Ergonomic Assessment Using Human Pose Estimation: A Real-Time Approach with YOLO and BlazePose

John Davi D. C. Pires[1], Marcelo Costa Oliveira[1], Baldoíno Fonseca [1]
Márcio Ribeiro [1], Bruno Antônio F. Giglio[2] and Alexandre Xavier Calzado [2]

[1]Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Caixa Postal 57.072-970 – Maceió – AL – Brazil

[2]Serviço Social da Indústria Departamento Regional de Alagoas
Caixa Postal 57.055-000 – Maceió – AL – Brazil

{jddcp,oliveiramc,baldoino,marcio}@ic.ufal.br

{bruno.giglio, alexandre.calzado}@sistemafiea.com.br

Abstract. *Musculoskeletal disorders (MSDs) are a major cause of workplace injuries, often linked to poor posture and repetitive movements. Traditional assessments like RULA are subjective and labor-intensive. To overcome these limitations, this study employs computer vision for automated ergonomic evaluation. Using YOLO for person detection and BlazePose for pose estimation, the system analyzes workplace postures via surveillance cameras, generating RULA scores to assess ergonomic risks. Tested on the MPI-INF-3DHP dataset, it achieves MPJPE-PA values between 243.91 mm and 295.70 mm. This scalable, objective approach enhances workplace safety. Future work will refine accuracy, handle real-world variations, and integrate real-time feedback.*

1. Introduction

Musculoskeletal disorders (MSDs) are a leading cause of injuries in industries, affecting muscles, tendons, ligaments, and peripheral nerves [ErgoPlus 2024]. These conditions are caused by repetitive tasks, poor posture, and biomechanical loads, leading to pain and decreased work productivity. In Brazil, work-related musculoskeletal disorders (WMSDs) have become an issue, with over 67,000 reported cases between 2007 and 2018, affecting women and workers aged 40–49 years [Oliveira and Santana 2024].

Traditional ergonomic assessment methods, such as the Rapid Upper Limb Assessment (RULA), are used in industrial settings due to their simplicity and effectiveness [Roman-Liu 2014]. These observational techniques rely on trained evaluators to estimate joint angles and assign risk scores based on formulas, helping to identify hazardous postures. However, these methods present limitations: they are time-consuming, subjective, and require experienced raters, leading to inconsistencies in the evaluations [Li and Xu 2019]. Some researchers have attempted to address these challenges using wearable sensors like inertial measurement units (IMU) and electromyography (EMG), but these approaches can be intrusive and are not suitable for large-scale monitoring [Li and Xu 2019]. There is growing interest in using computer vision techniques to automate posture assessment in an objective and scalable manner.

Human Pose Estimation (HPE) enables real-time ergonomic assessment by detecting and mapping body joints from images or video frames [Gong et al. 2016]. Deep learning advances have improved the accuracy of 2D and 3D posture analysis. Studies have applied HPE for ergonomic evaluation, including a deep learning-based method for estimating RULA scores from 2D poses [Chidambaram et al. 2023]. However, existing methods often struggle with occupational tasks due to limited training data and posture bias. To overcome these issues, this work proposes an automated ergonomic assessment system using surveillance cameras and a streaming approach for continuous monitoring. The system also generates individual and sector-level reports to enhance workplace safety and reduce MSD risks.

The remainder of this paper is organized as follows. Section 2 presents related works on ergonomic assessment and human pose estimation techniques. Section 3 describes the methodology adopted for system development, including the camera setup and data processing pipeline. Section 4 details the experimental setup and results obtained in real workplace scenarios. Finally, Section 5 concludes the paper and discusses future research directions.

2. Related Works

Recent studies have applied deep learning and computer vision to ergonomic risk assessment. [Chidambaram et al. 2023] proposed a CNN-based system using RGB images to estimate RULA scores with high accuracy. However, their method targets controlled environments and lacks real-time performance or integration with workplace infrastructure. In contrast, our approach enables continuous, real-time posture assessment using surveillance cameras in industrial settings, without wearable devices or depth sensors, and includes automated individual and sector-level ergonomic reporting.

[Li and Xu 2019] proposed a deep learning method for estimating RULA scores directly from 2D human poses projected from 3D data (Human3.6M) [Ionescu et al. 2014], reducing the subjectivity of manual assessments by computing joint angles and training a neural network to predict risk levels. While their approach targets accuracy in controlled settings, our system prioritizes scalability and real-world use by integrating detection, posture analysis, and report generation in a continuous pipeline. Evaluated on the MPI-INF-3DHP [Mehta et al. 2017] dataset, it achieved MPJPE-PA values between 243.91 mm and 295.70 mm, confirming its effectiveness in practical environments.

3. Materials and methods

The proposed system is a web-based application developed using FastAPI [Ramírez 2024]. The primary objective is to schedule ergonomic analyzes during working hours and generate comprehensive reports based on RULA assessments, providing both general and individual insights into workers' postures. In this study, a total of 6,151 samples from the MPI-INF-3DHP data set (Monocular 3D Human Pose Estimation In The Wild Using Improved CNN Supervision) [Mehta et al. 2017]. The approach follows a top-down pipeline [Zheng et al. 2023], shown in Figure 1.

The data used in this study is collected from a surveillance camera installed in the workplace. The camera in use is the Intelbras VIP-1230-B-G4, a Full HD (1920×1080)

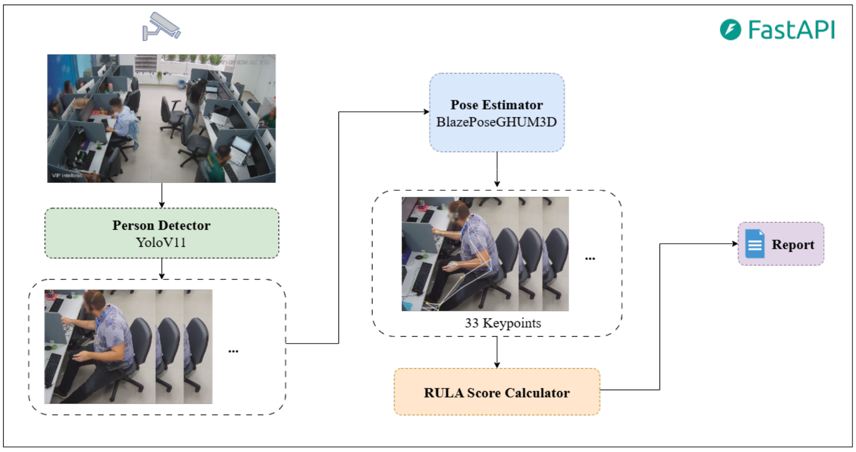

Figure 1. A diagram containing all the process behind the our system.

device that provides sufficient resolution to accurately detect a person's body and estimate their pose.

3.1. Person Detection

The system utilizes YOLOv11 [Khanam and Hussain 2024] to identify and localize individuals within a workplace environment captured by surveillance cameras. This step is crucial for isolating workers from the background and ensuring that only relevant subjects are analyzed, as shown in Figure 2. The use of YOLO enhances the accuracy of pose estimation by focusing on one person at a time and enables independent analysis of each individual present in the work environment.

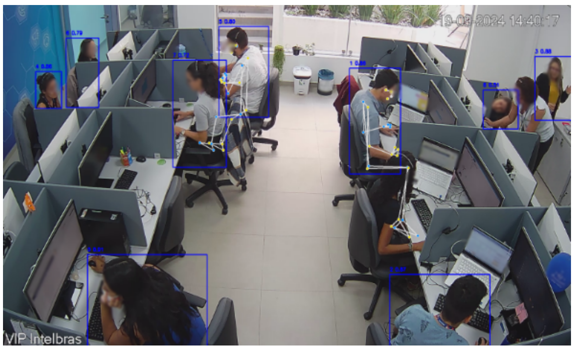

Figure 2. An image from the surveillance camera with bounding boxes representing the people detected by YOLOv11

3.2. Pose Estimation

Once individuals are detected, the system applies BlazePoseGHUM3D to extract 33 key body landmarks, detailed in Figure 3 enabling a detailed posture analysis. The estimated skeletal structure is overlaid onto the original image, allowing for a visual representation of the worker's posture.

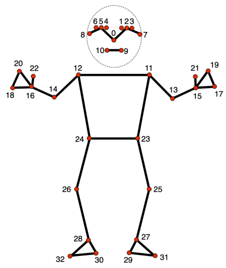

Figure 3. Visualization of BlazePose's 33 key points in human pose estimation.

3.2.1. RULA Score Calculation

Based on the extracted body landmarks, the system employs the Rapid Upper Limb Assessment (RULA) method to evaluate ergonomic risks. This process quantifies potential strain on workers and identifies postures that may contribute to musculoskeletal disorders. The implementation follows the approach proposed by [Nayak and Kim 2021], utilizing vector calculus to analyze limb positions based on pairs of detected landmarks.

3.3. Report Generation

The analyzed ergonomic data is compiled into structured reports for supervisors. These reports include both individual assessments and sector-wide evaluations, enabling targeted interventions to improve workplace ergonomics.

On a broader level, a graph displays the distribution of RULA scores across the entire environment. On an individual level, the report includes the RULA score count for each person, along with a timeline highlighting moments when an individual remained in a poor posture for longer than a user-defined threshold (Figure 4).

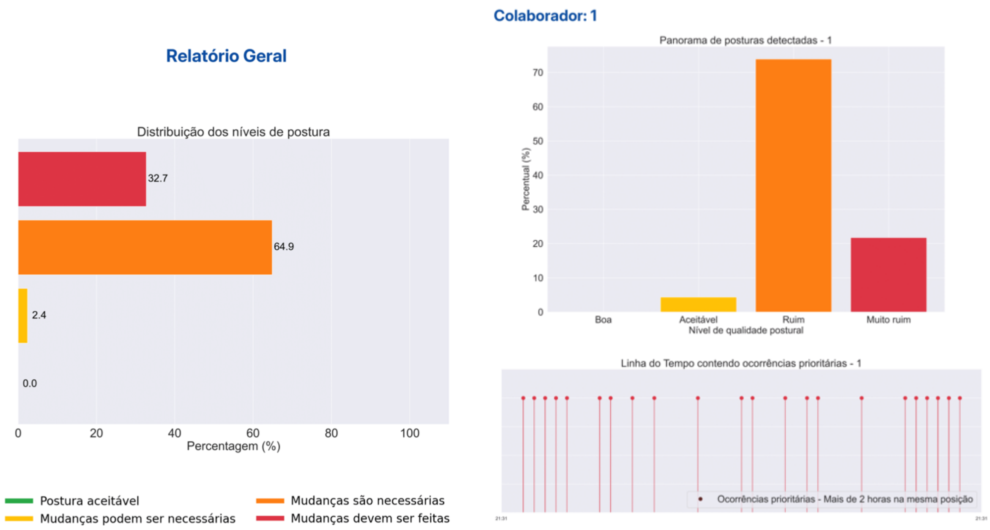

Figure 4. Graphs addressed in the report generated by the application.

4. Results and discussion

The model's performance was evaluated using two key metrics: MPJPE (Mean Per Joint Position Error) and MPJPE-PA (Mean Per Joint Position Error with Procrustes Alignment). The errors calculated for these metrics were calculated by turning 33 key points from the Blazepose model into 17 key points to fit the MPI-INF-3DHP dataset. The results are shown in the table below.

Metric	**TS1**	**TS2**	**TS3**	**TS4**
MPJPE (mm)	1,331.35	1,430.55	1,492.30	1,441.32
MPJPE-PA (mm)	247.44	295.70	284.97	250.27

Table 1. Model Performance on MPI-INF-3DHP Dataset Across Test Sets

The results show that the model achieves an MPJPE between 1,331.35 mm and 1,492.30 mm, with a refined MPJPE-PA range of 243.91 mm to 295.70 mm after alignment. While there is room for improvement in absolute joint position accuracy, the MPJPE-PA values indicate that the model effectively captures relative joint positions, making it well-suited for ergonomic assessments. With further refinements, such as fine-tuning on task-specific data, the system can achieve even greater precision for workplace posture evaluation.

An approach for human pose estimation is to use a single RGB-D camera (e.g. Microsoft Kinect) to estimate 3D pose [Plantard et al. 2017]. They developed a method for ergonomic assessment in real workplaces using Kinect data, improving posture evaluation and musculoskeletal risk assessments in cluttered environments. The method mitigates occlusions and enhances accuracy in sub-optimal conditions by applying a correction algorithm to Kinect's skeleton data. However, this solution may struggle with dynamic or fast-paced movements due to the limited depth range of the RGB-D camera.

To overcome this limitation, an approach based solely on color images is preferable. Deep Learning-based Human Pose Estimation (HPE) enables automated, objective posture analysis using standard RGB images, eliminating the need for depth sensors. For example, [Abobakr et al. 2019] used RGB-D images and convolutional neural networks (CNNs) to implement a semi-automated ergonomic assessment system, achieving a mean absolute error of 3.19° and 89% accuracy in predicting RULA grand scores. However, the paper does not evaluate performance under real-time conditions or dynamic movements, which could impact the robustness of the method.

To overcome the limitations of traditional ergonomic assessments, computer vision methods have gained traction. BlazePose, a lightweight CNN-based model, enables real-time posture evaluation using RGB images without depth sensors, as demonstrated in [Chidambaram et al. 2023] and previously discussed in Section 2.

5. Conclusion and Future Works

In conclusion, this work presents an automated ergonomic assessment system that leverages YOLO and BlazePose for real-time human pose estimation and RULA scoring. The proposed system effectively monitors worker postures in dynamic environments, offering scalable, objective, and efficient ergonomic evaluations. The results demonstrate promising accuracy in predicting RULA scores and assessing ergonomic risks. Future work will focus on enhancing the system's robustness under varying real-world conditions, such as diverse lighting environments and occlusions. Additionally, exploring the integration of real-time feedback for workers and supervisors could further optimize the system's impact on workplace safety. Further improvements in model accuracy and real-time processing will help expand the applicability of this system in broader industrial settings.

References

Abobakr, A., Nahavandi, D., Hossny, M., Iskander, J., Attia, M., Nahavandi, S., and Smets, M. (2019). Rgb-d ergonomic assessment system of adopted working postures. *Applied Ergonomics*, 80:75–88. Epub 2019 May 25.

Chidambaram, V., Mohan, G., .M, V., and Kanchan, B. (2023). Ergonomic investigations on novel dynamic postural estimator using blaze pose and transfer learning. *Ergonomics*, 67:1–35.

ErgoPlus (2024). Definition of musculoskeletal disorder (msd). `https://ergo-plus.com/definition-of-musculoskeletal-disorder-msd/`.

Gong, W., Zhang, X., Gonzàlez, J., Sobral, A., Bouwmans, T., Tu, C., and Zahzah, E.-h. (2016). Human pose estimation from monocular images: A comprehensive survey. *Sensors*, 16(12).

Ionescu, C., Papava, D., Olaru, V., and Sminchisescu, C. (2014). Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339.

Khanam, R. and Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements.

Li, L. and Xu, X. (2019). A deep learning-based rula method for working posture assessment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):1090–1094.

Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., and Theobalt, C. (2017). Monocular 3d human pose estimation in the wild using improved cnn supervision. In *3D Vision (3DV), 2017 Fifth International Conference on*. IEEE.

Nayak, G. K. and Kim, E. (2021). Development of a fully automated rula assessment system based on computer vision. *International Journal of Industrial Ergonomics*, 86:103218.

Oliveira, P. G. d. and Santana, V. S. (2024). Space-time analysis of work-related musculoskeletal disorders in brazil: an ecological study. *Cadernos de Saúde Pública*, 40(7):e00141823.

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

Ramírez, S. (2024). Fastapi. `https://fastapi.tiangolo.com`.

Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for msd risk assessment. *Applied Ergonomics*, 45(3):420–427.

Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., and Shah, M. (2023). Deep learning-based human pose estimation: A survey.