Online Action Recognition for Human Risk Prediction with Anticipated Haptic Alert via Wearables

Cheng Guo[1,2], Lorenzo Rapetti[1], Kourosh Darvish[3], Riccardo Grieco[1], Francesco Draicchio[4], Daniele Pucci[1,2]

***Abstract*— This paper proposes a framework that combines online human state estimation, action recognition and motion prediction to enable early assessment and prevention of worker biomechanical risk during lifting tasks. The framework leverages the NIOSH index to perform online risk assessment, thus fitting real-time applications. In particular, the human state is retrieved via inverse kinematics/dynamics algorithms from wearable sensor data. Human action recognition and motion prediction are achieved by implementing an LSTM-based *Guided Mixture of Experts* architecture, which is trained offline and inferred online. With the recognized actions, a single lifting activity is divided into a series of continuous movements and the *Revised NIOSH Lifting Equation* can be applied for risk assessment. Moreover, the predicted motions enable anticipation of future risks. A haptic actuator, embedded in the wearable system, can alert the subject of potential risk, acting as active prevention device. The performance of the proposed framework is validated by simulating the execution of a lifting task, while the subject is furnished with the iFeel wearable system. The source code for this paper is available here.**

I. INTRODUCTION

Work-related low-back disorders (WLBDs) still represent a societal challenge that threat the health conditions of working adults [1]. Among the large variety of their causes, payload lifting tasks in industrial environments play a pivotal role to determine poor ergonomic conditions that favor WLBDs [2]–[4]. In this context, ergonomics techniques to assess the quality of work conditions emerged, albeit based on qualitative questioners that are often costly and inconvenient to apply for dynamically changing work environments. It is then essential to develop quantitative, scalable systems that online monitor human ergonomics and that potentially alert the worker before endangering health conditions. This paper proposes a framework that combines wearables, learning-based methods and traditional lifting ergonomics to enable early assessment and prevention of worker biomechanical risk during lifting tasks execution. The haptic actuator integrated inside the wearables provides vibrotactile feedback to ensure the safety awareness of the worker.

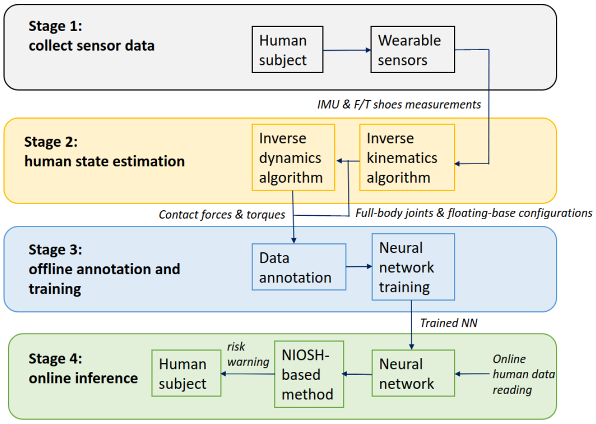


Fig. 1: An overview of the proposed framework.

The *Revised NIOSH Lifting Equation* (RNLE) is a renowned tool for assessing two-handed manual lifting ergonomics – it is published by the National Institute for Occupational Safety and Health (NIOSH) [5], [6]. The RNLE defines a *Recommended Weight Limit* (RWL) and a *Lifting Index* (LI) based on payload weight, which may lead to reliable risk assessment for WLMDs [4]. Unfortunately, approximately 35% of lifting tasks and 63% of workers can not be assessed by means of the RNLE due to its limited number of parameters and system constraints [7]. To overcome such limitations, further approaches have been proposed to assess lifting-related risks, e.g., *L5-S1 Internal Forces* [8], *Mechanical Energy Consumption* [9] and *Muscles Co-Activation* [10]. However, these offline ergonomics evaluation tools are not flexible enough to be used directly in an unstructured work environment.

As an attempt towards online human ergonomics evaluation, observational methods – like the *Rapid Entire Body Assessment (REBA)* and *Rapid Upper Limb Assessment (RULA)* – are leveraged for human-robot interaction [11]. The human data are measured by wearable sensors and an estimation of motion's ergonomics is provided by automatically fulfilling the worksheet. More recently, real-time tools for tracking whole-body joint compressive forces during robot interactions are employed in [12]. Analogously, the overloading joint torques can be computed using the displacement of the Center of Pressure in heavy lifting tasks with visual feedback [13]. For manual lifting tasks, the existing attempts are either overly generic, e.g. [11], or limited by hardware settings, e.g. portability restriction [13]. They also lack the ability to predict risks in advance and to alert the worker beforehand that biomechanical risks endanger health conditions.

*This work was supported by the Italian National Institute for Insurance against Accidents at Work (INAIL) ergoCub Project.

[1] Artificial and Mechanical Intelligence at Istituto Italiano di Tecnologia, Center for Robotics Technologies, Genova, Italy. email: `firstname.lastname@iit.it`

[2] Machine Learning and Optimisation, The University of Manchester, Manchester, United Kingdom.

[3] Mathematical and Computational Sciences, University of Toronto Mississauga, Canada.

[4] Department of Occupational and Environmental Medicine, Epidemiology and Hygiene, INAIL, Rome, Italy.

To recognize actions and predict human motions, often the approach is to tackle the two issues separately. Action recognition can be tackled as a classification problem by applying various supervised learning methods [14], [15], while motion prediction is more often regarded as a regression problem that can be addressed by means of generative adversarial networks [16], graph convolutional networks [17], dropout autoencoder LSTM [18] and etc. However, the *Guided Mixture of Experts* can resolve these two issues simultaneously [19], having the potential to simplify the architecture for motion prediction and risk assessment.

This paper proposes a learning-based approach that enables predictions of worker biomechanical risk during lifting tasks with anticipated haptic feedback. We use IMU-based sensing systems that show some advantages over vision-based approaches when used for human motion tracking due to easier calibration and a more convenient application in wider, partially occluded spaces. Our sensing suit requires 10 IMUs and a pair of Force/Torque shoes from which state estimation can be solved as *Inverse Kinematics* and *Inverse Dynamics* problems. More specifically, the contribution of this paper is threefold. First, we develop a system that can online monitor human ergonomics in the context of lifting activities. To do so, we propose a framework that integrates both the human state estimation algorithm and human action/motion prediction method, enabling the RNLE to not only estimate but also predict lifting risk continuously. Second, we adapted the *Guided Mixture of Experts* (GMoE) approach for recognizing a set of predefined actions that compose a complete lifting activity. The GMoE network is trained on a data set collected in a laboratory environment. Finally, we validate the proposed framework via an experimental analysis conducted in real-time lifting tasks.

The paper is organized as follows. In Section II we introduce the underlying technologies used in our research. In Section III the proposed framework is clarified, including the implementation of RNLE-based human lifting ergonomics monitoring system. Section IV presents an experimental analysis conducted on a human subject. At last, Section V concludes the paper.

II. BACKGROUND

A. Wearable Sensors

Sensing technologies are used to collect inputs from the environment by measuring physical quantities. In this research, we employed *iFeel*, a wearable sensors system developed by Istituto Italiano di Tecnologia (IIT) to monitor human states and provide responses. The system integrates both motion capture and force/torque sensing. Motion capture aims at tracking and recording the motion, based on inertial measurement unit (IMU) sensors. IMUs ensure high-frequency data and low latency, making *iFeel* suitable for real-time motion tracking. F/T sensors are used for measuring and regulating contact forces/torques when interacting with the environment.

B. Human Modeling and State Estimation

The human is modeled as a floating-base multi-rigid-body dynamic system [20]. The system configuration is represented by $q = (q_b, s)$, where q_b implies the floating-base pose (position and orientation) w.r.t. the inertial frame $\mathcal{I}$ and s is the joint position vector. The system velocity and acceleration are denoted by ν and $\dot{\nu}$ respectively. The n+6 equations describing human motion with n_c applied external wrenches is [21]:

$$M(q)\dot{\nu} + C(q,\nu)\nu + g(q) = B\tau + \sum_{k=1}^{n_c} J_k^T(q) f_k^c, \qquad (1)$$

where $M(q)$ and $C(q,v)$ represent respectively the mass and Coriolis effect matrix. $g(q)$ is the vector of the gravitational term. B is a selector matrix for joint torques τ. J_k is the *Jacobian* mapping the system velocity with the *k-th* link velocity that is associated with the external wrench f_k^c.

To estimate in real time the system configuration q and its velocity ν, a *dynamical inverse kinematics optimization* approach is proposed in [22]. The idea is to minimize the distance between the computed state configuration $(q(t), \nu(t))$ with the target measurements. First, the measured velocity is corrected using a rotation matrix. Then, to compute the state velocity, the constrained inverse differential kinematics for the corrected velocity vector is solved as a QP optimization problem. At last, the state velocity is integrated to obtain the configuration $q(t)$. For the base estimation, force/torque measurements are applied to determine the location of contacts. Then base estimation can be solved as part of the *dynamical inverse kinematics framework* [23].

In [20], the estimation of the human dynamics is performed by means of a Maximum-A-Posteriori (MAP) algorithm. The overall system dynamics can be reshaped to an equivalent compact matrix form. In this (Gaussian) domain, the vector of human kinematics/dynamics quantities can be regarded as stochastic variables. Given the measurement reliability, the solution is computed by maximizing the probability of this kinematics/dynamics vector.

C. Guided Mixture of Experts

The problem of simultaneous human action recognition and motion prediction is solved jointly by a learning-based approach proposed in [21], namely the *Guided Mixture of Experts* (GMoE). Given the past human states x_{k-i}, external forces f_{k-i}^c and hidden states r_{k-i}, the next optimal human state x_{k+1}^* can be formulated as:

$$x_{k+1}^* = \mathcal{H}^*(x_k, ..., x_{k-N}, f_k^c, ..., f_{k-N}^c, r_k, ..., r_{k-N}), \quad (2)$$

where the optimal mapping $\mathcal{H}^*$ is learned from human demonstration. By recursively applying equation (2), we can predict the future human states for the time horizon T.

In terms of r_{k-i}, we only consider human symbolic actions as the hidden states for simplification and estimate it as a classification problem. Hence, equation (2) can be further rearranged as:

$$\tilde{a}_{k+1} = \mathcal{D}_1^*(x_k, ..., x_{k-N}, f_k^c, ..., f_{k-N}^c) \ , \tag{3a}$$

$$\tilde{x}_{k+1}, \tilde{f}_{k+1}^c = \mathcal{D}_2^*(x_k, ..., x_{k-N}, f_k^c, ..., f_{k-N}^c, \tilde{a}_{k+1}) \ . \tag{3b}$$

where $\tilde{a}_{k+1}$ denotes the estimated human next action, $\mathcal{D}_1^*$ and $\mathcal{D}_2^*$ are two optimal mappings to learn.

Integrating the idea of Mixture of Experts (MoE), the *gating network* is guided to learn mapping $\mathcal{D}_1^*$ as a classification problem for recognizing human actions, while each *expert network* learns $\mathcal{D}_2^*$ as a regression problem to predict human motions associated with each specific action.

D. Revised NIOSH Lifting Equation

The *Revised NIOSH Lifting Equation* (RNLE) consists of the following two empirical equations:

$$\text{RWL} = \text{LC} \cdot \text{HM} \cdot \text{VM} \cdot \text{DM} \cdot \text{AM} \cdot \text{FM} \cdot \text{CM} \ , \tag{4a}$$

$$\text{LI} = W_{payload} / \text{RWL} \ . \tag{4b}$$

Equation (4a) determines a *Recommended Weight Limit* (RWL) for a specific task. Each factor in the equation is either from a qualitative assessment or from geometrical measurements weighted by a multiplier. More precisely, *LC* is the load constant (23kg), *HM* is the horizontal multiplier, *VM* is the vertical multiplier, *DM* is the vertical traveling distance multiplier, *AM* is the asymmetry multiplier, *FM* is the frequency multiplier and *CM* is the coupling multiplier.

The *Lifting Index* (LI) provides an estimate of the physical stress level, which is obtained in equation (4b) by dividing the payload weight $W_{payload}$ by the recommended weight limit. A LI smaller than 1.0 implies a safe condition for working healthy employees, while a higher value of LI denotes an increasing risk of work-related injuries.

III. PROPOSED FRAMEWORK

In this research, we propose a four-stage framework, as illustrated in Figure 1, that integrates methods introduced in Section III for continuously estimating lifting risks and monitoring human ergonomics in a real-time application. For clarification, a more detailed data flow of working pipelines is shown in Figure 2.

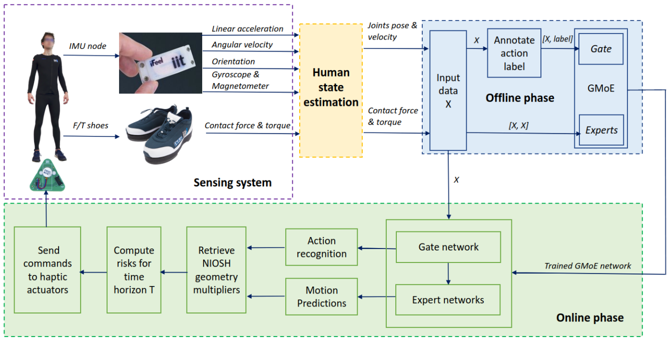


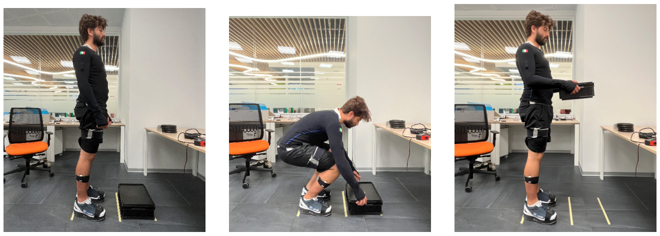
Fig. 2: Data flow of the proposed framework, composed of an online and offline phase.

Firstly, human kinematic measurements and ground contact force/torque are collected by *iFeel node* and *F/T shoes*. The sensor data are then regarded as the *targets* for estimating human full-body joints/floating-base configurations (e.g. positions and velocities) and external feet wrenches (e.g. forces and torques) via Inverse Kinematics (IK) and Inverse Dynamics (ID) algorithms. Afterwards, the outputs of the *human state estimation* module are manually annotated according to pre-defined action labels for training the GMoE network offline. During the online inference phase, combining the outputs of GMoE and IK/ID modules, the *NIOSH-based method* module is able to provide risk predictions for a given time horizon and thus send commands to haptic actuators worn by human subject.

A. Data Preparation

To apply GMoE for a lifting task scenario, we build a data set of in total 900 seconds, note that the data sampling frequency is 0.01 (*iFeel* sensor measurements are streamed every 10 ms), hence about 94456 data frames are available. The data set consists of two volunteers executing three types of lifting tasks repetitively, each task lasts for 150 seconds. For simplification, during each task, the volunteer is asked to naturally lift a 3kg payload to a certain height without twisting the upper trunk. The lifting height ranges from 68cm to 92cm, while the other variables (e.g., horizontal distance, payload weight and etc.) maintain the same.

Assume the human subject starts with a *standing* pose, a very natural sequence of actions during a single lifting activity is *squatting*, *rising* and back to *standing* pose again. The lifting risks are most likely to happen during *squatting* and *rising* phases. To apply the NIOSH equation, we need to identify when is the starting and ending moment of each action, such that we can determine the origin and destination status of human subject. For this purpose, we divide a single lifting activity into three continuous phases, each represented by a specific action, as demonstrated in Figure 3.



(a) *standing* (b) *squatting* (c) *rising*

Fig. 3: The three phases composing the lifting activity.

Since manual labeling is very expensive, we hence developed an autonomous tool to improve annotation efficiency. For the purpose of labeling, the estimated human whole-body data are visualized via a URDF model. In the meanwhile, the data are streamed in a terminal with a fixed frequency. By observing the action change of the URDF model, an action label is carefully assigned to the current data frame. As long as no new label is given by the user, the following data frames are considered to belong to the previous action. More precisely, the border between *standing* and *squatting* lies in the observation of a tendency of bending knees. Once the *squatting* action is reaching the end, the action label will soon be assigned as *rising* after observing the ascent of the pelvis. The accomplishment of *rising* is recognized by observing a totally erect trunk, hence the label will be assigned as *standing* once again. In the end, all the annotated data are divided into three subsets, 70% for training, 20% for validation, and the last 10% for test.

B. GMoE for Lifting Activity

For the purpose of simultaneous action recognition and motion prediction, we adopt the network model proposed in [21]. Since three actions are considered in our case, the implemented GMoE architecture thus consists of three similar *expert* networks and one *gate* network as illustrated in Figure 4. The input layer is of size 10x74, where 10 represents the window size for reading past data frames, 74 is the number of input features, consisting of 31 joint positions, 31 velocities, and 12 contact forces/torques. The output layer size of the *gate* network is 3x50, where 3 denotes the action categories, and 50 means the predicted maximum action probabilities in future frames. It should be noted, when creating time series data in practice, the sampling rate of input data is set as 3 to be aligned with inference time, such that the period between two data frames in an input sequence is 30ms. Therefore the total prediction time horizon is 1.5 seconds. Similarly, the output size of each *expert* network is 3x50x43, where 31 joints' positions and 12 foot wrenches are considered (in total 43 output features), excluding the joints' velocities.

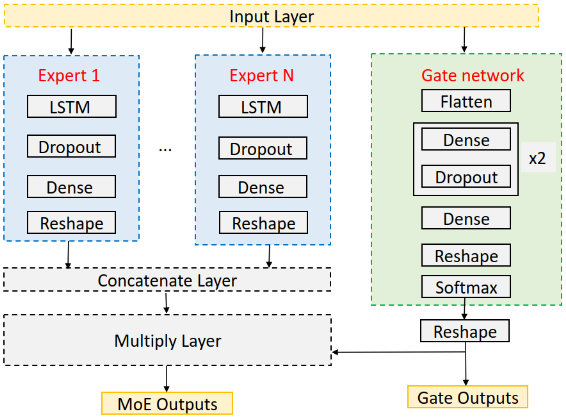


Fig. 4: Adapted structure of Guided Mixture of Experts architecture for action recognition and motion prediction.

During the training phase, the loss function L_1 associated with *gate* network and loss function L_2 associated with *expert* network are chosen as categorical cross-entropy loss and mean squared error loss, respectively. The total loss function L for GMoE is expressed as a linear combination of L_1 and L_2:

$$\begin{aligned} L &= b_1 L_1 + b_2 L_2 \\ &= -\frac{b_1}{2M}\sum_{t=1}^{T}\sum_{j=1}^{M}\sum_{i=1}^{N} a_i^{j,t} log(\tilde{a}_i^{j,t}) \\ &+ \frac{b_2}{2M}\sum_{t=1}^{T}\sum_{j=1}^{M} \| \sum_{i=1}^{N} \tilde{a}_i^{j,t} \tilde{\boldsymbol{y}}_i^{j,t} - \tilde{\boldsymbol{y}}^{j,t} \|_2 \end{aligned} \tag{5}$$

where b_1 and b_2 are manually chosen for the convergence of both classification and regression problems (in this case, b_1 is 1.0 and b_2 is 0.5 for faster convergence of *gate* network), T is the prediction time horizon, M is the total number of data frames, N is the number of experts, scalar value $a_i^{j,t}$ and vector $\tilde{\boldsymbol{y}}_i^{j,t}$ denote for human action and motion ground truth associated with i-th action and j-th data frame at time instance t in the future, operator $\tilde{}$ represents prediction values of both action recognition and future motions. To update the network weights, *Adam* optimizer is applied with epsilon equals 1e-6. Moreover, early stopping technique and adaptive learning rate are used to avoid overfitting or local optimum.

C. Risk Prediction and Haptic Alert

As mentioned in Section III-A, action recognition is used to determine the origin and destination moment of each action during a single lifting activity. Once the origin status is identified, each following moment can be considered a temporary destination status, which makes it possible to use NIOSH equation to compute risk at that moment. Until next action is detected, the NIOSH equation can be applied repeatedly without violating any constraints. Furthermore, by making use of predicted motions, we are also capable to predict potential risks in the future for a given time horizon. The process of estimating and predicting risks is demonstrated in Algorithm 1. Once any potential risk is detected, a signal will be sent to the haptic actuator mounted on the human's back. The signal strength corresponds to the predicted risk level. The human can thus take appropriate measures based on the vibrotactile feedback, i.e., to abort the task immediately or adjust only the lifting posture.

In practice, the action transition cost about 0.5s, which affects the accuracy of action detection. To retrieve more precise NIOSH variables, we implement an approach to compensate action change delay. At each moment, when the probability of previously recognized action is growing, the current action label maintains the same. Once the probability decreases over a pre-defined threshold, we consider the action transition already starts. Then we search for the action label whose probability increases also over a threshold.

As shown in Algorithm 1, from predicted motions we can update the human URDF model in simulator and retrieve geometry values to compute NIOSH variables H, V and D. Assume that the middle point of human hands is always overlapped with the Center of Mass (CoM) of the payload, H can be thus represented as the horizontal distance between the position of the CoM of human hand w.r.t. the frame attached to human foot, while V is computed by using the

Algorithm 1 Risk prediction using RNLE

```
Require: action at $t_0$: $A_{t_0}$, action at $t$: $A_t$, motion prediction at $t$ for future N steps: $M_t^{t+N}$, human origin status at $t_0$: $S_{t_0}$, NIOSH variables: $A$, $C$, $F$
Ensure: risk prediction at $t$ for future N steps: $R_t^{t+N}$
  Initialize $R_t^{t+N}$
  while $True$ do
      if $A_t$ is not $A_{t_0}$ then                ▷ Detect next action
          $A_{t_0} \leftarrow A_t$
          $S_{t_0} \leftarrow getHumanStatus(M_t^{t+N}[0])$
      end if
      for each item $i$ in $M_t^{t+N}$ do
          $S_t \leftarrow getHumanStatus(M_t^{t+N}[i]))$
          $H, V, D \leftarrow getVariables(S_{t_0}, S_t)$
          $R_t^{t+N}.append(RNLE(H, V, D, A, C, F))$
      end for
      return $R_t^{t+N}$
  end while
```

vertical position of the human hand w.r.t. the human foot:

$$H = \frac{H_{LeftHand}^{LeftFoot} + H_{RightHand}^{RightFoot}}{2} \ , \tag{6a}$$

$$V = \frac{V_{LeftHand}^{LeftFoot} + V_{RightHand}^{RightFoot}}{2} \ . \tag{6b}$$

and vertical traveling distance is denoted as $D = V_t - V_{t_0}$, where V_t and V_{t_0} represent the vertical distance at the destination and origin moment, respectively. For simplification, asymmetry angle A is not considered in our case, hence AM constantly equals 1. Lifting frequency is computed as the average number of lifts per minute over a 15-minute period. The coupling situation is considered as *Fair*.

IV. VALIDATION

A. Experimental Setup

To validate the performance of the proposed framework for assessing lifting risk in a real-time application, an experimental analysis is performed in a laboratory environment. A healthy volunteer is asked to perform three different lifting tasks corresponding to varied risk levels. In this setup, the participant's kinematics state is collected using *iFeel*, which is composed of a set of *iFeel-Nodes* (including sensors and actuators) and a central processing unit *iFeel-Station* (a micro-controlled board). The system operates for whole-body motion tracking via *iFeel-Nodes* that are mounted in predefined locations of the *iFeel-Suit*. Each *iFeel-Node* contains a 9-DoF IMU that provides absolute orientation and sensor-based velocity fusion data at a rate of 100 Hz. Once detecting any possible risks, a signal is sent to the haptic actuator of the *ifeel node* mounted on the human waist. The ground reaction forces and torques are retrieved using *iFeel-Shoes* equipped with F/T sensors integrated in the front and rear parts. The collected human data are streamed and resampled via YARP middleware [24] at a rate of 100Hz. Moreover, as mentioned in section II-B, human is modeled as a 66-DoFs floating-base multi-rigid-body system. However, for simplification, only 13 joints (e.g. T9T8, Right shoulder and etc.) are considered in our case, thus a reduced 31-DoFs URDF model is applied for simulation and visualization. The programs run on a 64-bit i7 2.6GHz laptop which is equipped with 32 GB RAM, Intel(R) UHD Graphics and Ubuntu 20.04 LTS.

TABLE I: Experimental lifting task variables of RNLE.

Task type	RNLE variables							RNLE results	
	H_origin (cm)	H_end (cm)	V_origin (cm)	V_end (cm)	D_origin (cm)	D_end (cm)	L (kg)	RWL_origin (kg)	LI
	HM_origin	HM_end	VM_origin	VM_end	DM_origin	DM_end		RWL_end (kg)	
Task 1	47	63	8	68	60	60	3	5.84	0.51
	0.53	0.40	0.80	0.98	0.90	0.90		5.40	0.56
Task 2	47	63	8	80	72	72	7	5.71	1.23
	0.53	0.40	0.80	0.99	0.88	0.88		5.33	1.31
Task 3	47	63	8	92	83	83	10	5.64	1.77
	0.53	0.40	0.80	0.95	0.87	0.87		5.06	1.98

The parameters of performed lifting tasks are listed in Table I. Additionally, asymmetry angle A equals zero (AM = 1.0), coupling quality is *Fair* (CM = 0.95), and lifting frequency is controlled as 7 lifts/min (FM = 0.7). Moreover, the payload is evenly distributed inside a square box.

During the experiment, the participant is asked to repeat each task three times in a steady and natural way, such that no jerks appear during lifting. The participant should avoid twisting the upper trunk so that the assumption of zero asymmetry angle is fulfilled. Furthermore, the participant is required to hold the box with both hands while his feet maintaining in a fixed position. The lifting activity is executed slowly, hence every single execution can be regarded as independent from the others.

B. Results Analysis

In this section, we first performed a variety of quantitative evaluations of the adapted GMoE model using an additionally collected unseen dataset, which is in a total of 15248 frames. Then we conducted a qualitative analysis based on the results of previously designed online experiments.

1) Quantitative Evaluation of Action Recognition: In order to assess the action classification performance, a confusion matrix associated with three human lifting actions is presented in Figure 5. Based on this confusion matrix, metrics such as *Accuracy*, *Precision*, *Recall* and *F1 score* can be further retrieved. *Accuracy* is the number of correct predictions of all N categories divided by the total number of predictions, as shown in Equation 7, where *total* means the number of all tested samples.

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{total} \tag{7}$$

Precision refers to the proportion of correctly predicted positive instances out of all instances predicted as positive, while *Recall* measures the proportion of correctly predicted positive instances out of all actual positive instances, as shown in Equation 8a and 8b, where i means each class.

$$Precision = \frac{TP_i}{TP_i + FP_i} \ , \tag{8a}$$

$$Recall = \frac{TP_i}{TP_i + FN_i} \ . \tag{8b}$$

TABLE II: Performance metrics for assessing GMoE model recognizing multi-class human lifting actions.

	Accuracy	**Precision**	**Recall**	**F1 score**
standing	/	0.890	0.969	0.928
rising		0.898	0.869	0.883
squatting		0.925	0.816	0.867
average	0.899	0.904	0.885	0.893

F1 score can be interpreted as a harmonic mean of the *Precision* and *Recall* as shown in Equation 9.

$$F1 = \frac{2 * Presicion * Recall}{Precision + Recall} \tag{9}$$

Table II summarizes the experimental results of these metrics for each single category classification. As we can see, *squatting* has the highest accuracy of 0.925, which indicates that the model has a low rate of falsely labeling instances as this action. On the contrary, *standing* has a relatively low accuracy. This is mainly because the transition period between *rising* and *standing* can be quite ambiguous (also partly due to the fact that the annotated border depends on human judgment), such that it can be hard for the model to distinguish these two phases exactly. Furthermore, both *squatting* and *rising* have relatively lower *Recall* values than *standing*. As explained before, the ambiguity between *rising* and *standing* leads to some false labeling of *standing* when they are actually *rising*. Also, the similarity between the motion patterns of *squatting* and *rising* (they are basically reversed) results in the confusion of them.

2) *Quantitative Evaluation of Motion Prediction:* In the following, we report the performance of GMoE regarding the task of motion prediction. For the sake of simplicity, two typical joints that can reflect the human motion patterns during a lifting task are chosen, namely, the left knee and right elbow. The rotational angles around the y-axis of these two joints during a period of about 5500 frames are demonstrated in Figure 6. The ground truths are depicted in black curves, while the predicted rotational angles at the future time steps 0, 19 and 49 are shown with blue, orange and yellow curves, respectively.

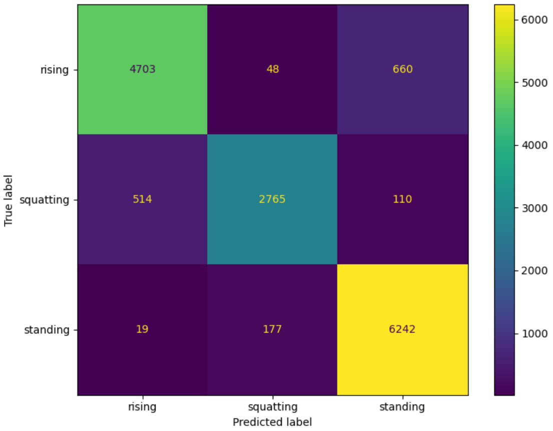


Fig. 5: Confusion matrix for the classification of three human lifting actions.

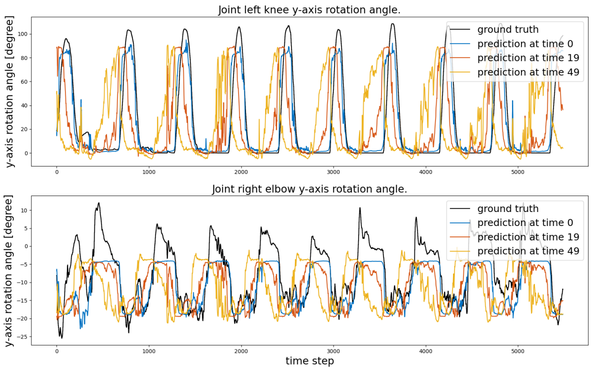


Fig. 6: Multi-time-step predictions of the y-axis rotation angle of the joint left knee and right elbow.

It can be observed from both rows in Figure 6 that the predicted y-axis rotational angle at a future time step 0 (blue curves) and the ground truths (black curves) basically coincide, despite the gaps at peaks. The amplitude differences at peak positions can be more easily observed for the right elbow joint. The predicted rotational angles at future time steps 19 and 49 exhibit a leading phase compared with the ground truth, where the phase differences should match the corresponding prediction time steps. It should be noted that the predictions at future time step 49 suffer more from uncertainties, which is reflected by the frequently appearing sharp fluctuations. This may be due to the fact that the model only has very limited historical information, yet to make a further prediction in the future, it is apparently insufficient to solely rely on this short period of history. Another interesting fact is that the model seems to perform worse in predicting the motions of the right elbow joint. A possible reason could be that the movements of the right elbow are also affected by the pose of the pelvis, while the knees have a more independent thus also more predictable motion pattern.

3) *Qualitative Evaluation:* To further evaluate the effectiveness of the proposed framework, we analyze qualitatively the results of *Task 2* (shown in Table I) as an example. A complete process of *rising* is demonstrated in Figure 7. As shown in the first row, the motions of both real human subject and simulated models are captured. The grey model reconstructs the human motion at current time *t* from sensor measurements, while the red model represents the predicted human motion at future time *t+0.6s* (in the experiment we output the maximum future 20 data frames, recalling the period between each data frame is 30ms, thus the prediction time is 0.6s). The correspondingly recognized actions at each moment are presented in the second row. The black, blue and red solid curves denote the probability of action *rising*, *squatting* and *standing*, respectively. In the third row,

we demonstrate the predictions of rotation degrees of left knee joint around y-axis for future 1.5s (maximum 50 future data frames), associated with the round dot curves. In the meanwhile, the blue curve stands for the ground truth of left knee joint rotation values. Figures in the last row demonstrate the lifting index during the *rising* action. The red curve and grey dot curve represent the risk value at the current time and future 0.9s (namely 30 data frames), respectively.

As shown in the picture at the top left in Figure 7, the human is almost finishing the action *squatting* at t=10.9s, and as the red model indicates, at the future time t=11.5s, the human model would probably be rising up a little bit. The recognized action at t=10.9s is still *squatting*, therefore no lifting risk is detected and the haptic actuator remains silent. As for the rotation angle of the left knee joint, it also reaches a peak value of about 100 degrees and it's going to decrease soon. When time t becomes 11.2s, it can be observed that the gray model is reaching the pose as predicted at t=10.9s. In the meanwhile, the action transition already started, thus we can see that the lifting risk grows from zero to 0.7 (hence a slight haptic alert appears), and as the predictions show, the risk value at t=11.8s should be equal to 1.0. Then at t=11.9s, the human is reaching the table and intends to put the payload on it. At this moment, the action is still recognized as *rising* with maximum probability. Moreover, the currently estimated lifting index is around 0.9 (corresponding to a medium haptic warning), which almost equals the value predicted at t=11.2s. At the final time t=12.4, apparently the *rising* action is completed, and the human subject is getting back to *standing* pose. Therefore the probability of *rising* starts to decrease. Correspondingly, the lifting index returns back to zero again.

4) Failure Cases: We present some failure cases here to reveal the limitations of the current system. As explained in Section III-A, the GMoE network is trained on a 15-mins data set that consists of basic lifting tasks. Hence, a very typical unsuccessful scenario is when completely unseen motion patterns appear in the online application, e.g., trunk twisting and overhead lifting. In such cases, precise action detection can become an issue, let alone predict risks. Another challenge lies in the restrictions of the NIOSH equation. For example, the system is not applicable to collaborative lifting tasks where multiple workers are present. Moreover, the noise and perturbations accumulated over time in online applications also have a great effect on the accuracy of the GMoE model. We hypothesize that the retrievement of unprecise NIOSH variables is also a notable limitation. This is the main reason for improving the swiftness of action detection and the accuracy of motion predictions.

C. Discussions

In comparison to risk assessment approaches proposed in literature [11]–[13], the main advantage of the proposed framework lies in its ability to early assessment and prevention of biomechanical risks faced by workers during realistic lifting tasks, by utilizing a learning-based approach and wearable sensing system. Despite training on a relatively small data set, we have shown that our model is able to generalize well to unseen data, as analyzed in IV-B.1 and IV-B.2. We also demonstrate robust qualitative performance during the live demo presented in IV-B.3. It is worth mentioning that although humans can feel muscular fatigue in the long term, the causal action is often neglected due to the lack of real-time quantitative ergonomic feedback. Therefore the anticipated haptic alerts play a vital role in improving the risk awareness of workers while performing heavy lifting tasks.

V. CONCLUSIONS

In this paper, we presented a framework that integrates wearable sensing, human state estimation, human action/motion prediction and NIOSH index for real-time manual lifting applications. Through online recognition of human actions, the execution of a single lifting activity can be segmented into a series of continuous parts. The commencement of each sub-action is considered the initial human state, with subsequent moments within this sub-action being regarded as temporary destination states. With the help of motion prediction, future human status can also be obtained. Hence RNLE can be applied to assess risks within the predicted time horizon. The vibrotactile feedback enables anticipated alert on the predicted lifting risks. The performance of the framework is tested in an experimental lifting scenario using the iFeel wearable system.

Future work should first address the problem of generalization by expanding the current lifting data set, such that more complex realistic lifting tasks can be considered. By improving the performance of GMoE model, a more precise retrieval of NIOSH geometry variables could be expected. It would also be interesting to include upper trunk twisting and overhead lifting in order to utilize the NIOSH equation. Moreover, a learning-based ergonomics assessment approach could be another promising topic.

REFERENCES

[1] H. D. of Biomedical and B. Science, *Work practices guide for manual lifting*. US Department of Health and Human Services, Public Health Service, Centers ..., 1981, no. 81-122.

[2] P. P. F. Kuijer, J. H. Verbeek, B. Visser, L. A. Elders, N. Van Roden, M. E. Van den Wittenboer, M. Lebbink, A. Burdorf, and C. T. Hulshof, "An evidence-based multidisciplinary practice guideline to reduce the workload due to lifting for preventing work-related low back pain," *Annals of occupational and environmental medicine*, vol. 26, no. 1, pp. 1–9, 2014.

[3] M.-L. Lu, T. R. Waters, E. Krieg, and D. Werren, "Efficacy of the revised niosh lifting equation to predict risk of low-back pain associated with manual lifting: a one-year prospective study," *Human factors*, vol. 56, no. 1, pp. 73–85, 2014.

[4] T. R. Waters, M.-L. Lu, L. A. Piacitelli, D. Werren, and J. A. Deddens, "Efficacy of the revised niosh lifting equation to predict risk of low back pain due to manual lifting: expanded cross-sectional analysis," *Journal of Occupational and Environmental Medicine*, pp. 1061–1067, 2011.

[5] T. R. Waters, V. Putz-Anderson, A. Garg, and L. J. Fine, "Revised niosh equation for the design and evaluation of manual lifting tasks," *Ergonomics*, vol. 36, no. 7, pp. 749–776, 1993.

[6] T. R. Waters, V. Putz-Anderson, and A. Garg, "Applications manual for the revised niosh lifting equation," 1994.

[7] P. G. Dempsey, "Usability of the revised niosh lifting equation," *Ergonomics*, vol. 45, no. 12, pp. 817–828, 2002.

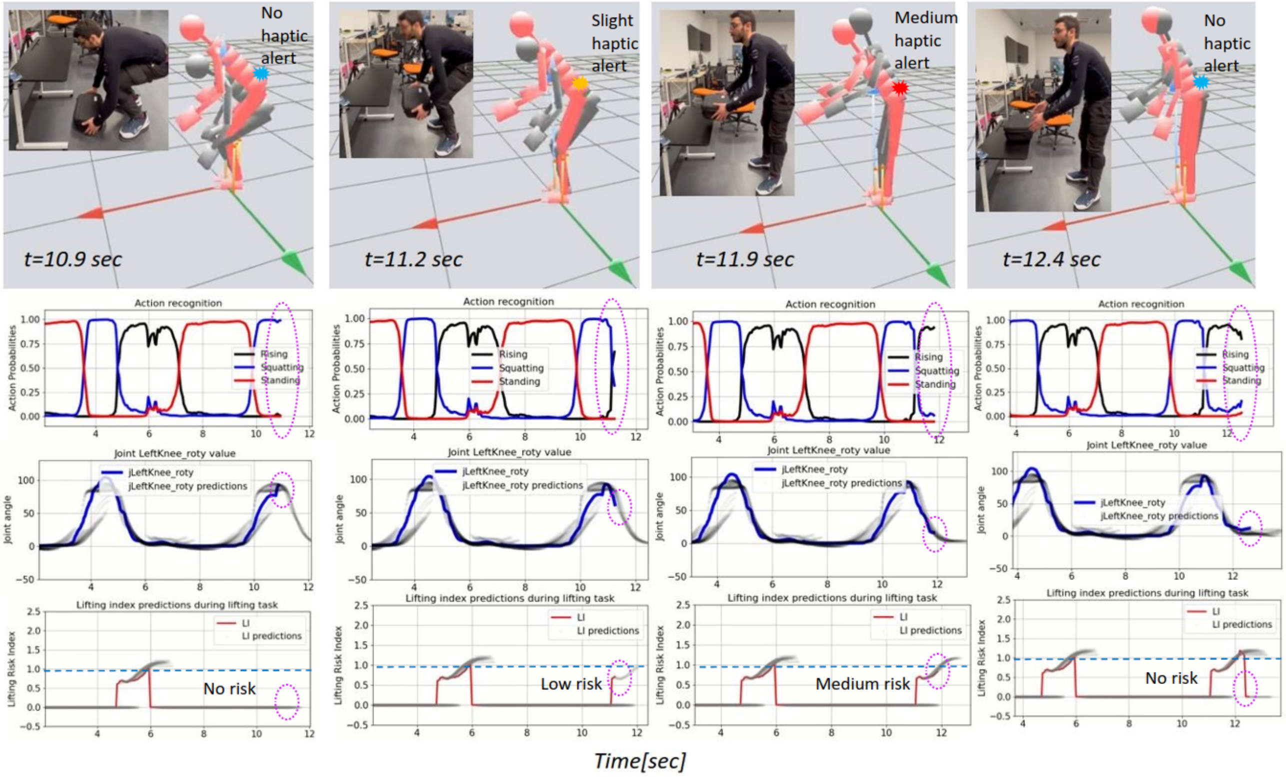


Fig. 7: Experimental results of the online action recognition and risk prediction architecture. The first row shows pictures of the sensorized subject during the task execution and virtual model visualization with estimated (gray) and predicted (red) configuration. In the second row, it is shown the action prediction probability. In the third row, ground truth and prediction of the left knee joint rotation angle are depicted. The bottom row shoes lifting index for the period till prediction time horizon.

[8] S. A. Lavender, G. B. Andersson, O. D. Schipplein, and H. J. Fuentes, "The effects of initial lifting height, load magnitude, and lifting speed on the peak dynamic l5/s1 moments," *International Journal of Industrial Ergonomics*, vol. 31, no. 1, pp. 51–59, 2003.

[9] A. Ranavolo, T. Varrecchia, M. Rinaldi, A. Silvetti, M. Serrao, S. Conforto, and F. Draicchio, "Mechanical lifting energy consumption in work activities designed by means of the "revised niosh lifting equation"," *Industrial health*, vol. 55, no. 5, pp. 444–454, 2017.

[10] A. Ranavolo, S. Mari, C. Conte, M. Serrao, A. Silvetti, S. Iavicoli, and F. Draicchio, "A new muscle co-activation index for biomechanical load evaluation in work activities," *Ergonomics*, vol. 58, no. 6, pp. 966–979, 2015.

[11] A. Shafti, A. Ataka, B. U. Lazpita, A. Shiva, H. A. Wurdemann, and K. Althoefer, "Real-time robot-assisted ergonomics," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1975–1981.

[12] L. Fortini, M. Lorenzini, W. Kim, E. De Momi, and A. Ajoudani, "A real-time tool for human ergonomics assessment based on joint compressive forces," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 1164–1170.

[13] L. Fortini, W. Kim, M. Lorenzini, E. De Momi, and A. Ajoudani, "A framework for real-time and personalisable human ergonomics monitoring," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 101–11 107.

[14] R. Zhao, W. Xu, H. Su, and Q. Ji, "Bayesian hierarchical dynamic model for human action recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7733–7742.

[15] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.

[16] A. Hernandez, J. Gall, and F. Moreno-Noguer, "Human motion prediction via spatio-temporal inpainting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7134–7143.

[17] W. Mao, M. Liu, M. Salzmann, and H. Li, "Learning trajectory dependencies for human motion prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9489–9497.

[18] P. Ghosh, J. Song, E. Aksan, and O. Hilliges, "Learning human motion models for long-term predictions," in *2017 International Conference on 3D Vision (3DV)*. IEEE, 2017, pp. 458–466.

[19] K. Darvish, E. Simetti, F. Mastrogiovanni, and G. Casalino, "A hierearchical architecture for human–robot cooperation processes," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 567–586, 2020.

[20] C. Latella, S. Traversaro, D. Ferigo, Y. Tirupachuri, L. Rapetti, F. J. Andrade Chavez, F. Nori, and D. Pucci, "Simultaneous floating-base estimation of human kinematics and joint torques," *Sensors*, vol. 19, no. 12, p. 2794, 2019.

[21] K. Darvish, S. Ivaldi, and D. Pucci, "Simultaneous action recognition and human whole-body motion and dynamics prediction from wearable sensors," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE, 2022, pp. 488–495.

[22] L. Rapetti, Y. Tirupachuri, K. Darvish, S. Dafarra, G. Nava, C. Latella, and D. Pucci, "Model-based real-time motion tracking using dynamical inverse kinematics," *Algorithms*, vol. 13, no. 10, p. 266, 2020.

[23] P. Ramadoss, L. Rapetti, Y. Tirupachuri, R. Grieco, G. Milani, E. Valli, S. Dafarra, S. Traversaro, and D. Pucci, "Whole-body human kinematics estimation using dynamical inverse kinematics and contact-aided lie group kalman filter," *arXiv preprint arXiv:2205.07835*, 2022.

[24] G. Metta, P. Fitzpatrick, and L. Natale, "Yarp: yet another robot platform," *International Journal of Advanced Robotic Systems*, vol. 3, no. 1, p. 8, 2006.