Detection of upper limb abrupt gestures for human–machine interaction using deep learning techniques

Journal of the Brazilian Society of Mechanical Sciences and Engineering (2024) 46:227
https://doi.org/10.1007/s40430-024-04746-9

TECHNICAL PAPER

Detection of upper limb abrupt gestures for human–machine interaction using deep learning techniques

Elisa Digo[1] · Michele Polito[1] · Stefano Pastorelli[1] · Laura Gastaldi[1]

Received: 7 June 2023 / Accepted: 22 January 2024

Abstract
In the manufacturing industry the productivity is contingent on the workers' well-being, with operators at the center of the production process. Moreover, when human–machine interaction occurs, operators' safety is a key requirement. Generally, typical human gestures in manipulation tasks have repetitive kinetics, however external disturbances or environmental factors might provoke abrupt gestures, leading to improper interaction with the machine. The identification and characterization of these abrupt events has not yet been thoroughly studied. Accordingly, the aim of the current research was to define a methodology to ready identify human abrupt movements in a workplace, where manipulation activities are carried out. Five subjects performed three times a set of 30 standard pick-and-place tasks paced at 20 bpm, wearing magneto-inertial measurement units (MIMUs) on their wrists. Random visual and acoustic alarms triggered abrupt movements during standard gestures. The recorded signals were processed by segmenting each pick-and-place cycle. The distinction between standard and abrupt gestures was performed through a recurrent neural network applied to acceleration signals. Four different pre-classification methodologies were implemented to train the neural network and the resulting confusion matrices were compared. The outcomes showed that appropriate preprocessing of the data allows more effective training of the network and shorter classification time, enabling to achieve accuracy greater than 99% and F1-score better than 90%.

Keywords Human–machine interaction · Collaborative robot · Manipulation · Deep learning · MIMUs · Abrupt movements

1 Introduction

In today's manufacturing industry, it is increasingly common to perform operations in which a collaborative robot, or more generally an automated machine, cooperates with operators, sharing the workspace at the same times. This coexistence necessitates specific emphasis to security issues in physical human–machine interaction, according to Industry 5.0 paradigm that features workers at the center of the production process [1]. Indeed, the human–machine interaction is becoming an emerging research field focused on optimizing the effectiveness, applicability, and performance of working conditions [2–5].

Reaching [6], pick-and-place [7], and assembly [8] are examples of typical industrial collaborative tasks, where humans' gestures are characterized by repetitive and controlled movements, mainly of upper limbs, whose kinetics is consistent with a standard operational production protocol. However, inattention, unexpected situations, and unforeseen contingencies might cause the operator to perform abrupt gestures that deviate from standard movements, leading to possible unsafe and improper interaction with the machine. Tracking human motion during handling operations and prompt identification and characterization of abrupt events are of paramount relevance for a safe and efficient human–machine collaboration.

Different technologies can be adopted to track the human motion and possibly recognize the human gestures [7, 9–13]. Traditional optical systems allow direct tracking of human postures with high accuracy and precision. However, some characteristics of these systems such as encumbrance, possible occlusion, and high cost do not make them well suited for industrial settings. On the contrary, inertial measurement sensors are low-cost, portable, easy to wear, and minimally invasive [7, 11–14]. When human motion recognition is

Technical Editor: Marcelo Areias Trindade.

✉ Elisa Digo
elisa.digo@polito.it

[1] Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Corso Duca Degli Abruzzi 24, Turin, Italy

Published online: 16 March 2024

performed through inertial data, actions can be classified and labelled based on motion patterns arising from accelerometers and gyroscopes. In this context, deep learning technique is emerging as a major component of research in artificial intelligence, in light of its capacity to generalize across problems, its ability to scale with input data [15], and to extract and learn features directly from raw data [16]. Many studies have exploited inertial sensors combined with deep learning techniques to recognize the human upper body motion [16–20]. The human motion recognition is easier when tasks have predictable kinetics, but abrupt movements occurrence could jeopardize prompt tracking accuracy and decrease effectiveness and safety in human–robot shared task execution [21]. Abrupt gestures are highly variable and characterized by uncertain kinematic and dynamic ranges, execution patterns, and number of involved human segments [22]. Castellote et al. [23] analyzed the voluntary reaction to a startling auditory stimulus in tasks in which the main requirement is the accuracy. Their results demonstrated that this kind of stimulus speeds up only the first part of the movement. Kirschner and colleagues [24] defined a human involuntary motion as a rapid hormonal reaction resulting in a fast and uncontrolled movement. In this case, unexpected robot motions were performed to cause human involuntary movements. The results of this work demonstrated that hands-on user training can increase cognitive and physical safety in human–robot interaction. Görür and colleagues [25] have proposed a robot decision-making framework to anticipate human's task related availability, intent, capability, and true need for help during the collaboration. Authors have found out the ability of that system to ensure a proper task achievement, while reducing the number of unnecessary information given to the human. In the work of Rosso and colleagues [22], the focus was the estimation of four features based on the kinematics of impulsive gestures. In detail, a methodological study was developed to characterize this kind of movements through an inertial sensor fixed on the wrist. Although all these works have assessed the characteristics and the effects of abrupt movements on task performance, methods for the ready identification of these events have not yet been effectively studied.

The main question of the research project referred to in this paper is: by tracking the kinematics of the operator's wrists in a typical handling task, is it possible to readily distinguish between standard and abrupt gestures through deep learning technique? To give an answer, an experimental workplace setup has been designed to study pick-and-place tasks executed by five participants wearing inertial sensors on their upper body. Both standard and abrupt movements were performed during trials. A recurrent neural network was implemented to analyze acceleration signals recorded by the inertial sensors. Four different data pre-classification methodologies to train the neural network were tested, with the overall aim of increasing its performance toward effective real-time gesture classification.

2 Materials and methods

2.1 Experimental set-up and protocol

A workstation was simulated to execute a typical industrial pick and place task. As shown in Fig. 1, the set-up was composed of a box containing 30 golf balls and four stations: S0, S1, S2, and S3. Stations S0, S1 and S2 were at the table level, while S3 was at a height of 30 cm with respect to the table.

Participants were asked to sit in front of the table, to pick a ball at a time from the box, and to place it into a specific station hole, whose sequence was defined by the lighting of green LEDs positioned near each station (standard movements). Each task was composed of 30 pick-and-place gestures (Fig. 2a). During the task, acoustic or visual alarms were randomly generated through the activation of a sound buzzer and the lighting of red LEDs, respectively. When this situation occurred, participants were instructed to perform an abrupt movement. In detail, when the acoustic alarm was generated, participants were asked to vertically extend the arm as fast as feasible (Fig. 2b). Instead, when a visual alarm

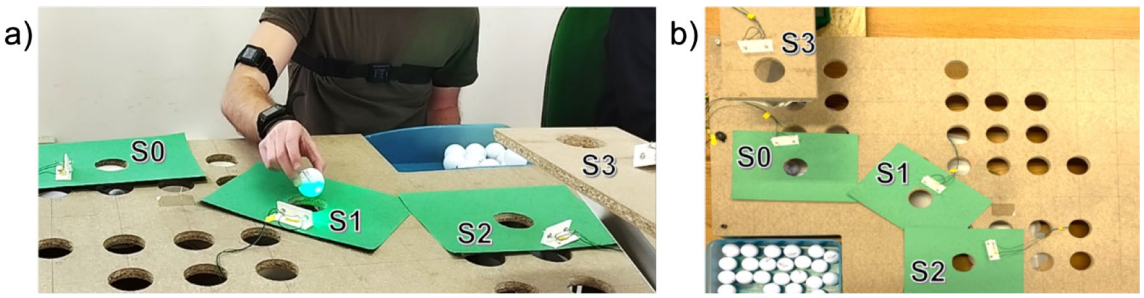

Fig. 1 Experimental set up: **a** 3D view; **b** top view

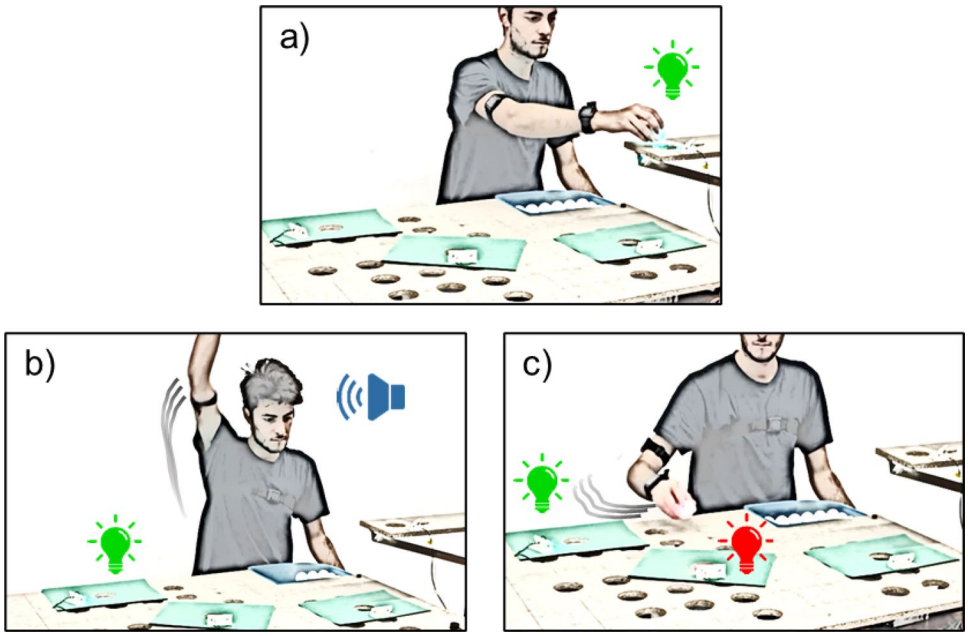

Fig. 2 Experimental protocol: **a** standard movement – green LED; **b** abrupt movement – acoustic alarm (buzzer); **c** abrupt movement – visual alarm (red LED)

was generated, participants were instructed to place the ball in the station corresponding to the activated red LED as fast as possible (Fig. 2c). For each participant, the task was repeated three times: Trial 1 with the right hand and with the trunk frontal with respect to the table; Trial 2 with the left hand and with the trunk frontal with respect to the table; Trial 3 with the left hand and with the trunk lateral with respect to the table. Inside each task, four random alarms were emitted, two visual and two sonorous.

The experimental set-up was adapted to the anthropometric characteristics of each participant. To this purpose, a board with holes (diameter of 6 cm) was made and positioned on the table to define the stations S0, S1, and S2. Moreover, a second board with just one hole (diameter of 6 cm) was added to the set-up to create the S3 station (Fig. 1). In detail, the more appropriate holes for S0, S1, and S2 stations were chosen inside a related group of 10 distributed holes. Instead, the distance of S3 station from each participant was regulated before the test.

The sound buzzer generating the acoustic alarms was fixed in the top left area of the table, whereas a pair of LEDs (one green and one red) was positioned near each station.

All the signals (lighting of green LEDs, activation of acoustic and visual alarms) were generated and controlled by an Arduino Nano microcontroller (Arduino, Italy), with processor ATmega328, clock speed of 16 MHz, and operating voltage of 5 V.

The code was written through an integrated development environment. The block diagram of the code is reported in Fig. 3. The first part of the code controlled the lightening of a green LED, corresponding to a station, every 3 s (20 bpm). It was repeated for 30 cycles, corresponding to the 30 golf balls that had to be picked-and-placed.

The first 6 cycles were standard, while for the cycles between the 7 and 30th, four cycles were randomly identified in which to activate twice the acoustic and visual alarms. The acoustic alarm was activated in correspondence of the green LED lighting, whereas the visual alarms were generated 500 ms after the green LED was turned on. Moreover, a control was implemented to avoid the overlap of visual and acoustic alarms. The sequence of green LEDs and of the alarms was updated and saved.

Five healthy participants (3 males and 2 females) with no musculoskeletal or neurological diseases were recruited for the experiment through a written informed consent. Their anthropometric data (mean ± standard deviation) are reported in the following: age of 37.8 ± 15.8 years, Body Mass Index of 20.7 ± 0.9 kg/m^2, upper arm length of 0.34 ± 0.03 m, and forearm length of 0.27 ± 0.01 m. Three subjects were right-handed, two were left-handed. The study was approved by the Local Institutional Review Board and all procedures were conformed to the Helsinki Declaration. Five wireless magneto-inertial measurement units (MIMUs) of an inertial sensors system (Opal™ APDM, USA) were positioned on participants' right upper arm (RUA), right forearm (RFA), sternum (STR), left upper arm (LUA), and left forearms (LFA) through bands supplied by the APDM kit (Fig. 4). Each sensor contains a tri-axial accelerometer (range ± 200 g), a tri-axial gyroscope (range ± 2000 deg/s), and a tri-axial magnetometer (range ± 8 Gauss). These units were positioned aligning their x-axes with the longitudinal axes of the corresponding human segments, while the y and z-axes were oriented as reported in Fig. 4. MIMUs

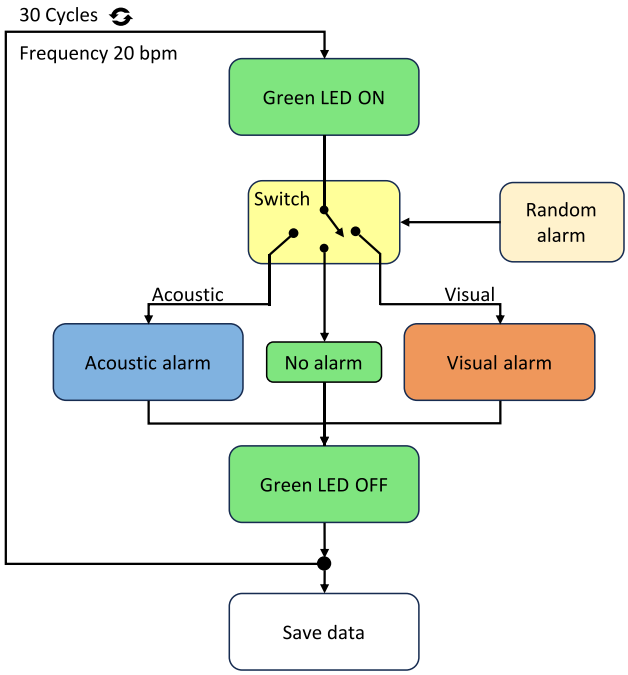

Fig. 3 Flow-chart of the developed Arduino code

communicated to a PC via Bluetooth. Data were acquired through the proprietary software Motion Studio™ (APDM, USA) at the sampling frequency of 200 Hz. The synchronization of the adopted systems was obtained with a voltage trigger of 5 V sent from Arduino to an input port of the Opal Synch Box (APDM), through a BNC connector.

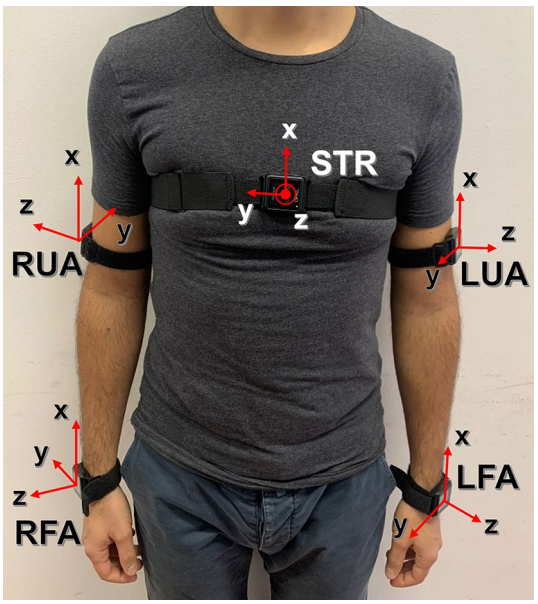

Fig. 4 MIMUs positioning on upper body and their reference frames

2.2 Data analysis

Data analysis was conducted with custom Matlab® routines on a machine with the following characteristics: Intel Core i7-11850H (8 core, from 2,50 GHz to 4,80 GHz), NVIDIA T600 (4 GB, GDDR6). The observation of collected signals showed a more significant involvement of forearms due to the task selected for the experiment. Accordingly, in this work, the analysis concentrated only on forearms accelerations. Accelerations measured with RFA and LFA MIMUs were postprocessed to remove constant gravity acceleration and then to estimate the magnitude of accelerations. Single movements were extracted segmenting the total acquisition into 3 s windows. Since the signal was well-paced by the lighting of the green LEDs at 20 bpm, the onset of each cycle was accordingly identified. Windows corresponding to abrupt movements were detected by the turning on of alarms. Considering the standard movements, the association between each LED and a specific station (S0, S1, S2, and S3) was recognized and windows were divided into four groups. The mean values and the standard deviation (std) of forearms accelerations were calculated for each trial and each station, first intra-subject and then inter-subjects.

Once the segmentation was performed, a Long Short-Term Memory recurrent neural network [26] was adopted to distinguish between standard movements and abrupt movements (Fig. 5). The network had the following characteristics: 1 input layer, 1 hidden layer of 100 hidden units, 2 output layers, Adam optimization, 100 epochs.

Four different methodologies were adopted to pre-classify movements to be fed to the neural networks:

(a) N3A3 methodology. Each three-second window was labeled as standard or abrupt, whether the alarm turned (Fig. 6a). For each trial of each subject, this windowing led to 26 standard movements and 4 abrupt ones.
(b) N1A1 methodology. All windows, both standard and abrupt, were divided into three sub-windows of one second (Fig. 6b). Considering all the trials for each subject, this led to 78 standard movement windows and 12 abrupt ones for each trial.
(c) N1A1S methodology. Each window was divided into three sub-windows of one-second duration each. Moreover, an additional condition was introduced for the potential abrupt windows by evaluating the std value with respect to a threshold. In detail, only sub-windows with standard deviation above 1.5 m/s^2 were considered as abrupt, while the other ones were excluded from the classification (Fig. 6c). This led to 78 standard movements and 12 potential abrupt ones, for each trial. Considering all subjects and all trials, a total of 1170 standard movements and 180 potential abrupt movements were isolated. Based on the standard deviation

constraints, only 137 unexpected movements were considered as abrupt movements.

(d) N1A1SO methodology. It was obtained applying the N1A1S methodology, but with some changes in the recurrent neural network (Fig. 6c). In detail, the number of hidden units was reduced from 100 to 30. As a consequence, the time of classification was reduced from 20 to 5 ms, making the methodology closer to a future real-time condition application. The amount of standard and abrupt isolated movements was the same of the N1A1S methodology.

For all methodologies, the label acceleration windows were divided in training (TR) and test (TE) sets, according to the following indications:

(a) For the N3A3 methodology, 120 windows were used for TR (60 standard, 60 abrupt), whereas 630 ones were used for TE (604 standard, 26 abrupt).

(b) For the N1A1 methodology, 120 windows were used for TR (60 standard, 60 abrupt), whereas 2130 ones were used for TE (2018 standard, 112 abrupt).

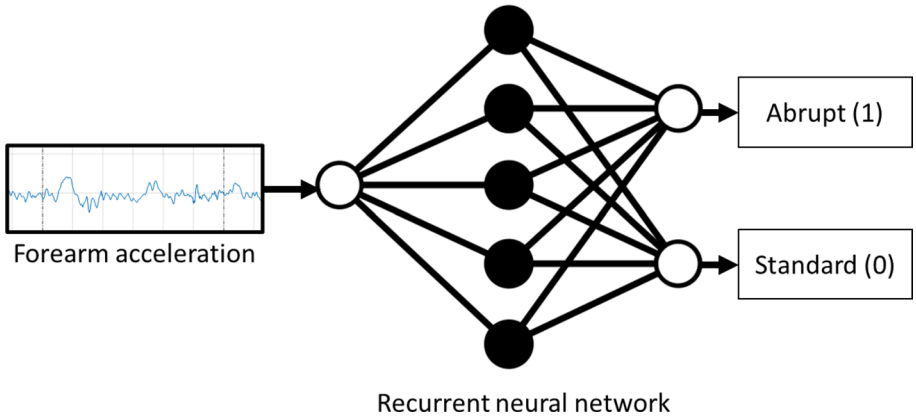

Fig. 5 Classification of gestures in abrupt and standard ones

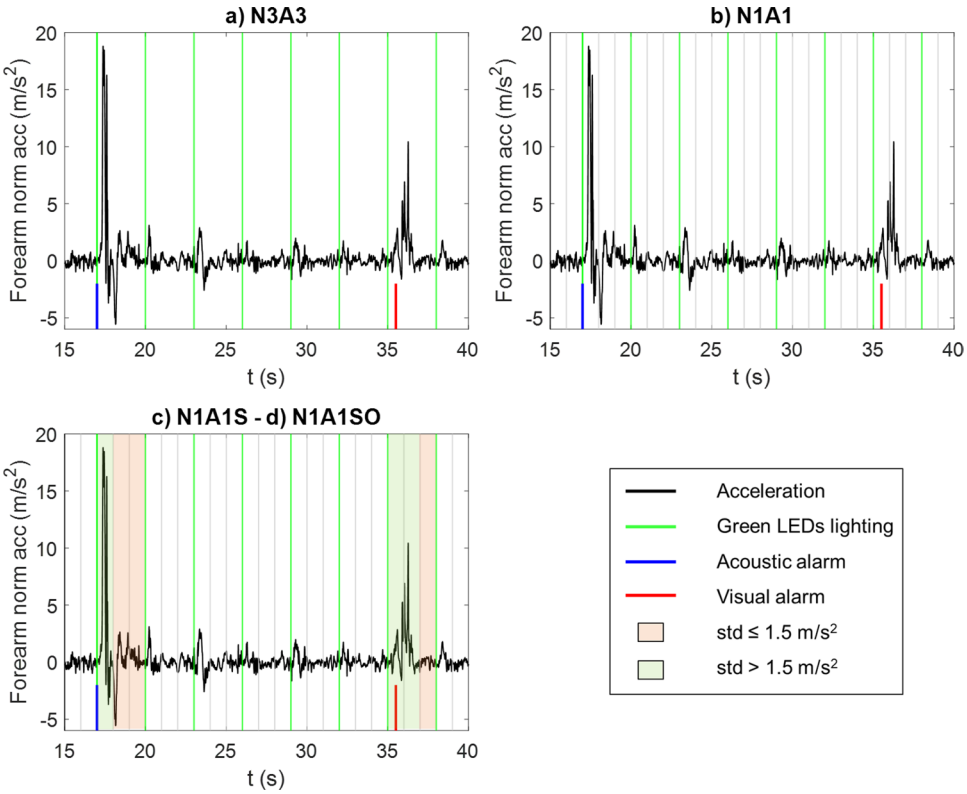

Fig. 6 Pre-classification methodologies. For a better understanding of the procedures, only a portion of the signal including two of the four generated abrupt movements is represented. **a** N3A3 methodology; **b** N1A1 methodology; (**c**–**d**) N1A1S and N1A1SO methodologies

(c) For the N1A1S methodology, 120 windows were used for TR (60 standard, 60 abrupt), whereas 2009 ones were used for TE (1932 standard, 77 abrupt).
(d) For the N1A1SO methodology, 120 windows were used for TR (60 standard, 60 abrupt), whereas 2009 ones were used for TE (1932 standard, 77 abrupt).

Once the classification was concluded with all the methodologies, the comparison between real and predicted classes was performed building confusion matrices [27]. Moreover, starting from the values of these matrices (true positives = TP, false positives = FP, true negatives = TN, false negatives = FN), the overall performance of the classification was quantified through the scores reported in equations from 1 to 4 [7, 28]. The accuracy represents an overall index of correct classification. The precision points out if a movement identified as abrupt is truly abrupt, whereas the recall quantifies the amount of identified abrupt gestures. The F1-score is the harmonic mean of precision and recall.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

3 Results

Inter-subject mean and standard deviation values of forearm accelerations are reported in Table 1 for each type of trial and for both standard and abrupt movements. Results related to the standard movements are reported separately for the four directions.

In Fig. 7, Trial 3 (with the trunk lateral with respect to the table) of the subject 05 is chosen as an explicative example, because one abrupt movement occurred in each direction. In detail, the average trends and standard deviation bounds of acceleration collected during standard movements are represented for each direction. In the same graph, the acceleration trends of the abrupt movements are overlaid for all directions.

Confusion matrices obtained from the classification results of the neural network for all methodologies are depicted in Fig. 8. Percentage values of accuracy, precision, recall, and F1-score calculated from confusion matrices for all the methodologies as indices of classification performance are reported in Table 2.

4 Discussions

The comparison between standard and abrupt acceleration trends (Fig. 7) highlights that the two signals are different in terms of shape, variation, and magnitude. Indeed, the amplitude of abrupt accelerations is definitively higher than the one of standard accelerations. In addition, as pointed out by the standard deviation for each direction (Fig. 7), subjects performed very repetitive standard movements, simulating typical industrial tasks. On the contrary, given the nonrepetitive property inherent in unexpected gestures, depending on subject, stimulus, and specific trial, a comparison among abrupt movements is meaningless. The same considerations can be made analyzing inter-subject mean and std values reported in Table 1. Indeed, accelerations related to standard movements (range 1.07–1.49 m/s^2) are on average equal to a third of the accelerations related to abrupt movements (range 3.17–3.54 m/s^2.

As described in literature [22, 29], abrupt movements are highly unpredictable and characterized by a shorter duration. Accordingly, this paper suggests a more complex tool for distinguishing between standard and abrupt movements. Data pre-processing influences classification results, as demonstrated comparing indices in Table 2 obtained from confusion matrices of N3A3, N1A1, and N1A1S methodologies (Fig. 8). In detail, the reduction of the window duration from 3 to 1 s (from N3A3 to N1A1) promotes a substantial increase of accuracy (+ 79.01%), precision (+ 70.18%), and F1-score (+ 62.77%) and a decrease of recall (− 20.60%). Hence, even if the N1A1 methodology is less sensitive than N3A3, it is more accurate and precise, and it promotes a

Table 1 Average value of forearms accelerations on windows of 3 s

Station	Standard acceleration mean ± std (m/s^2)				Abrupt acceleration mean ± std (m/s^2)
	S0	S1	S2	S3	S0–S1–S2–S3
Trial 1	1.31 ± 0.26	1.37 ± 0.14	1.29 ± 0.27	1.46 ± 0.22	3.54 ± 0.23
Trial 2	1.24 ± 0.18	1.49 ± 1.17	1.37 ± 0.22	1.30 ± 0.08	3.17 ± 0.27
Trial 3	1.07 ± 0.28	1.42 ± 0.20	1.21 ± 0.25	1.18 ± 0.23	3.26 ± 0.81

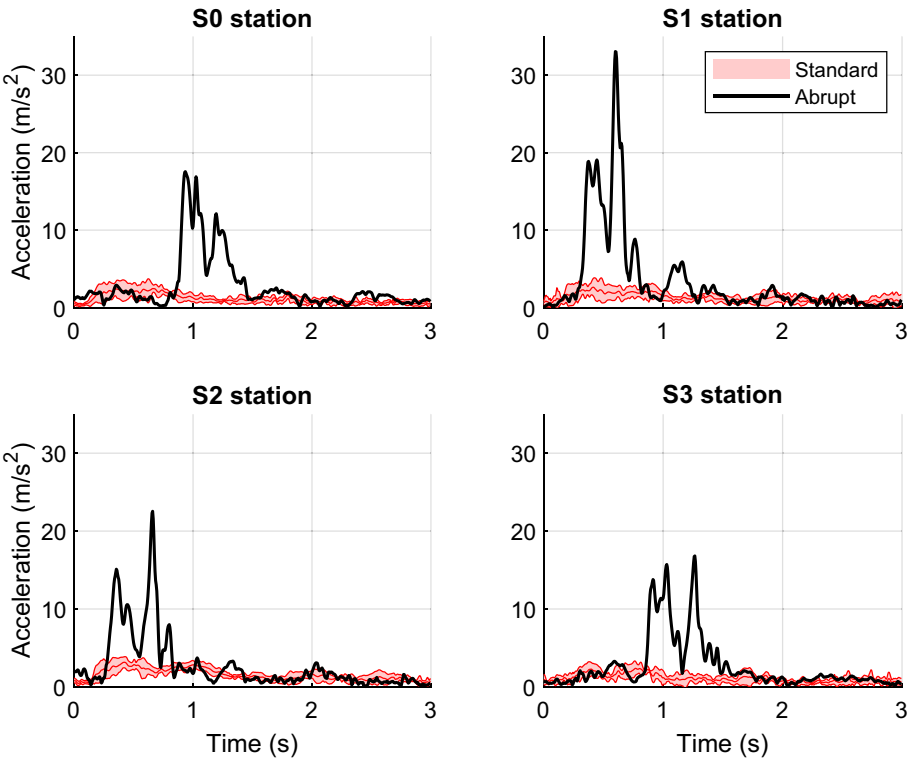

Fig. 7 Movement accelerations: standard average trend vs abrupt trend

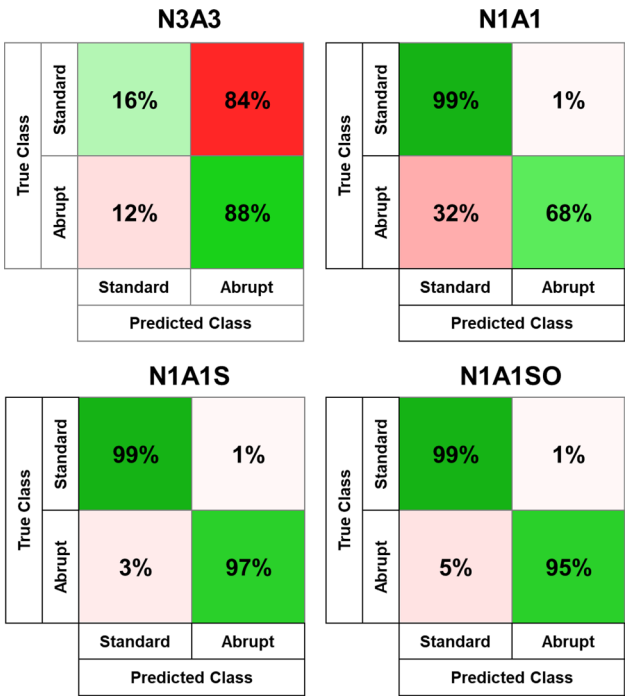

Fig. 8 Confusion matrices for the four tested methodologies

general improvement of classification (F1-score). Considering the introduction of a constraints on standard deviation and hence a finer pre-classification of abrupt gestures (from N1A1 to N1A1S), there is an increase of all indices (accuracy + 2.16%, precision + 10.72%, recall + 29.54%, and F1-score + 19.88%). Overall, an appropriate pre-processing of data allows obtaining a more effective network training and hence higher values on the principal diagonal of the confusion matrices (Fig. 8). In addition, since the recall is higher than the precision in almost all cases, it is more frequent that a standard movement is classified as an abrupt one than vice versa. This aspect is important for the safety of the operator in scenarios of collaborative robotics.

Another relevant aspect to consider in a context of collaborative robotics is the need to reach real-time conditions. The time necessary to distinguish between a standard and an abrupt movement is influenced by both the windowing procedure and the neural network performance. Moving from N1A1S to N1A1SO methodologies reduces the classification time of the network (from 20 to 5 ms) while generating a confusion matrix which is comparable to the one associated to N1A1S method (Fig. 8). This situation is also visible from values in Table 2: even if the precision has a slight increase

Table 2 Classification scores estimated from confusion matrices

Methodologies	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
N3A3	18.89	4.33	88.46	8.26
N1A1	97.09	74.51	67.86	71.03
N1A1S	99.25	85.23	97.40	90.91
N1A1SO	99.25	86.90	94.81	90.68

(+1.67%) and the recall has a slight decrease (−2.59%), the accuracy and the F1-score remain almost the same. This aspect suggests that the classification time can be reduced without negatively influencing the results.

The main limit of this study is represented by the restricted number of involved subjects, which is not appropriate when applying deep learning techniques, although partially mitigated by the large number of tests. In this regard, current activities consist in investigating other typical industrial tasks, also exploring the other collected signals in addition to forearms accelerations (angular velocities of forearms and signals of upper arms and trunk). Then, the same experimental campaign is going to be extended to around 100 subjects in order to have a homogeneous sample in terms of gender and age. Moreover, considering that the project is thought for industrial scenarios of collaborative robotics, current efforts aim reducing the duration of windows in which the signal is fragmented. This aspect can guarantee lower classification times and hence make the process closer to real-time feasibility.

5 Conclusions

Considering the importance of the operator's safety in the workspace shared with a robot or machine, the aim of this study was to identify possible human abrupt movements during a typical repetitive industrial task. A recurrent neural network was fed with forearms accelerations measured with MIMUs and exploited to distinguish between standard and abrupt gestures. All the results obtained in this work testify that the chosen deep learning network and the developed pre-classification methods for MIMUs accelerations are promising for identifying human abrupt movements in handling tasks. Current efforts are devoted to segment the signal into successively smaller windows to approach to real-time conditions while guaranteeing the same classification scores.

6 Supplementary Material

Authors will provide the acquired raw inertial data on request.

Acknowledgements Authors would like to thank Simona Anelli for her contribution in data collection and analysis.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. 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. https://doi.org/10.1016/j.jmsy.2021.10.006
2. Almusawi ARJ, Dulger LC, Kapucu S (2018) Online teaching of robotic arm by human–robot interaction: end effector force/torque sensing. J Brazil Soc Mech Sci Eng 40(9):1–14. https://doi.org/10.1007/s40430-018-1358-3
3. Jha A, Chiddarwar SS, Alakshendra V, Andulkar MV (2017) Kinematics-based approach for robot programming via human arm motion. J Braz Soc Mech Sci Eng 39(7):2659–2675. https://doi.org/10.1007/s40430-016-0662-z
4. Losey DP, McDonald CG, Battaglia E, O'Malley MK (2018) A review of intent detection, arbitration, and communication aspects of shared control for physical human–robot interaction. Appl Mech Rev 70(1):1–19. https://doi.org/10.1115/1.4039145
5. Melchiorre M, Scimmi LS, Mauro S, Pastorelli SP (2020) Vision-based control architecture for human–robot hand-over applications. Asian J Control 23(1):105–117. https://doi.org/10.1002/asjc.2480
6. Lin CL, Wang MJJ, Drury CG, Chen YS (2010) Evaluation of perceived discomfort in repetitive arm reaching and holding tasks. Int J Ind Ergon 40(1):90–96. https://doi.org/10.1016/j.ergon.2009.08.009
7. Digo E, Antonelli M, Cornagliotto V, Pastorelli S, Gastaldi L (2020) Collection and analysis of human upper limbs motion features for collaborative robotic applications. Robotics 9(2):33. https://doi.org/10.3390/ROBOTICS9020033
8. 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(1):5700–5705. https://doi.org/10.1016/j.ifacol.2017.08.1121
9. Weitschat R, Ehrensperger J, Maier M, Aschemann H (2018) Safe and efficient human-robot collaboration part I: estimation of human arm motions. In: Proceedings - IEEE international conference on robotics and automation, IEEE, 2018, pp 1993–1999. doi: https://doi.org/10.1109/ICRA.2018.8461190.
10. Wang Y, Ye X, Yang Y, Zhang W (2017) Collision-free trajectory planning in human-robot interaction through hand movement prediction from vision. In: IEEE-RAS international conference on humanoid robots, pp 305–310. doi: https://doi.org/10.1109/HUMANOIDS.2017.8246890.
11. Digo E, Gastaldi L, Antonelli M, Pastorelli S, Cereatti A, Caruso M (2022) Real-time estimation of upper limbs kinematics with IMUs during typical industrial gestures. Procedia Comput Sci 200(2019):1041–1047. https://doi.org/10.1016/j.procs.2022.01.303
12. Digo E, Antonelli M, Pastorelli S, Gastaldi L (2021) Upper limbs motion tracking for collaborative robotic applications. Adv Intell Syst Comput 1253:391–397. https://doi.org/10.1007/978-3-030-55307-4_59
13. Antonelli M, Digo E, Pastorelli S, Gastaldi L (2021) Wearable MIMUs for the identification of upper limbs motion in an industrial context of human-robot interaction. In: Proceedings of the 18th international conference on informatics in control, automation and robotics, ICINCO 2021, 2021, pp 403–409. doi: https://doi.org/10.5220/0010548304030409
14. Digo E, Pastorelli S, Gastaldi L (2022) A narrative review on wearable inertial sensors for human motion tracking in industrial scenarios. Robotics 11(6):138. https://doi.org/10.3390/robotics11060138
15. Sengupta S et al (2020) A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl Based Syst 194:105596. https://doi.org/10.1016/J.KNOSYS.2020.105596
16. Añazco EV, Han SJ, Kim K, Lopez PR, Kim TS, Lee S (2021) Hand gesture recognition using single patchable six-axis inertial measurement unit via recurrent neural networks. Sensors 21(4):1–14. https://doi.org/10.3390/s21041404
17. Jiang Y, Song L, Zhang J, Song Y, Yan M (2022) Multi-category gesture recognition modeling based on sEMG and IMU signals. Sensors 22(15):5855. https://doi.org/10.3390/s22155855
18. 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. Int J Pharma Med Biol Sci 6(4):114–118. https://doi.org/10.18178/ijpmbs.6.4.114-118
19. Luktuke YY, Hoover A (2020) Segmentation and recognition of eating gestures from wrist motion using deep learning. In: Proceedings - 2020 IEEE international conference on big data, big data, pp 1368–1373, doi: https://doi.org/10.1109/BigData50022.2020.9378382.
20. Kim M, Cho J, Lee S, Jung Y (2019) Imu sensor-based hand gesture recognition for human-machine interfaces. Sensors 19(18):1–13. https://doi.org/10.3390/s19183827
21. Devin S, Alami R (2016) An implemented theory of mind to improve human-robot shared plans execution. In: ACM/IEEE international conference on human-robot interaction. pp 319–326, 2016, doi: https://doi.org/10.1109/HRI.2016.7451768.
22. Rosso V, Gastaldi L, Pastorelli S (2022) Detecting impulsive movements to increase operators' safety in manufacturing. Mech Mach Sci 108:174–181. https://doi.org/10.1007/978-3-030-87383-7_19
23. Castellote JM, Valls-Solé J (2015) The StartReact effect in tasks requiring end-point accuracy. Clin Neurophysiol 126(10):1879–1885. https://doi.org/10.1016/j.clinph.2015.01.028
24. Kirschner RJ, Burr L, Porzenheim M, Mayer H, Abdolshah S, Haddadin S (2021) Involuntary motion in human-robot interaction: effect of interactive user training on the occurrence of human startle-surprise motion. In: ISR 2021 - 2021 IEEE international conference on intelligence and safety for robotics, pp 28–32, https://doi.org/10.1109/ISR50024.2021.9419526
25. Görür OC, Rosman B, Sivrikaya F, Albayrak S (2018) Social cobots: anticipatory decision-making for collaborative robots incorporating unexpected human behaviors. In: ACM/IEEE international conference on human-robot interaction, pp 398–406 doi: https://doi.org/10.1145/3171221.3171256.
26. Van Houdt G, Mosquera C, Nápoles G (2020) A review on the long short-term memory model. Artif Intell Rev 53(8):5929–5955. https://doi.org/10.1007/s10462-020-09838-1
27. Düntsch I, Gediga G (2019) Confusion matrices and rough set data analysis. J Phys Conf Ser 1229(1):012055. https://doi.org/10.1088/1742-6596/1229/1/012055
28. D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić, "Multi-label Classifier Performance Evaluation with Confusion Matrix," pp. 01–14, 2020, doi: https://doi.org/10.5121/csit.2020.100801.
29. M. Polito, E. Digo, S. Pastorelli, and L. Gastaldi, *Deep Learning Technique to Identify Abrupt Movements in Human-Robot Collaboration*, vol. 134 MMS. 2023. doi: https://doi.org/10.1007/978-3-031-32439-0_9.