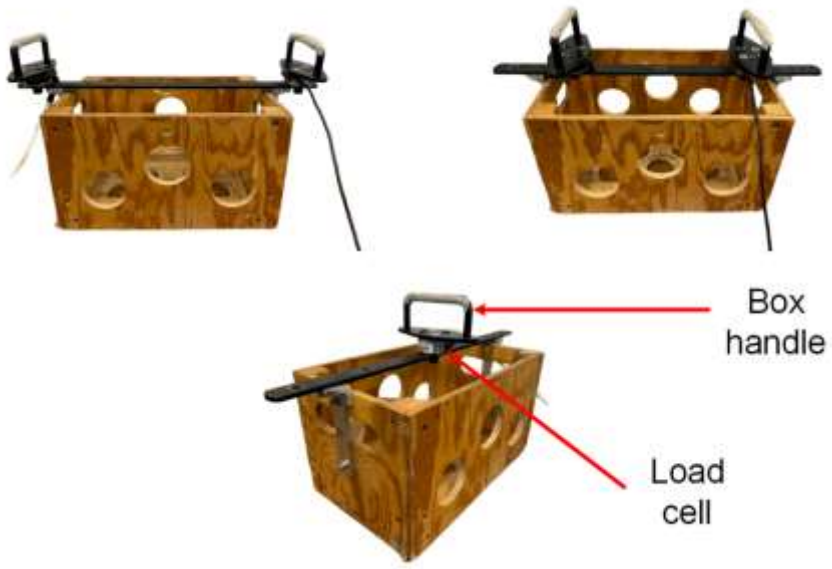

Figure 3.1. Illustrations showing load cell attachment on top of the box for the three hand configurations: broad (top left), narrow (top right), and one-hand (bottom).

Estimations of external hand forces completed as a multi-staged data processing effort, including data collection and processing, feature selection, data segmentation, and hand force prediction, followed by model evaluation (Figure 3.2). The latter stages are discussed in detail subsequently.

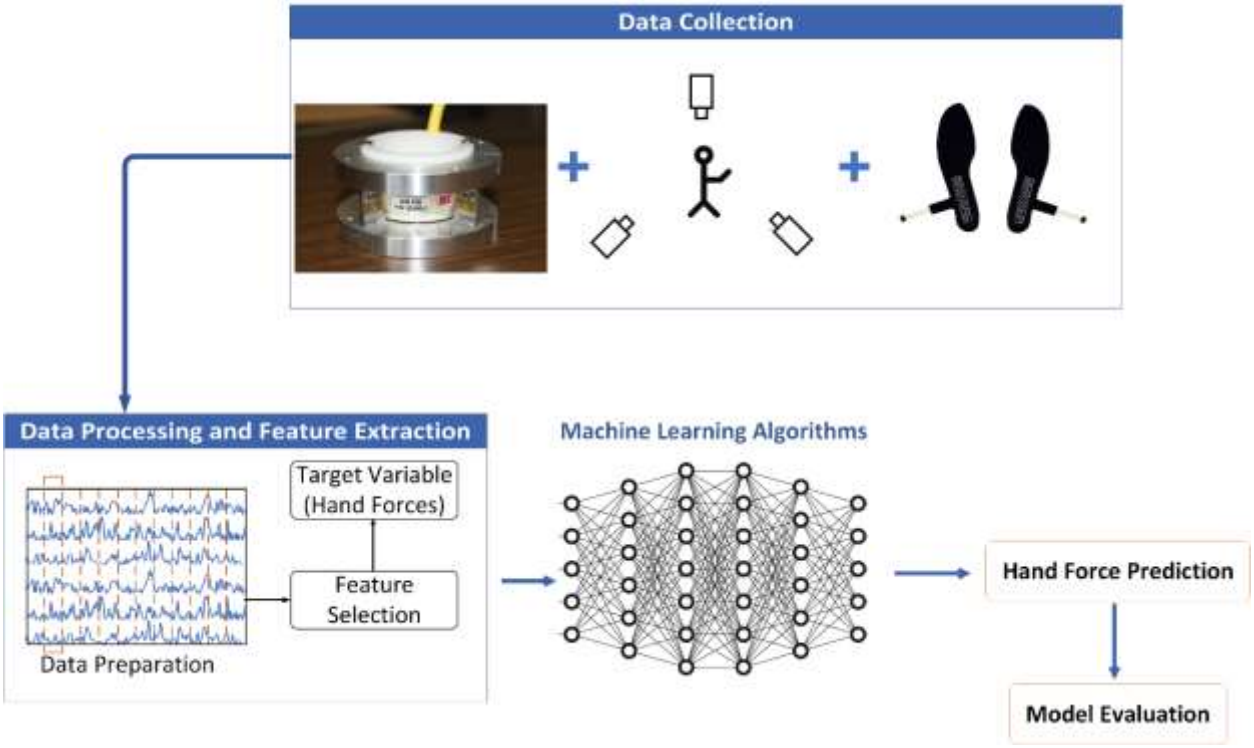

Figure 3.2. Overview of the hand force prediction process.

Data Processing and Feature Selection

Data Processing

MMC data (tri-axial positions and quaternion joint rotations) were processed as highlighted in section 2.5.1. Data recorded from the IPM system were processed using myoResearch software (v.3.20, Noraxon, Scottsdale, AZ, USA). Specifically, IPM data were extracted from the eight zones of each IPM sensor, including the 2-D center of pressure and total insole pressure. Data obtained from the load cells and IPM systems were filtered using a moving root-mean-square window to reduce sensor noise and other artifacts; respective window sizes of 300 and 200 ms were used, which were determined using residual analysis (Winter, 2009). Subsequently, load cell and IPM data were downsampled to 30 Hz to match the sampling rate of whole-body kinematics. All offline data processing was completed using MATLAB (MathWorks

Inc., Natick, MA, USA). Tri-axial hand forces in a local (box-centered) coordinate system (Figure 3.3) served as the target variables for subsequent machine learning prediction models. Filtered joint kinematics, IPM data, and hand force components were all normalized using the Min-Max-Scaler function (Pedregosa et al., 2011), since machine learning models are sensitive to unscaled data (Zhou et al., 2022). Both training and evaluation data (see below) were transformed to a (0,1) range.

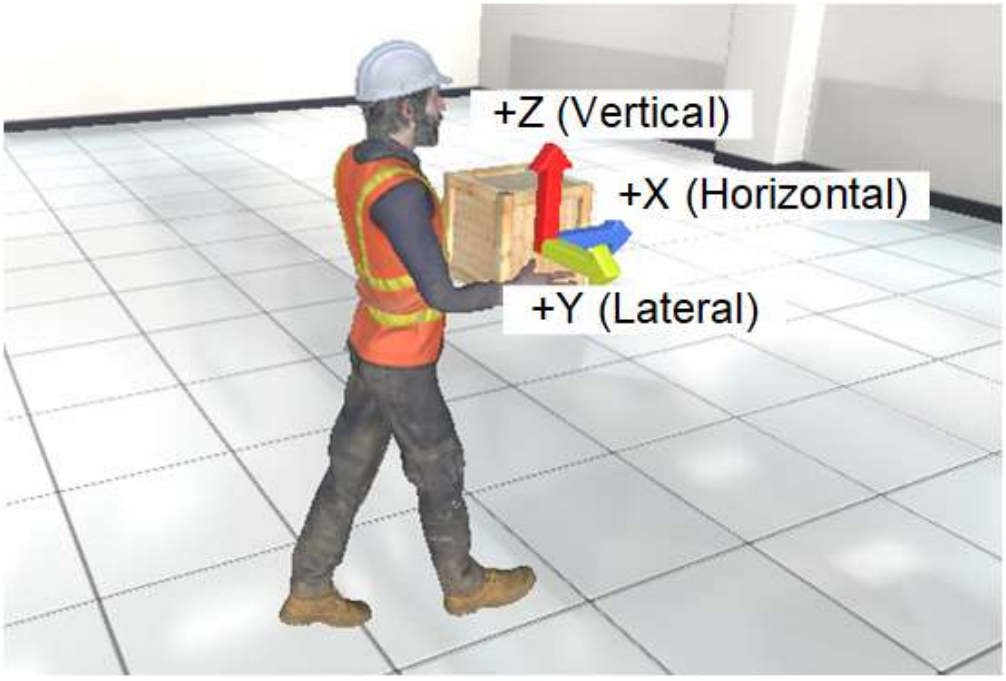

Figure 3.3. Illustration of the local coordinate system of the box manipulated by the participant. Note that all hand forces in this study were in box-centered coordinate system.

Feature Selection

Features are distinct and measurable attributes of data that serve as input for machine learning algorithms. Here, we considered three feature sets as input for the prediction models. The first feature set included whole-body kinematics (W-B), specifically time series of the tri-axial joint positions and quaternion joint rotations. Each data frame yielded a 119-dimensional

input vector. The second feature set included data from the IPM systems (IPM_{ONLY}) with a 22-dimensional vector of IPM data (from each pair of eight zones of the IPM system, 2-D center of pressure, and total insole pressure). The third feature set included data from whole-body kinematics along with data from the IPM systems (W-B + IPM). Each data frame yielded a 141-dimensional vector, comprising the W-B data along with the IPM_{ONLY} data.

Hand Force Prediction Models

Machine Learning Models

Two recurrent neural network (RNN) models were used to estimate the time series of external hand forces acting on participants while they performed the MMH tasks. Two specific models were selected: Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Units (BGRU). A Bi-LSTM approach is useful for modeling complex contextual relationships in sequential data. A Bi-LSTM has three gates – input, forget, and output – that modify the Bi-LSTM cell state (Hochreiter & Schmidhuber, 1997). Bi-LSTM models are suitable for regression problems, as they combine input from the past to the future, a forward LSTM, and from the future to the past, a backward LSTM. BGRU models are created by combining forward and backward gated recurrent unit networks. Similar to Bi-LSTM, a BGRU model has two gates – control and reset – that modify which information to retain, forget, and update (Cho et al., 2014). BGRU models have fewer learnable parameters; thus, these models can often converge readily and learn long-term dependencies more efficiently with more limited datasets (Fu et al., 2016; Greff et al., 2016).

Model Architecture

The overall model architecture included an input layer, a machine learning model (i.e., Bi-LSTM or BGRU), an attention layer, dropout functions, and output layers (Figure 3.4). All model implementations were done using Python 3.10.11 (https://www.python.org).

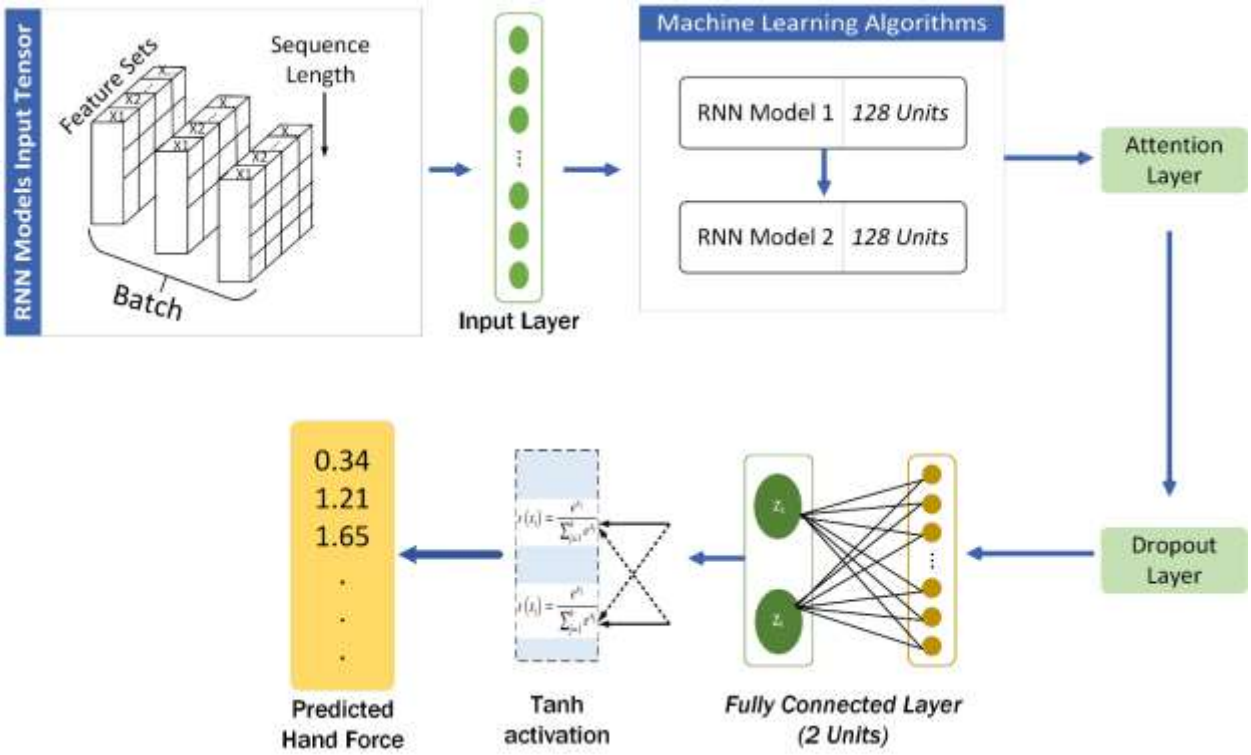

Figure 3.4. Overall architecture of the recurrent neural network models for hand force prediction.

Similar to the process described in Chapter 2, each machine learning model takes as input a multidimensional matrix based on the sequence length and feature dimensions. Sequence length is the time duration of the input data, which was fixed at 30, corresponding to 1 second of data collection (30 Hz). This sequence length was selected to match the length used for our MMH task classification analysis (Chapter 2). An attention layer was added to dynamically assign different weights (attention scores) to different parts of the input sequence. Adding such layer to the RNN model could enhance model performance, by ensuring the model focuses on

relevant information while processing sequential data. Specifically, a "Bahdanau Attention Mechanism" layer was implemented (Cho et al., 2014; Sutskever et al., 2014). Recall that the respective feature dimensions were 119, 141, and 22 for the W-B, W-B + IPM, and IPM_{ONLY} feature sets. Finally, the batch size was set at 32, which was determined as the optimal size during initial testing. In brief, the input layer was connected to each machine learning model, consisting of an attention layer, a dropout function, dense layer, an optimizer, a loss function, and a regressor layer. Predictions were made sample-to-sample, corresponding to 30 estimates for each second of data (Figure 3.5).

Separate time series, regression-based models were developed to predict external hand forces for tasks that were completed bilaterally (two-handed) or unilaterally (one-handed). Models for bilateral force predictors estimated bilateral hand forces for tasks completed in both the narrow and broad hand configurations, while models for unilateral force prediction were used for the one-hand task configuration.

Model Training and Evaluation

A leave-one-subject-out method was used to train and evaluate the machine learning models, by using each subject as a "fold." In other words, data from 35 participants were used for training, and data from one participant were used for evaluation, and this process was iterated over all 36 participants. This evaluation method mirrors real-world training and testing scenarios, in which a model is trained on a dataset of known subjects and then evaluated on a completely new, unseen subject (Jordao et al., 2018).

RNN models have several hyperparameters, which are parameters used to control the learning process, model complexity, and regularize the strength of the model. Bi-LSTM and BGRU hyperparameters include the number of hidden layers, learning rate, training rate, dropout rate,

optimizer, and decay rate. Hyperparameter values were determined here through an empirical tuning process involving manual adjustments to each parameter and subsequent evaluation of the resulting model performance. From this tuning process, hidden layer number, learning rate, dropout rate, optimizer, and decay rate were fixed as 320, 0.000005, 0.5, adaptive learning rates, and 0.00000008, respectively.

Performance of the prediction models was assessed using two metrics that were derived separately for the X, Y, and Z axes in each trial: 1) root-mean-square error (RMSE), as a measures of "typical" error; and 2) peak error (PE), the difference between the respective 95th percentile levels of the actual and predicted hand forces (Figure 5), as a measure of error in estimating the largest hand loads. Both metrics were obtained for each 1-second sequence of model input data. For models involving a bilateral hand configuration, the metrics were calculated separately for the left and right hands. Note that mean absolute error was explored initially as an addition performance metric; results for this metric are not reported here as the values were highly correlated with RMSE.

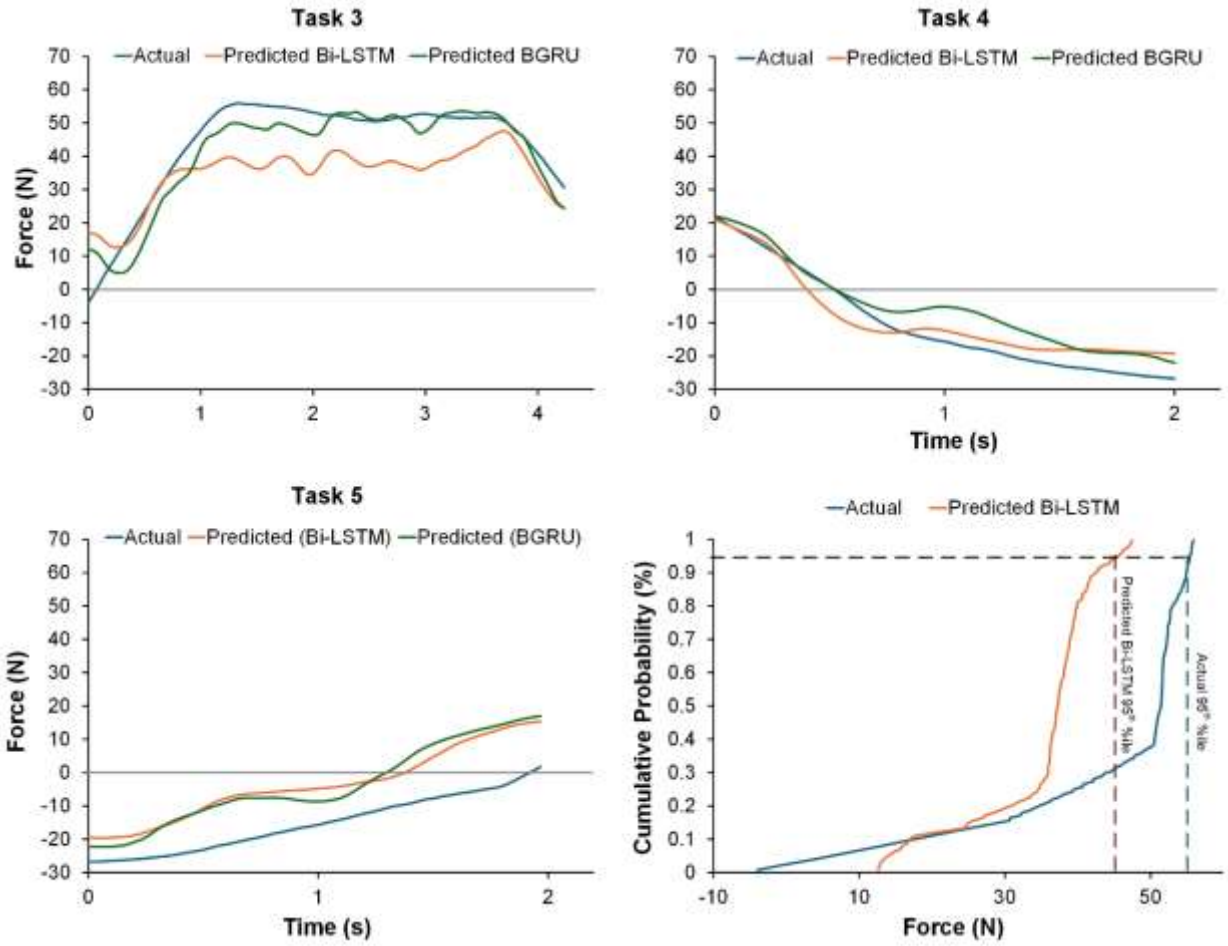

Figure 3.5. Examples of force predictions in the vertical direction using an attention-based recurrent neural network approach. In this example, the feature set included both kinematics and kinetics. Bottom-right image shows an example of the 95th percentile for actual hand force and hand forces predicted using a Bi-LSTM model in Task 3.

Statistical Analyses

Separate four-way, repeated measures analyses of variance (ANOVAs) were used to assess the effects of *Mass, MMH task, Feature set,* and *RNN model* on both performance metrics (i.e., RMSE and PE). For each ANOVA model, biological sex (*Sex*) was included as blocking effect. All statistical analyses were performed with R (version 4.3.0, R Foundation for Statistical Computing, Vienna, Austria), using the restricted maximum likelihood (REML) method.

Parametric model assumptions were verified, and statistical significance was determined when $p < 0.05$. Significant interaction effects were explored using simple-effects testing, and *post hoc* paired comparisons were completed using the Dunn-Sidak procedure. We present interaction effects only up to second order, both since it is difficult to interpret higher-order effects and since only a single higher-order effect was significant (see Appendix). Furthermore, our presentation of the result emphasizes main effects, given the extensive statistical results obtained and, more importantly, since nearly all interaction effects were relatively minor (i.e., effects of one factor were generally consistent across levels of another factor). Where interaction effects were more substantial, these effects are noted.

3.4 Results

ANOVA results are summarized in Tables 3.1 – 3.6 Appendix A. There were main or interaction effects of *Mass, MMH tasks,* and *Feature set* for both performance metrics (RMSE and PE). Detailed results are provided below.

Bilateral Hand Force Prediction Performance

Horizontal Force (X-axis) Performance Metrics

There were significant main effects of *Mass, MMH task,* and *Feature set* on RMSE and PE for the right hand (Figures 3.6 and 3.7). In the 6 kg condition, both RMSE and PE were significantly smaller compared to the 9 and 12 kg conditions. Tasks 4 and 5 yielded significantly larger RMSE and PE compared to all other MMH tasks. Using the IPM_{ONLY} feature set led to significantly larger RMSE and PE than when using either the W-B or W-B + IPM feature sets. For the left hand, there were also main effects of *Mass, MMH task,* and *Feature set* on RMSE and PE, and the patterns of results were similar to those found for the right hand (Figures 3.18 and 3.19 Appendix A). However, left-hand RMSE and PE were also significantly affected by *Sex* and *RNN model*, respectively. Using the Bi-LSTM model led to significantly smaller mean PE (4.62 N) compared to using the BGRU (4.77 N), and RMSE was significantly larger among males (5.53 N) vs. females (4.73 N).

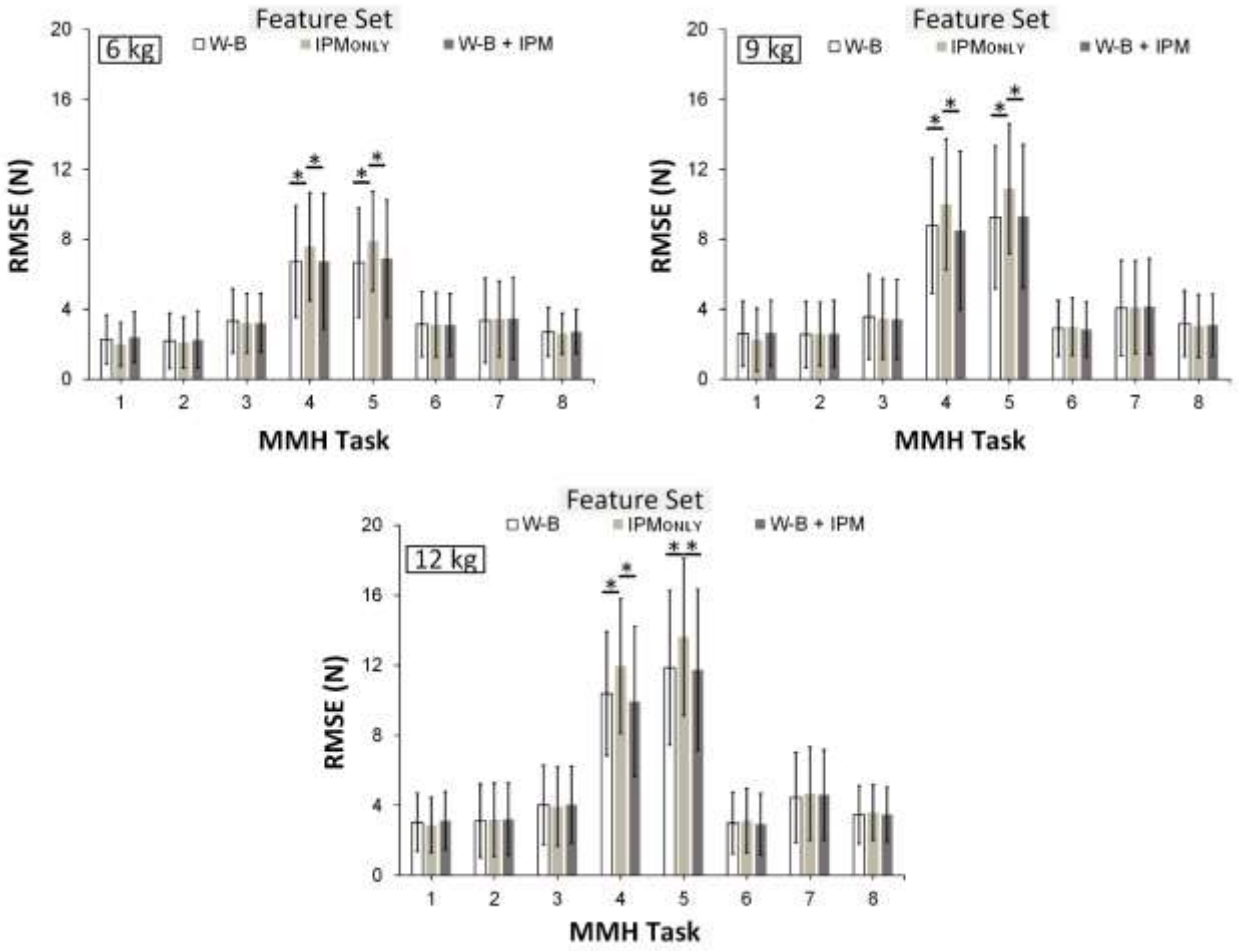

Figure 3.6. Root Mean Square Error (RMSE) in predicted forces for the right hand in the horizontal (X) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions. Here and in subsequent figures, error bars indicate standard deviation and the symbol * indicates a significant difference between paired means connected by the corresponding bar.

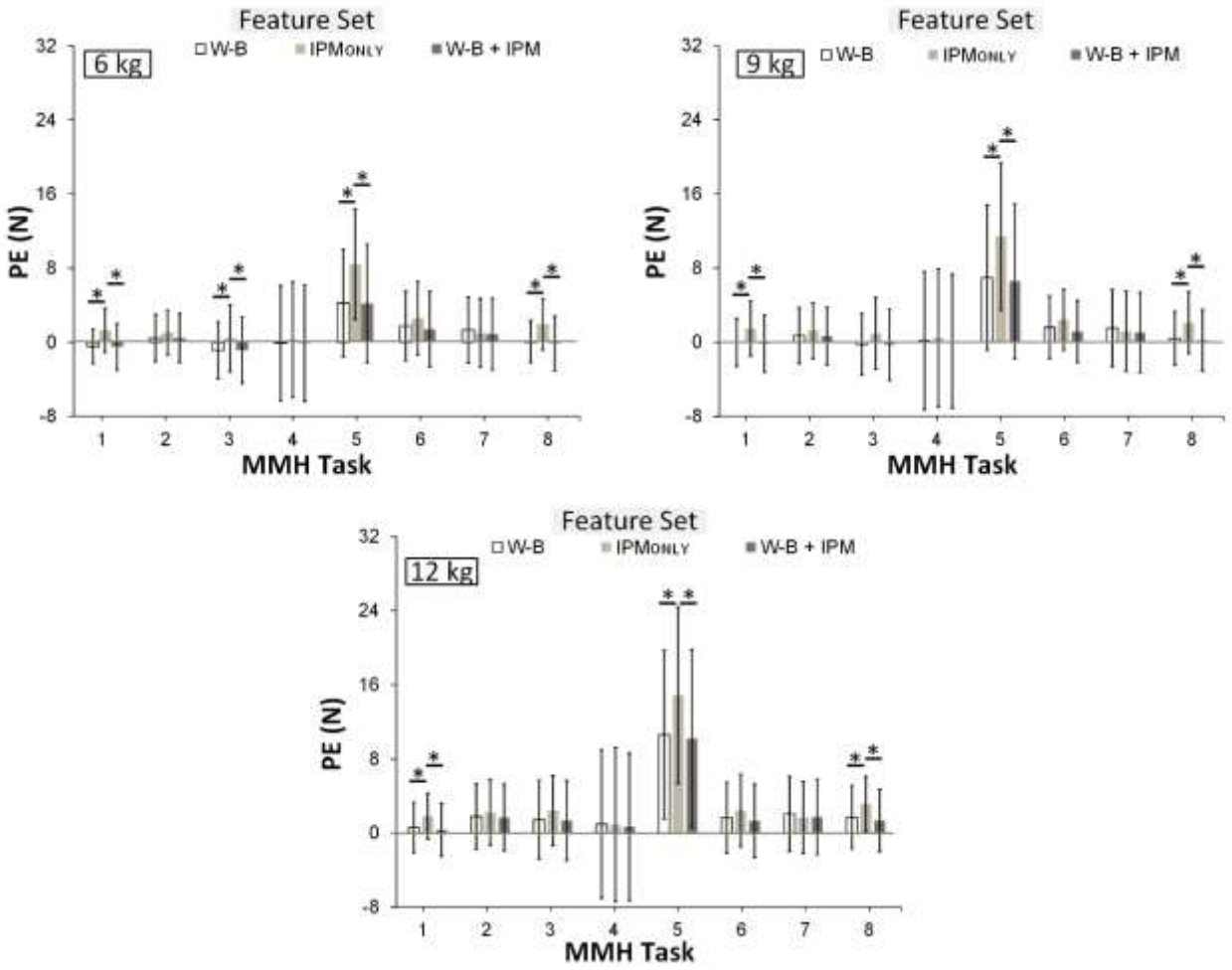

Figure 3.7. Peak Error (PE) in predicted forces for the right hand in the horizontal (X) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions. Here and below, PE > 0 and PE < 0 signify overpredicted and underpredicted peak hand forces, respectively.

Lateral Force (Y-axis) Performance Metrics

RMSE: Right hand RMSE was significantly affected by *Mass*, *MMH task*, and *Feature set* (Figure 3.8). In the 12 kg condition, RMSE was significantly larger compared to the 6 kg and 9 kg conditions. Tasks 4 and 5 had significantly smaller RMSE than all other MMH tasks.

Compared to using the W-B and W-B + IPM feature sets, using the IPM_{ONLY} feature set led to significantly poorer RMSE.

PE: There were significant main effects of *Mass, MMH task, Feature set,* and *RNN model* on right-hand PE (Figure 3.9). During the 6 kg condition, PE was significantly smaller than either the 9 kg or 12 kg conditions. Compared to other MMH tasks, Task 3 and Task 6 led to significantly larger PE. Using the IPM_{ONLY} feature set led to significantly smaller right hand PE, compared to the W-B feature set. While using either RNN model yielded similar PE for the right hand, using the Bi-LSTM model led to slightly smaller PE (2.6 vs 2.7 N) compared to using the BGRU model.

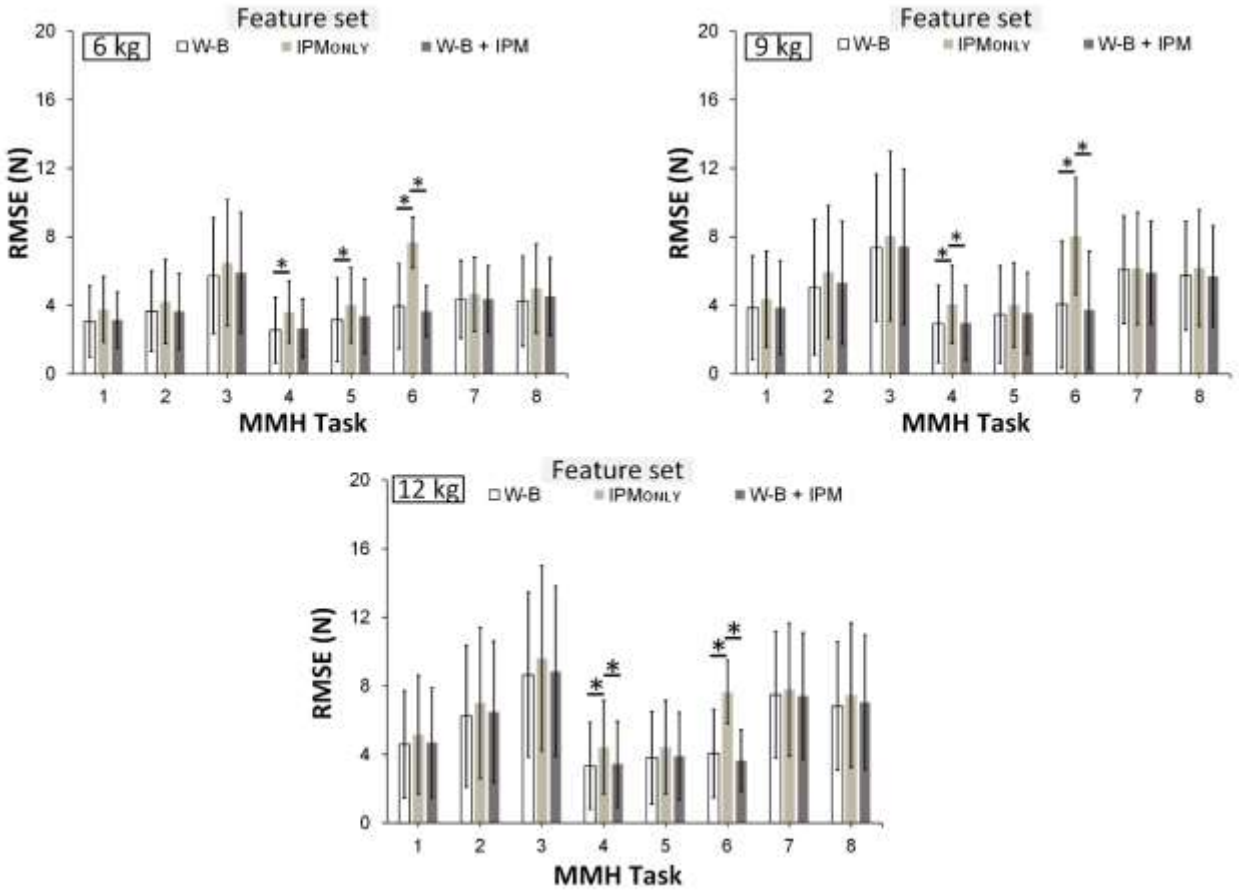

Figure 3.8. Root Mean Square Error (RMSE) in predicted forces on the right hand, in the lateral (Y) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 6, 9, and 12 kg mass conditions.

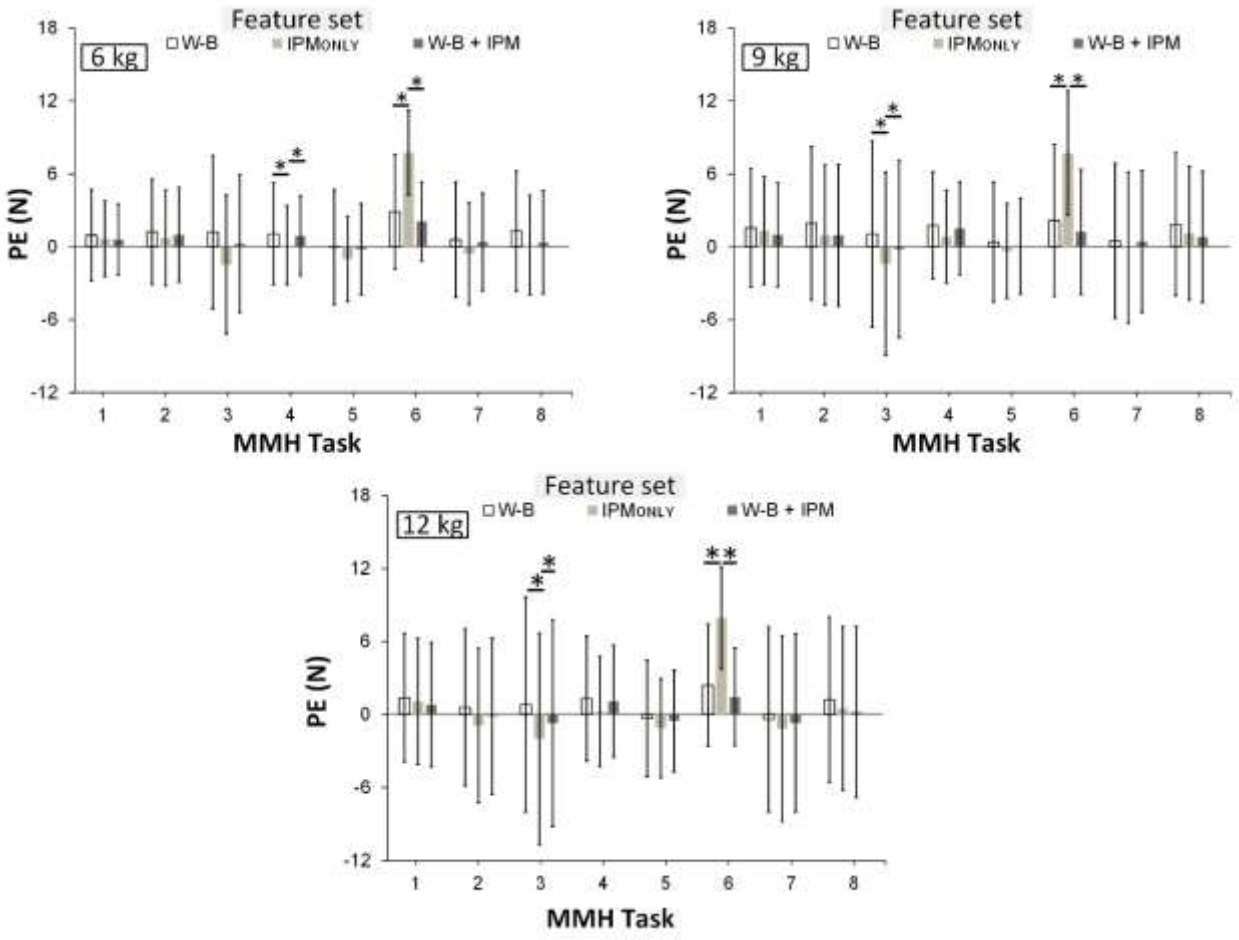

Figure 3.9. Peak Error (PE) in predicted forces on the right hand, in the lateral (Y) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 6, 9, and 12 kg mass conditions.

For the left-hand, both RMSE and PE were significantly affected by *Mass, MMH task,* and *Feature set*, while only left-hand PE was significantly affected by *RNN model* (Figures 3.20 and 3.21 Appendix A). The pattern of results was similar to that of the right hand.

Vertical Force (Z-axis) Performance Metrics

Both right hand RMSE and PE were significantly affected by *Mass*, *MMH task*, and *Feature set*. Also, *Sex* was significantly affected by RMSE, and *RNN model* was significantly affected by PE.

RMSE: In the 9 kg condition, RMSE was significantly smaller compared to the 12 kg condition. Tasks 4 and 5 yielded significantly poorer RMSE, compared to all other MMH tasks. Using W-B and W-B + IPM feature sets led to significantly smaller RMSE than using IPM_{ONLY} (Figure 3.10). RMSE was significantly smaller among females vs. males (15.2 vs. 16.7 N). The interaction effect of *Mass* × *MMH task* × *Feature set* on RMSE was substantial. Using the IPM_{ONLY} feature set yielded significantly larger RMSE, compared to 6 kg condition. Note that the magnitude of such differences varied between *MMH task* and *Feature set.* All simple effects were significant, except for the effect of *Mass* in the Tasks 4 ($p = 0.3$), 5 ($p = 1.0$), 6 ($p = 1.0$) for the W-B, IPM, and W-B + IPM feature set conditions

PE: During the 9 kg condition, PE was significantly smaller compared to the 6 and 12 kg conditions. Tasks 3, 5, and 6 yielded significantly smaller PE compared to all other MMH tasks. Compared to the W-B and W-B + IPM feature sets, using IPM_{ONLY} yielded significantly poorer PE (Figure 3.11). For a given RNN model, using BGRU model yielded significantly smaller PE (2.2 N) vs. Bi-LSTM model (2.5 N). The interaction effect of *Mass* × *MMH task* was substantial. Using any of the feature sets led to underprediction of PE in the 6 kg condition across the MMH tasks, while using any of the feature sets led to overprediction of PE in the 9 and 12 kg conditions in most of the MMH tasks. All simple effects were significant, except for the effect of *Mass* in the Tasks 4 ($p = 0.74$), 5 ($p = 0.93$), and 6 ($p = 0.45$) conditions.

For the left-hand RMSE and PE were both significantly affected by *Mass*, *MMH task*, and *Feature set* (Figures 3.22 and 3.23 Appendix A). Additionally, RMSE was significantly affected by *Sex*, and PE was significantly affected by *RNN model*. The pattern of results was similar to that of the right hand. Left hand PE also differed with *Sex*, being significantly smaller among females (0.9 N) vs. males (3.7 N).

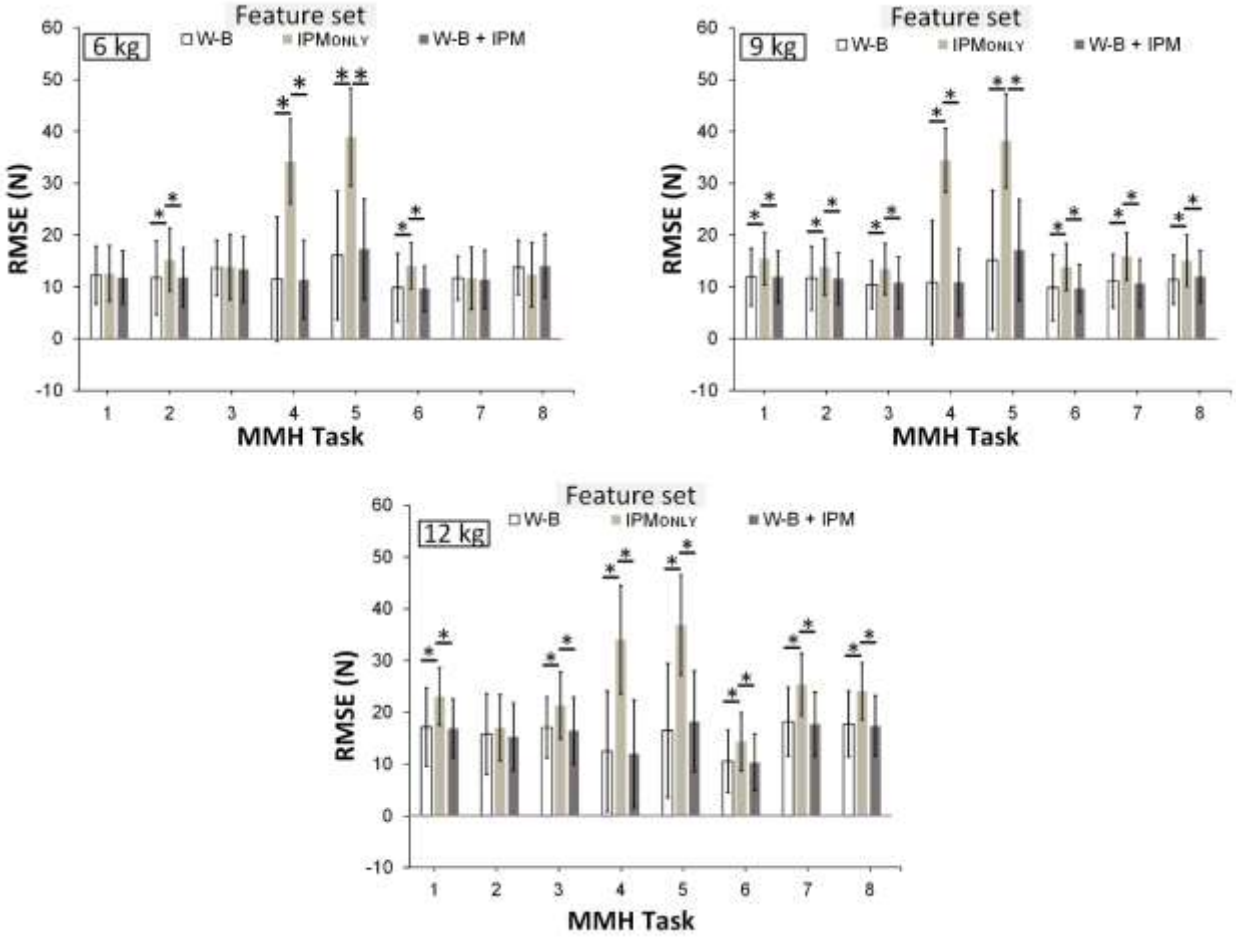

Figure 3.10. Root Mean Square Error (RMSE) in predicted forces on the right hand, in the Vertical (Z) direction, for different levels *MMH task* and *Feature set* shown separately for the 6, 9, and 12 kg mass conditions.

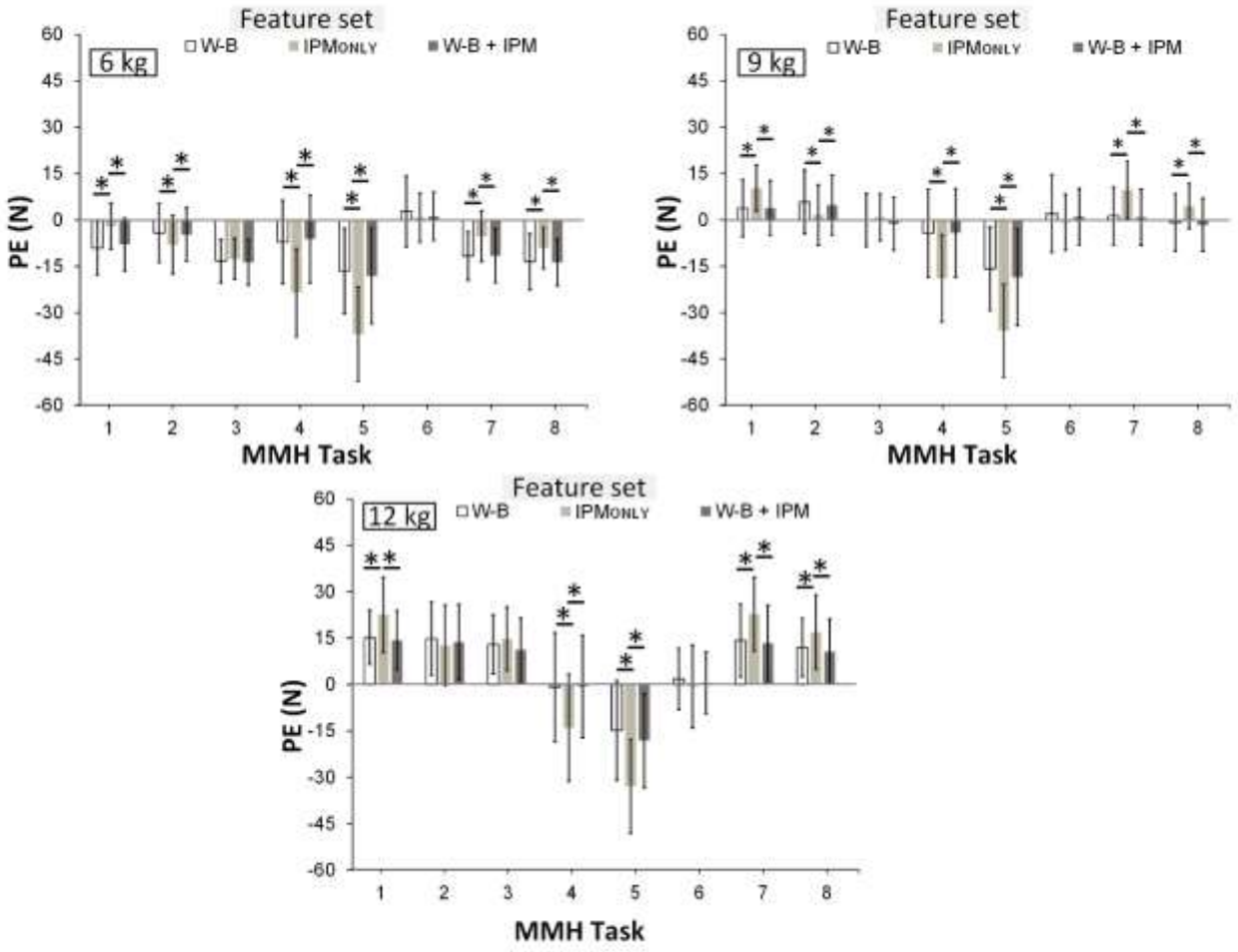

Figure 3.11. Peak Error (PE) in predicted forces on the right hand, in the Vertical (Z) direction, for different levels of *MMH task* and *Feature set* shown separately for the 6, 9, and 12 kg mass conditions.

Unilateral Hand Force Prediction Performance

Horizontal Force (X-axis) Performance Metrics

RMSE and PE were significantly affected by *Mass*, *MMH task*, and *Feature set* (Figures 3.12 and 3.13). During the 5 kg condition, RMSE and PE were significantly smaller, compared to the 7 and 9 kg conditions. Tasks 4 and 5 resulted in significantly larger RMSE and PE vs. all

other MMH tasks. Also, using the IPM_{ONLY} feature set led to significantly larger RMSE and PE compared to W-B and W-B + IPM feature sets.

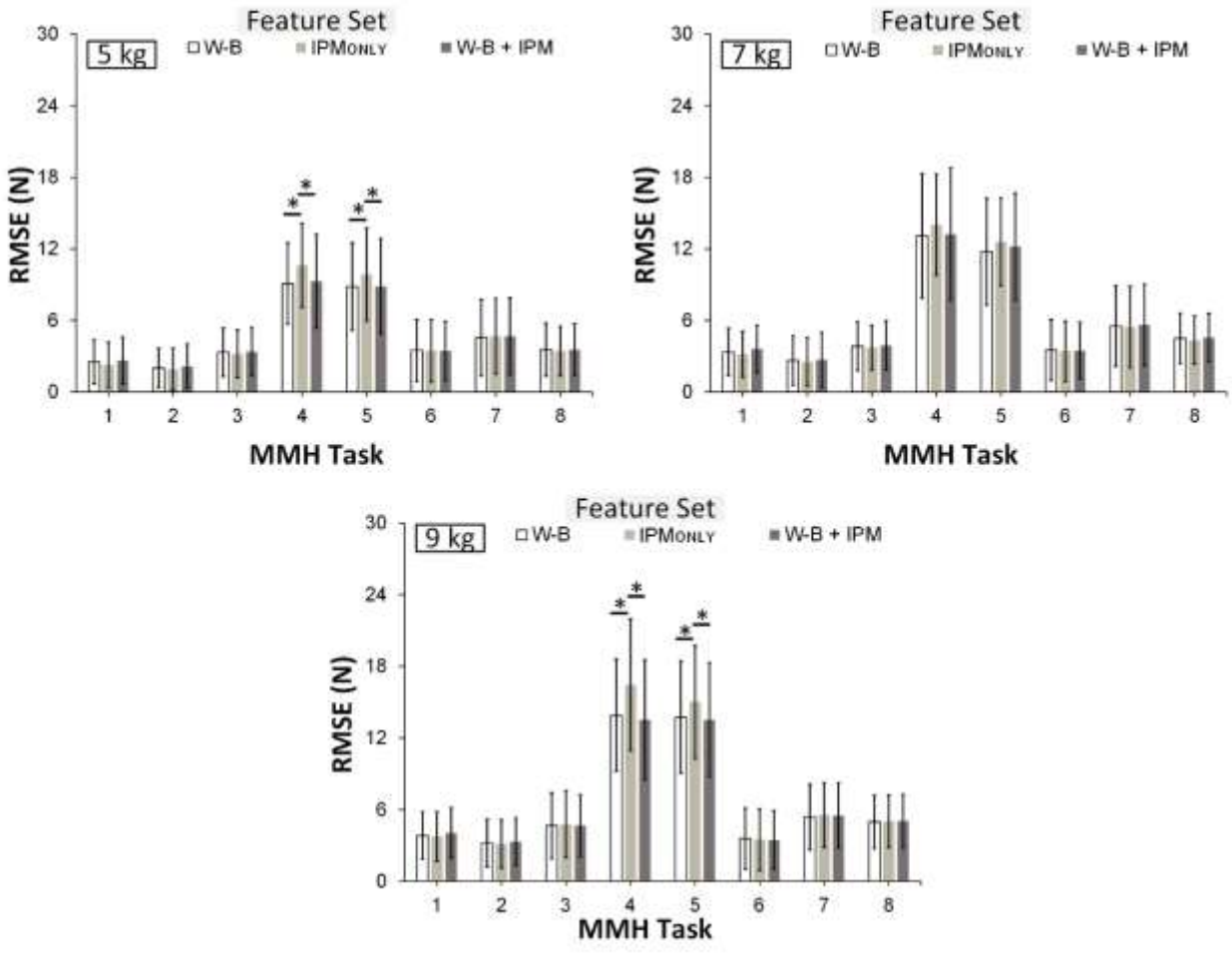

Figure 3.12. Root Mean Square Error in predicted forces in the horizontal (X) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7 and 9 kg mass conditions.

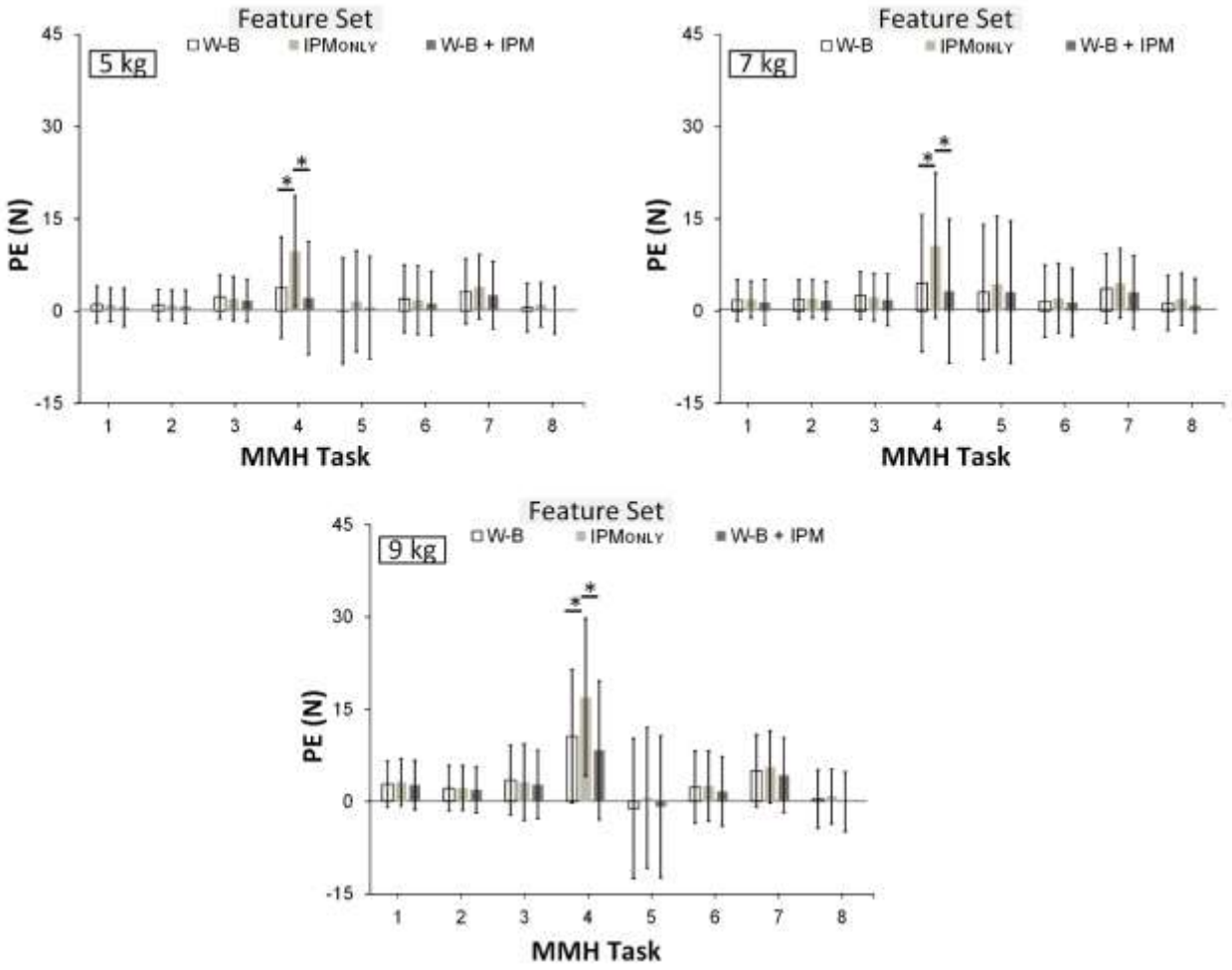

Figure 3.13. Peak Error (PE) in predicted forces in the horizontal (X) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7, and 9 kg mass conditions.

Lateral Force (Y-axis) Performance Metrics

There were significant main effects of both RMSE and PE on *Mass, MMH task, and Feature set* (Figures 3.14 and 3.15). Further, there was significant main effect of *Sex* on PE. In the 5 kg condition, both RMSE and PE were significantly smaller than in the 7 and 9 kg conditions. Tasks 4 and 5 exhibited smaller RMSE and PE in comparison to all other MMH tasks. Additionally, using the IPM_{ONLY} feature set yielded significantly larger RMSE and PE

compared to both W-B and W-B + IPM feature sets, and among males, PE was significantly smaller (1.6 vs. 1.3 N).

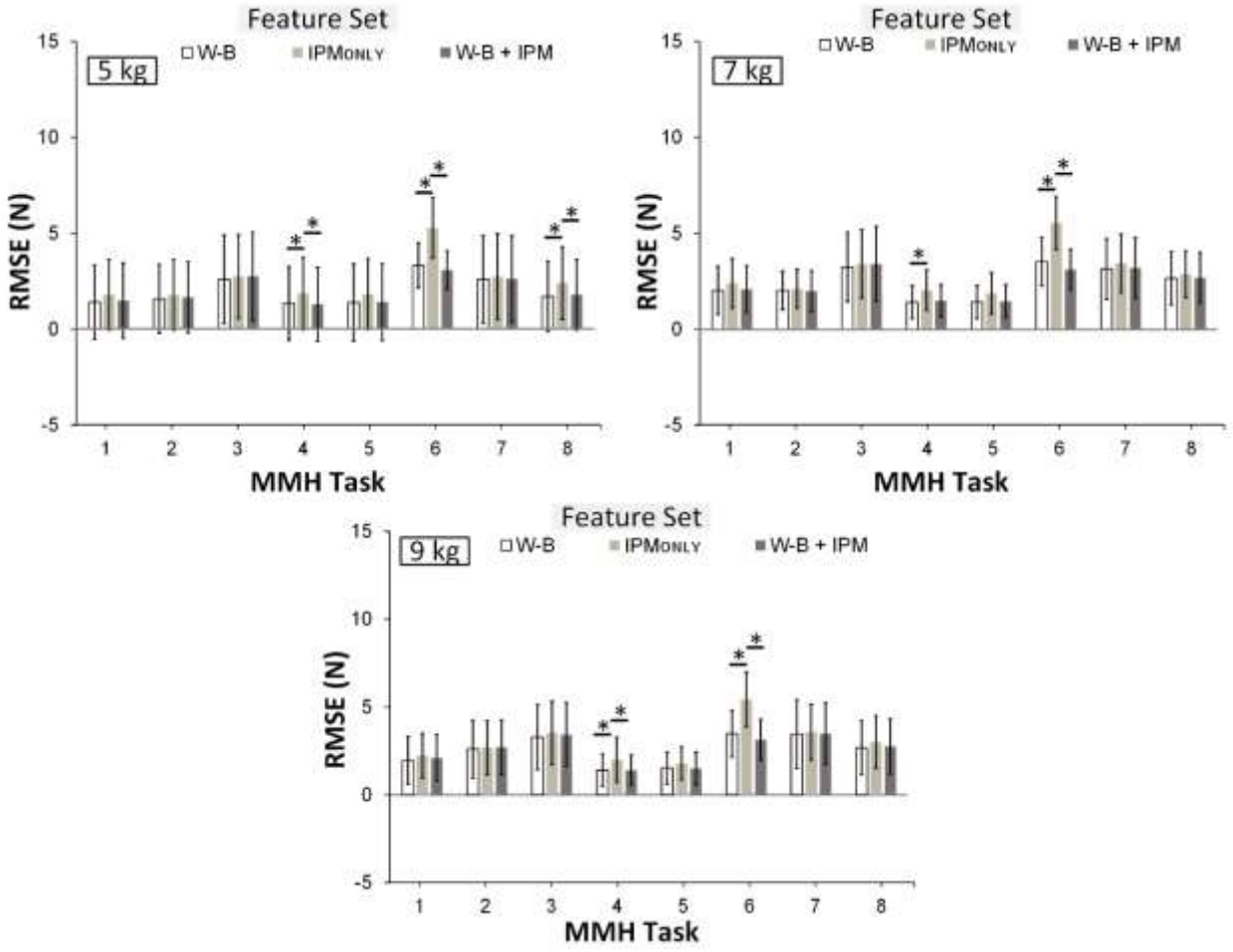

Figure 3.14. Root Mean Square Error in predicted forces in the lateral (Y) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7 and 9 kg mass conditions.

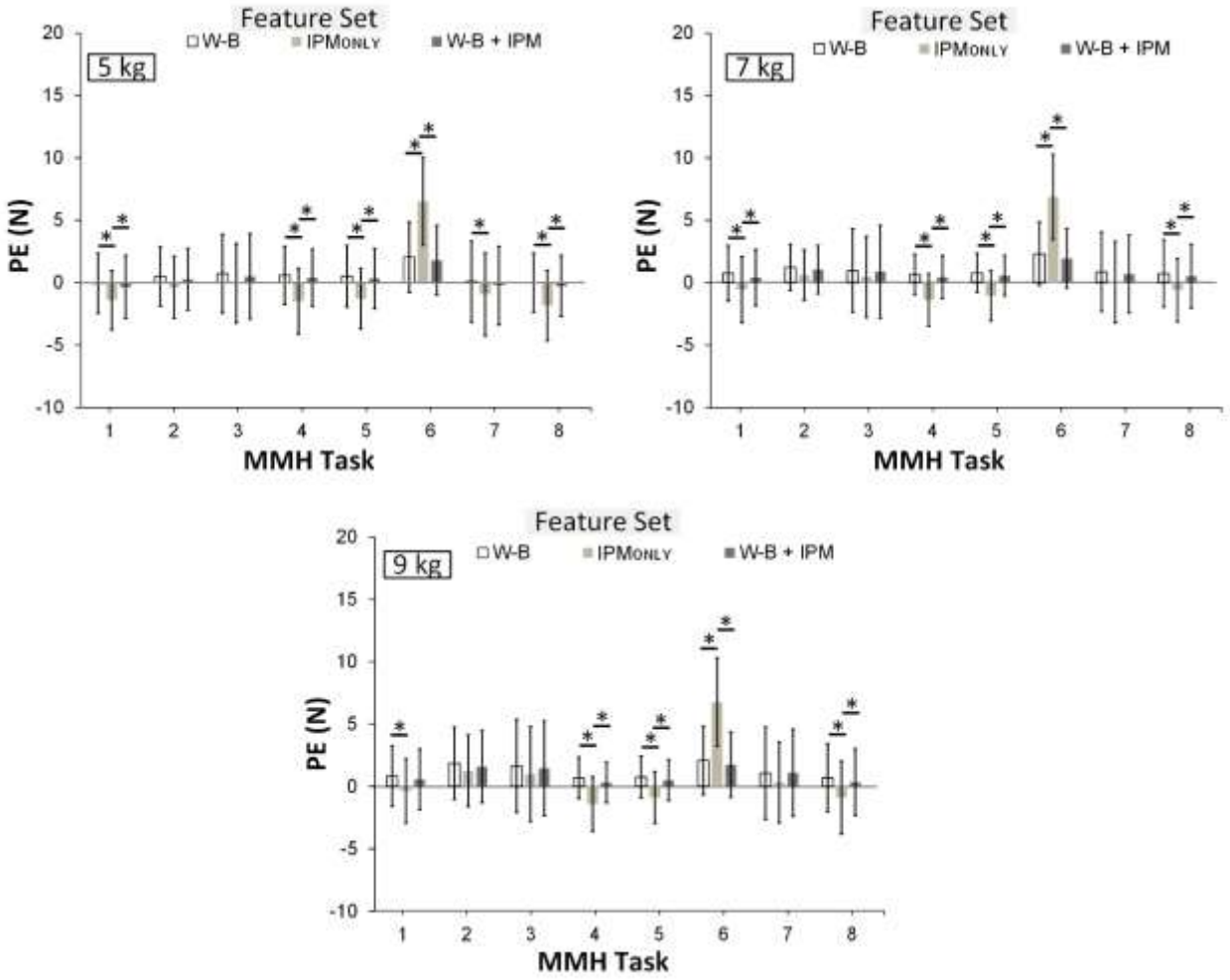

Figure 3.15. Peak Error (PE) in predicted forces in the lateral (Y) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7, and 9 kg mass conditions.

Vertical Force (Z-axis) Performance Metrics

There were significant main effects of *Mass, MMH task, Feature set, and Sex* on both RMSE and PE for unilateral hand (Figures 3.16 and 3.17).

RMSE: In the 9 kg condition, RMSE was significantly poorer compared to the 5 and 7 kg conditions. Tasks 4 and 5 yielded a significantly larger RMSE, compared to all other MMH Tasks. Using IPM_{ONLY} feature led to significantly larger RMSE than either W-B or W-B + IPM feature sets. RMSE was significantly smaller among females (18.3 N) vs. males (21.0 N).

Additionally, the interaction effect of *Mass* × *MMH task* × *Feature set* on RMSE was substantial. Using the IPM_{ONLY} feature set yielded significantly larger RMSE, compared to 6 kg condition. Note that the magnitude of such differences varied between *MMH task* and *Feature set.* All simple effects were significant, except for the effect of *Mass* in the Tasks 4 ($p = 0.17$), 5 ($p = 1.0$), 6 ($p = 1.0$) for the W-B, IPM, and W-B + IPM feature set conditions.

PE: In the 9 kg condition, hand forces were overpredicted (i.e., PE > 0), while hand forces were underpredicted (i.e., PE < 0) in the 5 kg condition. Tasks 3 and 8 yielded smaller PE compared to all other MMH tasks. For a given feature set, PE was small overall (-0.26 – 0.92 N), however, using W-B + IPM feature led to significantly smaller PE compared to IPM_{ONLY} and W-B feature sets. Among males, PE was significantly larger compared to females (-0.9 vs 1.9 N). Moreover, the interaction effect of *Mass* × *MMH task* was substantial. Across the MMH tasks, using any of the feature sets led to underprediction of PE in the 6 kg condition across the MMH tasks, while using any of the feature sets led to overprediction of PE in the 9 and 12 kg conditions in most of the MMH tasks. All simple effects are significant, except for the effects of *Mass* in Tasks 4 ($p = 1.0$), 6 ($p = 1.0$) conditions.

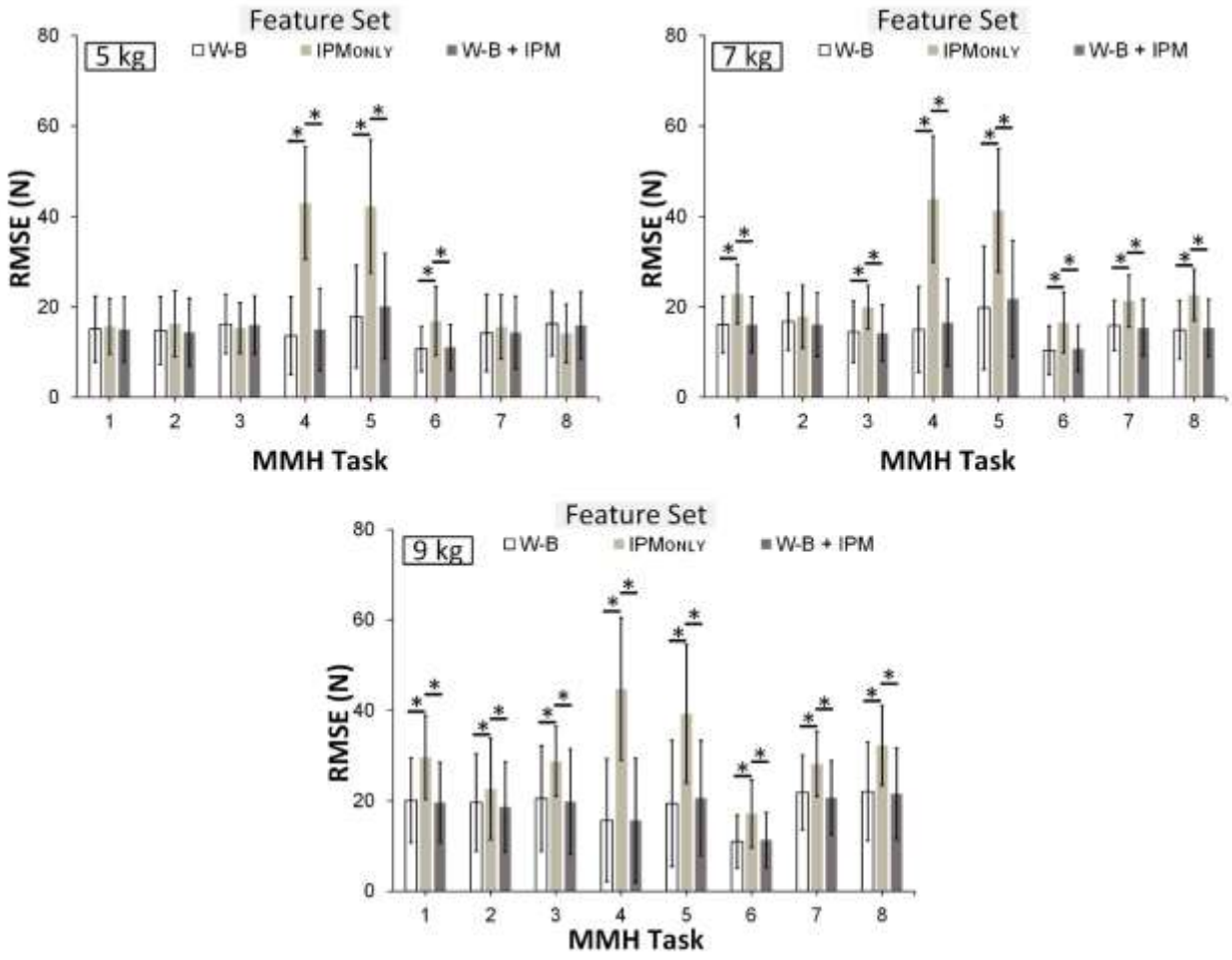

Figure 3.16. Root Mean Square Error in predicted forces in the vertical (Y) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7 and 9 kg mass conditions.

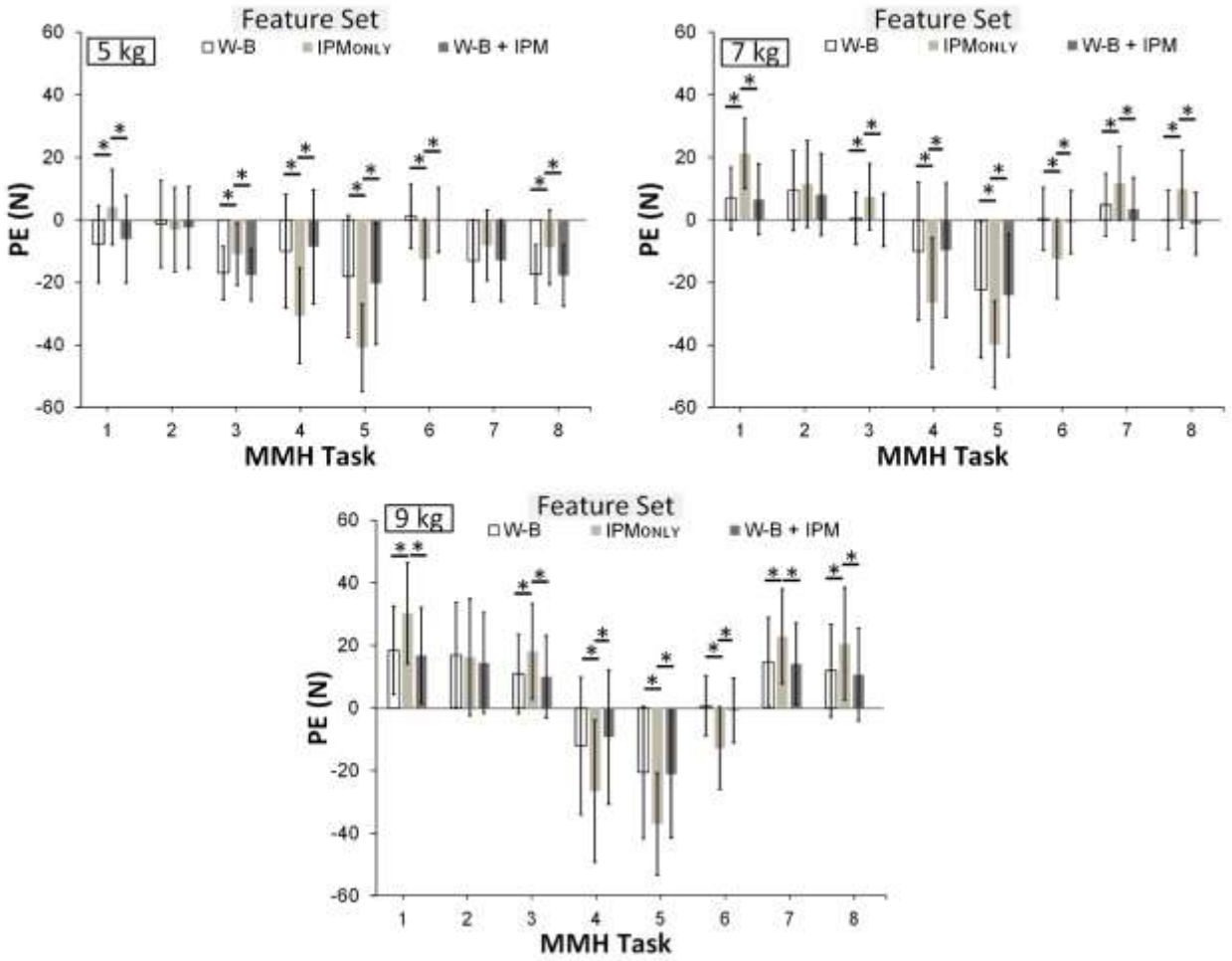

Figure 3.17. Peak Error (PE) in predicted forces in the vertical (Z) direction, for different levels of *MMH task* and *Feature set*, shown separately for the 5, 7, and 9 kg mass conditions.

3.5 Discussion

We investigated the feasibility of using other measures (i.e., kinematics from MMC and kinetics from IPM) for estimating external hand forces and the effectiveness of various machine learning algorithms. Integrating data from MMC and IPM, and leveraging machine learning algorithms, could enhance the efficiency of occupational exposure assessment methods, particularly by helping to estimate hand forces with the need for often-challenging direct measurements. Across the MMH tasks, feature sets, mass conditions and RNN models, both

right- and left-hand performance metrics in the horizontal direction were relatively small: RMSE ranged from 2.0 – 15.4 N, and PE ranged from -0.9 – 14.9 N. In the lateral direction, mean RMSE and PE for the right and left hands were 2.6 – 9.6 N and -2.0 – 7.9 N, respectively. However, RMSE and PE for both hands in the vertical direction were relatively larger and varied more substantially depending on MMH tasks, mass conditions, feature sets, RNN models, and between males and females. Specifically, RMSE for the right and left hands was 9.64 – 38.9 N, and PE was between -35.8 and 22.8 N. Similar patterns of results were observed for unilateral hand force prediction. In the horizontal and lateral directions, RMSE and PE were relatively small, ranging from 1.9 – 16.4 N and -1.8 – 9.7 N, respectively. However, in the vertical direction, both RMSE and PE were larger, with these metrics ranging from 10.3 – 44.7 N and -40.8 – 30.3 N, respectively.

Hand Force Predictions in different MMH tasks

Right and left hand mean RMSE in the vertical direction was 11.4 – 23.7 N, and mean PE in hand force prediction for the right and left hands ranged from -35.8 – 22.8 Nm, with both values depending on the specific MMH task (Figures 3.10 and 3.11). Earlier studies have shown that the performance of machine learning models often depends on the specific model and task type (Barazandeh et al., 2017; Kim & Nussbaum, 2014; Mokhlespour Esfahani, 2018). In these studies, several MMH tasks were simulated, such as symmetric and asymmetric lifting, pushing and pulling, and the findings indicated that machine learning models frequently misclassify tasks with similar motions (e.g., pushing vs. pulling). While these studies are not directly comparable to the current study, we saw a similar task-dependency, in that RMSE was higher in Tasks 4 and 5 (box pushing and pulling) by 68 – 108% compared to other MMH tasks.

In addition, right and left-hand RMSE was substantially larger for Tasks 4 and 5, compared to other MMH tasks, when predicting external hand forces in the horizontal direction (Figure 3.6) More specifically, Tasks 4 and 5 conditions led to substantially larger RMSE (8.2 – 9.7 N) than other MMH tasks (2.6 – 4.0 N). One possible explanation for why these particular Tasks had larger RMSE than other MMH tasks in the horizontal direction is that the kinematic and ground reaction feature sets may not capture all the relevant details needed to predict hand forces accurately. It is important to distinguish between pushing and pulling because they impose different biomechanical demands on the musculoskeletal system (Hoozemans et al., 2004). Acquiring more detailed input data by exploring additional sensor modalities could help enhance the ability of machine learning models to predict external hand forces during pushing and pulling tasks. Data extracted from a pair of tactile gloves, for example, were used to predict hand loads during MMH tasks involving different types of lifting, with an overall mean absolute error between 1.74—1.80 N (Zhou et al., 2023). Each pair of gloves had 65 pressure sensors to map the pressure exerted on the palmer side of the side. Using such devices could provide more information, improving the performance of machine learning algorithms in estimating external hand forces.

Use of Different Feature Sets on Hand Force Prediction

Hand force prediction performance varied substantially depending on the feature set. Compared to W-B and W-B + IPM, using the IPM_{ONLY} feature set increased RMSE substantially in the right hand by up to 7, 22, and 60% in the horizontal, lateral, and vertical directions, respectively. Similarly, using the IPM_{ONLY} feature set increased RMSE for one hand force prediction in the horizontal, lateral, and vertical directions by 5, 22, and 60%, respectively, compared to W-B feature set. One possible explanation for why using the IPM_{ONLY} feature set

led to larger RMSE is that while IPM quantifies ground reaction forces, it does not provide information on joint dynamics, making it challenging for RNN models to interpret how specific movements contribute to ground reaction forces. Another reason could be that the IPM might have moved slightly within the participants' shoes. Though not directly comparable, Jönsson et al. (2019) compared ground reaction forces estimated by force plates and IPM during quadriceps resistance exercises. In their study, they found differences of up to 8.3% in the two systems when measuring center-of-pressure at different surfaces. This difference was attributed to the slight displacement of IPM when performing the resistance exercises and to the inherent difference between the two systems when measuring the center of pressure at different surfaces. Thus, the types and fit of footwear could have substantially affected pressure patterns, making it difficult to generalize the weight across participants.

Effects of Load Mass on Hand Force Prediction

Our models generated larger mean RMSE when predicting right- and left-hand external hand forces for tasks done using the larger 12 kg mass. Specifically, in the 6 and 9 kg conditions, the right and left-hand RMSE in the vertical direction was 29% smaller compared to the 12 kg condition (Figures 3.10 and A.5). We saw similar patterns of results in the horizontal and lateral directions. Yet, for unilateral hand force prediction, larger RMSE was observed in the 7 kg and 9 kg conditions vs. the 5 kg condition. Specifically, in the 7 and 9 kg conditions, unilateral RMSE in the vertical direction was substantially larger (7 kg = 19.0 N; 9 kg =22. 6 N) compared to the 5 kg condition (17.5 N). These patterns of results for unilateral force prediction were similar in the horizontal and lateral directions.

The reason for this discrepancy in the effect of mass on RMSE between unilateral and bilateral predictions is not clear to us. A potential reason is that the unilateral models lacked

sufficient data to generalize well to unseen data. Specifically, our bilateral hand force prediction model was trained using ~51% more data than the unilateral hand force model. Consequently, although the RMSE was smaller in the 5 kg condition, the PE was substantially larger in the 5 kg condition (-8.3 N) compared to the 7 kg condition (1.5 N).

We suggest that future work is required to ensure the accuracy of machine learning models in capturing hand force patterns, particularly peak positions. It is important to identify peak hand forces because such loads are crucial for identifying and mitigating extreme conditions that may be associated with the development of WMSDs. Additionally, further investigation is warranted to better comprehend why prediction performance varies across unilateral and bilateral tasks and to ensure more consistent model performance across such tasks.

Differences Between Males and Females

Among females, right- and left-hand RMSE were smaller, by up to 11%. Also, left-hand PE among females was substantially smaller (0.9 N) than among males (3.71 N). This result was similar to what we found in Chapter 2 for task classification. Recall that the mean precision of MMH task classifiers decreased by up to 4% among males depending on the specific MMH task. There might be some effects of individual characteristics (i.e., *Sex*) in our model. Further investigation is warranted to quantify the sex-based differences when predicting external hand forces using machine learning models. Quantifying these differences could promote the development of an ergonomic tool that prioritizes inclusivity and equity in the workplace.

Limitations

A few limitations of this study need to be noted. First, participants were volunteers who did not have experience in performing MMH tasks. Experienced workers often use different approaches from novice workers when performing MMH tasks (Authier et al., 1996; Lee &

Nussbaum, 2012; Plamondon et al., 2014; Plamondon et al., 2010). Thus, caution should be exercised when generalizing this result to experienced workers. Second, participants in this study were all relatively young and healthy (i.e., 18 – 39 years old and exercised at least 3 three times weekly) and thus do not represent the full age of the working population. Therefore, caution is also warranted when extrapolating the findings to other populations, especially older individuals or those with musculoskeletal disorders. Third, while a range of MMH tasks were simulated, each task encompassed a limited selection of configurations, such as only two starting positions. Fourth, it is unclear whether the adjustments in our setup to accommodate individual anthropometry contributed to observed differences between males and females. Future work should investigate why inputs from MMC led to different performance between males and females.

Practical Implications of Our Findings

For future applications, we suggest that the BGRU model and W-B + IPM feature set could be used as the base approach for predicting external hand forces. We recommend the BGRU model over the Bi-LSTM because it yielded smaller PE (2.2 vs. 2.5 N) when predicting external hand forces in the vertical direction. Also, the BGRU model has a simplified architecture, consisting of two bidirectional gates, making it more computationally efficient and easier to train with large datasets (Khandelwal et al., 2016; Zhang & Xu, 2021). The W-B + IPM feature set is suggested, as it often yielded small PE for both the unilateral and bilateral prediction models.

Using MMC and IPM together with RNN-based models to predict external hand forces is a novel approach and warrants some comments. One advantage is that using data from MMC and IPM provides a non-intrusive approach to obtain whole-body kinematics and kinetics. Second,

RNN-based models effectively captured the complex and non-linear relationships between predictors and hand forces in the time series, resulting in reasonable RMSE. However, other aspects should be considered, such as some MMH tasks (e.g., pulling and pushing) and conditions involving unilateral hand forces. Third, the leave-one-subject-out validation approach used here mimics real-world scenarios by employing unseen data to evaluate the performance of the model, potentially indicating model performance in practice.

3.6 Conclusions

Three-dimensional, dynamic hand forces are needed for comprehensive and effective physical exposure assessment. Yet, these forces are difficult to obtain *in situ* partly due to the need for specialized equipment and expertise. We evaluated combining data from MMC and IPM, together with machine learning algorithms, to predict 3D hand forces for both one-handed and two-handed MMH tasks. Our results suggest that hand force predictions in both the vertical and horizontal directions were notably less accurate during pushing and pulling tasks. Also, in most of the tested conditions, using the IPM_{ONLY} feature set increased RMSE up to 60%. Overall, our findings indicate that the proposed approach has the potential to predict dynamic external hand forces, without direct measurement, offering a balance of simplicity and non-intrusiveness in quantifying physical exposure. Future work is needed, though, to improve the performance of the model in predicting hand forces during push and pull task conditions.

3.7 References

Arjmand, N., Gagnon, D., Plamondon, A., Shirazi-Adl, A., & Lariviere, C. (2010). A comparative study of two trunk biomechanical models under symmetric and asymmetric loadings. *Journal of Biomechanics*, *43*(3), 485-491.

Authier, M., Lortie, M., & Gagnon, M. (1996). Manual handling techniques: Comparing novices and experts. *International Journal of Industrial Ergonomics*, *17*(5), 419-429.

Barazandeh, B., Bastani, K., Rafieisakhaei, M., Kim, S., Kong, Z., & Nussbaum, M. A. (2017). Robust sparse representation-based classification using online sensor data for monitoring manual material handling tasks. *IEEE Transactions on Automation Science and Engineering*, *15*(4), 1573-1584.

Bernard, B. P., & Putz-Anderson, V. (1997). Musculoskeletal disorders and workplace factors; a critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back.

Burdorf, A. (1995). Reducing random measurement error in assessing postural load on the back in epidemiologic surveys. *Scandinavian Journal of Work, Environment & Health*, 15-23.

Burdorf, A., & van der Beek, A. (1999). Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. *Scandinavian Journal of Work, Environment & Health*, *25*, 25-30.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *Am J Ind Med*, *53*(3), 285-323.

Fu, R., Zhang, Z., & Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. 2016 31st Youth academic annual conference of Chinese association of automation (YAC),

Gallagher, S., Hamrick, C. A., Love, A. C., & Marras, W. S. (1994). Dynamic biomechanical modelling of symmetric and asymmetric lifting tasks in restricted postures. *Ergonomics*, *37*(8), 1289-1310.

Garg, A., Chaffin, D. B., & Freivalds, A. (1982). BIOMECHANICAL STRESSES FROM MANUAL LOAD LIFTING: A STATIC VS DYNAMIC EVALUATION. *IIE Transactions (Institute of Industrial Engineers)*, *14*(4), 272-281.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.

Hlucny, S. D., & Novak, D. (2020). Characterizing human box-lifting behavior using wearable inertial motion sensors. *Sensors*, *20*(8), 2323.

Hochreiter, S., & Schmidhuber, J. J. N. c. (1997). Long short-term memory. *9*(8), 1735-1780.

Hoozemans, M. J. M., Kuijer, P. P. F. M., Kingma, I., van Dieën, J. H., de Vries, W. H. K., van der Woude, L. H. V., Veeger, D. J., van der Beek, A. J., & Frings-Dresen, M. H. W. (2004). Mechanical loading of the low back and shoulders during pushing and pulling activities. *Ergonomics*, *47*(1), 1-18.

Jönsson, M., Munkhammar, T., Norrbrand, L., & Berg, H. E. (2019). Foot centre of pressure and ground reaction force during quadriceps resistance exercises; a comparison between force plates and a pressure insole system. *Journal of Biomechanics*, *87*, 206-210.

Jordao, A., Nazare Jr, A. C., Sena, J., & Schwartz, W. R. (2018). Human activity recognition based on wearable sensor data: A standardization of the state-of-the-art. *arXiv preprint arXiv:1806.05226*.

Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An empirical exploration of recurrent network architectures. International conference on machine learning,

Khandelwal, S., Lecouteux, B., & Besacier, L. (2016). *Comparing GRU and LSTM for automatic speech recognition* LIG].

Kim, S., & Nussbaum, M. A. (2014). An evaluation of classification algorithms for manual material handling tasks based on data obtained using wearable technologies. *Ergonomics*, *57*(7), 1040-1051.

Koppelaar, E., & Wells, R. (2005). Comparison of measurement methods for quantifying hand force. *Ergonomics*, *48*(8), 983-1007.

Kumar, S. (1993). Perception of posture of short duration in the spatial and temporal domains. *Applied Ergonomics*, *24*(5), 345-350.

Lee, H., Yang, K., Kim, N., & Ahn, C. R. (2020). Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. *Automation in Construction*, *120*, 103390.

Lee, J., & Nussbaum, M. A. (2012). Experienced workers exhibit distinct torso kinematics/kinetics and patterns of task dependency during repetitive lifts and lowers. *Ergonomics*, *55*(12), 1535-1547.

Lee, T.-H. (2015). The effects of load magnitude and lifting speed on the kinematic data of load and human posture. *International Journal of Occupational Safety and Ergonomics*, *21*(1), 55-61.

Lim, S. (2024). Exposures to Several Risk Factors can be Estimated from a Continuous Stream of Inertial Sensor Measurements during a Variety of Lifting-Lowering Tasks. *Available at SSRN 4684350*.

Lim, S., & D'Souza, C. (2019). Statistical prediction of load carriage mode and magnitude from inertial sensor derived gait kinematics. *Applied Ergonomics*, *76*, 1-11.

Mathiassen, S. E., & Winkel, J. (1991). Quantifying Variation in Physical Load Using Exposure-Vs-Time Data. *Ergonomics*, *34*(12), 1455-1468.

Mokhlespour Esfahani, M. I. (2018). *Development and Assessment of Smart Textile Systems for Human Activity Classification* Virginia Tech].

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Plamondon, A., Delisle, A., Bellefeuille, S., Denis, D., Gagnon, D., Larivière, C., & Group, I. M. R. (2014). Lifting strategies of expert and novice workers during a repetitive palletizing task. *Applied Ergonomics*, *45*(3), 471-481.

Plamondon, A., Denis, D., Delisle, A., Larivière, C., Salazar, E., & group, I. M. r. (2010). Biomechanical differences between expert and novice workers in a manual material handling task. *Ergonomics*, *53*(10), 1239-1253.

Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., & Kaufman, J. (2001). Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics*, *44*(6), 588-613.

Steinebach, T., Grosse, E. H., Glock, C. H., Wakula, J., & Lunin, A. (2020). Accuracy evaluation of two markerless motion capture systems for measurement of upper extremities: Kinect

V2 and Captiv. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *30*(4), 291-302.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.
U.S. Bureau of Labor Statistics. (2021). *Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work*. bls.gov. Retrieved May 7 from https://www.bls.gov/data/home.htm
van der Beek, A. J., & Frings-Dresen, M. (1998). Assessment of mechanical exposure in ergonomic epidemiology. *Occupational and Environmental Medicine*, *55*(5), 291-299.
Wang, M., Zhao, C., Barr, A., Fan, H., Yu, S., Kapellusch, J., & Harris Adamson, C. (2023). Hand posture and force estimation using surface electromyography and an artificial neural network. *Human Factors*, *65*(3), 382-402.
Wells, R., Norman, R., Neumann, P., Andrews, D., Frank, J., Shannon, H., & Kerr, M. (1997). Assessment of physical work load in epidemiologic studies: common measurement metrics for exposure assessment. *Ergonomics*, *40*(1), 51-61.
Wells, R., Van Eerd, D., & Hägg, G. (2004). Mechanical exposure concepts using force as the agent. *Scandinavian Journal of Work, Environment & Health*, 179-190.
Wiktorin, C., Hjelm, E. W., Winkel, J., Köster, M., & Group, S. M. I. S. (1996). Reproducibility of a questionnaire for assessment of physical load during work and leisure time. *Journal of Occupational and Environmental Medicine*, 190-201.
Wiktorin, C., Karlqvist, L., Winkel, J., & Group, S. M. I. S. (1993). Validity of self-reported exposures to work postures and manual materials handling. *Scandinavian Journal of Work, Environment & Health*, 208-214.
Winkel, J., & Mathiassen, S. E. (1994). Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. *Ergonomics*, *37*(6), 979-988.
Winter, D. A. (2009). Biomechanics and motor control of human movement. In (Third ed., pp. 49-50). John wiley & sons.
Yang, H., Haldeman, S., Lu, M. L., & Baker, D. (2016). Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey. *J Manipulative Physiol Ther*, *39*(7), 459-472.
Yang, K., Ahn, C. R., & Kim, H. (2020). Deep learning-based classification of work-related physical load levels in construction. *Advanced Engineering Informatics*, *45*, 101104-101104.
Zhang, T., & Xu, R. (2021). Performance comparisons of Bi-LSTM and Bi-GRU networks in Chinese word segmentation. Proceedings of the 2021 5th International Conference on Deep Learning Technologies,
Zhou, G., Aggarwal, V., Yin, M., & Yu, D. (2022). A Computer Vision Approach for Estimating Lifting Load Contributors to Injury Risk. *IEEE Transactions on Human-Machine Systems*.
Zhou, G., Lu, M.-L., & Yu, D. (2023). Tactile Gloves Predict Load Weight During Lifting With Deep Neural Networks. *IEEE Sensors Journal*.

Appendix A – ANOVA Results and Additional Figures

Table 3.1. Summary of ANOVA results for the effects of *Mass, MMH task, Feature set, RNN model*, and *Sex* on bilateral hand force prediction performance metrics s (RMSE and PE) in the horizontal direction. Entries are *F* values (*p* values) and significant effects are highlighted in bold font.

	Right Hand		Left Hand	
Effect	RMSE	PE	RMSE	PE
Mass (M)	**1560.09 (<.0001)**	**1611.95 (<.0001)**	**1871.84 (<.0001)**	**1780.66 (<.0001)**
MMH task (T)	**7722.18 (<.0001)**	**3806.12 (<.0001)**	**9147.47 (<.0001)**	**5934.18 (<.0001)**
Feature set (F)	**77.91 (<.0001)**	**1578.35 (<.0001)**	**20.19 (<.0001)**	**1349.9 (<.0001)**
RNN model (R)	0.01 (0.922)	0.68 (0.408)	0.16 (0.694)	**18.69 (<.0001)**
Sex (S)	0.81 (0.376)	0 (0.957)	**8.45 (0.006)**	3.27 (0.08)
M × T	**237.94 (<.0001)**	**209.34 (<.0001)**	**221.94 (<.0001)**	**162.69 (<.0001)**
M ×F	**7.28 (<.0001)**	0.18 (0.949)	1.6 (0.171)	0.93 (0.448)
T × F	**47.78 (<.0001)**	**191.77 (<.0001)**	**26.44 (<.0001)**	**89.71 (<.0001)**
M × R	0.05 (0.948)	0.21 (0.807)	0.07 (0.929)	0.14 (0.87)
T × R	0.36 (0.927)	0.63 (0.736)	0.58 (0.773)	1.25 (0.273)
F × R	0.1 (0.903)	**4.01 (0.018)**	0.06 (0.945)	2.94 (0.053)
M × S	**19.09 (<.0001)**	**4.06 (0.017)**	**32.19 (<.0001)**	**34.08 (<.0001)**
T × S	**48.72 (<.0001)**	**34.65 (<.0001)**	**176.82 (<.0001)**	**115.6 (<.0001)**
F × S	0.78 (0.457)	**21.23 (<.0001)**	**17.37 (<.0001)**	**9.71 (0)**
R × S	0.55 (0.458)	2.55 (0.111)	0.03 (0.857)	1.18 (0.277)
M × T × F	1.36 (0.097)	0.43 (0.996)	0.58 (0.964)	0.4 (0.998)
M× T × R	0.09 (1)	0.09 (1)	0.04 (1)	0.05 (1)
M × F × R	0.05 (0.996)	0.03 (0.998)	0.33 (0.858)	0.13 (0.97)
T × F × R	0.39 (0.978)	0.55 (0.902)	0.79 (0.682)	1.21 (0.258)
M × T × S	**4.69 (<.0001)**	**3.81 (<.0001)**	**8.93 (<.0001)**	**6.71 (<.0001)**
M × F × S	0.52 (0.721)	0.52 (0.719)	0.48 (0.754)	0.15 (0.962)
T × F × S	**2.11 (0.009)**	**3.52 (<.0001)**	**3.41 (<.0001)**	1.32 (0.187)
M × R × S	0.17 (0.847)	0.09 (0.916)	0.02 (0.984)	0.13 (0.882)
T × R × S	0.27 (0.966)	0.93 (0.478)	1.76 (0.09)	1.73 (0.097)
F × R × S	0.05 (0.952)	0.69 (0.501)	0.26 (0.772)	**6.61 (0.001)**
M × T × F × R	0.08 (1)	0.1 (1)	0.09 (1)	0.06 (1)
M × T × F × S	0.28 (1)	0.13 (1)	0.22 (1)	0.17 (1)
M × T × R × S	0.1 (1)	0.02 (1)	0.03 (1)	0.03 (1)
M × F × R × S	0.04 (0.997)	0.04 (0.997)	0.09 (0.985)	0.08 (0.99)
T × F × R × S	0.45 (0.958)	0.55 (0.902)	1.37 (0.16)	0.41 (0.972)
M × T × F × R × S	0.05 (1)	0.05 (1)	0.07 (1)	0.03 (1)

Table 3.2. Summary of ANOVA results for the effects of *Mass, MMH task, Feature set, RNN model* and *Sex* on bilateral force prediction performance metrics in the lateral direction. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

	Right Hand		Left Hand	
Effect	RMSE	PE	RMSE	PE
Mass (M)	**1313.22 (<.0001)**	**1805.89 (<.0001)**	**472.64 (<.0001)**	**435.31 (<.0001)**
MMH Task (T)	**1131.66 (<.0001)**	**357.25 (<.0001)**	**645.72 (<.0001)**	**194.47 (<.0001)**
Feature set (F)	**460.5 (<.0001)**	**118.34 (<.0001)**	**30.28 (<.0001)**	**89.94 (<.0001)**
RNN model (R)	0.93 (0.334)	**10.1 (0.002)**	2.77 (0.096)	**6.23 (0.013)**
Sex (S)	0.62 (0.436)	0.15 (0.699)	1.11 (0.3)	2.1 (0.156)
M × T	**74.46 (<.0001)**	**105.89 (<.0001)**	**34.53 (<.0001)**	**32.7 (<.0001)**
M ×F	0.5 (0.739)	2.65 (0.031)	0.53 (0.711)	1.48 (0.206)
T × F	**84.66 (<.0001)**	**118.64 (<.0001)**	**14.92 (<.0001)**	**32.82 (<.0001)**
M × R	0.04 (0.959)	0.29 (0.745)	0.01 (0.989)	0.04 (0.964)
T × R	0.83 (0.563)	0.97 (0.452)	0.28 (0.96)	0.22 (0.98)
F × R	0.48 (0.618)	**11.34 (<.0001)**	0.17 (0.843)	**8.53 (0)**
M × S	**4.94 (0.007)**	**11.46 (<.0001)**	**35.18 (<.0001)**	**63.95 (<.0001)**
T × S	**22.39 (<.0001)**	**24.57 (<.0001)**	**65.14 (<.0001)**	**78.64 (<.0001)**
F × S	**139.06 (<.0001)**	**22.44 (<.0001)**	**5.09 (0.006)**	0.07 (0.928)
R × S	1.39 (0.239)	**8.18 (0.004)**	0.17 (0.683)	0.12 (0.731)
M × T × F	0.46 (0.994)	0.84 (0.706)	0.1 (1)	0.27 (1)
M× T × R	0.08 (1)	0.1 (1)	0.02 (1)	0.02 (1)
M × F × R	0.17 (0.956)	0.03 (0.998)	0.28 (0.889)	0.08 (0.989)
T × F × R	0.17 (1)	1.27 (0.22)	0.14 (1)	0.15 (1)
M × T × S	**2.42 (0.002)**	**2.39 (0.003)**	**4.52 (<.0001)**	**6.03 (<.0001)**
M × F × S	1.66 (0.156)	0.37 (0.831)	0.27 (0.899)	0.25 (0.908)
T × F × S	**7.91 (<.0001)**	**5.77 (<.0001)**	0.53 (0.916)	1.67 (0.055)
M × R × S	0.54 (0.58)	0.62 (0.536)	0.03 (0.972)	0.04 (0.958)
T × R × S	0.14 (0.995)	0.38 (0.917)	0.22 (0.981)	0.44 (0.876)
F × R × S	0.26 (0.769)	**5.47 (0.004)**	0.05 (0.95)	**3.3 (0.037)**
M × T × F × R	0.06 (1)	0.08 (1)	0.04 (1)	0.03 (1)
M × T × F × S	0.39 (0.998)	0.44 (0.996)	0.1 (1)	0.17 (1)
M × T × R × S	0.06 (1)	0.06 (1)	0.03 (1)	0.06 (1)
M × F × R × S	0.07 (0.991)	0.38 (0.824)	0.21 (0.931)	0.1 (0.982)
T × F × R × S	0.72 (0.752)	0.49 (0.938)	0.28 (0.996)	0.3 (0.994)
M × T × F × R × S	0.07 (1)	0.05 (1)	0.04 (1)	0.04 (1)

Table 3.3. Summary of ANOVA results for the effects of *Mass, MMH task, Feature set, RNN model* and *Sex* on bilateral hand force prediction performance metrics in the vertical direction. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

	Right Hand		Left Hand	
Effect	RMSE	PE	RMSE	PE
Mass (M)	**1551.4 (<.0001)**	**8265.72 (<.0001)**	**1527.25 (<.0001)**	**8206.39 (<.0001)**
MMH Task (T)	**1581.59 (<.0001)**	**201.6 (<.0001)**	**1683.48 (<.0001)**	**260.33 (<.0001)**
Feature set (F)	**5825.25 (<.0001)**	**271.36 (<.0001)**	**5381.21 (<.0001)**	**368.79 (<.0001)**
RNN model (R)	0.39 (0.533)	**9.9 (0.002)**	0.02 (0.888)	**13.32 (0)**
Sex (S)	**6.01 (0.02)**	3.99 (0.054)	**6.58 (0.015)**	**4.55 (0.04)**
M × T	**136.15 (<.0001)**	**665.36 (<.0001)**	**115.04 (<.0001)**	**619.85 (<.0001)**
M ×F	**59.86 (<.0001)**	**6.51 (<.0001)**	**68.47 (<.0001)**	**7.18 (<.0001)**
T × F	**898.02 (<.0001)**	**99.64 (<.0001)**	**794.6 (<.0001)**	**53.09 (<.0001)**
M × R	0.55 (0.579)	0.22 (0.806)	2.4 (0.091)	0.07 (0.931)
T × R	0.66 (0.708)	0.62 (0.742)	0.49 (0.844)	0.94 (0.474)
F × R	1.17 (0.309)	1.47 (0.23)	0.68 (0.507)	0.07 (0.93)
M × S	**11.45 (<.0001)**	**2.71 (0.066)**	**8.2 (0)**	**7.55 (0.001)**
T × S	**19.15 (<.0001)**	**102.23 (<.0001)**	**59.49 (<.0001)**	**193.83 (<.0001)**
F × S	2.51 (0.081)	**183.54 (<.0001)**	**7.95 (0)**	**141.78 (<.0001)**
R × S	0.23 (0.633)	0.12 (0.726)	1.3 (0.254)	1.61 (0.204)
M × T × F	**20.97 (<.0001)**	1.35 (0.103)	**17.08 (<.0001)**	1.41 (0.075)
M× T × R	0.14 (1)	0.04 (1)	0.38 (0.981)	0.03 (1)
M × F × R	0.05 (0.995)	0.18 (0.951)	0.2 (0.937)	0.19 (0.944)
T × F × R	0.43 (0.964)	0.33 (0.99)	0.45 (0.957)	0.47 (0.95)
M × T × S	**3.56 (<.0001)**	**5.56 (<.0001)**	**2.44 (0.002)**	**4.24 (<.0001)**
M × F × S	**8.73 (<.0001)**	0.63 (0.645)	**2.66 (0.031)**	0.88 (0.477)
T × F × S	**2.87 (0)**	**14.88 (<.0001)**	**5.99 (<.0001)**	**24.33 (<.0001)**
M × R × S	0.1 (0.906)	0.01 (0.989)	0.27 (0.765)	0.11 (0.898)
T × R × S	0.7 (0.669)	0.44 (0.877)	0.74 (0.636)	0.45 (0.87)
F × R × S	0.58 (0.563)	**4.23 (0.015)**	0.25 (0.778)	**3.22 (0.04)**
M × T × F × R	0.09 (1)	0.07 (1)	0.07 (1)	0.07 (1)
M × T × F × S	**1.99 (0.001)**	0.54 (0.976)	1.35 (0.104)	0.43 (0.997)
M × T × R × S	0.06 (1)	0.06 (1)	0.05 (1)	0.05 (1)
M × F × R × S	0.48 (0.751)	0.12 (0.977)	0.36 (0.838)	0.2 (0.941)
T × F × R × S	0.16 (1)	0.51 (0.927)	0.22 (0.999)	0.46 (0.955)
M × T × F × R × S	0.21 (1)	0.07 (1)	0.14 (1)	0.08 (1)

Table 3.4. Summary of ANOVA results for the effects of *Mass, MMH task, Feature set, RNN model* and *Sex* on unilateral force prediction performance metrics in the horizontal direction. Entries are *F* values (*p* values) and significant effects are highlighted in bold font.

Effect	RMSE	PE
Mass (M)	**825.77 (<.0001)**	**828.22 (<.0001)**
MMH Task (T)	**4940.23 (<.0001)**	**2428.79 (<.0001)**
Feature set (F)	**16.46 (<.0001)**	**532.65 (<.0001)**
RNN model (R)	0.11 (0.7415)	3.19 (0.0743)
Sex (S)	0.60 (0.4444)	0.14 (0.7064)
M × T	**96.31 (<.0001)**	**74.76 (<.0001)**
M ×F	**5.72 (0.0001)**	0.68 (0.6035)
T × F	**16.23 (<.0001)**	**79.24 (<.0001)**
M × R	0.23 (0.7944)	0.12 (0.8906)
T × R	0.15 (0.9935)	0.22 (0.9801)
F × R	0.14 (0.8721)	**5.067 (0.0063)**
M × S	**5.09 (0.0062)**	**9.53 (0.0001)**
T × S	**45.21 (<.0001)**	**51.85 (<.0001)**
F × S	2.55 (0.0785)	**10.52 (<.0001)**
R × S	0.06 (0.8153)	1.30 (0.2543)
M × T × F	**1.70 (0.012)**	0.34 (0.9996)
M× T × R	0.16 (0.9998)	0.10 (1)
M × F × R	0.14 (0.9663)	0.09 (0.9858)
T × F × R	0.48 (0.9441)	0.52 (0.9234)
M × T × S	**2.31 (0.0036)**	**3.14 (0.0001)**
M × F × S	1.11 (0.3512)	0.67 (0.6144)
T × F × S	0.44 (0.9625)	0.48 (0.9456)
M × R × S	0.07 (0.9353)	0.04 (0.9644)
T × R × S	0.12 (0.9974)	0.19 (0.9883)
F × R × S	0.09 (0.9103)	**4.73 (0.0089)**
M × T × F × R	0.08 (1)	0.08 (1)
M × T × F × S	0.52 (0.9828)	0.25 (1)
M × T × R × S	0.05 (1)	0.07 (1)
M × F × R × S	0.14 (0.9674)	0.03 (0.9986)
T × F × R × S	0.25 (0.998)	0.9 (0.5641)
M × T × F × R × S	0.08 (1)	0.09 (1)

Table 3.5. Summary of ANOVA results for the effects of *Mass*, *MMH task*, *Feature set*, *RNN model* and *Sex* on unilateral force prediction performance metrics in the lateral direction. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

Effect	RMSE	PE
Mass (M)	**183.36 (<.0001)**	**307.24 (<.0001)**
MMH Task (T)	**740.65 (<.0001)**	**450.04 (<.0001)**
Feature set (F)	**247.49 (<.0001)**	**32.79 (<.0001)**
RNN model (R)	0.03 (0.8615)	0.37 (0.5443)
Sex (S)	0.12 (0.7269)	**5.0 (0.0321)**
M × T	**14.06 (<.0001)**	**21.30 (<.0001)**
M ×F	0.84 (0.499)	0.92 (0.4493)
T × F	**47.61 (<.0001)**	**202.88 (<.0001)**
M × R	0.25 (0.7766)	0.04 (0.957)
T × R	0.13 (0.9966)	0.24 (0.9746)
F × R	0.17 (0.8405)	2.14 (0.1175)
M × S	**64.98 (<.0001)**	**57.85 (<.0001)**
T × S	**22.77 (<.0001)**	**31.35 (<.0001)**
F × S	1.86 (0.156)	**3.92 (0.0199)**
R × S	0.04 (0.8482)	**9.67 (0.0019)**
M × T × F	0.50 (0.9871)	0.18 (1)
M× T × R	0.06 (1)	0.04 (1)
M × F × R	0.14 (0.968)	0.09 (0.9856)
T × F × R	0.19 (0.9995)	0.29 (0.9951)
M × T × S	**6.14 (<.0001)**	**6.55 (<.0001)**
M × F × S	0.66 (0.6212)	0.58 (0.676)
T × F × S	**6.20 (<.0001)**	**3.23 (<.0001)**
M × R × S	0.20 (0.8228)	0.30 (0.7372)
T × R × S	0.10 (0.9982)	0.75 (0.6297)
F × R × S	0.15 (0.8615)	**3.29 (0.0371)**
M × T × F × R	0.05 (1)	0.02 (1)
M × T × F × S	0.37 (0.9991)	0.29 (0.9999)
M × T × R × S	0.08 (1)	0.03 (1)
M × F × R × S	0.15 (0.9632)	0.07 (0.9905)
T × F × R × S	0.10 (1)	0.62 (0.8536)
M × T × F × R × S	0.07 (1)	0.05 (1)

Table 3.6. Summary of ANOVA results for the effects of *Mass, MMH task, Feature set, RNN model* and *Sex* on unilateral force prediction performance metrics in the vertical direction. Entries are *F* values (*p* values), and significant effects are highlighted in bold font.

Effect	RMSE	PE
Mass (M)	**678.25 (<.0001)**	**3082.5 (<.0001)**
MMH Task (T)	**677.79 (<.0001)**	**813.11 (<.0001)**
Feature set (F)	**2725.88 (<.0001)**	**22.55 (<.0001)**
RNN model (R)	0 (0.9651)	0.04 (0.8274)
Sex (S)	**6.49 (0.0155)**	**7.28 (0.0108)**
M × T	**59.61 (<.0001)**	**276.66 (<.0001)**
M ×F	**51.03 (<.0001)**	**5.25 (0.0003)**
T × F	**348.94 (<.0001)**	**163.88 (<.0001)**
M × R	0.12 (0.8847)	0.25 (0.7789)
T × R	0.13 (0.9958)	0.36 (0.9248)
F × R	0.10 (0.9007)	1.74 (0.1753)
M × S	**45.64 (<.0001)**	**34.88 (<.0001)**
T × S	**13.31 (<.0001)**	**82.71 (<.0001)**
F × S	**15.83 (<.0001)**	**20.34 (<.0001)**
R × S	0.20 (0.6536)	**6.95 (0.0084)**
M × T × F	**11.42 (<.0001)**	0.74 (0.8393)
M× T × R	0.07 (1)	0.08 (1)
M × F × R	0.38 (0.8196)	0.1 (0.9837)
T × F × R	0.37 (0.982)	0.46 (0.9543)
M × T × S	**4.22 (<.0001)**	**2.50 (0.0016)**
M × F × S	1.36 (0.2455)	1.04 (0.3842)
T × F × S	**1.77 (0.0364)**	**3.94 (<.0001)**
M × R × S	2.49 (0.0831)	0.12 (0.8896)
T × R × S	0.28 (0.9602)	0.39 (0.9067)
F × R × S	0.42 (0.6586)	0.40 (0.673)
M × T × F × R	0.09 (1)	0.10 (1)
M × T × F × S	0.55 (0.9733)	0.45 (0.9947)
M × T × R × S	0.16 (0.9998)	0.06 (1)
M × F × R × S	0.14 (0.9674)	0.04 (0.9971)
T × F × R × S	0.27 (0.9965)	0.25 (0.9975)
M × T × F × R × S	0.10 (1)	0.08 (1)

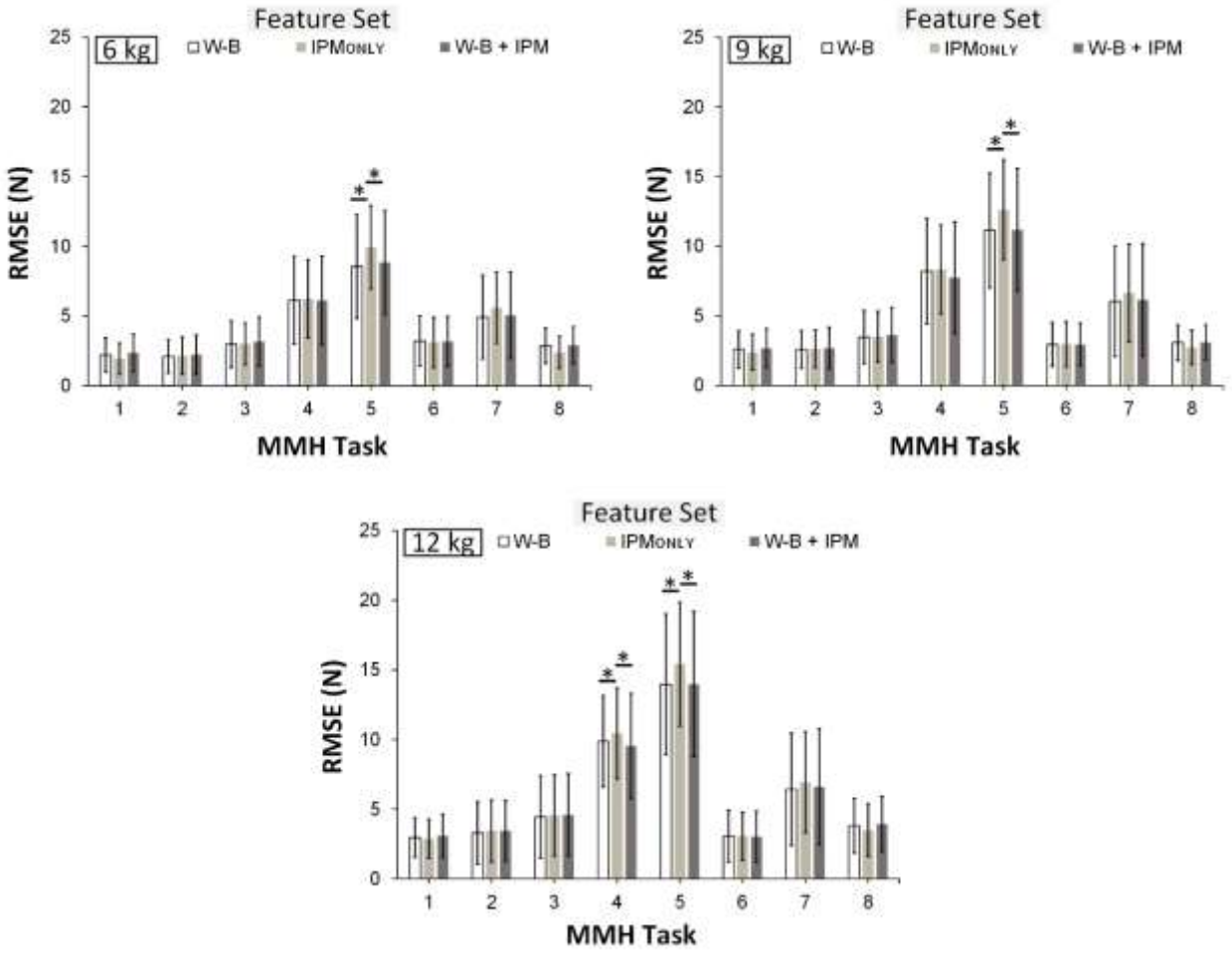

Figure 3.18. Root Mean Square Error (RMSE) in predicted forces for the left hand in the horizontal (X) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

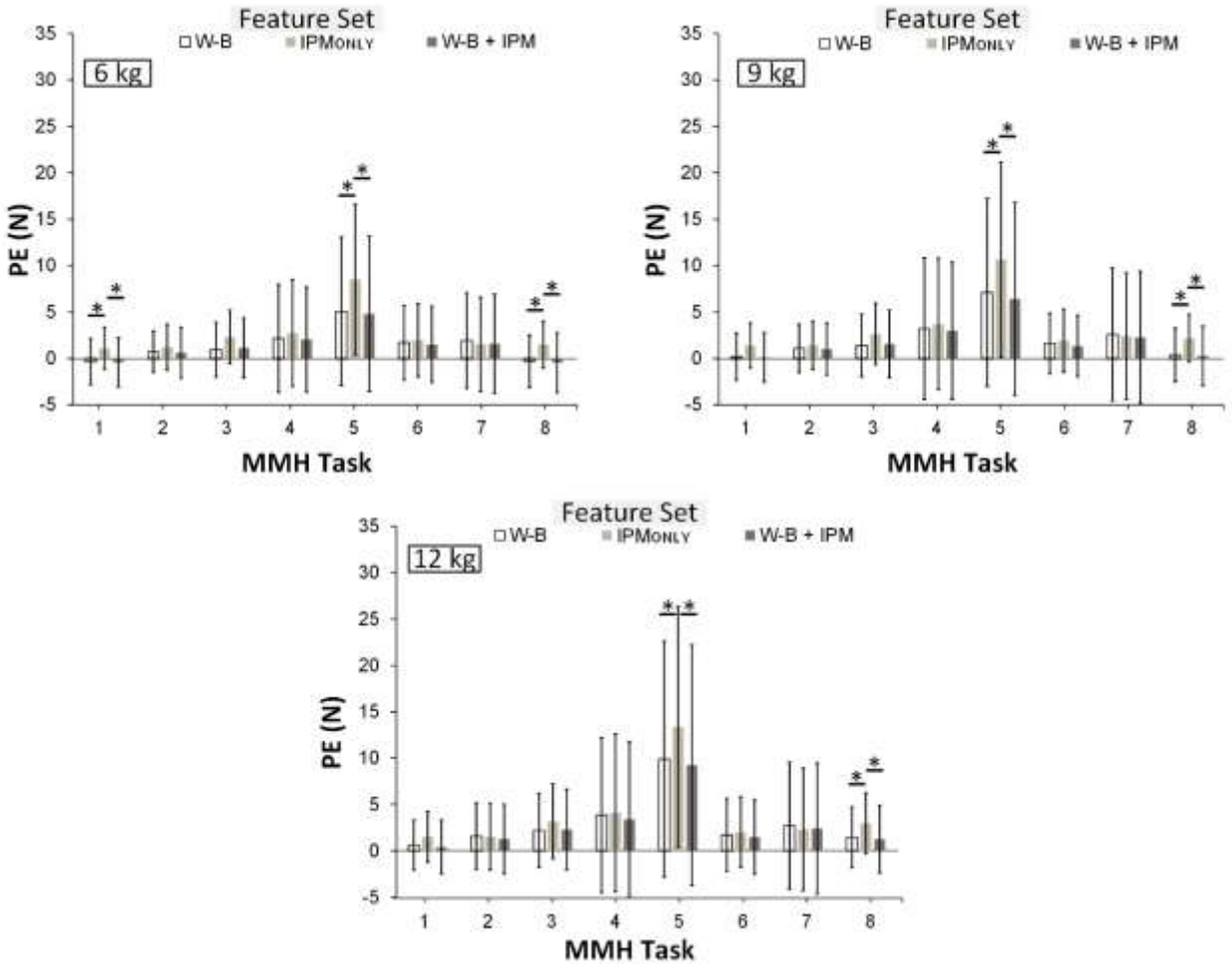

Figure 3.19. Peak Error (PE) in predicted forces for the left hand in the horizontal (X) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

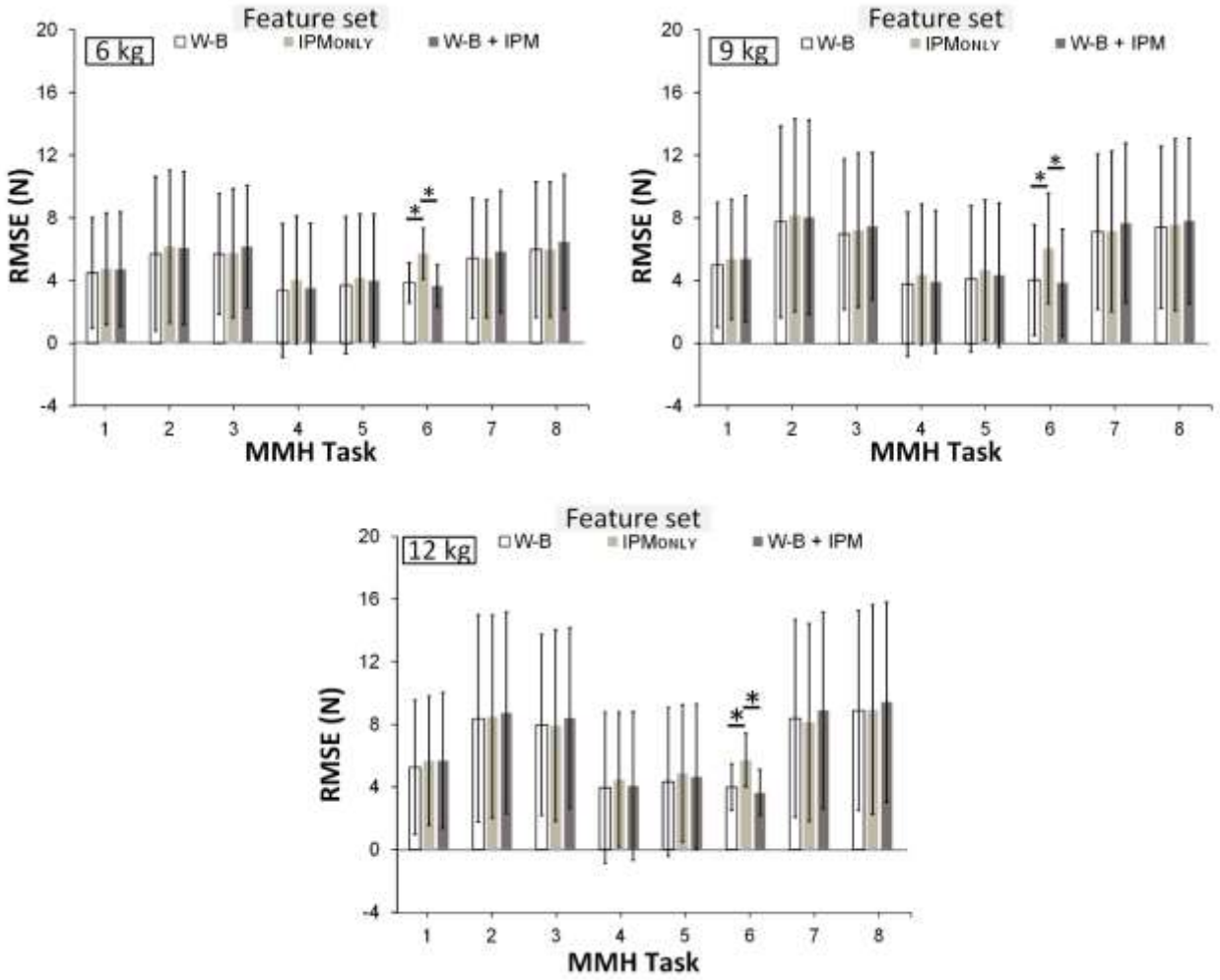

Figure 3.20. Root Mean Square Error (RMSE) in predicted forces for the left hand in the lateral (Y) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

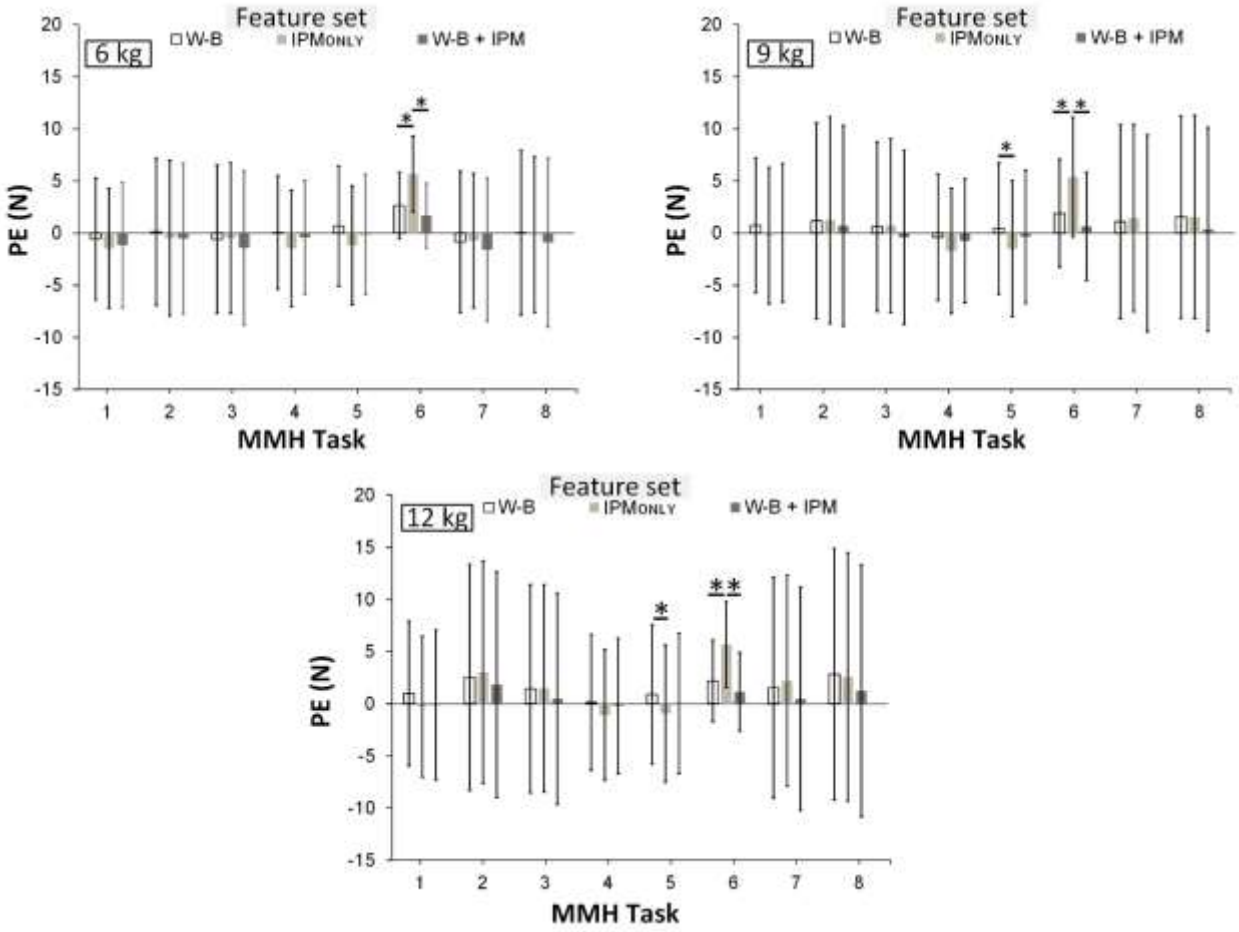

Figure 3.21. Peak Error (PE) in predicted forces for the left hand in the lateral (Y) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

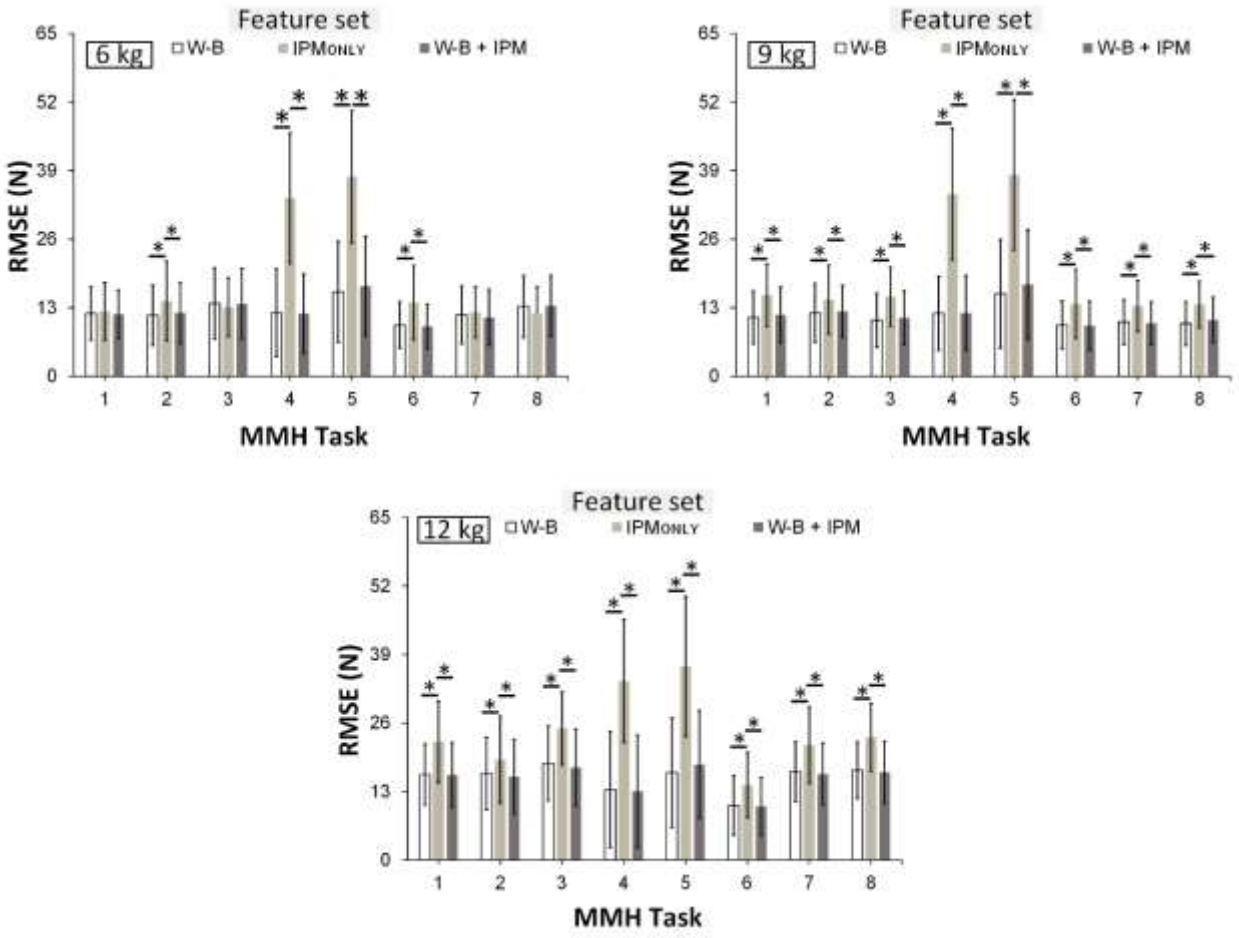

Figure 3.22. Root Mean Square Error (RMSE) in predicted forces for the left hand in the vertical (Z) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

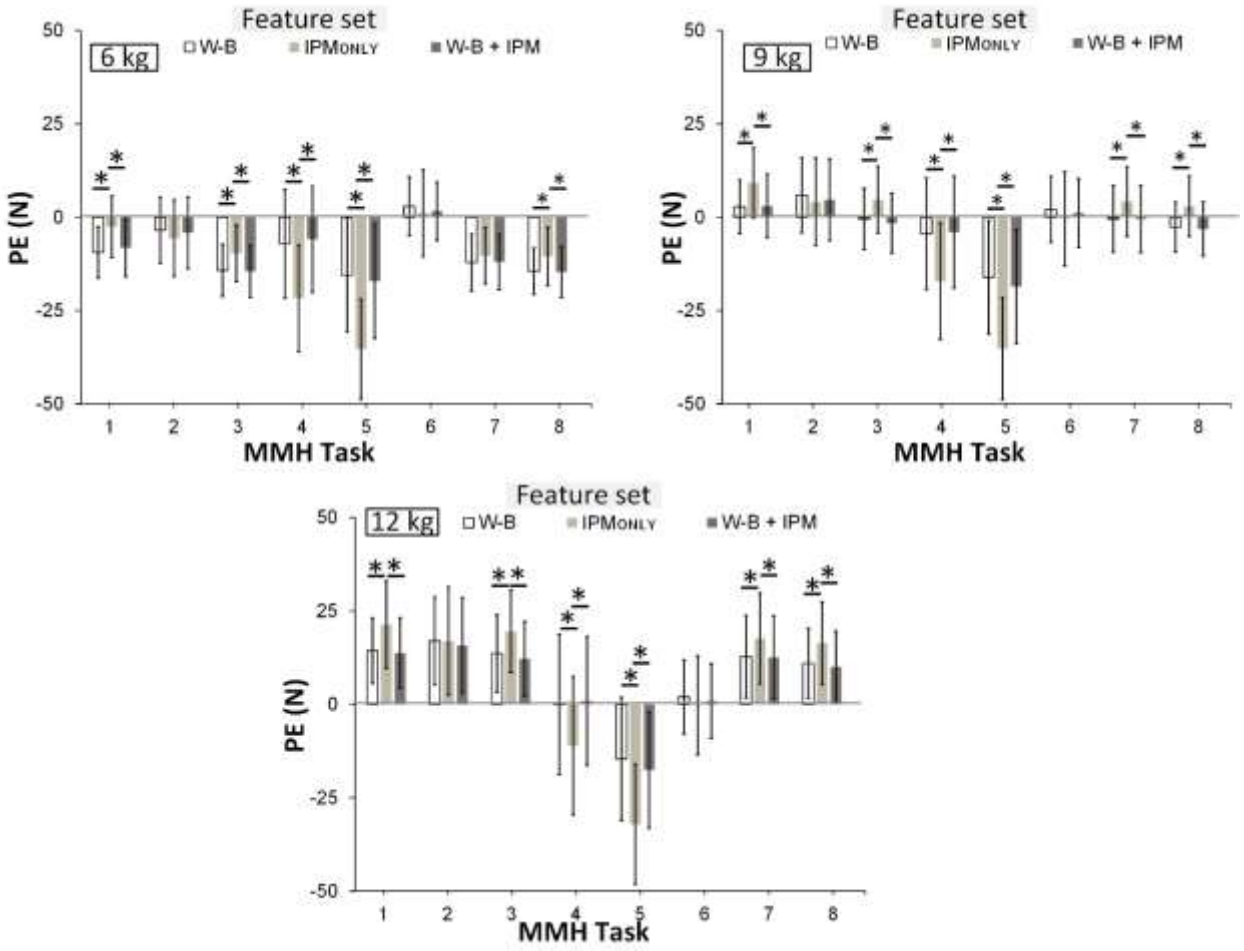

Figure 3.23. Peak Error (PE) in predicted forces for the left hand in the vertical (Z) direction. Results are presented for different levels of *MMH task* and *Feature set*, and are shown separately for the 6, 9, and 12 kg mass conditions.

4. Estimating Dynamic Spine Loads during Diverse Manual Material Handling Tasks using Data from Markerless Motion Capture and In-sole Pressure Measures.

4.1 Abstract

Physical workload is a risk factor for developing work-related musculoskeletal disorders. Biomechanical models are a valuable tool to quantify the impact of specific physical workloads, enabling accurate measurement of individual and job-related differences. However, biomechanical models are seldom used in the field due in part to the difficulty in collecting the requisite kinematics and kinetics data. We evaluated an approach for estimating spine loads on data obtained from markerless motion capture (MMC) and in-sole pressure measurement systems (IPM). We compared estimates from a musculoskeletal model that was driven by several combinations of inputs, including kinematics from MMC, kinematics from a whole-body inertial measurement unit (IMU) system, and kinetics from direct measures of hand loads. Predicted peak lumbosacral forces differed depending on the input data, although the magnitude of these differences was typically fairly small. Such effects, though, were also dependent on the specific MMH task and the load mass handled, and they differed between males and females. In conclusion, our findings support the use of MMC and kinetics from predicted hand forces as input for musculoskeletal model in estimating biomechanical demands.

4.2 Introduction

Physical workload has been identified as a key risk factor contributing to work-related musculoskeletal disorders (WMSDs; Burdorf & Sorock, 1997; Hoogendoorn et al., 1999; Radwin et al., 2001). Physical workloads are often quantified in three main dimensions – level (load), frequency, and duration in different body regions (van der Beek & Frings-Dresen, 1998; Winkel & Mathiassen, 1994). Several methods exist to quantify physical workload, and these methods can be categorized broadly as self-reports, observation techniques, and direct measurements (David, 2005; Jansen & Burdorf, 2003; Spielholz et al., 2001; van der Beek & Frings-Dresen, 1998). However, many of these methods, even direct measurements, often provide relatively crude exposure measures, primarily due for a failure to consider all three of the main dimensions of physical exposure (Burdorf, 1992; Winkel & Westgaard, 1992). This lack of information makes it challenging to obtain comprehensive risk assessments. Such methods also lack completeness and potential representativeness because, in part, of the dynamic nature of many work environments (Burdorf & van der Beek, 1999). Therefore, there is a need for better measures of physical workload that are more complete and representative of exposure to physical workload.

Biomechanical models are a valuable tool to quantify the impact of specific physical workloads more precisely, enabling accurate measurement of individual and job-related differences (Garg et al., 1982). Biomechanical models yield physical workloads in terms of joint force or torque (biomechanical demands) by combining data from external exposure (kinematics and kinetics) and anthropometrical details. Biomechanical models have been used to evaluate physical workload in many different contexts. For example, such models have been used to analyze biomechanical demands during manual material handling jobs for over four decades

(Chaffin & Baker, 1970; Garg & Chaffin, 1975; Neumann et al., 2001; Seo et al., 2016). However, biomechanical models are rarely used among practicing ergonomists. This lack of use is partly due to the difficulty in collecting the necessary kinematics and kinetics data *in situ*. As an example, it is challenging to obtain precise measurement of exerted forces (either hand forces or ground reaction forces) in the workplace for input into biomechanical models.

Recent advancements in technologies, particularly in sensors (e.g., inertial measurement units [IMUs], in-sole pressure measurement [IPM] systems, markerless motion capture [MMC] systems) and machine learning algorithms, offer the possibility of monitoring body kinematics and kinetics in diverse work environments with minimal worker interference. Some studies have demonstrated this feasibility by using sensor data and machine learning models to perform biomechanical analysis and quantify physical workload. As an example, kinematic features generated from MMC were used to estimate L5/S1 joint moments during various lifting tasks (Wang et al., 2021). In other investigations, physical workload has been quantified via biomechanical models using data from whole-body IMUs and IPM systems during bricklaying activity (Yu et al., 2019), and using data from MMC and IPM system during simulated lifting tasks (Wang et al., 2023).

However, these and similar approaches (Liu, 2019; Mehrizi et al., 2017) still have an important limitation, specifically the use of static biomechanical models to quantify physical workloads. Applying static biomechanical models to dynamic activity such as lifting assumes that the effects of linear and angular accelerations are negligible, yet this assumption may underestimate the actual workload on various body segments and joints (Wagner et al., 2007). In fact, the compressive force and peak moment at the low back estimated by a dynamic model during lifting can be two to three times higher compared to what is obtained with a static model

(Bush-Joseph et al., 1988; Garg et al., 1982). Second, the relative benefits of using different input data to drive musculoskeletal models remain unclear. Specifically, it is unclear whether data from MMC can be used as input to accurately estimate biomechanical demands.

The aim of this study was to evaluate the use of MMC-based inputs for assessing biomechanical demands during MMH tasks, using predicted spine loads obtained from a full-body, dynamic musculoskeletal model. The model was driven using kinematics and kinetics from both gold standard inputs and estimates derived from a MMC system (Chapter 3). We evaluated the effects of these different inputs based on estimates biomechanical loading at the lumbosacral level, since mechanical loading and strain at this level are associated with spine injury risk (Coenen et al., 2014; Norman et al., 1998). For example, peak and cumulative loading of the low back has been identified as a risk factor of WMSD (Da Costa & Vieira, 2010; Kumar, 2001). We expected that driving a full-body musculoskeletal model using different inputs would yield substantially different estimates of spine loads. We also anticipated that the magnitudes of these effects would depend substantially on the specific MMH task, the specific task conditions, and differ between males and females. Completing this work was intended demonstrate whether data from MMC and estimated hand forces can be used to perform accurate biomechanical analysis of internal reaction forces.

4.3 Methods

Participants

A convenience sample of 18 participants (sex-balanced), recruited from the university and local community, completed the study. Respective means (SD) of age, body mass, and stature were 26 (4.2) years, 78.9 (16.9) kg, and 175.9 (7.6) cm for males; and 26 (4.4) years, 66.2 (8.9) kg, and 169.2 (7.0) cm for females. All participants reported being physically active (i.e.,

exercising at least twice a week) and with no musculoskeletal disorders within the past year. The research reported herein complied with the tenets of the Declaration of Helsinki, and the study protocol was approved by the Institutional Review Board at Virginia Tech. Informed consent was obtained from all participants prior to any data collection.

Experimental Design and Procedures

Participants completed scenarios of simulated MMH tasks involving combinations of two MMH task conditions. Specific MMH tasks simulated were symmetric box lifting from floor (Sym_Floor) and knee (Sym_Knee) heights; and asymmetric box lifting the hip height (Asym_Hip). We focused our analysis on these three tasks given that processing all of the tasks simulated in Chapter 2 was excessively time-consuming. All tasks were performed using three levels of box *Mass*, and two levels of *Hand Configuration*. More detailed descriptions are provided in Chapter 2. The presentation of *Hand Configuration* was counterbalanced using 3 × 3 balanced Latin Squares. Within a given *Hand Configuration*, the presentation order of *Box Mass* was counterbalanced using 3 × 3 balanced Latin Squares.

Instrumentation

Whole-body kinematics and hand forces data were available as described in Chapters 2 and 3. In addition, whole-body kinematics were monitored using a wearable inertial motion capture system (Noraxon Ultium, Noraxon, Scottsdale, AZ, USA). The Noraxon Ultium consists of 16 IMUs, positioned on the head, upper thoracic, lower thoracic, pelvic, and bilaterally on the upper and lower arms, hands, thighs, shanks, and feet. Data capturing was done using myoResearch software, at 200 Hz. To minimize temporal offsets between data obtained from different systems, the IMU system was synchronized with the other two data sources (hand forces and insole pressure measurement) using an analog output signal generated from LabView.

Whole-body kinematics, exported in Biovision Hierarchy (BVH) format from iPi Motion Capture Studio and Noraxon myoResearch, were low-pass filtered (6 Hz cutoff; 4th order Butterworth; bidirectional) to remove sensor noise and other artifacts (Winter, 2009). BVH provides skeleton hierarchy information, motion data, and rotational joint data. This is a popular motion data format and has wide applications in modeling and animation (Dai et al., 2010). Of note, kinematics data from Noraxon were downsampled to 30 Hz to match the sampling rate of the MMC system.

Biomechanical Modeling Approach

A biomechanical model was developed in the AnyBody^{TM} Modeling System (AMS; Version 7.3, AnyBody Technology, Aalborg, Denmark). AMS was preferred for this study because it is easy to maintain and develop, since it uses a text-based input method, has a rich biomechanical model repository, and can simulate musculoskeletal dynamics (Damsgaard et al., 2006).

A full-body musculoskeletal model was employed from the AnyBody Model Repository (Lund et al., 2020) to estimate reactive lumbosacral shear and compression forces. Participant anthropometry (i.e., height and body mass) was used to scale body segment masses and lengths in the AMS. The full-body model was created by combining five body models – lumbar spine, thoracic spine, cervical spine, shoulder arm, and leg models. In the model, trunk muscles included the rectus abdominus, external and internal obliques, erector spinae, multifidus, quadratus lumborum, psoas major, and latissimus dorsi. In brief, the model consists of 72 segments, 432 degrees-of-freedom, and 339 reaction forces. Internal reaction forces were computed using inverse dynamics analysis (top-down approach), and by using the min/max

approach to solve the problem of kinetic redundancy and estimate muscle activations (Damsgaard et al., 2006).

Three sets of input to this musculoskeletal model were created. Input #1 was the "gold standard" and consisted of IMU kinematics and actual hand forces. Input #2 consisted of MMC kinematics and actual hand forces (MMC + Actual Hand Force). Input #3 was formed from MMC kinematics and hand forces estimated as described in Chapter 3 (MMC + Predicted Hand Force). For the latter, the model developed by combining kinematics and kinetics feature sets from MMC and an in-sole pressure system (W-B + IPM), respectively, and a bi-directional gated recurrent neural network (BGRU), was used to estimate hand forces. We used the W-B + IPM feature set because it often yielded smaller peak errors (up to 43%) for bilateral hand force predictions compared to using only the kinematics feature set. Also, we employed the BGRU model over the Bi-LSTM because the former resulted in smaller peak errors (2.2 vs. 2.5 N) when predicting external hand forces in the vertical direction.

For all three Inputs, the BVH file exported from either iPi Motion capture studio or Noraxon myoResearch was imported into AMS and fitted to the 72-segment full-body musculoskeletal model. Then, a linked-segment model with 432 degrees of freedom was reconstructed. An established approach that uses virtual markers to align the linked-segment model with the musculoskeletal model was employed. This approach minimizes linear distances between virtual markers and the musculoskeletal model (Karatsidis et al., 2018; Skals et al., 2017). Moreover, the tri-axial hand forces time series served as external hand forces on the musculoskeletal model. For each Input, the external hand force was attached bilaterally to the wrist node of the full-body model.

Outcome Measures and Statistical Analyses

Three-dimensional reaction forces at the lumbosacral (L5-S1) discs were extracted, including axial compression (F_{COMP}), mediolateral shear (F_{ML}), and anteroposterior shear (F_{AP}) forces. All F_{ML} data were converted to absolute values, assuming that there were not important differences between left- and right-directed values, similar to the approach of Madinei and Nussbaum (2023). Peak (95^{th}-ile) and cumulative forces of each reaction force were extracted and served as outcome measures. Outcome measures were obtained separately for each MMH task, mass, and hand configuration.

Separate four-way, repeated measures analyses of variance (ANOVAs) were used to assess the effects of *MMH task, Mass, Hand Configuration,* and *Input* on model-estimate peak and cumulative spine forces. For each ANOVA model, biological sex (*Sex*) was included as blocking effect. Significant interaction effects were explored using simple-effects testing, and *post hoc* paired comparisons were completed using the Tukey's HSD procedure. All statistical analyses were performed using JMP Pro (SAS, Cary, NC) using the restricted maximum likelihood (REML) method. Parametric model assumptions were verified, and statistical significance was determined when $p < 0.05$. Summary data are reported as least-square means (with standard deviations as intervals) from statistical model fits. Given the study aims, the subsequent presentation of results emphasizes only the main and interaction effects of *Input*.

4.4 Results

ANOVA results are summarized in Table 4.1 Appendix A. There were main or interaction effects of *Input* for both outcome measures (peak and cumulative forces). Detailed results are provided below.

Axial compression force

Peak F_{COMP} was significantly affected by *Input* and the *MMH task* × *Input* interaction. Across the MMH tasks, peak F_{COMP} was up to 15% smaller using MMC + Predicted Hand Force vs. using the Ground Truth input (Figure 4.1). All simple effects were significant, with differences between *MMH tasks* significant for all *Inputs* ($p < 0.0004$) and differences between *Inputs* were significant for all *MMH tasks* ($p < 0.0001$). Notably, there were no main or interaction effects of *Input* on cumulative F_{COMP}. However, cumulative F_{COMP} was slightly smaller (up to 2.3%) using MMC + Predicted Hand Force compared to the other two inputs (Figure 4.2).

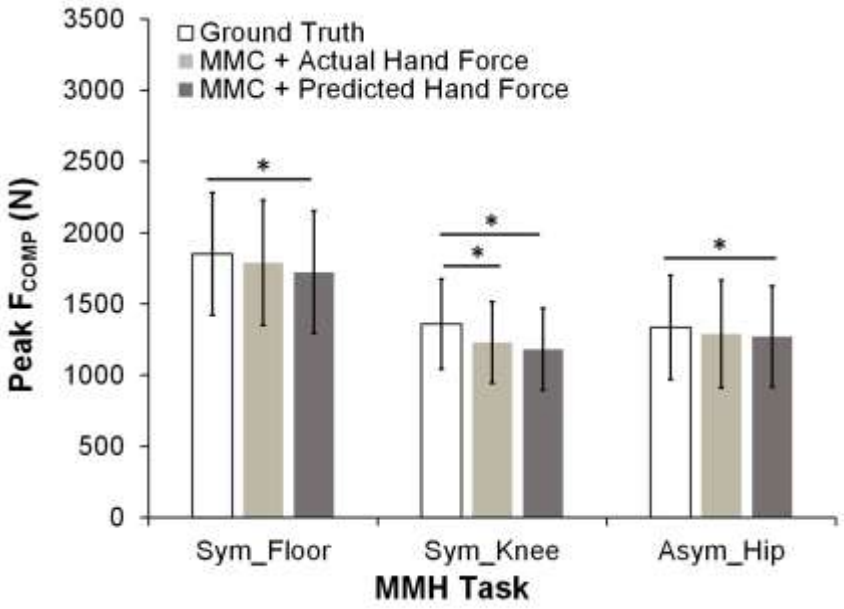

Figure 4.1. *MMH task* × *Input* interaction effect on peak F_{COMP}. Note that * (here and below) indicates a significant difference between paired means connected by the corresponding bar. Error bars (here and below) indicate standard deviations.

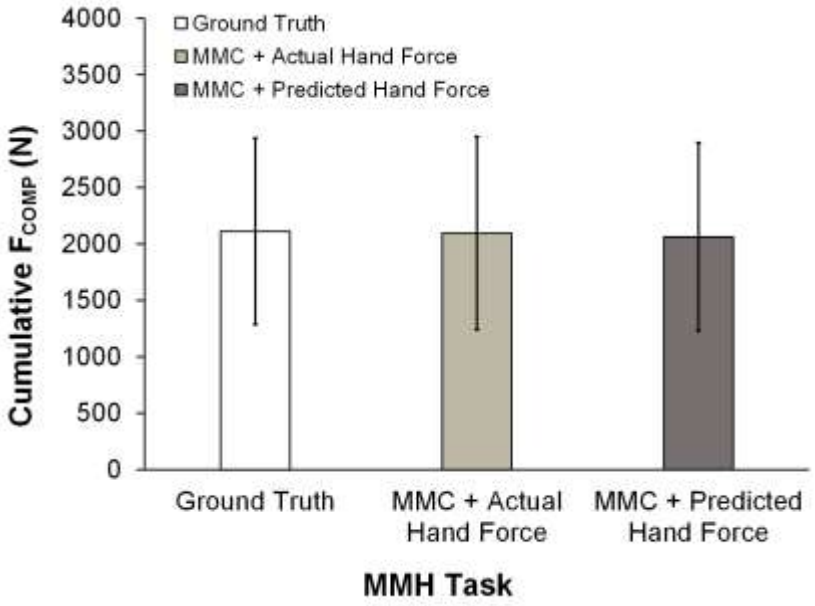

Figure 4.2. *Input* main effect on cumulative F_{COMP}.

Mediolateral shear force

The *Input* main effect and *MMH task* × *Input*, and *Sex* × *Input* interaction effects were significant for F_{ML} peak force (Figure 4.3). Using MMC + Actual Hand Force and MMC + Predicted Hand Force inputs led to significantly larger peak F_{ML} by 23 – 29% in Sym_Floor condition compared to using the Ground Truth input. However, using MMC + Actual Hand Force and MMC + Predicted Hand Force inputs yielded smaller peak F_{ML} of up to 24% in Asym_Hip condition compared to using Ground Truth input. All simple effects were significant, with *MMH task* significant for all *Input* ($p < 0.0001$), and *Input* significant for all *MMH task* ($p < 0.0001$). Among males, peak F_{ML} was significantly smaller, by up to ~15%, when using MMC + Predicted Hand Force input. However, among females, peak F_{ML} was ~14 – 18% larger when using either MMC + Actual Hand Force input or MMC + Predicted Hand Force inputs. Only the simple effects of *Input* for both *Sexes* were significant ($p < 0.0003$).

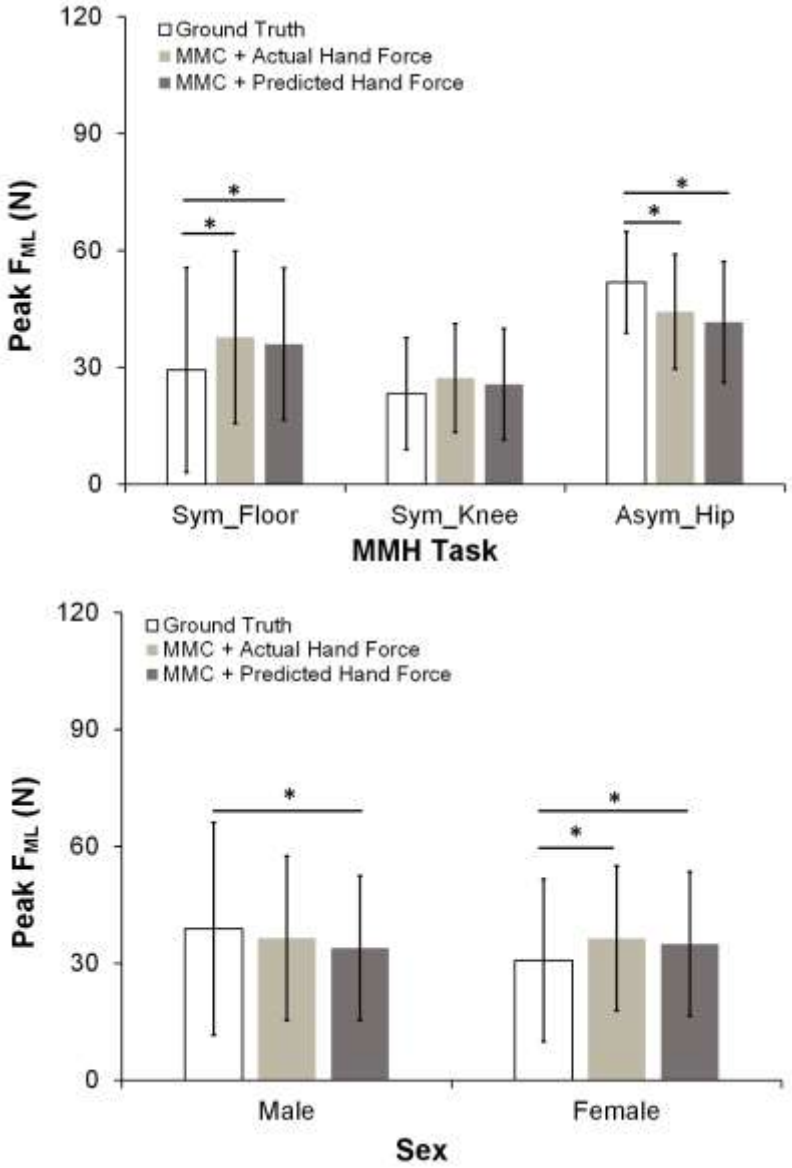

Figure 4.3. *MMH task* × *Input*, *Mass* × *Input*, and *Sex* × *Input* interaction effects on peak F_{ML}.

There was a significant main effect of *Input*, and interaction effects of *MMH task* × *Input* and *Sex* × *Input* on cumulative F_{ML} (Figure 4.4). In the Asym_Hip condition, using MMC + Actual Hand Force and MMC + Predicted Hand Force inputs led to significantly smaller cumulative F_{ML} by up to 21% compared to the Ground Truth input. All simple effects were significant except for the effect of *Input* in the Sym_Floor and Sym_Knee conditions ($p = 0.24$). Cumulative F_{ML} was ~11 – 19% smaller among males, depending on the specific input. Simple

effects were significant for the effect of *Input* among males ($p < 0.0001$) and the effect of *Sex* when using Gound Truth as an input ($p = 0.019$).

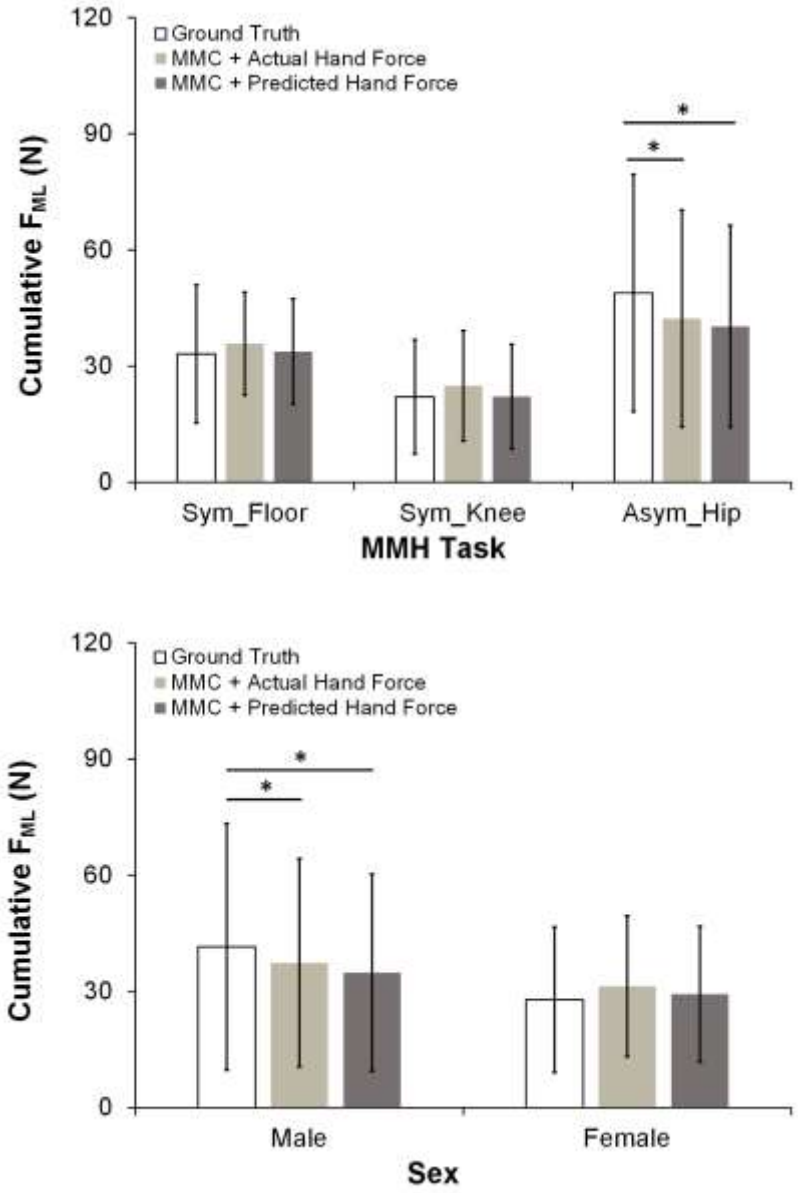

Figure 4.4. *MMH task* × *Input* and *Sex* × *Input* interaction effects on cumulative F_{ML}.

Anteroposterior shear force

There was a significant main effect of *Input*, and interaction effects of *MMH task* × *Input*, *Mass* × *Input*, and *Sex* × *Input* on peak F_{AP} (Figure 4.5). In Sym_Floor condition, using MMC +

Predicted Hand Force input led to significantly reduced peak F_{AP} by up to ~9% compared to using Ground Truth input. Also, in the Sym_Knee condition, using MMC + Predicted Hand Force input yielded significantly smaller peak F_{AP} up to ~9% compared to using Ground Truth input. All simple effects were significant except for the effect of *Input* in Asym_Hip condition ($p = 0.41$). In the 6 and 9 kg condition, using MMC + Predicted Hand Force input led to smaller peak F_{AP} up to ~11% than using the Ground Truth input. Simple effects of *Mass* for *Input* were significant ($p < 0.0001$). Peak F_{AP} was ~5 – 13% smaller among males and females when using MMC + Predicted Hand Force input compared to using Ground Truth input. Simple effects were significant except for the effect of *Input* among males ($p = 0.15$) and the effect of *Sex* when using ground truth input ($p = 0.19$).

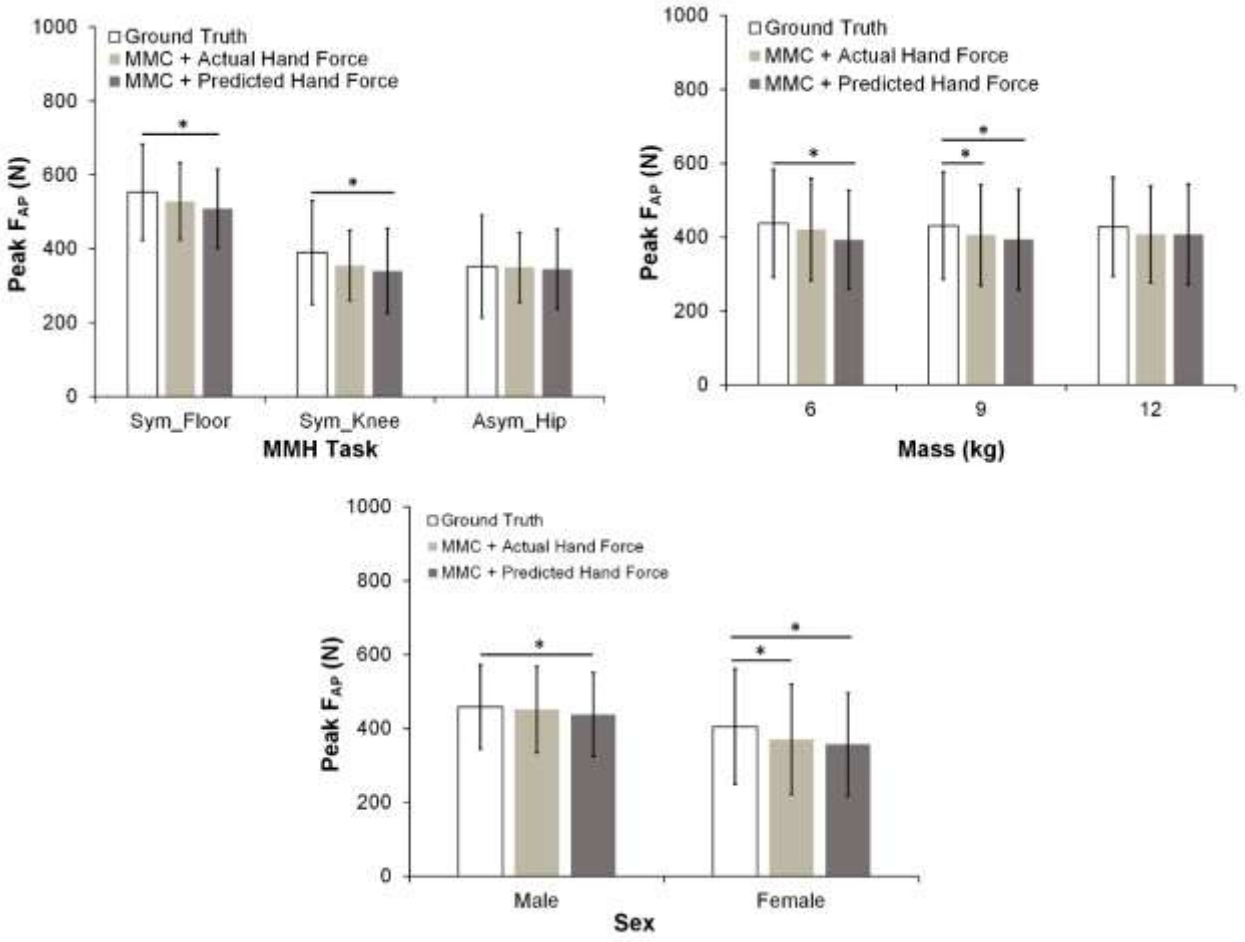

Figure 4.5. *MMH task* × *Input* and *Sex* × *Input* interaction effects on peak F_{AP}.

The *Input* main effect and *Mass* × *Input*, and *Sex* × *Input* interaction effects were significant for cumulative F_{AP} (Figure 4.6). Also, the interaction effect of *Mass* × *Input* on cumulative F_{AP} approached significant (p = 0.054). Across mass conditions, there were no significant differences between the three inputs. Yet, in the 6 kg condition, using MMC + Predicted Hand Force input led to smaller cumulative F_{AP} by 5%, using it led to larger cumulative F_{AP} by 3% in the 12 kg condition. Among males, using MMC + Actual Hand Force yielded significantly larger cumulative F_{AP} of up to 8%, while using MMC + Predicted Hand Force led to significantly smaller cumulative F_{AP} of up to 7% among females compared to using ground truth input. Simple effects were significant, except for the effect of *Sex* using ground

truth as an input (p = 0.15). Across MMH tasks, there were no significant differences between the three inputs. However, in Asym_Hip condition, using any of the inputs led to larger cumulative F_{AP} by 5 – 7%. All simple effects were significant except for the effect of *Input* in Sym_Floor (p = 0.16) and Sym_Knee (p = 0.97) conditions.

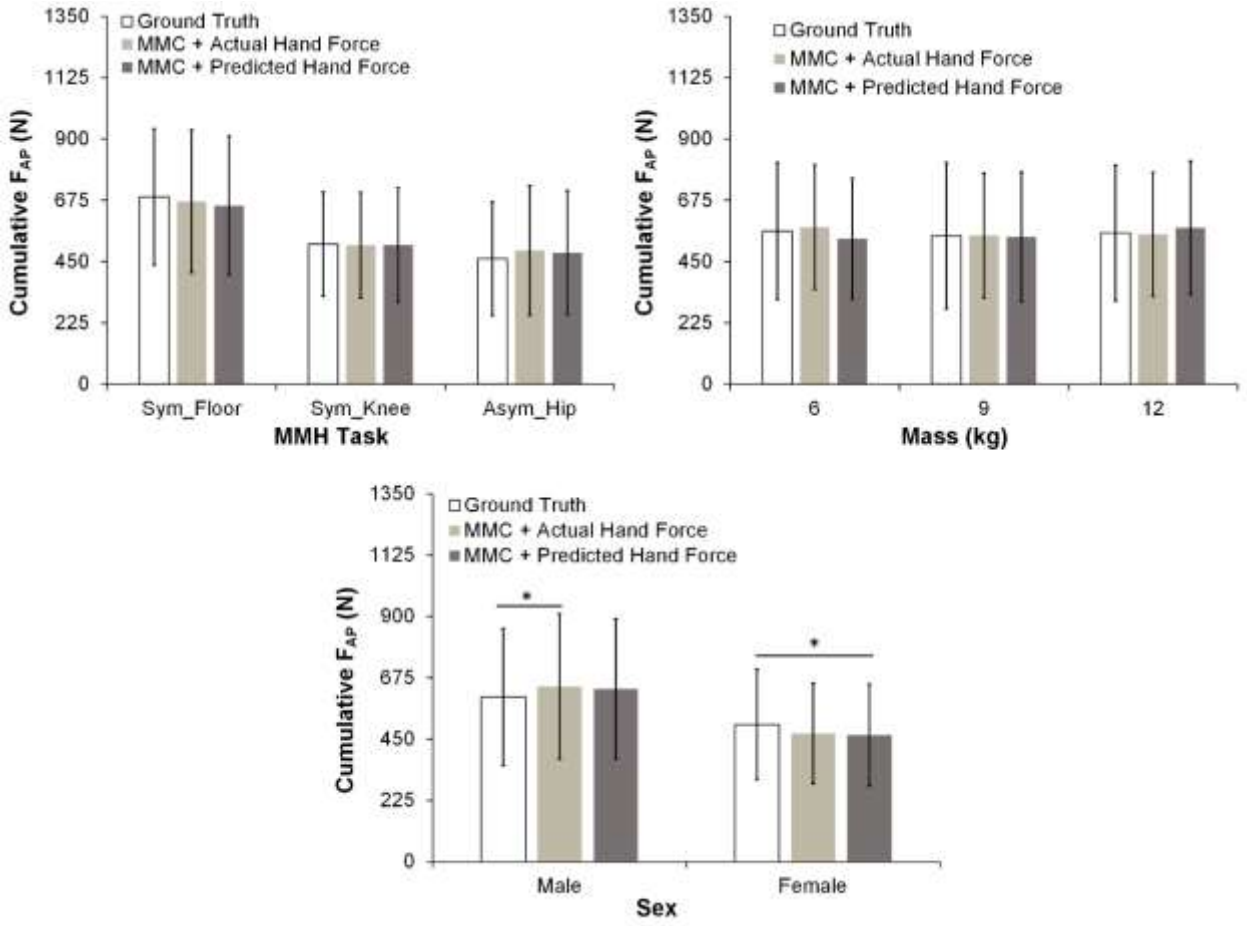

Figure 4.6. *MMH task* × *Input* and *Sex* × *Input* interaction effects on cumulative F_{AP}.

4.5 Discussion

We found that that spine loads, estimated using a musculoskeletal model, differed when either MMC + Actual Hand Force or MMC + Predicted Hand Force were used as model inputs.

Also, we found that the magnitudes of these effects varied depending on the specific MMH task, mass condition, and hand configuration, and differed between males and females.

Compared to using Ground Truth input, using MMC + Predicted Hand Force vs. Ground Truth input clearly led to an underestimation of peak F_{COMP}, up to 15% across the MMH tasks (Figure 4.1), and an overestimation of peak F_{ML}, by 23% only in the Sym_Floor condition (Figure 4.3). Also, MMC + Predicted Hand Force input resulted in a smaller F_{AP}, up to 9% in the Sym_Floor and Sym_Knee conditions, and up to 11% in the 6 and 9 kg conditions (Figure 4.5). The differences in lumbosacral reaction forces estimated with the musculoskeletal model when using Ground Truth vs. MMC + Predicted Hand Force as input likely resulted from errors in predicting hand forces. Recall from Chapter 3 that we suggested using a BGRU model and W-B+IPM feature set as the base model for predicting external hand forces. However, using the BGRU model and W-B + IPM feature set led to root-mean-square-error (RMSE) between 4.6 – 13.3 N when predicting bilateral hand forces.

We also found differences in the estimated lumbosacral forces between males and females. Particularly among males, using MMC + Predicted Hand Force inputs led to smaller peak F_{ML} up to ~15% compared to using the Ground Truth input (Figure 4.3). Interestingly, using MMC + Actual Hand Force and MMC + Predicted Hand Force inputs led to larger peak F_{ML} (up to 18%) among females compared to using Ground Truth inputs.

To be candid, the reason why the effect of different inputs differ between males and females are unclear. One reason for such divergence could be explained by sex differences in kinetics and whole-body kinematics which affects estimated spine loads. In Chapter 2, we found that RMSE was substantially smaller among females when predicting hand forces in the horizontal direction (4.73 N vs. 5.53 N) and vertical direction (15.2 vs. 16.7 N). Consequently,

we anticipated a larger difference in estimated lumbosacral forces when using MMC + Predicted Hand Force vs. Ground Truth as inputs to the musculoskeletal model. This expectation is based on the observed larger RMSE when predicting hand forces in both the horizontal and vertical directions among males. Recall we found that males and females typically use different kinematics to perform symmetric lifting in Chapter 1. Biomechanical demands estimated from an electromyography-based musculoskeletal model indicated that lumbar forces depended on postures (Gallagher et al., 1994), emphasizing the influence of kinematics on the outcome of musculoskeletal models.

Beyond variations in kinetics and whole-body kinematics, sex-based anatomical differences (e.g., distinct moment arms of specific muscles) could account for the disparities between the observed results and our initial expectations. A musculoskeletal model estimated larger spinal compression forces (up to 4.7%) among females during various symmetrical lifting tasks (Ghezelbash et al., 2016). Yet, when employing a sex-specific musculoskeletal model that considered segmental mass distributions and muscle physiological cross-sectional areas, males demonstrated larger spine compression forces (up to 25.3%) than females during diverse lifting tasks (Firouzabadi et al., 2021). Notably, these contrasting findings underscore the multifaceted nature of biomechanical factors influencing spine loading across biological sex. Females tend to have smaller muscle physiological cross-sectional areas (Anderson et al., 2012; Marras et al., 2001), which has been associated with lower muscular strength, and consequently lower spine tolerance limits. For example, female lifting strength was estimated to range between 48 and 70% of males (Kumar & Garand, 1992; Marras et al., 2002). To create an effective and efficient ergonomic tool for the workplace, it is essential to develop a sex-specific musculoskeletal model that considers crucial anatomical distinctions between males and females.

Can we use MMC-based Kinematics and Predicted Hand Forces to Estimate LS/S1 Forces?

An optimization-based modeling approach was used to estimate lumbosacral spine forces using three types of data inputs: 1) kinematics from IMUs and kinetics from actual hand loads; 2) kinematics from MMC and kinetics from actual hand loads; and 3) kinematics from MMC and kinetics from predicted hand loads. Our results indicate that using MMC + Predicted Hand Forces as input to musculoskeletal model yields biomechanical demand estimates closer to those obtained using Ground Truth as input. For example, using MMC + Predicted Hand Force as inputs to drive the musculoskeletal model led to smaller peak F_{COMP} than using Ground Truth as input, though, the magnitude of this difference was rather small (~2.69 – 4.30% of maximum peak F_{COMP}). Also, the cumulative F_{COMP} estimated using MMC + Predicted Hand Force was closer to those estimated using Ground Truth as input (0.36 – 0.84% of maximum cumulative F_{COMP}).

Limitations

Some limitations of this study need to be noted. First, the musculoskeletal model relies on several assumptions, including using the min/max approach for solving the muscle recruitment problem when estimating lumbosacral forces. Use of the min/max approach in the AMS could affect estimated lumbosacral forces by potentially under-utilizing muscles with small moment arms (Damsgaard et al., 2006). Hence, future work should explore comparisons between the output from the AMS and other musculoskeletal models that employ *polynomial criteria* or *soft saturation criterion* to solve muscle recruitment problems. Although we used IMU-based kinematics as our gold standard, reconstructing motion data from IMUs may introduce errors from soft-tissue artifacts. It is unclear whether the adjustments in our setup to accommodate individual anthropometry contributed to observed differences between males and females. Future

work should investigate why inputs from MMC led to different performance between males and females

4.6 Conclusion

Physical workload is a risk factor for both acute trauma injuries and cumulative musculoskeletal disorders. Biomechanical models could quantify the impact of physical workloads, enabling accurate measurement of individual and job-related differences. We evaluated using MMC for assessing biomechanical demands by comparing outputs from a musculoskeletal model driven by kinematics from IMU, kinetics from actual hand loads, and kinematics from MMC and kinetics from actual and predicted hand loads. Using kinematics from MMC and kinetics from actual and predicted hand loads as inputs led to biomechanical demands closer to those estimated using the ground truth. However, in most of the tested conditions, magnitudes of such differences were dependent on MMH tasks, mass conditions, and differs between males and females. Overall, our results suggest that MMC + Predicted Hand Force can be used as input to drive musculoskeletal models to estimate biomechanical demands.

4.7 References

Anderson, D. E., D'Agostino, J. M., Bruno, A. G., Manoharan, R. K., & Bouxsein, M. L. (2012). Regressions for estimating muscle parameters in the thoracic and lumbar trunk for use in musculoskeletal modeling. *Journal of Biomechanics*, *45*(1), 66-75.

Burdorf, A. (1992). Exposure assessment of risk factors for disorders of the back in occupational epidemiology. *Scandinavian Journal of Work, Environment & Health*, 1-9.

Burdorf, A., & Sorock, G. (1997). Positive and negative evidence of risk factors for back disorders. *Scandinavian Journal of Work, Environment & Health*, 243-256.

Burdorf, A., & van der Beek, A. (1999). Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. *Scandinavian Journal of Work, Environment & Health*, *25*, 25-30.

Bush-Joseph, C., Schipplein, O., Andersson, G., & Andriacchi, T. (1988). Influence of dynamic factors on the lumbar spine moment in lifting. *Ergonomics*, *31*(2), 211-216.

Chaffin, D. B., & Baker, W. H. (1970). A biomechanical model for analysis of symmetric sagittal plane lifting. *AiiE Transactions*, *2*(1), 16-27.

Coenen, P., Gouttebarge, V., van der Burght, A. S., van Dieën, J. H., Frings-Dresen, M. H., van der Beek, A. J., & Burdorf, A. (2014). The effect of lifting during work on low back pain: a health impact assessment based on a meta-analysis. *Occupational and Environmental Medicine*, *71*(12), 871-877.

Da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: a systematic review of recent longitudinal studies. *American Journal of Industrial Medicine*, *53*(3), 285-323.

Dai, H., Cai, B., Song, J., & Zhang, D. (2010). Skeletal animation based on BVH motion data. 2010 2nd International Conference on Information Engineering and Computer Science,

Damsgaard, M., Rasmussen, J., Christensen, S. T., Surma, E., & De Zee, M. (2006). Analysis of musculoskeletal systems in the AnyBody Modeling System. *Simulation Modelling Practice and Theory*, *14*(8), 1100-1111.

David, G. C. (2005). Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine*, *55*(3), 190-199.

Firouzabadi, A., Arjmand, N., Pan, F., Zander, T., & Schmidt, H. (2021). Sex-Dependent Estimation of Spinal Loads During Static Manual Material Handling Activities—Combined in vivo and in silico Analyses. *Frontiers in Bioengineering and Biotechnology*, *9*, 750862.

Gallagher, S., Hamrick, C. A., Love, A. C., & Marras, W. S. (1994). Dynamic biomechanical modelling of symmetric and asymmetric lifting tasks in restricted postures. *Ergonomics*, *37*(8), 1289-1310.

Garg, A., & Chaffin, D. B. (1975). A biomechanical computerized simulation of human strength. *AiiE Transactions*, *7*(1), 01-15.

Garg, A., Chaffin, D. B., & Freivalds, A. (1982). Biomechanical stresses from manual load lifting: a static vs dynamic evaluation. *IIE transactions*, *14*(4), 272-281.

Ghezelbash, F., Shirazi-Adl, A., Arjmand, N., El-Ouaaid, Z., Plamondon, A., & Meakin, J. (2016). Effects of sex, age, body height and body weight on spinal loads: Sensitivity analyses in a subject-specific trunk musculoskeletal model. *Journal of Biomechanics*, *49*(14), 3492-3501.

Hoogendoorn, W. E., Van Poppel, M. N., Bongers, P. M., Koes, B. W., & Bouter, L. M. (1999). Physical load during work and leisure time as risk factors for back pain. *Scandinavian Journal of Work, Environment & Health*, 387-403.

Jansen, J. P., & Burdorf, A. (2003). Effects of measurement strategy and statistical analysis on dose-response relations between physical workload and low back pain. *Occupational and Environmental Medicine*, *60*(12), 942-947.

Karatsidis, A., Jung, M., Schepers, H. M., Bellusci, G., de Zee, M., Veltink, P. H., & Andersen, M. S. (2018). Predicting kinetics using musculoskeletal modeling and inertial motion capture. *arXiv preprint arXiv:1801.01668*.

Kumar, S. (2001). Theories of musculoskeletal injury causation. *Ergonomics*, *44*(1), 17-47.

Kumar, S., & Garand, D. (1992). Static and dynamic lifting strength at different reach distances in symmetrica] and asymmetrical planes. *Ergonomics*, *35*(7-8), 861-880.

Liu, M. (2019). *Video-based human motion capture and force estimation for comprehensive on-site ergonomic risk assessment*

Lund, M., Tørholm, S., Jensen, B., Galibarov, P., Dzialo, C., Iversen, K., Sarivan, M., Marra, M., & Simonsen, S. (2020). *The AnyBody Managed Model Repository (AMMR)*. In *Zenodo, Version*

Madinei, S., & Nussbaum, M. A. (2023). Estimating lumbar spine loading when using back-support exoskeletons in lifting tasks. *Journal of Biomechanics*, *147*, 111439.

Marras, W., Jorgensen, M., Granata, K., & Wiand, B. (2001). Female and male trunk geometry: size and prediction of the spine loading trunk muscles derived from MRI. *Clinical Biomechanics*, *16*(1), 38-46.

Marras, W. S., Davis, K. G., & Jorgensen, M. (2002). Spine loading as a function of gender. *Spine*, *27*(22), 2514-2520.

Mehrizi, R., Xu, X., Zhang, S., Pavlovic, V., Metaxas, D., & Li, K. (2017). Using a marker-less method for estimating L5/S1 moments during symmetrical lifting. *Applied Ergonomics*, *65*, 541-550.

Neumann, W., Wells, R., Norman, R., Frank, J., Shannon, H., Kerr, M. S., & Group, O. W. (2001). A posture and load sampling approach to determining low-back pain risk in occupational settings. *International Journal of Industrial Ergonomics*, *27*(2), 65-77.

Norman, R., Wells, R., Neumann, P., Frank, J., Shannon, H., & Kerr, M. (1998). A comparison of peak vs cumulative physical work exposure risk factors for the reporting of low back pain in the automotive industry. *Clinical Biomechanics*, *13*(8), 561-573.

Radwin, R. G., Marras, W. S., & Lavender, S. A. (2001). Biomechanical aspects of work-related musculoskeletal disorders. *Theoretical Issues in Ergonomics Science*, *2*(2), 153-217.

Seo, J., Lee, S., & Seo, J. (2016). Simulation-based assessment of workers' muscle fatigue and its impact on construction operations. *Journal of Construction Engineering and Management*, *142*(11), 04016063.

Skals, S., Rasmussen, K. P., Bendtsen, K. M., Yang, J., & Andersen, M. S. (2017). A musculoskeletal model driven by dual Microsoft Kinect Sensor data. *Multibody System Dynamics*, *41*(4), 297-316.

Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., & Kaufman, J. (2001). Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics*, *44*(6), 588-613.

van der Beek, A. J., & Frings-Dresen, M. (1998). Assessment of mechanical exposure in ergonomic epidemiology. *Occupational and Environmental Medicine*, *55*(5), 291-299.

Wagner, D. W., Reed, M. P., & Rasmussen, J. (2007). Assessing the importance of motion dynamics for ergonomic analysis of manual materials handling tasks using the AnyBody Modeling System. *SAE Transactions*, 2092-2101.

Wang, H., Xie, Z., Lu, L., Li, L., & Xu, X. (2021). A computer-vision method to estimate joint angles and L5/S1 moments during lifting tasks through a single camera. *Journal of Biomechanics*, *129*, 110860.

Wang, Z., Wang, W., Chen, J., Zhang, X., & Miao, Z. (2023). Posture Risk Assessment and Workload Estimation for Material Handling by Computer Vision. *International Journal of Intelligent Systems*, *2023*.

Winkel, J., & Mathiassen, S. E. (1994). Assessment of physical work load in epidemiologic studies: concepts, issues and operational considerations. *Ergonomics*, *37*(6), 979-988.

Winkel, J., & Westgaard, R. (1992). Occupational and individual risk factors for shoulder-neck complaints: Part II—The scientific basis (literature review) for the guide. *International Journal of Industrial Ergonomics*, *10*(1-2), 85-104.

Winter, D. A. (2009). *Biomechanics and motor control of human movement*. John Wiley & Sons.

Yu, Y., Li, H., Umer, W., Dong, C., Yang, X., Skitmore, M., & L Wong, A. Y. (2019). Automatic Biomechanical Workload Estimation for Construction Workers by Computer Vision and Smart Insoles.

Appendix A – ANOVA Results and Additional Figures

Table 4.1. Summary of ANOVA results for the effects of *MMH task, Mass, Hand Configuration, Lift Origin, Input, and Sex* on peak and cumulative lumbosacral forces: FCOMP, FML, and FAP. Entries are F values (p values) and significant effects are highlighted in bold font.

	Peak			Cumulative		
Effect	F_{COMP}	F_{ML}	F_{AP}	F_{COMP}	F_{ML}	F_{AP}
Task (T)	**1034.91 (<.0001)**	**330.38 (<.0001)**	**1151.12 (<.0001)**	**332.17 (<.0001)**	**233.17 (<.0001)**	**283.81 (<.0001)**
Mass (M)	2.32 (0.099)	**7.72 (0.001)**	1.62 (0.198)	**3.32 (0.036)**	**5.6 (0.004)**	2.13 (0.119)
Hand Configuration (H)	**37.49 (<.0001)**	1.26 (0.262)	**28.55 (<.0001)**	**22.14 (<.0001)**	1.67 (0.196)	**24.76 (<.0001)**
Input (I)	**50.58 (<.0001)**	**3.02 (0.049)**	**34.53 (<.0001)**	**2 (0.136)**	**3.89 (0.021)**	**0.4 (0.67)**
Sex (S)	2.68 (0.121)	0.28 (0.604)	3.79 (0.069)	3.95 (0.064)	2.82 (0.113)	4.91 (0.042)
T × I	**4.74 (0.001)**	**25.97 (<.0001)**	**7.27 (<.0001)**	0.43 (0.785)	**7.4 (<.0001)**	2.33 (0.054)
M × I	1.45 (0.216)	1.78 (0.131)	**2.38 (0.05)**	1.22 (0.299)	0.76 (0.549)	**2.67 (0.031)**
I × S	0.21 (0.807)	**16.6 (<.0001)**	**7.35 (0.001)**	1.65 (0.193)	**9.85 (<.0001)**	**10.98 (<.0001)**
H × I	0.06 (0.938)	0.01 (0.987)	**0.63 (0.53)**	0.03 (0.968)	0.05 (0.952)	0.34 (0.711)
M × T × I	0.88 (0.536)	0.56 (0.813)	1.31 (0.235)	0.34 (0.951)	0.43 (0.903)	0.58 (0.796)
M × H × I	0.09 (0.987)	0.06 (0.994)	0.29 (0.882)	**0.01 (1)**	0.09 (0.987)	0.08 (0.989)
H × T × I	0.65 (0.625)	0.51 (0.729)	0.42 (0.792)	0.07 (0.991)	0.39 (0.813)	0.08 (0.988)
M × I × S	0.26 (0.903)	0.12 (0.976)	0.71 (0.587)	0.14 (0.966)	0.17 (0.955)	0.33 (0.856)
H × I × S	0.25 (0.777)	0.34 (0.711)	0.14 (0.871)	0.1 (0.907)	0.46 (0.631)	0.12 (0.889)
T × I × S	0.2 (0.938)	**1 (0.406)**	0.87 (0.481)	0.29 (0.886)	1.36 (0.246)	0.51 (0.728)
M × H × T × I	0.12 (0.999)	0.12 (0.998)	**0.07 (1)**	**0.04 (1)**	**0.09 (1)**	**0.08 (1)**
M × H × I × S	0.19 (0.941)	0.18 (0.949)	0.42 (0.795)	0.06 (0.994)	0.09 (0.984)	0.23 (0.922)
M × T × I × S	0.18 (0.994)	0.18 (0.993)	**0.1 (0.999)**	**0.06 (1)**	0.25 (0.981)	**0.07 (1)**
H × T × I × S	0.11 (0.98)	0.17 (0.955)	0.11 (0.981)	0.05 (0.995)	0.25 (0.91)	0.11 (0.981)
M × H × T × I × S	0.21 (0.988)	0.36 (0.941)	0.22 (0.988)	**0.04 (1)**	0.19 (0.992)	**0.08 (1)**

5. Chapter 5. Conclusions

Manual material handling (MMH) tasks have been associated with developing WMSDs. Minimizing the frequency and intensity of handling objects is an ideal solution. Yet, MMH is an integral part of many industry sectors, including manufacturing, construction, warehousing, and distribution. Physical exposure assessment can help identify high-risk tasks, guide the development and assessment of ergonomic interventions, and contribute to understanding exposure-risk relationships. Physical exposure can be assessed using self-assessment, observational methods, and direct measurements. Nonetheless, implementing these methods *in situ* can be challenging due to differences in outcome measures, diagnostic criteria, and difficulty in interpreting and assessing results. Hence, there is a critical need to improve physical exposure assessments to protect workers and save costs.

This dissertation assessed the accuracy of a markerless motion capture system (MMC) to quantify physical exposures during MMH tasks using three studies. Specifically, the first study investigated the performance of an MMC system, together with a machine learning algorithm, for classifying diverse MMH tasks during a simulated complex job (Chapter 2). In the second study, the feasibility of predicting dynamic hand forces using alternative measures, such as kinematics from MMC and/or in-sole pressure systems, coupled with machine learning algorithms (Chapter 3). Finally, in the third study, we systematically evaluate MMC for assessing biomechanical demands by comparing outputs from a full-body musculoskeletal model driven by kinematic and kinetics from gold standard input and estimates derived from the MMC (Chapter 4). Overall, the findings of these studies demonstrated the potential of using MMC to classify several common occupational tasks and to estimate the associated biomechanical demands for a given worker (automatically and with minimal physical contact). Additionally, the methods

developed here can help stakeholders rapidly assess an individual worker's exposure to physical demands during diverse tasks.

Findings from the study on the performance of using data from an MMC system to classify MMH tasks indicated that the use of MMC appears to have a clear potential for physical exposure assessment. Using kinematic data from MMC and machine learning algorithms led to satisfactory performance in classifying MMH tasks and task conditions (e.g., mean precision of 85 – 97%). In most cases, though, our results showed that the classification performance depended on feature sets and MMH tasks. Further, important sex-specific differences were apparent regarding the performance of the machine learning algorithms in classifying the MMH tasks and task conditions.

Dynamic hand forces are required by biomechanical modeling software to perform top-down inverse dynamics analysis and to estimate internal forces. In the second study, we introduced a novel approach to predict dynamic hand forces by combining kinematics from MMC and ground reaction forces from in-sole pressure systems with machine learning algorithms. We tested the feasibility of this approach for quantifying dynamic hand forces during several MMH tasks, and when using two hands vs using one hand. Findings here show that the performance of the machine learning algorithm when predicting hand forces in the horizontal, lateral, and vertical directions was reasonable. However, such performance differs based on MMH tasks, feature sets, and mass conditions. Specifically, tasks involving pushing and pulling had the largest RMSE errors when predicting hand forces in the horizontal and vertical directions. Using feature sets from both MMC and in-sole pressure system led to smaller RMSE, compared to using feature sets only from in-sole pressure system. Also, the performance differed between males and females, with females having smaller RMSE up to 11% overall. Overall, the

proposed approach has the potential to predict dynamic external hand forces, without direct measurement. Our method could be integrated with postural exposure assessment tools available in industry (e.g., Intenseye, 2023; TuMeke, 2023; VelocityEHS(R), 2022) to create a comprehensive risk assessment tool.

Past methods likely underestimate biomechanical exposures because they are not applicable to MMH tasks or do not consider the effects of dynamics on body segments. Thus, the third study systematically evaluated using data obtained from MMC and in-sole pressure measurement systems to estimate biomechanical demands. We compared estimates from a full-body musculoskeletal model that was driven using kinematics from inertial measurements units and kinetics from actual hand forces (ground truth); kinematics from MMC and kinetics from actual hand forces; and kinematics from MMC and kinetics from predicted hand forces. Using kinematics from MMC and kinetics from predicted hand forces often yields biomechanical demands closer to those estimated using the ground truth. Yet, the magnitude of such differences varies depending on the MMH tasks, and mass conditions, and differs between males and females. Overall, our findings support using MMC and kinetics from predicted hand forces as input for a musculoskeletal model in estimating biomechanical demands.

Overall limitations

A few limitations of this work should be noted. First, participants included only young, healthy adults who did not have experience in performing MMH tasks. Therefore, caution should be taken when generalizing the results to older and experienced participants. Second, while we simulated a wide range of MMH tasks, each task only included a few selections of configurations. Third, the task performed here was relatively short, thus, it is unclear if the methods developed here could be generalized to more prolonged tasks. Fourth, the current tasks were simulated in a controlled laboratory. Therefore, the relevance to actual work settings

warrants further investigation. Fifth, each task were analyzed independently. However, continuously analyzing MMH tasks may yield different results than analyzing each task independently due to potential task interdependence in the work process (Harari et al., 2018).

Future Work

The current research sought to facilitate effective and efficient physical exposure assessment in the workplace. Our results have demonstrated the feasibility of using data from MMC and machine learning models to classify several MMH tasks. We also showed a non-intrusive approach to predicting external hand forces using whole-body kinematics from MMC and kinetics from IPM. Our findings also support using MMC and kinetics from hand forces as input to the musculoskeletal model to estimate lumbosacral forces. However, in most cases, our findings depend on the specific MMH tasks, and task conditions, and differ between males and females. For example, in Chapter 2, we found that our model often "confuses" the carrying task as the pushing task; and in Chapter 3, hand force prediction accuracy was worse in pushing and pulling tasks. Previous studies have stated that machine learning model performance often depends on the task type. Future work should explore using other sensors to supplement data from MMC to improve task classification and hand force prediction accuracy. Physical exposure assessment has three components – intensity, duration, and frequency. However, we focused on intensity. Future work should incorporate duration and frequency in addition to intensity to have a more comprehensive physical exposure assessment. As an example, future work should estimate the time duration over which the peaks of actual and predicted hand forces occur and calculate percentage errors by comparing duration estimates of actual hand forces to predicted hand forces. We also found differences in estimating lumbosacral forces between males and females. Quantifying sex bias in models is important to promote the development of an ergonomic tool that prioritizes equity and inclusiveness in the workplace. Thus, future work

should explore the feasibility of developing a sex-specific model to reduce sex bias when quantifying lumbosacral forces. Future work should also investigate using only kinematics data from MMC to estimate biomechanical demands.

Overall, the work here provided new methods to assess physical exposures in the workplace and demonstrate the potential of using non-invasive technologies (MMC and IPM systems) to collect detailed data and quantify physical exposures *in situ.* Specifically, our findings here support using MMC to classify several common occupational tasks and biomechanical analysis. Future research and development will support establishing exposure-response relationships and facilitate specific interventions needed in the workplace.

REFERENCES

Harari, Y., Riemer, R., & Bechar, A. (2018). Factors determining workers' pace while conducting continuous sequential lifting, carrying, and lowering tasks. *Applied Ergonomics*, *67*, 61-70.

Intenseye. (2023). *Ergonomics AI*. Retrieved April from https://www.intenseye.com/products/ergonomics

TuMeke. (2023). *Risk Suite*. Retrieved April from https://www.tumeke.io/product/tumeke-suite

VelocityEHS(R). (2022). *Motion Capture Technology Makes Ergonomics Assessments Simple*. Retrieved April from https://www.ehs.com/2022/12/motion-capture-technology-makes-ergonomics-assessments-simple/