Human Factors and Wearable Technologies, Vol. 175, 2025, 41–48
https://doi.org/10.54941/ahfe1006357

Improvement of Safety in Collaborative Robotics Through Prompt Detection of Abrupt Movements

Greta Di Vincenzo, Michele Polito, Elisa Digo, Laura Gastaldi, and Stefano Pastorelli

Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Italy

ABSTRACT

The rise of Industry 5.0 has redefined human-robot collaboration, leveraging the strength of robot in performing repetitive tasks with precision and the one of humans in expressing decision-making abilities. Ensuring safety requires a real-time recognition of human gestures, especially when abrupt movements due to inattention and unexpected circumstances occur. This study integrates wearable magneto-inertial measurement units (MIMUs) with deep learning techniques to distinguish between human standard and abrupt gestures during an industrial task. A Long Short-Term Memory neural network was trained on MIMU acceleration data from 60 participants performing a pick-and-place task with induced abrupt gestures. After pre-processing and segmenting signals into overlapping windows, 90% overlap was found optimal due to its high performance and short time of classification. The testing of the system on new participants demonstrated its high reliability (balanced accuracy of 91%, macro F1-score of 90%, specificity of 97%, and recall of 85%) in real-time gesture detection.

Keywords: Human-robot interaction, Gesture recognition, Deep learning, MIMUs, Safety

INTRODUCTION

The transition from Industry 4.0 to Industry 5.0 has introduced some new pillars giving robotics a broader perspective. Technology-driven processes give way to a human-centric approach, in which the safety, the well-being, and the inclusion of workers within the work environment are prioritized (Xu et al., 2021). The collaboration of humans and robots in shared workspaces enhances their respective and complementary strengths. While robots execute repetitive and high-precision tasks with speed and consistency, humans coordinate the collaboration system expressing their essential cognitive skills such as decision-making, problem solving, and adaptability (Zafar et al., 2024). To ensure an effective and safe collaboration, the robot system must constantly recognize human gestures and promptly react to them. This requirement extends beyond repetitive standard gestures typical of manipulation tasks in a workshop station (Bortolini et al., 2017; Lin et al., 2010), also including abrupt movements caused by inattention or external circumstances unrelated to the work task (Digo et al., 2024b)

In recent years, artificial intelligence (AI) has been rapidly gaining traction in industrial environments, as it represents a new opportunity for collaboration within human-robot systems. This capability enhances safety by reducing risks and promoting more effective human-robot interaction. Recent advancements in AI have significantly improved human–robot collaboration through the development of cognitive models collecting and processing information from both the environment and the human operator. Subsequently, the acquired information is translated into data promoting the dynamic adaptation of the robot behaviour (Zafar et al., 2024). One of the most suitable technologies for human motion analysis in a work environment are magneto-inertial measurement units (MIMUs), which offer portability, lightweight design, and independence from specific laboratory settings. When combined, MIMUs and AI open the door to addressing complex challenges related to real-time human gesture recognition by robots. According to literature, the integration of MIMUs and AI provides a cost-effective, versatile, and reliable tool for the identification of human activities in collaborative robotics systems (Liu et al., 2018; Ordóñez et al., 2016).

Even if the scientific community has explored the recognition of abrupt human movements with MIMUs (Digo et al., 2024b; Polito et al., 2023; Rosso et al., 2022), the integration of AI techniques for a real-time recognition of abrupt gestures still receives relatively little focus. Accordingly, the aim of this study was to exploit the integration of MIMUs with deep learning techniques to implement a real-time algorithm for the detection of human abrupt movements in collaborative robotics scenarios. The final outcome consisted in utilizing data from a MIMU fixed on the human forearm to distinguish between standard and abrupt gestures during a typical industrial task. Specifically, a Long Short-Term Memory (LSTM) neural network was first trained on a previously collected dataset including MIMUs signals from sixty subjects during a traditional pick-and-place task interspersed with randomly induced abrupt gestures. Subsequently, the trained network was tested online on data recorded from five new participants.

EXPERIMENTAL TEST

The analyses in this study were conducted using a previously collected dataset involving sixty healthy participants of working age (Digo et al., 2024a). The experimental protocol consisted of performing a typical industrial pick-and-place task while seated at a table. A custom set-up including two boards with holes positioned at varying distances horizontally from the participant and vertically from the table was specifically realized (Digo et al., 2024b). Each participant performed three different trials: (i) FR_r – sitting frontally with respect to the table and using the right arm (Figure 1a); (ii) FR_l – sitting frontally with respect to the table and using the left arm (Figure 1b); (iii) sitting laterally with respect to the table and using the left arm (Figure 1c). In each trial, a sequence of 30 pick-and-place movements was executed picking a ball at a time from a box on the table and placing it into a hole corresponding to the lighten up of a green LED. Among these movements, 4 abrupt movements were randomly triggered by 2 visual alarms (light-up of a red LED) and 2 acoustic alarms (sound of a buzzer). In case of the visual alarm (activated 500 ms after the green LED was turned on), participants were asked to place the ball in the hole corresponding to the specific red LED as fast as possible. In case of the acoustic alarm (activated when the green LED was turned on), participants were asked to raise the arm involved in the test as fast as possible. During the tests, one MIMU (Opal™ V2R, APDM, USA.) was fixed on subjects' forearm used to pick and place the balls during the test.

(a) (b) (c)

Figure 1: Experimental set-up and protocol: a) FR_r trial; b) FR_l trial; c) LA_l trial.

Due to the effective capacity in modelling sequences and time-dependent data (Ordóñez et al., 2016), a LSTM neural network was trained and tested through linear accelerations recorded with MIMUs on participants' forearms for the distinction between standard and abrupt movements. This network was developed using the high-level library Keras (Python, USA). As shown in Figure 2, the LSTM architecture is characterized by four layers: (i) Input Layer with 100 time-steps and 1 feature; (ii) LSTM layer with 100 hidden units; (iii) Dropout layer with the user-defined rate equal to 0.5; (iv) Dense layer with a single neuron characterized by a sigmoid activation function. Moreover, the threshold for binary classification was set to 0.9 to address class imbalance, as windows with abrupt gestures occur in only about 5% of cases compared to standard gestures.

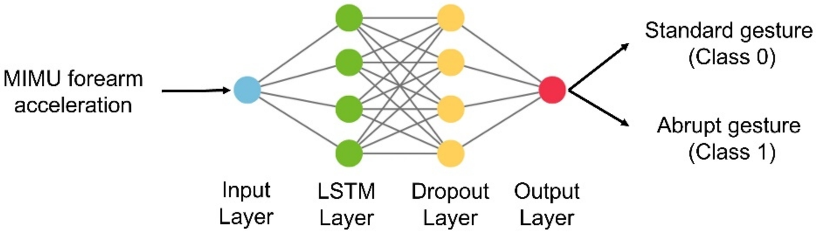

Figure 2: Architecture of the LSTM neural network.

For the training of LSTM, MIMUs linear accelerations were processed removing the gravity component, calculating the norm, and segmenting the resulting signals into 3-second windows corresponding to single movements and labelled as either normal (label = 0) or abrupt (label = 1). Each 3-second window was further divided into six sub-windows of 0.5 s each. Considering the occurrence of acceleration peaks in correspondence of visual or acoustic alarms, only the 2^{nd}, the 3^{rd}, and the 4^{th} sub-windows were labelled as abrupt. Subsequently, these sub-windows were split into two groups: the 80% was used for training, and the remaining 20% for testing. The training set was further divided using k-fold cross-validation, with $k=5$ (Wong and Yeh, 2020).

To approach real-time recognition, a sliding windows method was adopted (Imanzadeh et al., 2024; Xiang et al., 2024). Each window has a fixed length, and it advances incrementally based on a defined step size. Since the step size is shorter than the window length, consecutive windows overlap. In this study, the window length was fixed at 0.5 s, while the overlap percentage and consequently the step size were varied to assess their impact on network performance. Overlap percentages of 50%, 75%, 90%, 95%, and 99% were tested. Once the signal was segmented, labels were assigned to each window with the same criteria adopted for the training. Segmented data were provided as input to the network, comparing the output to the actual movement through the estimation of performance metrics: balanced accuracy, macro F1-score, specificity, and recall (Cullerne Bown, 2024; Rivera et al., 2017). In addition, the time required to analyze a single window was estimated dividing the average inference time (time to classify data for a single subject) by the number of windows per subject.

The testing of LSTM neural network was conducted in real-time with five additional subjects performing the same protocol with the same set-up. A block diagram of the real-time detection flow is represented in Figure 3. Once the communication with the sensors is established, MIMUs raw data are read and stored. The pre-processing phase includes removing gravitational acceleration, calculating the acceleration norm, and segmenting the data into overlapping windows. The pre-processed data are fed into the network for recognition. To visually represent the output of the classification, a green or red window is displayed if a standard or abrupt gesture is recognized, respectively. As shown by the blue arrow in Figure 3, the recognition of the just-completed gesture i occurs simultaneously with the data streaming of the upcoming movement $i + 1$.

By comparing the actual sequence of gestures with the predicted one, a confusion matrix was generated, allowing again the computation of performance metrics (balanced accuracy, macro F1-score, specificity, and recall). Additionally, the time required for data pre-processing and movement classification was calculated and averaged both within and across subjects.

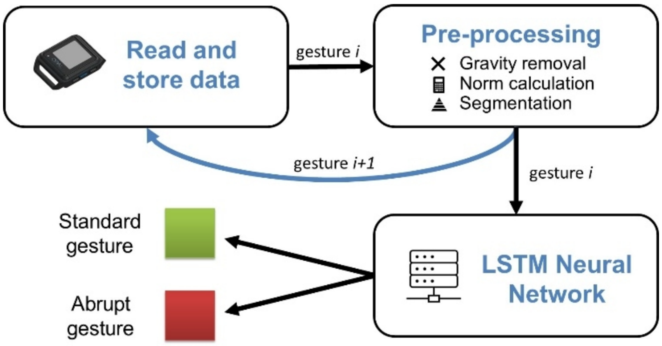

Figure 3: Block diagram summarizing the real time detection of movements.

RESULTS AND DISCUSSIONS

Figure 4 presents the percentage performance metrics (balanced accuracy in blue, macro F1-score in red, specificity in green, and recall in yellow) as a function of the overlap percentage.

Table 1 reports the average time (ms) required to analyze a single window for different overlap percentages. Among the two overlap percentages (90% and 95%) with an average processing time below 1 ms, 90% was selected as the most suitable due to its superior metrics: a balanced accuracy of 87%, a macro F1-score of 78%, a specificity of 86%, and a recall of 89%.

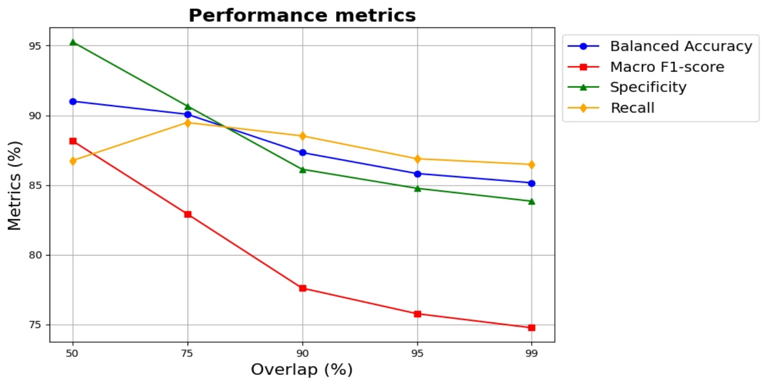

Figure 4: Performance metrics for different overlap percentages.

Table 1: Average time to analyse a single window for different overlap percentages.

Overlap (%)	Average Time to Analyze a Single Window (ms)
50%	1.30
75%	1.11
90%	0.84
95%	0.86
99%	1.18

Figure 5 presents the confusion matrix obtained from the real-time test of the detection system. In this test, only 12 standard movements out of 390 (3%) were misclassified as abrupt gestures. From a safety perspective, failing to recognize an abrupt gesture could lead to a potential collision. In this test, 9 abrupt gestures out of 60 (15%) were misclassified as normal ones. However, performance metrics calculated from the confusion matrix were always at least equal to 85% (balanced accuracy of 91%, macro F1-score of 90%, specificity of 97%, and recall of 85%), testifying an excellent classification (Rivera et al., 2017).

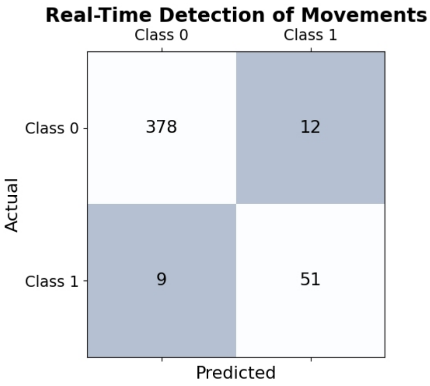

Figure 5: Confusion matrix with classification outcomes of the real-time detection of movements.

Table 2 reports the pre-processing and classification times for all five subjects involved in the test phase. Pre-processing times are always below 10 ms, indicating that the operations performed on raw data to prepare them for the LSTM neural network are not computationally expensive. Regarding classification times, they are always below 265 ms, meaning that the network requires approximately 9% of the total movement duration to perform the classification.

Table 2: Pre-processing times and movement classification times for all subjects (mean ± standard deviation).

Subject	Pre-Processing Time (ms)	Movement Classification Time (ms)
01	8.90 ± 1.87	263.8 ± 65.2
02	9.12 ± 2.14	265.1 ± 95.6
03	9.13 ± 2.45	251.3 ± 67.9
04	8.83 ± 2.26	253.7 ± 70.6
05	9.04 ± 2.49	264.7 ± 70.2
Inter-subjects	9.00 ± 2.26	259.7 ± 74.9

CONCLUSION

The aim of this study was to integrate MIMUs with deep learning techniques to recognize human movements in real-time, thereby improving human-robot collaboration in terms of both efficiency and safety. Specifically, a LSTM network was trained on forearm MIMU acceleration data to distinguish between standard and abrupt gestures occurring during a typical industrial task. Overall, results demonstrate the effectiveness of this approach in identifying abrupt movements under conditions closely approximating real-time. Ongoing research efforts focus on refining the signal acquisition and processing to minimize streaming delays. The objective is to enhance the system's responsiveness, ensuring that abrupt movements are detected at their earliest onset and hence safety is guaranteed.

REFERENCES

Bortolini, M., Ferrari, E., Gamberi, M., Pilati, F., Faccio, M., 2017. Assembly system design in the Industry 4.0 era: A general framework. IFAC-PapersOnLine 50, 5700–5705.

Cullerne Bown, W., 2024. Sensitivity and Specificity versus Precision and Recall, and Related Dilemmas. J Classif 41, 402–426.

Digo, E., Polito, M., Caselli, E., Gastaldi, L., Pastorelli, S., 2024a. Dataset of Abrupt and Standard Industrial Gestures (DASIG). https://doi.org/10.5281/zenodo.13927735

Digo, E., Polito, M., Pastorelli, S., Gastaldi, L., 2024b. Detection of upper limb abrupt gestures for human–machine interaction using deep learning techniques. Journal of the Brazilian Society of Mechanical Sciences and Engineering 46.

Imanzadeh, S., Tanha, J., Jalili, M., 2024. Ensemble of deep learning techniques to human activity recognition using smart phone signals. Multimed Tools Appl 83, 89635–89664.

Lin, C. L., Wang, M. J. J., Drury, C. G., Chen, Y. S., 2010. Evaluation of perceived discomfort in repetitive arm reaching and holding tasks. Int J Ind Ergon 40, 90–96.

Liu, H., Fang, T., Zhou, T., Wang, Y., Wang, L., 2018. Deep Learning-based Multimodal Control Interface for Human-Robot Collaboration. Procedia CIRP 72, 3–8.

Ordóñez, F. J., Roggen, D., Liu, Y., Xiao, W., Chao, H.-C., Chu, P., 2016. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors 2016, Vol. 16, Page 115 16, 115.

Polito, M., Digo, E., Pastorelli, S., Gastaldi, L., 2023. Deep Learning Technique to Identify Abrupt Movements in Human-Robot Collaboration 73–80.

Rivera, P., Valarezo, E., Choi, M.-T., Kim, T.-S., 2017. Recognition of Human Hand Activities Based on a Single Wrist IMU Using Recurrent Neural Networks. International Journal of Pharma Medicine and Biological Sciences 6, 114–118.

Rosso, V., Gastaldi, L., Pastorelli, S., 2022. Detecting Impulsive Movements to Increase Operators' Safety in Manufacturing. Mechanisms and Machine Science 108 MMS, 174–181.

Wong, T. T., Yeh, P. Y., 2020. Reliable Accuracy Estimates from k-Fold Cross Validation. IEEE Trans Knowl Data Eng 32, 1586–1594.

Xiang, L., Gu, Y., Gao, Z., Yu, P., Shim, V., Wang, A., Fernandez, J., 2024. Integrating an LSTM framework for predicting ankle joint biomechanics during gait using inertial sensors. Comput Biol Med 170, 108016.

Xu, X., Lu, Y., Vogel-Heuser, B., Wang, L., 2021. Industry 4.0 and Industry 5.0—Inception, conception and perception. J Manuf Syst 61, 530–535.

Zafar, M. H., Langås, E. F., Sanfilippo, F., 2024. Exploring the synergies between collaborative robotics, digital twins, augmentation, and industry 5.0 for smart manufacturing: A state-of-the-art review. Robot Comput Integr Manuf 89, 102769.