

Article

# Action Recognition for Human–Robot Teaming: Exploring Mutual Performance Monitoring Possibilities

Shakra Mehak <sup>1,2,\*</sup> , John D. Kelleher <sup>3</sup> , Michael Guilfoyle <sup>1</sup>  and Maria Chiara Leva <sup>2,\*</sup> <sup>1</sup> Pilz Ireland Industrial Automation, T12 AW80 Cork, Ireland; m.guilfoyle@pilz.ie<sup>2</sup> School of Food Science and Environmental Health, Technological University Dublin, D07 H6K8 Dublin, Ireland<sup>3</sup> School of Computer Science and Statistics, Trinity College Dublin, D02 PN40 Dublin, Ireland; john.kelleher@tcd.ie

\* Correspondence: s.mehak@pilz.ie (S.M.); mariachiara.leva@tudublin.ie (M.C.L.)

**Abstract:** Human–robot teaming (HrT) is being adopted in an increasing range of industries and work environments. Effective HrT relies on the success of complex and dynamic human–robot interaction. Although it may be optimal for robots to possess all the social and emotional skills to function as productive team members, certain cognitive capabilities can enable them to develop attitude-based competencies for optimizing teams. Despite the extensive research into the human–human team structure, the domain of HrT research remains relatively limited. In this sense, incorporating established human–human teaming (HhT) elements may prove practical. One key element is mutual performance monitoring (MPM), which involves the reciprocal observation and active anticipation of team members’ actions within the team setting, fostering enhanced team coordination and communication. By adopting this concept, this study uses ML-based visual action recognition as a potential tool for developing an effective way to monitor the human component in HrT. This study utilizes a data modeling approach on an existing dataset, the “Industrial Human Action Recognition Dataset” (InHARD), curated specifically for human action recognition assembly tasks in industrial environments involving human–robot collaborations. This paper presents the results of this modeling approach in analyzing the dataset to implement a theoretical concept that can be a first step toward enabling the system to adapt dynamically. The outcomes emphasize the significance of implementing state-of-the-art team concepts by integrating modern technologies and assessing the possibility of advancing HrT in this direction.



**Citation:** Mehak, S.; Kelleher, J.D.; Guilfoyle, M.; Leva, M.C. Action Recognition for Human–Robot Teaming: Exploring Mutual Performance Monitoring Possibilities. *Machines* **2024**, *12*, 45. <https://doi.org/10.3390/machines12010045>

Academic Editor: Dan Zhang

Received: 5 December 2023

Revised: 27 December 2023

Accepted: 6 January 2024

Published: 9 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** human–robot interaction; human–robot collaboration; teaming AI; human–robot teaming; ML; human factors

## 1. Introduction

### 1.1. Human–Robot Teaming: A System Modeling Perspective

The concept of HrT depends on merging the capabilities of humans and robots. From a system modeling perspective, a human–robot team can be regarded as a system comprising two subsystems: a human subsystem and a robot subsystem. Each subsystem has its own characteristics and capabilities, and the interaction between these two subsystems determines the overall performance of the system. In order to model the human–robot team as a system, it is necessary to understand the capabilities and limitations of each subsystem, as well as how they can interact and cooperate to achieve a common goal.

Humans possess intrinsic flexibility, cognition, and problem-solving skills [1], whereas robots offer high accuracy, speed, and repeatability [2]. As the ability of robots to act intelligently and the potential for them to be installed without cages increases, manufacturers have developed guidelines and design criteria such as autonomy and mechanical design [3–6]. The purpose of these guidelines and design criteria is to ensure the safety and reliability of the system. However, HrT brings unique challenges to these guidelines. Due to cognitive/computational and physiological differences between robots and humans [7], robots

need to be programmed with the ability to predict and comprehend the tasks/intentions of human teammates. Humans acquire the ability to predict the behaviour of other humans over time [8], but robots must be explicitly trained on how to do this. In addition, when confronted with the volatility of real-world applications, robots cannot rely on a human teammate to stick to the defined task [9], nor can they consistently predict how their human partner would act when something goes wrong. The potential solution to this problem is to equip the robot with explicit models of their human teammates. Ideally, these models should learn the generalized features of human behavior without requiring individuals to act in a certain manner. To this end, machine learning (ML) offers emerging techniques for constructing cognitive models and behavioral blocks [10], providing a vital perspective for human–robot collaborative activities. By modeling the human–robot team from a system perspective, it is possible to gain a better understanding of how the team functions as a whole and identify potential improvements or enhancements to the system.

The human–robot interaction (HRI) field has advanced, and the need for applications requiring humans and robots to form a team and collaborate to solve complicated issues has increased [11,12]. Therefore, addressing the teaming configurations and elements and showing the path to incorporate them using advanced algorithms is necessary. However, determining the optimal teaming configuration for HrT can be complicated, particularly due to the unpredictable nature of HRI, as well as the diverse range of tasks and environments that human–robot teams may encounter.

### *1.2. Trust in Human–Robot Teaming*

Trust plays a vital part in collaborative efforts, particularly in the context of HrT. In teams where tasks are related and mutually dependent, the efficacy relies primarily on the trust between the human and robot team members. This mutual trust is a cornerstone of teamwork and is important for the success of HrT [13].

A human's trust in a robot involves a belief in the robot's proficiency [14]; as demonstrated by the robot's capacity to understand and conform to human preferences and the ability to make a valuable contribution to a common objective. In exploring the factors that influence trust in HRI, Hancock et al. [15] highlighted that various robot characteristics and performance-based factors are particularly crucial. Their research suggests that trust in HRI is most significantly influenced by aspects related to the robot's ability and performance. Therefore, manipulating and improving these performance aspects can have a substantial impact on the level of trust established between humans and robots.

In order for robots to establish trust and work effectively alongside human teammates, we argue that gaining an understanding of human behavior would indeed enhance the robot's adaptability and behavior, which are key performance factors. This would involve analyzing contextual patterns and predicting human task performance based on this analysis. By doing so, robots can be better equipped to anticipate the needs and actions of their human counterparts, leading to trust building and, ultimately, effective teaming. The study by Hancock et al. [15] stated that higher trust is associated with higher reliability. Therefore, the implication is that if a capacity for performance monitoring in a robot is to be developed, it also needs to be highly reliable; otherwise, it could potentially have a detrimental effect on the overall HRI trust dynamic. Hancock et al. included team collaboration characteristics and tasking factors as relevant factors of the HRI trust dynamic. However, further specification of the team- and task-related effects could not be provided because of an insufficient number of empirically codable studies. To gain more insights, additional research experiments are required in this field.

Overall, trust in HrT is a dynamic element that should facilitate smooth interactions, effective task execution, and adaptive coordination between human and robot team members. The dataset we used for this preliminary analysis did not contain data regarding trust and performance from the perspective of the human operator, as it was collected solely for providing action recognition capabilities. However, as stated in our study, it is considered a stepping stone for future research.

### 1.3. Translating Human–Human Teaming Characteristics into a Human–Robot Teaming Setting

Teamwork is a collection of each team member's interlinked thoughts, behaviors, and emotions required for the team to function [16]. Teammates offer emotional support; individuals usually feel more assured when communicating with others who share a similar experience. Moreover, humans enjoy the sense of belonging that comes with being a team member, especially in a well-functioning team [17].

The main elements of teamwork include coordinating teammates' tasks, anticipating needs based on knowledge of assigned roles, adaptability, team orientation, backup behaviors, closed communication, and mutual performance monitoring. The aforementioned teaming elements are extracted from the Big Five teamwork model, a theoretical foundation for implementing team learning theory in HhT structures [18].

Despite the fact that many researchers have discussed teaming features in the context of HhT, relatively few studies [9,19–25] have underlined their importance in HrT. Incorporating HhT features into a HrT framework may increase the team's efficacy and efficiency. By applying these elements, the team members may be better able to understand each other's strengths and weaknesses, communicate more effectively, and work together toward shared goals and objectives. Moreover, the concept may improve overall productivity and reduce the risk of errors or miscommunication. These teaming components are seen as universally applicable to collaborative processes [18]. It is imperative to adapt these elements to optimize the efficacy of HrT. To this end, we have reinterpreted these HhT elements in the context of HrT and identified current methods that hold promise for prospective implementation while also highlighting the distinct challenges associated with these approaches:

Team leadership roles can be dynamically allocated to robots or humans depending on the task. For instance, a human can lead a task requiring creativity and decision making, and a robot with advanced data processing capabilities can lead a task requiring large data analysis. In this regard, multi-agent systems [26] can be employed to assign roles. However, the challenge lies in human safety, and humans may resist being led by a robot.

Team orientation is important to align the objectives of both human and robot team members. The field of social robotics can be instrumental in this regard. Utilizing Natural Language Processing (NLP) models and social signal processing can equip robots with social intelligence [27]. However, the implementation of NLP models requires substantial data and model training.

Backup behaviors involve team members supporting each other when required. In HrT, robots can be programmed to support human team members in tasks that are hazardous or difficult for humans. Scheduling algorithms [28] can be used to optimize backup behaviors in HrT. Additionally, multi-agent systems [26] can be used to develop a backup behavior framework. Developing algorithms that can assess the needs of human team members and switch tasks accordingly requires detailed design and testing to ensure functional team collaboration.

Adaptability is a key aspect that entails the ability of a team to adjust to environmental or task-related transformations. To expedite adaptation among team members, shared mental models are employed to modify task requirements. Advanced ML methods, such as reinforcement learning [29] and online optimization algorithms [30], can enable robots to adapt to environmental changes. However, the challenge lies in ensuring that the adaptation is timely, contextually appropriate, and aligns with the preferences of the human team members.

Mutual performance monitoring involves understanding the performance of team members to ensure that the team is working toward the goals. This requires a comprehensive understanding of each teammate's actions or tasks. Sensors and real-time data analytics can be employed to monitor teammates' actions. Robots can be equipped to interpret teammates' actions by integrating cognitive abilities using ML algorithms [31]. The development of reliable algorithms demands a significant amount of training data and

involves feature selection, model selection, and hyperparameter tuning to optimize the performance of these algorithms.

However, the potential implementation of these teaming elements in HrT should be investigated separately. This article considers the implementation of *MPM* as a key missing ingredient in HrT structures.

#### 1.4. The Need for Mutual Performance Monitoring in Human–Robot Teaming

Fundamental to the development of HrT is the question of how well robots can engage in implicit collaboration, which is the process of synchronizing with team members' actions based on understanding what each teammate is most likely to do or is doing. It has always been challenging for team researchers to specify cognitive skills or attributes that enable teams to engage, adapt, and synchronize task-relevant information [32]. However, researchers have discovered that *MPM* is a practical teaming component that contributes to successful teaming [17]. *MPM* is the capacity to monitor the tasks of other teammates while executing one's own tasks [18]. It is an important aspect of teaming that can be characterized in the context of HrT as the reciprocal understanding and monitoring of team actions, progress, and outcomes. Here, we consider the innate human ability to understand and interpret teammates' actions and apply this concept to the team relations between humans and robots. However, it is important to note that although robots may lack the natural capabilities humans possess, they have the potential to acquire them through artificial intelligence (AI) and programming. Through *MPM*, team members can gain valuable insights into their intentions and task-related challenges, enabling effective communication and shared decision making for successful task accomplishment. Although humans possess the innate ability to interpret their teammates' actions or intentions [11], not all social skills may be necessary for robots. However, certain task-related cognitive abilities can empower robots to function as effective team members in collaborative settings. Further, teaming abilities involve robots' responsiveness to adaptable team dynamics and individual preferences. Robots should be equipped with the ability to adjust their behavior, communication style, and task allocation to their human teammates' preferences and work styles [9]. In this regard, the modeling of *MPM* can be a step forward in achieving adaptability, allowing for enhanced teamwork and collaboration.

To this end, this study proposes the implementation of action recognition as a pivotal method for realizing *MPM* in human–robot team environments. Action recognition is a field of ML/computer vision that involves the identification and classification of human actions from video data. It has emerged as a sophisticated and promising approach in contemporary research due to its potential applications in various domains, such as surveillance and human–computer interaction [33]. In the literature, “human action” is often translated as visual action [34], highlighting the significance of visual cues in this field. However, action recognition goes beyond just visual aspects and encompasses a broader range of modalities, such as depth information, pose estimation, and other modalities, enabling a more comprehensive understanding of human actions in diverse scenarios [35]. The process of action recognition involves several steps, including feature extraction, representation learning, and classification. Integrating action recognition into *MPM* provides a data-driven and objective framework for assessing teammates' actions. The system can understand their objectives and preferences by developing a computational algorithm that translates human teammates' actions into a trained model, enabling the robot counterpart to effectively plan and execute subsequent tasks. However, selecting an appropriate ML algorithm [33], developing an efficient feature extraction or learning process, and optimizing model training and evaluation are the foremost considerations. Moreover, it is imperative to consider additional factors, like computational cost and the ability to process data in real time, to guarantee the system's feasibility and efficacy [30]. This involves selecting algorithms and designing processes that balance computational efficiency and recognition accuracy, thus facilitating real-time interactions in dynamic environments without compromising performance. In particular, implementing a well-



designed visual action recognition system is crucial for enabling a robot to accurately comprehend its teammates' actions. Thus, this paper seeks to address the following research questions:

- RQ1: Can action recognition be used in human–robot teaming to pave the way for an important element of human teamwork, i.e., *MPM*?
- RQ2: What types of sensor data and ML algorithms can be deployed for its practical implementation?

This study adopts an ML paradigm to realize the conceptual framework and explore its execution using an existing dataset (InHARD) [36] to address our research questions, allowing us to carry out this study as a feasibility study. The organization of this paper is as follows. Section 2 presents a comprehensive literature review. Section 3 discusses the conceptual framework, the dataset adopted along with its experimental design context, and the approach employed for the model implementation and configurations. Section 4 explains the model results, discusses the limitations, and presents the findings obtained. Finally, Section 5 concludes this paper by highlighting its contributions and future implications and suggesting potential research studies in this direction.

## 2. Related Works

Previous studies have explored diverse facets of HrT and their significance in improving teaming performance and mitigating errors in collaborative tasks. The broader topics of HrT discussed in the literature include trust, the shared mental model, task or role allocation, communication, and team coordination and adaptability. These elements of HrT play an imperative role in promoting effective HrT. Yet, the introduction and implementation of *MPM* remains an unexplored teaming component within HrT settings.

There is extensive literature on communication, coordination, and adaptability in HrT [20,37,38]; core challenges, including communication modality and frequency, still need to be resolved. For instance, a study [39] on explanation-based communication was conducted to evaluate four communication conditions, and the experiments indicated the need for more robust and efficient communication strategies for successful teaming. Another component of effective teaming is encapsulated by trust, a multifaceted element in collaborative work. The authors defined trust in the context of teamwork as the assured dependence on a team member to accurately execute actions, even in uncertain and risky circumstances, without the need for surveillance or control [40]. Numerous studies have underscored the importance of trust in developing collaboration and guaranteeing the effortless performance of collaborative tasks within HrT environments [8,23,41]. For instance, the authors of [41] introduced the Trust Inference and Propagation (TIP) model, concentrating on trust modeling in multi-human multi-robot teams. The model evaluates both direct and indirect experiences with robots and theoretically proves that trust develops through repeated interactions.

In human–robot team studies, researchers have sometimes used shared mental models to measure situational awareness in human–robot teams [17,42,43]. One of these models was presented in [43], where a capability-aware shared mental model (CASMM) was introduced that uses tuples to break down tasks into sets of procedures related to complexities. The model dynamically combines the task grouping ideas extended by humans and the AI model via negotiation, fostering a better cohesive mental model among teammates. Concurrently, in [44], the authors explored the importance of achieving a common, meaningful, and timely understanding of the context in which humans and machines act and interact. The study investigated how AI-related approaches of belief and reasoning based on ontologies can enable knowledge sharing among all team members, both human and machine, thereby attaining a high level of interoperability between heterogeneous entities. However, upon critical analysis, these studies often assume that agents intrinsically possess the ability to adopt shared mental models of task routines when collaborating in a team. In practical scenarios, the assurance of a shared mental model may not always be guaranteed, particularly in ad hoc teams [45], where agents might adhere to divergent patterns. Further-

more, measuring situational awareness in human–robot teams poses distinct challenges. The inherent constraint of robotic counterparts in expressing their beliefs and perceptions to their human teammates adds another layer of complexity to this evaluative process.

### *Research Gap*

Although extensive research has been conducted on different aspects of HrT, a significant gap exists in developing reliable HrT methodologies that can adapt to the diverse and ever-changing nature of human–robot teams. Existing studies predominately focus on optimizing task allocation [46], refining interaction mechanisms [4], or offering theoretical perspectives [23]. However, current studies often disregard equipping robots with cognitive abilities using data modeling techniques, disregarding the differences between human and robot team members. There is a need for more comprehensive research that considers the distinct perspectives of robot agents in HrT. Specifically, a step we consider important is the capability of robotic agents to monitor and recognize tasks carried out by human team members.

## **3. Materials and Methods**

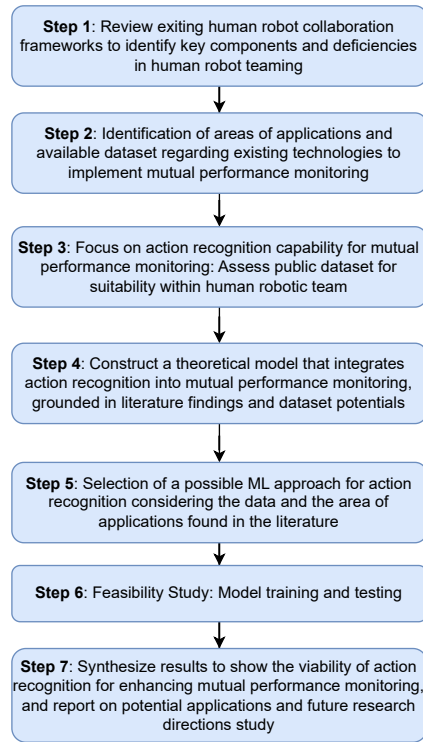
This study presents the framework for implementing *MPM* within the HrT paradigm. Figure 1 outlines our methodological approach to studying *MPM* in human–robot teaming, starting from the review of the original HhT framework to the final synthesis of data-driven insights. This study begins with an integrated review of existing human–robot collaboration frameworks and relevant literature on action recognition. This foundation facilitates the process of identifying and evaluating public datasets and technologies relevant to our research aims. The subsequent steps in Figure 1 explain our process of developing a theoretical model, selecting appropriate ML algorithms, and analyzing data to derive meaningful insights into the dynamics of *MPM* implementation within HrT.

The following section discusses the details of our methodology.

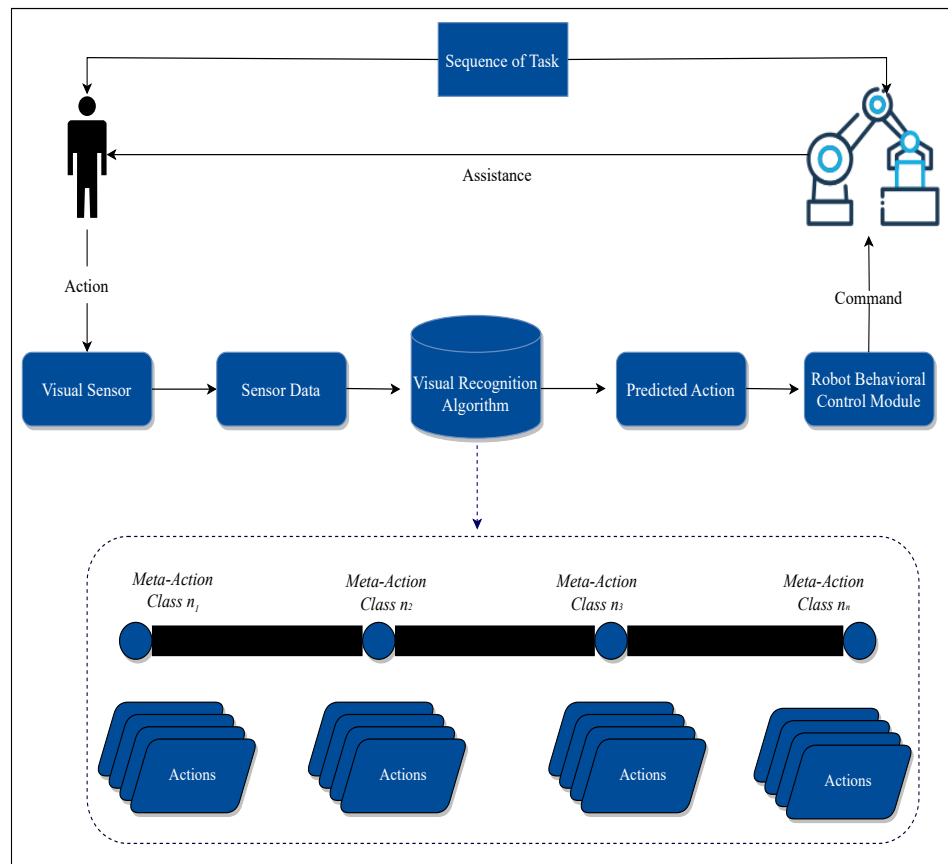
### *3.1. Conceptual Foundation*

To implement action recognition within the context of *MPM*, we conceptualize a scenario in which a human and a robot collaboratively work on a manufacturing task, such as assembling a complex product. The assembly station has vision sensors, a robot, and a human teammate. The robot is connected to a vision system that monitors human tasks and predicts actions through a predictive model, as illustrated in Figure 2. The model used in the task recognition system anticipates actions based on observed action sequences and contextual information. As the task progresses, the task recognition algorithm continuously analyzes the vision inputs to track human actions. The system identifies the actions, recognizes the sequence of task steps being executed, and assesses the quality and progress of each step. Moreover, if the system predicts that a human team member will require a tool or assistance in the next step, it can prepare the necessary resources or communicate the need to the robot, ensuring a smooth workflow. To validate this concept, we explore the potential of visual recognition as a powerful tool for realizing *MPM* in practical HrT scenarios.

The following sections discuss the data description and modeling of *MPM*, using industrial datasets and deep learning methodologies.



**Figure 1.** Methodological framework for studying mutual performance monitoring in human–robot teaming.



**Figure 2.** Mutual performance monitoring in human–robot teaming with task/action recognition. Visual sensors capture real-time human actions, facilitating data transfer to the computational algorithm for human action recognition.

### 3.2. Dataset Description and Experimental Design

The dataset used in this study is the “Industrial Human Action Recognition Dataset” (InHARD) [36], curated specifically for human action recognition assembly tasks in industrial environments involving human–robot collaborations. The dataset was selected for its comprehensive multi-modal sensor data and diverse HRI scenarios, which are essential for accurately modeling and analyzing complex HrT dynamics. The dataset’s extensive annotations and metadata provide insights for developing an exact understanding and predictive capacity in MPM.

In the InHARD dataset, participants are tasked with assembling a component by following sets of instructions and using the UR10 robotic arm, screws, hooks, and tools like a screwdriver. The data used in this study were collected using image-based cameras. The dataset contains over 2 million frames and fourteen industrial action classes gathered from 16 distinct individuals, providing more than 4800 diverse action samples captured across 38 videos. To ensure comprehensive coverage of actions and address occlusion challenges, the videos were captured from three different viewpoints: a top view, a left view, and a right view. Each frame in the video is annotated with the action performed by the subject. The data labeling is performed at the frame level, with each frame assigned an action label. The assembly task consists primarily of seven operations, each requiring approximately 15 actions. Throughout a complete assembly, the human performs between 100 and 180 actions, each taking between 0.5 and 27.0 s [36].

The number of data samples varies across the classes, with some having a higher representation than others (Table 1). In order to address the issue of class imbalance caused by fewer data samples in certain classes, such as the *Take Subsystem*, we employed the temporal cropping data augmentation [47]. This technique randomly selects fixed-duration segments from the video sequences, generating multiple crops for the respective class. For the final implementation, we considered only 450 samples for all classes. To prepare the dataset, we transformed it into a feature vector using one-hot encoding labels. Subsequently, we divided the dataset into two arrays: one for training, which accounted for 80 percent of the data, and another for model testing, which comprised the remaining 20 percent. We used stratified sampling when selecting the test data to ensure that the training and test sets had the same class label distributions. This partitioning enabled us to effectively assess the performance of the trained model. For model evaluation, we used accuracy and a confusion matrix [31,48] to understand the model’s performance and ability to correctly classify instances.

Our experiments used a hardware setup comprising an Intel Core i9-9900X processor, 64 GB of RAM, and two NVIDIA Geforce RTX 2080TI graphic cards.

**Table 1.** Number of samples for each class.

Meta-Action Class Label	No. of Samples
Assemble System	1378
Consult Sheets	132
No Action	500
Picking in Front	456
Picking Left	641
Put Down Component	385
Put Down Measuring Rod	74
Put Down Screwdriver	416
Put Down Subsystem	77
Take Component	485
Take Measuring Rod	76
Take Screwdriver	420
Take Subsystem	39
Turn Sheets	224

### 3.3. Action Recognition Model Configurations Based on the InHARD Dataset

To implement *MPM*, we trained a deep learning model, specifically a 2D convolutional neural network (2DCNN) [49], on the InHARD dataset. We used a 2DCNN, as it excels in extracting spatial features from individual frames. This ability is important for understating body postures and fine-grained motion details given the dynamic HrT environments, as highlighted in the recent literature [33]. Moreover, 2DCNNs offer better computational efficiency compared to 3DCNNs [50], which is particularly important for real-time or resource-constrained applications.

The objective of our task was to recognize the action performed by a human based on an input image. We specifically designed the 2DCNN with certain task-related concerns. The model comprises two convolutional layers with 256 filters with a kernel size of  $3 \times 3$ , a stride of 1, and zero padding, each followed by a pooling layer. The activation function used in the network is the Rectified Linear Unit (ReLU) [51], promoting non-linearity and enabling better feature extraction. The fully connected layers have 512 and 256 neurons, which are designed to further process the features extracted by the convolutional layers. The final layer of the network is a softmax output layer, with the number of neurons equal to the number of action classes in the dataset. The softmax layer is connected to the last fully connected layer and is responsible for generating the predicted action class probabilities. During training, the Adam optimization algorithm was used with a learning rate of 0.001, a batch size of 32, and a momentum of 0.9. The model was trained until convergence, wherein the improvement in the validation loss was determined, and the training process was stopped when the improvement fell below the threshold.

## 4. Results and Discussion

In our study, we focused on the implementation of *MPM* in HrT through the use of action recognition. The primary aim of the study was to conceptualize and demonstrate how *MPM* could be enabled in such HrT settings with robots equipped to effectively understand and interpret human actions. To achieve this, we employed a 2DCNN to develop a human action recognition model capable of accurately identifying actions during the assembly task. The results from our experiments primarily centered around the classification capabilities of the 2DCNN model.

The model's performance was evaluated by generating accuracy and confusion metrics. The metrics provided valuable insights into our action recognition model's performance and its ability to distinguish different actions during the assembly process. Using the designated training and testing datasets, the model demonstrated an accuracy of  $\sim 82\%$  and  $\sim 73\%$  for the training and testing phases, respectively, as shown in Figure 3. In addition, we observed variations in the performance across different action classes, as presented in Table 2. The "Picking Left" action class exhibited the highest accuracy, with the model achieving an accuracy of 95%. This high accuracy can be attributed to the distinct visual cues and unique features associated with the "Picking Left" action. These cues included specific hand movements, object-grasping techniques, and body positions, allowing the model to accurately identify and classify this action. Conversely, the *No Action*, *Turn Sheets*, and *Put Down Measuring Rod* action classes exhibited relatively lower accuracy rates. This can be attributed to the inherent similarities between actions involving hand-object interactions and object manipulations. The model became confused between similar actions such as *Turn Sheets* and *Consult Sheets*. The challenges in accurately differentiating between these actions resulted in misclassifications and lower precision rates. The confusion matrix is presented in Figure 4, which compares the actual and predicted actions.



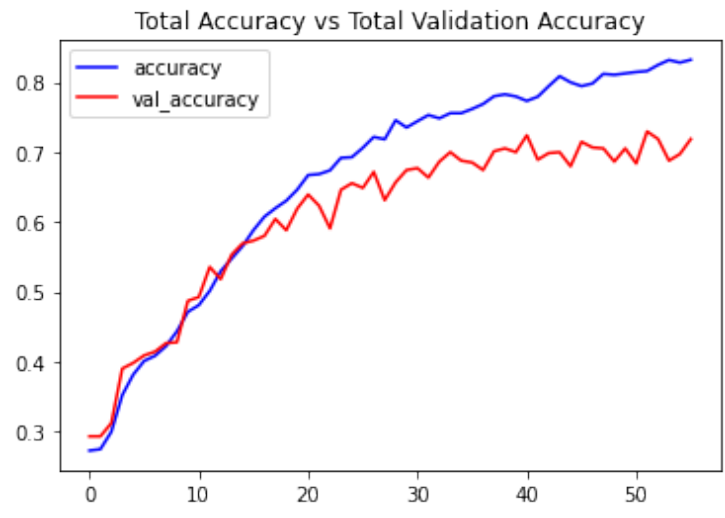


Figure 3. Accuracy graph for human action recognition on the InHARD dataset.

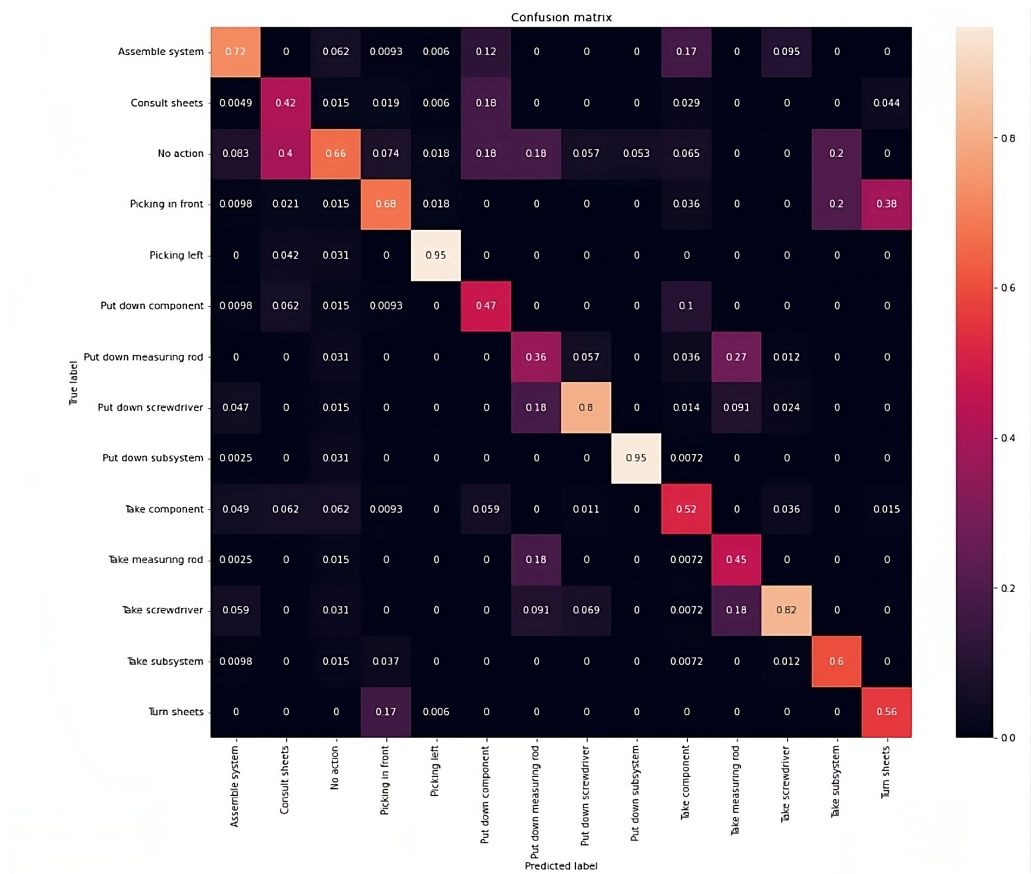
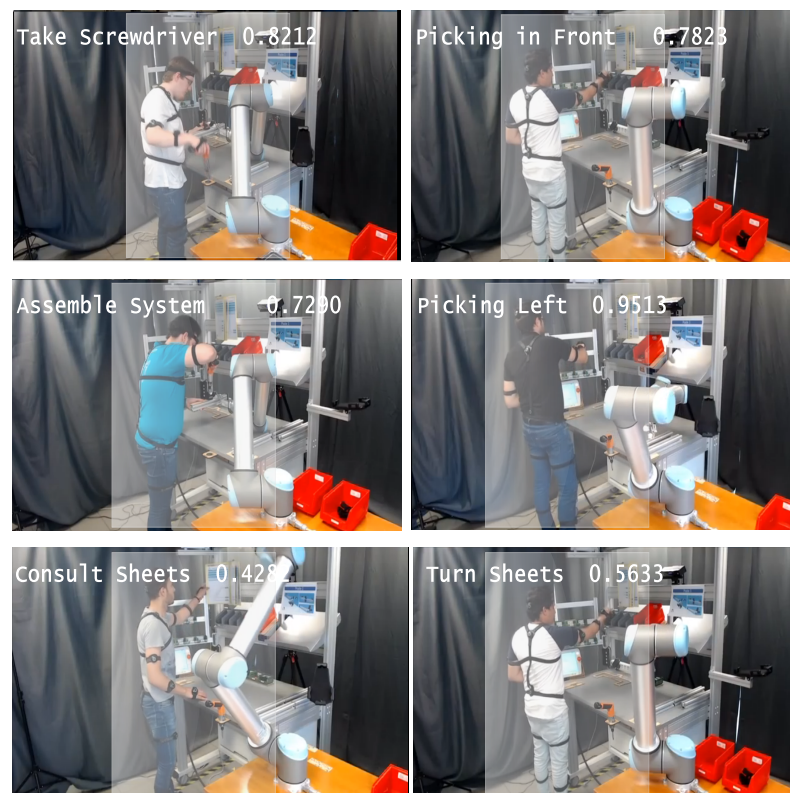


Figure 4. The confusion matrix for human action recognition using the InHARD dataset shows the classification results and is calculated by comparing the predicted action labels with the ground-truth labels for a given dataset.

**Table 2.** Accuracy results for each class.

Meta-Action Class Label	Accuracy %
Assemble System	~64
Consult Sheets	~58
No Action	~35
Picking in Front	~52
Picking Left	~95
Put Down Component	~78
Put Down Measuring Rod	~47
Put Down Screwdriver	~68
Put Down Subsystem	~90
Take Component	~69
Take Measuring Rod	~60
Take Screwdriver	~65
Take Subsystem	~88
Turn Sheets	~32

Despite some confusion in certain action classes, the overall scores indicate that the model learned to recognize and discern human action patterns within the testing data, enabling the accurate identification of various actions in the assembly setting. Importantly, the model also demonstrated its ability to generalize this knowledge and accurately interpret previously unseen data. Figure 5 shows the system's working outcomes, presenting the accuracy scores for different action classes. The outcomes demonstrate the system's proficiency in action recognition, offering valuable insights for implementing *MPM* in HrT applications.

**Figure 5.** System in operation: action recognition on InHARD test set.

While our study mainly focused on presenting classification results, it is important to acknowledge the broader implications of these findings. Our research contributes more than merely the attainment of classification accuracy. Instead, it serves as a key contribution to redefining the dynamics of HrT. The article advocates for a broader perspective on HrT,

urging future research to delve into the intricate components that define successful HhT and apply these insights to the realm of HrT using advanced algorithms, as we presented in this article. The approach represents a shift from viewing robots only as tools or independent agents to seeing them as integral and interactive components within human–robot teams.

#### *Findings and Limitations*

This study has delineated the pipeline, conceptual framework, and model essential for implementing *MPM* using action recognition in HrT settings. Our findings, derived from the data used in this study, reveal promising avenues for enhancing team coordination and task efficiency. The ML model, specifically designed for action recognition, suggests the prospect of real-time behavioral adaptation, facilitating responsive teamwork in human–robot teams. These results, while preliminary, indicate an increased level of trust and reliability in these interactions, aligning with prior research, such as that of Hancock et al. [15].

However, a significant limitation lies in the fact that the integration and empirical validation of the proposed approach remain unexplored. Furthermore, we have identified several other challenges regarding model implementation, including the need for a large amount of training data for deep learning models, the possibility of low accuracy for certain actions, and the sensitivity of action recognition systems to environmental changes. We have observed that the variations in accuracy are primarily due to the inherent limitations within the data used in our study, particularly when dealing with actions that share similar characteristics within the dataset. These similarities can sometimes lead to confusion in accurate recognition for machine learning models. However, resolving this issue effectively would require expanding our training datasets or integrating more advanced sensors, which are both resource-intensive solutions. Furthermore, the reliance on an ML model for action recognition, despite its innovative application, brings forth challenges in decoding complex human behaviors and the variability inherent in real-world settings.

Potential solutions to these limitations could involve using transfer learning techniques and domain adaptation [52], which lay the groundwork for future work. Transfer learning involves using the knowledge learned by a pre-trained action recognition system on one dataset as the starting point for training a new action recognition system on a different dataset. It can allow the new system to benefit from the knowledge learned by the pre-trained system, which can improve overall performance in the new environment or situation. Additionally, it may be possible to improve the accuracy of the action recognition system for certain actions using domain adaptation techniques. Domain adaptation involves modifying the training data or the system architecture to better suit the specific actions or environment under consideration. The constraints in terms of data availability and the accuracy of the captured information highlight the need for further research and development to enhance data collection studies.

#### **5. Conclusions and Future Directions**

The achievement of effective HrT conditions in HRI remains challenging, marked by various unresolved issues that parallel those found in human teamwork dynamics. However, a step toward improving teamwork conditions is implementing capabilities to enable *MPM* in HrT settings. Our paper demonstrates that *MPM* is a necessary step in this direction and can be achieved using current unobtrusive sensor technologies and the adoption of a suitable ML algorithm. In this regard, we trained an ML model for human action recognition using a state-of-the-art deep learning algorithm on an existing industrial-setting dataset. The results from the evaluation represent a substantial leap forward, marking a promising contribution in this domain. This approach was specifically introduced to facilitate performance monitoring in a robot teammate, allowing it to interpret and understand the actions of its human counterparts for enhanced teaming effectiveness. Additionally, we discussed the limitations inherent in our proposal.

Given the recent advancements in AI and the development of complex algorithms, realizing fundamental teaming components is becoming increasingly feasible. Building

on the outcomes of our study, we propose two sets of possible research avenues for future investigations: one set for immediate follow-ups and another with longer-term theoretical implications.

The immediate follow-up research avenues are described below:

- Proposal 1: Advanced ML techniques for developing visual recognition-based *MPM*—The integration and implementation of advanced ML models in *MPM* represent several unexplored areas. Specifically, we suggest exploring deep learning architectures, representation learning [53], and evolutionary computation techniques for adapting to environmental cues [54], and fusion techniques to overcome obstacles in *MPM* action recognition such as context information, model performance, and data-related issues. Deep learning architectures, adept at processing complex data, can accurately interpret context, enhancing *MPM*'s effectiveness.
- Proposal 2: Empirical validation—While this study has laid the groundwork for introducing the HhT element into HrT, conducting validation and evaluation studies is essential. The imperative nature of undertaking validation and evaluation studies cannot be overstated, as they are instrumental in generating empirical evidence concerning the effectiveness and implications of incorporating HhT components within real-world HrT settings.
- Proposal 3: Task-oriented action recognition for improving security in collaborative applications—Action recognition can be regarded as a practical method to improve security measures in HRC. It can function as a useful tool for detecting anomalies in behavior or performance that might signify potential security problems. For example, if a robot or human team member indicates activities or performance patterns that differ from established norms, this could be an early sign of a security breach. Repetition of such anomalies can be used as an indicator of a system being potentially compromised. Such deviations, once detected by the action recognition system, would prompt further investigation and appropriate response steps. This proactive strategy for security risks within HRC can manage immediate risks and contribute to the development of more resilient and secure collaborative systems.

The longer-term theoretical implications of our study that merit in-depth exploration are:

- Proposal 1: Further exploring big-five teaming elements in HrT—The intersection of established big-five HhT characteristics with the evolving landscape of HrT invites further empirical analysis. This study suggests exploring the potential alignment between the key characteristics of HhT, such as mutual performance monitoring, team orientation, backup behaviors, team leadership, and adaptability, and the distinctive skills exhibited by robots. This exploration has the potential to generate novel approaches for enhancing team performance.
- Proposal 2: Improving levels of safety and security in HrT—Considering the complexity of HrT, especially with ML for performance monitoring, future research should focus on enhancing safety and security levels. The example explored in this paper regarding action recognition shows the potential to redefine safety proximity criteria, and at the same time, provides an additional tool for identifying possible security breaches. It introduces the potential of a more context-aware robotic intelligent system. As stated by Schaefer et al. [55], incorporating context-driven AI is important for advancing future robotic capabilities, thereby promoting the development of situational awareness, calibrating trust, and enhancing team performance in collaborative HrT. Furthermore, the application of *MPM* based on human action recognition is likely to be highly advantageous in safety-critical settings such as shared manufacturing cells. By using human data, the system can foresee possible threats and take proactive steps to prevent hazardous situations. Hence, the system can improve the performance, safety, and efficiency of the production cell by accurately predicting human behaviors. Therefore, the system can make more informed decisions about its actions and better understand the context of the manufacturing process. However, this requires improv-



ing the reliability and robustness of the AI algorithm underpinning these functions beyond the values achieved by the confusion matrix discussed in this paper.

In conclusion, the motivation for this study is not only to investigate *MPM* in HrT but also to pave the way for the formulation of novel ideas by utilizing current technological possibilities to develop real teamwork capabilities in HrT.

**Author Contributions:** Conceptualization, S.M. and M.C.L.; methodology, S.M.; software, S.M.; validation, J.D.K. and M.G.; formal analysis, S.M.; investigation, J.D.K. and S.M.; resources, M.G.; data curation, S.M.; writing—original draft preparation, S.M.; writing—review and editing, J.K.; visualization, S.M.; supervision, J.D.K. and M.C.L.; project administration, M.C.L. and M.G.; funding acquisition, M.G. and M.C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Collaborative Intelligence for Safety-Critical Systems (CISC) project. The CISC project received funding from the European Union’s Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant, agreement no. 955901 <https://www.ciscproject.eu/> accessed on 30 March 2023.

**Data Availability Statement:** The dataset used in this study is publicly available in Zenodo at <https://doi.org/10.5281/zenodo.4003541>, accessed on 30 March 2023 reference number [36].

**Conflicts of Interest:** S.M. and M.G. are employed by Pilz Ireland Industrial Automation, Cork, Ireland. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Devlin, S.P.; Flynn, J.R.; Riggs, S.L. Connecting the big five taxonomies: Understanding how individual traits contribute to team adaptability under workload transitions. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Baltimore, MD, USA, 3–8 October 2021; SAGE Publications Sage CA: Los Angeles, CA, USA, 2018; Volume 62, pp. 119–123.
2. Wolf, F.D.; Stock-Homburg, R. Human-robot teams: A review. In Proceedings of the International Conference on Social Robotics, Golden, CO, USA, 14–18 November 2020; pp. 246–258.
3. Martinetti, A.; Chemweno, P.K.; Nizamis, K.; Fosch-Villaronga, E. Redefining safety in light of human-robot interaction: A critical review of current standards and regulations. *Front. Chem. Eng.* **2021**, *3*, 32. [[CrossRef](#)]
4. Tuncer, S.; Licoppe, C.; Luff, P.; Heath, C. Recipient design in human–robot interaction: The emergent assessment of a robot’s competence. *AI Soc.* **2023**, 1–16. [[CrossRef](#)]
5. Mutlu, B.; Forlizzi, J. Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. In Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, Amsterdam, The Netherlands, 12–15 March 2008; pp. 287–294.
6. Harper, C.; Virk, G. Towards the development of international safety standards for human robot interaction. *Int. J. Soc. Robot.* **2010**, *2*, 229–234. [[CrossRef](#)]
7. Hoffman, G.; Breazeal, C. Collaboration in human-robot teams. In Proceedings of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA, 20–22 September 2004; p. 6434.
8. de Visser, E.; Parasuraman, R. Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *J. Cogn. Eng. Decis. Mak.* **2011**, *5*, 209–231. [[CrossRef](#)]
9. Gombolay, M.C.; Huang, C.; Shah, J. Coordination of human-robot teaming with human task preferences. In Proceedings of the 2015 AAAI Fall Symposium Series, Virginia, 12 November 2015.
10. Tabrez, A.; Luebbbers, M.B.; Hayes, B. A survey of mental modeling techniques in human–robot teaming. *Curr. Robot. Rep.* **2020**, *1*, 259–267. [[CrossRef](#)]
11. Zhang, Q.; Lee, M.L.; Carter, S. You complete me: Human-ai teams and complementary expertise. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 30 April–5 May 2022; pp. 1–28.
12. Webber, S.S.; Detjen, J.; MacLean, T.L.; Thomas, D. Team challenges: Is artificial intelligence the solution? *Bus. Horizons* **2019**, *62*, 741–750. [[CrossRef](#)]
13. Lewis, M.; Sycara, K.; Walker, P. The role of trust in human-robot interaction. *Found. Trust. Auton.* **2018**, *117*, 135–159.
14. Guo, Y.; Yang, X.J. Modeling and predicting trust dynamics in human–robot teaming: A Bayesian inference approach. *Int. J. Soc. Robot.* **2021**, *13*, 1899–1909. [[CrossRef](#)]
15. Hancock, P.A.; Billings, D.R.; Schaefer, K.E.; Chen, J.Y.; De Visser, E.J.; Parasuraman, R. A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Factors* **2011**, *53*, 517–527. [[CrossRef](#)]
16. Onnasch, L.; Roesler, E. A taxonomy to structure and analyze human–robot interaction. *Int. J. Soc. Robot.* **2021**, *13*, 833–849. [[CrossRef](#)]
17. Albon, R.; Jewels, T. Mutual performance monitoring: Elaborating the development of a team learning theory. *Group Decis. Negot.* **2014**, *23*, 149–164. [[CrossRef](#)]



18. Salas, E.; Sims, D.E.; Burke, C.S. Is there a “big five” in teamwork? *Small Group Res.* **2005**, *36*, 555–599. [[CrossRef](#)]
19. Ma, L.M.; IJtsma, M.; Feigh, K.M.; Pritchett, A.R. Metrics for human-robot team design: A teamwork perspective on evaluation of human-robot teams. *ACM Trans. Hum.-Robot Interact. (THRI)* **2022**, *11*, 1–36. [[CrossRef](#)]
20. You, S.; Robert, L. Teaming up with robots: An IMO (inputs-mediators-outputs-inputs) framework of human-robot teamwork. *Int. J. Robot. Eng.* **2018**, *2*. [[CrossRef](#)]
21. Guznov, S.; Lyons, J.; Pfahler, M.; Heironimus, A.; Woolley, M.; Friedman, J.; Neimeier, A. Robot transparency and team orientation effects on human-robot teaming. *Int. J. Hum. Interact.* **2020**, *36*, 650–660. [[CrossRef](#)]
22. Yasar, M.S.; Iqbal, T. Robots That Can Anticipate and Learn in Human-Robot Teams. In Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Sapporo, Hokkaido, Japan, 7–10 March 2022; pp. 1185–1187.
23. De Visser, E.J.; Peeters, M.M.; Jung, M.F.; Kohn, S.; Shaw, T.H.; Pak, R.; Neerinx, M.A. Towards a theory of longitudinal trust calibration in human-robot teams. *Int. J. Soc. Robot.* **2020**, *12*, 459–478. [[CrossRef](#)]
24. Shah, J.; Breazeal, C. An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Hum. Factors* **2010**, *52*, 234–245. [[CrossRef](#)] [[PubMed](#)]
25. Gervasi, R.; Mastrogiacomo, L.; Maisano, D.A.; Antonelli, D.; Franceschini, F. A structured methodology to support human-robot collaboration configuration choice. *Prod. Eng.* **2022**, *2022*, 1–17. [[CrossRef](#)]
26. Dahiya, A.; Aroyo, A.M.; Dautenhahn, K.; Smith, S.L. A survey of multi-agent Human-Robot Interaction systems. *Robot. Auton. Syst.* **2023**, *161*, 104335. [[CrossRef](#)]
27. Lemaignan, S.; Cooper, S.; Ros, R.; Ferrini, L.; Andriella, A.; Irisarri, A. Open-source Natural Language Processing on the PAL Robotics ARI Social Robot. In Proceedings of the Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, Stockholm, Sweden, 13–16 March 2023; pp. 907–908.
28. Hari, S.K.K.; Nayak, A.; Rathinam, S. An approximation algorithm for a task allocation, sequencing and scheduling problem involving a human-robot team. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2146–2153. [[CrossRef](#)]
29. Singh, S.; Heard, J. Human-aware reinforcement learning for adaptive human robot teaming. In Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Sapporo, Hokkaido, Japan, 2022; pp. 1049–1052.
30. Tian, C.; Xu, Z.; Wang, L.; Liu, Y. Arc fault detection using artificial intelligence: Challenges and benefits. *Math. Biosci. Eng.* **2023**, *20*, 12404–12432. [[CrossRef](#)] [[PubMed](#)]
31. Naser, M.; Alavi, A. Insights into performance fitness and error metrics for machine learning. *arXiv* **2020**, arXiv:2006.00887.
32. Chakraborti, T.; Kambhampati, S.; Scheutz, M.; Zhang, Y. Ai challenges in human-robot cognitive teaming. *arXiv* **2017**, arXiv:1707.04775.
33. Huang, X.; Cai, Z. A review of video action recognition based on 3D convolution. *Comput. Electr. Eng.* **2023**, *108*, 108713. [[CrossRef](#)]
34. Rodomagoulakis, I.; Kardaris, N.; Pitsikalis, V.; Mavroudi, E.; Katsamanis, A.; Tsiami, A.; Maragos, P. Multimodal human action recognition in assistive human-robot interaction. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 2702–2706.
35. Kong, Y.; Fu, Y. Human action recognition and prediction: A survey. *Int. J. Comput. Vis.* **2022**, *130*, 1366–1401. [[CrossRef](#)]
36. Dallel, M.; Havard, V.; Baudry, D.; Savatier, X. Inhard-industrial human action recognition dataset in the context of industrial collaborative robotics. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 7–9 September 2020; pp. 1–6.
37. Seraj, E. Embodied Team Intelligence in Multi-Robot Systems. In Proceedings of the AAMAS, Auckland, New Zealand, 9–13 May 2022; pp. 1869–1871.
38. Perzanowski, D.; Schultz, A.C.; Adams, W.; Marsh, E.; Bugajska, M. Building a multimodal human-robot interface. *IEEE Intell. Syst.* **2001**, *16*, 16–21. [[CrossRef](#)]
39. Chiou, E.K.; Demir, M.; Buchanan, V.; Corral, C.C.; Endsley, M.R.; Lematta, G.J.; Cooke, N.J.; McNeese, N.J. Towards human-robot teaming: Tradeoffs of explanation-based communication strategies in a virtual search and rescue task. *Int. J. Soc. Robot.* **2021**, *14*, 1117–1136. [[CrossRef](#)]
40. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [[CrossRef](#)]
41. Guo, Y.; Yang, X.J.; Shi, C. TIP: A Trust Inference and Propagation Model in Multi-Human Multi-Robot Teams. *arXiv* **2023**, arXiv:2301.10928.
42. Natarajan, M.; Seraj, E.; Altundas, B.; Paleja, R.; Ye, S.; Chen, L.; Jensen, R.; Chang, K.C.; Gombolay, M. Human-Robot Teaming: Grand Challenges. *Curr. Robot. Rep.* **2023**, *4*, 1–20. [[CrossRef](#)]
43. He, Z.; Song, Y.; Zhou, S.; Cai, Z. Interaction of Thoughts: Towards Mediating Task Assignment in Human-AI Cooperation with a Capability-Aware Shared Mental Model. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–29 April 2023; pp. 1–18.
44. Demir, M.; Cohen, M.; Johnson, C.J.; Chiou, E.K.; Cooke, N.J. Exploration of the impact of interpersonal communication and coordination dynamics on team effectiveness in human-machine teams. *Int. J. Hum.-Interact.* **2023**, *39*, 1841–1855. [[CrossRef](#)]
45. Zhang, Y.; Williams, B. Adaptation and Communication in Human-Robot Teaming to Handle Discrepancies in Agents’ Beliefs about Plans. In Proceedings of the International Conference on Automated Planning and Scheduling, Prague, Czech Republic, 8–12 July 2023; Volume 33, pp. 462–471.

46. Schmidbauer, C.; Zafari, S.; Hader, B.; Schlund, S. An Empirical Study on Workers' Preferences in Human–Robot Task Assignment in Industrial Assembly Systems. *IEEE Trans. Hum.-Mach. Syst.* **2023**, *53*, 293–302. [[CrossRef](#)]
47. Wang, L.; Ge, L.; Li, R.; Fang, Y. Three-stream CNNs for action recognition. *Pattern Recognit. Lett.* **2017**, *92*, 33–40. [[CrossRef](#)]
48. Hossain, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
49. Gholamrezaii, M.; Almodarresi, S.M.T. Human activity recognition using 2D convolutional neural networks. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; pp. 1682–1686.
50. Stamoulakatos, A.; Cardona, J.; Michie, C.; Andonovic, I.; Lazaridis, P.; Bellekens, X.; Atkinson, R.; Hossain, M.M.; Tachtatzis, C. A comparison of the performance of 2D and 3D convolutional neural networks for subsea survey video classification. In Proceedings of the OCEANS 2021: San Diego–Porto, San Diego, CA, USA, 20–23 September 2021; pp. 1–10.
51. Taye, M.M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* **2023**, *11*, 52. [[CrossRef](#)]
52. Shi, Y.; Li, L.; Yang, J.; Wang, Y.; Hao, S. Center-based transfer feature learning with classifier adaptation for surface defect recognition. *Mech. Syst. Signal Process.* **2023**, *188*, 110001. [[CrossRef](#)]
53. Wang, Y.; Liu, Z.; Xu, J.; Yan, W. Heterogeneous network representation learning approach for ethereum identity identification. *IEEE Trans. Comput. Soc. Syst.* **2022**, *10*, 890–899. [[CrossRef](#)]
54. Liu, Z.; Yang, D.; Wang, Y.; Lu, M.; Li, R. EGNN: Graph structure learning based on evolutionary computation helps more in graph neural networks. *Appl. Soft Comput.* **2023**, *135*, 110040. [[CrossRef](#)]
55. Schaefer, K.E.; Oh, J.; Aksaray, D.; Barber, D. Integrating context into artificial intelligence: Research from the robotics collaborative technology alliance. *Ai Mag.* **2019**, *40*, 28–40. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.