scientific reports

Check for updates

OPEN

Validation of computer vision-based ergonomic risk assessment tools for real manufacturing environments

Thomas Agostinelli[1], Andrea Generosi[1]✉, Silvia Ceccacci[2] & Maura Mengoni[1]

This study contributes to understanding semi-automated ergonomic risk assessments in industrial manufacturing environments, proposing a practical tool for enhancing worker safety and operational efficiency. In the Industry 5.0 era, the human-centric approach in manufacturing is crucial, especially considering the aging workforce and the dynamic nature of the entire modern industrial sector, today integrating digital technology, automation, and sustainable practices to enhance productivity and environmental responsibility. This approach aims to adapt work conditions to individual capabilities, addressing the high incidence of work-related musculoskeletal disorders (MSDs). The traditional, subjective methods of ergonomic assessment are inadequate for dynamic settings, highlighting the need for affordable, automatic tools for continuous monitoring of workers' postures to evaluate ergonomic risks effectively during tasks. To enable this perspective, 2D RGB Motion Capture (MoCap) systems based on computer vision currently seem the technologies of choice, given their low intrusiveness, cost, and implementation effort. However, the reliability and applicability of these systems in the dynamic and varied manufacturing environment remain uncertain. This research benchmarks various literature proposed MoCap tools and examines the viability of MoCap systems for ergonomic risk assessments in Industry 5.0 by exploiting one of the benchmarked semi-automated, low-cost and non-intrusive 2D RGB MoCap system, capable of continuously monitoring and analysing workers' postures. By conducting experiments across varied manufacturing environments, this research evaluates the system's effectiveness in assessing ergonomic risks and its adaptability to different production lines. Results reveal that the accuracy of risk assessments varies by specific environmental conditions and workstation setups. Although these systems are not yet optimized for expert-level risk certification, they offer significant potential for enhancing workplace safety and efficiency by providing continuous posture monitoring. Future improvements could explore advanced computational techniques like machine learning to refine ergonomic assessments further.

Keywords Ergonomic Risk Assessment, Motion capture, Deep learning, Computer vision, Work-related MSDs, Ergonomic monitoring

In the evolution of industrial sectors, Industry 4.0 and Industry 5.0 represent transformative phases characterized by increasing digitization and human-centric approaches, respectively. Industry 4.0, which is still in the developmental stages, focuses on the integration of digital technologies such as the Internet of Things (IoT), artificial intelligence, and big data analytics to enhance industrial processes. As we transition to Industry 5.0, the emphasis shifts to incorporating human skills and expertise alongside advanced technologies to promote sustainability and resilience[1,2].

In this context, the role of human workers, especially in manufacturing sector, is pivotal, emphasizing the need for human-centric approaches to ensure flexibility, creativity, and problem-solving[3–5]. The challenge lies in creating suitable working conditions tailored to individual capabilities, going beyond the traditional view of "workers" as a homogeneous group. As the workforce's age increases and working conditions evolve, understanding and accommodating the variability in workers' capabilities becomes imperative. In fact, it is well known how aging contributes to a decline in work capacity due to physiological changes (e.g., a 65-year-old

[1]Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, via Brecce Bianche, 12, Ancona 60131, AN, Italy. [2]Department of Education, Cultural Heritage and Tourism, Università degli Studi di Macerata, via Luigi Bertelli, 1, Macerata 62100, MC, Italy. ✉email: a.generosi@univpm.it

worker has about half the physical work capacity of a 25-year-old[6]), and that employment rates are expected to rise over the next decades among people over the age of 55 (i.e., people over 55 will increase to 66.7% percent by 2060[7]). The dynamic nature of modern manufacturing - characterized by frequent changes in workstations and tasks and the introduction of collaborative robots - poses a significant challenge in maintaining correct postures and in mitigating the risk of work-related MSDs[8]. In this scenario, industries will increasingly have to account for human variability and anticipate workers' behaviours, while the traditional concept of a uniform workforce is being replaced by a need for tailored solutions that predict individual behaviours and specific work-related risks. The urgency to address these challenges is underscored by the persistent prevalence of work-related MSDs, as highlighted in the 2019 European Risk Observatory Report[9], thus, effective health and work-related MSD risk assessing methods and tools are crucial[10]. To mitigate ergonomic risks and enhance worker well-being, it is essential to consider the unique characteristics and performance of every worker. This requires the development of robust and cost-effective tools capable of directly monitoring working postures, continuously assessing ergonomic risks during work activities, informing workers in case of incorrect postures, and providing insights to managers to improve the workspace and the individuals' working conditions.

Ergonomists generally use ergonomic risk assessment methods to monitor and decrease ergonomic risks associated with work-related MSDs. Some of these methods involve evaluating risks through either direct on-site observation or posthumous analysis of previously recorded videos, focusing on workers while they perform their tasks.

Such methods primarily exploit standardized observational-based tools[11] such as Rapid Entire Body Assessment (REBA)[12], Rapid Upper Limb Assessment (RULA)[13], Ovako Working posture Analysing System (OWAS)[14], and Occupational Repetitive Actions Index (OCRA Index)[15]. This approach necessitates the involvement of an experienced ergonomist who directly observes workers' actions, either in person or through video recordings. The collection of data required for calculating the risk index is typically time-consuming because it involves subjective observation or a straightforward estimation of projected joint angles (e.g., elbow, shoulder, knee, trunk, and neck) by analysing videos or pictures, and, additionally, it is strictly subjective, so it does not guarantee the repeatability of measurements. In this scenario, addressing these limitations is crucial to developing more efficient and adaptable methods for ergonomic risk assessment in dynamic work environments. A viable solution might be the introduction of new methods and tools for automated or semi-automated ergonomic postural assessment. To this end, MoCap systems can be used to collect data accurately and quantitatively.

The paper is structured as follows: Sect. 2 presents a literature review, providing an overview of existing ergonomic risk assessment tools and methodologies. Section 3 details the materials and methods used in this study, including a performance comparison of different 2D RGB MoCap tools taken from existing literature and data collection procedures. Section 4 outlines the industrial case studies investigated. Section 5 reports the collected results. Section 6 discusses the results, limitations and future developments. Finally, Sect. 6 draws conclusions.

Literature review

Sensor and marker-based MoCap systems

The first attempts done with MoCap systems to assess ergonomic risk were based on the use of high-end technologies, leveraging optical devices (i.e., markers) or inertial sensors mounted on the tracked person's body[16–18].

Today, the most reliable commercially available options (e.g., Xsense[19], Vicon Blue Trident[20], Vicon Nexus[21], OptiTrack[22]), still use this approach based on sensor and marker-based optical systems. Although robust, these systems are costly, have setup limitations, and are challenging to use in real working environments[23]. They often require wearable sensors or markers, making them invasive and potentially altering the operator's normal behaviour. Operators wearing such sensors are always fully aware that they are being monitored: the risk is that they could behave correctly on purpose, masking any ergonomic flaws in the workstation in which they are working. Moreover, marker-based tools are highly expensive and require the use of a multiple-camera setup, which is almost always impossible in a real manufacturing environment, while sensor-based systems can also be expensive and, more importantly, they require a trained person to aid the operators in wearing the sensors in specified locations around the body and to perform a calibration procedure before every measuring session. All these factors ended up severely limiting their applicability for daily use in a factory setting[24].

AI-based MoCap systems

By leveraging Machine Learning technologies, and in particular Deep Learning, it is possible to develop models that can recognize the human figure and the position of its body joints through pattern recognition techniques. The advantage of AI-based MoCap systems over sensor and marker-based ones lies in the fact that the former don't require the operator to wear any specialized device, being it a marker or a sensor, and that they only need the use of one or two common RGB cameras (e.g., webcams or smartphones' cameras). One of the most used Deep Learning models trained to track the human body is OpenPose[25], a tool developed by Carnegie Mellon University that shows remarkable accuracy and high robustness against occlusion[26] but at the cost of being highly computationally intensive[27]. Tf-pose-estimation[28] started as a fork of the OpenPose open-source project and, thanks to its light computational requirements, it can be run even on mobile devices[29], showing at the same time good performances in pose recognition in a laboratory environment[30]. Another competitor to Openpose is MediaPipe[31], a complete suite of body feature detection models (e.g. pose, hand, iris, face) proposed by Google[32,33].

Recent advances include the development of systems such as the "Smart Ergonomic Explorer" (SEE)[34] and the "Quick Capture" system[35], both of which utilize Convolutional Pose Machines (CPM) for enhanced posture detection accuracy. The SEE system uses a proactive strategy-oriented model, integrating a rapid entire body assessment method with CPM. It is capable of capturing data in a real-world setting using a smartphone-based setup, thus improving the system's applicability and reducing the costs associated with high-end MoCap tools. Both systems demonstrated significant reliability when compared to manual evaluations. These systems have the potential to revolutionize markerless postural assessment, enhancing its accessibility and applicability[10,30]. Indeed, 2D RGB MoCap systems, although less accurate, boast portability and flexibility, reducing costs and adapting to various positioning configurations without the need to recalibrate each time.

MoCap systems validation

However, despite the promising potential of these systems, their validation has primarily occurred under controlled laboratory conditions, while limited experimentation in real working environments has been carried out so far. Moreover, based on the information available through this study, there is no systematic literature review or meta-analysis focusing on an evidence-based comparative assessment of the various RGB MoCap systems exploiting 2D RGB cameras for ergonomic assessment, in different real industrial settings. For example, in[36] the authors only performed the post-training validation of an ad hoc AI risk level classifier, without testing the resulting tool on simulated or real work tasks. On the other hand[37], presented a system capable of automatically selecting the appropriate ergonomic risk assessment tool (namely REBA, RULA, OWAS) based on the type of activity to be assessed and then automatically performing the assessment. However, the authors validated the angles measuring tool's accuracy, but they did not assess whether the tool provides accurate ergonomic risk assessments. In[38] the authors performed the validation of an ergonomic risk assessment tool in a real manufacturing environment and compared it with manual assessments. They reported statistics such as the Intra-Class Correlation (ICC) between human ergonomists and the average score of the manual and automated assessments, but they did not report the ergonomic scores directly.

The existing studies cannot be easily compared, as they test the proposed tools in very different working contexts (e.g., construction and industrial manufacturing), considering different ergonomic assessment tools and different experimental study designs. Moreover, there is a lack of consistency in reporting the results. Consequently, the deployment of these technologies across diverse industrial contexts still prompts a crucial question: are current RGB MoCap systems mature enough to be confidently used for ergonomic assessments in the multifaceted landscape of Industrial Manufacturing?

In order to provide an answer to this question, this research provides the results of experimentations performed in real industrial manufacturing, that aims to assess the limitations and potential of current RGB MoCap systems by comparing the outcomes of their application in different production lines, which differ in terms of spatial layout, automation level, workstations' proximity level, positioning of picking racks in relation to the operator, operator's mobility level during tasks. To this end, we compared the results of ergonomic risk assessments carried out using both a semi-automatic RGB MoCap system and manual analysis.

The objective of this paper is twofold. First, it seeks to assess the effectiveness of 2D RGB MoCap systems in conducting ergonomic risk analyses across various industrial manufacturing environments. It also compares the accuracy and reliability of semi-automated ergonomic analyses performed using these systems with those of traditional manual ergonomic assessments.

The second objective is to evaluate the current maturity of MoCap RGB systems as accurate measurement tools for ergonomic workstation validation, or whether they are more appropriate for continuous monitoring, alerting ergonomists to potential risk situations requiring further analysis.

To achieve these goals, the paper begins by presenting a structured benchmarking of the performance of different MoCap RGB systems, as reported in the literature. It then details the evaluation of one of these benchmarked systems in a real-world manufacturing context, providing insights into its performance within an actual production facility. The evaluation considers key environmental factors such as operator mobility, lighting conditions, and plant layout, all of which influence the effectiveness of the systems under review.

Materials and methods

To carry out a comparative evaluation of the various 2D RGB MoCap systems for ergonomic evaluation in the literature, we first proceeded with a survey of the state of the art for this topic. Table 1 reports the results of this survey, conducted through the academic research database Google Scholar, restricted to the period from 2017 (i.e., the year of the introduction of the first AI-based body tracking models[25]) to the present, and using the following query: "RGB Motion Capture systems" AND "ergonomic risk assessment" AND "industrial manufacturing" AND "computer vision" AND "continuous monitoring" AND "dynamic environments" FROM 2017 TO 2023.

Analysing them, we noticed that currently, available research papers report results in an extremely heterogeneous manner, making it challenging to objectively compare the performance achievable by the various tools. This difficulty mainly arises because some studies report comparative results between ergonomic values predicted by the proposed tool and values collected through a ground-truth approach (i.e., predicted values compared with actual values obtained through a manual assessment performed by ergonomists), while others report only statistical indicators of accuracy and reliability, i.e., accuracy or Cohen's kappa. Consequently, we decided to adopt the accuracy score, computing it for the Ergonomic Risk Assessment (ERA henceforth) Score and Risk Level. It was decided that Cohen's kappa would not be used, as this would have resulted in a loss of reliability: regardless of the ergonomic method used, in fact, most of the reported resulting indices fell into the same two or three score or risk levels. This resulted in an unbalanced distribution of indices, with a greater number of them in the intermediate levels.

Paper	CV Tool	Environment	Methodology	Dataset
[39]	-	Laboratory	An approach in which researchers obtained view-invariant features from 2D pose and compute the ERA from it	Six people during training and one during validation, all of them involved in a task simulation in a lab environment
[40]	Proprietary	Laboratory	Training of a classifier that takes as input the 17 landmark locations produced by OpenPose and outputs an action level	Training and validation achieved with Human 3.6 dataset
[38]	OpenPose	Laboratory	Training of a classifier that takes as input the 17 landmark locations produced by OpenPose and outputs an action level	Training was achieved with a mixture of Human 3.6 dataset and videos of a weightlifting task
[41]	OpenPose	Real environment	Classic approach in which angles computed through 2D estimated pose are used to obtain an ERA	Five videos, chosen to test for occlusion (although it is moderate), multi-person viewpoint and back viewpoint
[30]	tf-pose-estimation	Laboratory	Classic approach in which angles computed through 2D estimated pose are used to obtain an ERA	Six people performing five static postures in a lab environment
[42]	Proprietary model	Real environment	Classic approach in which angles computed through 2D estimated pose are used to obtain an ERA	11 postures with people performing basic activities either indoor or outdoor
[43]	tf-pose-estimation	Real environment	Classic approach in which angles computed through 2D estimated pose are used to obtain an ERA	5 postures with people performing working tasks
[37]	OpenPose	Real environment	Proposal of an automatic ERA method selection based on activities present in the analysed task and subsequently ERA scoring	15 YouTube videos including six basic working activities
[44]	MediaPipe	Real environment	2D pose estimated through MediaPipe are send to a 3D Inverse Kinematic on Unity to predict a 3D pose. Ergonomic risk assessment is then obtained from the 3D pose	6 typical woodworking postures are reproduced in laboratory and filmed through an iPhone 11 Pro camera. The videos are shot so that they always include operators' faces
[38]	OpenPose	Real environment	Decision Tree trained to predict the ergonomic risk level in the analysed working posture	11 sewing machine operators' videos taken from two synchronised viewpoints with a smartphone

Table 1. Articles in Google Scholar from 2017 to 2023 proposing RGB MoCap systems. For each paper Table 1 details as follows: • Authors and publication year. • Employed Computer Vision tool. • Type of test environment. • Methodological approach. • Type of dataset used for tool validation.

Finally, some papers do not present statistical indicators or complete comparative results. For the sake of completeness, we have also included these studies, although we did not calculate any statistical indicators for them.

Tools benchmarking

Yan et al.[39] presented a camera view-invariant ERA approach using 2D poses, validated with video data from six individuals simulating construction tasks in a laboratory, achieving 87–89% accuracy (for both ERA scores and ERA risk levels) compared to IMU sensors.

Li et al.[40] proposed a CV-based system for RULA risk levels from 2D poses, using the Human 3.6 dataset[41,42] for validation, achieving 87/89% accuracy (for both ERA scores and ERA risk levels). Li et al.[36] improved this with OpenPose, assessing risk levels with 97% accuracy using static posture images from the Human 3.6 dataset and a laboratory-based weight-lifting task simulation.

Massiris Fernandez et al.[43] employed OpenPose for ERA in a real construction environment with five videos of workers, demonstrating perfect agreement via unweighted Cohen's kappa.

Agostinelli et al.[30] used tf-pose-estimation for RULA scores, validated on images of six individuals in five static poses taken in a laboratory, showing an 80% accuracy for both ERA Score and Risk Level.

Nayak et al.[44] used a proprietary human detection and recognition model for RULA-based ERA in real environments, involving 11 postures with people in indoor/outdoor activities, achieving a 45% accuracy for ERA scores and 64% accuracy for ERA risk levels.

Generosi et al.[45] focused on manufacturing environments with tf-pose-estimation, assessing five postures with people working, revealing 60% accuracy for ERA scores and 80% accuracy for risk levels.

Lin et al.[37] introduced a task-based ERA method selection using OpenPose, with a dataset comprising 15 YouTube videos of working activities in manufacturing, validating angle measurement accuracy against a Vicon Motion Capture system.

Jeong et al.[46] used MediaPipe and a 3D Inverse Kinematic model for ergonomic assessment in woodworking, using a dataset of six typical postures filmed through an iPhone in a real environment, achieving a 29% accuracy for ERA score and 86% accuracy for ERA risk level.

Lastly, Su et al.[38] developed a Decision Tree with OpenPose for ergonomic risk in sewing operations, using videos of 11 sewing machine operators from a real manufacturing environment, comparing manual and automatic REBA scores. Table 2 presents a more detailed description of the papers described above.

In most of the reviewed papers, the context of the use of the tool presents limitations. In cases where the images or videos used to validate the tool are acquired in the laboratory, the lighting and framing conditions of the operator are optimal (i.e., good illumination over the whole subject without overexposed or underexposed areas, operator fully framed and without major occlusions, as in[30,36]) and so they can differ a lot from the what could be found in a real manufacturing shop floor. In other cases, publicly available datasets such as Human 3.6 are used, but these do not depict people engaged in work tasks (e.g[36,40]).

Results of other studies, which referred to videos shot by the authors in real work environments (as in[44]) or to videos acquired from publicly available sources (e.g. YouTube in[37]), are very difficult to replicate as the context, in which the videos were taken, is rarely described with an in-depth level of detail. Moreover, in most of such

Paper	Context	ERA method	Measure	Reported Results	Computed Results (ERA Score)	Computed Results (ERA Risk Level)
39	Simulated dynamic postures typical of a construction environment	OWAS	Detection accuracy compared to ground truth data obtained from IMU sensors	Accuracy = 87–89%	*	Accuracy = 87–89%
40	Generic static postures from Human 3.6 dataset	RULA	Validation of the classifier is carried out on the Human 3.6 dataset through a comparison of the results obtained with the proposed tool and the ones obtained by computing ERA scores directly on 3D poses (data included in the HUMAN 3.6 dataset)	Confusion matrix on ERA Score (see the paper for detailed results)	Accuracy = 87–89%	Accuracy = 87–89%
36	Generic static postures from Human 3.6 dataset plus a video that was taped in a laboratory environment	RULA	No methodology validation was carried out. Only the classifier was validated through a k-fold cross-validation	*	*	Accuracy = 97%
41	Two outdoor construction videos taped by the authors and three publicly available outdoor working activities	RULA	Comparison between manual and automatic assessment through Cohen's kappa statistics	Unweighted Cohen's kappa = 1,00	*	*
30	Static postures from a laboratory environment	RULA	Comparison of RULA scores in terms of median and RMSE between ergonomist manual assessment and automatic assessment	Automatic and Manual ERA Score Results (see the paper for more details)	Accuracy = 80%	Accuracy = 80%
42	Full body (i.e., not occluded) videos of people involved in different indoor and outdoor working activities	RULA	Comparison between manual and automatic evaluation through one-way ANOVA and ICC	Automatic and Manual ERA Score Results (see the paper for more details)	Accuracy = 45%	Accuracy = 64%
43	Videos taped by the authors in a washing machine manufacturing environment	RULA	Comparison between manual and automatic evaluation through percentage differences	Automatic and Manual ERA Score Results (see the paper for more details)	Accuracy = 60%	Accuracy = 80%
37	15 YouTube videos of people working in several working environments	REBA, RULA, OWAS	The ERA tool is not validated. OpenPose is validated against Vicon for the accuracy of angle measurement	*	*	*
44	Videos taped by the authors of a person working in a woodworking environment	REBA	Posture-by-posture comparison between manual and automatic assessment	Automatic and Manual ERA Score Results (see the paper for more details)	Accuracy = 29%	Accuracy = 86%
38	Videos taped by the authors of 11 sewing machine operators	REBA	Comparison between manual and automatic assessment through average predictions. ICC between human ergonomists	ICC between 0,81 and 0,96. Average manual score 5,67 (SD = 1,6). Average automatic score 4,49	*	*

Table 2. Benchmarked tools presented in the assessed papers.

considered videos, there are lighting and occlusion situations similar to the laboratory context. So, the results of these studies cannot properly inform industrial decision-makers regarding the effective potential and limits of these solutions based on the features of their own production facilities.

In order to provide industries with useful information to better understand the level of reliability and accuracy that systems of this type can provide in the multifaceted manufacturing landscape, there is the need for further studies aiming to compare systems' performance in different application contexts (e.g., differing from each other in terms of the spatial layout of the considered production lines, in the number and arrangement of workstations, in the level of visual obstacles surrounding the workers, and the automation level).

However, the presented tools' source code is not generally publicly available. Even where the model architecture is described in detail and the trained model weights are provided (e.g[44]). , , performing a functioning implementation of the tool is difficult also for people with advanced deep learning skills. This lack of access to the tools limits the reproducibility of the experiments documented in the papers for further research efforts. This is why we adopted the tool described in[30,45] for the experimentation in this paper. Based on the results of the reported benchmarking, the tool demonstrated to perform with accuracy and reliability generally comparable with those of the other proposed systems (laboratory: ERA score accuracy = 80%, ERA risk levels accuracy = 80%; industrial manufacturing setting: ERA score accuracy = 60%, ERA risk levels accuracy = 80%). To the best of our knowledge, it is the only tool for which results are available from tests conducted both in a laboratory and in a real industrial manufacturing working environment (i.e., a washing machine assembly line). Moreover, the validation contexts are described in sufficient detail in the original papers.

The 2D RGB MoCap system used in the experimentation

The tool applied for the experimentation is the one selected from the benchmark discussed in the previous section, and it is the result of a long-term research work started in 2019, as documented in[30,45]. The current developed technology is useful to support ergonomists to easily and efficiently assess human factors in manufacturing, in particular, OCRA, REBA and RULA indexes. The tool, described in[45], exploits novel algorithms to predict the position of human body joint points and calculate angles between limbs. This system includes the implementation of the tf-pose-estimation tool (MCU model) to track human skeletal joints and Google Mediapipe's hand tracking model to enable the recognition of hand landmarks. The angles between body segments are assessed by firstly evaluating the predicted body orientation in each frame, with Eq. 1:

$$\alpha = \arcsin\left(\frac{x}{l \cdot 0.8}\right) \tag{1}$$

where α is the angle between the sagittal plane of the subject and the direction of the camera, x is the length of the segment related to the shoulder, and l the length of the segment related to the spine. Then, according to the predicted orientation, for each frame the coordinates $(x,\ y)$ of the respective keypoints are used to compute the angles between each body segment $i - j$ and $j - k$ using specific algorithms (Eq. 2). The formula for estimating the angle between these segments is based on the scalar product between the vector formed by the first segment $i - j$ and the one formed by the second segment $j - k$ (Eq. 3), and the cross product between the norms of these vectors (Eq. 4).

$$\theta\ = \left(\arccos\left(\frac{\gamma}{\delta}\right)\right) \times\ \left(\frac{180}{\pi}\right) \tag{2}$$

$$\gamma\ = (x_j - x_i) \times\ (x_k - x_j) + (y_j - y_i) \times\ (y_k - y_j) \tag{3}$$

$$\delta\ = \left(\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}\right) \times\ \left(\sqrt{(x_k - x_j)^2 + (y_k - y_j)^2}\right) \tag{4}$$

Mediapipe has been mainly used to calculate wrist angles. This is achieved by determining the angle between the segments formed by the elbow and wrist coordinates (obtained from tf-pose-estimation) and those between the wrist and middle finger MCP (obtained from Mediapipe). From these angles, the software can calculate the ergonomic indices RULA, OCRA, and REBA semi-automatically through three specialised modules. The automation degree is driven by how much the index is based on angles rather than external factors that are difficult to calculate by video analysis (e.g., the weight of objects handled by an operator). A web-based GUI (Graphical User Interface) facilitates the ergonomist's manual intervention. Generosi et al.[45] describes in detail the GUI, the underlying index calculation algorithms, and the tool's modular architecture. The tool requires at least one video recording of the operator to perform the joint's angle extraction. The relative positioning of the camera and operator is not binding, although the lateral framing allows the system to obtain the best results in terms of reliability of ergonomic risk assessment, given that most of the angles considered by the implemented protocols are extracted from the side of the human figure. If the working environment layout permits it, it is possible to acquire multiple viewpoints of the worker (e.g., a lateral and a frontal/rear one) and feed them to the software for improved accuracy. The ergonomist could record the operators from one angle and then record them from another using the same camera used before. However, by recording the videos at two separate times there is a possibility of running into behavioural bias of the framed operators, who might behave differently between the first and second takes, undermining the goodness of the ergonomic assessments. Moreover, in the case of two videos recorded at different moments, editing them to synchronise the beginning of the recorded working cycle between the two (or more) videos is necessary. The ergonomist can choose to edit the videos with an external video editing software or use the integrated web interface, which provides a Graphical User Interface (GUI) with basic video editing tools and an interface to set up the parameters for the ergonomic risk assessments. Figure 1 shows the system's architecture. Once uploaded the videos, the Motion Analysis System module analyses them, synchronizing if they were not natively synchronized, collecting joint position data from tf-pose-estimation and the MediaPipe hand tracking module and then proceeding to compute the angles from these data. After extracting the joint angles, the next module (Human Work Analysis) computes the required ergonomic score (e.g., RULA, REBA, OCRA) based on how the ergonomist sets up the analysis. Finally, the tool integrates a database for storing and retrieving raw angles data and previous ergonomic risk assessments.

Camera synchronization

The Camera Synchronization module is a critical component designed to ensure accurate and reliable ergonomic assessments by aligning multiple video streams when they have not been natively synchronised during the recording. This module is essential for aligning video recordings captured from different angles or at different times, ensuring that the analysis of human motion is based on synchronized data. Following the approach presented in[47], we calculate the vertical speed of joint keypoints from each camera and determine the optimal time offset where the correlation between these speeds is highest. Once the optimal offset is determined, the script adjusts the video frames, either trimming or padding them to achieve synchronization. If the initial synchronization results are not satisfactory, the script can reset the offsets and allow for re-synchronization with adjusted parameters. This allows for the alignment of frames across different videos, ensuring that movements captured from various viewpoints are temporally consistent. The script processes directly the tf-pose-estimation keypoints data from each camera view, referring to a configuration file detailing environment setting, e.g. the camera acquisition frame rate. After the videos are synchronized, they will feed again the tf-pose-estimation and Mediapipe modules for the other steps.

Industrial case studies

This section presents different industrial case studies conducted to evaluate the performance and practicality of 2D RGB Motion Capture systems in manufacturing environments. For this purpose, the tool proposed in Sect. 3 has been employed as a benchmarking instrument, given that it has demonstrated comparable accuracy to that of other systems evaluated in the relevant literature. For this study we ensured adherence to all ethical guidelines and regulatory requirements for experiments involving human subjects. The study was conducted according to the guidelines of the.

Declaration of Helsinki and approved by the Ethics Committee of Università Politecnica delle Marche (Prot.n. 0100472). Informed consent was obtained from all participants prior to participation.

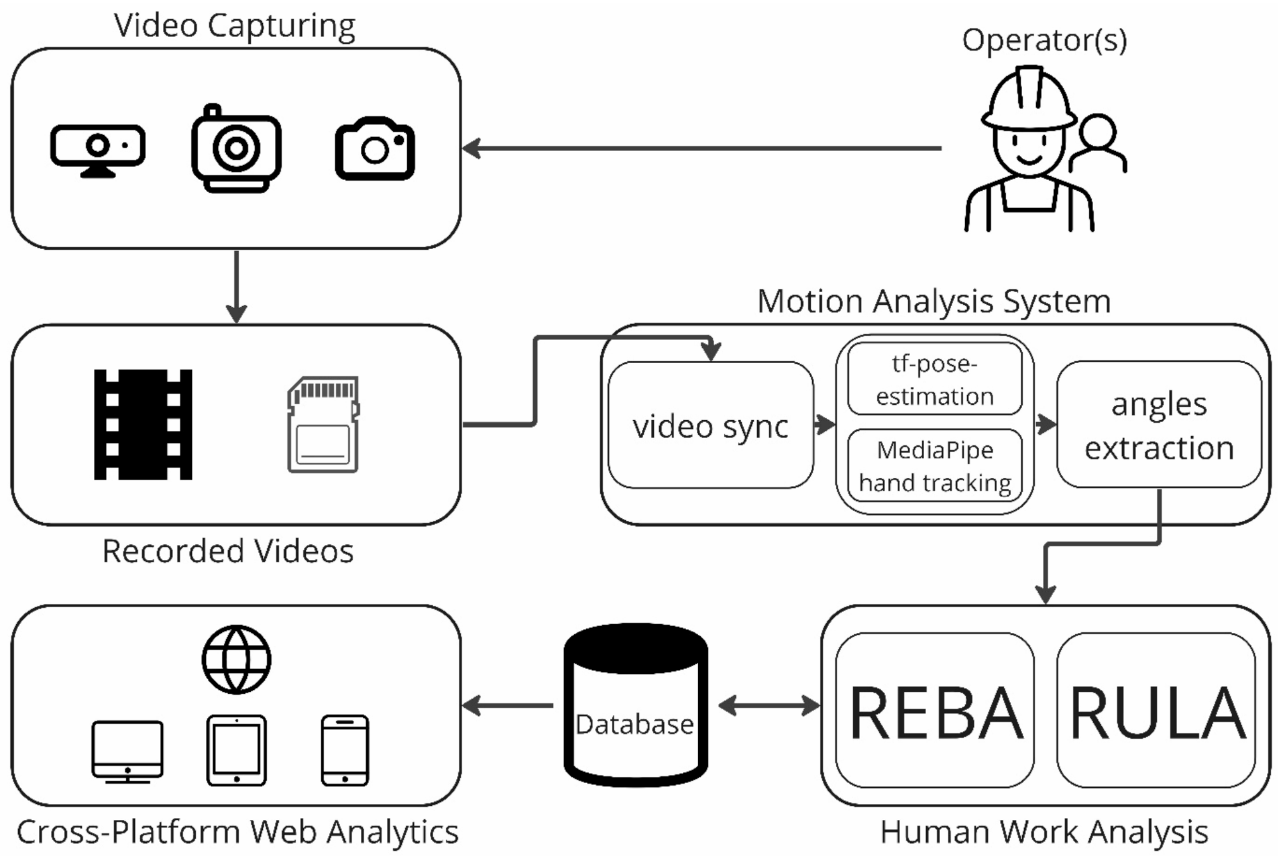

Fig. 1. High level architecture of the tool exploited for the case studies.

The context

The examined contexts belonged to a manufacturing company producing kitchen extractor hoods. They differed in terms of production lines' layout, the level of visual obstacles surrounding the workers, and the level of automation.

Line L1

Line 1 (L1) was an assembly line for a finished product with a linear layout and a medium-low level of automation. The conveyor belt's speed allowed operators to handle it, not dictating the timing of the production line. The output was 1 piece per minute, with five workstations (visible in Fig. 2):

- L1W1: the operator manually picked semi-assembled parts from a cart that came from another production line. She installed various components on these parts, also manually picked from boxes arranged around and behind her.
- L1W2: the operator took an expanded polystyrene tray and placed it on a roller table which is mounted on a mobile cart. He positioned the assembly coming from the previous station onto the tray along with other components to be later installed. Finally, he pushed the tray with its contents off the roller table onto a conveyor belt placed at the same height.
- L1W3: the operator awaited the arrival of the tray, picks up the components on the tray, and mounts them onto the semi-assembled part to produce the final product. For this task, she used a suspended pneumatic screwdriver placed in front of her.
- L1W4: the operator waited for the finished product and then cleaned some glass components with a cloth. She then connected the product's electrical plug to a test bench to check the operation of the spotlights which were mounted on the assembly.
- L1W5: the operator manually lifted the finished product (weighing 6.8 kg) from the conveyor belt and placed it inside a cardboard packaging placed on the floor behind him.

Positioning the cameras to monitor workers along this line, capturing at least two shots for each operator (i.e., a rear view and a side view) was quite difficult, as shown in the layout in Fig. 3. At all stations (except L1W4), there was a rack or a box stack behind the workers. As a result, only the worker at station L1W4 had an optimal rear view: the rear views of workers in L1W1 and L1W3 could only be taken obliquely, while the height of the rack and the boxes stack partially occluded the worker in L1W2 and L1W5. Regarding the workers' side views, the side cameras could be optimally positioned only in L1W3 and L1W4 workstations. In L1W5 workstation,

Fig. 2. Line L1 workstations.

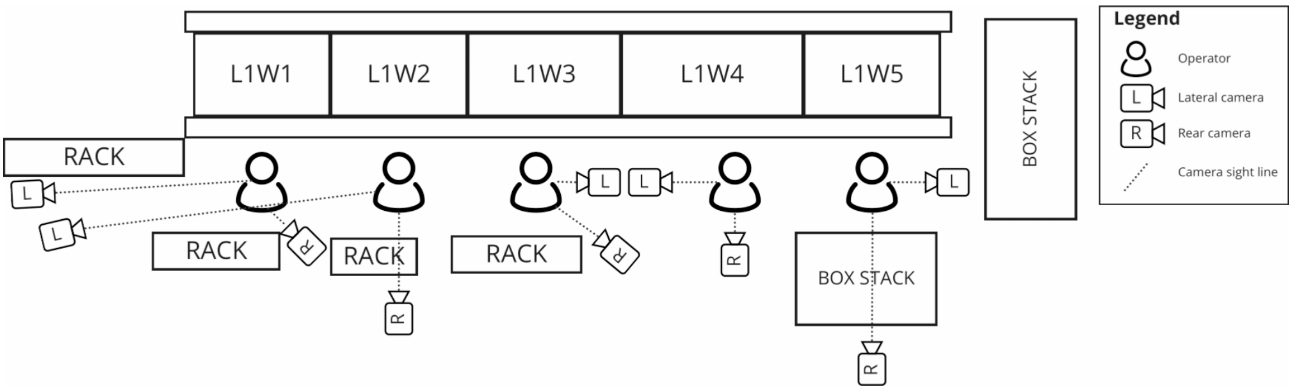

Fig. 3. Line L1 layout.

a tall boxes stack placed to the right limited the available space for camera positioning. The result is that only a very close side shot was possible, so the subject was often only partially filmed. Due to spatial organization, lateral camera views for both L1W1 and L1W2 workstations were shot obliquely. Moreover, the lateral view for the worker in L1W2 resulted partially occluded by the operator working in L1W1 workstation.

Line L2

Line 2 (L2) consists of a single bending press (visible in Fig. 4). There was no automation, and the operator moved around a lot, often going in and out of the frame. Additionally, the positioning of the bending machine close to a very powerful artificial light source made it difficult to position the cameras to ensure they were not blinded by the light. The production output was about 0.3 pieces per minute (one piece every 2 min and 50 s).

In L2W1 the operator manually picked up, from the rack behind him, a flat perforated steel sheet weighing about 2.5 kg. Through multiple passes on the various dies installed on the bending press he deformed the sheet metal to create a metallic box. Upon completion of the process, he placed the box on the rack situated to left of the operator. The operator supported the weight of the sheet throughout the entire work cycle and operated the press via a pedal switch. The spatial organization of the work area (visible in Fig. 5) limited the possibility of taking side shots of the operator: cameras could not be placed on the right side because there was a forklift-reserved lane that had to be kept obstacles-free. On the left side, the positioning of the camera was hindered by the presence of a rack, so that the side view of the worker could only be taken obliquely. Instead, there was enough space between the worker and the rack behind him to easily position the camera for an optimal rear view of the operator.

Line L3

Line 3 (L3) was an assembly line for a finished product, consisting of two parallel sub-lines: the main line and the supply line, both with a linear layout. The automation level was medium, featuring a conveyor belt similar to Line 1, with additional conveyor belts delivering components from the supply line to the main line, and a rail-guided manipulator to assist operators in moving heavy assemblies. The output was 0.75 pieces per minute (one

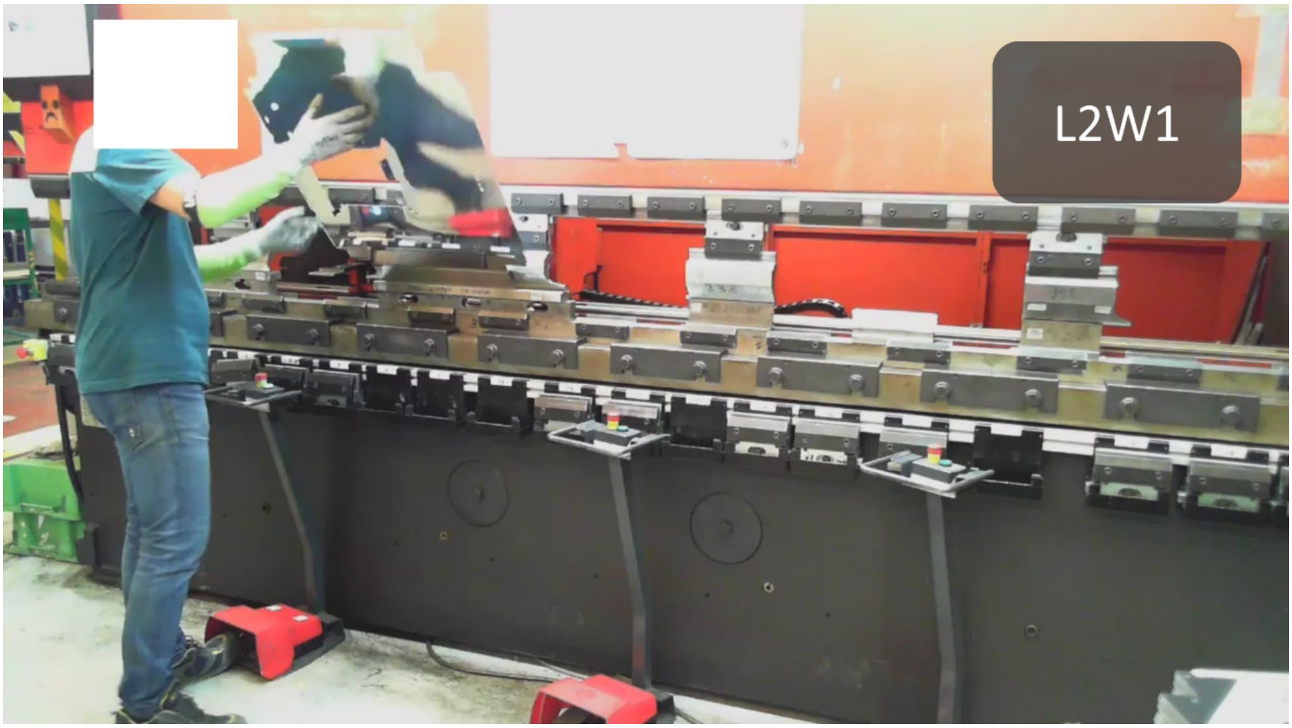

Fig. 4. Line L2 workstation.

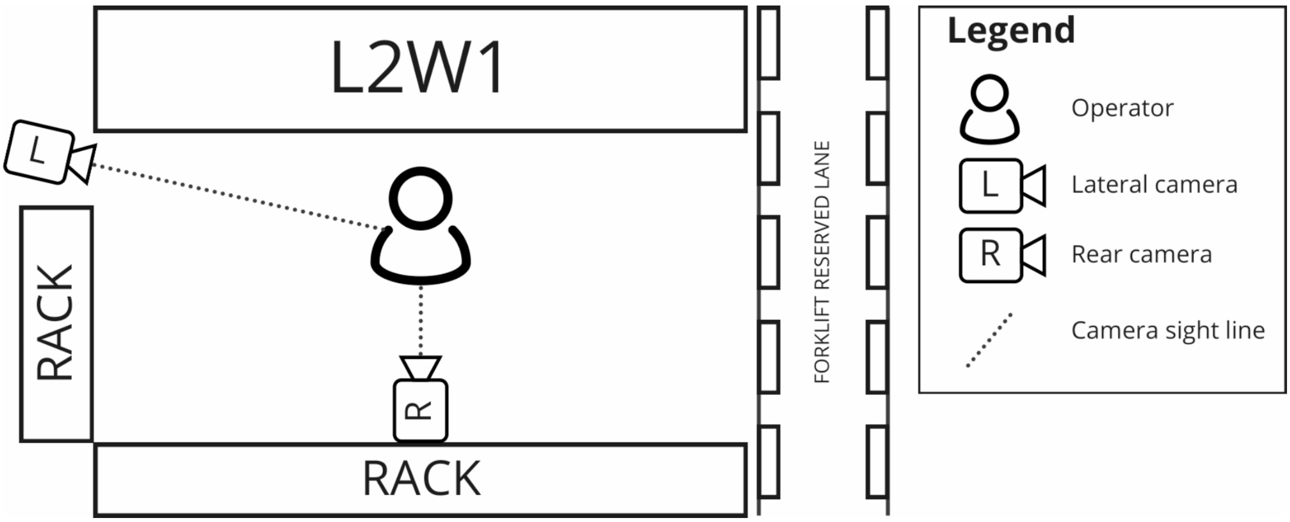

Fig. 5. Line L2 layout.

piece every 1 min and 20 s). The main line had five workstations, while the supply line had two workstations (visible in Fig. 6):

- L3W1: the operator received, from the supply line, a polystyrene tray loaded with sub-assemblies necessary for the finished product. She assembled these components using a suspended pneumatic screwdriver placed in front of her.
- L3W2 and L3W3: at both stations, operators awaited the arrival of the output from the previous workstation and installed additional components on it, using a suspended pneumatic screwdriver. The necessary components were delivered via conveyor belts from the supply line.
- L3W4: the operator used a rail-guided manipulator to pick up the finished product coming from the previous workstation and placed it on a measurement bench for testing purposes. The electrical connection between the product and the measurement bench, as well as the testing process, occurred automatically.

Fig. 6. Line L3 workstations.

- L3W5: the operator used another rail-guided manipulator to pick up the tested product and placed it inside a cardboard packaging that was supplied by the supply line.
- L3W6: the first workstation of the supply line. The operator manually picked cardboard boxes from a box stack, he opened them, and then arranged cardboard dividers inside of the boxes. Finally, he placed the boxes on a roller conveyor that supplied L3W5.
- L3W7: the operator manually picked the components that are necessary for workstations L3W1 to L3W3 from various carts placed behind him, and then placed them on respective roller tables, ensuring that the main line is always stocked.

Given the layout of the production line (visible in Fig. 7), placing cameras to monitor the worker at the L3W2 workstation would have affected the normal production workflow, so that no shots were taken for the L3W2 workstation. The presence of obstacles and the proximity of some of the workers from each other limited the possibility to capture adequate rear and side shots for all the operators. The side view of the operator working at the L3W1 station was limited by the presence of a rack on the operator's left side and by another operator on the right, so that only a highly obliquely side shot from the left was possible. The high proximity of L3W2 and L3W3 operators made a side shot of both operators unfeasible, as there was no space to adequately place a side camera. The rack behind the first three operators partially occluded the first and the third operators, making only an oblique rear shot feasible, while it completely occluded the second operator. Finally, L3W6 and L3W7 workstations' operators were surrounded by boxes stacks that partially occluded them, except for the L3W7 operator's right side. Cameras were placed to avoid as many occlusions from the boxes as possible.

Data collection

For each production line, we carried out the monitoring using RGB action cameras (AKASO Brave 7 LE with a 20 MPx sensor) which were mounted on tripods. The cameras were strategically positioned behind and to the side of the monitored operators, whenever possible. Camera placement was determined based on the layout of the production lines and the presence of any obstructing obstacles. No alterations were made to the layout to ensure that the cameras were optimally positioned, preserving the normal workflow of the operators. The monitoring spanned a single work shift, capturing footage of one operator at each workstation. In total, 13 operators were engaged during the analysed work shift across the three production lines, but one operator was excluded from the experiment due to the inability to position cameras at that workstation, attributable to the high number of obstacles. We, therefore, analysed 12 operators, 6 male, and 6 females, with an average age of 35.54 years.

For each workstation, we selected the video clip depicting the work cycle with the highest-quality footage, based on considerations such as minimizing unrelated individuals passing through the frame and ensuring the operator remained fully visible for most of the time.

Ergonomic evaluation procedure

The ergonomic evaluation procedure consisted of the following steps, detailed below:

- Manual selection of the most hazardous postures from the video footage.
- Manual extraction of the angles from the selected postures.
- Automatic extraction of the angles from the entire video footage, selecting only the angles from the frames matching those containing the selected postures.
- REBA and RULA analysis of both the manual and automatic angles.

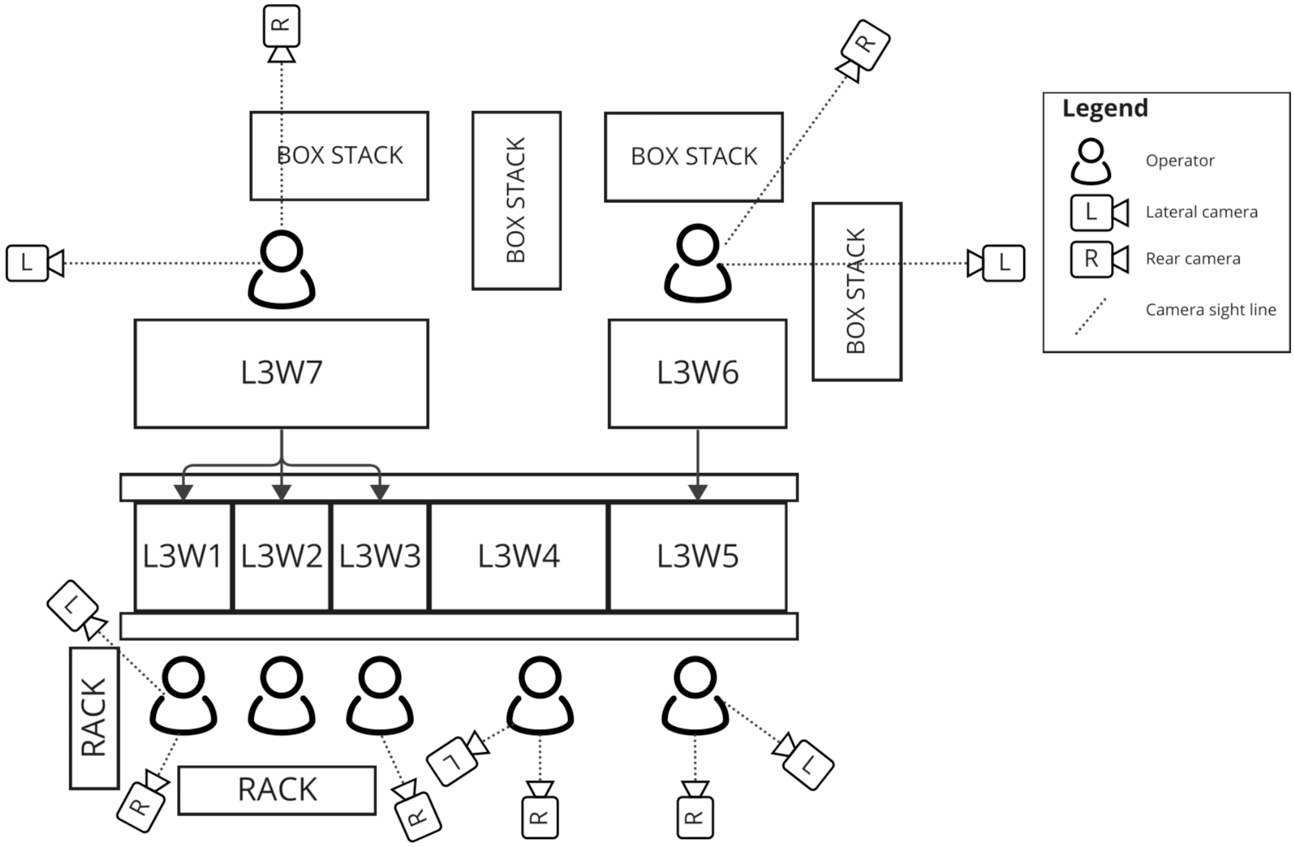

Fig. 7. Line L3 layout.

In the manual analysis, a panel of four expert ergonomists identified the most hazardous postures for each video or multisided video clip. This resulted in the selection of 23 frames for the first line (L1), 5 for the second line (L2), and 22 for the third line (L3), totalizing 50 video frames. For each frame, the ergonomists independently extracted the angles of the postures manually, resulting in four joint angles for each posture joint. The median of these four angles was calculated to obtain a single angle for each posture joint.

For automatic angle extraction, the Motion Analysis module extracted angles throughout the entire work cycle for each assessed workstation, providing results for each frame. Only the angles from the frames selected by the ergonomists were retained for comparison between manual and automatic results. The collected angles are visible in Fig. 8, while Table 3 highlights which angles have been extracted from which point of view and indicates on which anatomical plane these angles were projected.

The ergonomic risk assessment was then carried out using both REBA and RULA methodologies for both the manual and automatic angles. REBA and RULA assign scores to each analyzed body segment according to the joint's angle. Thresholds are applied to these angles to classify the ergonomic risk level associated with each posture. Specifically, for each joint angle, predefined thresholds are used to determine whether the angle represents a low, medium, or high risk for musculoskeletal disorders.

For example, in the RULA protocol, a shoulder flexion exceeding 45 degrees is classified as a moderate risk, and a score of 2 is assigned, while a flexion angle greater than 90 degrees is classified as a high risk, resulting in a score of 3[12]. Similarly, for the REBA protocol, trunk flexion between 20 and 60 degrees corresponds to a moderate risk level, yielding a score of 3, whereas trunk flexion exceeding 60 degrees is considered a high risk, with a score of 4[13]. These thresholds help in quantifying the ergonomic risk based on the extent of deviation from a neutral posture, thereby allowing both manual and MoCap-derived measurements to be consistently evaluated.

Individual Scores are adjusted by adding modifiers that account for intense efforts, the nature of motion or activity, frequency of repetition, and hand-tool coupling quality (in REBA's case) or intense efforts and muscle utilization frequency (in RULA's case). These adjusted Scores are then integrated using reference tables to derive a final score indicating the MSD Risk Level for the evaluated task. REBA's output is the REBA Grandscore, ranging from 1 (minimum) to 15 (maximum). Similarly, RULA's output is the RULA Grandscore, ranging from 1 (minimum) to 7 (maximum).

For consistency of evaluation, the REBA and RULA analyses of the selected postures were performed by the Human Work Analysis system's module for both automatic and manual angles. In addition, to perform the REBA and RULA analyses for the automatic angles, we manually set the parameters that are not automatically acquired by the system, thus the angle extraction is fully automatic but the ergonomic risk assessment in its

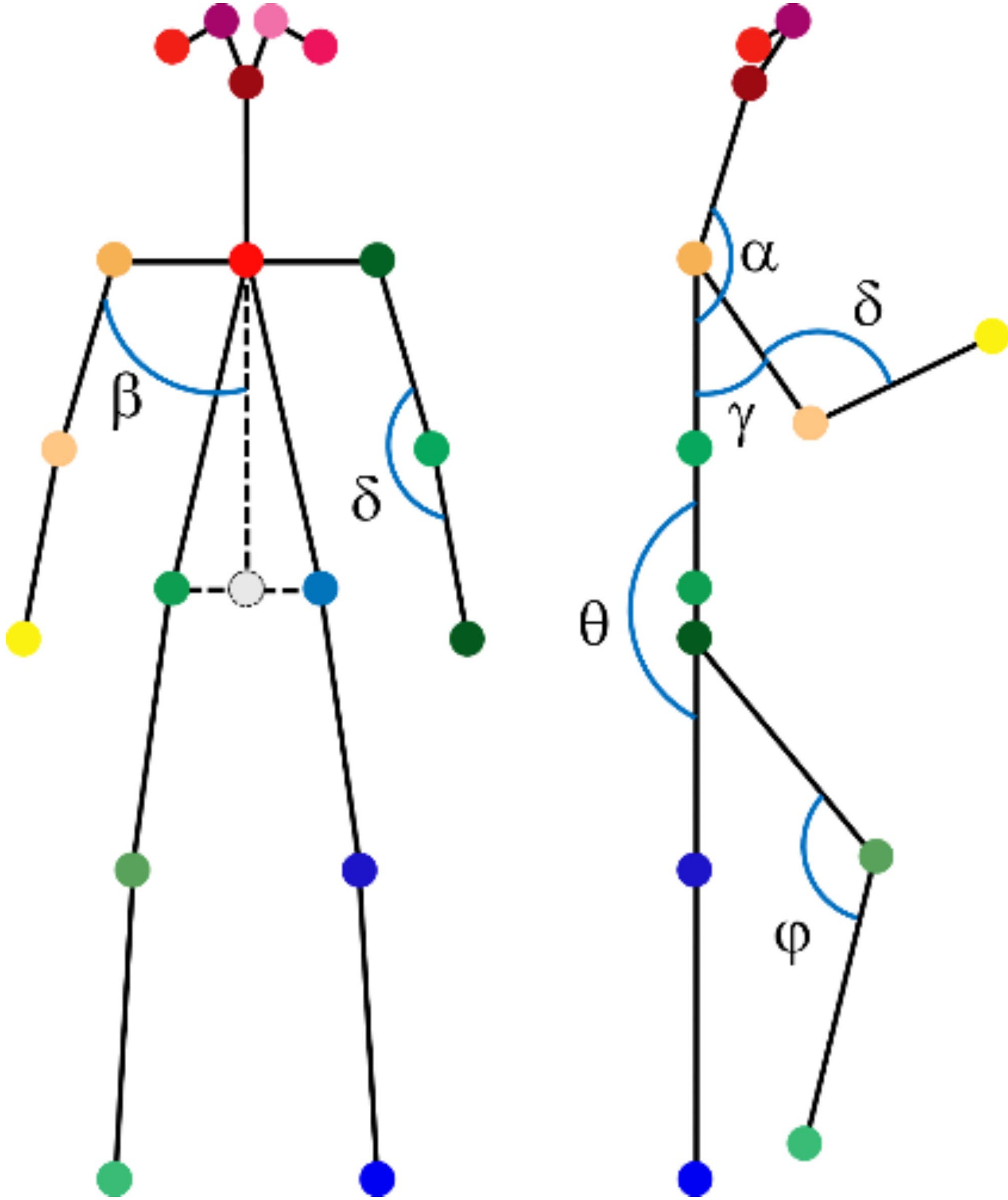

Fig. 8. Angles extracted from the skeleton.

entirety should be considered semi-automatic. The values were taken from the manual analyses carried out by the ergonomists. To evaluate the performances obtained in REBA and RULA semi-automatic applications for each workstation, we calculated both Accuracy and RMSE to compare results obtained with the semi-automatic assessment and results obtained with manual assessment. In the first case, Accuracy has been calculated as the percentage value of correct assessments for both Scores and Risk Levels, while RMSE has been calculated through Eq. 5:

Angle	Rear camera	Lateral Camera	Anatomical Plane
Neck flexion/extension (α)		X	Sagittal (longitudinal)
Shoulder abduction (β)	X		Coronal (frontal)
Shoulder flexion/extension (γ)		X	Sagittal (longitudinal)
Elbow flexion/extension (δ)	X	X	Sagittal (longitudinal)/Coronal (frontal)
Trunk flexion/extension (θ)		X	Sagittal (longitudinal)
Knee bending angle (ϕ)		X	Sagittal (longitudinal)

Table 3. Cameras from which the angles were measured and anatomical planes on which such angles are projected.

Score	Median	Standard Deviation
Semi-auto Grandscore	3	1.35
Manual Grandscore	5	1.39
Semi-auto Risk	2	0.61
Manual Risk	3	0.79

Table 4. Descriptive statistics for the REBA analysis.

Score	Median	Standard Deviation
Semi-auto Grandscore	6	1.27
Manual Grandscore	7	1.73
Semi-auto Risk	3	0.42
Manual Risk	3	0.60

Table 5. Descriptive statistics for the RULA analysis.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad (5)$$

where P_i is i-th the predicted value (semi-automatic assessment), O_i is the i-th observed value (manual assessment) and n is the total number of observations.

Results

In this section, we present the results of our experimentation. Subsections 5.1, 5.2, and 5.3 detail the REBA and RULA analyses for production lines L1, L2, and L3, respectively, in terms of Accuracy, and RMSE. Subsection 5.4 will present the results of the comparative analysis between the three production lines involved in this paper and the papers previously identified in the literature, in terms of ERA Score and Risk Level accuracy. Tables 4 and 5 show the descriptive statistics (median and standard deviation) for the REBA and RULA analysis respectively.

Line L1

For the L1 line, the system provided a correct Risk Level estimate for 15 out of 23 postures (65.22%) using both REBA and RULA analysis, while the Score evaluation accuracy reached only the 33.78% for REBA and 30.43% for RULA (Table).

Figure 9 shows that, in the case of workstation L1W3, the system computed the correct risk index for all considered postures (REBA Risk Level RMSE = 0.00; RULA Risk Level RMSE = 0.00). On the other hand, workstation L1W2 is the one for which the system was the least accurate (REBA Risk Level RMSE = 0.77; RULA Risk Level RMSE = 1.61).

In L1W2 the rear camera was partially occluded by the rack behind the operator, who moved around a lot within the workspace during the task, leading to the worst results of the line. This may have resulted in the system underestimating the analysis for both the REBA and RULA analyses. In the case of the L1W4 workstation, the operator was not always fully framed by the side camera, as she was often too close to the camera due to the many movements she made in the workstation to perform the tasks. In the case of the L1W5 workstation, the results are also not sufficiently accurate. This may be due to the operator often standing too close to the side camera, while the framing of the rear camera was rather angled to the transverse plane (i.e., the anatomical plane that divides the body into superior and inferior sections), in order to avoid occlusions from the rack behind the operator.

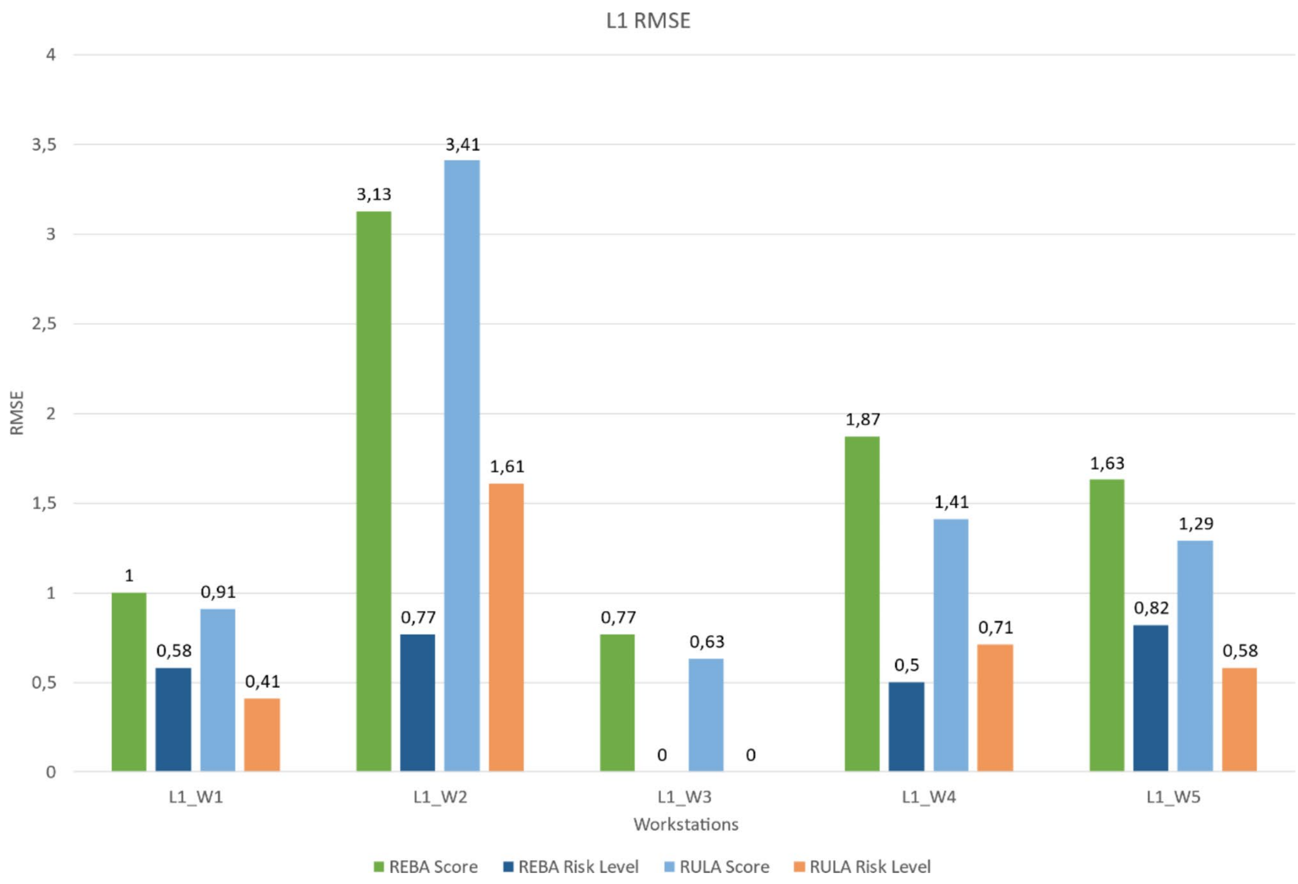

Fig. 9. Line L1 REBA and RULA Score and Risk Level RMSE.

Line L2

As showed in Table, in the case of the L2 line, the system provided a correct Risk Level estimate for 4 out of 5 observed postures (80.00%) with the REBA analysis, whereas 2 out of 5 postures (40.00%) were correctly predicted with the RULA analysis.

Figure 10 shows that the estimates of the Risk Levels are not completely correct using either the REBA or the RULA analysis (REBA Risk Level RMSE = 0.89; RULA Risk Level RMSE = 0.77).

This may be due to the fact that the side camera, which was already sub-optimally placed as it was angled in relation to the sagittal plane of the operator (i.e., the anatomical plane that divides the body into right and left sections), was often partially occluded by the rack. In addition, the side camera suffered from glare from artificial lighting and the resulting image was often overexposed. This may have compromised the system's ability to correctly predict angles.

Line L3

As it can be seen in Table 6, in the case of the L3 line, the system gave 16 correct Risk Level predictions out of 22 (72.73%) using the REBA analysis and 12 correct predictions out of 22 (54.55%) using the RULA analysis. Scores Accuracy values show instead significantly lower values.

As shown in Fig. 11, in the case of workstations L3W3 and L3W6, the system provided the correct risk index for all considered postures (REBA Risk Level RMSE = 0.00; RULA Risk Level RMSE = 0.00). On the other hand, workstation L3W5 proved to be the one for which the system provided the least accurate results (REBA Risk Level RMSE = 1.00; RULA Risk Level RMSE = 1.20).

The risk indices estimated by RULA analysis for workstations L3W1 and L3W4 (RULA Risk Level RMSE = 0.71 for both) and those estimated by both RULA and REBA analysis for workstation L3W7 (REBA Risk Level RMSE = 0.71; RULA Risk Level RMSE = 0.71) are also not completely correct. Regarding the risk index estimates, the worst results provided by the system can be observed in relation to workstations L3W5 and L3W7.

For the L3W5 workstation, the low system's accuracy might be due to the frequent occlusions of the operator's left side - particularly the arms - as the operator often put his arms inside the boxes he was setting up. Also, for the L3W7 workstation, the operator's shot was often occluded. In addition, the operator's lower body was only partially framed by the footage.

As a result, the system was not always able to correctly predict the segments corresponding to the upper limbs: although the postures adopted by the operator were often asymmetrical, the system often interpreted

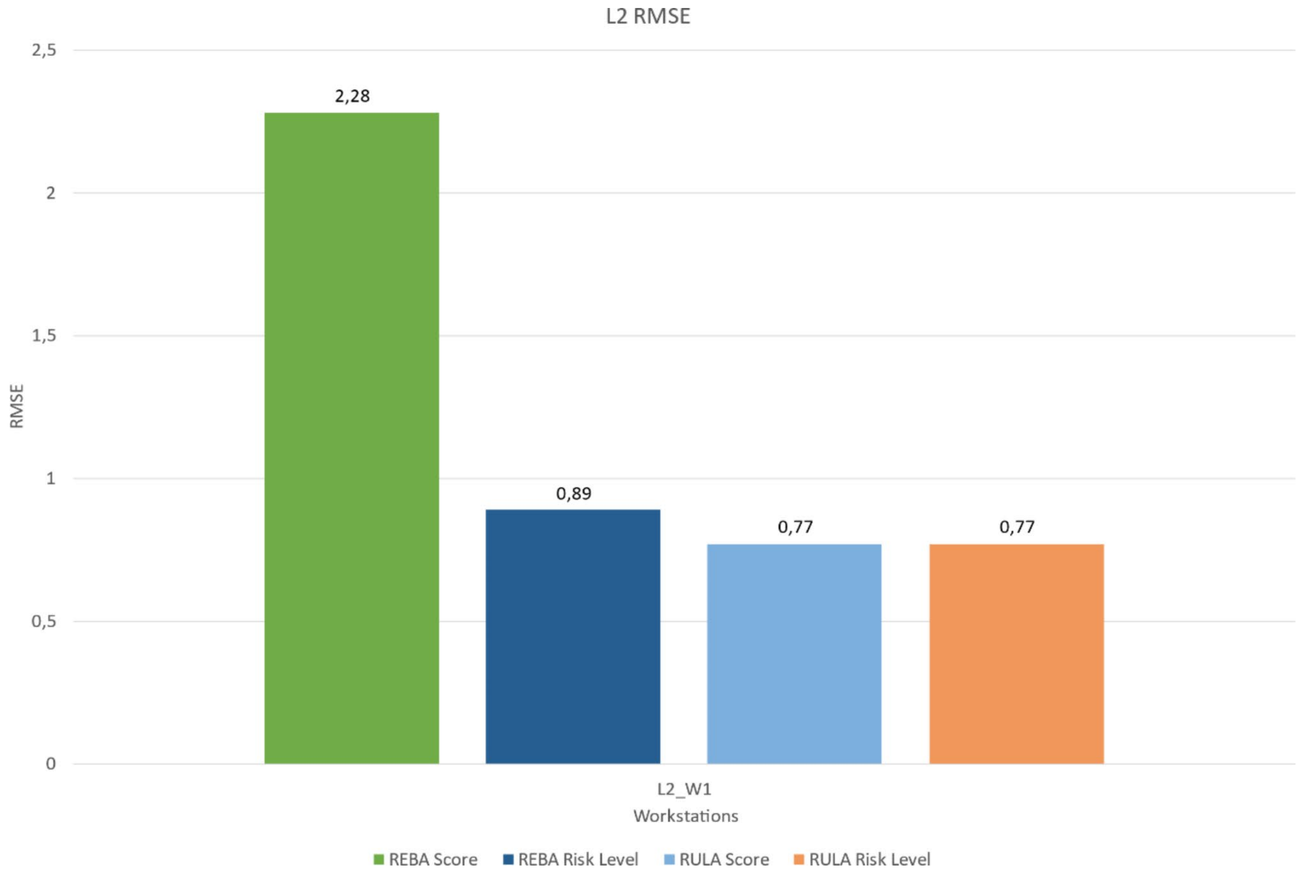

Fig. 10. Line L2 REBA and RULA Score and Risk Level RMSE.

Line	ERA Score Accuracy	ERA Risk Level Accuracy
L1 REBA	34.78%	65.22%
L1 RULA	30.43%	65.22%
L2 REBA	60.00%	80.00%
L2 RULA	40.00%	40.00%
L3 RULA	9.09%	72.73%
L3 REBA	22.73%	54.55%

Table 6. ERA score and risk level accuracy for every line.

them as symmetrical or vice versa. Similarly, the predictions of the angles associated with the lower segments are not always correct, and consequently errors are also observed in the estimates of some postures.

Approaches comparison

Figure 12 illustrates a comparison of the performance of three ERA tools from Nayak & Kim, Generosi et al., and Jeong & Kook, alongside the ERA system proposed in this paper, similar systems implemented in the three case studies (L1, L2, and L3) using both REBA and RULA analyses. The left side of the graph focuses on the Accuracy of these systems in calculating ergonomic scores, while the right side presents the accuracy in determining ergonomic Risk Levels. We only included Nayak & Kim, Generosi et al. and Jeong & Kook because (i) the experiments they report are conducted in a real environment like the ones reported in this paper and (ii) either they reported their Accuracy values, or we were able to calculate them. From the left section of the graph (ERA Score Accuracy), it is evident that the systems evaluated vary significantly in performance. Nayak & Kim's system shows an Accuracy of 45%, Generosi et al. reports a 60.00% Accuracy, while Jeong & Kook features a lower 29.00% Accuracy. When applied to the case studies, the L1 REBA analysis achieves an Accuracy of 34.78%, while the L1 RULA registers slightly lower at 30.43%. The accuracy increases notably in L2 REBA at 60.00%, with L2 RULA also slightly increases at 40.00%. Finally, the L3 REBA achieves the lowest Accuracy at 9.09% while L3 RULA achieves an Accuracy of 22.73%.

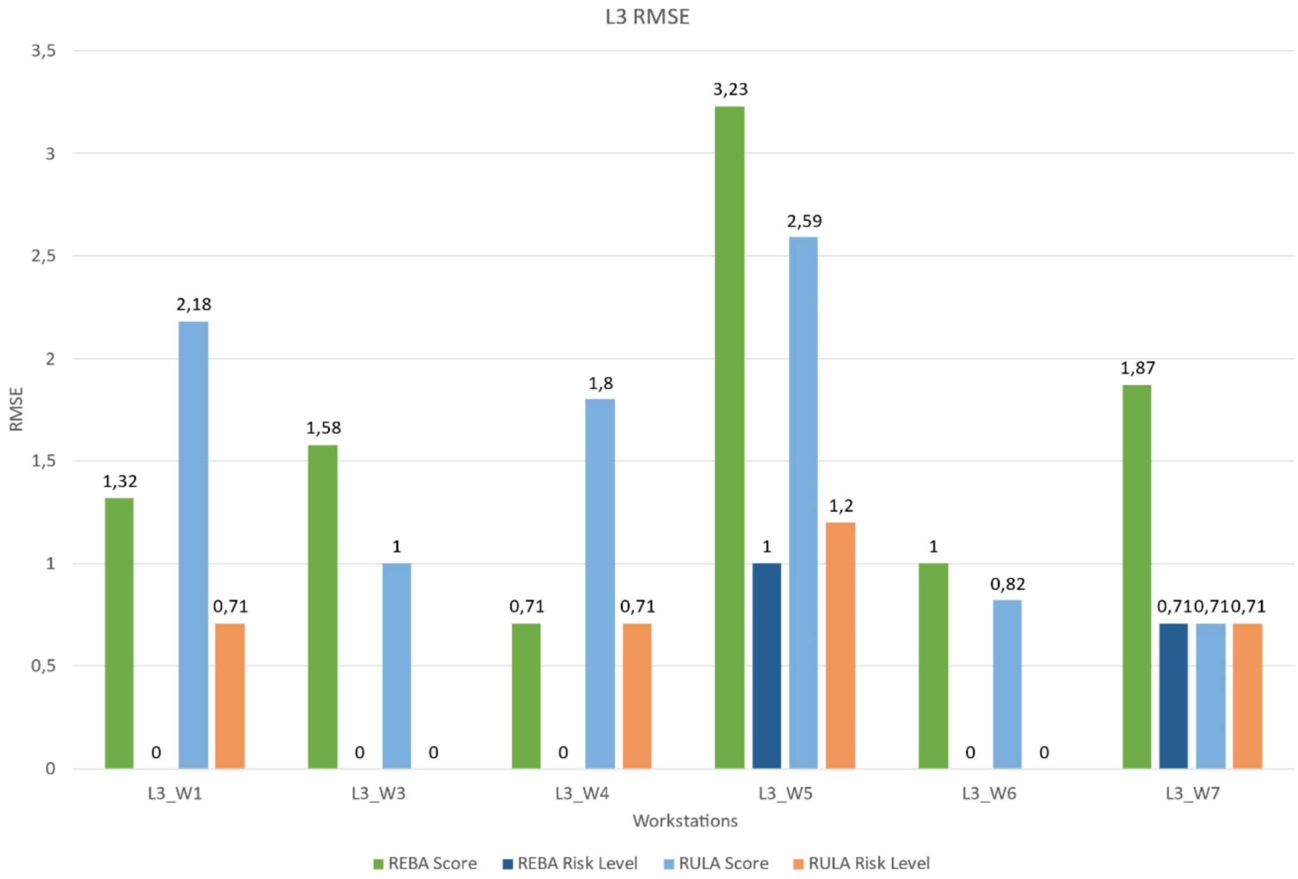

Fig. 11. Line L3 REBA and RULA Score and Risk Level RMSE.

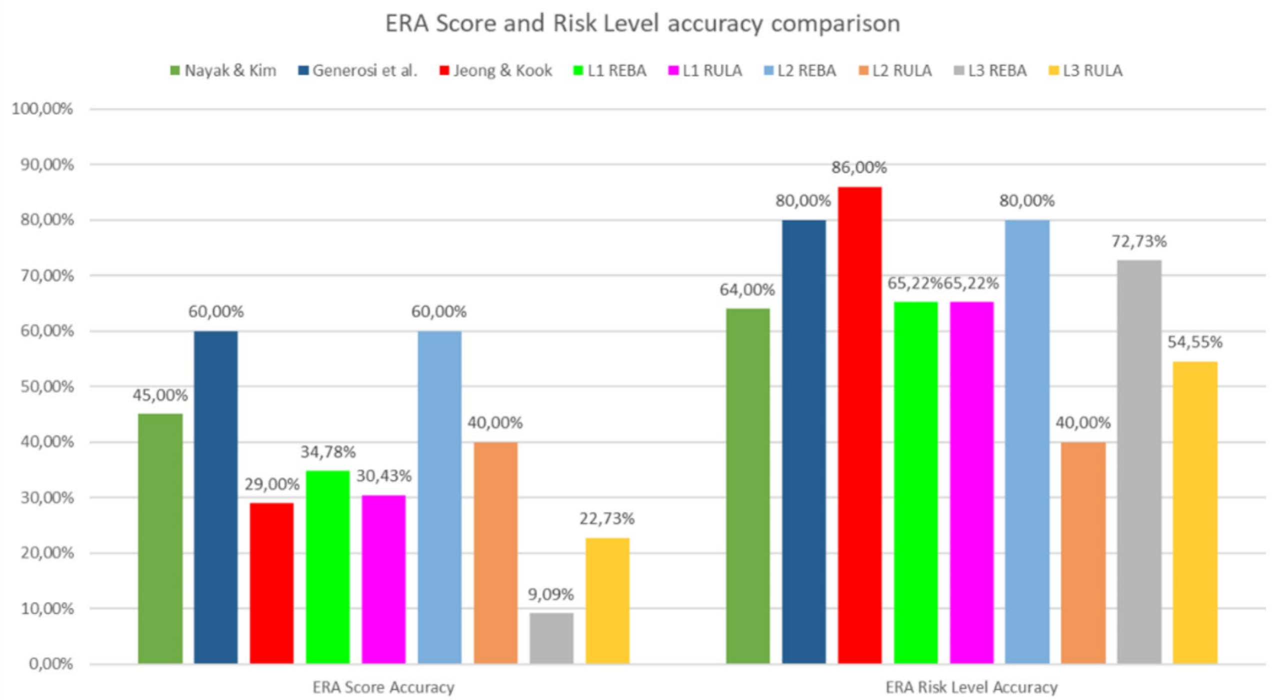

Fig. 12. ERA Score and Risk Level accuracy comparison.

In terms of ERA Risk Level accuracy (right side of the graph) both the L1 REBA and L1 RULA demonstrates an Accuracy of 65.22%. The L2 REBA shows 80.00% accuracy, while L2 RULA scores a lower Accuracy, at 40.00%. Finally, L3 REBA achieves an Accuracy of 72.73%, while L3 RULA records an accuracy of 54.55%. Looking at the systems found in literature, we notice that Nayak & Kim's system shows an accuracy of 64.00%, while both Generosi et al. and Jeong & Kook scored a higher accuracy of, respectively, 80.00% and 86.00%.

Discussions

Performance and limitations of 2D RGB MoCap systems

The reported results highlight two aspects of the tested RGB MoCap system in industrial manufacturing environments. On one hand, the system's ability to accurately predict the ERA Score does not appear to be particularly high: in three cases, the RMSE value exceeds 2 (L2W1 REBA, L3W1 and L3W5 RULA), while in another three cases (L1W2 REBA, L1W2 RULA and L3W5 REBA) the RMSE exceeds 3. In all the other cases, the RMSE exceeds 1, and only in six out of 24 cases the RMSE value for the ERA Score is lower than 1. Nevertheless, in most cases the system has a Risk Level RMSE of less than 1 or even equal to 0, except for L1W2 and L3W5 for the RULA method and L3W5 for the REBA method.

The other aspect that emerges from the results is the high variability of the tool performances due to the external conditions (e.g., spatial layout, automation level, workstations' proximity level, positioning of picking racks in relation to the operator, operator's mobility level during tasks, employed ergonomic method): the system's Accuracy in predicting the ERA Scores ranges from 9.09% (L3 REBA) to 60.00% (L2 REBA), while the Accuracy corresponding to the ERA Risk Level fluctuates between 40.00% (L2 RULA) and 80.00% (L2 REBA).

Although it cannot be definitively asserted that the other 2D RGB MoCap systems proposed in literature would exhibit the same variability if subjected to the same conditions proposed by this experimentation, it has been nevertheless shown that the performances, both in terms of ERA Score and ERA Risk Level, are comparable. This is supported by the fact that these tools are mostly based on the same Deep Learning models (OpenPose CMU and Mediapipe BlazePose) for the body landmarks detection.

Therefore, the system appears sufficiently accurate to provide a rough indication of whether a more detailed analysis of the workstation is required.

Challenges and pathways for enhancement

The discussed limitation could be overcome by providing fixed camera locations integrated along the lines at the time of line design. This would allow more than two cameras to be integrated at each location, thus potentially improving system's accuracy. More studies are needed to assess how possible variables (e.g., presence of other persons in the frame, distance of the operator from the camera, degree of occlusion of the operator's image, lighting conditions, angle of the frame with respect to the frontal, sagittal and transverse planes, degree of overlap of the operator's image with that of other persons in the frame) might impact on the system's accuracy.

Furthermore, the results provided by the system were only compared with those obtained by ergonomic experts through manual analysis. Although manual analysis is the most widely used system to carry out ergonomic analysis to date, it has its limitations, as well as the various proposed MoCap systems. The problem lies in the difficulty of unambiguously identifying a 3D posture from 2D images depicting it from different viewpoints, due to the perspective distortion introduced.

A viable solution to overcome actual limitation of MoCap systems might be that proposed in[39], based on the identification of view-invariant features that allow the 3D posture to be recognised regardless of the viewpoint from which it is observed, albeit that methodology was tested in the construction environment rather than in the manufacturing industry. An alternative approach could be to create a synthetic dataset of the postures deemed ergonomically unsuitable by the different ergonomic risk assessment methods (e.g. REBA, RULA), captured from different viewpoints. The resulting dataset could serve as a basis for training a network capable of classifying postures according to an ERA score, bypassing the calculation of angles affected by the perspective distortion problem. The development of this methodology and related tools, as well as a comparative analysis with the current approach that involves the calculation of angles, will be the subject of future works.

Conclusion

The paper reports the results of an experimentation carried out in an industrial manufacturing context of a state-of-the-art 2D RGB MoCap system, comparing the results of its application in diverse production lines and workspaces differing each other for the layout, the environmental lighting, the workers' mobility inside the working spaces and the presence of occluding obstacles located around the operators.

The results of the experiments showed a large variability in the tools' accuracy as a function of the characteristics of the lines and workstations, as well as the configuration and positioning of the cameras. In general, the results obtained do not allow us to state that the considered system, as well as other similar systems, has sufficient reliability to effectively support the ergonomic analysis carried out by experts to certify workstations' Risk Level. However, the Accuracy level demonstrated by the system in terms of ERA Risk Level allows us to argue that such systems may be useful to continuously monitor the operators during their work activities and can represent a helpful tool to speed up the ergonomists' job. In fact, the introduction of such systems can potentially revolutionise the way in which ergonomic analysis is carried out in the industrial manufacturing context, bringing benefits both in terms of improving the safety and well-being of workers and in terms of increasing the efficiency and competitiveness of the company.

Such systems do not seem capable of providing ergonomists with reliable and precise ergonomic assessments or of collecting specific data on individual workers or specific workstations. However, they appear to be sufficiently accurate to implement a network of cameras along the production lines in order to monitor the statistical trends of the ERA Scores and, more importantly, the ERA Risk Levels. A monitoring network designed as proposed would be vital to limit the risk of work-related MSDs as it would allow the early identification of ergonomic discomfort situations, alerting ergonomists to the need for intervention.

Such approach can improve the efficiency, productivity and competitiveness of manufacturing companies, particularly those characterised by work environments with frequent changes of operators at workstations and tasks. It can also lead to a reduction in costs associated with accidents, illnesses and medical expenses related to non-ergonomic working conditions, as employees would be working in optimal conditions.

In order to improve the accuracy of current MoCap systems, it is necessary to overcome their limitations, mainly due to the difficulty of unambiguously identifying a 3D posture from 2D images depicting it from different viewpoints, due to the perspective distortion introduced. To this end, one possible avenue is the development of a new methodology that allows ergonomic analyses to be carried out, bypassing the calculation of angles from 2D postures, by training a machine learning-based system capable of associating a posture with its ergonomic evaluation, using synthetic datasets containing postures (inspired, for example, by the synthetic dataset presented in[43,48]).

Data availability
The data generated and analysed during the current study are public available in the following repository: https://github.com/boso94/ergonomicAnalysis.

Received: 31 May 2024; Accepted: 8 November 2024
Published online: 13 November 2024

References
1. Aheleroff, S., Huang, H., Xu, X. & Zhong, R. Y. Toward sustainability and resilience with industry 4.0 and industry 5.0. *Front. Manuf. Technol.* **2**, 951643 (2022).
2. Garrido, S., Muniz Jr, J. & Ribeiro, V. B. Operations management, sustainability & industry 5.0: a critical analysis and future agenda. *Clean. Logist Supply Chain*. **10**, 100–141, (2024).
3. Kadir, B. A. & Broberg, O. Human-centered design of work systems in the transition to industry 4.0. *Appl. Ergon.* **92**, 103334 (2021).
4. Romero, D., Stahre, J. & Taisch, M. The operator 4.0: towards socially sustainable factories of the future. *Comput. Ind. Eng.* **139**, 106128 (2020).
5. Sgarbossa, F., Grosse, E. H., Neumann, W. P., Battini, D. & Glock, C. H. Human factors in production and logistics systems of the future. *Annu. Rev. Control*. **49**, 295–305 (2020).
6. European Commission. The 2015 ageing report: economic and budgetary projections for the 28 EU member state. Retrieved July 5. from (2021). https://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf
7. Ilmarinen, J. Physical requirements associated with the work of aging workers in the European union. *Exp. Aging Res.* **28**, 7–23 (2002).
8. Ciccarelli, M., Papetti, A. & Germani, M. Exploring how new industrial paradigms affect the workforce: a literature review of Operator 4.0. *J. Manuf. Syst.* **70**, 464–483 (2023).
9. European Agency for Safety and Health at Work. Work-related musculoskeletal disorders: prevalence, costs and demographics in the EU. Retrieved November 21. from (2023). https://osha.europa.eu/en/publications/msds-facts-and-figures-overview-prevalence-costs-and-demographics-msds-europe
10. Ciccarelli, M. et al. SPECTRE: a deep learning network for posture recognition in manufacturing. *J. Intell. Manuf.* **34**, 3469–3481 (2023).
11. Joshi, M. & Deshpande, V. A systematic review of comparative studies on ergonomic assessment techniques. *Int. J. Ind. Ergon.* **74**, 102865 (2019).
12. Hignett, S. & McAtamney, L. Rapid entire body assessment (REBA). *Appl. Ergon.* **31**, 201–205 (2000).
13. McAtamney, L. & Corlett, E. N. RULA: a survey method for the investigation of work-related upper limb disorders. *Appl. Ergon.* **24**, 91–99 (1993).
14. Karhu, O., Kansi, P. & Kuorinka, I. Correcting working postures in industry: a practical method for analysis. *Appl. Ergon.* **8**, 199–201 (1977).
15. Occhipinti, E. OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*. **41**, 1290–1311 (1998).
16. Battini, D., Faccio, M., Persona, A. & Sgarbossa, F. New methodological framework to improve productivity and ergonomics in assembly system design. *Int. J. Ind. Ergon.* **41**, 30–42 (2011).
17. Vignais, N. et al. Innovative system for real-time ergonomic feedback in industrial manufacturing. *Appl. Ergon.* **44**, 566–574 (2013).
18. Vignais, N., Bernard, F., Touvenot, G. & Sagot, J. C. Physical risk factors identification based on body sensor network combined to videotaping. *Appl. Ergon.* **65**, 410–417 (2017).
19. 'Xsense'. Retrieved November 23. from (2023). https://www.xsens.com/motion-capture
20. 'Vicon Blue Trident'. Retrieved November 23. from (2023). https://www.vicon.com/hardware/blue-trident/
21. 'Vicon Nexus'. Retrieved November 23. from (2023). https://www.vicon.com/software/nexus/
22. 'Optitrack'. Retrieved November 23. from (2023). https://optitrack.com/
23. Manghisi, V. M. et al. Automatic ergonomic postural risk monitoring on the factory shopfloor–the ergosentinel tool. *Procedia Manuf.* **42**, 97–103 (2020).
24. Schall, M. C., Sesek, R. F. & Cavuoto, L. A. Barriers to the adoption of wearable sensors in the workplace: a survey of occupational safety and health professionals. *Hum. Factors J. Hum. Factors Ergon. Soc.* **60**, 351–362 (2018).
25. Cao, Z., Simon, T., Wei, S. E. & Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7291–7299, (2017).
26. Ota, M. et al. Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait Posture*. **80**, 62–67 (2020).
27. da Silva Neto, J. G., Teixeira, J. M. X. N. & Teichrieb, V. Analyzing embedded pose estimation solutions for human behaviour understanding. *Anais Estendidos do XXII Simpósio De Realidade Virtual e Aumentada*. **SBC**, 30–34 (2020).
28. 'GitHub - tryagainconcepts/tf-pose-estimation: Deep Pose Estimation implemented using Tensorflow with Custom Architectures for fast inference.' Retrieved January 30. from (2024). https://github.com/tryagainconcepts/tf-pose-estimation
29. 'Wie funktioniert unsere LINDERA Algorithmus? 3 Top Fakten'. Retrieved January 30. from (2024). https://lindera.de/technologie/wie-funktioniert-lindera-algorithmus/

30. Agostinelli, T. et al. Preliminary validation of a low-cost motion analysis system based on RGB cameras to support the evaluation of postural risk assessment. *Appl. Sci.* **11**, 10645 (2021).
31. 'MediaPipe' Google for Developers. Retrieved January 30, from (2024). https://developers.google.com/mediapipe
32. Vakunov, A., Chang, C. L., Zhang, F., Sung, G., Grundmann, M., & Bazarevsky, V. Mediapipe hands: On-device real-time hand tracking. *In Workshop on Computer Vision for AR/VR*. **2** (4), 5 (2020).
33. Duy Khuat, B., Thai Phung, D., Thi Thu Pham, H., Ngoc Bui, A. & Tung Ngo, S. Vietnamese sign language detection using Mediapipe. *Proc. 2021 10th Int. Conf. Softw. Comput. Appl.* **162**, 165 (2021).
34. Lee, Y. C. & Lee, C. H. SEE: a proactive strategy-centric and deep learning-based ergonomic risk assessment system for risky posture recognition. *Adv. Eng. Inf.* **53**, 101717 (2022).
35. Li, Z., Zhang, R., Lee, C. H. & Lee, Y. C. An evaluation of posture recognition based on intelligent rapid entire body assessment system for determining musculoskeletal disorders. *Sensors*. **20** (16), 4414 (2020).
36. Li, L., Martin, T. & Xu, X. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Appl. Ergon.* **87**, 103138 (2020).
37. Lin, P. C., Chen, Y. J., Chen, W. S. & Lee, Y. J. Automatic real-time occupational posture evaluation and select corresponding ergonomic assessments. *Sci. Rep.* **12**, 2139 (2022).
38. Su, J. M., Chang, J. H., Indrayani, N. L. D. & Wang, C. J. Machine learning approach to determine the decision rules in ergonomic assessment of working posture in sewing machine operators. *J. Saf. Res.* **87**, 15–26 (2023).
39. Yan, X. et al. Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Adv. Eng. Inf.* **34**, 152–163 (2017).
40. Li, L. & Xu, X. A deep learning-based RULA method for working posture assessment. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, 63, 1090–1094, (2019).
41. Ionescu, C., Li, F. & Sminchisescu, C. Latent structured models for human pose estimation *International Conference on Computer Vision*, 2220–2227, (2011). (2011).
42. Ionescu, C., Papava, D., Olaru, V. & Sminchisescu, C. Human3. 6m: large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1325–1339 (2013).
43. MassirisFernández, M., Fernández, J. Á., Bajo, J. M. & Delrieux, C. A. Ergonomic risk assessment based on computer vision and machine learning. *Comput. Ind. Eng.* **149**, 106816 (2020).
44. Nayak, G. K. & Kim, E. Development of a fully automated RULA assessment system based on computer vision. *Int. J. Ind. Ergon.* **86**, 103218 (2021).
45. Generosi, A., Agostinelli, T., Ceccacci, S. & Mengoni, M. A novel platform to enable the future human-centered factory. *Int. J. Adv. Manuf. Technol.* **122**, 4221–4233 (2022).
46. Jeong, S. & Kook, J. CREBAS: computer-based REBA evaluation system for wood manufacturers using MediaPipe. *Appl. Sci.* **13**, 938 (2023).
47. Pagnon, D., Domalain, M. & Reveret, L. Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness. *Sensors*. **21**, 6530 (2021).
48. Seo, J. & Lee, S. Automated postural ergonomic risk assessment using vision-based posture classification. *Autom. Constr.* **128**, 103725 (2021).

Author contributions

Manuscript draft: T.A., A.G., S.C. Study conception: T.A., A.G. Software: T.A., A.G. Data processing: T.A., A.G. Statistical data processing: T.A., S.C. Data interpretation: T.A., A.G., S.C. Manuscript revision: T.A., A.G., M.M. Supervision: M.M.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.