JOHANNES KEPLER
UNIVERSITY LINZ

Submitted by
Georgios Sopidis

Submitted at
Institute of Pervasive Computing

Thesis Supervisor & First Evaluator
Univ.-Prof Dr. Alois Ferscha

Second Evaluator
Univ.-Prof Dr. Paul Lukowicz

Month Year / April 2025

Cognitive Augmentation for Manual Assembly

Doctoral Thesis

to obtain the academic degree of

Doktor der Technischen Wissenschaften

in the Doctoral Program

Technische Wissenschaften

JOHANNES KEPLER
UNIVERSITY LINZ
Altenbergerstraße 69
4040 Linz, Austria
www.jku.at
DVR 0093696

Abstract

Activity recognition combined with artificial intelligence is a vital area of research, ranging across diverse domains, from sports and healthcare to smart homes. In the industrial domain, and the manual assembly lines, the emphasis shifts to human-machine interaction and thus to human activity recognition (HAR) within complex operational environments. Developing models and methods that can reliably and efficiently identify repetitive human activities, traditionally just categorized as either simple or complex activities, remains a key challenge in the field.

Limitations of the existing methods and approaches include their inability to consider the contextual complexities associated with the performed activities. To address these limitations, this thesis introduces the concept of cognitive augmentation. This approach integrates human activity recognition with advanced AI approaches and feedback mechanisms, that aim to i) enhance awareness and knowledge of the performed activities and tasks ii) optimize workflows, and iii) support worker decision-making.

Creating different levels of activity abstractions allows for a more nuanced comprehension of activities and describes the underlying characteristics and patterns. This work proposes a hierarchical taxonomy of activities, categorized into atomic, micro, meso, macro, and mega levels, as a framework for understanding and analyzing industrial workflows. The approach is derived from real world observations of industrial assembly processes and reflects the task composition implemented in realistic operations that align with existing literature. Additionally, it presents a guidance system for sensing in manufacturing to apply artificial intelligence (AI) through the outlined abstraction levels in various processes. Each level of abstraction introduces distinct requirements in sensing, data processing, modeling, and feedback, making the taxonomy a practical starting tool for the development of AI-based HAR systems.

The implementation of these concepts is demonstrated through real-world prototypes, including a smartwatch for activity monitoring, a smart helmet for cognitive and safety assistance, and depth imaging for macro-level activity recognition. These systems bridge the gap between theoretical models and practical applications, and utilize multi-modal sensor data, such as IMUs and depth cameras, to classify and quantify activities, detect anomalies, and provide real-time feedback to workers.

The augmentation of human cognitive abilities with technologies such as AI, aims to guide and optimize industrial assembly, particularly in uncontrolled non-laboratory environments, by i) shaping workflows to enable structured data analysis and ii) highlighting correlations across various levels throughout the assembly progression. The expected impact includes improving task accuracy, reducing human errors, assisting worker training, and enhancing the quality control of final products in complex industrial settings.

Acknowledgements

I would like to sincerely thank all those who have supported, guided, and inspired me during the adventure of completing this thesis.

First and foremost, I express my deepest gratitude to my supervisor, Prof. Alois Ferscha, for the guidance and support throughout my doctoral journey and for providing me with the opportunity to explore the fascinating world of research.

I would also like to thank my mentors and colleagues at the Institute of Pervasive Computing and Pro2Future, whose collaboration, advice, and constructive discussions provided me with valuable insights and shaped my research experience and growth.

I extend my heartfelt appreciation to my beloved friends and peers who have constantly provided encouragement, motivation, and much-needed humor and smiles over this challenging period.

Finally, and most importantly, I thank my family for their unconditional love, endless support, and continuous belief in me. Their constant presence has been my greatest source of strength and inspiration.

Thank you all for contributing to this significant milestone in my life!

–This work has been supported by the FFG COMET K1 Center "Pro^{2}Future II" (Cognitive and Sustainable Products and Production Systems of the Future), Contract No. 911655

Contents

1 Introduction

1.1 Framing Cognitive Augmentation

It is apparent that technological developments over the last few years have significantly increased the amount of data people generate in their daily lives. In particular, sensors and other smart devices that perceive changes produce data for various purposes, one of which is to benefit humans by assisting them. The increasing interest in supporting humans is explored in the field of human activity recognition (HAR), which offers the opportunity to identify movements and actions of the human body through data from different sensors. The collected information is used as input to the appropriate models and therefore can provide an understanding of the activities that lead to reasoning, awareness, and decision-making in everyday situations [1] and eventually lead to cognition.

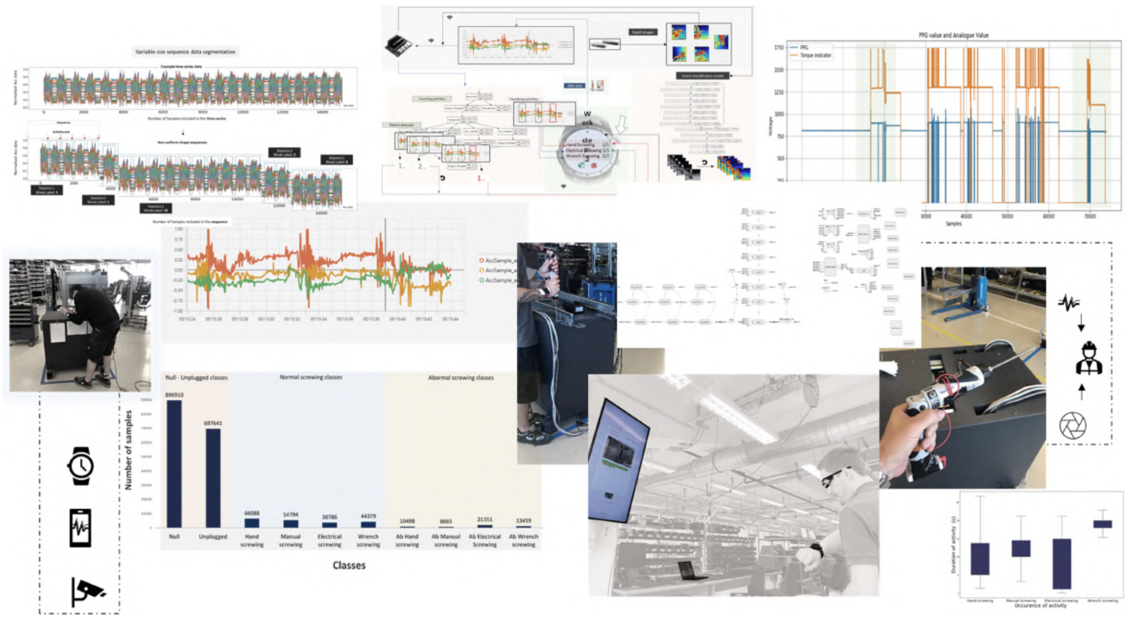


Figure 1.1: The figure demonstrates an overview of the components and methods explored in this thesis on cognitive augmentation for manual assembly. It showcases key visuals, including sensor outputs, model performance results, and industrial application scenarios, which are elaborated upon in subsequent sections. Furthermore, it shows AI-driven tools that enhance worker performance, streamline workflows and enable intelligent human-machine collaboration in industrial environments.

Engel et al., [2] define cognition in the context of task and process automation to be the process of developing knowledge and understanding where sensory input is transformed, reduced, elaborated, stored, recovered, and used. Various types of sensors are involved in these processes such as visual sensors or wearable inertial measurement units (IMUs) and stationary or mobile sensors as shown in Figure 1.1. Each application requires a suitable type, based on environmental factors, process specifications, and personal desires. Accordingly, each application utilizes algorithms that fit the type of data that are created based on the sensor selection, for HAR and human behavior recognition (HBR).

To enhance those, the academic community is very active in researching innovative technologies regarding sensor designs, and traditional and non-traditional sensor types [3]. Furthermore, several efforts aim to improve existing methodologies or address recent issues, for the detection and classification of activities using supervised or unsupervised machine learning models. However, despite the benefits of those methods, they often occur in controlled settings with less flexibility and complexity than in real-world scenarios. This limitation further extends to industrial workflows where their implementation is even more limited, and face challenges in adapting to dynamic environments and task variability. Yet, it is expected that the next generation of cognitive processes will learn based on accumulated knowledge and will be able to assist humans in everyday tasks and augment the capabilities of machines [4]. These cognitive systems and technologies utilize self-learning algorithms and require human intelligence such as planning, reasoning, and learning and can simulate the way the human brain works [5].

Recent findings from the authors in [6], address the potential of human-AI systems in augmenting human performance. Their meta-analysis shows that true human-AI synergy is harder to achieve and on average, human–AI combinations often underperform compared to the best of humans or AI alone. However, augmentation through such systems enhances human capabilities since human-AI systems are better than humans working alone as visualized in Figure 1.2.

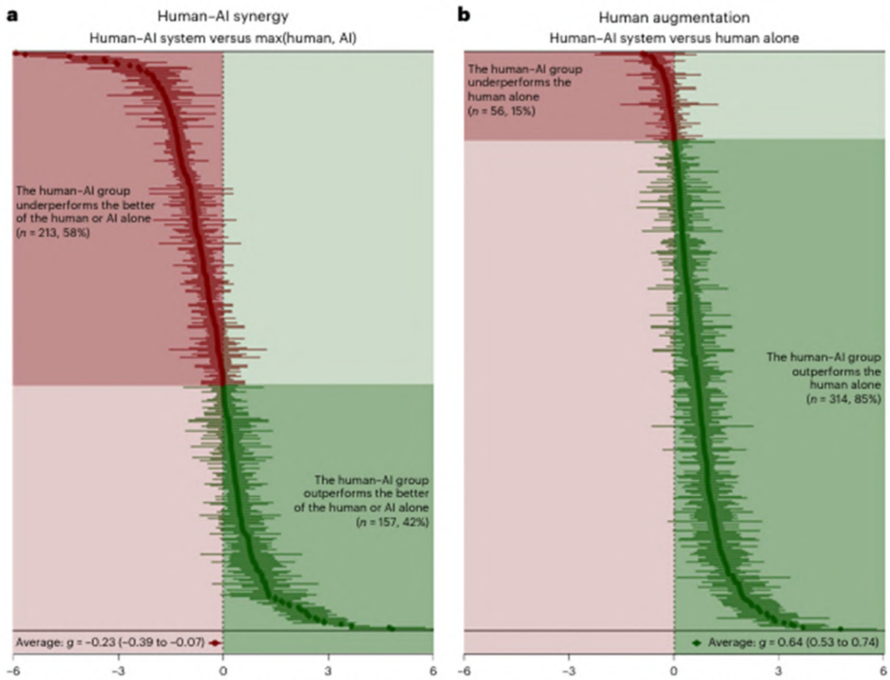


Figure 1.2: Source: Reproduced from [6]. The figure shows that human-AI systems improve human performance. Their research shows that true synergy is harder to achieve. Panel (a): Human-AI systems often fail to outperform the best of either alone (human or AI). Panel (b): In most cases, collaborating with AI makes humans better at tasks and the human–AI group outperforms the human alone in 85% of cases.

While physical tools can enhance human performance, equipment capable of processing and transforming information can enhance human cognitive performance [7]. As visualized in Figure 1.3 the human in industrial manufacturing processes holds an important role that cannot be replaced but needs to be elevated by incorporating the necessary tools and methods [8]. When such tools that involve unsupervised, deep, machine learning techniques are used in a collaborative manner, **human cognitive performance is augmented** [7]. Cinel et al., [9] describe human cognitive enhancement as improving cognitive processes that help generate knowledge and understanding of the environment. In their work, Marois et al, highlight that **"the goal of cognitive enhancement is to augment human capacities to facilitate task performance, instead of reducing the resources necessary to carry out the task"**. This approach favors the matching of human capacities to tasks and technologies that prove efficient in daily work-life situations with high cognitive demands or where traditional practices prove unable to reduce errors [10].

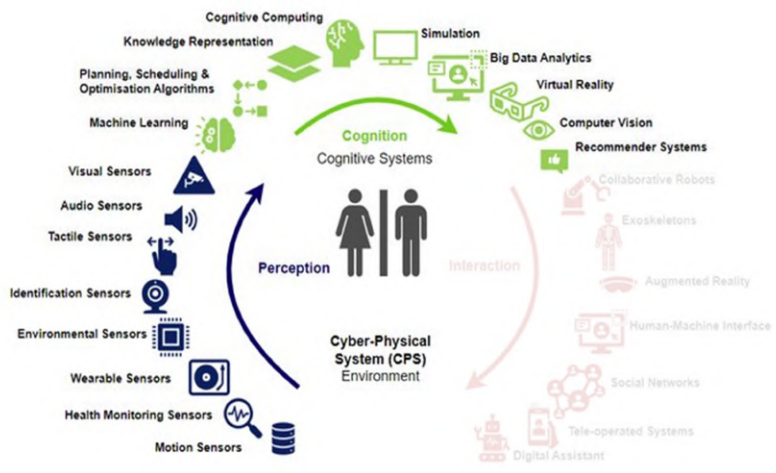


Figure 1.3: Source: Reproduced and adapted from [8]. The figure visualizes human-machine collaboration within cyber-physical systems (CPS) in smart manufacturing deployments, highlighting its main sectors, such as perception, cognition, and interaction. The blue and green regions represent the primary focus of this research work, which covers sensor-based perception and cognitive decision-making through AI, machine learning, and data analysis. The red region, which includes interaction technologies such as collaborative robots and human-machine interfaces, is less prominent as it falls out of the scope of this work.

Cognitive Augmentation extends the concepts of traditional HAR and HBR approaches by integrating the data captured by those sensors into advanced systems. These systems can support humans in real time to improve task performance and decision-making in industrial settings. Especially in the manual assembly workflows where the correct execution of activities is highly important to achieve high-quality standards.

To date achievements have been made in entertainment and sports to track athletes' performance in real-time [11, 12, 13], in healthcare to monitor patients [14] and industrial applications [15] which transform and develop due to advances in artificial intelligence and sensor technology creating systems that continuously learn and adapt to complex tasks and changing environments. Nevertheless, there is still a need for progress in minimizing potential errors due to dynamic environments and task variability of manual workflows that can be achieved through cognitive augmentation.

1.2 Positioning in Science

In recent years the **industrial** field has progressed and expanded in ways that the daily work tasks for humans become difficult. Previously, automating processes, or replacing humans with robots were considered meaningful solutions that would lead to the desired result but this is not the case in all different types of manufacturing. In the industrial **assembly processes**, the high degree of product customization, flexibility in task execution, dynamic working environments, and production quality goals demand **human-centered** lines. Figure 1.4, presents an initial map of the relative research that has been conducted in the field of human-centered AI.

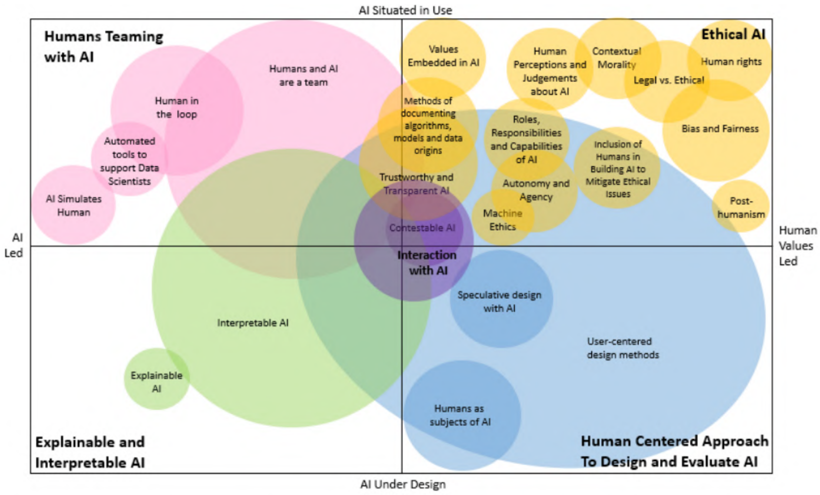


Figure 1.4: Source: Reproduced from [16]. The figure presents a map of the research landscape of Human-Centered AI (HCAI), illustrating key areas of focus in the field. The bubble sizes show the relative number of research papers in each area, while the color coding represents major research domains.

In this field, workers can find **assistance** to complete their **daily tasks** in systems that are designed to promote collaboration between **humans and AI** as visualized in Figure 1.5. As it is shown in the image manual, hybrid and automated assembly have their advantages and disadvantages with manual assembly being high in variability, and flexibility but low in quality and productivity. In this context, **cognitive augmentation** can improve the workers' performance in real time especially when it is seamlessly integrated, contributing to the **optimization of workflows**. Such improvements can be achieved by understanding a range of activities that are performed in an industrial environment.

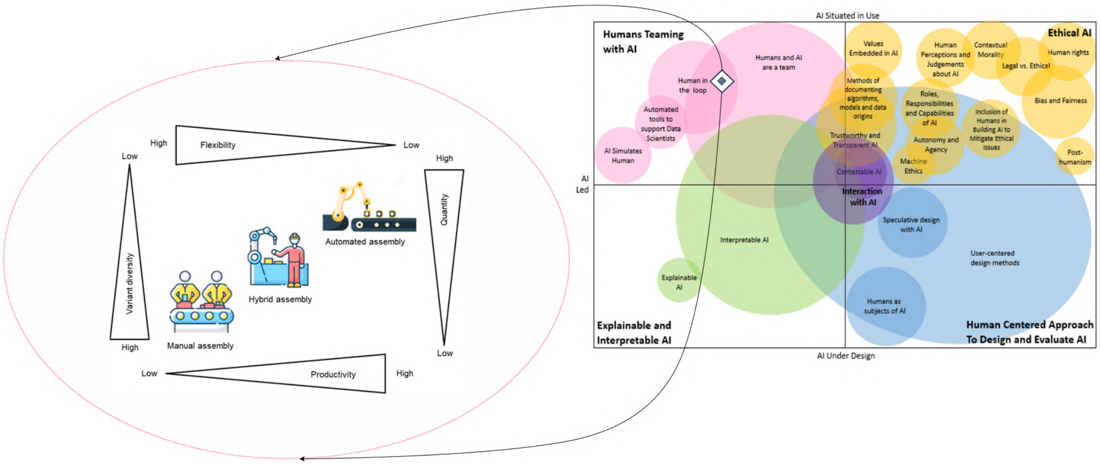


Figure 1.5: Source: Reproduced and adapted from [16]. The graph visualizes the related topics that are addressed in this thesis and places the work within the general frame of research for assistance systems in manufacturing workflows as presented in [16], (right part of the image). The research in this thesis aligns with the "Humans Teaming with AI" quadrant (pink), where humans remain in the loop and collaborate with AI systems. The left side of the image visualizes key aspects of manual, hybrid, and automated assembly processes, with an emphasis on the trade-offs between flexibility, productivity, variant diversity, and quantity.

This work employs wearable and non-wearable sensors in combination with state-of-the-art machine learning algorithms to detect fundamental human activities during assembly tasks that received limited attention in the field of human activity recognition (HAR). This thesis lays the groundwork for developing comprehensive cognitive augmentation systems and highlights the need to integrate them into employees' daily work life(DWL). Additionally, aims to modernize and digitize industrial processes enhancing human-machine cooperation to achieve a cognitive symbiosis and optimize production efficiency.

1.3 The Need for Augmentation

Industrial manufacturing and assembly processes constitute an important piece of modern societies, providing innovation, economic growth, and technological advancements [17].

The human expertise in those procedures remains important since they involve multiple manual assembly tasks and activities [18, 19]. Despite the prevailing focus on locomotion activities in human activity recognition (HAR), it is necessary to recognize that human industrial activities involve complexities far beyond locomotion activities, which need to be explored. Figure 1.6 shows the already known advantages (in productivity, variant diversity) and an example of the aim, which is to also improve productivity and quality in the manual assembly processes as humans remain an essential part of an assembly.

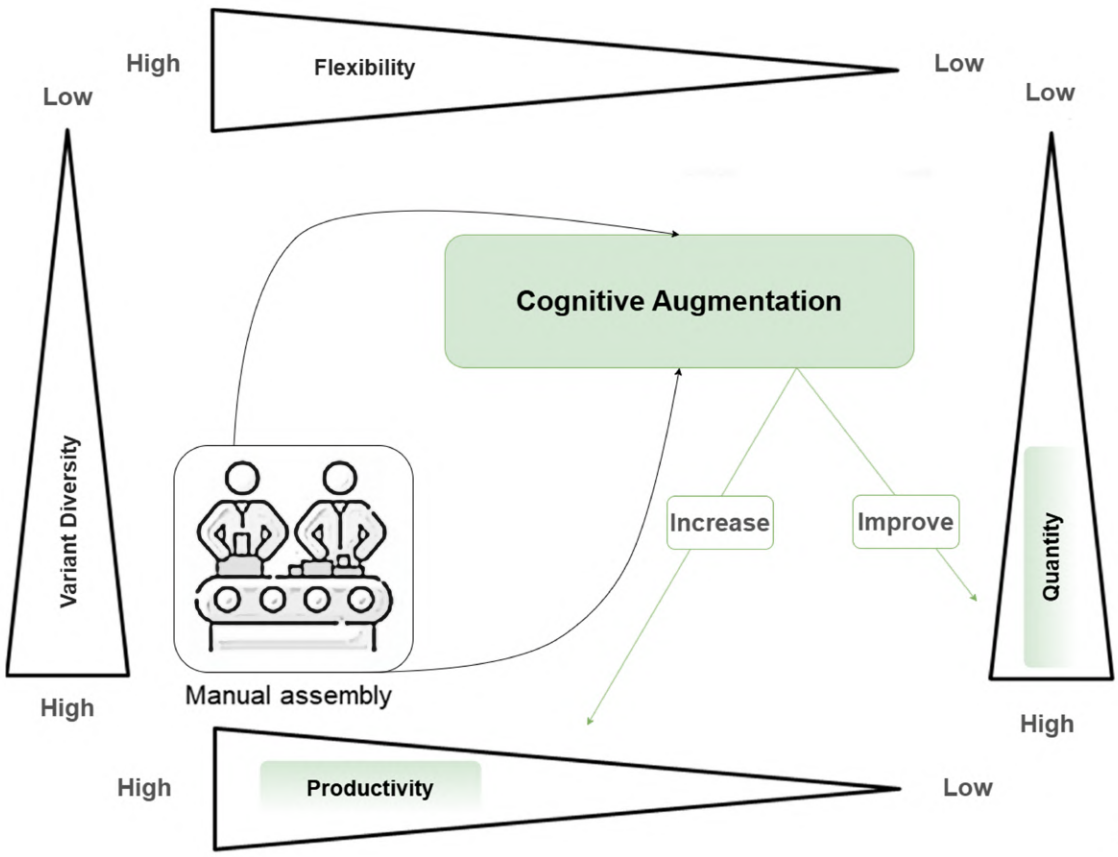


Figure 1.6: Source: Reproduced and adapted from [20]. Based on the work by the authors, the figure presents the intention of this work to improve productivity and quality for manual assembly through the research and implementation of cognitive augmentation. In the green triangles, one can see the values as expected to impact the manual assembly, improve the quantity, and increase productivity adding to the commonly known advantages of manual assembly.

In the HAR field, the correct execution of activities is essential to:

(i) achieve production goals,

(ii) ensure product quality,

(iii) and maximize operational efficiency.

Therefore, advanced techniques are required to analyze the activities and optimize complex tasks and workflows.

Research shows that work-related errors in manual industrial manufacturing are mostly (70% of the time) associated with task complexity, instructions, difficulty, training, and tools. Examples of production defects consist of missing components and examples of "human errors" that include assembly-related errors [20], such as incomplete execution of the activity, components omitted, and operation omitted [21]. These errors frequently occur during short critical actions such as fasting or tightening, which are typically unmonitored due to being short and weak in sensor data (– see Section 5.1). This thesis refers to these actions as micro-activities. To cope with that, studies for human activity recognition have been conducted for assistant systems related to assembly processes of industrial automation. In these studies the authors:

- apply human error classification strategies for missing or faulty applied components [22],
- use hand-operated tools in controlled environments [23],
- or employ mostly mixtures of stationary and wearable sensors [15].

While existing research highlights wearable sensors as technologies that support workers, given their potential to enhance productivity and increase efficiency, they are mainly used to collect health and movement data from less complex user activities, to improve health and ergonomics [24]. Therefore, little is known about the extent to which individual human, complex, and short-in-duration activities (e.g., screwing) in an actual factory assembly process can be recognized solely by wrist-worn inertial measurement units (IMUs) and their exploitation to reduce failures that occur during a manual assembly workflow.

Consequently, it is important to explore the potential of specific-for-the-task-sensors and devices using intelligent algorithms to achieve, cognitive augmentation in supporting industrial assembly processes and addressing emerging demands.

1.4 Thesis Objectives and Contributions

This thesis explores how advanced wearable and non-wearable sensor-based systems combined with state-of-the-art machine learning algorithms can be employed to recognize, classify, and count industrial activities in different stages of assembly workflows. The goal is to achieve workflow recognition and promote cognitive augmentation of human workers to understand and optimize assembly processes. Additionally, aims to support industries that face challenges in modernizing older equipment due to high replacement costs or removal difficulties with cognitive augmentation systems that integrate seamlessly into existing workflows and equipment.

To systematically address the challenges of recognizing activities, this thesis proposes a hierarchical taxonomy of industrial activities that separates assembly processes into different levels of abstraction and guides the AI system design across each level. The taxonomy provides a conceptual framework but also offers a standardized approach and a practical tool for guiding the design of AI-based human activity recognition (HAR) systems. Each level of abstraction introduces distinct recommendations for the selection of sensors, models, preprocessing techniques, and feedback approaches during the development of such systems. These categories which are commonly found in the literature form foundational points of AI-HAR-based system design since they determine how systems perceive interpret and respond to human activities (– see Section 4.4).

Five levels of an industrial process are realized, which are derived from empirical observation of real industrial assembly workflows. In contrast with binary or two-level classification approaches in HAR (e.g., simple/complex or fine/coarse), the proposed framework introduces intermediate levels of abstraction, complementing and extending existing multi-level approaches. The levels of the taxonomy present differences in activity characteristics, sensor, and AI model selection, and facilitate modular system design across workflow stages, support customized feedback based on activity level (– see Section 4.1, and Section 4.2).

A core concept that is analyzed is the micro-activities during industrial tasks due to their critical role in providing insights into the quality and efficiency of tasks. The focus on micro-activities arises from the limited existing research in real industrial settings, their complexity, difficulty in recognition, and relevance for real-time error correction during their execution. These repetitive activities with a short duration are handled by the proposed deep learning architectures and captured by appropriate sensors. Given their minimal intrusive nature, wearable sensors such as inertial measurement units (IMUs) are used to track task-specific movements of workers in real-time to enhance process awareness and improve performance. Additionally, the thesis investigates the practical implementation and benefits of prototyped solutions for cognitive assistance in real applications. To achieve this, the research shows the need to collect data from real industrial settings to realistically address complexities, variability, and unpredictable human behavior.

Building on this foundation, the key contributions and the expected impact of this work are presented below.

Key contributions include:

- Development of an activity taxonomy, addressing a research gap identified in the literature that integrates atomic, micro, meso, macro, and mega-level activities into a structured hierarchy. The taxonomy proposes categories suitable for industrial applications that are derived through empirical observations of real-world assembly tasks.
- Design of AI systems based on this taxonomy for wearable and non-wearable sensor data that focus on robust recognition and classification of assembly activities and guidelines for sensor placement, data preprocessing methods, model selection, and feedback design related to each abstraction level.
- Utilization of low-cost wearable systems that enable mobility in working space, scalable deployment, and adaptability in various industrial settings.
- Real industrial data collected during the normal working hours of the assembly workers where the complexity, unpredictability, and contextual variations of human activity can be captured and addressed directly.

- Detection of activity patterns and comparison of recognition models to identify and classify complex tasks and activities during industrial assembly processes.
- Research on model architecture, window size, sliding rate, and weighting techniques to ensure better performance.
- Model implementation for variable-length sequences to handle the dynamic and unpredictable nature of industrial tasks based on real data (non-laboratory).
- Handling weak labels to address the challenges of incorrect or tedious annotation tasks in industrial settings.
- Integration of a counting method as part of the learning process to track the number of completed activities supporting real-time workflow monitoring and quality assurance.
- Optimization for fewer training parameters to ensure suitability for deployment on devices with limited computational power and energy resources.
- Demonstration of the conceptual unobtrusive assistance system prototypes, to support workers effectively without interfering with their tasks that provide real-time cognitive assistance, task-level feedback, and safety enhancement.

This approach focuses on industrial activity analysis and optimization by identifying the structure of industrial workflows and assembly processes through which workers benefit from cognitive augmentation in complex operational environments.

As a result, the **expected impact includes:**

- Support and guidance to the workers in real-time particularly to novices through real-time feedback.
- Optimization of the task performance by notifying the worker of incomplete steps.
- Quality control and quality assurance, by recognizing and counting performed actions.

Thus, minimizing the errors in the final product suggests less rework time for workers and less additional costs for the employer. Furthermore, the collaborative behavior of humans and machines could reinforce the confidence of newly employed people and reduce the duration of their training. Furthermore, the proposed hierarchical taxonomy and feedback mechanisms may provide structured training support and skill enhancement, promoting progressive learning and effective performance monitoring.

The scalability of this approach is addressed by the use of tools such as low-cost wearable technologies, adaptable deep learning models, and versatile feedback systems, extending their applicability beyond specific workflows into broader industrial contexts. Finally, this work intends to enable the development of recognition technologies and the awareness of industrial assembly processes in uncontrolled environments, allowing researchers and practitioners to design systems, sensors, and algorithms that align with the required level of abstraction to enhance workers' capabilities.

1.5 Organization of Thesis

Chapter 1 introduces the concept of cognitive augmentation in regard to manual assembly workflows, discussing the need for such improvements in human intelligence in the industrial field. Furthermore, the main research challenges are identified, leading to the formulation of the research questions that are presented in Chapter 3.

Chapter 2 presents an overview of the cognitive augmentation landscape in industrial environments and defines the key concepts by providing the theoretical background of the terminology used throughout the thesis. Moreover, it focuses on state-of-the-art approaches and methodologies in cognitive computing, HAR, sensor systems, assistance systems, and models to optimize assembly workflows.

Chapter 3 outlines the research methodology design for cognitive augmentation in industrial settings. It discusses the connection between research and industry, key challenges, technologies, and approaches, followed by the research questions and an overview of the methodology used in this study.

Chapter 4 provides more details for the structuring of industrial settings and the sensing mechanisms that are frequently used in those. It continues with the explanation of methods for the data collection process and introduces i) the development of a taxonomy to categorize activities as well as ii) how this taxonomy can be used to guide the design of AI systems, in manual assembly workflows.

Chapter 5 builds on the previous chapter and explores the degree of understanding of certain activities of the assembly process. It describes the experimental setting, the data processing pipeline, and the models that were used to identify and classify work steps in a real assembly workflow. Furthermore, it explores new approaches to segment or split time-series data in order to use them to count activities that are part of the workflow. As part of the experiment, the details of an in-lab real use case are analyzed and the development of a deep learning model to address the aforementioned work is described.

Chapter 6 reflects on the results of the previous chapters by applying the described theoretical concepts to real prototype applications that were implemented throughout this research. It identifies the goals and challenges encountered and outlines their expected impact.

Chapter 7 summarizes the key findings and contributions of the cognitive augmentation of manual assembly workflows subsequent to the taxonomy development, AI system guidance, activity classification, and counting in order to enhance human capabilities in industrial assemblies.

2 Background

2.1 Framing the Background

Industrial tasks become more complex, and thus require more support to the workers, to achieve daily work goals, especially during tasks with high cognitive load. Figure 2.1 shows causes for manufacturing errors, which often arise from such complexity [25].

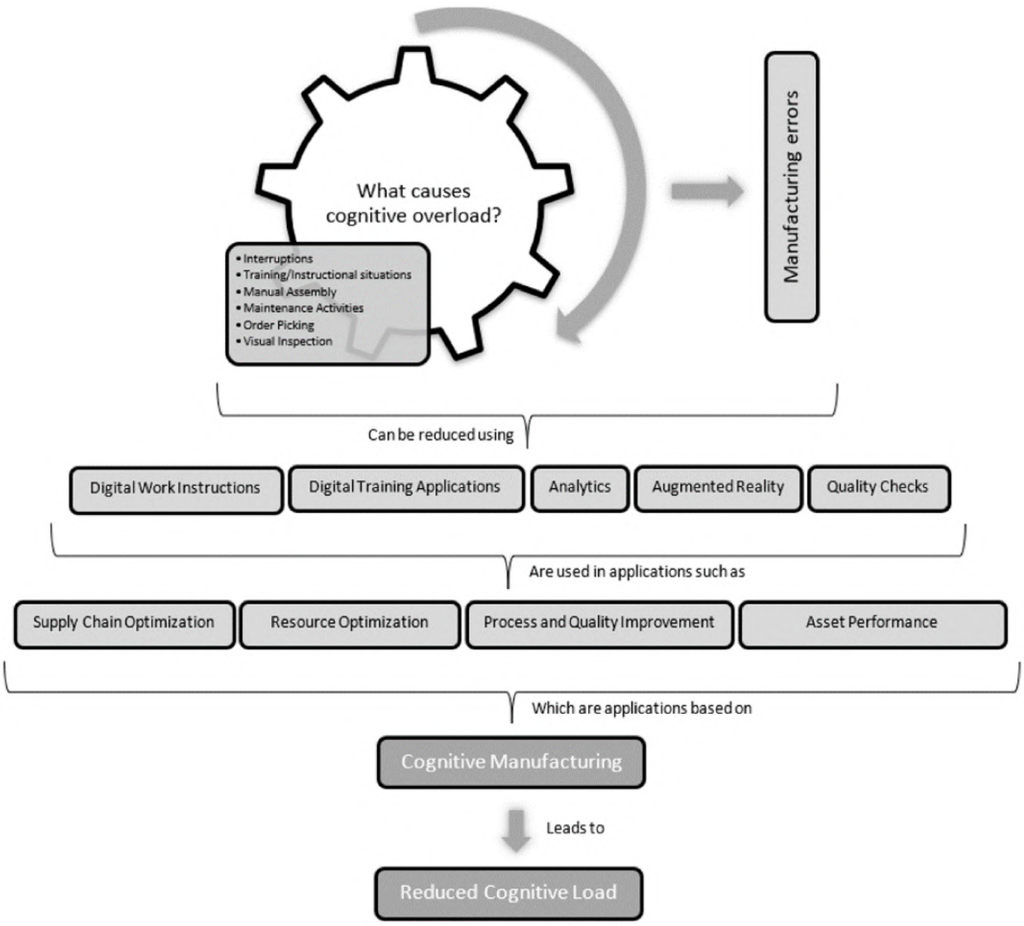


Figure 2.1: Source: Reproduced from [25]. The diagram illustrates the causes of cognitive overload which leads to manufacturing errors. Interruptions, manual assembly, training/instructional situations, order picking, and visual inspection among others are found to contribute to manufacturing errors. Accordingly, solutions to reduce it, are visualized such as digital work instructions, augmented reality, analytics, and quality checks to mitigate cognitive overload. These solutions support applications in supply chain optimization, resource optimization, process, and quality improvement, and asset performance, as discussed by the authors, to minimize cognitive load and manufacturing errors.

Three types of cognitive load are considered in the cognitive load theory:

(i) the intrinsic (related to the complexity of a topic or a task),

(ii) extraneous(related to how the information of a task is presented),

(iii) and germane(associated with the mental effort to create new knowledge and skills)

based on the work by the authors in [26]. In the manufacturing processes, intrinsic load is usually high due to the assignment of complex and variable tasks to the human workers, and extraneous load is observed due to the lack of technological support, to create a clear structure and order for the precise execution of a workflow [25]. Support is usually offered to them with printed or digitally visualized instructions and steps to follow, however, cognitive assistance goes beyond the basic tools that are provided, offering real-time awareness, guidance, and feedback to optimize the human-centered complex industrial workflows [27, 28].

Understanding an operator's cognitive load is important during human-machine interaction or execution of tasks. Tasks with high cognitive load, especially in industrial settings like assembly tasks, are characterized by the mental effort required to process large amounts of information and maintain concentration while performing physically demanding tasks due to i) task complexity, ii) the need to follow detailed instructions and iii) rapid decision-making, often worsened by high-paced environments [29]. In their work, Biondi et al., relate the impact of high cognitive load with increasing assembly task completion times, resulting in affecting the manufacturing cycle times [30]. According to another study, with 75 workers and engineers participating, the cognitive performance of assemblers is influenced by factors such as task design, timing, physical demands, motivation, teamwork, and assembly interface design, with positive and negative effects on the final result [31].

As it becomes clear, the increasing complexity and high demands of industrial tasks require providing sophisticated cognitive assistance to the workers. At the same time, progress in technology and computing systems demonstrates positive results in supporting human activities through its evolution.

2.2 Advancements in Industrial Processes

2.2.1 Transition to Pervasive Computing

Computing systems have always been used to support human activities, but in recent years significant changes and transformations in their functionality have led to the development of different approaches to address the growing complexity and the needs of society [32]. In the past, computing was developed on large centralized mainframe computers. However, advances in the fields of microchips, clusters, packet-switching technologies, and GUIs helped switch to Personal Computers (Pcs) [33]. Furthermore, globalization of network standards increased worldwide communication and data sharing [34]. This change provided businesses and industries with the necessary tools to integrate sensors and actuators with built-in network connectivity to delegate tasks to remote computing devices for processing, memory, and storage purposes [35]. The growth of technologies such as the Internet of Things (IoT) and edge computing expanded the competencies of traditional network nodes that distribute capabilities across decentralized systems [36]. These computing models that evolved from mainframes to decentralized architectures reinforced the rise of pervasive computing, where technology is seamlessly integrated into everyday environments.

Pervasive Computing

Pervasive or ubiquitous computing are two terms used to describe the integration of computing technology seamlessly into daily life. "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [37]. In these technologies, the devices operate in the background without the users consciously interacting with them [38].

Nowadays, these two concepts overlap and are sometimes used interchangeably with pervasive computing focusing on a system that reacts based on actions, using sensing technologies [39] and intelligent algorithms that process data and make decisions in real-time without explicit direct input [40]. The authors in [41] mention that pervasive computing, also known as ubiquitous computing, is a growing field that enables the embedding of computational capabilities into everyday devices to minimize the consumption of resources (such as battery, memory, and CPU usage) while enhancing the devices' ability to collect, process, and communicate various data types.

As technology evolves, pervasive systems become more context-aware and able to adapt to their environment and input to improve human experiences and lifestyles. The integration of pervasive technologies like sensors in industrial assembly environments to capture real-time data on human activities through HAR systems assists in the optimization and improvement of the overall assembly process.

2.2.2 Cognitive Systems in Industrial Processes

Enhancing task accuracy, work effectiveness, and error management are among the main focus of research studies that aim to overcome limitations in human cognitive abilities. Cognitive augmentation or enhancement explores ways to improve human cognition "given its potential for reducing errors in complex operational environments, but also for occupational psychology to improve work performance, mitigate risks, and improve job stress/well-being", as mentioned by the authors in [10]. In this context, systems that can modify their behavior are implemented to meet human capabilities, and the term "cognitive system" is used, to describe new software or hardware solutions that model human intelligence and provide continuous information about the status of production steps [5]. These cognitive systems are based on sensor networks that are capable of real-time data acquisition, interaction, and understanding of processes to assist in daily tasks. The authors in [42] emphasize the use of wearable cognitive assistants to improve the ability of employees to meet job demands while supporting autonomy and providing adaptive feedback, resulting in more efficient workflows and clearer communication.

Cognitive Assistance in industry

Following the progress of industry towards flexible, sustainable, and human-centered industry paths, new models and approaches are needed for the future, where collaborative Human-AI systems are expected to play a key role [43]. In Industry 5.0 humans are crucial for developing productive systems [15] and are essential in the workplace [44]. In addition, with the increasing diversity and complexity of tasks in industrial operations, the demand for cognitive assistance to employees also increases. Furthermore, interaction between Humans and AI rises as well, whether it is intermittent, proactive, or continuous [45, 27]. As visualized in Figure 2.2 the intermittent paradigm expects the human operator to initiate the interaction, while in the proactive paradigm, the cognitive assistant plays an active role as it begins the interaction. This work focuses more on the continuous aspect of cognitive assistance with the system working in the background providing online information to the operator.

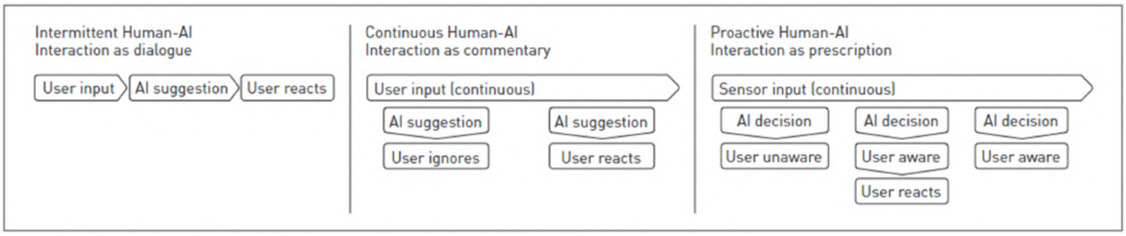


Figure 2.2: Source: Reproduced from [45]. The figure illustrates a visual representation of human-AI interaction across three paradigms discussed by the authors. In the *Intermittent Interaction* (Dialogue) based on the user's input, the AI provides suggestions, and the user reacts. In the *Continuous Interaction* (Commentary) the AI provides ongoing suggestions based on continuous user input, which the user may ignore or react to. In the *Proactive Interaction* (Prescription) the AI autonomously makes decisions based on continuous sensor input, with the user either being unaware or aware and reacting when aware.

The leverage of wearables and machine learning models in order to convert sensor data into meaningful insights for activity recognition leads to exploring the human activity recognition (HAR) field in order to understand how these technologies can enhance worker awareness and productivity in industrial settings. Additionally, worker assistance systems simplify interactions between humans and complex machines, reinforcing both the physical and cognitive skills of employees [46]. These systems offer diverse functionalities that support operators in their daily work, as mentioned by the authors in [47], including:

- **Increased:** physical and cognitive support, speed and productivity, quality control, comfort and convenience, ergonomics, worker capacity, worker safety, worker integration, and location independence.
- **Decreased:** mental stress, language barriers, and search times.

- **Enabled:** health control.

Recommendations derived from the work of Marl et al., [47], correlate assistance systems with specific user groups in the industry, as appropriate, with sensor-based systems found to benefit the flexible or unskilled worker. They consider the worker's skill level, age, physical abilities, experience, and the variety of work to provide matching cognitive systems to the requirements of each identified group. Continuing in the same direction Pokorni et al., propose a Cognitive Assistance System - Quality Function Deployment (CAS-QFD) that focuses on the worker but aligns both worker benefits as well as company benefits [48].

The implementation of cognitive technologies in a construction machine manufacturer was discussed in [18], where the authors demonstrate how human-centered approaches can enhance productivity in manual assembly environments by integrating sensing and guidance devices to support workers in high-variety tasks. As stated in [28], "Cognitive systems are capable of human-like actions such as perception, learning, planning, reasoning, self- and context-awareness, interaction, and performing actions in unstructured environments". This framework also presents an understanding of the potential of cognitive technologies to enhance worker capabilities by placing humans at the center of assembly tasks. Therefore, human remains a very important factor in the loop of assembly tasks as their activities have an impact on the result of the workflow process. These activities are explored in the context of human activity recognition (HAR), which provides insights into the execution of tasks and the interactions of workers with tools and components within their environment to optimize industrial assembly workflows.

2.3 Human Activity Across Domains

Human activity recognition (HAR) is the research area that involves developing systems to identify and categorize human activities based on sensor data. This field has applications across various domains such as industry, sports, healthcare, and smart homes, where understanding human actions can lead to better user experiences, improved safety, and optimized performance.

Important steps and many results have already been published in the field of activity recognition and assistance systems to date, where daily activities that people are willing to monitor through recognition systems can be an asset for their DL. Useful information can be collected and used as an aid in healthcare, well-being, sports life [49] and in other personal or professional areas of everyday life [50, 51, 52]. In activity recognition, there are two main categories associated with the types of devices used: i) body-worn sensors, and ii) statically deployed sensors. In these categories, options such as video or IMU or combinations thereof provide the necessary data for each case. Complementary methods to the IMUs were proposed in [53, 54], where the authors explore alternative sensing modalities and fusion techniques to enhance the activity recognition performance in monitoring leg motion activities and inspection activities in a production environment.

Wearable sensor-based activity recognition has emerged as a promising approach for monitoring behaviors and promoting healthier lifestyles. The authors in [55], investigate the abilities of six kinds of sensors to recognize 14 different human activities in a kitchen scenario that were performed by 10 volunteers and achieve an accuracy of 92.44%. In [56], Wang et al. give an introduction to the state-of-art sensor modalities in HAR in health care, where traditional machine learning and deep learning approaches were applied. Bian et al. in their work [57] provide a comprehensive survey of human activity recognition (HAR) sensing techniques, categorizing them into five classes and highlighting their strengths and limitations, which can inform the selection of appropriate sensing modalities for AI-assisted activity recognition in various settings. Bulling et al. [58] discuss the key research challenges that human activity recognition shares with general pattern recognition and identify those challenges that are specific to human activity recognition i.e. the inter-class similarity, intra-class variability, and null class problem.

2.3.1 Advances In Activity Classification

Multiple studies the recent years, intended to detect human activities by utilizing conventional approaches that rely on feature extraction and statistical methods. However the advance in deep learning has been fundamental in encouraging its use in complex projects of computer vision, speech recognition, automated translation, the medical field, the robotics industry, and many others [59] offering high-quality modeling and comprehension of data, along with challenges [60].

With the increasing adoption of new technologies, devices, and sensors, people generate continuously more and more data to support their daily activities. Researchers can use this sensor data to identify the human body's actions and movements for human activity recognition, or HAR, as it is more commonly known. Various sensor types collect data in those settings, such as the ones that use video and inertial measurement units (IMUs). Sensors provide the means to capture data related to human activities, which can be used to develop machine-learning models for human activity recognition (HAR) and human behavior recognition (HBR). Achievements have been made in sports and entertainment [11, 12, 13], industrial applications [15], and healthcare [14].

Meanwhile, the academic community is actively researching innovative sensor technologies for human activity and behavior recognition, including new sensor designs, applications of traditional sensors, and the usage of non-traditional sensor types [3]. Many studies are currently being conducted to improve existing approaches or solve newly identified problems for the detection and classification of activities with supervised or unsupervised techniques. The majority of research studies and applications in the field of HAR have, up to this point, focused on detecting activities, such as walking, standing, and sitting, as well as other daily living (DL) activities [61], and analyzing their characteristics to generate new insights [62].

A variety of deep learning models had been implemented using wearables in [63] with a focus on recognizing human activities in DL. Zhao et al. [64] proposed a deep network architecture using residual bidirectional (LSTM) where the Opportunity and UCI public domain data sets were used for the experiment. Results in the work by the authors in [65], where deep learning models were compared in terms of accuracy, memory requirements, and speed for activity recognition classification tasks on ten public data sets, showed the superiority of CNN architectures with respect to the other networks. Similarly, as discussed by the authors in [66], the results show better performance for repetitive tasks in two public data sets for CNN models and the bi-LSTM model for the Opportunity data set. As mentioned in the work of the authors in [56] "LSTMs are recommended to recognize short activities that have natural order while CNN is better at inferring long-term repetitive activities. The reason is that RNN uses the time-order relationship between sensor readings, and CNN is capable of learning features contained in recursive patterns".

Bohra et al. [67] implemented a high-accuracy HAR classification model that outperforms state-of-the-art prediction accuracy on raw signal data and is suitable for deployment in resource-constrained environments of smart wearables and IoT devices. Similarly, Guinea et al. [68] create image representations of the time-series data to take advantage of the strengths that convolutional neural networks (CNNs) have shown when dealing with image data with their experimentation being limited to publicly available datasets that do not pose the complexities of industrial assembly tasks. Likewise, Khaertdinov et al. [69] investigated HAR using Deep Triplet Networks but relied solely on benchmark datasets rather than real-world data, limiting their model's applicability to practical scenarios. The study in [70] uses wearable-based sensing for human activity recognition, and deep learning to enhance feature extraction and classification accuracy. While wearable devices provide valuable motion data, the authors' focus remains on locomotion activities. On the other hand, Gupta et al. [71] recognize human activities beyond just locomotion, and demonstrate the potential of their models for activity classification that was tested with the WISDM dataset. Although these studies highlight the effectiveness of deep learning in HAR, their application on controlled datasets might have an impact on their generalization ability to real-world applications in practical deployment.

In [72] by Jiang et al. a novel activity image is constructed out of accelerometer and gyroscope signals which then was used as input for the deep convolutional neural networks (DCNN) for automated learning of features and evaluated in three public data sets [73, 74, 75]. Yang et.al [76] investigated the multi-channel time series data acquired from body-worn sensors, with their proposal of a CNN to increase recognition accuracy when applied to the Opportunity activity recognition data set. Data obtained from a smartphone (UCI-HAR data set) was used in [77] to utilize their architecture of a temporal convolutional network for recognizing 6 activities. Ordonez et al. [78] proposed a combination of convolutional and LSTM recurrent units, which is appropriate for multi-modal wearable sensors and eliminates the requirement for expertise in designing features.

The proposed deep network in [79] could handle various sequence measurements from different body-worn devices separately, improving the effectiveness of CNN through convolutions per sensor channel and per body-worn device. The evaluation of their deep learning architecture is performed and evaluated on three data sets, the Opportunity data set, the Pamap2 data set, and an industrial data set, outperforming the state-of-the-art approaches. The authors in [80] used the previous datasets, and included in their study two additional datasets, to investigate the use of adversarial learning to improve generalization and robustness in human activity recognition using wearable sensors via Self-KnowledgE Distillation. In contrast, Diaz et al. [81] employed low-resolution infrared (LRIR) sensors to develop a privacy-preserving approach based on a prototype recurrent convolutional network (PRCN) with few-shot learning for recognizing locomotion activities where one or two individuals engage in daily tasks.

In this thesis, Long Short-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs) are employed for classification tasks. Unlike conventional approaches that rely on preprocessed data, the models in this work are trained on raw data that are collected in a real industrial environment during normal worker shifts using the minimum necessary number and type of sensing devices. This approach forces the models to learn and identify activities directly from unprocessed raw data, without any prior feature engineering or feature extraction.

2.3.2 Advances In Activity Counting

Counting is one valuable skill that a human uses daily in diverse activities and tasks, for simple or more complicated tasks that might benefit from technological assistance in various fields. According to the conducted literature review, the majority of studies in this area mostly used video-capturing sensors and have been conducted in the sports or medical sector.

Fang et al. [82] in their study, explore the possibility of counting the number of items in a display and raise the question, "Can a recurrent neural network learn to count things?" with their findings favoring a positive answer. While they also use an LSTM model, the developed model in the subsequent Section 5.3.4, takes as input raw IMU data relating to human activities in daily life. In a different setting, the authors of [83] propose to count repetitive activities in a video by sight and sound using an audiovisual model which differs, from the approach presented in this dissertation, among others in the choice of the sensors, since the aim is to use body-worn sensors. In [84], the MM-Fit dataset is introduced, which contains data from inertial sensors and ambient video sensors capturing full-body workouts. A single 3d accelerometer worn at the chest is employed in [85] to recognize four types of workouts and count repetitions after the workout is firstly determined and classified by their algorithm. Another study focused on fitness exercises is the one in [86] where the authors select camera sensors for realizing their approach to do workout repetition counting. The researchers in [87] designed and implemented a body capacitance-based sensor and employed a residual deep convolutional network that uses dilated convolutions for recognizing and counting gym workouts. While their approach had competitively high counting accuracy, the below-proposed approach opts for sensors available in devices of daily use such as smartwatches or smartphones, and utilizes raw data. An error of ± 1 repetition in 91% of the performed sets for Cross-fit, exercises is achieved in [88] where the authors used a vibration signal during their data collection and trained a neural network for counting that relied on whether an input window containing a repetition start. However, the model that will be presented in the following Chapter 5.3.4, uses only the weak labels as target data for the variable size input.

Weakly labeled data

Weakly labeled data can be beneficial for deep learning algorithms in certain situations and refer to data that is only partially labeled, meaning that it has some form of annotation, but not all the information is present. This type of data is less expensive and time-consuming to obtain than fully labeled data, and it can be used to train deep learning models in a semi-supervised manner. The authors in [89], propose an attention-based convolution neural network to process weakly labeled human activities and recognize them. The dataset contains information only about the type of activity that occurred in a sequence of sensor data. A weakly labeled dataset was also included in their Dual Attention Network For Multimodal Human activity recognition Using Wearable Sensors from the writers of [90] where they blend channel attention and temporal attention on a CNN, for multimodal HAR. The activities that are contained in the dataset are walking, jogging, jumping, going upstairs, or going downstairs, and have a significant difference from the activities that we explore and the way that we create our training dataset.

Raw data

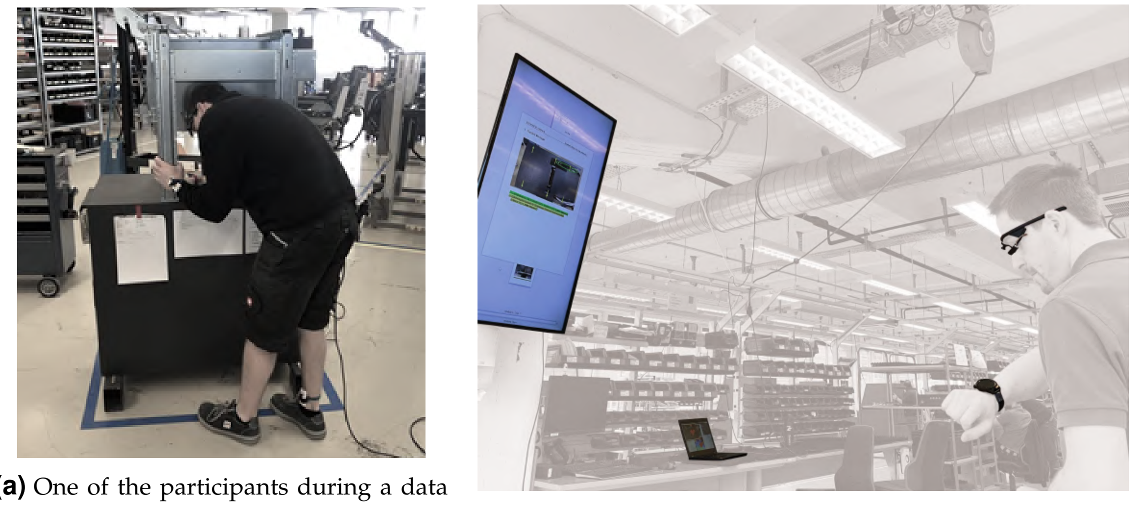
Raw data as input for the models has the advantage of reducing the need for pre-processing techniques, which can be a time-consuming and resource-intensive task. When working with raw data, the model can automatically learn useful features from the data, which can save computational resources and reduce the risk of human error. Important contributions have been made by Shen et al. in [91], where they propose a workout tracking system that uses smartwatches to accurately and efficiently track both cardio and weightlifting workouts without the need for user input. Their counting strategy begins with detecting and labeling weightlifting sessions, followed by a naive peak detection algorithm based on auto-correlation results. They filter out non-repeating signals and calculate the number of repetitions by counting detected peaks. Likewise, Prabhu et al. in [92] also based their approach on classifying the activities before counting with a peak detector method. Their research aims to identify the most effective artificial intelligence model for repetition counting in LME exercises to be used in wrist-worn rehabilitation programs.

In their work, Taborri et al. in [93], implemented the following algorithms, one for recognizing activities based on SVMs and one for counting actions related to workers in the industry. Twenty-three body-worn sensors collected data from the participants which were divided into windows of 0.6s and features such as mean, standard deviation, maximum, and minimum were computed for each activity. Physical exercises for indoor and outdoor environments can be recognized with the real-time segmentation and classification algorithm from the writers of [94]. The method they propose requires one sample of data for each target exercise, however once more the counting relies on accurate classification of the activities. Examples of data with the few-shot learning were employed by Nishino et al. [95] to recognize workouts using a wearable sensor including data augmentation and diversification techniques for their data to achieve repetition counting.

In the presented approach, raw data are used, as input to the deep learning models, from public datasets containing human activities. This data is divided into segments of variable size without filtering and with weak labels that utilize only the number of repetitions for each sequence as the target.

2.3.3 Applications of HAR in Industrial Processes

Several studies as well, dealt with the support of employees in their DWL by implementing various human-machine collaboration systems. However, studies that record their own real-world data for application-specific use cases remain limited. The figures 2.3a, 2.3b illustrate examples of wearable devices used for data collection and application in industrial settings. Stiefmaier et al. in [96] monitored the contribution of wearable multi-modal sensor systems for activity recognition in production environments and more specifically in car manufacturing. The authors in [97] select masonry as a task to record acceleration data from wrist-worn sensors in a laboratory. Walking or stationary activities are observed in [98] and their classification depends on a shoe-based movement sensor. Makaewa et al. in [99] reported on an effort to detect activities from workers "by estimating the lead time of each period of an operation process" in an unsupervised manner.



(a) One of the participants during a data recording collection, wearing IMU sensors and the eye-tracker.

(b) One of the participants during a test session focusing on the devices of the assistant system in his workspace.

Figure 2.3: The figures show an exemplary visualization during a test assembly task and the assistant system's components where a participant interacts with wearable devices for capturing data and uses feedback devices to get support in his workspace.

The results of clustering frequent micro actions such as screwing in industrial environments, are investigated in [100], while a multi-modal approach utilizing a smart armband and a camera for activity recognition in assembly tasks is described in [101], similar to the previous study but with different sensors. In the same context, Qingxin et al. [102] propose unsupervised methods for factory activity recognition using body-worn accelerometer data, and focus on the challenge of obtaining labeled sensor data for each worker. While their argument is correct, supervised learning remains the preferred approach when labeled data is available, as it typically yields higher classification accuracy.

2.4 Complexity in Industrial Processes

In industrial assembly, human activity recognition requires significant attention due to the critical role humans play in the process [103]. Assembly tasks, characterized by intricate processes and workflows in different fields, are the focus of HAR, which aims to develop models for understanding and classifying these activities, ranging from simple actions to complex, multi-step processes [104]. However, while this categorization assists in assessing complexity, it may overlook factors such as context and granularity. By acknowledging the connection between industrial assembly and HAR, it can be explored how insights from activity recognition research can enhance the optimization of manufacturing and assembly operations.

2.4.1 Basic to Advanced Task Categorization

Two main approaches prevail in the literature: binary class and non-binary multi-class approaches, as shown in Table 2.1, which groups research publications from various domains, such as manufacturing, into recognized activity abstraction categories.

Table 2.1: This table presents an overview of the existing literature in the domain of human activity recognition, presenting a quantitative distribution of papers across different abstraction levels and domains. The dominance of publications at the simple/complex level across various domains suggests a significant focus on understanding activities in this distinction. On the other hand, the comparatively small number of publications in some areas, such as group activities, suggests possible topics for more investigation and study. Furthermore, the existence of publications at various levels of abstraction highlights the complexity of human activity recognition research and emphasizes the necessity for sophisticated methods of activity analysis and classification.

Abstraction Levels							
Approaches:	**Binary**				**Non-Binary**		
Domain	**Atomic-Simple, Complex-Composite**	**Composite-Gross, Fine-Grained**	**Micro, Macro**	**Low, High-Level**	**Gestures, Actions, Interactions, Group Activities**	**Atomic action, Primitive Task, Task**	**Various, Other Terminology**
Manufacturing, Robotics, Construction	[105, 106, 107]	[108]	[109]	[110, 111]		[112, 113, 114, 115, 116, 117]	[118, 119]
Healthcare							[120, 121, 122]
Sports					[123]		[124, 125]
ADLs	[126, 127, 128, 129, 130, 131, 132, 133]			[127]			
IADL	[134, 132]	[135]	[134]				
Group					[136]		
Other Domain, Unspecific HAR	[137, 138, 133, 139, 140, 141, 131, 142, 104, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152]	[153]					[154, 1, 155, 156]

Twofold Task Categorization

As mentioned by the authors in [105, 106, 107, 127], compared to **simple activities, complex activities** are composed of actions and are much more complex and semantically consistent with a human's real life. Ramanujam et al. [144] attempted a division of public HAR datasets that apply deep learning techniques into simple and complex activities, categorizing them into conventional and hybrid models. Nevertheless, the problem remains without the additional distinctions because any task that is more complicated than simple is automatically labeled as complex. Meanwhile, Bouton et al. [143] investigates complex daily activities in sedentary settings such as remote work or study environments, concluding that gaining insights into an individual's daily activities can help develop applications that improve their well-being and overall health. The work from [132] exhibits higher detection accuracy and less 'inter-subject variability' as mentioned by the authors since simple types of activities are performed similarly across users. Additionally, they observed confusion between simple and complex activities such as "cooking and standing" or "sitting on sofa and lying on sofa". Chen et al. in [140] separate human activities into simple human activities (SHAs) and complex human activities (CHAs), where SHAs may be recognized with an accelerometer, whereas CHAs need multi-modal sensor data.

The distinction between activities is crucial in HAR; however, existing taxonomies for human activity abstraction levels are often limited to simple and ambiguous categories that cannot capture the complexity and diversity of human activities across various domains [11, 133, 138, 139, 141, 147, 148, 149, 150, 151, 152, 157, 158, 159, 160, 161]. Zhang et al. in [162] explored complex activities such as eating, which involves a variety of movements, and reported that they increase the challenge and difficulty for HAR methods. For example, an activity such as "cooking" involves actions like "cut", "take", or "mix" in different order and frequency. The authors in [109, 134] refer to these actions as "micro-activities" or **micro-motions** and characterize the complex activities as "macro activities" or **macro-motions**. In their work, they discuss how micro-motions serve several purposes in understanding macro-motions. This includes i) confirmation of the execution of all necessary micro-motions; ii) facilitating evaluation of their sequence correctness; and iii) recognizing differences among macro-activities or their execution variations.

Different types of activities require different granularity levels, as mentioned by the authors in [128]. In their study, they utilized the terminology "complex activities" of daily living (ADLs), which are built on top of simple activities and convey more specific contextual information, and "simple or coarse-grained activities", such as walking, sitting, and cycling, which may be directly assessed from an inertial sensor unit.

The terms "coarse-grained and fine-grained activities" and "low-level and high-level activities" are also used to describe different aspects of activity abstraction. The division between **coarse-grained or gross** and **fine-grained activities**,which reflects the level of detail or granularity in activities, appears in industrial manufacturing [108] but also in various other domains [153] including daily activities [135]. Coarse-grained activities are broad categories or high-level actions, while fine-grained activities are more specific and detailed. Fine-grained activities provide a richer understanding of the activity by capturing smaller sub-actions or variations within a broader category. However, this division may not explicitly address complexity or context.

On the other hand, the distinction between **low-level** and **high-level activities** refers to the hierarchical organization of activities [110, 111]. Low-level activities are typically microscopic or elementary actions that form the building blocks for higher-level activities. High-level activities encompass a collection of lower-level actions or represent more abstract concepts. This division emphasizes the hierarchical relationship between activities and can be useful for understanding the composition and structure of complex activities. Investigating this topic is interesting since, in most settings, only a few activities are considered (less than 10), and only a few of them discuss hierarchical dependencies between activities on lower and higher levels [163].

Previous research has proposed various levels of activity abstractions, ranging from activities such as walking and sitting to motions such as lifting and grasping. The hierarchical approach to human activity recognition involves recognizing simpler activities initially and using them to recognize higher-level activities. The representation of high-level activities is based on the sub-events or sub-activities that serve as observations derived from the higher-level activity [136]. The use of sub-events not only makes the recognition process computationally tractable and conceptually understandable but also reduces redundancy in the recognition process by reusing recognized sub-events multiple times. An example given by the authors in [136] is that the high-level activity of "fighting" may involve detecting a sequence of sub-events such as punching and kicking interactions.

By distinguishing between activity abstractions, researchers can develop methods or approaches that are tailored to the specific level of abstraction [100, 164], thereby improving activity recognition models.

One main finding derived from the authors in [125] is that "activities involving several body parts are more easily recognizable and allow for shorter window sizes". Yet, they state that the recognition accuracy for such complex activities is still low. This is due to several factors like i) the inter-class similarity, ii) the difficulty in defining each activity and its boundaries, and iii) the lack of open datasets. Furthermore, activity abstractions provide means of interpretability that enable comprehension of the behavior of the assisted user or system. The ability to identify the appropriate level of abstraction is critical for

i) developing effective activity recognition models, ii) better understanding and gaining insights into human behavior, and iii) modeling human behavior in a given context.

Several ongoing challenges persist, including issues such as incomplete information within activity datasets, insufficient contextual details accompanying activity data, and the complexities involved in modeling composite activities [104]. Composite activities, as highlighted in the literature by [1, 135], pose particular challenges due to their composition of multiple shorter activities and the associated temporal decomposition required for their recognition. Kulsoom et al., in [155], extends the concept of composite activities and presents additional categorizations based on operation types, namely concurrent, sequential, and interleaved activities. Moreover, despite the growing interest in interaction recognition, the recognition of activities involving groups of people [123] has received relatively less focus, as identified by Morshed et al. [156].

In their work, the authors in [13] state that "the existing taxonomies in the field of activity recognition, while valuable, exhibit limitations in their categorization by not encompassing sufficient distinct activity categories, thereby indicating the need for further refinement and improvement". The presented studies provide evidence of the prevalence of binary distinctions and the limitations in existing methodologies. Thus, describing activities as simple or complex, high or low level, and fine or coarse-grained hinders a detailed enough understanding of the underlying components and structures of human activities, the different sub-components of activity, and how they relate to each other. Therefore, there is a research gap in exploring beyond the simple–complex dichotomy, which can be addressed by incorporating additional levels of activity abstraction.

Multifold Task Categorization

The hierarchical framework provided by Kuutti et al. [165] offers insights to address the limitations of binary classifications in human activity recognition. In this, the authors explain how activities consist of chains of actions, with each action comprising individual operations, where individual actions become comprehensible only when viewed within the larger context of the activity. Recent findings by Miranda et al. [142] suggest that a promising strategy for dealing with complex HAR is to model activities as a series of dependent atomic actions. In this context, Saguna et al. [146] focus on the semantics of the application and explain it as a fundamental activity unit that cannot be further decomposed.

Following the previous explanation and using comparable methods in the existing literature, we refer to the framework of **Atomic Actions, Primitive Tasks, Tasks** that comprises three hierarchical levels: task, primitive task, and atomic action [112, 113, 114, 115, 116, 166]. An assembly task is decomposable into multiple continuous primitive tasks, which can be executed sequentially or in parallel. Similarly, each primitive task can be further broken down into continuous atomic actions, which are also capable of sequential or parallel execution. While the proposed framework addresses the binarization encountered in the literature regarding activities, our approach proposes different classifications for levels or actions involved in certain activity stages, respectively. The proposed taxonomy provides a more refined classification for some actions than existing frameworks to cover the complexities of human activity in industrial contexts. Hence, it is suggested that certain actions, as defined in their context, can be further decomposed or integrated into a broader hierarchy of activities.

The challenge in activity recognition lies in the generalization of models across diverse contexts [159] through recognizing patterns in sensory input. Conventional approaches often oversimplify activities as either simple or complex without accounting for contextual variations. Furthermore, human activity recognition tends to be associated with locomotion activities, and often instead of employing categorizations based on factors like complexity or granularity, many studies opt for a broad approach to activities without specific classification [118, 120, 121, 122, 124, 154]. An additional aspect was addressed by the authors in [131] where they provide insight into the characteristics of complex activities, highlighting that complex activities exhibit longer duration, comprise a combination of simple activities, and encompass multiple behaviors. These complex activities often possess high-level semantics, such as daily activities like cooking, cleaning, or industrial assembly tasks.

Examining binary and multi-class approaches revealed that certain categories of activities often receive less attention, as they are often processed using the same models designed for other categories. However, these activities are important in the recognition and classification process and require customized methods due to their distinct attributes and complexities. Hence, it is critical to distinguish the nuanced differences between activities across various contexts and categorize them appropriately to leverage context-specific characteristics for understanding user behavior to improve activity recognition systems in real-world settings.

3 A Research Design for Cognitive Augmentation

3.1 Entanglement of Research and Modern Industries

The digitization of manufacturing is often combined with continuous development and the demand for improved efficiency, quality control, and adaptability especially in discrete manufacturing and assembly processes [167]. However, while many manufacturers are interested in, or are planning to include AI in their workflows, only a few have applied it successfully until recently [168]. Many industries cannot entirely modernize their older equipment since it cannot easily be removed or is costly to replace. As a result, equipping their operations with additional systems can provide assistance in their workflows which utilize sensors, actuators, and AI-driven solutions to augment worker capabilities, reduce errors, and enhance quality assurance. In other cases, the task itself cannot be automatized due to complexity reasons and this requires the human abilities strictly involved in the process while progressing to intelligent manufacturing [169].

3.2 Key Challenges in Industrial Domain

As industrial processes continue to evolve the need to address emerging issues and challenges also grows proportionally, requiring sophisticated solutions that are also identified and stated in the literature. Alqahtani et al. explored causes that affect manual assembly processes and revealed factors such as low training level, experience, poor workplace layout, and fatigue, among others to determine the number of occurring human errors and the quality of a final assembly product [20]. Their findings contribute to the decision-making of companies and organizations regarding their strategies to reduce errors, reduce the cost of rework, and provide assistance and guidance to employees. Supporting and guiding systems, particularly those implemented in ubiquitous flexible and mobile devices such as wearable sensors, offer potential solutions for production efficiency and improved worker performance.

Although there are examples of wearable sensor applications in the field of HAR, they are mainly focused as mentioned above on sports, or health fields, leaving space to investigate their use in the industrial field [170] as presented in Figure 3.1. Specifically, the image shows that the application of wearables in the industrial sector is minimal with a value of 0.53% compared to their application in fitness or other areas where it exceeds 30%. The

Figure 3.1: Source: Reproduced from [170]. The figure illustrates the major areas and applications of wearable devices in HAR as presented by the authors. The Overview of wearable devices in human activity recognition (HAR) shows that wearables have an increasing popularity in various fields, with fitness (37.65%) and lifestyle (32.92%) being the most common while the application in industry is only 0.53%. Moreover, the figure shows the most common placements and the distribution of body locations for wearable sensors on the body, with the wrist (59.28%) selected as the most frequently used placement.

authors of [171, 172], discuss the opportunities and challenges of artificial intelligence (AI) and the limited application of wearables in the industrial domain. This issue is closely related to the findings in [173] where the authors mention the important factor of low-cost sensors for assistive and collaborative technologies in workplaces. Apart from that, the adaptability and flexibility of an assistant system, support the operator's freedom and mobility inside his workspace. Therefore, he can perform more complex activities, while bottlenecks and efficiency issues in industrial lines and manufacturing can be identified [174].

Another important aspect is the limited (inadequate) availability of datasets from industrial environments, specifically data from employees performing their daily tasks [174] which could support the development of more models and increase the robustness and efficiency of the existing ones. Additionally, exploring alternative approaches for data segmentation and the potential of deep learning-based models in production lines [176] could further enhance the applicability of cognitive systems in manual assembly workflows. Figure 3.2 visualizes the percentage of each data source that is used for HAR applications in the literature provided by the authors in [175]. Their findings show that CCTV (52%) and mobile sensors (26%) are the most frequently used data sources for human activity recognition (HAR), while other sources like wearable body sensors (8%), YouTube videos

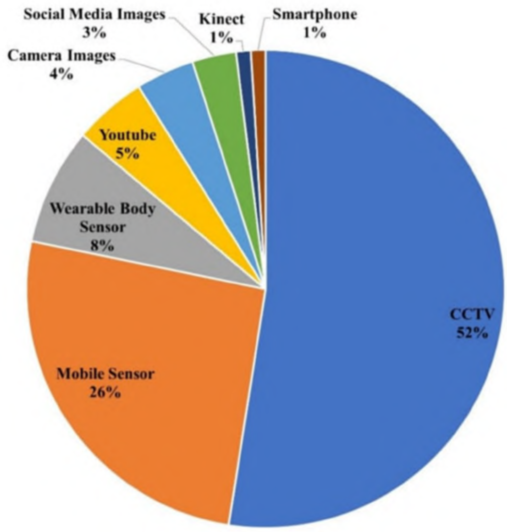


Figure 3.2: Source: Reproduced from [175]. The figure demonstrates the distribution of data sources that are used for human activity recognition (HAR), highlighting CCTV (52%) and mobile sensors (26%) as the most common sources. The wearable body sensors are used less frequently (8%) based on the work from the authors in [175]

(5%), and social media images (3%) are used less often. Some studies also combine multiple sources to improve accuracy.

Fahle et al. provide additional support for research in this direction highlighting two valuable limitations in their systematic literature review on machine learning methods for manufacturing processes: "As a result of this paper, two major gaps in the research were revealed, shown by the **very few to none existing paper on the topics of assistance systems** and learning factory training concepts. Although **assistance systems have been the focus of research for the past years, the implementation of ML in these systems have not been widely mentioned in literature."** [177]. Subsequently, the authors in [27] identify the need for :

(i) further case studies applications of worker assistance systems,

(ii) methodologies for deciding the appropriate one for specific user groups, and

(iii) a methodology for suitability evaluation of worker assistance systems.

Figure 3.3: The figure presents a visualization of the identified research gaps in human activity recognition (HAR) and assistive technologies, highlighting key challenges such as the need for: low-cost sensors for assistive technologies, for understanding complex industrial tasks, for human-AI collaboration protocols, issues related to datasets availability, wearables in industry, deep learning approaches in production, case studies for assistance, and error detection in manual assembly. The related literature is available in [6, 20, 27, 170, 171, 172, 173, 174, 176, 178, 179, 177, 180].

These aspects align with the research challenges depicted in Figure 3.3. The figure visualizes a condensed and organized summary of the identified gaps as a result of a literature review on activity recognition in industrial places. Furthermore, it highlights the opportunity for systematic approaches in designing and implementing HAR-based assistive technologies in industrial settings.

In the development of HAR systems that are robust, practical, and worker-centric, it is essential to move beyond lab-based evaluations and study human activity directly within industrial environments. This is necessary due to the specific challenges and the inconsistent conditions that are found in those settings. Industrial environments include real-world complexity that cannot be simulated in labs such as variable lighting, background noise, and workspace constraints that are hard to replicate. This affects also the sensor-capturing process, signals, and overall data quality while ground truth labeling can also be easier addressed in the lab setting than in real settings.

Another important aspect is the unpredictable human behavior which is much more restricted in the lab where participants follow pre-defined steps. In contrast, real workers adapt, skip steps, multitask, interact with coworkers, and interrupt or change workflows on the fly which increases the complexity of the task analysis. Finally, the evaluation of cognitive augmentation systems regarding usability, acceptability, or long-term system robustness, can only be assessed when they are deployed in the actual industrial setting and under real conditions, capturing user fatigue, and distractions, which labs fail to reproduce. This work addresses that need by conducting extensive data collection and system development within a manufacturing environment.

3.3 Key Technologies & Approaches

In 2022, ElMaraghy et al. wrote in their study that "Cognitive technologies perform AI-based supplementary tasks that help make better decisions and complete objectives and tasks that usually require human intelligence, such as planning, reasoning, and learning". Figure 3.4 visualizes how to move from data to cognitive adaptation and from smart to cognitive manufacturing systems. As it is stated, the Adaptive Cognitive Manufacturing Systems (ACMS) paradigm integrates intelligent technologies to improve resilience, responsiveness, and adaptability, to create innovative and transformative changes in modern manufacturing systems and reach cognition [5].

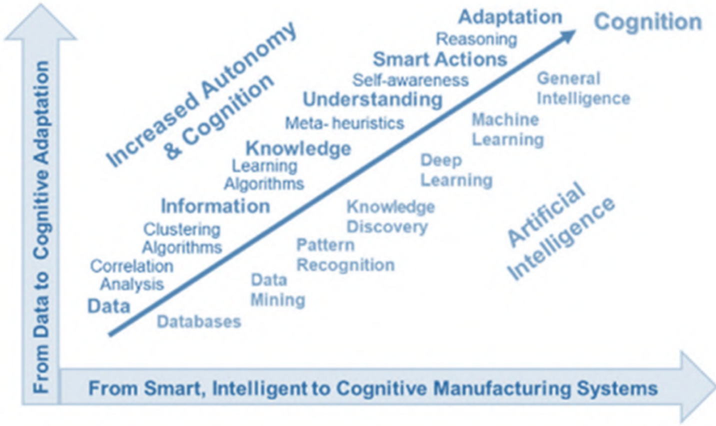


Figure 3.4: Source: Reproduced from [5]. The figure illustrates the evolution from data to cognition in manufacturing systems. The authors highlight key stages that lead from data to cognitive adaptation and from smart to cognitive manufacturing systems.

A more recent definition in [181] outlines cognitive manufacturing as "intelligent cyber-physical manufacturing capable of perception, decision-making, and reacting by utilizing information obtained throughout the whole product life cycle." The authors also provide an overview of the current capabilities, trends, and technologies used in the field. Among others, they identify trends such as sensor fusion techniques and diverse sensor inputs for data acquisition, as well as supervised machine learning and deep learning approaches.

As previously discussed, humans remain a central element in the discrete manufacturing and assembly processes. Therefore, it is essential to analyze and understand their actions within these systems to enhance efficiency in modern manufacturing environments. Figure 3.5 illustrates the taxonomy of existing literature in HAR systems presenting the application areas, data sources, techniques, and challenges associated with recognizing and analyzing human activities across different environments as it was realized in [175].

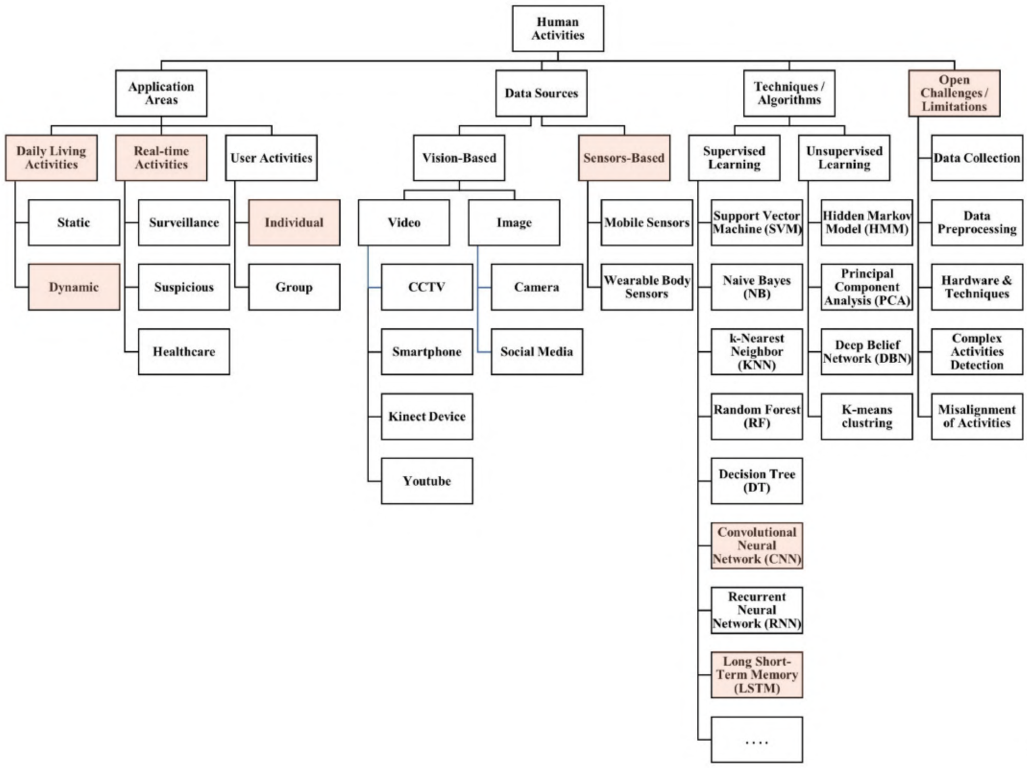


Figure 3.5: Source: Reproduced from [175]. The figure illustrates the categorization of human activities based on application areas, data sources, techniques, and open challenges as discussed by the authors in their work. The highlighted categories (shaded in color) indicate the specific aspects addressed in this research. These refer to dynamic daily work activities that are performed by individuals and can be identified in real-time by sensor-based devices. This work uses supervised deep-learning approaches to address open challenges such as data collection, preprocessing, and complex activity detection.

In this thesis, a combination of these technologies and approaches will be used to address the open challenges that are described in Figure 3.5 and the challenges discussed in the previous section. Specifically, the focus is on the data collection scheme that was followed, the preprocessing techniques, the utilized sensors, and the models that were implemented to augment the cognitive capabilities of a worker. The activity recognition chain ARC framework proposed by Roggen et al., [182, 78] is well-suited in activity detection for repetitive and manual operations in dynamic work environments. This processing pipeline is visualized in Figure 3.6.

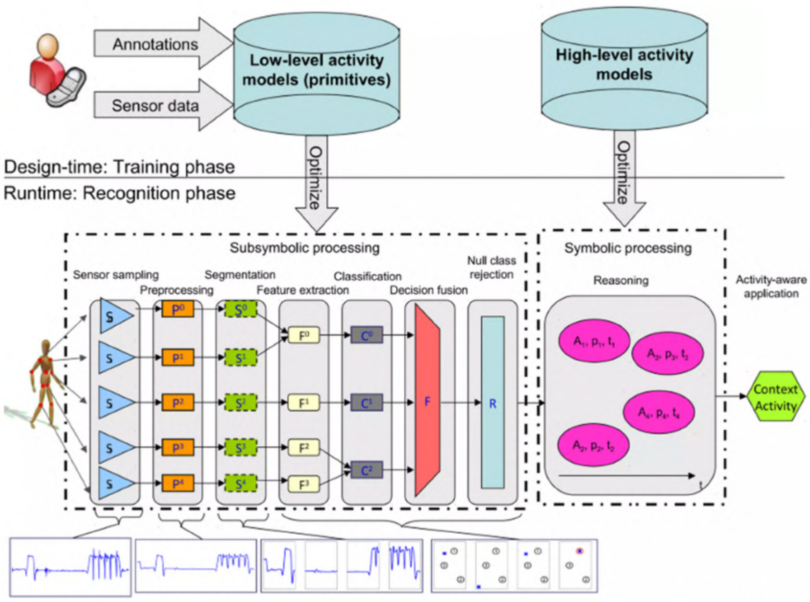


Figure 3.6: Source: Reproduced from [182]. The figure illustrates the Activity-Recognition Chain (ARC) that is used to identify human activities from sensor data as outlined by the authors in their work. Raw sensor data are mapped to action primitives (events) through signal processing and machine learning techniques. Subsequently, complex activities, are identified by analyzing detected events.

3.4 Research Questions

Based on the former issues identified in the literature and in practical industrial applications, the goal of this thesis is to explore how to enhance the abilities of the worker in daily assembly tasks by investigating the following challenges:

- **RQ1: How can structured levels of activity abstraction in manufacturing workflows be defined to enhance cognitive assistance on assembly processes?**

  Distinct assembly hierarchies of activity abstractions exist within the workflow of a manufacturing process, each contributing uniquely to the overall assembly sequence, and efficiency as observed in the order of occurring activities and the characteristics of the produced IMU signals.
  Related Publications by Author:[183, 18]

- **RQ2: How can various sensor modalities (*wearable and non-wearable*) be utilized to recognize and categorize manual activities in industrial settings, to support AI system design for enhanced human-machine collaboration? (across abstraction levels?)**

  Building on the understanding of activity hierarchies, the next step is to explore the tools and techniques for capturing these activities and the practical application of sensors in industrial settings, addressing the technical aspects for effective activity recognition and classification by optimizing sensor placement, employing robust data preprocessing techniques, and selecting appropriate machine learning models. This integration enhances collaborative intelligence by providing accurate, real-time activity recognition, which supports human workers in complex tasks, improves safety, and increases overall productivity
  Related Publications by Author: [183, 160, 184]

- **RQ3: What methods can be employed to detect and extract information on key activities within manual assembly processes to assess the recognition of task execution in industrial workflows?**

  This question investigates the methods for extracting meaningful information from key activities(micro), which are the smallest detectable units of action that achieve a defined objective, crucial for detailed analysis and understanding of the assembly process.
  Related Publications by Author: [185, 186, 100, 160]

- **RQ4: How can machine learning techniques be leveraged to count repetitive actions in time-series data from industrial workflows, addressing ambiguous annotations and varying task durations?**

  Having identified the micro-activities, the next challenge is to automate their quantification. This question examines the use of advanced machine learning techniques, particularly Deep Neural Networks, to count repetitions of micro-activities using raw IMU data, ensuring efficient monitoring.
  Related Publications by Author: [187]

3.5 Research Methodology Overview

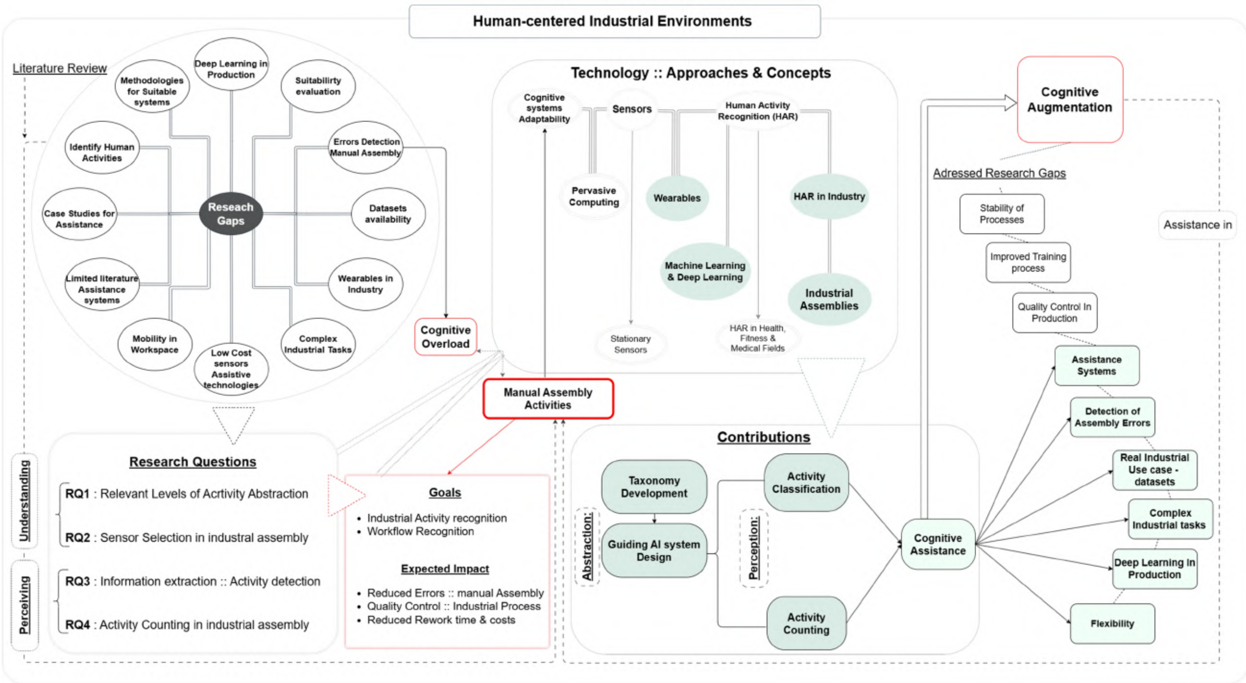


Figure 3.7: The figure outlines the thesis structure which focuses on cognitive augmentation within manual assembly workflows. It is organized into five main sections: Research Gaps, Research Questions (RQ), Technology and Approaches, Goals and Impact, and Contributions. The figure utilizes color coding to emphasize the core elements of the thesis with red, with green highlighting the technologies and approaches that were employed and are directly connected to the contributions. The lighter green emphasizes the aspects that cognitive assistance supports through literature and is also addressed in this thesis. The identified gaps through the literature review and from the insights provided by industrial experts highlight high cognitive load as being responsible for manual assembly errors among other issues such as error detection, industrial dataset unavailability, and the need for assistive systems with low-cost sensors. The research questions follow the gaps guiding the selection of technologies and approaches toward the goals and expected impact. Finally, the contributions are visualized that feature the taxonomy development, the activity classification, and the activity counting that bridge theory and practice in cognitive assistance for industrial assembly tasks.

This section outlines the research methodology that was followed to provide cognitive augmentation through assistance in human-centered industrial environments following the approach described in [188]. In Figure 3.7 the road map is divided into sections that connect to each other and contribute to the addressing of the challenges of manual assembly activities. The diagram organizes the research process into key areas: research gaps as discussed in Chapter 2 and in Section 3.2, research questions as presented in Section 3.4, the thesis aims with the expected impact and contributions as described in Section 1.4. The technologies and approaches that were used to support the practical research are briefly noted in Section 3.3 and elaborated in the respective section for each topic.

This thesis uses a combination of methods providing qualitative insights and quantitative data analysis. For the realization of the research problems of this thesis apart from the research gaps that were identified in the literature, scientific exchange and discussions were conducted with industrial experts to identify problems and challenges in their manual assembly lines. Identified challenges led to the formation of research questions and hypotheses to achieve goals.

In the context of this thesis and based on the existing literature and experts' opinion, a taxonomy was developed to categorize different levels of complexity in activity abstraction within the industrial assembly. The selection of models and tools is guided by the developed taxonomy, which recommends practices and methods to apply those concepts and support workers in real-world industrial settings. Various models for classification and counting tasks were explored, that leverage supervised machine learning algorithms and use multiple types of raw input data. Their performance was evaluated using standard metrics such as accuracy, precision, recall, F1-score, mean absolute error (MAE), and confusion matrices, along with visualizations to analyze results and compare them with the ground truth annotations. Additionally, some models were tested in real-world industrial settings at the industrial site to assess their practical applicability. All the insights were also properly communicated to the research community and the different stakeholders involved in this thesis, including the industry partners. The details of the model selection, training process, and evaluation methodology are described in detail in the corresponding dedicated chapter for each topic.

Assistance to the employee to reduce his cognitive load which is strongly related to manual assembly errors, is among others one common focus between industry and academia. Addressing this challenge requires state-of-the-art algorithms that can employ low-cost sensors for HAR, flexibility in the workstation, and complex activity recognition in industrial workspaces to automate activity recognition, enable workflow recognition, and provide real-time information, feedback, and cognitive assistance to assembly workers. Therefore, they become aware of the type and number of occurred activities, make better decisions during the process, and reduce their cognitive load. Significant impact is expected as result of the worker's cognitive augmentation, in the quality control of the final product, the reduction of human errors, and the reduction of rework time and costs.

4 Abstraction in Cognitive Augmentation

4.1 Introduction to Activity Abstraction

At the core of industrial operations lie the activities performed by individuals, or individuals in combination with machines or automated systems [189]. Cognitive augmentation aligns with and considers the challenges in the field of industrial assembly tasks, where technology will be applied to support workers, reduce cognitive overload, and stabilize processes.

According to the authors of [190], manual assembly tasks have a clear distinction regarding complexity. Objective assembly complexity refers to inherent properties of the assembly process, such as the number and dependencies of components. In contrast, perceived assembly complexity is subjective and influenced by personal capabilities and experience. From the manipulation of individual components to the orchestration of complex production lines, each activity plays a crucial role in shaping the outcome of the manufacturing process. Yet, the complexity and diversity of these activities present difficulties for researchers, engineers, and workers in the field who want to understand them better to enhance performance and could benefit from a standardized approach.

Traditional approaches for activity analysis have often relied on recognizing few activities and in simplistic categorizations, such as simple versus complex activities [104, 191]. Simple activities, such as walking, running, sitting, or standing, are extensively researched, whereas everything else is categorized as complex. Consequently, the recognition of complex activities remains relatively unexplored, as highlighted by the authors in [137]. While these frameworks provide a basic understanding of activity complexity, they prove ineffective in capturing the nuanced structure and hierarchical organization of industrial tasks or miss the depth and detail needed to analyze and categorize industrial activities. Activities can be categorized along a scale of increasing complexity, as discussed by Schneider et al. [126], where simple activities occur for a short period, while complex activities may occur for more prolonged periods. Furthermore, Peng et al. [130] state that the features designed for simple activities are poor at representing complex activities.

This work proposes a novel hierarchical activity abstraction framework inspired by real industrial assembly operations, tailored to the specific level of abstraction, ensuring that the activity model fits the task in industrial contexts. Within this framework, distinguishing between activity abstractions becomes critical since, e.g., distinct statistical, temporal, and spatial properties appear at different levels of abstraction that go beyond the binary spectrum or the existing non-binary approaches. For example, activities such as grasping or positioning are short in duration, repetitive, and localized. In contrast, activities, such as assembling a full module or operating across multiple stations, span longer timeframes and involve broader spatial interactions.

These distinctions can lead to more efficient sensor placement, information capture, feature extraction, resource utilization, and architecture design. Beyond improving recognition performance, this structured approach aims to enhance the interpretability, and modularity of AI-based cognitive augmentation systems in real-world manufacturing environments, and facilitate scalability by providing a systematic framework to adapt to larger or more complex scenarios. The recognition models can be independently trained and reused for each level without the need to replace the full pipeline, while the information from a combination of models at different levels can improve the feedback to the user and the interpretability by providing clear guidance (e.g., 3 out of 4 actions occurred during the assembly of process X). Additionally, the assistance can be customized based on the experience level of the operator offering personalized and detailed feedback to the novice worker and more abstract progress tracking to the advanced worker.

4.2 Structuring Abstraction – Development of Activity Taxonomy

A taxonomy is proposed using a hierarchical activity abstraction structure that combines the simplicity of the simple–complex division with the clarity of the non-binary approaches to study activities at different levels of granularity and abstraction along with their respective sub-divisions. Additionally, this work extends beyond taxonomy development to offer practical applicability and relevance in real-world industrial assembly scenarios, where workers can benefit from cognitive augmentation to enhance efficiency, reduce errors by providing feedback and information, and streamline daily work tasks.

4.2.1 Industrial Assembly Process

In industrial manufacturing, assembly processes are essential for joining multiple components to produce functional products [103]. Whether it is assembling automobiles, electronic devices, or machinery, the efficiency and effectiveness of the assembly process impacts the success of the manufacturing operation. Workflow optimization depends on an understanding of the hierarchical structure and relationships between different phases of assembly. Every stage of the assembly process on a manufacturing floor, typically involving a frame (e.g., chassis) where modules are systematically assembled to construct complex products, requires coordination, precision, and adherence to defined work instructions. In [192], the authors describe the hierarchical nature of complex product assembly data, highlighting three granularity levels of product, assembly, and part. They state that the refined management of assembly processes and hierarchical organization of assembly data can be achieved by decomposing complex assembly activities into more detailed activities.

Building upon this framework, the analysis of real-world assembly processes revealed the hierarchical structure observed in complex assembly scenarios [103, 185, 192, 193]. Moreover, this analysis aligns with findings from the literature regarding the roles of modular assembly and units or sub-assemblies in the assembly process. By breaking down complex assembly tasks into modular units, manufacturers can

(i) enhance efficiency and flexibility in their production workflows,

(ii) separate design tasks into distinct units,

(iii) and simplify the design process

while prioritizing product customization [194, 195] and responsiveness [196]. The authors in [197] highlighted the importance of the number of modules, the joining sequences between modules, and the tolerance management issues in car body design. In assembly processes, modules are self-contained subsystems designed to be constructed, examined, manufactured, and developed independently from the overall system for independent integration. This ensures interchangeability, standardization, and re-usability, with clear interfaces and relative independence from other modules [198].

In addition, this work recommends adding the post-assembly processes [199] as a level to the assembly process hierarchy where the concept of the final product is contextualized within the scope of the specific manufacturing line. This indicates that what constitutes the final product may vary depending on the manufacturing line. For instance, what serves as the final product on one manufacturing line might be considered a sub-assembly or component in another line. With this feature, industrial processes are observed, covering every stage of the manufacturing process, from components to finished products, by performing the required manual activities for each stage. Therefore, the assembly chain example presented in Figure 4.1, for understanding various industrial workflows, does not extend to subsequent stages of product development or distribution but focuses on the assembly process of a product, which is completed on the production line.

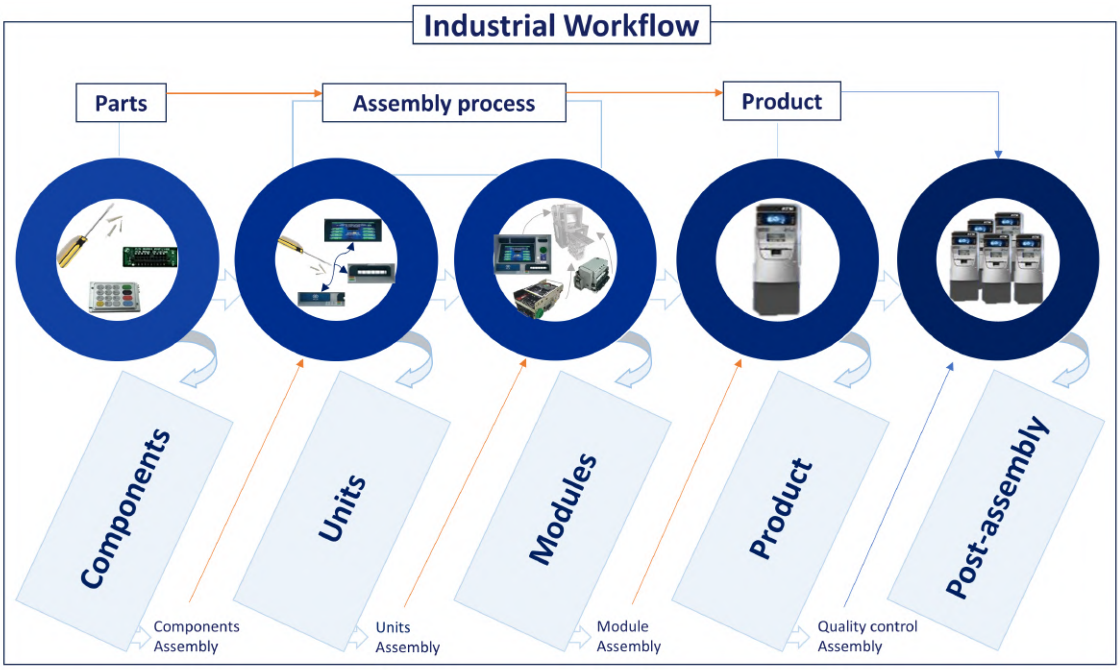


Figure 4.1: This figure presents a hierarchy of an exemplary assembly process: components, units, modules, products, and post-assembly. Additionally, it demonstrates how these stages are interconnected and how activities and tasks flow inside a real assembly scenario from components to the final product. Each stage builds upon the previous one, with components being assembled into units, units into modules, modules into the final product, and finally, the product being integrated into the production line.

4.2.2 Abstraction Levels of the Taxonomy Framework

The proposed taxonomy presented in Figure 4.2 introduces five stages atomic, micro, meso, macro, and mega providing a hierarchical framework for analyzing and categorizing activities within the assembly process of discrete manufacturing products to understand the progression of assembly tasks. In this framework, the term “activities” is employed for the different levels to ensure consistency, adopting a unified terminology throughout our analysis. Moreover, the systematic use of the terms atomic, micro, meso, macro, and mega, denotes the distinctions between activities at various stages of the assembly workflow.

Atomic: This level includes the smallest self-contained operations or steps that can be performed by a human in a manual product assembly scenario involving the basic components or tools for discrete or singular manipulations within a broader action or activity. Examples of such atomic operations or sub-tasks that contribute to completing a particular action could be grasping the screw or tool, positioning the screw or tool, turning the screw or tool, and releasing the screw or tool.

Micro:This level includes the smallest recognizable actions within a task or process that serve a specific goal, detectable using sensors or observation, requiring tools to execute certain tasks like joining or attaching components. They are composed of a sequence of atomic steps, generating singular activities, including operations that can be executed by a single individual without extensive planning or coordination. They are characterized by their relatively short duration and repetitive pattern, and they can be performed independently or as part of a larger activity as an individual work step. At the micro-level of assembly, individual tasks are executed by integrating multiple atomic operations, each involving distinct components, tools, or parts. The sensor data required to detect screwing as an activity can be limited to the movement and orientation of the tool, as this information alone contains indicative signals of the screwing action. An example of a micro-activity is the entire process of picking up a tool, a screw, and a component, positioning the screw in the part, and using the tool with rotating movements to tighten a screw.

Meso:This level includes a collection of coordinated actions and operations from the micro-level and/or atomic level, forming a coherent sequence to achieve an objective driven by a particular motive. This level bridges the gap between fundamental atomic and micro-level processes toward increasingly complex processes, representing a crucial stage in assembly workflows to accomplish intermediate goals typically undertaken by a single individual or a small group of individuals. At the meso-level, sub-assemblies, and modules of the product are prepared through the aggregating micro-level activities, which are each designed with specific functions and aims. They are self-contained units that meet intermediate ends, which are essential steps toward the final assembly. An example of a meso-level task would be to prepare the sub-assemblies of a module, such as the cash box for an ATM.

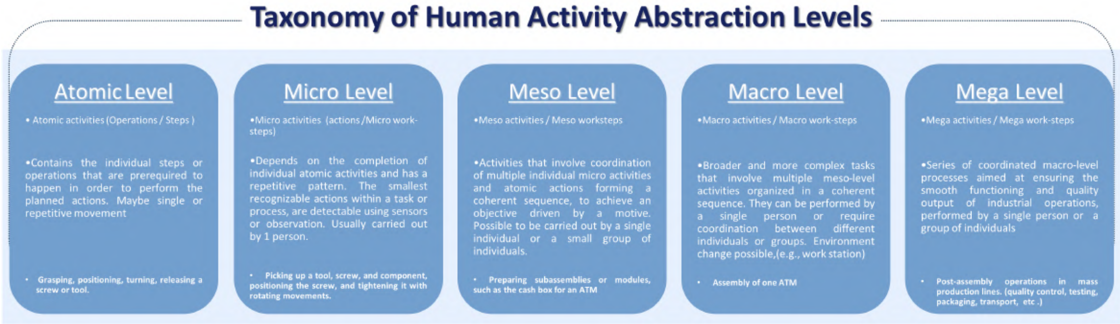


Figure 4.2: The figure presents a visualization for the proposed taxonomy. At the atomic level, individual assembly activities are considered as singular tasks involving basic operations or manipulations on discrete components or tools. The micro-level aggregates multiple atomic operations into coherent sequences, representing actions within the assembly process. Larger assembly tasks are formed at the meso-level by combining multiple micro-level activities, often involving the assembly of sub-components or partial assemblies. The macro-level encompasses entire assembly processes, including stages such as the assembly of major components or modules. Finally, the mega-level represents the overall assembly process, incorporating post-assembly activities such as quality control checks, packaging, or final inspection.

Macro:At the macro-level, tasks evolve into broader and more complex tasks, encompassing multiple steps and components from the previous meso-level. This process involves coordinating multiple meso-level activities, which include the assembly of sub-assemblies and modules prepared in earlier stages. Macro-activities can involve sequences of actions, including those demanding higher-level cognitive functions like decision-making and planning. These tasks might require coordination between individuals, groups, or machines (robots) where the outcome of this coordination results in the accomplishment of a complex goal, such as the full assembly of a product (e.g., the ATM) depicted in Figure 4.1.

Mega: This level comprises a series of coordinated macro-level processes to ensure the smooth functioning and quality output of industrial operations. In a broader sense, mega-level activities may involve various tasks conducted by humans beyond the assembly line operation yet included in the industrial workflow. These could include packaging products for safe transport and inspection, testing functionality, or conducting quality-control checks. These activities contribute to the overall goal of achieving efficient and effective production processes on a large scale. Examples include the post-assembly operations that exist in mass production lines of individual products, such as cars or other complex items assembled from multiple modules or sub-assemblies.

Following the presented hierarchical structure of complex processes in industrial manufacturing, focusing on the assembly of products, these concepts align with the industrial example illustrated in Figure 4.1. This example provides an overview or a reference to the presented approach rather than a comprehensive representation of all applications included in an industrial setting.

In the assembly of automated teller machines (ATMs), the hierarchical structure of assembly processes is evident across separate abstraction levels. At the component level, where atomic activities are involved, screws, tools, and individual parts like display panels form the foundational building blocks of the ATM. These components are then manipulated and used to create the required sequential work steps that contribute to developing units at the micro-level. They often involve repetitive tasks such as tightening motions with rotating movements using tools. While the micro-level focuses on one specific single action every time, the meso-level relies on the coordination of multiple actions to complete larger modules like cash-handling systems or user interface systems that are assembled independently during that level. All modules, units, and the remaining components are assembled into a complete ATM system at the product or macro-level. Subsequently, in the post-assembly stage, which represents the mega-level, the final product is integrated into the manufacturing line, undergoing additional processes such as quality control, software installation, functionality testing, and packaging.

In addition to the hierarchical structure demonstrated within the ATM assembly, the presented activity recognition taxonomy extends beyond this specific application to cover various industrial contexts. To validate its versatility, a hierarchical formulation that represents assembly processes using equations has been developed. This conceptual representation, through equations, serves as an initial tool for the activity recognition taxonomy, providing insights into the intricacies of assembly tasks to understand the relationships between activities within industrial assembly workflows. In this model for assembly processes, a set of symbols is utilized to quantify various attributes at different hierarchical levels. The specific step of the assembly activity, performed at a given level, is indicated by the symbol A_{level}, and the time spent on each step of the activity is indicated by the symbol t_{level}. Moreover, a_{level} represents the kinematic characteristics associated with the assembly step or operations at that level, like acceleration or angular velocity. The equation P_{level} expresses the overall occurred activities at a specific level within the taxonomy framework. Throughout the model, summations aggregate the contributions of individual activities, providing an analysis of assembly workflows.

Hierarchical formulation of our taxonomy in an assembly process

Attributes:

- P_{level} : Activity on level of taxonomy.
- A_{level}: Step of assembly activity.
- t_{level}: Time spent on step of assembly activity.
- a_{level}: Kinematics properties on step of assembly activity

$$P_{\text{atomic}_i} = f(A_i, t_i, a_i)$$

$$P_{\text{micro}_j} = \{P_{\text{atomic}_j}\}_{j=1}^{n}$$

$$P_{\text{meso}_k} = \{P_{\text{micro}_k}\} \cup \{P_{\text{atomic}_k}\}_{k=1}^{m}$$

$$P_{\text{macro}_l} = \{P_{\text{meso}_l}\} \cup \{P_{\text{micro}_l}\} \cup \{P_{\text{atomic}_l}\}_{l=1}^{p}$$

$$P_{\text{mega}_q} = \{P_{\text{macro}_q}\} \cup \{P_{\text{micro}_q}\} \cup \{P_{\text{PA}_q}\} \cup \{P_{\text{atomic}_q}\}_{q=1}^{r}$$

4.3 Comparative Analysis with SOTA

This section compares the introduced taxonomy to the current state-of-the-art (SOTA) approaches that exist in the literature, while the following section presents the application of the taxonomy to real-world industrial scenarios. After reviewing relevant publications outlined in Section 2, a categorization methodology was generated for an assembly scenario, to visualize the differences between binary approaches, non-binary approaches, and the proposed one. Figure 4.3, illustrates a simplified process assembly of an ATM illustrating some key steps that capture a subset of the entire assembly, as the complete one typically involves numerous additional steps and complexities. The example is derived from original scenarios and real-world use cases documented in prior work [185, 186, 160, 100, 164] by the author, discussing ATM and digger assemblies in more detail.

Approaches		Binary			Non-binary	
Task	Activities	Simple / Complex	Fine-grained / Coarse-grained	Low level / High level	Atomic Action / Primitive Task / Task	Atomic / Micro / Meso / Macro / Mega
Lifting, a tool / attaching a screw to the frame	Lifting, Grasping, Placing, Turning	Simple	Fine-grained	Low level	Atomic Action	Atomic
Performing a single type of operation, on a component	Bolting, Securing, Tightening, Screwing	Complex	Fine-grained	Low level	Atomic Action	Micro
Performing multiple types of operations on components to complete a larger task. e.g., Assembling a cash handling system	Lifting, Aligning, Fitting, Fastening, Routing, Connecting, Securing, Insulating	Complex	Coarse-grained	High level	Primitive Task	Meso
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Performing multiple types of operations on components to complete a larger task. e.g., Assembling screen, keyboard,etc.	Lifting, Aligning, Fitting, Fastening, Routing, Connecting, Securing, Insulating	Complex	Coarse-grained	High level	Primitive Task	Meso
Performing multiple types of operations on components to complete a larger task. e.g., Assembling front panel module	Lifting, Aligning, Fitting, Fastening, Routing, Connecting, Securing, Insulating	Complex	Coarse-grained	High level	Primitive Task	Meso
Complete all the mounting processes of the workstep. e.,g., Installing sub-assembled components. ATM assembly	Mounting, Securing, Connecting,	Complex	Coarse-grained	High level	Task	Macro
Quality Control of multiple products,Inspection ATM systems, Operating the assembly line, Safety, performance of assembled product	Diagnosing, Testing, inspection, Analyzing, Troubleshooting, Verifying, Supervising, Coordinating, Managing, Optimizing	Complex	Coarse-grained	High level	-	Mega

Figure 4.3: The table illustrates a simplified ATM assembly process, derived from a real industrial use case [185], showcasing activities across different assembly levels: atomic, micro, meso, macro, and mega. It serves as a comparative analysis with existing approaches for activity categorization, highlighting how each level contributes to the overall process. Specific activities are provided for clarity, offering insights into the hierarchical organization of assembly tasks.

In simple/complex, fine-grained/coarse-grained, and low-level/high-level classifications, a two-stage categorization method between detailed and broader classifications as outlined in the related work is observed. These approaches emphasize the distinction between the initial assembly tasks as sub-actions and their combination as broader categories. For instance, in the simple/complex approach, a screwing process already belongs to the complex category, since it is considered much more complicated compared to lifting a hand or grabbing an object. Additionally, in the fine-grained/coarse-grained, low-level/high-level categories, lower-level classifications consist of activities similar to screwing, as the primary stage, and continue afterward to more demanding activities.

At the atomic level, activities involve manipulating individual tools and components through discrete or singular manipulations. These actions serve as the building blocks for more complex tasks. For instance, lifting a tool or grasping a single screw constitutes an atomic action. Moving to the micro-level, tasks involve the execution of specific actions that combine interactions between at least two elements from the atomic level to form a single type of action. For example, screw tightening involves grasping, lifting, and turning a screw with a tool. The micro-level is characterized by repetitive actions, such as repeatedly rotating a screw to fix it tight. At this level, a difference can be noticed between the proposed methodology and the non-binary approaches reported in the literature. In the proposed methodology, the screwing process occupies a different level compared to lifting or grasping because it is considered a combination of those actions.

Transitioning to the meso-level, activities become more comprehensive, involving the combination of micro-level tasks to accomplish intermediate goals. For instance, assembling an ATM's front panel system requires the integration of micro-level activities. At the macro-level, larger components and subsystems, previously prepared at the meso-level, are brought together to construct the final product. This step involves coordinating multiple meso-level activities, each contributing essential components or modules designed with specific functions and aims. For example, assembling the ATM at the macro-level entails integrating pre-assembled sub-parts, such as the cash-handling system and front panel to the chassis into a cohesive structure.

The proposed methodology considers the assembly process as an integral part of a larger manufacturing ecosystem rather than an isolated event. In this regard, at the mega-level, the focus extends beyond assembly to encompass broader activities such as quality control, inspection, and overall process optimization, as opposed to existing approaches that often overlook such human tasks. In the second example of an industrial assembly process, the focus shifts from screwing processes to welding processes within the context of car assembly. This hierarchical framework presented in Figure 4.4 remains the same, including the five levels: atomic, micro, meso, macro, and mega, illustrating the progression of tasks.

Approaches		Binary			Non-binary	
Task	Activities	Simple / Complex	Fine-grained / Coarse-grained	Low level / High level	Atomic Action / Primitive Task / Task	Atomic / Micro / Meso / Macro / Mega
Holding two pieces together to be welded	Holding, Aligning, Positioning	Simple	Fine-grained	Low level	Atomic Action	Atomic
Performing a single type of operation, on a component	Arc initiation, Arc maintenance, Weld penetration, Arc termination	Complex	Fine-grained	Low level	Primitive Task	Micro
Performing multiple types of operations on components to complete a larger task subassembly/module	Component positioning, Joint alignment, Welding sequence execution, Quality assurance	Complex	Coarse-grained	High level	Primitive Task	Meso
Welding all subassemblies/modules together to form the final product	Subassembly integration, Structural welding, Quality control, Final inspection	Complex	Coarse-grained	High level	Task	Macro
Conducting quality control checks on welded components	Visual inspection, Dimensional measurement, Weld strength testing, Defect detection	Complex	Coarse-grained	High level	-	Mega

Figure 4.4: The table illustrates welding processes in car assembly, presenting the hierarchical framework of tasks, and showcasing activities across different assembly levels: atomic, micro, meso, macro, and mega. It serves as a comparative analysis with existing approaches for activity categorization, highlighting how each level contributes to the overall process and showing how individual actions aggregate into more complex tasks across the assembly line. Specific activities are provided for clarity, offering insights into the hierarchical organization of assembly tasks.

4.4 Application of the Taxonomy for Guiding AI System Design

This section proceeds to the practical application of the taxonomy by presenting its deployment in real-world industrial scenarios. The analysis identifies and proposes key categories to guide or support practitioners in the research and implementation of AI applications and systems in industrial assembly processes, as previously presented in Figure 4.5. Recommendations are provided for designing AI systems customized to meet the specific requirements of individual use cases and address the design of an AI system as a collective contribution of multiple elements beyond just the AI model itself. These elements encompass but are not limited to **i)** sensor placement, sensor types, and sensor mobility [15, 200, 201, 199, 202, 191, 203, 204, 205, 206, 207, 208, 209, 186, 210, 211, 212, 155, 213, 214, 215, 216], **ii)** sampling rate, duration, frequency of actions [217, 218, 219, 185, 160, 220, 221, 222, 223], **iii)** preprocessing techniques, window size, models to use [224, 225, 137, 217, 226, 227, 208, 211, 228, 223, 119, 229, 230, 231, 129, 117, 232], and **iv)** interaction and feedback mechanisms [233, 234, 207, 235, 236, 237].

Category	Level / *Explanation*	Atomic	Micro	Meso	Macro	Mega
Sensor Placement	*Positioning of sensors within an environment for data collection*	stationary	wearable	Mix (stationary + wearable)	stationary	stationary
Sensors Types	*Types of sensors used for data collection*	**Part picking sensors**, RFID, weight sensors, proximity pick-by-light systems, EMG(wearble)	**IMU**, Force/Torque Sensor Wearable sensors, ACC, GYR, Visual sensors	Visual sensors, Mocap System	Visual Sensors, sensor fusion	Multi-modal sensors, GSI
System Mobility	*Flexibility of the system's physical mobility*	Low	High	Moderate / Low	Low	Low
Sampling Rate	*Rate at which sensor data is collected*	High (100Hz - 1000Hz), (or MHz for RFID)	Medium Moderate (50Hz - 200Hz)	Low (30Hz - 60Hz), 30 fps	Low (10Hz - 30Hz), 30 fps	Variable (10Hz - 30Hz)
Duration	*Time span of the experiments or observations*	Short tasks, Short (Seconds)	Short to Medium tasks, Moderate (Minutes)	Medium tasks, Moderate to Long (Minutes-Hours)	Medium to Long tasks (Hours)	Long tasks (Hours)
Frequency of Actions	*Rate at which activities occur*	Frequent	Frequent	Medium	Low	Variable
Preprocessing Techniques	*Methods applied to raw sensor data, Signal filtering, Basic feature extraction, Low-pass filtering*	Signal filtering, Basic feature extraction, Min, Max, Avg,Std Peak detection, PCA	Signal segmentation and filtering(Butterworth etc.), Feature engineering, Domain-specific end-to-end features.	Data augmentation techniques such as rotation, flipping, or cropping. Image filtering and geometric normalization	Edge detection image denoising, contrast enhancement,	Data Fusion, ML, Fusion of multi-modal data, Global context analysis
Window Size	*Size of the temporal window for data processing*	No windows or Short windows	Short / Variable size windows / sliding window approach	Longer / fixed-size windows / sliding window approach	Sequence-based windows	Long and adaptive windows
Models to Use	*Types of models employed for activity recognition*	Basic classifiers (SVM, Decision Trees), Clustering, Binary classification	Machine Learning, e.g., Random forests, CNNs, Lstms	Computer vision models / Deep Architectures.	Attention models, Rule-based Systems, CNNs, RNNs	Advanced AI Models, Anomaly detection algorithms, Ensemble models,Rule-based Systems, Reinforcement Learning
Feedback mechanisms	*Information or signals about the output or performance of a system.*	Immediate On-screen Alerts, Led indicators	Real-time (AR) displays, wearable devices with haptic feedback, interactive digital work instructions	Process Optimization Alerts Workstation Monitoring screens, e.g., manufacturing execution systems (MES)	Performance Analytics Dashboard Production Monitoring, dashboarding tools, KPI monitoring	Predictive Maintenance and Optimization System (ERP) systems, Reports

Figure 4.5: This figure illustrates key characteristics across atomic, micro, meso, macro, and mega-levels of assembly activity recognition systems. Each group of related elements is color-coded, and each line represents a different category, ensuring distinctions between aspects. The figure highlights variations that are important in the overall design of an AI system, such as sensor placement, types of sensors used, system mobility, sampling rate, duration of experiments, frequency of actions, preprocessing techniques, models employed for activity recognition, window size for data processing, and feedback mechanisms. Associated recommendations are provided for each category and level to serve as a starting point for the development of AI models under the "Models to Use" category, which is related to industrial assembly.

As task abstraction increases, so does the cognitive and spatial complexity of the activity, requiring system configurations that adapt accordingly. These categories are foundational in the development of AI-based HAR systems in industrial environments as repeatedly observed in the existing literature. Each one directly influences and controls how the systems perceive human activities, interpret human behavior, and respond to humans during the production flow. Sensor placement, types, and mobility (e.g., on the wrist, on tools, at workstations, or in the facility), determine what kind of data can be captured and how accurately, especially in dynamic environments. The sampling rate, duration, and action frequency define the level of detail the system captures, and the period that each activity is observed, in order to detect and interpret different types of tasks. Preprocessing techniques (e.g., filtering, segmentation, multimodal fusion), and windowing convert noisy, high-frequency, and heterogeneous sensor streams into structured, informative input data that models can reliably use. Model architectures (e.g., lightweight classifiers, temporal neural networks, ensembles), determine how well the system learns patterns over time, recognizes different activities, and performs in terms of recognition accuracy and generalization. Feedback mechanisms (e.g., real-time vibration, visual summaries, production dashboards), are essential for providing context-aware responses to the operators.

At the atomic level, where activities involve individualized tool interactions and component manipulations, sensors are typically mounted on the parts, components, or body, ensuring the capture of object movements and interactions. These identification sensors (ID) encompass a diverse range, including RFID, weight sensors, proximity-aware, pick-by-light systems, or EMG, allowing for multifaceted data collection. Preprocessing techniques applied to raw sensor data involve signal filtering, feature extraction, peak detection, and PCA approaches, which, in the scope of this work, are selected for refining the data for subsequent analysis. Machine learning models, such as basic signal processing and classifiers like SVMs and decision trees, are employed to recognize basic manipulation tasks characteristic of this level, sometimes without the need for temporal windows.

Transitioning to the micro-level, where activities entail more specific tasks involving interactions with parts from the previous atomic-level, a similar sensor setup is maintained with an increased focus on hand manipulations and sequences of atomic-level tasks. Preprocessing techniques, if applied, become more refined, incorporating signal segmentation and feature engineering within time-series data to capture sequential actions. Machine learning models evolve to include CNNs and RNNs capable of analyzing more complex interactions, while short or adaptive windows facilitate the analysis of sequential actions and interactions. Despite these advancements, the focus remains usually on single users engaging in short-duration tasks with high-frequency actions.

As the analysis progresses to the meso-level, activities involve performing specific tasks using larger components, which typically occur at the main assembly station or sub-assembly stations. Sensor placement may extend to encompassing the assembly workstation, capturing interactions between larger components or modules. In order to facilitate comprehensive data capture, vision sensors, depth cameras, and Mocap systems are introduced. In addition, preprocessing techniques become more sophisticated and memory extensive, incorporating feature extraction, windowing, augmentation techniques, image filtering, and object recognition. Computer vision and deep learning models, such as CNNs and LSTMs, are employed to recognize complex assembly patterns with longer or adaptive windows enabling the analysis of entire assembly sequences and sub-tasks.

At the macro-level, where activities encompass complete assembly processes and integration tasks, the sensor setup may cover the entire assembly workstation or multiple adjacent workstations, combining sensor systems [216] to capture interactions across multiple sub-assemblies or components. This level requires advanced AI models, ensemble models, and rule-based systems to analyze complex assembly workflows effectively. Long and adaptive windows, depending on the type of employed data, facilitate the analysis of entire assembly processes and integration tasks, while the subject number may extend to multiple users or groups engaged in longer assembly processes and system integration.

Finally, at the mega-level, activities involve comprehensive manufacturing processes and the optimization of production workflows across the entire factory floor or assembly facility. Sensor placement becomes more varied, capturing interactions across the entire facility, while preprocessing techniques involve global context analysis, advanced machine learning, and data fusion techniques. Advanced AI models, deep learning architectures, and custom ensemble models are employed to analyze complex manufacturing processes and optimize production workflows. Variable window sizes and sampling rates are tailored to capture diverse manufacturing activities and system performance metrics across the facility with infrastructure supporting real-time data processing and the optimization of manufacturing operations.

4.5 Discussion on Activity Abstraction

The employment of the multiple-level activity abstraction scheme offers several compelling benefits. Firstly, it facilitates a better understanding of complex systems by breaking them into smaller-level activities. This breakdown allows researchers and industry professionals to gain insights into the individual components that constitute the larger system, leading to a more detailed understanding of workflows, training programs, and automation systems. Consequently, it becomes easier to identify areas for improvement, troubleshoot problems, and optimize overall system performance. Apart from that, recognizing such complex activities is also important for daily living activities because it can enable tracking digital well-being, providing context-aware user experiences and notifications, and allowing better content recommendations [128].

Based on the study of assembly processes, certain activities may appear similar to atomic tasks but differ in nature and purpose. For instance, while activities like walking between workstations or reading assembly instructions are essential, they are either locomotion activities or cognitive tasks rather than physical manipulations of components. Although locomotion activities are inherently part of the overall activity framework in assembly tasks, individuals typically maintain a static posture while executing them, focusing on manipulation and precise task execution over physical movement. On the other hand, operating machinery or inspecting finished products involves complex activities beyond the scope of atomic actions. These tasks encompass a range of actions, such as controlling machinery or evaluating product quality, that require cognitive judgment and coordination. While they contribute to the overall process, they are distinct from the discrete singular actions associated with atomic tasks that would be under the latter stage in the presented hierarchy.

Within the domain of human activity recognition in industrial assembly tasks, the ability to recognize micro-activities, such as screwing, without the requirement to detect separately every underlying sub-action emerges as an interesting aspect. This is due to the goal-oriented nature and foundational role that micro-activities play in the hierarchical structure of assembly processes. Even though a micro-activity can be divided into smaller steps, recognizing the indicative pattern of the micro-activity reduces the need to understand separately the independent atomic actions except when tasks specifically require identifying manipulations at the atomic-level. The rationale for this is that detecting all the sub-actions would require more sensors and computing capacity, which might not essentially provide more meaningful information about the task at hand. Instead, the focus is on detecting the overall activity and its characteristics (such as duration, frequency, etc.), which can be captured using a smaller number of sensors and analyzed more efficiently while indicative of the underlying atomic-level components.

Regarding the duration of activities, each level exhibits a distinct timescale in a particular scenario. Micro-activities, which are characterized by repetitive actions, typically have shorter duration. As activities progress to higher levels, incorporating multiple previous activity levels, the duration increases proportionally, reflecting the cumulative complexity and scale of tasks involved. For instance, a meso-level activity comprising 10 micro-activities has a longer duration, which is calculated by aggregating the duration of each micro-level and atomic-level activity. This temporal progression also highlights the hierarchical nature of assembly tasks and underscores the incremental development of products across different levels. Nevertheless, there may be some slight overlap in the activity duration due to the variety of activities in the industry. Therefore, while the timescale can indicate each level, a scenario analysis is required for interpretation.

Additionally, atomic, micro, meso, macro, and mega-level activity abstractions contribute to a common understanding that bridges the gap among people of diverse roles and expertise in the assembly process, facilitating effective communication and knowledge transfer to achieve greater productivity, execution accuracy, and scalability. Each successive abstraction level builds upon actions that exhibit variations, i) such as the number of individuals involved, ii) the execution station utilized, iii) the overarching goals they serve, iv) and the complexity of the tasks they encompass. For instance, at the atomic or micro-levels, activities may involve individual actions such as grasping a component or tightening a screw. These actions are relatively simple and executed by individual workers at specific workstations. As we move to higher levels such as the meso or macro and mega-levels, micro-activities become more complex, involving coordination among multiple workers or machines across different stations to achieve larger production goals. Furthermore, the end goals of micro-activities differ across levels. At the micro-level, the focus may be on completing discrete assembly tasks, whereas at the macro and mega-levels, micro-activities contribute to broader objectives such as optimizing production efficiency or meeting customer demand.

The hierarchy of assembly stages can also benefit both the arrangement of sensors and data analysis. By grouping the assembly process into distinct stages, it is possible to deploy wearable sensors, such as IMUs or visual sensors, to gather relevant information. These stages can provide real-time data awareness of worker movements and interactions with components or data collection tools at each level. In addition, researchers can identify and consider more factors related to specific cases. For instance, experiments at the atomic level may occur in controlled lab environments, ensuring precise data collection. However, as assembly complexity evolves at the macro and mega-levels, experiments transition to real-world factory settings, introducing factors that may include limited lighting, occlusion effects, noise, and other environmental variables along with the system's obtrusiveness. The complexity of tasks also impacts the frequency of actions, with activities occurring more frequently at lower assembly levels, such as the atomic and micro-levels, compared to the macro and mega-levels, where actions occur less frequently. Additionally, privacy concerns rise with increasing complexity reflecting the diverse and extensive nature of data collection across the factory floor while underscoring ethical and practical considerations in system development. Addressing potential data privacy implications led to the prioritization of less invasive and privacy-friendly sensors for the applicable levels, considering the individual's privacy rights.

These differences underline the need for tailored development measures and regulations compliance, as assembly activities progress from lower to higher levels of complexity, which plays a significant role in user acceptance, system deployment, and overall effectiveness in real-world assembly environments. Moreover, considerations surrounding data processing and power consumption grow across assembly levels. While at the atomic and micro-levels, data processing and sensor cost are low, progressing to the macro and mega-levels, there is an increasing requirement for computational and finance resources, especially if the generated data will be stored.

Overall, it is reasonable that while the aim is to try to create a versatile and robust framework, there may be limitations to cover each case in all different domains of HAR. Consequently, this method is presented for industrial assembly processes, but it may need adjustments to describe activities in other domains. Additionally, some specific use cases may fall within the boundaries of the described levels. In such cases, the suggestion is to draw recommendations from both levels and address with a hybrid approach the complexities of the case. Although extensive, it is acknowledged that these recommendations cannot serve as the sole method or rigid rules due to the individual requirements, challenges, and outcomes inherent to each unique case. However, they provide a valuable starting point and a preliminary framework for navigating the complexities of designing and implementing AI technologies in industrial settings.

4.6 Findings on Abstraction – Summary

Based on the findings presented in this chapter, the research questions can be addressed as follows:

RQ1: How can structured levels of activity abstraction in manufacturing workflows be defined to enhance cognitive assistance in assembly processes?

This work aims to enhance the understanding and support of human activities within industrial contexts. To achieve this, a comprehensive taxonomy was developed, that spans from the atomic to the mega-levels , for categorizing human activities in assembly processes for discrete product manufacturing which is inspired by the sequential order of operations in real industrial assembly chains.

(i) Atomic Level: Smallest, self-contained manual operations (e.g., grasping, positioning).

(ii) Micro Level: Recognizable single activities composed of atomic steps (e.g., screwing, welding).

(iii) Meso Level: Intermediate tasks that involve assembling sub-components (e.g., preparing ATM sub-assemblies).

(iv) Macro Level: Comprehensive assembly processes that integrate modules into final products.

(v) Mega Level: Encompasses broader manufacturing processes, including post-assembly tasks like quality control and packaging.

This taxonomy offers a hierarchical structure that facilitates better decision-making for the user, process optimization, and system design that addresses the binarization of «simple» and «complex» activities and complements existing non-binary approaches. It is derived from empirical observations of real-world industrial assembly workflows. that reflect the natural decomposition of manual tasks into progressively larger and more complex units. Additionally, separating tasks into manageable steps reduces confusion, optimizes processes by applying specific task AI-driven automation, and enhances the training of workers by guiding them through simple to complex tasks. Moreover, it enables a more granular analysis, troubleshooting problems, and identifies inefficiencies, that lead to improved worker performance through lower cognitive load. Lastly, differences in the approaches are demonstrated with a comparison between the proposed taxonomy and the existing categorization schemes through an example of a real assembly scenario (e.g. 4.3).

RQ2: How can various sensor modalities (wearable and non-wearable) be utilized to recognize and categorize manual activities in industrial settings, to support AI system design for enhanced human-machine collaboration?

Building upon this taxonomy, specific recommendations are provided for designing AI systems tailored for activity recognition in industrial assembly tasks as presented in Table 4.5. Although each use case may be unique with specific requirements and goals, the fundamental human activities and physical manipulations of the tools and components across assemblies typically remain similar. These recommendations examine, among others, sensor placement, preprocessing techniques, and model selection across different levels of activity abstraction. The second part of this chapter outlines how different sensor modalities (wearable and non-wearable) can be effectively deployed across different abstraction levels to optimize AI-driven activity recognition to provide cognitive assistance in various industrial tasks.

These guidelines link sensor modalities (wearable and non-wearable) with appropriate models and specific characteristics of the tasks. They also relate system components such as sensor setup, preprocessing steps, and feedback methods, to the complexity levels of the activities. In this way the taxonomy provides a practical framework to guide real-world AI system design, moving beyond conceptual models.

5 Perception and Reasoning in Cognitive Augmentation

5.1 Introduction

After the presentation of the taxonomy in the previous chapter and the guidance table for AI system design, this chapter focuses on the topics of the detection and counting of repetitive key activities in industrial assembly. The emphasis is on investigating the sensor-based approaches and models that were introduced in the previous chapter to extract information and classify activities within industrial assembly workflows. Furthermore, state-of-the-art methods are developed to handle variations in task execution and duration, address ambiguous annotations, and quantify actions within time series data.

Two main challenges are addressed in the industrial processes through the development of cognitive augmentation systems:

(i) recognizing and classifying key activities—particularly at the micro-level, due to their critical role in providing insights into the quality and efficiency of tasks, and

(ii) counting the number of repetitions of such activities to support workflow tracking and quality assurance.

The recognition of micro activities requires models that are able to learn from short, weak, and noisy patterns often in imbalanced data. For this purpose, the classification task used supervised learning and fixed-size time windows that allowed the models to detect short, repetitive tasks based on motion patterns captured via IMU signals. In contrast, the counting task which is important for workflow tracking and quality assurance is addressed with variable-length sequences using weak labels that specify only the total number of activity repetitions within a segment. This design avoids the need for precise annotations and represents more realistically real settings, where activity boundaries are ambiguous and annotation is resource-intensive.

Motivation

An automated understanding of work-steps in industrial assembly work is important for assistive guidance technologies in employee-machine collaboration. While process tracking is important and offers a broad view of task progress, the errors typically occur at the micro activity level. Each micro-activity (e.g., tightening a screw, inserting a connector) is a small action that has a big impact on the final product result. A single missed screw can cause downstream issues, from functional failure to safety recalls. Forgotten or incorrectly installed parts, and missing or non-tightened screws during assembly, which are expensive and time-consuming to repair, are some common mistakes that are addressed with this approach.

As discussed in the previous chapter, atomic activities refer to fundamental, indivisible motor actions, such as "grasping" or "turning". In contrast, micro activities are compositions of atomic actions that achieve a specific goal and task. Thus, micro activities occur at the first level at which industrial tasks become verifiable (e.g., a screw is either fastened or not). Going further, the meso level may include the coordination of several micro activities, such as the assembly of a panel that could consist of X number of micro activities.

However, the detection of micro activities is challenging due to their short duration and their high similarity across tasks (e.g., hand-screwing vs. tool-screwing). Often their weak or small patterns in the sensor data compared to other activities, lead to confusion in their recognition by conventional HAR detection methods which are designed for longer and more distinct activities. The challenge is further supported by their limited availability in public datasets which makes it more difficult to progress with the research in industrial tasks.

In this work the focus is on a seamless embedding of non-impeding Inertial Measurement Units (IMUs), worn on the body or integrated into tools and devices, allowing for unobstructed monitoring of tools' usage patterns. Therefore, understanding the activities that occurred and thus recognizing the assembly work steps. The activity detection at the micro level will provide real-time feedback by confirming that a screw has been applied, alerting if a step was skipped or incomplete, and informing about the completed actions.

The hypothesis is that a system capable of high-level detection for micro activities in an assembly line, utilizing IMUs and neural networks, will i) reduce the error rate in the final product, ii) assist the workers in real-time scenarios by performing quality control iii) understand the stages of the assembly workflow. The results of this study are evidenced by empirical observations of work-step executions by (a) hand screwing, (b) screwdriver screwing, (c) machine screwing, (d) wrench screwing, with the size of the null class being disproportionally dominant in the data set. Deep Learning models including Long Short-term Memory (LSTM) and Convolutional Neural Network (CNN) architectures are

Figure 5.1: The image illustrates a worker performing a fastening operation using a hand tool, highlighting the importance of human intervention in complex assembly processes.

evaluated while presenting the challenges encountered during the conducted research and experiments.

The aim is to identify micro activities of employees during the assembly, for assistance purposes in their daily complex tasks. This can be achieved by using mobile wearable devices and hand-operated tools to correct mistakes while they occur and avoid errors during later stages of assembly. Instead of providing the worker with full process level monitoring, micro activity tracking offers focused support by targeting the actions that occur in real-time. This approach is intended to reduce cognitive load and help workers stay concentrated on the task at hand.

Human Error in Industrial Assembly Application

Errors in industrial places are prone to happen, either because of fatigue of the workers or due to the repetitive nature of the job in an assembly manufacturing line. The indirect detection of errors such as a missing screw or a non-tightened screw is the objective of this section, during the assembly of an automated teller machine (ATM). This machine requires a large number of human activities and many components to be integrated during its assembly. The later a defect or an error is noticed on the assembly line, the more costly it will be to repair it, especially if the "error" continues to another assembly station. On the contrary, if the error is detected while it happens, the operator can correct it instantly and prevent the defective product from,

(i) moving to the next stage on the assembly line,

(ii) reaching the market as a device of inadequate quality,

(iii) needing additional rework time by the worker.

To achieve that, this experiment focuses on the detection of micro activities of workers in their DWL projects and tasks, intending to develop a smart and adaptive system that will be able to detect those activities and provide confirmation for completion of the task or notify for bottlenecks along the process.

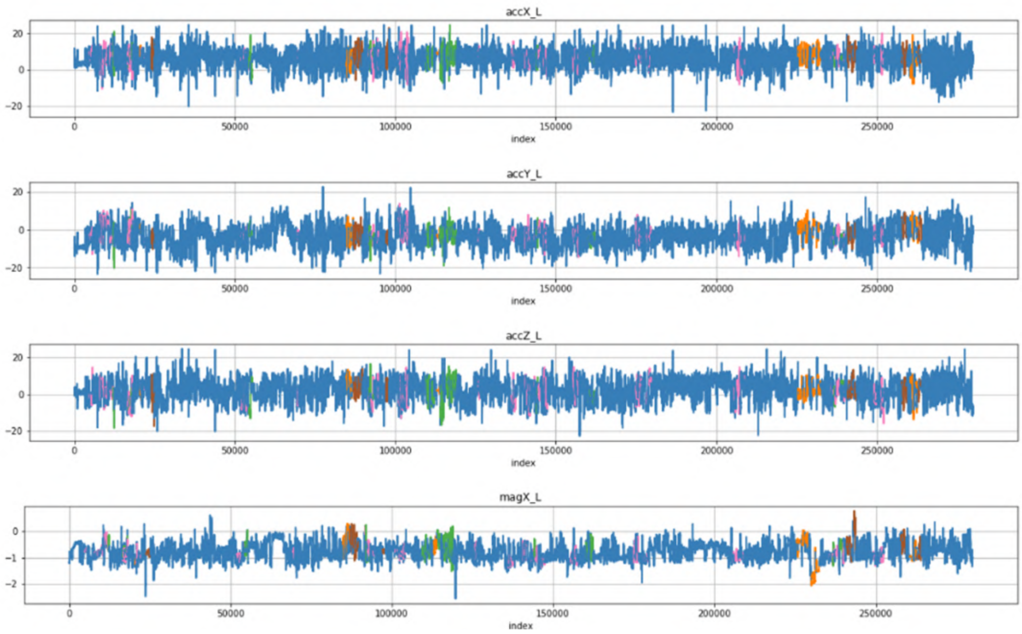


Figure 5.2: The figure presents exemplary time-series data from an industrial assembly process, showing accelerometer (accX-L, accY-L, accZ-L) and magnetometer (magX-L) signals. The highlighted colored segments indicate detected activities of interest within the null class (blue) in the worker's motion patterns. Each color represents a different activity identified in the IMU signals.

The study is based on data collected exclusively from IMU sensors, as visualized in Figure 5.2, that are worn by a small group of employees, to detect the number of screws that are used, while assembling an ATM utilizing different tools. Every work-step has several screwing actions that are predetermined and necessary to complete. The system will detect the occurring screwing activities, compare them with the required activities, and provide the appropriate message to the operator. Investigating the level of micro activity detection that the neural networks in combination with the IMUs can achieve, to assist workers in their DWL, is important for the novices, along with the experienced ones.

During this research, various challenges were encountered regarding the collected data, the implemented deep learning models, and the categorization and quantification of the activities. The following sections present the system architecture, the approaches to dealing with the occurred difficulties, and the results for the micro activity recognition and action counting in industrial work scenarios, contributing to the goal of workflow recognition and human cognitive augmentation.

5.2 Activity Recognition approach

Workflow recognition and acknowledgment of human's cognitive state can support the level of consciousness and awareness for employees in the industry. In this regard, investigating the collaborative behavior of humans and machines with cognitive abilities, by developing machines that can actively adapt to the user's behavior and needs, can prove beneficial for many reasons. Workplace safety, simplification of industrial processes, immediate response to changing or challenging conditions, improved employee efficiency, reduced downtime, and proper use of equipment, to name a few.

5.2.1 Experimental Design

Case Study Overview

In this experiment, the purpose is to perform indirect work-step recognition in the DWL of workers, by identifying micro activities and provide confirmation of each complete work-step or provide notifications that help them focus on possible errors. This study is a part of a larger use case which is described in Figure 5.3. In this image, one can see how the system is organized, with the sensor modalities, and the collected data used as input to the algorithms in order to understand the activities before the reasoning takes place. The decision of the system will be communicated to the user through a mobile personal actuator, which is a smartwatch, or through a wall-mounted monitor, in cases where the assembly process is static.

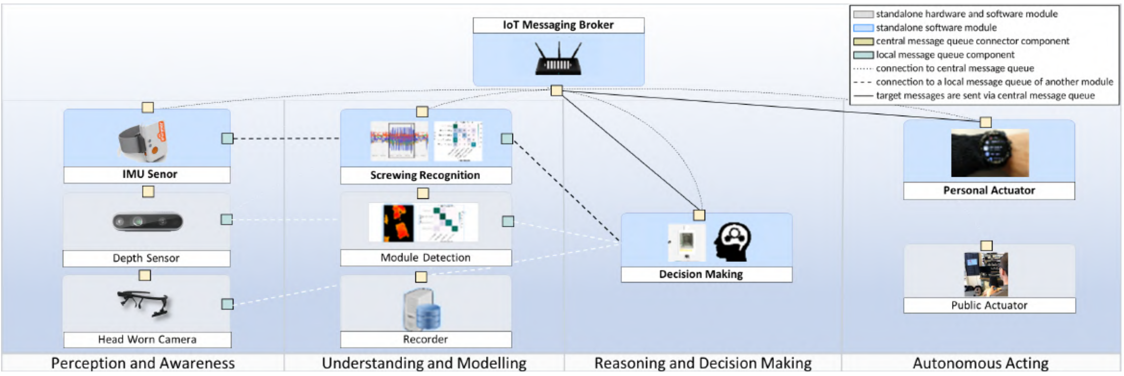


Figure 5.3: This figure visualizes the system architecture of the larger network that includes the micro-activity (light-blue boxes) recognition part as one of its key components. The data collected by several sensors will be used as input to the algorithms and result in a complete workflow detection application. A smartwatch was chosen as a feedback modality.

The framework follows a multi-sensor approach where three IMU sensors are worn on the body of the operator and one is placed on one of the tools that is used. The two sensors in the hands provide data on hand movements and are placed in both hands in order to have a more complete view of the data, regarding the micro activity screwing recognition. In addition, it supports the prevention of creating a biased model regarding the use of the left or right hand. The tool sensor strengthens the detection of the electrical screwdriver while the body sensor gives information about the position of the worker and if he is moving, considering that during motions such as walking the probability of performing some screwing activity is reduced. All of the sensors have direct wireless communication with a server that collects data, except for the eye-tracker which is cable-connected. The data are transferred to the server where the "understanding and modeling" starts and the "reasoning and decision making" takes place.

In the presented setting, the data collection took place in the production line of an industrial manufacturing company, during the working hours of the workers. The data acquisition was followed by the preprocessing, later the learning and the classification of the activities occurred similar to the approach in [96, 182] and finally the notification through the feedback modality – see Figure 5.3. These context-aware activities with a small range of motion, – referred to as micro activities in the presented taxonomy– are more likely to be detected by the deep learning approaches than by shallow approaches as stated in [63].

Wearable Sensors Installation

Several modalities can be utilized in order to perform human activity recognition such as image-capturing sensors, inertial measurement units, depth sensors, or smartphones. In this case, the interest is in micro activities such as small movements of hands. Subsequently, the choice to obtain the data from IMU sensors was unidirectional, since it was required the device to be attached to the wrist of the user, similar to the approach in [238]. This arises from the fact that this position of the sensor is generally accepted by employees in manufacturing, as less disturbing and intrusive and provides descriptive data for complex small-scale activities of hands. Stationary sensors such as depth sensors (e.g. kinect) are not suitable for similar assembly tasks as the worker often has to perform activities inside the machine where the sensor has limited or no visibility. The IMU sensors contain an accelerometer, a gyroscope, and a magnetometer, thus keeping costs low and ensuring minimal interference. In addition to the small size of data that they produce, they are already contained in many devices of daily use such as smartphones, smartwatches, cameras, and others.

They can work as standalone devices in terms of mobility, since the sensors that contain them can function without cables, which is important when the worker moves around in his working space to collect parts required for the assembly. Their ability to establish a connection to the main system via Bluetooth supported the desired freedom of movement which is important for the user. Yet another crucial issue is the user's privacy and protection of his personal data. In this case, IMUs are considered to be more compliant with the General Data Protection Regulation (GDPR), less invasive, and less unobtrusive compared to image-capturing sensors, in the DWL of the operator. Further, IMUs are suitable and able to maintain continuous data recordings while on the move.

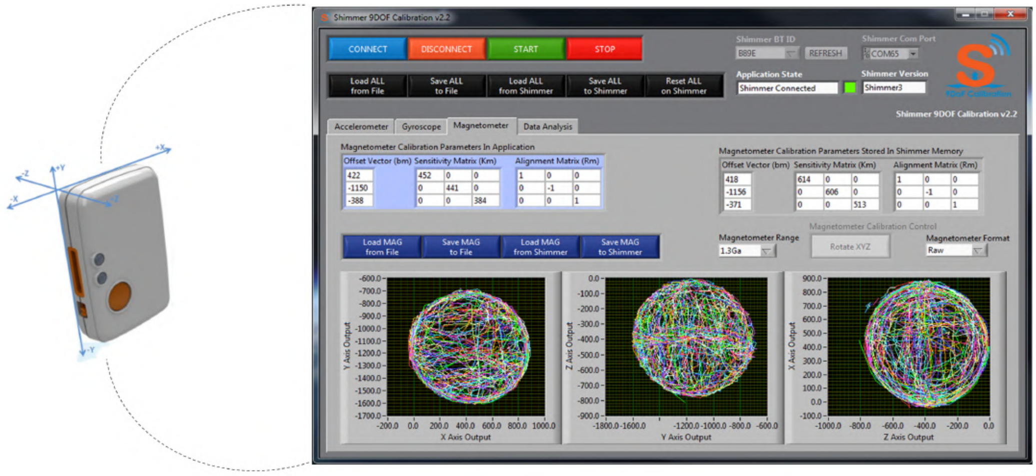


Figure 5.4: The figure shows the Shimmer sensor and next to it the 9DOF Calibration Interface. The interface displays the magnetometer calibration parameters as an example, based on the user manual that is provided by the official distributor [239].

The selected device in Figure 5.4 is from the company Shimmer which produces wireless wearables sensors. The sensor is a low-cost platform and the model was the Shimmer3 GSR+ Unit which contains besides the sensors, a MSP430 micro-controller (24MHz, MSP430 CPU), Bluetooth Radio – RN-42, Integrated 8GB micro SD card, 450mAh rechargeable Li-ion battery. The figure also shows an example of the user interface in order to calibrate the accelerometer, gyroscope, and magnetometer of the device for the capturing of the activities. This calibration has a big impact on the scale, quality, and reliability of the collected data as well as the results of the models.

5.2.2 Data Processing Pipeline

The data for the study were recorded simultaneously, for each class, in the workplace of the collaborating company Figure 2.3a, during the working hours of their employees. In the recording sessions that took place over a period of one year, novice and experienced workers executed the normal workflow process for the ATM assembly. Each worker was equipped with four calibrated IMUs, in order to produce comparable data on the same scale. One on the wrist of each hand, one on the right ankle, and the fourth was installed on one of the tools that was used. The raw data from those sensors were stored in CSV (Comma-separated values), files and provided the input for the machine learning models.

Moreover, a head-worn camera system was utilized to record image data, which was used exclusively as an annotation mechanism for the IMU data and not for the classification part. The received data were sent and monitored on a notebook connected via Bluetooth to the IMUs and via cables to the eye-tracker. The activities were recorded with a sampling rate of 60Hz and their annotation was performed by a single researcher..

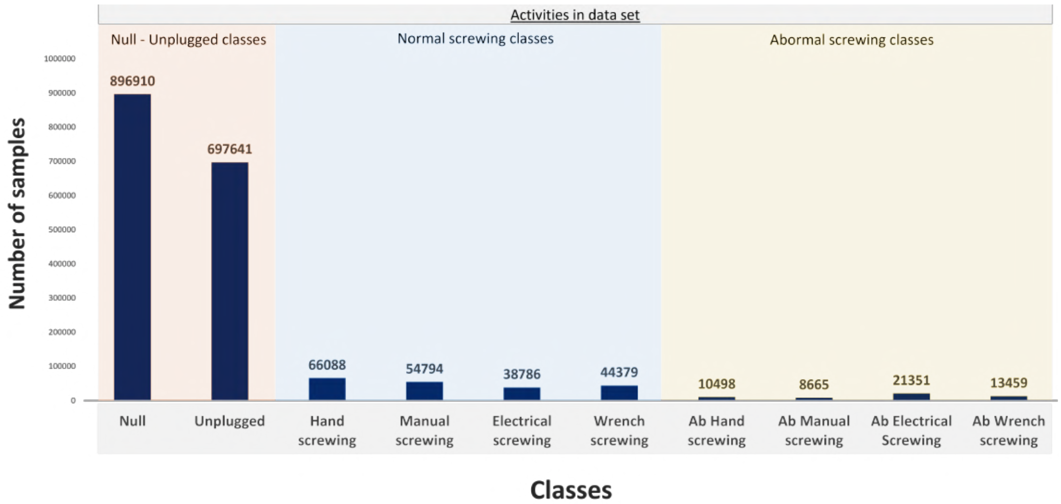


Figure 5.5: The figure shows the sum of the samples per class used as input to the algorithms for the complete data set of this study. The imbalance is evident between the classes with null and unplugged classes prevailing. The classes with normal activities contain all the properly executed activities and the abnormal section in the graph provides an overview of the activities that were initially problematic for the recognition.

In Figure 5.5 an overview of the data that were recorded and used as input for the models' training, is given. For each class, the number of samples collected from each session can be seen. Twenty recording sessions occurred with 3 participants, where sixteen of them followed the normal assembly workflow and the rest recorded one of the four screwing activities at a time. That resulted in a total of 10 hours of recording sessions collected on different days and during the employees' working hours according to their individual work schedules during their shifts at the production site.

Each session consisted of the four activities of interest which were i) hand screwing, ii) tool screwing, iii) manual screwing, iv) wrench screwing, and the null class. The activities of interest covered almost 10.6% of the whole data set, while the abnormal activities took about 6% of the data set. An abnormal class was created for each activity that included all the signals that were not clear, less descriptive, or were deviating from the signals that were realized as the "proper" ones. Figures 5.6, 5.7, 5.8, 5.9, present the variations of the patterns that were included in each class and some of them had to be moved to the corresponding abnormal class as a result of the inconsistent style of a participant performing an activity of interest. Given that each recording was performed in real working conditions, it was normal to experience unexpected distractions that led to improper execution during the assembly.

Figure 5.6: The figure shows the hand screwing pattern: The annotated normal pattern is on the left, and the abnormal or divergent pattern is on the right.

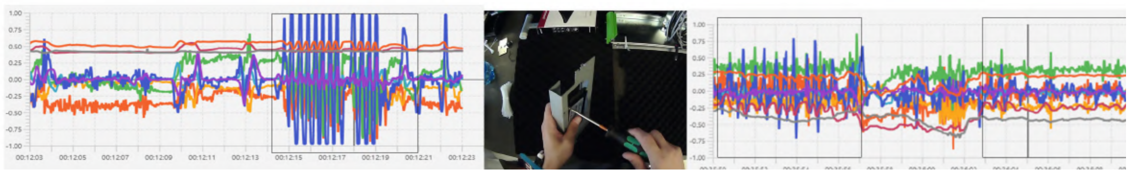


Figure 5.7: The figure shows the manual screwing pattern: The annotated normal pattern is on the left, and the abnormal or divergent pattern is on the right.

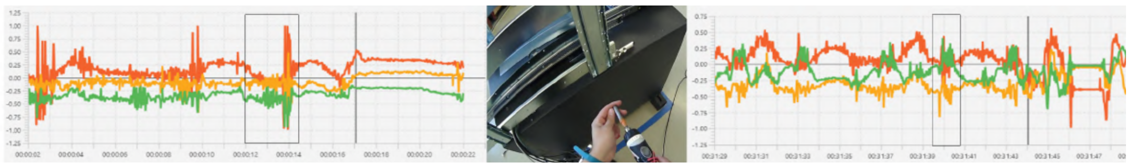


Figure 5.8: The figure shows the electrical screwing pattern: The annotated normal pattern is on the left, and the abnormal or divergent pattern is on the right.

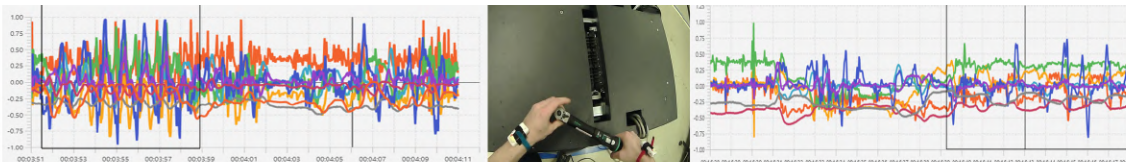


Figure 5.9: The figure shows the wrench screwing pattern: The annotated normal pattern is on the left, and the abnormal or divergent pattern is on the right.

Unscrewing activities that were not used in the training occupy almost 3.6% and the rest of the data set, almost 80 %, was the null class, consisting of multiple not screwing activities of the worker related to the assembly of the ATM.

The null class had the biggest size since screwing activities occurred rarely during the assembly process. Activities that belong to the null class are for instance the cable installation and mounting for communications and electrical processes of the ATM, collection of screws from the shelves, and walking to pick up and mount parts of the product. Workers with different levels of expertise participated in the study and performed the activities with their individual style of screwing, with no predefined order of the activities. That was an additional difficulty for the models in the learning phase, however it contributed to the generalization of the models.

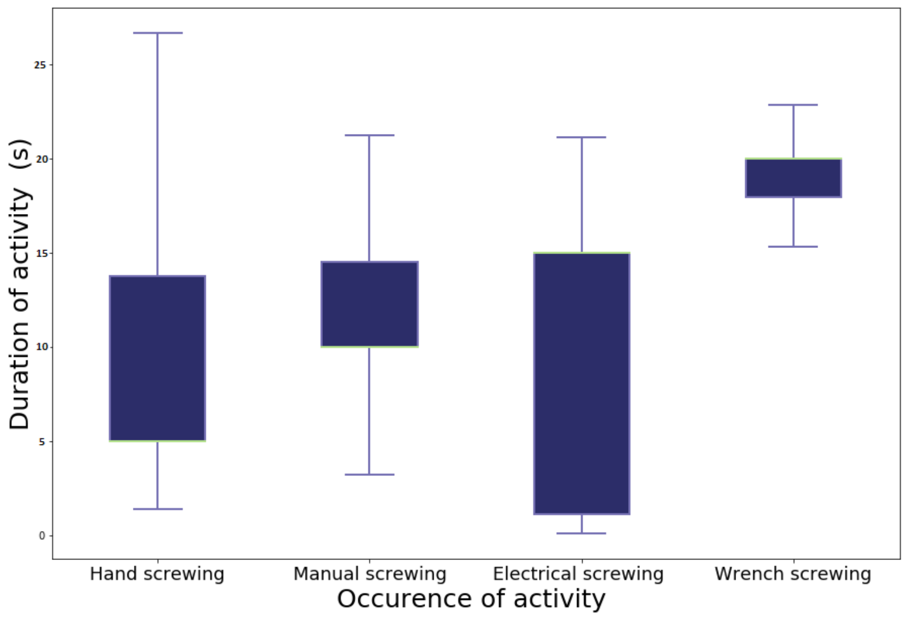


Figure 5.10: The figure shows the duration for each activity of interest is visualized. The minimum and maximum values given below include the outliers for the classes: hand(1): min: 0.272s - max: 23.494s, manual(2): min: 0.34s - max: 28.152s, el. screwdriver(3): min: 0.017s - max: 8.245s, wrenching(4): min: 0.408s - max: 24.021s.

An unplugged class was created from the null class data, which consisted of non-screwing activities that the worker was performing while wearing the IMU sensors but without wearing the eye-tracker. Thus, the data concerned with those activities were removed in order to reduce the size of the redundant activities of the null class and the imbalance with the rest of the classes. In Figure 5.10 the diagram and the minimum and maximum duration of activities can be seen, to have a better understanding of the diversity of data, in terms of duration for every class.

Data Information

The original raw data were then compiled into 2-D CSV files, that consisted of 39 columns for each sample. The columns were the timestamp, the class label, the accelerometer, the gyroscope, the magnetometer for XYZ for each one of the 4 sensors (2 wrists, ankle, electrical screwdriver), and one additional column for the electrical screwdriver that produced two values on/off as visualized in Figure 5.11. This approach (value, on/off) was integrated from the beginning of the study, but the models were tested for their performance initially without it. The data collected by this sensor produced a weak signal, as the tool was mainly static during this task, and the activity had a short duration.

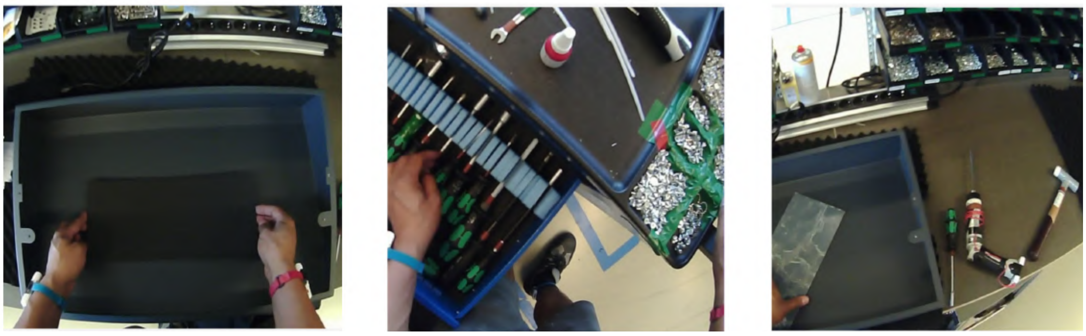


Figure 5.11: The figure presents the IMU sensor placement for data collection during the industrial assembly tasks. Starting from left to right, sensors were positioned on the left and right hands (wrist-worn), on the leg (ankle-worn), and on the tool (attached to the power screwdriver) to capture worker movements and tool activities from multiple sources during the task execution.

The machine learning system was based mainly on deep network architectures, therefore the data were not preprocessed further. However, normalization techniques were applied to obtain the same scale of [-1, 1] for all values and reshaping of the data in a 3-D array [samples, time steps, features] as required by the deep learning algorithms.

5.2.3 Algorithm Development

Machine learning models

Variations of deep learning models that are state of the art in HAR and fit this supervised classification problem, from simple RNN (Bi-dir LSTM) and CNN to more complex attention models and mechanisms [240] were implemented and tested for their effectiveness. Examples of the models used in this study are presented in Figure 5.12 where some of these models are custom-developed architectures inspired by existing designs, while others are established implementations based on prior literature. A residual bi-directional LSTM [64] with two stacked recurrent layers and n-input/2 units was used, where n-input was the input array's number of features and 0.1 the dropout of the output layer.

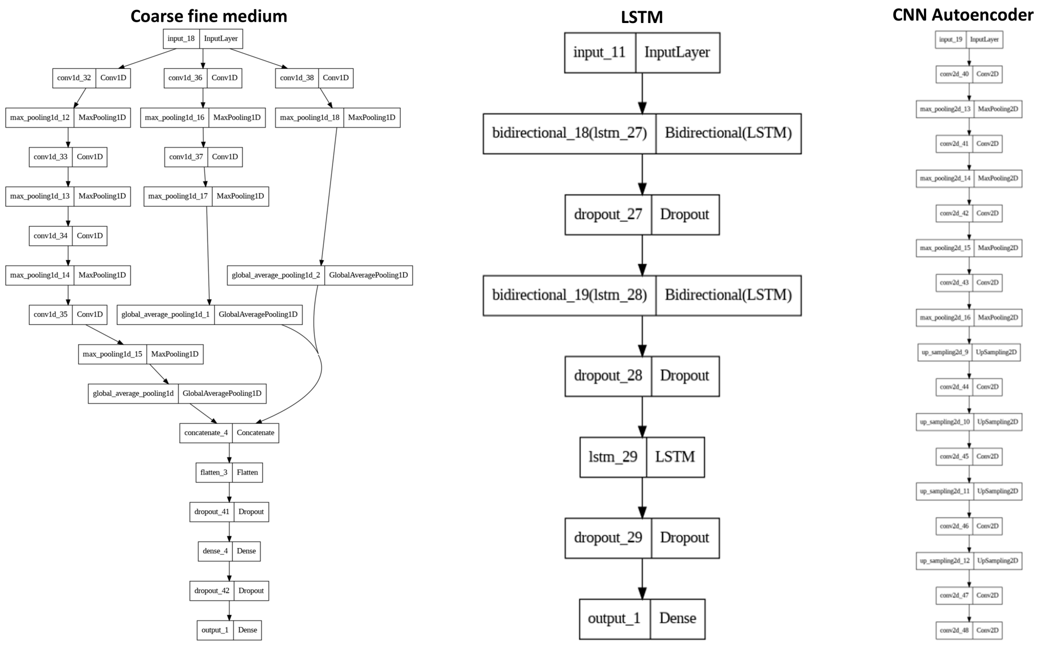


Figure 5.12: The figure shows an overview of various models that were utilized in the study. Among them, starting from left to right are representations of the coarse fine medium model, the LSTM model, and the CNN autoencoder. These models form essential components of the system architecture and include both custom-developed solutions and established designs based on prior literature [241, 64, 242].

For the CNN approach three models were implemented, one basic CNN model, one CNN auto encoder architecture, and the coarse fine medium model which presented the highest results for this use case. The basic CNN consisted of a pooling layer after each one of the four convolutional layers, followed by a fully connected layer and the dropout layer. The CNN autoencoder is based on the encoder-decoder structure with convolutional and pooling layers for dimensionality reduction and avoidance of overfitting. The selected number of layers for the models and their parameters were decided after testing for the most efficient outcome. The more complicated model that was used in this study, was the coarse fine model [241], in which "Three parallel CNNs are used for local feature extraction, and later they are fused in the classification task stage. The whole CNN scheme is based on a feature fusion of a fine-CNN, a medium-CNN, and a coarse-CNN" that is getting concatenated and flattened for the final classification –see Figure 5.13.

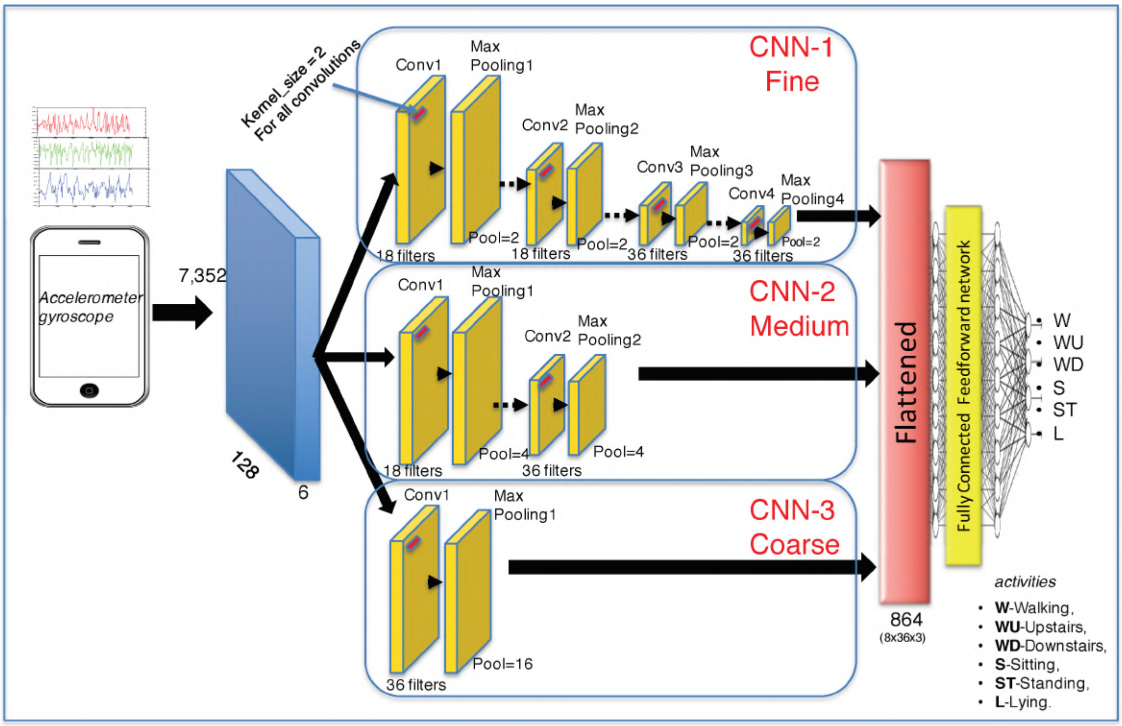


Figure 5.13: Source: Reproduced from [241]. The figure shows the coarse fine medium topology for the convolutional neural network that was implemented for the study and was developed by the authors in their work.

The training data sets that were used as input for the models refer to the total number of workers, thus the aim was a person-independent model. The testing data consisted of one data set of the participants, that the model has not seen before. The data were formatted to fit the shape required by the deep learning models for the supervised deep learning problem and were segmented by the sliding window method. Window sizes from 1s to 20s were explored to fit the multiple time lengths of the activities while shifting values of 0, 0.05, 0.25, and 0.5 were tested, as it is proposed by [243] where the correct window size for more complex activities is studied. During the learning, a number of parameters such as the batch size, the learning rate, iterations, the dropout, etc., were examined to determine the best values for each.

The commonly used regularized cross entropy [244] was chosen as the loss function. To penalize large weights during training, a l2 regularization term was applied to the loss function, where λ was its coefficient. To mitigate over-fitting issues, the dropout regularization [245] was used during training, which randomly dropped neuron units from the neural network. The CNN and LSTM architectures mentioned in the previous sections were constructed using TensorFlow [246] and Keras libraries [247]. During the process, the NumPy library was used for numerical computations, and the Pandas library for dataset handling in Python [248, 249]. The Adam was used as an optimizer for the models, the rectified linear (Relu) as the activation function and the learning rate was set to 0.0001 since those parameters were found to produce better results.

5.2.4 Use Case Optimizations

In a survey that was conducted in [60], common drawbacks for HAR are mentioned, such as the class imbalance problem, the inter-class similarity, or the intra-class variability that require in each case particular methods to handle it. Similar complications occurred in this case of micro activity recognition and consequently assembly step recognition. Some of the approaches to deal with existing issues include:

(i) increasing the data for the classes of interest by collecting data in a more controlled setting,

(ii) utilizing extra sensor modalities for higher and broader capturing of the activities,

(iii) implementing methods for data augmentation to enlarge the diversity of the training set,

(iv) applying weighting techniques for better optimization of the models.

From the chart in Figure 5.5 it is obvious that there is a class imbalance problem since all the activities of interest together represent only 10% of the data. However, the considerably large size of the null class compared to the other classes was not the only impediment. Most of the confusion occurred, because in the null class exist activities that have similar patterns to one of the classes of interest.

Oversampling Techniques

One of the main solutions for class imbalance problems is the effort to enlarge the under-sampled classes with extra either "artificial" or "real" data. In this case, the collection of extra real data was implemented, for each class that did not have an adequate amount of data. In this scenario, the workers could perform the same activities in the same way that they do in their actual work time, with the only difference being that these activities would not be interrupted, by activities that are not of interest. Four recording sessions were scheduled, to record one activity of interest per session, one for every screwing class.

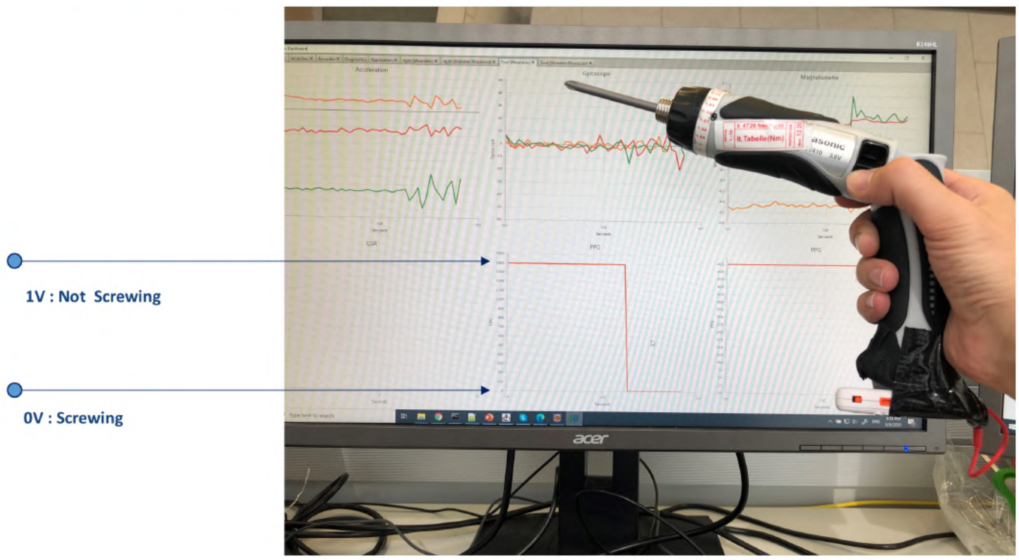


Figure 5.14: The figure presents the functionality of the sensor on the electrical screwdriver. The sensor detects screwing activity based on the button activation, where 0V indicates active screwing and 1V indicates inactivity.

The results after the integration of the additional data for the underrepresented classes of the data set, were partially efficient for most of the classes but not for all of them. The data from the sensor of the electrical screwdriver (valueP) that gives a value equal to zero when the machine is working and greater than 1 when it's not functioning proved to be appropriate –see Figure 5.14. The models were tested with and without this value to validate the level of detection for the specific activity.

Data Augmentation

Addressing the data imbalance in the dataset required the exploration of multiple methods that are recommended in the literature [250]. These methods are applied to the data to enrich the underrepresented classes but also to enhance the model's generalization and mitigate bias. In Figure 5.15, one part of the dataset is visualized under the transformation for rotation where the IMU data are rotated. In this case, when the sensors are placed with different orientations e.g., upside-down, the models are still robust enough to understand the correlation between the class and the inverted IMU signals of the readings. Furthermore, this can mitigate problems that occur because of the vibration or unexpected movements of the sensors.

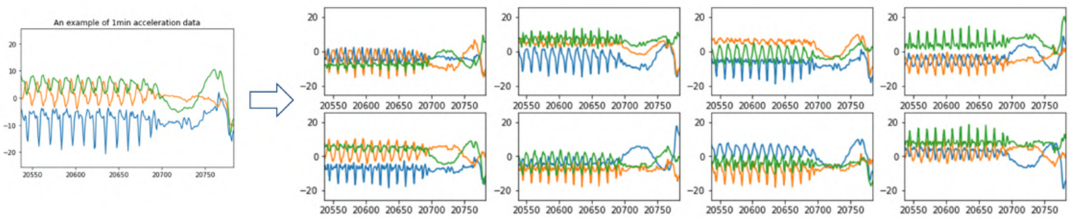


Figure 5.15: The figure presents the rotation as a data augmentation technique to introduce variability in IMU sensor data by simulating arbitrary changes in sensor orientation.

Additional transformations were applied, such as jittering, time-warping, etc to improve the model's and the dataset's capabilities. Jittering involves adding random noise to the original signal, to increase the model's robustness to small fluctuations while time-warping applies non-linear distortions to the time axis to account for different execution speeds of activities. Despite these efforts of artificially increasing the data, the model's performance did not significantly improve, and the real data collection that was presented in the previous section proved to work better for this use case.

Weighting Approaches

In time-series classification fixed-length windows are often used to extract information and patterns [251]. The most common approach is the sliding window where the window moves forward with a small overlap separating data across consecutive windows. The label for each window usually results from the majority of the labeled samples within it or by using as a label of the window, the label of the last time-step. This suggests that the correct window in terms of duration must be chosen for a specific activity. However, the issues arise if there are multiple activities to detect and present significant differences in their duration.

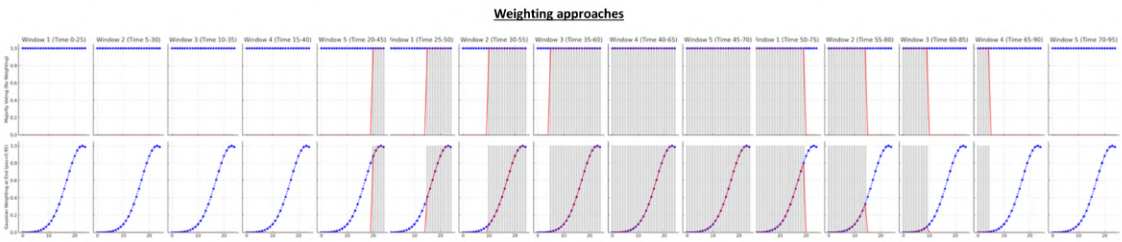


Figure 5.16: The figure presents a comparison of majority voting (top) and Gaussian weighting (bottom) across sliding windows. In the visualization, the blue dots represent the sample weights across each window, where uniform weights are assigned in majority voting. The red line shows the contribution of the samples to the window label. Gaussian weighting detects activity transitions earlier by assigning higher weights to the recent samples, while majority voting labels a window based on the majority of samples that exist in the window. This can delay the recognition until the activity dominates the window. Additionally, in majority voting, an activity is still labeled even after it has stopped, because older samples inside the window still contain it. This causes a delay in recognizing inactivity. Gaussian weighting adjusts faster by giving more importance to recent samples, ensuring the label updates as soon as the activity stops.

To improve activity detection, Gaussian weighting is applied inside the sliding window, as illustrated in Figure 5.16. A similar approach has been explored in financial time-series analysis, where Gaussian Weighting Reversion (GWR) has been applied to emphasize more recent data points for decision-making in portfolio selection [252] and inspired this implementation. This implementation prioritizes the most recent samples that enter the window, to detect the changes in activity earlier than with majority voting.

As the sliding window moves forward in time, Gaussian weighting enables the system to understand the new activities faster by focusing on the new samples that enter the window. The assignment of higher weights to later samples ensures that an activity is recognized as soon as it starts influencing the end of the window, and consequently, it is also faster to detect the end of it. In contrast in majority voting, the activity remains undetected until it becomes the dominant label in the window and past windows in time still carry the label of the previous activity. For example in the majority voting approach, the blue dots represent the assigned weights that are neutral for all points. There, the start of the activity has a low impact (red line, weight contribution) on the window to trigger detection, and is labeled as "inactive". Similarly, when the activity ends, the older samples that remain in the window continue to dominate the label and thus the window stays labeled as "active". In the second approach, the weighting adapts to such transitions and focuses on recent data that "enter" the window, recognizing faster the updates of the activity status.

Experimental Overview Configuration

Figure 5.17 below, presents the micro-activity road-map recognition that was followed from the beginning of the experiment until the final results that are visualized in the following sections. The experiment began with the evaluation of the models with benchmark data sets for human activity recognition followed by a performance comparison between those and the collected data for the study. The impact of non-screwing activities that belonged to the null class with similar patterns to the screwing classes and the class imbalance was investigated by removing the null class from the training and test data sets to examine the recognition rate between the rest of the classes. As the data collection happened over a long period, the models were trained gradually with more data in order to improve the results. Subsequently, in the data collection for the additional data in the minority classes, only one activity was recorded each time in each recording session, in contrast to the initial sessions where the daily assembly workflow was captured. Before the final validation, the electronic support refers to the button of the electrical screwdriver that records a signal every time the button is pressed during a screwing process.

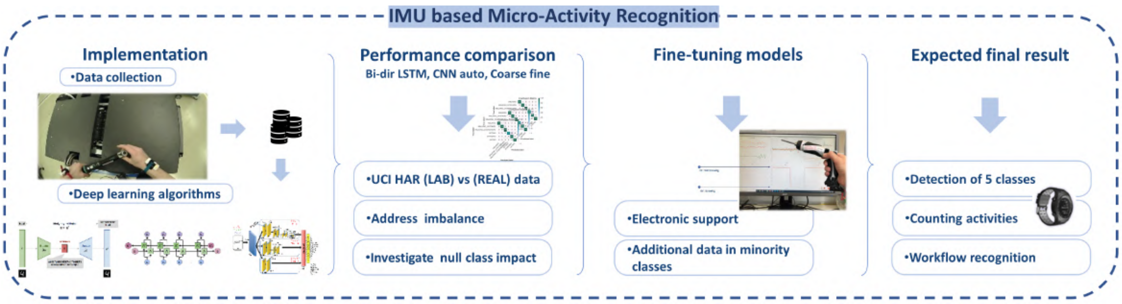


Figure 5.17: The figure demonstrates the suggested approach for Micro-activity recognition in steps. The implemented models were tested in the UCI HAR dataset before the captured real data at the industrial partner were used as input. An investigation into the impact of the null class was conducted and weighting techniques were applied to address the imbalance of the data, while algorithms were compared with each other. The additional data for the minority classes were obtained in independent recording sessions for each class before the electronic support was incorporated into the study. As a first step, five classes must be recognized and displayed on a wrist-worn modality e.g., a smartwatch. The final system should be able to count the occurred activities and understand the current stage of the workflow.

Manual assembly tasks in industrial human-machine collaborations can be regarded as a classification problem, hence accuracy, precision, recall, and F1-score (F1), are adopted to evaluate the results of the classifiers. It was necessary though to include balanced accuracy as a metric also, keeping in mind the imbalance between the classes. The recognition accuracy for each class is given as a percentage in the following confusion matrices, in the visualization of the results and the additional metrics on the summary table.

5.2.5 Completed vs Incomplete Activity

In industrial assembly, ensuring that tasks are fully completed is important for maintaining the quality of a product. Incomplete activities might lead to defects, or safety hazards, requiring rework time if they are late detected during the production process. However, tracking the status of the detected activity in real-time provides instant feedback and allows immediate corrective actions. Building upon the detection of activities, the following approaches were explored to monitor the task's completion:

- Torque-based detection using an electrical screwdriver – identifying relations in the generated IMU signals and the torque of an electrical tool for screwing.

- Acoustic-based detection using a microphone – capturing the characteristic click sound from screwing tools that indicates a screw is tightened.

The two approaches are designed to address different work-environment conditions and tool types. For example, the integrated sensor could be more reliable when it is used with electrical tools and in noisy environments. On the other hand, the audio-based approach could be more appropriate for non-electrical tools and less noisy settings, where the characteristic clicking noise can be effectively captured and analyzed.

Torque detection

During this experiment, a sensor that was able to measure the produced torque of the tool was developed by a hardware expert. The device was employed to examine correlations between the produced signals and determine if the final phase of the activity could be detected during the assembly of an ATM's parts. The "torque sensor" was integrated into the electrical screwdriver that is visualized in figure 5.18(left) and the IMU sensor transmitted data via the IMU to a server for further analysis. The graphs in figure 5.18(right) illustrate the correlations that were observed between the "PPG" value which indicates the functionality of the electrical screwdriver's button and the integrated "torque sensor".

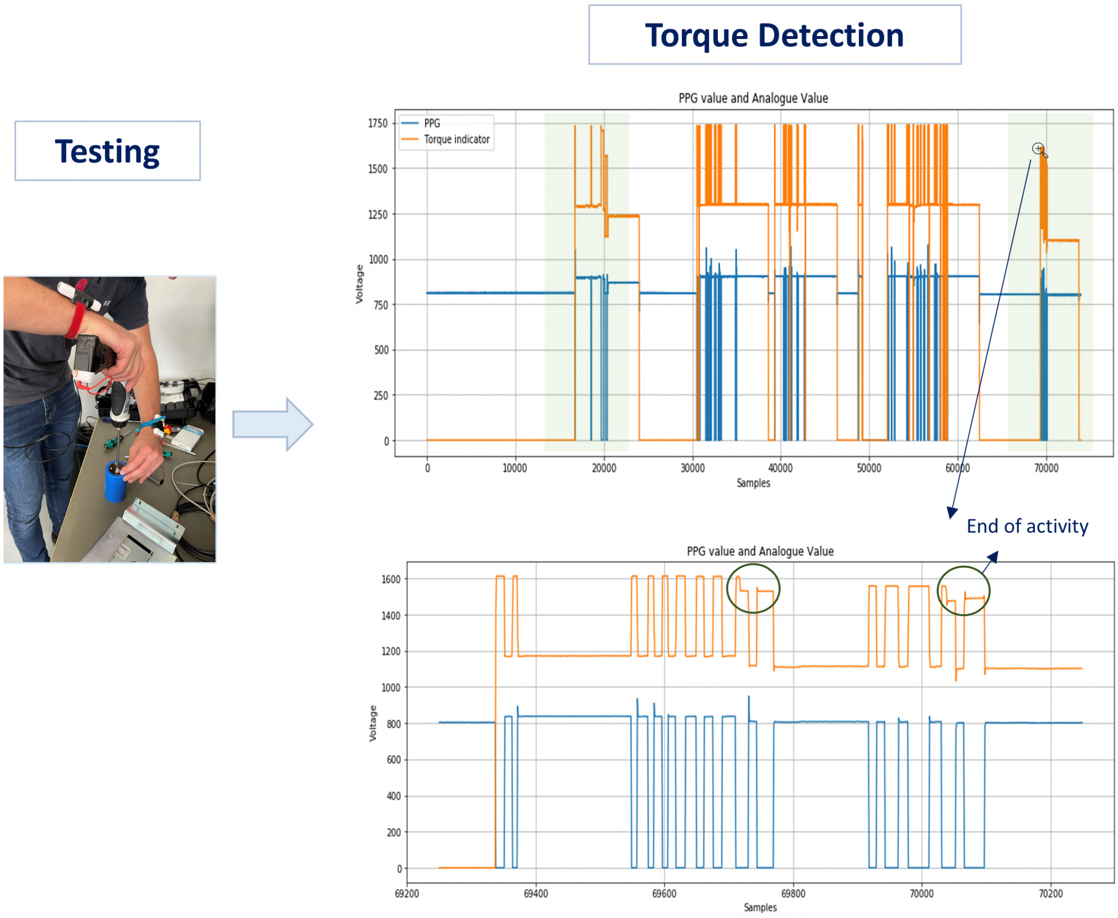


Figure 5.18: The figure visualizes the completed activity detected in the pattern of a custom torque sensor specifically developed for the use case. The orange lines show the torque values and the blue lines the use of the electrical tool. The orange ones that are always in the same scale before the drop at the end, suggest that the worker is either pressing and leaving the button in one screw or switching between multiple screws without reaching the end of the activity. In some cases, the button on a screw may be pressed slightly resulting in gradual or partial screwing. The drop in signal at the end of each sequence pattern represents the end of the activity.

The bottom graph in figure 5.18, visualizes a segment of the screwing activity where multiple start-and-stop instances of the tool can be observed. These remain within the same scale and indicate that the screwing activity took place, but the screw was not fully tightened. Instead, it was only partially inserted or screwed. A characteristic signal drop is recognized at the end of those patterns and represents that the tool reached the tightened phase or achieved the final torque marking the completion of the activity.

Audio based detection

In this setting, a torque wrench is used to fasten hoses during the assembly of a digger's hydraulic control unit by human workers. The wrench tool emits a clicking noise when a screw is fastened with the required force. This sound is the indicator that informs the worker that the activity is completed and the next activity can begin. However, it was observed in the data that the workers might fasten the same screw multiple times with the noise appearing in the data various times for one screw. A directional wireless microphone was employed in this setting that records clicking noises and is worn by the workers as a headset.

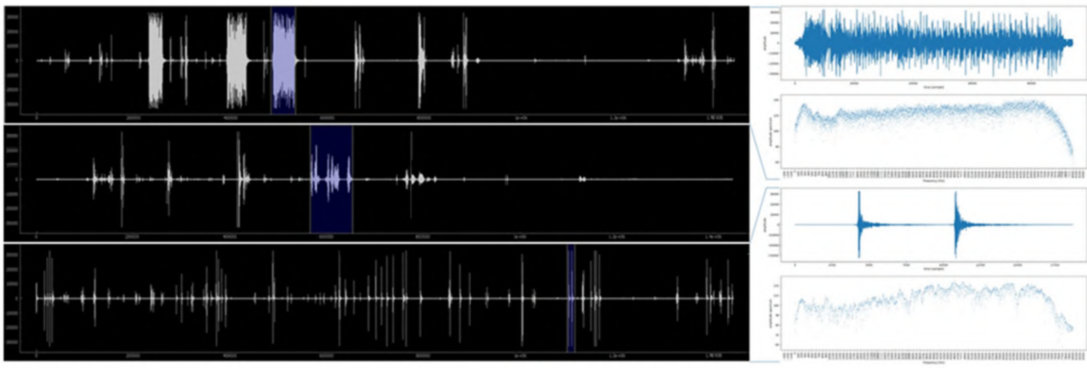


Figure 5.19: The figure presents examples from the collected audio data that were used by the authors in [160]. (Top Left) Signal from an electric screwdriver, (Top Right) Highlighted signal with corresponding frequency feature vector, (Middle Left) Signal containing human speech, and (Bottom Left) Double clicking of a torque wrench with its derived feature vector (Bottom Right). (*The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*)

The algorithm that was developed, clusters the clicking sounds based on time, to ensure that each group corresponds to a fastening action. The expected cluster count is predefined as the number of hoses to be fastened is known for each workpiece and workstep beforehand. **More information about the specific use case, can be found in subsection 6.2.4 and in [160].

5.2.6 Experimental Evaluation - Results

This subsection provides more details on the data collected for each class of interest in this study, the duration of activities, the tools used, and the patterns produced for each activity. The results visualization begins with the application of traditional machine learning approaches at the initial stages of the study.

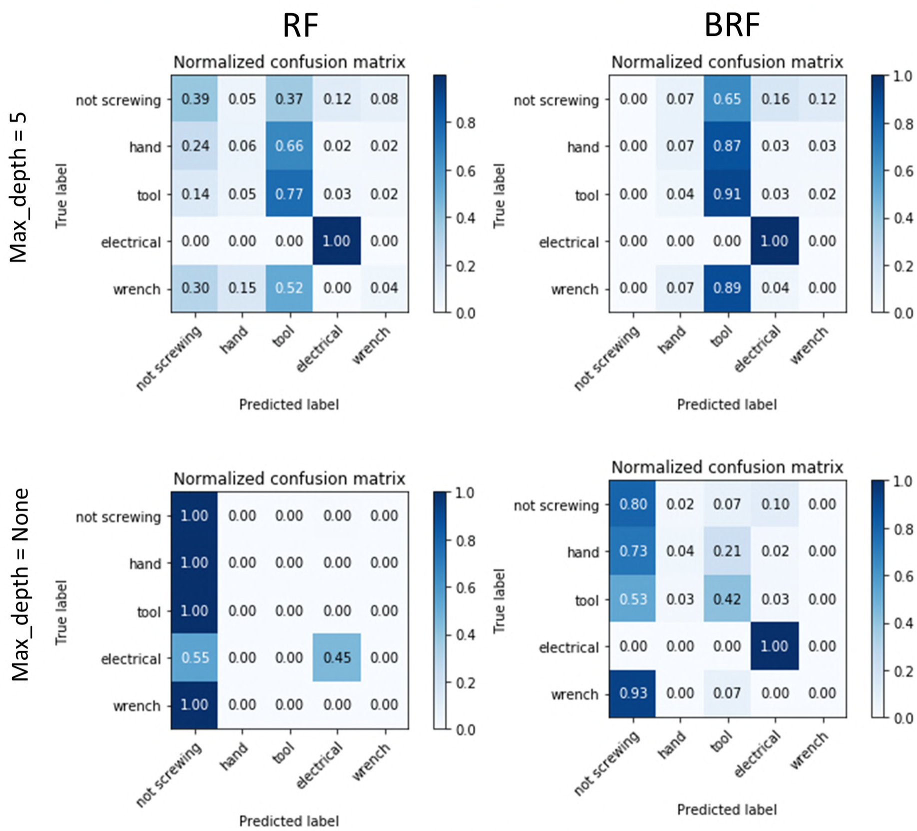


Figure 5.20: The figure shows initial results with traditional machine learning approaches. Specifically, a comparison of Random Forest (RF) and Balanced Random Forest (BRF) performance since there is a big imbalance on the data, under different maximum depth settings. The window size is 2000ms and the sliding rate 0.25.

Figure 5.20 visualizes the results of random forest (RF) and balanced random forest (BRF) in the dataset that was acquired from the ATM use case. While these models provided a basic understanding, their performance in this specific use case proved less than the expected. Consequently, deep learning approaches, which are frequently recommended in the literature for their ability to capture complex patterns and their simpler implementation process, were explored in subsequent stages of the study.

The implemented deep learning models were tested in the UCI HAR dataset before the collected data of the real use case were used as input. The reason for that is that in the initial phase of the experiment, public datasets offer more reliable benchmarks for model selection. The performance of the implemented models for the UCI HAR data set and for the collected data was evaluated for the same window size (2560ms) and sliding rate(0.5) to compare these data sets with each other. They were then trained with multiple combinations of window sizes and sliding rates for classification into five classes and the best results for each architecture are presented. The results for the UCI HAR dataset showed that the implementation of the coarse fine medium model was better than the others which is also the same for the captured data as visualized in figure 5.22c.

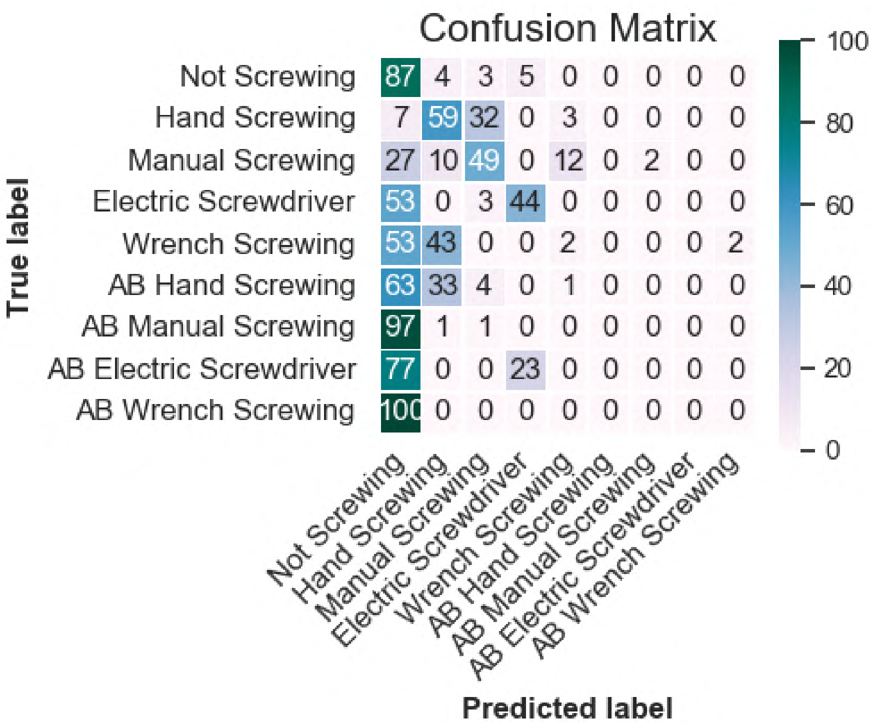
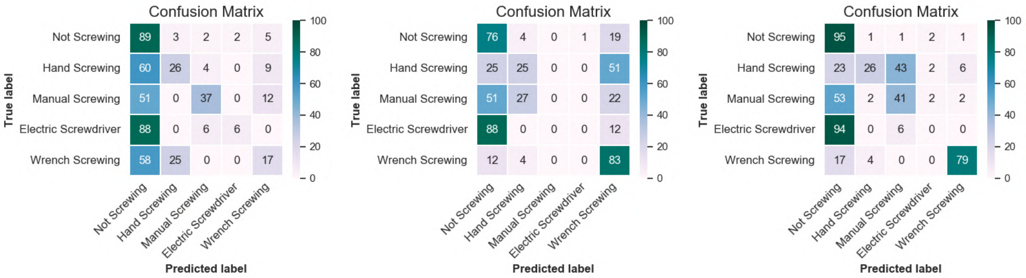


Figure 5.21: The figure shows a performance with abnormal classes using an LSTM model on the use case data. The performance metrics for the classification are the following: Accuracy: 78.51%, Bal. Accuracy: 26.82%, Recall: 78.51%, F1 score: 78.61% .



(a) **LSTM**
WS: 2560ms , SR:0.5

(b) **CNN Auto**
WS: 2560ms , SR:0.5

(c) **Coarse fine**
WS: 2560ms , SR:0.5

Figure 5.22: The figure presents a comparison between deep learning architectures with early collected data. The data set at this stage consists of data before the collection and addition of extra data to minority classes. The window size and sliding rate were chosen to be similar to the UCI HAR settings where the models were first tested. Additional metrics in table5.1.**WS: Window Size, SR: Sliding Rate.

The imbalance in the data set led to trivial recognition for most of the classes, however, it improved with:

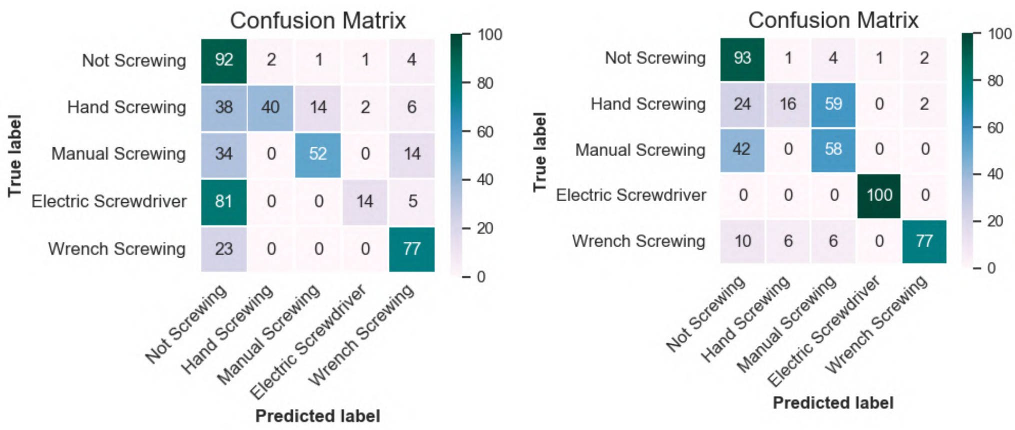
(i) the addition of data in the minority classes,

(ii) the application of sample and class weighting techniques,

(iii) and the selection of a smaller window size.

Figure 5.21, presents an LSTM model with abnormal (AB) classes in the detection process. Abnormal classes, such as AB Hand Screwing, AB Manual Screwing, AB Electric Screwdriver, and AB Wrench Screwing, represent outliers or unexpected variations in the dataset, which can make the classification task more challenging. In this first classification result, there are some classes that are detected sufficiently and others poorly. The values of the abnormal classes show that there is a big correlation to the null class, making those classes similar to each other or similar to the activities of the null class.

Furthermore, the tables present the results after the addition of data for minority classes Figure 5.23a, as well as, an evaluation of the electronic support Figure 5.23b. A significant improvement that shows complete recognition of the electrical screwdriver is visible in Figure 5.23b, mainly based on the button's electrical signal. In the same figure, one can see the confusion between hand and manual screwing as two classes that share a similar motion pattern. Better prediction is achieved for all the classes especially by merging the hand screwing and the manual screwing classes as the hand movement and the produced sensor signals are similar for them. The final result can be seen in Figure 5.24 where the recognition for all the classes has been improved. The summary Table 5.1 indicates that for the screwing activities of the presented study, with the small motion range usually a smaller window size is preferred.



(a) Coarse fine
WS: 2000ms , SR: 0.5

(b) Coarse fine
WS: 2000ms , SR: 0.5

Figure 5.23: (a) The figure visualizes examples of early results before the addition of data to minority classes to the main dataset. (b) The figure gives the results for the dataset including the additional data for minority classes and the valueP. The values of the metrics can be seen in table 5.1.

The deep learning models that are presented have been selected to highlight the progression of methodologies and the improvements during the process. The utilized "attention" models that were tested, were outperformed from other models in this use case. Subsequently, they are not presented in this thesis as they did not provide beneficial results.

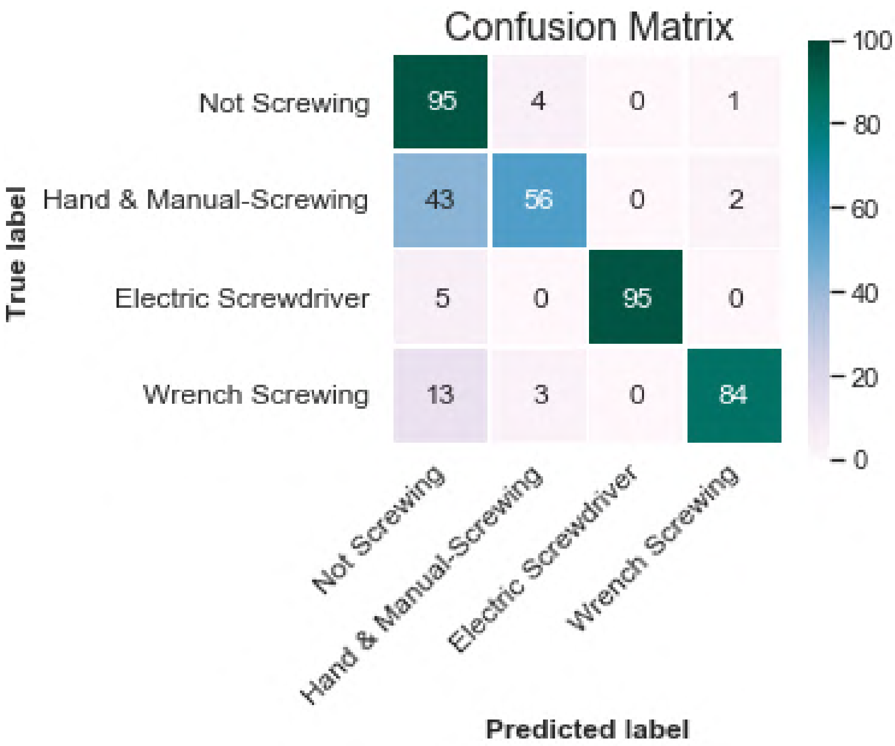


Figure 5.24: Final evaluation Coarse fine
Accuracy: 91.19%, Bal. Accuracy: 82.41 %, Recall: 91.19%, F1 score: 91.28%. WS: 2000ms, SR: 0.5

Table 5.1: The summary table gives an overview of the optimal results achieved for each experiment and presents the models with the best recognition performance in terms of metrics. The second column describes the training data used as input for the models. Org refers to the original data collected from the normal workflow. "Org+Minor" indicates the addition of data to minority classes and the val_P the addition of data for the electrical screwdriver.

A/A	Training Data	valueP	Classes	Architecture	Window_size	Sliding rate	Weighting	Accuracy	Balanced accuracy	Presicion	Recall	F1 score
Fig. 5.22a	Org data set	No	5	LSTM	2560	0.5		81.19	35.01	83.68	81.19	82.36
Fig. 5.22b	Org data set	No	5	CNN auto	2560	0.5		69.44	36.75	82.89	69.44	74.73
Fig. 5.22c	Org data set	No	5	Coarse Fine	2560	0.5		87.74	48.30	87.73	87.73	87.56
Fig. 5.23a	Org+Minor	No	5	Coarse Fine	1000	0.5	Standard	86.20	54.91	88.30	86.20	86.93
Fig. 5.23b	Org+Minor+val_P	Yes	5	Coarse Fine	5000	0.5	Standard	88.03	58.81	90.28	88.03	88.54
Fig. 5.24	Org+Minor+val_P	Yes	4	Coarse Fine	2000	0.5	Standard	**91.19**	**82.41**	**91.47**	**91.19**	**91.28**

Table 5.1 shows a summary of the best results of the conducted experiments, showing the accuracy, balanced accuracy, precision, recall, and F1 score for each setting. In the second column of the table is mentioned the data set that was used for the experiment, followed

by the architecture, the window size, and the sliding rate. The valueP refers to the use of the electronic helper that was implemented from the beginning of the study and used only in a later stage to demonstrate the performance of the models with and without this adjustment. The "Standard weighting", points to a set of class weights that were identical for all the experiments where the setting was applied, and assigned a lower weight to the null class to handle the imbalance. The number of the classes states the classes that were contained in the data set for each case. Higher results are not always achieved between the same window sizes and sliding rates, due to the different data used in each experiment and the difference in the duration of the activities Fig .5.10, in each class as well as, across separate classes. In the table is also apparent that the balanced accuracy increases with the addition of data and with the merge of the hand and the manual screwdriver. The different activities are shown in figures 5.6, 5.7, 5.8, 5.9.

5.2.7 Discussion on Activity Classification

Limitations with sensors and data annotation

For machine learning and particularly for deep learning a large amount of accurate data is important. This study started with two sensors. One on each hand but progressively additional sensors were used in order to increase the recognition rate. The assumption was that an ankle sensor would be helpful to detect when the worker is moving away from the workstation. Figure 5.25 visualizes, the test that was implemented using transfer learning trained on PAMAP2 [253], and tested on the collected ankle data from workers in their working space.

The sensor on the electrical screwdriver would help to identify when the specific tool is being used, considering that the wrist's sensor signals were too weak to be recognized by the algorithms. Due to the standstill character of the activity e.g., the user is keeping his hand almost stable for 0.5-1.5 seconds, the acceleration and angular speed do not change much, consequently, the confusion with activities from the null class was frequently observed.

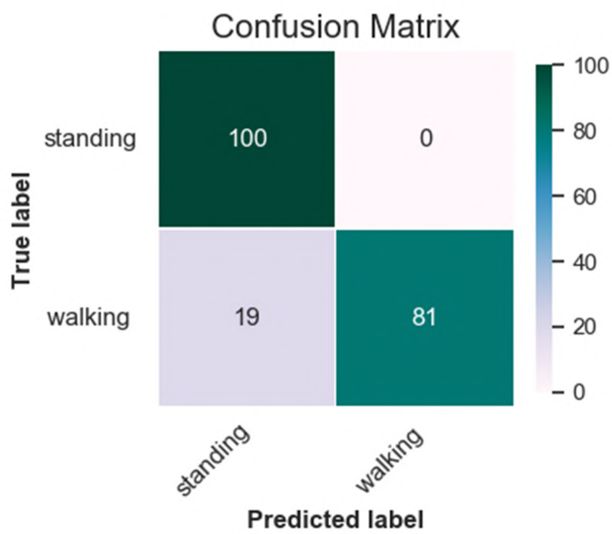


Figure 5.25: The confusion matrix shows the results of a coarse-fine-medium model trained on PAMAP2, and tested on Keba ankle sensor data. The model classifies standing vs. moving, providing insights into the status of task execution, as workers are typically stationary during certain activities. WS: 2000ms, SR: 0. Acc: 87.54%, Bal. Acc:90.43%.

The annotation of the data, where the eye-tracker was employed capturing the ground truth, is a process that requires great attention because it can affect the structure of the study. It is a time-consuming procedure and especially with IMU signals, it can be harder to find the exact beginning and end of an activity. Furthermore, for different repetitions of the same activity, there might be a variety of patterns to consider as the start and end. For example, there were multiple times that a) the temporal length of one activity was very different as a result of different sizes of screws or various screwing techniques e.g., one, could mount and fix one screw as one complete activity or someone else could perform the activity in two steps, in the first step mount a number of screws in multiple positions and in the second step fix them, b) the user was performing activities with non-correct tools, for example, hand screwing with the electrical screwdriver.

These behaviors could not be eliminated, due to the personal style of each individual performing the same activity. Some workers for instance use a screwdriver mainly by turning their wrists while other workers use their fingers. Inter-class similarity as an issue, was difficult to be avoided, as activities such as hand screwing or screwing with a manual screwdriver produce similar patterns. The additional data collection supported a higher level of recognition for the algorithms.

Proposed approaches to challenges

In the previous section, the drawbacks that were faced during this study were mentioned. However, the reason for most of the study's difficulties was the size of the null class and consequently, the variety of redundant activities contained in it. Increasing the amount of data assisted the classifier, but since the null class was also increasing, individual recording sessions for each class was a better approach.

Another method suggested in the literature was to reduce the null class by removing some samples (under-sampling) which in this use case decreased the overall accuracy. Oversampling techniques with and without data augmentation were also tested for the less populated classes, which was beneficial for some of the activities of interest, but nevertheless resulted in more confusion for the rest. The techniques that were tested were rotating, jittering and dynamic warping [250] and combinations of those.

Class and sample weighting methods were applied to force the model to pay higher attention to specific classes and samples, trying to achieve a desirable balance between the classes. By inspecting the raw signals, it was observed an entanglement between classes, namely the inter-class similarity that was mentioned before shows the association among the groups and the intra-class variability which is defined as the confusion within the same class. Regarding the inter-class similarity between different classes the same strategy was followed. Consistency in the signals annotated with the same label was decisive. Activities such as screwing with a hand or screwing manually with a screwdriver sometimes were producing comparable signals due to the similar movement of the hand.

Collecting extra data for each class of interest proved to be more useful in this use case, compared to other approaches suggested for imbalanced data sets. Individual data recordings for each one of the classes to enrich them and make the system more robust [58], were decided as a plan against the size of the null class. This experiment was again put in execution, at the site of the industrial company partner, with the difference that the worker was performing the same activity multiple times, e.g., screwing 10 screws with a wrench, unscrewing them, and screwing them again without interruption of further interfering and insignificant activities. The worker used his personal style and executed the order of activities, without any further instruction.

The tuning of parameters and hyper-parameters for each model was not an easy task to achieve and most of the attempts were based on suggestions or implementation of models that were found in literature and in the trial-error process. Furthermore, the hypothesis that an activity of interest is never executed during walking, led to the additional sensor at the worker's ankle detecting walking. This implementation could be carried out in two steps: 1) one model to detect walking and 2) one model to recognize the activities. This approach could be extended to integrate the part of the completed vs incomplete activity detection, improving the model's ability to distinguish between intermediate and fully completed actions.

5.2.8 Findings on Activity Classification – Summary

Based on the findings presented in this chapter, the research question can be addressed as follows:

RQ3: What methods can be employed to detect and extract information on key activities within manual assembly processes to assess the recognition of task execution in industrial workflows?

Key activities such as **micro-level activities** play a fundamental role in assembly processes, as described in the taxonomy in the previous chapter since they contribute to task execution and efficiency in DWL. The high importance of micro-activities as **goal-directed units in industrial assembly**, is directly linked to product quality and workflow accuracy. Their recognition can address key challenges such as **short activity duration**, inter-task similarity, and lack of **availability in public datasets**. Additionally, the findings demonstrate the potential of micro-activity detection to offer real-time support to the workers without overwhelming them, reducing their cognitive load and thus the manual assembly errors.

Sesnor-based activity recognition methods were implemented such as **IMU sensors** on workers, **wrists ankles, and tools** for recording the activities, along with an eye-tracker that was used only for ground truth annotation. Continuous and **real-time recordings** provided the data that were needed to be used by the **machine and deep learning models** to extract information and classify the activities. Approaches were selected in order to avoid extensive preprocessing steps and manual feature extraction that is typically required by traditional machine learning methods, thereby simplifying the overall workflow. After testing multiple models and parameters, the **coarse fine medium model** was able to identify and differentiate between key assembly activities within the workflow with an **accuracy of 91.19%** and increase the balanced accuracy to 82.41%. **Confusion matrices** provided insights for **misclassifications** guiding the path to optimize the models in order to improve the model's recognition.

The challenges that appeared such as **class imbalance, varying activity durations,** and the presence of uninformative "null" class activities, were addressed using a combination of approaches. **Weighting approaches** was applied to balance the classes, **approaches for assigning a label** to windows were implemented, **different window sizes**, and **overlaps to optimize segmentation**, methods for **detecting the end of the activity** were presented and activities that share similar features were combined to improve classification detection.

The integration of the models in the **selected feedback device** which is a smartwatch equipped with the same IMU sensors, will offer **instant feedback** through alerts, workflow tracking, and **personalized support** to workers based on their activity recognition patterns. These methods were used to **detect key activities** and enable an assessment of execution quality by tracking executed activities **or identifying** indicators for **errors** through deviations from the expected **workflow** in **real-time**.

5.3 Activity Counting

5.3.1 Introduction to Activity Counting

Following the presentation of the approach for identifying and categorizing activities that were outlined in the previous chapters, the next step in cognitive augmentation is to recognize how often these activities occur. This part is important as it provides quantifiable insights into the efficiency of manual assembly processes in real-time and the understanding of potential errors, reducing rework time and costs in the post-assembly steps. Leveraging machine learning techniques to analyze time-series data from industrial workflows enables accurate counting of such actions, in addition, to support in cases of ambiguous annotations and varying task durations. This chapter explores methodologies for counting and analyzing activity occurrences to optimize manual assembly and thus provide cognitive assistance to workers in manual assembly environments.

Motivation

Building upon this existing body of work, the scope of this study is to highlight the counting of events that occur in a given period in different activities of daily human life (DHL) or daily work life (DWL). The focus is on counting the end of an activity to determine the number of times that activity occurs. By centering the attention on event counting, the aim is to provide a comprehensive understanding of the frequency and occurrence of specific actions within the broader context of human activities. In daily activities, such as workouts or sports, it is critical to correctly segment and recognize the type of activity using a sophisticated model [254]; however, mainly classification models can offer such information. Besides that, it is important to acknowledge that different fields exhibit variations in sensor types, and signal characteristics, produced by these sensors and face different challenges.

“ Can a recurrent neural network learn to count things ? ”

Figure 5.26: "Can a recurrent network learn to count things?" In their work, this was the main question that was investigated by the authors in [82].

With small adjustments, counting with AI and IMU data can be used in the industry to solve a variety of problems. Some examples include:

(i) Sensor data analysis to monitor the performance of equipment, detect anomalies, and optimize operations;

(ii) Quality control to count the number of defects or errors and improve the quality of products and reduce costs

(iii) Safety monitoring to count the number of incidents to improve safety and reduce the risk of accidents in industrial environments.

For example, in an industrial setting where the tasks are more complicated, workers have many repetitive tasks to complete daily, such as screwing activities during assembly processes, which they occasionally miscount or forget to execute [185]. In this regard, the aim is to provide people with information and raise awareness about the number of completed activities.

As Kim et al. [255] stated in their work, counting is one ability that humans usually acquire from a young age, and while it appears to be a simple task, young people still need a long period to master it. Comparably, it is challenging to develop a model that can count the number of completed activities (CA) in a time period, based on data from Inertial Measurement Units (IMUs) or similar body-worn sensors. The term "completed activities", refers to the repetitive activities that can constitute a single work step in a workflow, e.g., the screwing of one screw, which is complete with the tightening of it.

The analyzed data for this research are sequences of varying lengths annotated with weak labels that serve as targets for the machine learning models. Traditional approaches for handling time-series data often involve dividing the data into fixed window lengths. However, due to the high variance in activity durations, using a small window for a long activity or a large window for a short activity can result in the loss of important information [251]. Up until recently, one commonly used approach required very large window samples that could fit all activity sizes inside, padding them with a value (typically zeros) and feeding them as input to the networks. Annotating data, on the other hand, is usually laborious and time-consuming, and requires considerable attention and precision. While the windowing and classification methods that were discussed in the previous chapter effectively detected activities, the new approach offers a more flexible, adaptive solution that minimizes manual adjustments and streamlines data annotation.

As an overview of the challenges motivating this work, the focus is on the following:

(i) spread of valuable information across consecutive sequences,

(ii) information loss caused by using a single, fixed window size for varying-duration activities,

(iii) limited flexibility of models that are more specific for particular data due to manual preprocessing methods, and

(iv) the topic of data annotation.

Proposed Methodology

To address the aforementioned issues, a model design is proposed that works with :

(i) variable length of data as input,

(ii) data that have some form of annotation but is not completely annotated, known as weakly labeled,

(iii) raw calibrated data that are normalized but not subjected to any further filtering, to reduce complexity, simplify the preprocessing stage, and develop a more robust model in unprocessed data to achieve counting,

(iv) the implemented counting method, integrated within the model's training process rather than simply incrementing the count of correctly classified instances, and

(v) fewer training parameters than existing model architectures in the literature to enable suitability for deployment on devices that have limited processing and power resources.

The approach aims to improve the model's ability to accurately count the activities performed by a user, rather than just detect them. To the best of the author's knowledge, this is the first study that investigates the counting of completed activities and tasks, in a way that goes beyond counting the quantity of previously correctly recognized activities from a classifier, employing an LSTM [256] for counting patterns in a sequence.

5.3.2 Counting Approach

Counting repetitions in a sequence is a fundamental problem in various fields, such as speech and image processing, bio-informatics, and cognitive science. Many methods for counting can be deployed, the majority of which require hand-crafted rules, feature extraction and statistical methods, or rule-based systems to manually count objects or events. As was described previously in 2.3.2, neural-network-based approaches can be used to count repetitions in a sequence with a combination of CNNs and RNNs or encoder-decoder architectures, trained on a labeled dataset of sequences, to learn a mapping between the input sequence and the number of repetitions. Focusing on a system that can handle weakly labeled data while being less reliant on human intervention and more automated can reduce the complexity of the counting process, make the model more robust, and provide flexibility for applying the method to a wide range of problems.

Weakly labeled data offers a cost-effective and efficient alternative to acquiring fully labeled data, as it is less expensive, time-consuming, and tedious. This type of data enables the use of semi-supervised learning approaches, which can be beneficial when considering the annotation cost associated with large and complex datasets. By leveraging weak labels, the model is encouraged to learn more generalized patterns in the data, leading to improved performance on unseen examples. For the experimental setup, IMU acceleration data are used. The data are from daily human activities performed at a quality control checkpoint in a car maintenance scenario that captures activities relevant to the inspection process. Thus, they represent real-world conditions by addressing the complexity of variable-length data, to develop robust and realistic models for activity counting and recognition tasks.

Despite the fact that counting repetitions in a sequence with variable-length data is more challenging than counting repetitions in a sequence with fixed-length data, this structure is more realistic because data from signals, time series, texts, and other sources have varying lengths. Using fixed-length tensors for the data can be efficient in certain situations because they simplify the problem, as the model only needs to process a fixed amount of data, regardless of the length of the input sequence. Furthermore, because the libraries and software tools required to build the model are more widely accessible, its implementation and deployment may be simpler. However, using variable length tensors can also be beneficial in many situations. They allow the model to handle input sequences of different lengths, which is important when dealing with complex real-world data and activities of various lengths. Additionally, variable-length tensors enable the model to process the entire input sequence at once, rather than only a fixed-length subset of it, which can be valuable when the position of the repetitions is not known in advance.

In this approach for counting, data from public datasets that contain data from human activities in the car manufacturing industry, recorded with IMU sensors, are used. Sequences of data that have a variable size are created and for each sequence, a weak label is obtained. The label shows the execution number of one type of activity observed in the sequences, which is fed into an LSTM regression model built with the ragged tensors. For each sequence with a random size that is used as input to the algorithm, one single count is predicted as the output.

5.3.3 Experimental Design

The data used for this study are part of the Skoda Public dataset [242], which includes repetitive activities regarded as single, discrete actions as opposed to continuous activities, such as walking or running. The dataset contains activity classes recognized from body-worn sensors during various car maintenance tasks. Each activity is associated with specific hand movements and interactions with different parts of the vehicle, such as opening doors, checking gaps, manipulating the steering wheel, and others that are visualized in figure 5.27.

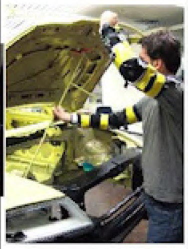
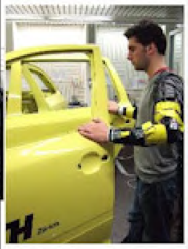
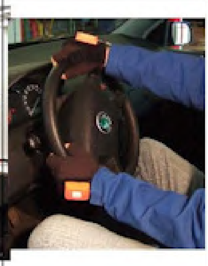
				
The user holds a notepad with his left hand and writes down a short sentence with his right hand.	The user opens the hood with his left hand and blocks it with a stick kept with his right hand.	The user removes the stick with his right hand while keeping the hood with his left hand then closes the hood with his left hand.	The user checks the gaps on the front door by sliding his left and right hand over the gaps. The two hands move simultaneously.	The user grabs the car left front door with his left hand while it is closed and opens it completely.
				
The user grabs the car left front door with his left hand while it is open and closes it completely.	The user grabs the car left front and back doors with his left and right hands than open and close completely and at the same time the two doors	The user checks the gaps on the trunk by sliding his left and right hand over the gaps. The two hands move simultaneously.	The user opens the trunk using both hands and then moves it up and down on the top of his head three times before closing it.	The user grabs the steering wheel with both hands and turns it clockwise and counterclockwise three times.

Figure 5.27: Source: Reproduced from [257]. The figure shows the available list of activity classes recognized from body-worn sensors during various car maintenance tasks. More information about the Skoda public dataset publication is available in [257].

Dataset Description

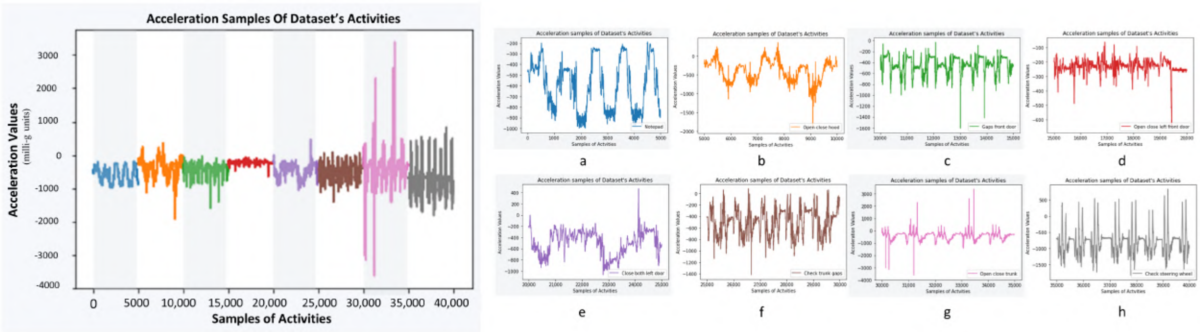


Figure 5.28: This figure shows examples of the signals that represent each class of the Skoda Public dataset with different colors. Starting from left to right (**a**) Notepad, (**b**) Open close hood, (**c**) Gaps front door, (**d**) Open close left door, (**e**) Close both left door, (**f**) Check trunk gaps, (**g**) Open close trunk, (**h**) Steering wheel. For all the activities, samples are taken from the X-axis accelerometer on the right hand. The activities of the "open hood" and "close hood" as well as the "open left door" and "close left door" are displayed together as "open close hood" and "open close left door", since they are always consecutive. The acceleration is provided in milli-g units.

The example signals of the manipulative gestures of the dataset that were performed in a car maintenance scenario, visualized in figure 5.28, are "write on notepad", "open hood", "close hood", "check gaps on the front door", "open left front door", "close left front door", "close both left door", "check trunk gaps", "open/close trunk", and "check steering wheel". These activities were recorded for about 3 h by 20 sensors placed on one subject's left and right upper and lower arms. For each sensor, there are acceleration values on the x, y, and z axes that are calibrated in milli-g units (1000 = earth gravity vector, which in S.I. units would be 0.001 g or 0.00981 m/s^2), and the sensor sample rate is approximately 98 Hz, as stated by the dataset's authors.

Sequence Annotation Pipeline

The objective of this work is to count how many times one activity happened in a period of time, e.g., detect in the data patterns how many times the person closed the hood in the activity "close engine hood". The dataset contains a dense label for each sample, which allows the detection of the end of the activity. The weak label targets are generated by "recording" one repetition for each task completion. Therefore, for the training phase, for every instance where a task is successfully completed, a single repetition sample is marked as “activity end”.

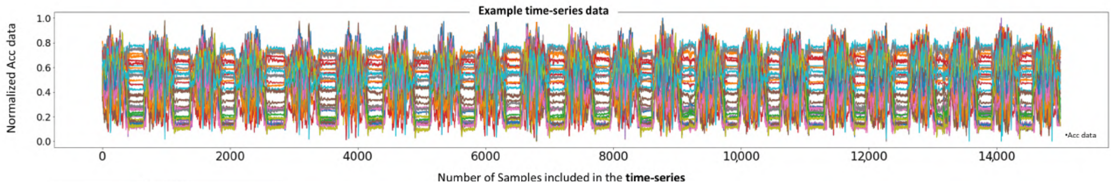


Figure 5.29: This figure presents an example of the time-series data with 14.000+ samples that will be divided into distinct example sequences of segmented acceleration data with varying duration, number of samples, and weak labels that have a range from 1 to 10 counted activities. The data are normalized between 0-1 as shown in y-axis.

This approach, creates weak annotations that indicate the presence of completed repetitions, allowing the model to learn and recognize the patterns associated with activity completion. Then the data are normalized with minmaxscaler [258] in a range of [0, 1] and divided into variable-size sequences. Figure 5.29 presents one part of the data before it gets segmented into variable length sequences. By using an algorithm to generate an array of random numbers, the entire dataset is split into segments, which define the sample length of the sequences. For example, if the aim is to generate 20 sequences of variable-length data, the algorithm will create 20 random numbers between 0 and the dataset’s maximum index value.

The labels for each sequence in this dataset are produced by the number of spotted endings or finished tasks in the sequences, where the last timestamp of each observed activity adds 1 count to the final label of each unique sequence. Figure 5.30 visualizes the division of a time series into sequences of variable size and how the weak labels are formed. The term “weak labels” in the presented method denotes the absence of data annotations that map the start and end of an event in the sequence. The number of activities in the sequence is the only information of the sequence that the model utilizes as a target value.

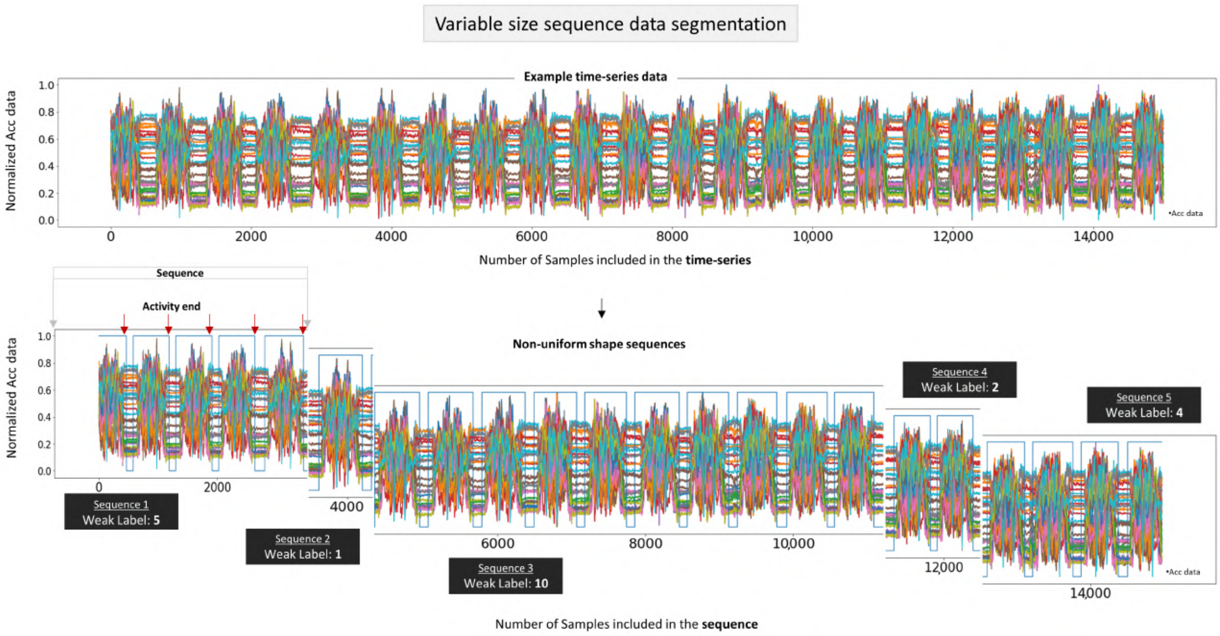


Figure 5.30: This figure presents the non-uniform shape input for the neural network. We visualize on top of the image time-series data that will be divided into five distinct example sequences of segmented acceleration data with varying duration, number of samples, and weak labels that have a range from 1 to 10 counted activities. The blue squared line shows the start and end of each activity within each sequence. The weak label is generated by the number of spotted endings (red arrow) inside each individual sequence.

The original data is replicated for each activity to provide the model with a larger dataset to train, without using data augmentation techniques for generating variation in the signal's patterns. This expanded dataset introduces greater variability in the unique length of activity sequences and the number of activities contained within them, thereby enhancing the model's robustness. The subsets of activities are divided into 600–900 sequences, where 100 of each type were left as a test dataset, as presented in Table 5.2 below, and 10% of each training dataset was used as a validation set. Consecutive activities, such as "open left front door" and "close left front door", were merged into one class, as explained in Figure 5.28. In this case, the algorithm must count +1 when one of the activities of interest is happening. In the last entry of the table with the label "combined activities", one can see results with 8000 training sequences, for a class that is generated with combined data from all the previous classes in a single one. Every activity that is not a null class will be counted in this scenario, to distinguish between an occurring activity and the null class without considering the type of activity. Finally, 8000 varying-length sequences from all classes were created, of which 7000 were used to train the network. Despite the more complex approach, employing variable-size sequences allows the most valuable information from the used data to be extracted without padding.

5.3.4 Counting Algorithm Development

A fixed window approach is a commonly used method for segmenting time-series data before using them as input to deep learning models. The segmentation is based on characteristics of the event that we want to identify, such as its periodicity, frequency, and length, among others. Despite their ease of implementation and interpolation with other libraries, fixed-length tensors with a predefined shape have limitations. For example, they are not well suited to handling non-uniform shape data, such as sequences of varying length, without the need for padding or truncation, which can result in additional noise to the data, increase in the computation time, information loss, and storage inefficiency, and it may not always be appropriate. To address the above issues, we used TensorFlow's ragged tensors [246], which support variable-length sequences of samples.

In this study, nine sub-datasets were used with our algorithm to count activities, with nine separate trainings for each subset. Figure 5.31 shows the architecture of the model that is used for the counting task. The acceleration data are separated into variable-length sequences, each of which comprises several activities and is used as input data, as was previously mentioned. The weak label that the model uses as target data is the total number of activities in each sequence. The model learns to relate acceleration data to the number of activities, so when we feed as input " new unseen" acceleration data of variable length, it outputs the number of spotted activities. The annotation provided no information about the location of activities within the sequence, nor did it provide any additional supporting details to guide the model.

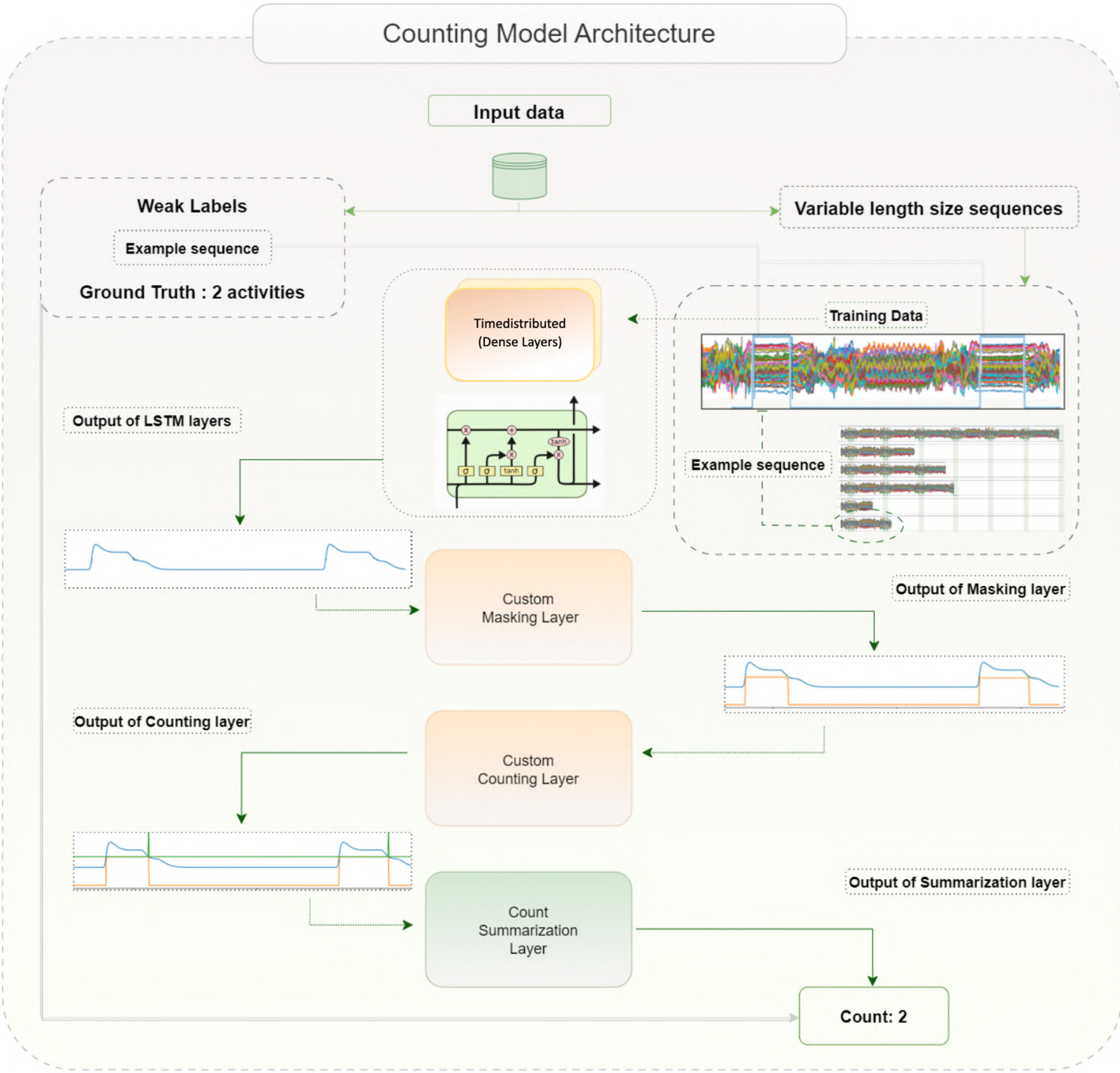


Figure 5.31: This figure presents an example of the model architecture that was used for the counting of activities. The acceleration data are divided into variable-length sequences and then used as input to the model. For each sequence, there is one weak label that is generated by the number of activities that are included in the sequence. Two time-distributed dense layers process each sensor reading independently before entering an LSTM layer where we get an output for each time step. Since the input data have a variable size, ragged tensors are employed for this task. The output of the LSTM part is inserted into a mask layer that detects values above a threshold and converts the signal into a square form before it continues to a layer that detects the created "edges" of the square shape and gives a final summation of all edges of the sequence to one single number.

A large grid search was deployed to explore the best combination of parameters for the number of layers, learning rate, batch size, optimizer, loss function, and activation functions to use in our network. Various hyper-parameters were explored to achieve the best performance. For the number of layers, experiments were conducted with configurations ranging from one to three time-distributed layers and one to four LSTM layers. For the learning rate, values such as 0.01, 0.001, 0.0001, and 0.00001 and different batch sizes, including 2, 4, 8, 16, 32, and 64 were tested. As for the optimizer, experiments occurred with Adam, RMSprop, and SGD while loss functions, such as mean squared error, mean absolute error, and Huber loss were evaluated. Finally, different activation functions, such as ReLU, tanh, and sigmoid, were tested to achieve the optimal performance for the task.

Figure 5.32: From left to right, this image presents the network's diagram of the counting model with the input (60 acceleration signals) of variable length and output of 1 number. Moreover, the hyper-parameters include Huber loss, ADAM optimizer, a learning rate of 0.0001, and a batch size of 2. An example of a learning curve for the training and validation sets demonstrates the model's performance during training. The x-axis represents the number of training epochs, while the y-axis represents the loss metric.

As shown in Figure 5.32, two dense layers were used at the beginning of the model to reduce the dimensions of the input data before entering the LSTM. The first dense layer is composed of 60 neurons (number of input signals), and the second consists of 2 neurons with the rectified linear unit (RELU) as an activation function. Two custom layers are then placed after an LSTM layer that outputs a three-dimensional sequence and has one neuron with a linear activation function. The custom masking layer thresholds the signal and converts the output of the LSTM to a more binary format, and then, the counting layer counts the regions where the signal value is not zero and summarizes them to one final number, as shown in Figure 5.31, of the output's graph.

After experimenting with several parameter values as mentioned above, a batch size of 2 with a learning rate of 0.0001 and the "Adam" optimizer were selected for optimizing the model. The algorithm's performance is evaluated using the Huber loss as the loss function, which is a combination of the mean squared error (MSE) loss function and the mean absolute error (MAE) loss function. This combination improves the performance of the model when outliers are present in the data, which is possible in our study because the input sequences were generated arbitrarily.

5.3.5 Experimental Evaluation - Results

In this study, deep learning approaches were applied to acceleration data for counting the number of activities in variable-length sequences, as presented in the model architecture. The ground truth in Figure 5.31 is two activities in the illustrated example sequence. It is evident that the LSTM outputs a signal with two peaks, which is then converted to a binary format by the masking layer, and counts +1 at the edge of each square area. Two is the final result predicted by the model for the specific input sequence. Figure 5.32 visualizes the learning curve of the training and a validation loss to present the model's performance during the training of the "notepad" dataset. After each training, the model was evaluated with unseen data sequences of the same class, and the results show that the model can predict very close to the weak label.

Table 5.2: This table lists the overall summary results for accuracy and MAE across all activity classes for the test datasets. As one can see, it contains nine separate datasets of activities. For each activity, the samples of the original dataset, the variable length training sequences created from the data, the range of the number of activities within the sequences, the test sequences, and the results such as the dataset accuracy and mean accuracy in 100 sequences of each test dataset are available. The counts' range shows the maximum number of activity counts contained in 100 sequences of different lengths of the test dataset. The mean percentage accuracy provides an overall assessment of the model's performance by displaying the average deviation of the predicted number of repetitions from the true value across all test sequences, whereas the dataset accuracy assesses the model's ability to predict the precise number of repetitions accurately. The last entry in the table represents a class where all the data from all classes were combined into one class and then split into variable-length sequences so that the model is trained on more complex data. In that case, the model learns a larger variety of patterns from all classes as a single activity class and must identify between the activity class and the null class to perform the counting.

Training A/A	Activity	No Of Samples in Original Dataset	Training Seq.	Range Counts in Test Seq.	Test Seq.	Test Dataset Accuracy	MAE	Mean % Accuracy in Test Seq.
1	Steering wheel	51,904	500	0–08	100	60/100	0.4	72.19
2	Check trunk gaps	70,000	500	0–07	100	89/100	0.11	91.66
3	Notepad	74,000	500	0–06	100	92/100	0.08	96.12
4	Open close hood	186,399	800	0–07	100	70/100	0.3	78.58
5	Open close left door	82,000	600	0–12	100	68/100	0.33	80.22
6	Gaps front door	60,000	500	0–09	100	84/100	0.18	90.38
7	Close both left door	72,000	500	0–06	100	75/100	0.25	79.22
8	Open close trunk	95,000	600	0–10	100	74/100	0.26	81.65
9	Combined activities	705,904	7000	0–11	1000	765/1000	0.242	81.29

Table 5.2 contains information regarding the training data and the results of all dataset activities. The discussion section provides further details about the results. The table shows in approximation the number of data samples contained in the original dataset for each class, the number of training and test sequences, the range of the number of activities included in different sequences, the accuracy in the test data, the mean absolute error, and the mean % accuracy of test sequences in the test data.

For two of the activities of the Skoda Dataset, example results are visualized, of the weakest and best cases of the model's predictions in variable-length data sequences. In Figure 5.33, one can see the graph showing the ground truth and prediction of the model for the activity "writing in notepad". There, the model was trained with 500 sequences of variable size and variety in the number range of contained activities. The findings reveal that the model predicts 92 out of 100 correctly, while the remaining 8 predictions have an error of one activity. Similarly, even though the model predicted less accurately for the dataset's "steering wheel" class, the predictions have a maximum error of one activity, as shown in Figure 5.34.

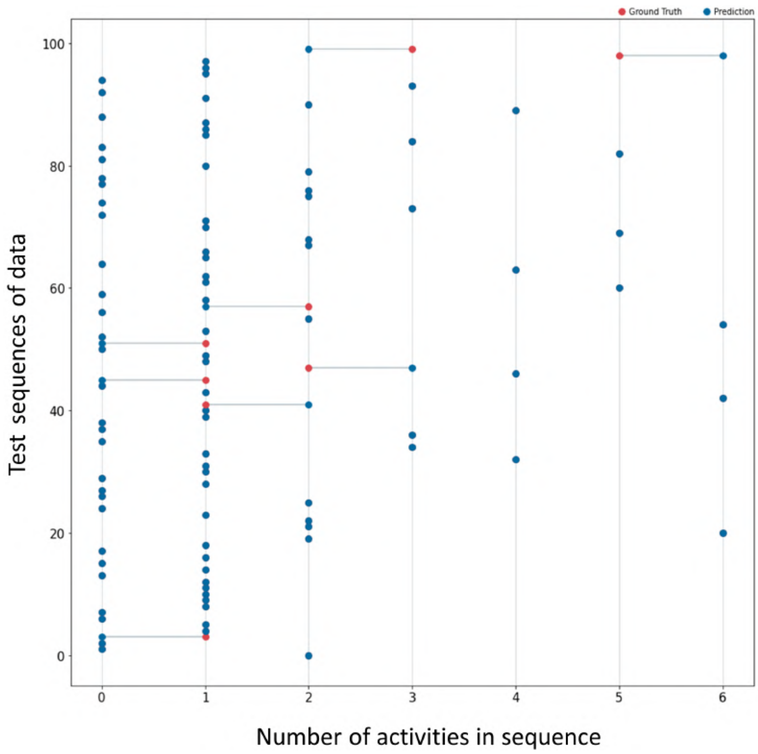


Figure 5.33: The figure presents the ground truth (red dots) and prediction (blue dots) of the model for 100 unseen sequences of data. The input data are from the notepad writing activity of the Skoda dataset. The algorithm for counting predicted accurately 92 out of 100 activities. A line connecting the two numbers shows the difference in the incorrectly predicted sequences. The largest error per sequence observed in the graph is 1 count.

The "notepad" class has the smallest MAE, 0.08, and the highest mean % accuracy, while the "steering wheel" class has the highest, 0.4, and the lowest mean % accuracy. In the "combined activities" class, the algorithm counts interesting activities in a sequence, regardless of the activity type, in a dataset consisting of all classes combined into a single one. The table lists 8000 training sequences of varying lengths, of which 7000 were used for the network's training. Randomly, 1000 sequences were kept as test data, and the number of activities in 765 out of 1000 was predicted correctly. From those 1000 sequences, 230 had an error of ± 1 counts, 3 of them an error of 2 counts, and 2 of them an error of 3 counts. As shown in the table, the mean accuracy is presented as a percentage, representing the average accuracy of 100 sequences from each test dataset. For example, for the "open close trunk" dataset, the mean accuracy for the 100 test sequences is 81.65%. This means that the accuracy of the model was found for each predicted sequence of this test dataset, and subsequently, provided an average estimate of the accuracy across all the test sequences to evaluate the performance of the model. The dataset accuracy evaluates the model's ability to predict the exact number of repetitions accurately, while the mean accuracy gives an overall measure of the model's performance by indicating the average deviation of the predicted number of repetitions from the true value across all test sequences.

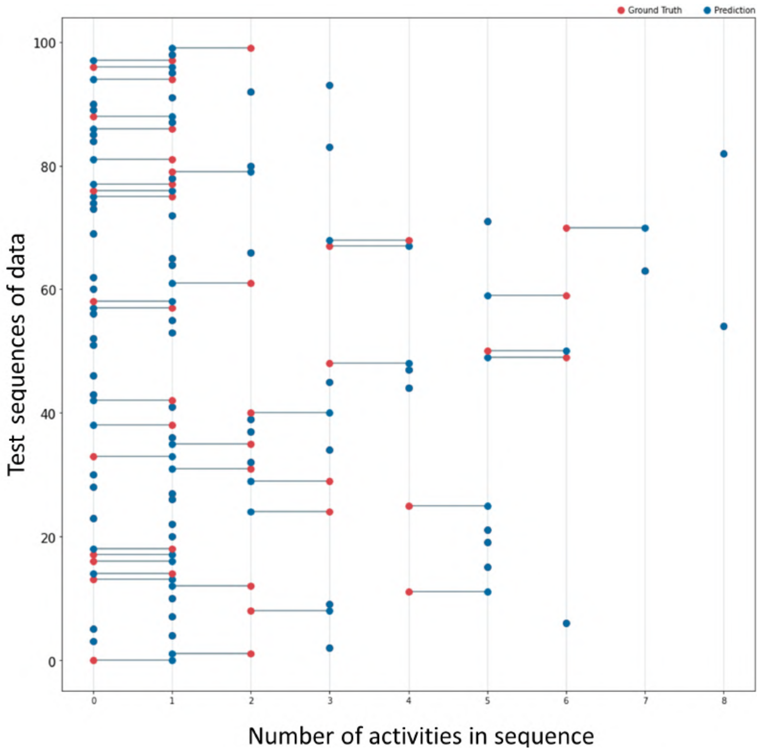


Figure 5.34: The figure presents the ground truth (red dots) and prediction (blue dots) of the model for 100 unseen sequences of data. The input data are from the steering wheel activity of the Skoda dataset. The algorithm for counting predicted accurately 60 out of 100 activities. A line connecting the two numbers shows the difference in the incorrectly predicted sequences. The largest error per sequence observed in the graph is 1 count.

5.3.6 Discussion on Activity Counting

The current work confirms that it can count interesting events in time series with more flexibility concerning the size of each input sequence from a model that uses :

(i) solely normalized data,

(ii) variable size length data,

(iii) weak labels, and

(iv) a deep learning approach.

According to this method's preliminary findings, when the model is trained for specific activities, the algorithm can accurately predict, in most cases, the exact number of times an event is repeated in a sequence. For some of the activities, the prediction is better than others. For example, activities such as "open close hood" and "open close left door" contain patterns of both opening and closing the object, which can possibly create a larger confusion for the model to recognize the pattern. The lower results were achieved for the steering wheel class. The wheel rotates three times in each direction, clockwise and counterclockwise, before switching. In this case, the orientation for each side within the same sequence may be contributing to the confusion, or the data may not be sufficient for the model's design and a deeper architecture or new data may be needed to capture the dependencies. Likewise, for the "combined activities" class, in each sequence, the model may contain patterns from one or more different activities that need to be counted. However, due to the weak labels, no other information about the activity type is available, except for the total number of events contained in each. At the end of the study it was observed that after a certain timestep, some minor synchronization misalignment appeared. This could also be a reason for the weaker results in some classes suggesting an opportunity for further refinement in future experiments. The consistently small error, typically within the range of ± 1, demonstrates the effectiveness of our current architecture in accurately counting activities. However, it also highlights the potential for further research and improvements to reduce this error even further and achieve even more precise activity counting results.

The model must be trained on a dataset of labeled sequences where the number of repetitions in each sequence is known, regardless of the architecture being used. To address the diversity in the different activities without padding and by utilizing the entire information of the sequence, weak labels that are less time-consuming and variable-size sequences are used. Nevertheless, it is important to note that using weakly labeled data does introduce certain limitations since the data are only partially labeled. A more comprehensive target for the model might be provided, for example, by a second model trained with information on the type of activity occurring, the location, the duration, or even data examples to use for training. A multi-task learning model that shares layers between signal reconstruction and the HAR tasks inspired by the presented model in [159], could be tested to explore the robustness and generalization of the model when applied to unseen users.

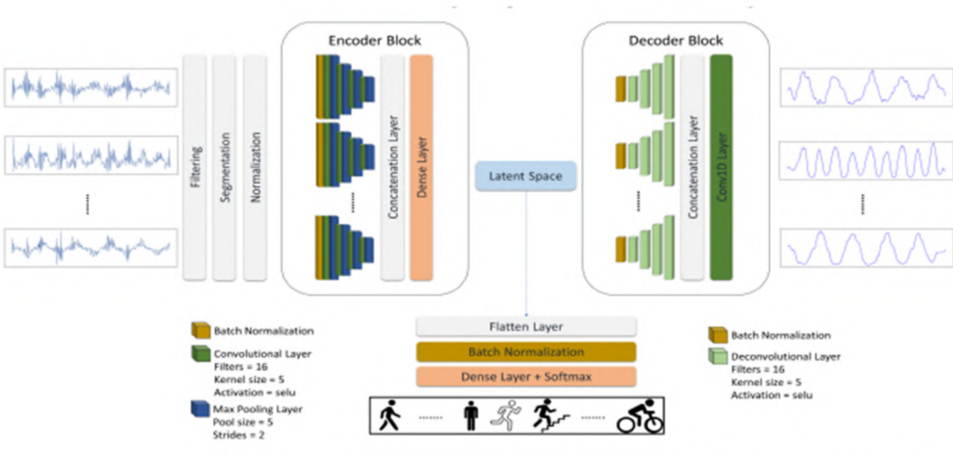


Figure 5.35: The figure presents the architecture of a proposed Multi-Task Learning Model (MCAE) by the authors including the present author in [159], combining unsupervised signal reconstruction with supervised human activity recognition (HAR) classification through a shared latent space.

In the presented work, a ragged tensor model was developed using calibrated data, that were normalized on a range of [0,1] to ensure a common scale. The calibrated version of the data is selected, because it is in S.I. units and can be replicated by anyone even though the raw data of the public dataset with our model provided comparable results. Additionally, the calibrated data support the use of any sensor that takes readings using the same units, not just the specific sensor that the dataset's authors used in their study. Despite the benefits of ragged tensors, such as efficient storage and easy handling of variable-length data, working with them proved to be challenging, requiring additional effort and consideration of alternative approaches, as some operations, libraries, and software tools outside of the TensorFlow environment are not currently sufficiently supported.

5.3.7 Findings on Activity Counting – Summary

Based on the findings presented in this chapter, the research question can be addressed as follows:

RQ4: How can machine learning techniques be leveraged to count repetitive actions in time-series data from industrial workflows, addressing ambiguous annotations and varying task durations?

Counting can be beneficial for a variety of fields, such as health care, sports and fitness, robotics applications, identifying actions and events in industrial settings, etc. Building on the classification performed derived from the introduced taxonomy, counting completes the recognition cycle by enabling **quantifiable detection** of activities. Especially in **industrial places**, the **repetitive tasks** are challenging to count because of inconsistent lighting, occlusions, noise, and diverse worker behaviors. In this work, the complexity starts with the **variability in the durations** of the activities and the labor-intensive process of generating annotations. To address these challenges machine learning techniques are used, and particularly **deep learning** approaches are employed to exploit the **raw data** minimizing the need for extra feature creation.

A **LSTM-based architecture** is designed to recognize and count the repetitive activities from raw, normalized **IMU acceleration data** to prevent the use of visually intrusive devices and the unwanted collection of identifiable personal information. The acceleration data are divided into **variable-length sequences** and then used as input to the model. For each sequence, there is **one weak label** generated by the number of activities included in the sequence. Two time-distributed dense layers process each sensor reading independently before entering an LSTM layer, where we obtain an output for each time step. The output of the LSTM part is inserted into a mask layer that detects values above a threshold and converts the signal into a square form. This processed signal then continues to a layer that detects the created "edges" of the square shape and provides a final summation of all edges in the sequence, resulting in a single output number.

LSTMs are well-suited for time-series data due to their ability to capture temporal dependencies and handle sequences of varying lengths. Furthermore, the variety in the duration of activities enables the use of **non-fixed-size windows** to avoid the loss of critical information and the bias toward a specific class(window based on class's duration).

Instead, **ragged tensors** are used to handle the non-uniform data shapes that are used as input to the models. The selected model is designed to work with the variable size data and achieves high **accuracy with a repetition margin of +-1** in most cases. The data preparation process is simplified with the use of **weakly labeled data** to reduce the manual annotation effort since this method can work with partially labeled data. Additionally, the model is designed to work on low-power devices since it demands **limited computational resources** making it practical for real-time deployment in industrial applications.

Integrating such a system into a cognitive augmentation pipeline for manual assembly can significantly improve the operators **awareness during the DWL tasks** and support the **error tracking** by counting and providing **real time feedback** on **task completion**.

6 Cognitive Augmentation

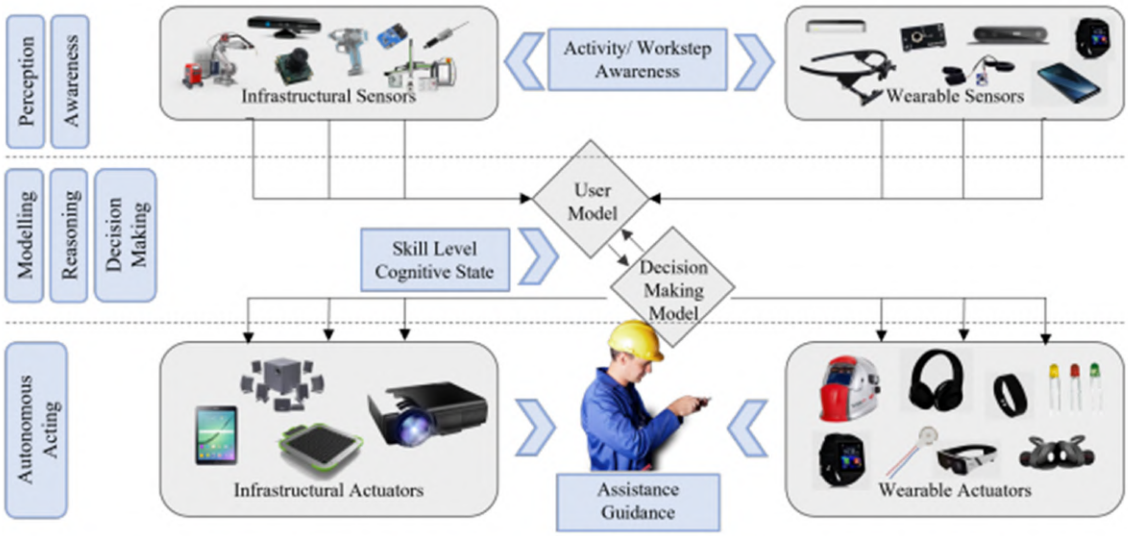


Figure 6.1: Source: Reproduced from [18]. The figure illustrates a three-layered cognitive architecture for supporting workers on the shop floor as presented in [18]. The system integrates perception and awareness through infrastructural or wearable sensors to facilitate decision-making and provides assistance via infrastructural and wearable actuators to enhance worker efficiency. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

This research seeks to expand knowledge on the topic of augmentation with cognitive systems. The term "cognition" heavily relies on domains of computer science and psychology, with the latter becoming a dominant theory to understand and explain intelligent behaviors that focus on the human mind in recent years [184]. Many contributions have been identified over the years that supported the formation of the artificial intelligence research field with Alan Newel being one of the pioneers stating that cognition in humans operates in rules that can be modeled by computers [259]. Advances in ICT technology and artificial intelligence enable the instantiation of cognitive systems, which are becoming increasingly important in industries and manufacturing processes to augment workers' capabilities. In [18] the authors, including the present author, present a 3-level framework for cognitive systems that include sensors, modeling, and actuators as presented in Figure 6.1 to create assistant systems.

This chapter presents a practical implementation of cognitive augmentation, through prototypes, that were conceptualized or developed based on the taxonomy and the approaches discussed in earlier chapters. The aim is to bridge the gap between abstraction and real- world industrial applications, by showcasing their role in assisting workers with activity detection, real-time feedback, and workflow optimization.

6.1 From Research to Augmentation Prototypes

In industrial workflows, manual assembly processes are subject to inefficiencies because of human error, lack of systematic datasets, and insufficient integration of advanced sensors. Moreover, while low-cost sensors and deep learning techniques have shown promising results, their application in real-world scenarios remains limited, especially in specific fields with data containing ambiguous annotations and varying task durations.

With this work the aim is to address these issues that are not well-documented, leaving a gap in literature and industry for practical implementation and evaluation. Table 6.1 below, shows the identified research gaps from the literature review and their connection to the research questions. Furthermore, it provides context for each research question and describes the recognized problems based on the research gaps. Accordingly, outlines the corresponding objectives by offering insights into the challenges.

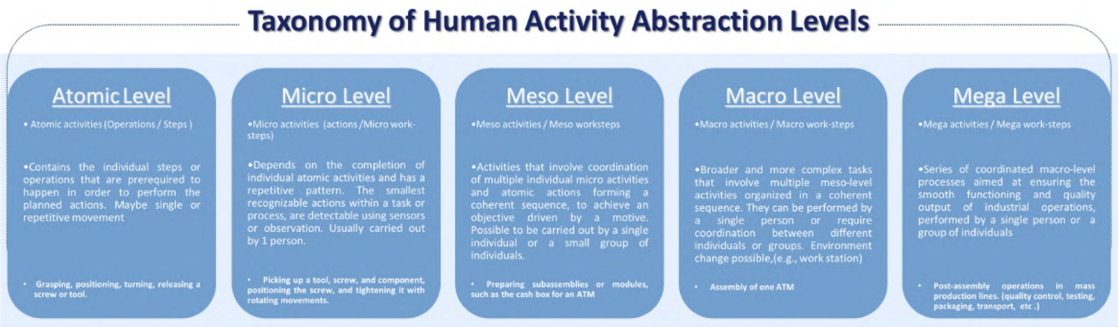


Figure 6.2: The figure presents the proposed taxonomy introduced in Chapter 4.

The combination of hierarchical activity abstraction levels as depicted in Figure 6.2 and the real-time recognition, classification, and counting of activities intend to build the foundation for creating cognitive augmentation systems. These systems integrate wearable and non-wearable sensor data, and advanced machine learning techniques, and enable real-time workflow monitoring and feedback to the worker.

In the subsequent sections, prototypes are introduced to demonstrate the feasibility of these concepts in real-world scenarios. These prototypes highlight the potential of cognitive augmentation systems across various industrial contexts.

Table 6.1: Research Gaps and Corresponding Research Questions.

Research Gaps	**Research Questions (RQs)**	**Context and Objective**
Errors in Manual Assembly Processes, Human Errors	RQ1: How can structured levels of activity abstraction in manufacturing work-flows be defined to enhance cognitive assistance on assembly processes?	**Context:** Current methods for categorizing activities are either too granular or too broad. Ambiguous categorization of activities to support cognitive systems in enhancing worker efficiency and support them with the right tools in each case based on the specific level. **Objective:** Propose a structured approach that organizes activities into levels and recognizes differences between them.
Datasets Availability, Wearables in Industry	RQ2: How can various sensor modalities (wearable and non-wearable) be utilized to recognize and categorize manual activities in industrial settings to support AI system design for enhanced human-machine collaboration?	**Context:** Wearables like IMUs and non-wearables like cameras have unexplored potential in recognizing manual activities. **Objective:** Develop methodologies to guide AI system design in activity recognition and categorization. Introduce specific tools, approaches, and methods, guiding both human understanding and AI development.
Complex Industrial Tasks, Low-Cost Sensors Assistive Technologies, Mobility in Workspace, Deep Learning in Production	RQ3: What methods can be employed to detect and extract information on key activities within manual assembly processes to assess the recognition of task execution in industrial workflows?	**Context:** Low-cost assistive technologies require robust methods for tracking task execution in diverse industrial workflows. **Objective:** Propose algorithms to identify and assess task performance. Focus on cost-efficiency, privacy, and scalability.
Deep Learning in Production, Case Studies for Assistance, Methodologies for Suitable Systems	RQ4: How can machine learning techniques be leveraged to count repetitive actions in time-series data from industrial workflows, addressing ambiguous annotations and varying task durations?	**Context:** Variability in task durations complicates annotation processes in time-series data. **Objective:** Leverage deep learning to count repetitive actions, improving annotation quality and activity recognition.
Suitability Exploration, Limited Literature Assistance Systems	All RQs: Address the gaps in developing prototype cognitive augmentation systems for assistance in daily work life.	**Context:** Limited research exists on the feasibility and integration of Cognitive augmentation systems into real-world workflows. **Objective:** Explore the design and feasibility of prototype cognitive augmentation systems, identifying key challenges and opportunities for implementation.

6.2 Empirical Paradigms – Prototypes in Action

Several practical applications were considered to better understand the presented concepts of the previous chapters and their mapping to the taxonomy and the research questions that are presented in Figure 6.3. The author of this thesis contributed to aspects of the works presented in Table 6.2, while the lead authors made the main contributions. Each use case demonstrates how cognitive augmentation systems address specific challenges in manufacturing workflows, based on hierarchical activity abstraction levels and and the proposed system design for AI systems.

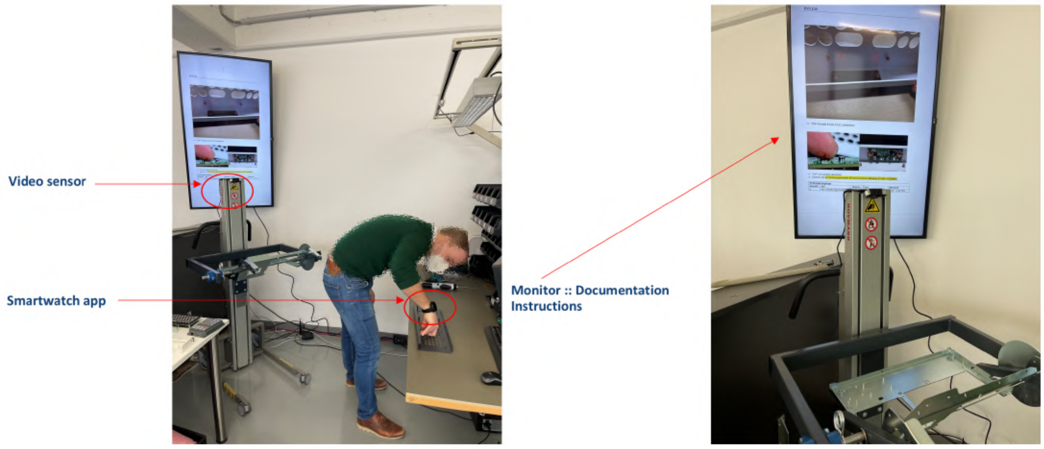


Figure 6.3: The figure visualizes exemplary sensors and actuators used in a lab setting to augment the operator's awareness and cognition. The setup integrates a smartwatch app, a video sensor, and a monitor displaying documentation instructions to assist the operator in real-time decision-making.

In addition to the taxonomy and theoretical background, expert insights were obtained through an industry-focused workshop, in which 20 participants expressed their opinions regarding the actuators that employees in the assembly sector consider suitable for their work. The results are presented in percentages to reflect the distribution of the individual's preferences, and their negative (light blue) and positive (dark blue) impressions were recorded as feedback based on what is useful and what is not.

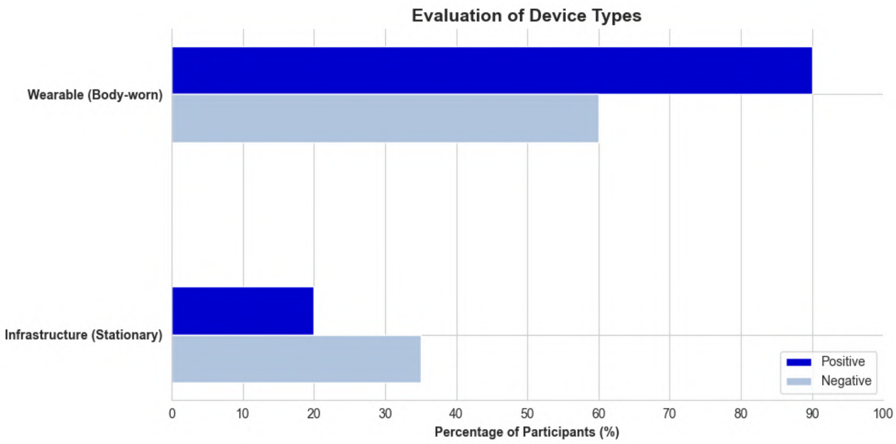


Figure 6.4: The figure presents the opinion of the participants on the evaluation of device type. The comparison is between wearable and stationary devices based on user feedback.

Figure 6.4 presents the evaluation findings regarding the device type preferred for the assembly task. Figure 6.5 outlines the preferred feedback method, highlighting visual and vibration signals. Figure 6.6 shows the participants' opinions on the overall usability of the selected device. Based on that, it was concluded that a body-worn smartwatch is best suited for the assembly task that considers micro-level, activities that could work mainly with visual and vibration signals. Additionally, actuators such as tablets or larger monitors could be available to display details about the workflow processes.

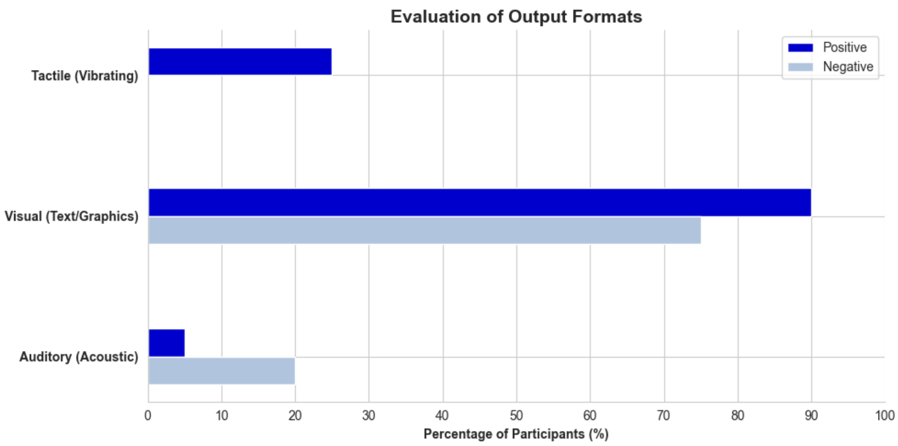


Figure 6.5: The figure presents the opinion of the participants on the evaluation of output actuators. The participants' preferences are visualized for tactile, visual, and auditory feedback.

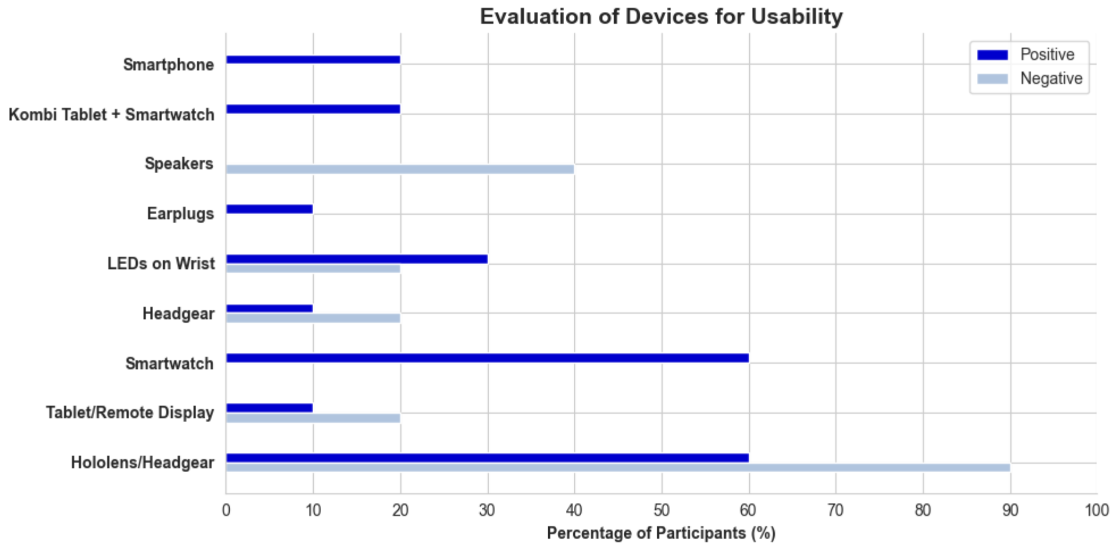


Figure 6.6: The figure presents the opinion of the participants on the evaluation of devices for usability. Positive and negative feedback are shown on different devices that could be used in an assembly process.

Figure 6.3 and Figure 6.7 visualize an example of the first use case, which demonstrates a smartwatch that can assist workers during product assembly. Specifically, this concept device of the first use case, describes how wearable devices can enhance task monitoring in dynamic assembly environments by classifying and counting repetitive activities in real-time. The device leverages the collected IMU and depth data and offers insights and notifications to the worker about micro or macro work steps. The next use case presents again a wearable sensor system using as input, the collected data from multiple sensor devices and traditional machine learning algorithms that occurred in an actual production line to achieve a realistic yet controlled working environment.

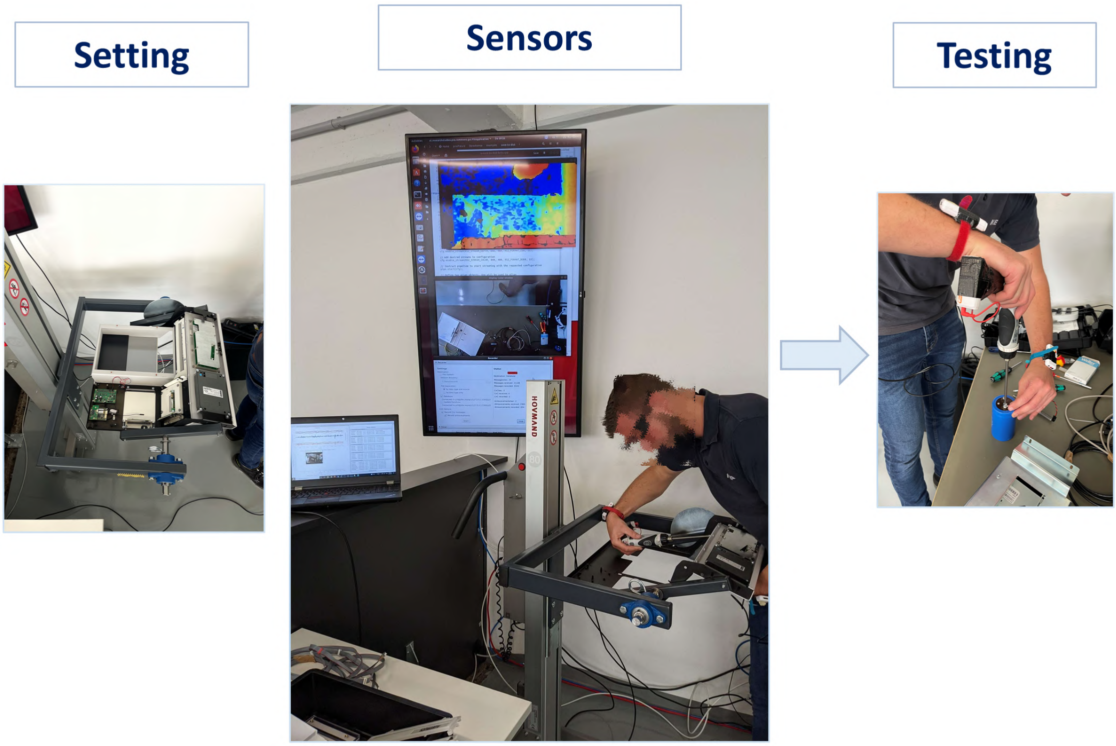


Figure 6.7: The figure presents the employee during an assembly task using sensors and actuators that were selected to be used in practical implementations for his cognitive augmentation.

A smart helmet was also designed to augment the capabilities of workers, equipped with several diverse sensors that offer important information to the operator through multiple modalities, including visual, haptic, and auditory feedback for industrial use cases. This type of helmet is used mainly in the welding industry and is particularly used for worker safety. Additionally, for the welding field, another application is introduced that analyzes arc welding techniques and recognizes the skill level of the worker. Finally, the last use case that is presented uses neural networks based on object detection and depth imaging to classify and monitor macro work-steps in production lines.

Table 6.2: Mapping of use cases to taxonomy levels, thesis chapters, research questions, and industrial contexts. The author of this thesis contributed to specific aspects of the following works, while the lead authors made the main contributions.

Use Case	**Taxonomy Level**	**Research Questions Addressed**	**Key Insights/Goals**
1. Smartwatch for Activity Monitoring and Counting [186]	Micro and Macro Activity Recognition	RQ1: Relevant levels of activity abstraction RQ2: Sensor selection RQ3: Activity detection RQ4: Activity counting	Combines IMU-based micro-activity recognition with macro-activity feedback for real-time assistance in assembly lines offering real-time activity detection and counting and workflow monitoring.
2. Smart Screwing in Industrial Assembly [160]	Micro Levels	RQ1: Relevant levels of activity abstraction RQ2: Sensor selection RQ3: Activity detection	Utilizes a combination of sensors and traditional machine learning algorithms to classify activities in an industrial assembly production facility
3. Smart Helmet for Cognitive and Safety Assistance [184]	Safety and Workflow Monitoring	RQ2: Sensor selection RQ3: Activity detection	Monitors worker movements and provides real-time safety alerts, enhancing situational awareness and workflow monitoring.
4. Smart Welding Assistance [161]	Micro Activity Recognition	RQ2: Sensor selection RQ3: Activity detection	Supports recognition of welding processes using low-cost sensors and traditional machine learning methods.
5. Depth Images for Macro Workstep Recognition [164]	Macro Activity Recognition	RQ1: Relevant levels of activity abstraction RQ3: Activity detection	Uses depth imaging to classify and monitor macro-level worksteps in production lines.

6.2.1 Smartwatch for Activity Monitoring and Counting

The objective of this study is to implement quality control of the smart manufacturing operation by providing guidance and support to novice workers in distracting situations or in case errors appear. Subsequently, creating customized models that can adapt to the employees in industrial IoT (IIoT) environments based on personal IMU and depth data. The goal is that the resulting prototype product will be an unobtrusive assistance system, embedded in a smartwatch to support workers in their daily work life (DWL).

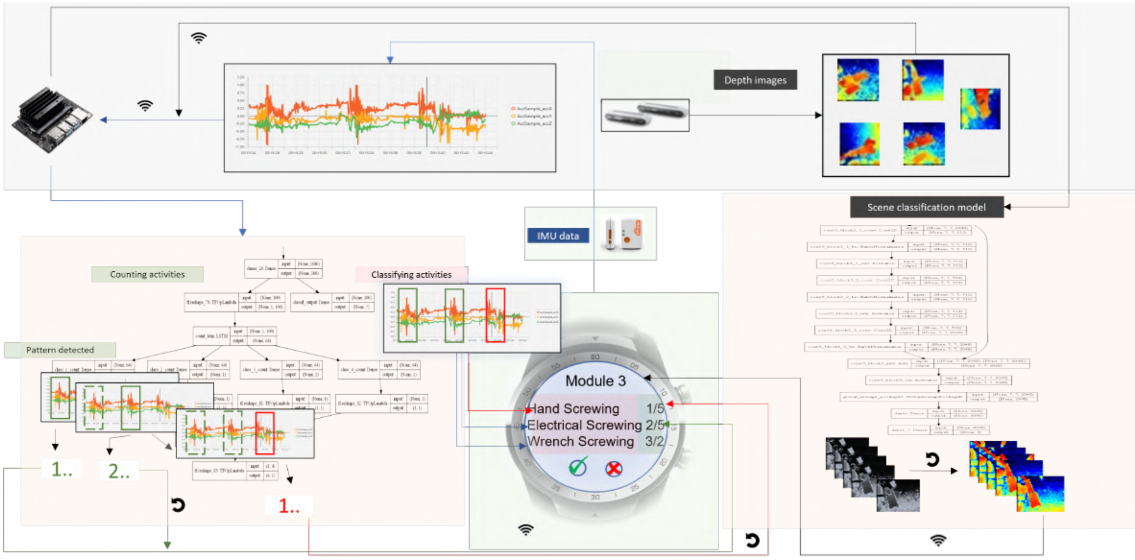


Figure 6.8: The figure shows the road map for an IIoT system proposed in this study showing the micro activity recognition in steps along with the workstep classification. The IIoT process begins with the initialization of the application on the smartwatch that enables the collection of the IMU and depth data as it is visible in the upper part of the image. The collected data is then sent to a CPU via Bluetooth and used as input to the deep learning models, which will have to predict the number of activities that constitute each module (left side of the watch in the image) and the correct module of the workflow in which the employees currently are (right side of the watch in the image). Individual deep models are trained with labeled micro activity data to learn to recognize similar patterns (classification part) and additionally perform scene classification by using as input labeled depth data that will be later combined in order to provide feedback through a wrist-worn smartwatch.

The whole process of the ATM assembly consists of many worksteps. Every workstep has several screwing actions that are predetermined and necessary to complete. The system will count the occurring screwing activities, to compare them with the required activities and provide the appropriate message to the operator. The aim is to i) create an IoT-based detection model to identify workflows and micro activities e.g., screwing detection in industrial processes, ii) provide support and guidance to the workers in real-time, particularly to novices, iii) perform quality control and quality assurance. Thus, assistance in minimizing the errors in the final product, suggests less rework time for workers, a faster error detection rate, and fewer additional costs for the employer.

Furthermore, the collaborative behavior of humans and IoT machines could reinforce the confidence of newly employed people and reduce the duration of their training. Connecting and managing these devices is mainly encouraged due to the IoT and the protocols used, in order to create discreet and unnoticeable industrial systems for employee support.

System Design

For the work step classification, we are using images coming from the depth sensor. Identifying the meso stages will help us know how many micro work steps are needed and how far workers have finished from the whole assembly workflow. The depth camera was mounted in a fixed position at a height of three meters with a top-down view. The view from the camera covered the ATM machine and the area where the worker is moving in to assemble the machine along with the tool table, Figure 6.9.

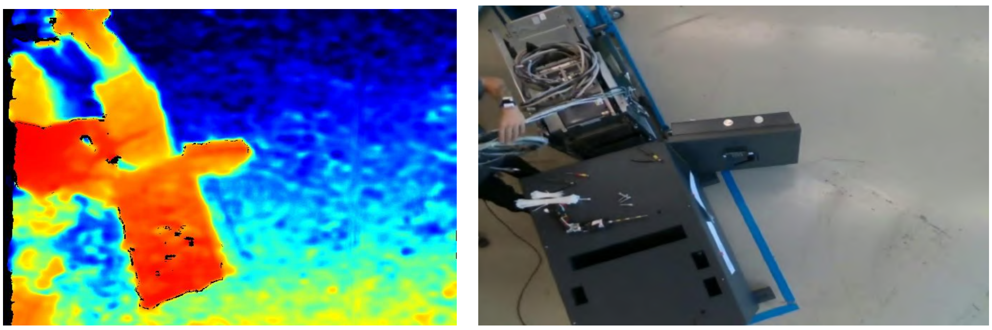


Figure 6.9: The figure visualizes the top-down view from the realsense camera, showing both RGB and depth images for an assembly work-step. In the right photo is apparent that the depth sensor keeps the employee's personal data secret.

For micro steps, data from the IMUs are used to identify micro activities of workers during the assembly of an ATM and provide confirmation of each complete micro work step or provide notifications that help workers focus on possible errors. The IMU sensors were mounted on workers' wrists and on an electrical screwdriver they use, as in Figure 6.10. The transmitted IMU data contain an accelerometer, a gyroscope, and a magnetometer, thus keeping costs low and ensuring minimal interference.

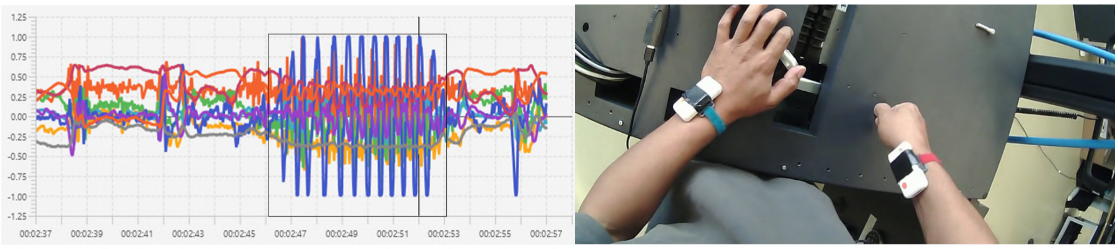


Figure 6.10: The figure presents an exemplary pattern in the data, for the class of hand screwing. For each class, the patterns produced by the specific actions of the worker were annotated and divided into 5 classes in order to use it in a classification model for supervised learning.

In the final stage, the user will be able to have the complete IoT system working in a smartwatch as visualized previously in Figure 6.3. Each IoT node transmits data to the IoT system, where the data from the IMUs are processed and the activity is recognized while the new workstep is classified based on depth images. The system will detect the activities, classify them as in Figure 6.11, and enumerate them in order to determine the stage of the process. Finally, the system will provide online visual feedback in its built-in monitor combined with vibrations for delivering messages in Figure 6.12. Information on the smartwatch includes how many micro activities are complete at each timestamp, how many activities are yet to be done and what is the current macro work-step, Figure 6.8.

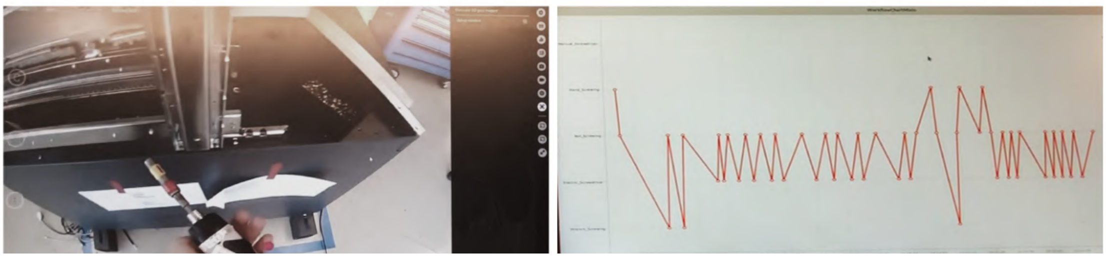


Figure 6.11: The figure presents a real test of the classification system in the assembly facility.

Currently, the IoT system is deployed in an industrial environment where the collected data are used to train and fine-tune the models. LSTM and CNN architectures are implemented to identify the activities and the worksteps for this supervised problem.

Figure 6.12: The figure presents the worker interacting with the system in the IIoT environment. A notification is received on the smartwatch for the completion of a macro workstep. The screen shows a more detailed report of the completed micro activities before the new macro step begins with a set of new tasks.

6.2.2 Smart Helmet for Cognitive and Safety Assistance

This use case demonstrates the design process of a cognitive headgear that can be employed in two industrial environments [184], visualized in Figure 6.13. The use case relates to research question 2 and presents a device consisting of various sensors and actuators with diverse functionalities for multi-modal assistance. In this work, the objective is to incorporate cognitive assistance and safety in products that the operators rely on, in their daily work life, during industrial processes. The head-worn device can support workers in welding applications and also provide navigation information through haptic, visual, and auditory feedback.

Industrial environments like welding and manufacturing plants require precision and efficiency. Additionally many tasks such as i) highly complex weld seams in 3 dimensions, ii) in case of welding tasks that are dislocated from the production hall, or iii) in case of one-shot tasks that need not be repeated, the human is preferred over a robot. However workers, still face challenges that have a direct impact on the final result of the process, e.g., badly executed weld seams may lead to earlier material failure. Furthermore, the distance between the welding torch and the material plays an important role in the welded product as well as the applied motion and the technique. Consequently is important to provide valuable information to the worker through, headgear-based techniques like skill level detection, workflow detection, and monitoring parameters implemented to create a smooth welding process.

Figure 6.13: The figure presents a prototypical implementation of a cognitive headgear for welding was presented in the work of [184] where the goal was to design and develop a cognitive product for industrial applications to support the operator in his daily work tasks. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

On the other hand, modern manufacturing plants are large and complex spaces with a mix of work-stations and diverse navigation paths that affect efficiency since workers must move usually within those environments as part of their tasks, to reach parts, tools, etc. A mobile assistant system can enhance navigation by identifying and recommending the most efficient paths, thereby augmenting the worker's ability to navigate the workspace effectively.

The prototype helmet consists of a head-worn device with many sensors and actuators. A computational unit and a battery are also required to support the stand-alone and mobile functionality that can perform computationally expensive tasks. The proposed prototype includes a belt where the Jetson Xavier and the battery can be mounted to reduce the weight on the head of the worker and address ergonomic concerns.

The human-centered system relies on a notification channel that puts the user in the loop and includes visual, auditive, and haptic feedback on the headgear. Tactile notifications are considered as the head is sensitive to haptic stimuli and provides faster reaction times. A centralized module decides which kind of modality will be activated based on context-based selection criteria. Visual feedback is enabled by LED arrays, which can be activated separately using a WS2801 chip. Haptic feedback is enabled by an array of vibration motors embedded into the wristband of the headgear. Auditory feedback is provided by off-the-shelf headphones using text-to-speech headphones if necessary.

This prototype of a welding helmet uses an eye-tracking sensor for skill-level and workstep detection, to ensure minimal distraction and minimal disruption to the welding process. The helmet uses algorithms to monitor the worker's state and workflow, providing haptic feedback when suboptimal routines or cognitive load are detected. The Visual feedback is considered not adequate for this task, as the workers must focus on the welding unit. For the second scenario, the navigation system functions in a digitalized environment and requires worker localization to be enabled, using an RGBD camera for environment modeling and a standard RGB camera for localization. Vibration motors and LED stripes provide visual feedback on the periphery of the field of view.

6.2.3 Smart Welding Assistance in Industrial Applications

The use case focuses on the development of an assistance system for smart arc welders in industrial manufacturing processes [161] visualized in Figure 6.14. Multiple methods exist for recognizing the important and crucial characteristics of a weld seam, however, most of them are based on image processing which would need a costly high-tech visual system. Additionally, the ambient light produced during the welding presents further challenges throughout the process in this approach. Hence, IMU sensors can fulfill the identified problems and improve the manufacturing process by assisting the user and supporting the company to stay competitive.

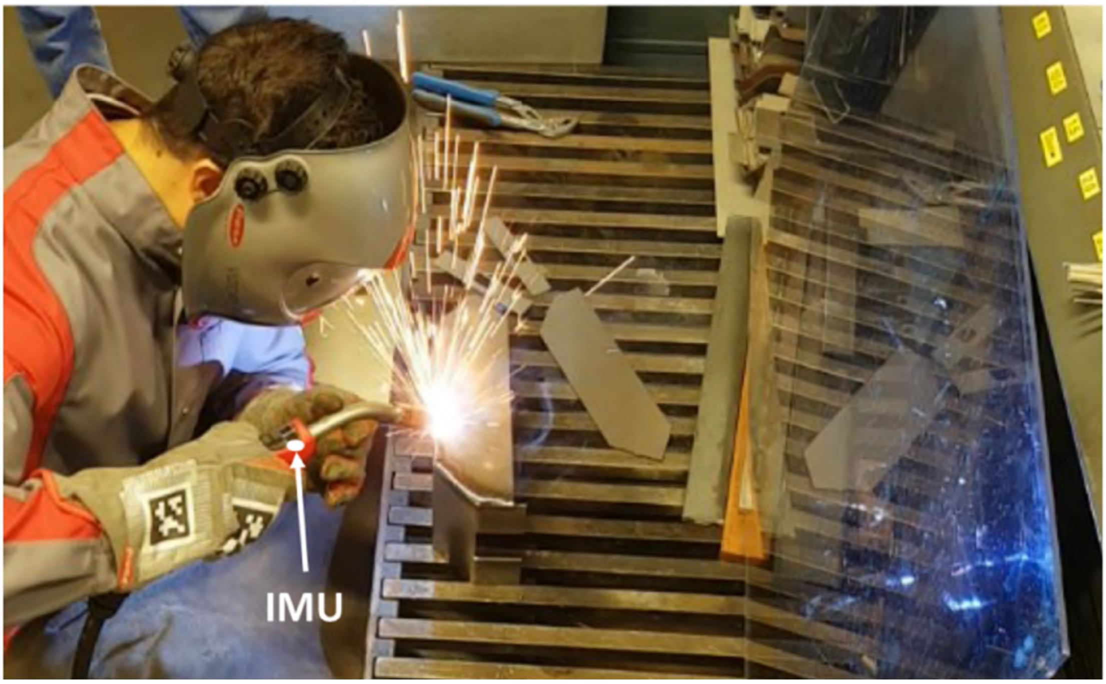


Figure 6.14: The figure presents an implementation of a smart arc welder to analyze arc welding techniques, equipped with an Inertial Measurement Unit (IMU) from the authors in [161]. The IMU tracks 3D acceleration, and rotational speed, to provide real-time motion data for skill assessment and process optimization. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

Through that system which is mounted on the arc welder and involves skill-level detection, the welder can get online feedback about the process. Visual, haptic, and auditory signals for feedback were tested for different working conditions while computed/raw features and deep learning approaches were implemented to assess the welder's skill level. Furthermore, the system would enable direct skill evaluation, provide online assistance, save extra production pipeline steps, counteract psychological stress, and save cost and effort in the training of novice workers.

6.2.4 Smart Screwing in Industrial Assembly

This use case focuses on developing a cognitive worker assistance system for industrial assembly tasks and uses the appropriate software and hardware components to provide real-time support to workers during multi-step processes. Acceleration sensors, physiological sensors, and microphones were employed to collect data in separate recording sessions, whereby each assembly process took around 40 minutes. The manufacturing process of a digging machine consists of multiple steps, one of which is to assemble the hydraulic center (meso level) manually by hand as visualized in Figure 6.15, due to the complexity of the task. A comparative study of modern machine learning approaches, including deep learning and ensemble classifiers, was conducted using raw data, handcrafted features, feature pre-processing, and raw features.

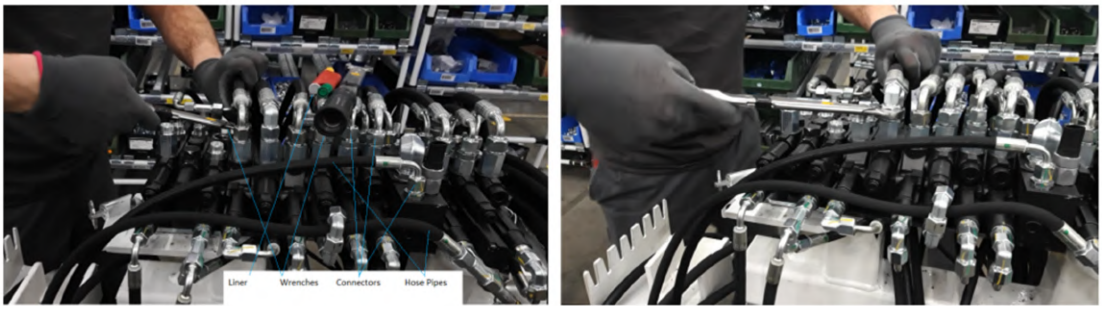


Figure 6.15: The figure shows a visualization of the hydraulic workpiece during assembly. Multiple hoses are connected in a specific order and direction and they are fastened using torque wrenches of various sizes. In the original process before the embedding of smart assistance, the operator used a marker to draw a line, indicating that the connection was completed. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

As explained in [160] during the production process multiple hoses and connectors are attached to the unit which are potentially different every time based on the received orders from customers. Consequently, the assembly workflow is not rigid since there are modifications and slight differences or individuals may execute steps in a different order. The two steps that remain the same are i) first a hose is screwed on by hand and ii) fastened with a wrench to ensure that the hose is secured with a specific torque. In Figure 6.16 the assembly steps of the hydraulic unit are visualized as recorded during the data collection at the industrial facility.

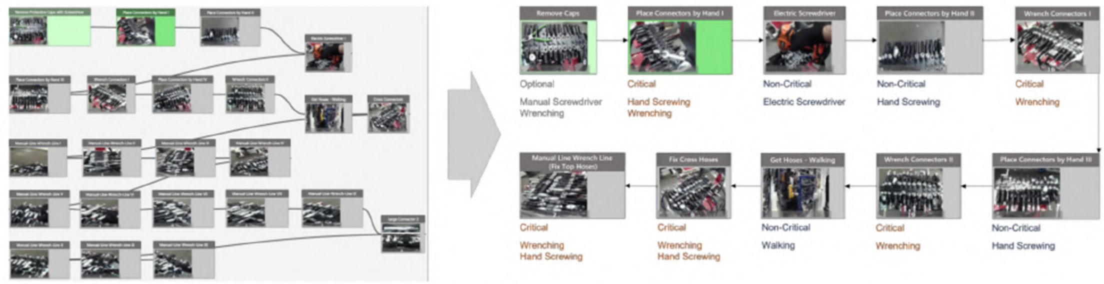


Figure 6.16: The figure shows a workflow representation of the assembly process. The original complex workflow (left) was simplified (right) for better research investigation. The critical steps of the process include wrenching actions, while non-critical and optional steps may or may not be represented in the data. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

The hydraulic unit requires the attachment of almost 50 hoses during its assembly and it is possible that workers neglect some of the steps or incorrectly execute some of them. In both situations, the mistake will be only identified at the last steps of the quality control and testing of the product or in the post-assembly (mega-level) which leads to a costly disassembly and reassembly of the product. Up until now the method that was used to identify a complete micro task was a colored line drawn in the hose and the connector. However, this approach proved to be ineffective due to the time delays it generated in the assembly process. Early estimations on this topic after consulting the company's workers and experts indicate that the elimination of the analog coloring method would lead to an increase in production efficiency by roughly 7 percent.

Real-time feedback to the workers will be used to assist workers in being aware of the current situation of their assembly workflows and reduce the time for rework. From various options that were examined, short auditory notifications through speakers for predicted errors were selected as the most appropriate for the given circumstances of the working environment. In this way, the system's complexity is kept low and additional overhead is minimized. For more detailed feedback on the process, a tablet was decided to be more suitable.

6.2.5 Depth Images for Macro Workstep Recognition

This use case highlights the use of sensor integration and real-time classification in a different domain and not in an assembly task. It demonstrates how humans can benefit in their workspace by cognitive augmentation systems while maintaining ethical considerations like privacy. The system is deployed in an industrial environment for classifying and localizing workers in real-time visualized in Figure 6.17(left), and will provide guidance and feedback to them during sequential manufacturing bending tasks. The main sensor used in this study is a depth sensor to ensure that during the tasks the worker's data remains private. The sensor is the Intel realsense and is mounted on the ceiling to monitor the worker's location and interactions. Additional information about the machine's state extracted by the bending machine was used by the deep learning models to achieve more accurate classification.

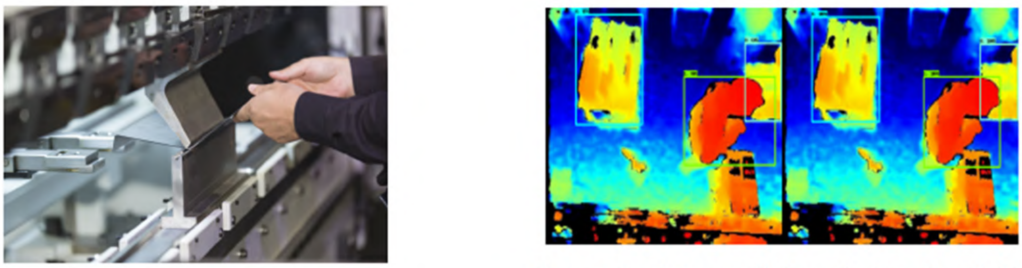


Figure 6.17: The figure visualizes a worker during a bending task in his workstation on the left. The right image visualizes the detection of the worker through the depth sensor using the presented classification model. Detecting worker work steps helps identify necessary worksteps and progress on a production line. It provides feedback to novice workers, guiding them through assembly steps. *The author of this thesis contributed to aspects of this work, although the main contributions were made by the lead authors.*

As mentioned in the related publication [164] the depth sensor locates the worker and objects in the scene, such as a computer display, a tool table, and a measuring tool table as shown in Figure 6.17(right). Areas of interest are defined based on previous objects and the bending area of the machine. Workers entering these areas, imply that they interact with an object, indicating the possible beginning of a new work step. The bending machine sends messages to the model with the current machine state or information on the computer display with a timestamp. Having all this information from both the depth sensor and the bending machine, the model classifies the worker's current work step based on the previous messages. The feedback is provided by automatically changing the display to the appropriate window for the current work step.

6.3 From Prototypes to Cognitive Augmentation

A demonstration of practical implementations reveals how the theoretical concepts regarding abstraction levels and AI design can be applied to industrial settings and improve them. These prototypes are developed within real-world manufacturing scenarios using real industrial data to ensure relevance and scalability. Those data were captured by low-cost wearables sensors (e.g., IMUs and other sensors) that are integrated into smartwatches or smart helmets to support cost-effective solutions.

The prototypes align with key research gaps, focusing on critical industrial challenges such as the i) availability of industrial datasets for developing models, ii) complex task workflow support iii) wearables and mobility in the workspace iv) low-cost assistive technologies v) deep learning approaches in industrial assembly vi) use case studies for assistive technologies vii) human error detection.

For all the previously presented use cases, real industrial data were recorded from the workers performing their tasks during their daily shifts. In this regard, datasets that capture real working conditions under a normal flow of work can be explored and provide insights that relate to the actual demands of the industry. Different types of data were collected such as IMU and camera data from ATM assembly, IMU, visual and auditory data from a central unit of a digging machine assembly, depth data from ATM assembly and a bending process as well as IMU data from a welding scenario. The IMU sensors provide accelerometer, gyroscope, and magnetometer data, while the camera data that are used in the use cases only for annotation purposes include recorded videos from the egocentric or environment's perspective. The employment of the specific type of sensors enhances the worker's autonomy and flexibility in being mobile in his workspace.

The assembly use cases concern the manufacturing of a product, therefore most of them contain data and activities that span from atomic to macro levels. These recordings focus on increasing the limited data about industrial use cases and assistive systems of this type while exploring the application of traditional machine learning and deep learning

Table 6.3: Mapping of identified research gaps to use cases. addressed (+), Partially addressed (pa, intermediate contribution), Not addressed (-).

Research Gap	Smartwatch for Activity Monitoring and Counting	Smart Helmet for Cognitive and Safety Assistance	Smart Welding Assistance: Insights and Observations	Smart Screwing in Industrial Assembly (Traditional ML Bernhard)	Depth Images for Macro Workstep Recognition
Human Activities (error)	+	+	+	+	+
Real Industrial Data	+	+	+	+	+
Wearables in Industry	+	+	-	+	-
Complex Industrial Tasks	+	+	+	+	+
Low-Cost Sensors & Assistive Technologies	+	pa	+	+	-
Mobility in Workspace	+	+	+	+	pa
Deep Learning in Production	+	pa	+	-	+
Limited Literature Assistance Systems	+	+	+	+	+
Case Studies for Assistance	+	+	+	+	+
Suitability Evaluation	pa	-	pa	-	-
Weak Labels	+	-	-	-	-

approaches. Table 6.3 presents details about the research gaps that relate to each use case.

In chapters 2 and 3, the link between cognitive overload and assembly errors was examined through a review of the literature. The findings indicate that manual assembly activities are significantly affected by high cognitive demands in dynamic environments with repetitive tasks. Furthermore, as it is stated in related work, workers who receive assistance from AI perform better, compared to their performance without assistance. Therefore, it was essential to explore strategies that can reduce or control cognitive load and also enhance task execution and decision-making.

As a result in this work, the concept of cognitive augmentation was presented to assist humans in their working environment with the use of advanced technologies such as machine learning, wearable sensors, and real-time feedback systems presented in Table

Table 6.4: Mapping of use cases to expected impact, including technical aspects.

Use Case	Sensors Used	ML/DL Models	Feedback Modality	Support to Workers in Real-Time	Task Outcome	Expected impact
Smartwatch for Classification and Counting	IMU (acc, gyr, mag), torque sensors, Depth camera	Deep learning	Visual, haptic feedback	Real-time activity detection	Activity count verification and Workstep detection	Reduce cognitive load, Reducing task duration, Improving task precision, Streamlining workflow steps
Smart Screwing Traditional ML	Accelerometer, microphone	Traditional ML (Decision Trees)	Auditory and Visual feedback	Real-time activity detection	Screwing error identification	Task precision monitoring, Streamlining workflow steps
Smart Helmet	IMU, camera sensor	Capable of DL and ML approaches	Visual, audio, haptic alerts	Safety alerts and guidance	Welding quality, navigation support	Improving task consistency, Minimizing worker fatigue
Smart Welding System	IMUs	DL (bi-dir LSTM)	Visual, audio, haptic alerts	Real-time process feedback	Welding quality monitoring	Reducing task variability, Improving task precision
Macro Workstep Classification	Depth camera, machine messages	DL (R-CNN)	Visual notifications	Workstep identification	Workflow stage progress	Streamlining workflow steps, Task analysis

6.4. Essentially the goal is to enhance cognitive abilities with systems that can unobtrusively support humans by providing personalized assistance, reducing mental effort, and mitigating errors in real time. The expected impact of cognitive augmentation extends beyond individual task execution. It promotes a safer and more efficient work environment, encourages skill development through feedback and insights, especially for novice workers, and contributes to organizational goals such as productivity, quality assurance, and workflow optimization.

With cognitive assistance, the aim is to address challenges in manual assembly and industrial workflows. One of the main advantages is the application of real-time error monitoring that detects potential mistakes and bottlenecks during tasks. Moreover, the instant feedback for correcting errors through visual, auditory, or haptic means facilitates the progression of tasks and is expected to keep stable or reduce the cognitive load. Additionally, cognitive assistance supports the tracking of activities, and counting completed tasks, to ensure the worker's consistent performance. The approaches regarding the detection could also be expanded to include worker safety by detecting hazardous conditions and providing navigational support in risky environments as already mentioned in the discussed use cases. Finally, cognitive assistance and augmentation improve training processes for novice workers, offering contextual, step-by-step instructions that can accelerate learning and skill acquisition.

6.4 Insights from Real-World Prototyping

The **prototypes of cognitive augmentation systems** in real industrial contexts provide a concrete demonstration of the proposed hierarchical activity abstraction model and system design guidance. They are built to **support workers on real workflows** and use real data for the training of their models and learning tasks. The used sensors, including IMUs, microphones, depth cameras, and torque sensors, are selected for real-time operations and fit the level of activity recognition required. The feedback that they provide also aligns with the proposed hierarchical levels and fits the demands of each industrial use case.

The smartwatch concept focuses on micro-level actions (e.g., screwing) demonstrating that micro-activity detection is feasible using IMUs in real-time. These low-cost wearable devices detect screwing activities using deep learning on raw IMU data and provide **immediate feedback** for the type and the quantity of the activity **to support error prevention** and task completion. The depth-based recognition, on the other hand, provides contextual augmentation through the identification of assembly stages that occur at meso and macro levels. The integration of these systems supports the workflow process and the assembly progression **without interrupting the operator** or being intrusive. Lastly, the smart helmet is a multi-sensory device that provides assistance and awareness in high-risk environments and its practical deployments could address different industrial contexts (screwing and welding).

The **real industrial data** that are used in all these cases help the models become more **robust** against task variability, data noise, and **unstructured real conditions**. Additionally, they confirm that cognitive augmentation systems can be built using **cost-effective technologies** such as wearables like IMUs and smartwatches that provide real-time support with minimal hardware; depth cameras and audio inputs that enable non-invasive support in collaborative spaces; and feedback systems that support both task execution (e.g., completing a screw) and situational awareness (e.g., environmental danger).

These prototypes demonstrate the **practical applicability** of the proposed theoretical framework and show that the hierarchical activity abstraction model can be matched to **appropriate sensor**, **model**, and **feedback** configurations in **industrial environments**.

7 Conclusion

Existing literature indicates a direct correlation between cognitive load and errors in manual assembly tasks. Additionally as discussed in the introduction, deep learning techniques can enhance human cognitive performance by improving environmental understanding and supporting task execution.

This dissertation explored the impact of cognitive augmentation on a worker's manual assembly tasks and its potential to enhance their performance in daily industrial processes by reducing cognitive load through workflow recognition and process awareness. The objective was to understand the differences between the individual stages of assembly processes and offer AI guiding assistance to develop systems that will be capable of improving workflow efficiency, reducing human error, and supporting decision-making through real-time feedback.

The proposed approaches are demonstrated through prototypes deployed in industrial settings following the developed taxonomy. The aim was to determine whether deep learning models can identify activities that are typically not the focus of the research community and remain undetected due to the lack of public datasets. To this end, the evaluation presents confusion matrices and graphs showing the model's ability to detect activities using raw real-world data for classification, while public datasets are used for the counting tasks, for which there was developed a dedicated model.

7.1 Insights in Cognitive Augmentation

The research presented in this thesis demonstrates the potential of cognitive augmentation in manual assembly workflows. The results of each method are presented in detail in the Abstraction Chapter (4), the Perception Awareness Chapter (5), and the Cognitive Augmentation Chapter (6).

This research contributes to improving the capabilities of AI-driven systems in industrial settings, fostering more efficient, safe, and human-responsive manufacturing processes. Consequently, a structured framework for analyzing and understanding human activities along with recommendations for AI system design, is provided to empower researchers and professionals in various industrial fields to develop more effective methods and tools for studying, modeling, and supporting human activities. –see Chapter 4.

Prior literature such as the work in [10], discusses behavioral cognitive enhancement and applications for operational domains, including real-time feedback, task monitoring, and attention support. Building on these concepts this work demonstrates how AI and sensor-based systems can augment human cognitive capabilities in manual assembly by:

(i) **enhancing task awareness** through real-time feedback and micro-activity tracking (e.g., immediate feedback, awareness at each step),

(ii) **reducing human errors** by identifying skipped or incomplete steps (e.g., missing screws),

(iii) **supporting training** and **quality assurance** by providing quantifiable feedback on activity execution (e.g., verifying the execution of required actions), and

(iv) **assisting in real-time decisions**, such as notifying the worker to repeat missed or incomplete steps.

A more detailed explanation of how these strategies align with and address each of the research questions presented earlier is provided in the corresponding chapter. Below is a summary of the key findings that reflect these strategies:

- **Hierarchical Activity Taxonomy** (addresses **RQ1**): A structured framework was developed, that categorizes industrial tasks into five levels (atomic, micro, meso, macro, and mega) to highlight the sequential nature of assembly operations. This taxonomy was derived from empirical observations of real-world industrial assembly workflows and aligns with established literature on modular and hierarchical assembly [103, 185, 192, 193]. Its aim is to support the development of reliable, robust, and suitable real-world deployment AI systems with a standardized framework, besides facilitating effective communication between experts and non-experts by breaking down tasks into manageable steps.

  Each level corresponds to distinct temporal properties, spatial scope, and activity complexity representing real-world workflows more accurately than binary or two-level activity classification systems and in a task-progressing way compared to non-binary approaches. Additionally, recognition models can be trained independently for specific levels that can be reused or extended without redesigning the full system.

Furthermore, higher-level workflows can be composed by combining lower-level classifiers, enabling upward scalability from atomic actions to complete processes.

Another key benefit is personalized cognitive support since the taxonomy allows systems to adapt their feedback and guidance strategies to a worker's expertise level. A novice worker, for example, might benefit from micro-level guidance with feedback after each screw tightening, whereas an experienced worker might only require progress indicators for the completion of macro tasks. Finally, the combination of models from different levels could enhance the interpretability and explainability of the system by identifying the number of micro activities that occurred and the macro level providing context in the workflow and supporting specific error detection and transparency.

- **AI System Design Recommendations** (addresses **RQ2**): Based on the developed taxonomy, specific guidelines were established for AI-driven activity recognition in industrial settings. The discussed categories are commonly found in the literature and guide the AI-HAR-based system design since they determine how systems perceive interpret and respond to human activities. They include sensor placement strategies, preprocessing techniques, and model selection tailored to the described abstraction levels. Wearable and non-wearable sensors, including IMUs and RGB or depth cameras, were discussed as means to capture robust data for activity recognition across different levels of task abstraction and machine learning models to classify the activities of interest. These recommendations provide a practical starting framework for designing adaptable, explainable, and context-aware human-machine collaboration systems.

To explore the taxonomy-based recommendations for the AI system design in practice, a use case was developed. The use case focused on one stage of the proposed taxonomy—the micro level—where real data and wearable sensors were used to classify assembly activities. The high importance of micro-activities as **goal-directed units in industrial assembly**, is directly linked to product quality and workflow accuracy since most of the errors occur in this stage. Their recognition can address key challenges such as **short activity duration**, inter-task similarity, and lack of **availability in public datasets**. The following findings highlight the core technological, methods, tools, and advancements achieved in this research:

- **IMU-Based Activity Recognition** (addresses **RQ3**): Developed a wrist-worn IMU system as part of a larger network to identify manufacturing activities, with early prototypes demonstrating promising performance.
- **Dataset Collection & Real-World Validation** (supports **RQ3**): A dataset was recorded using the IMU sensors to capture data from workers executing real assembly activities in an industrial setting. This dataset was used to train and evaluate the deep learning models without any extensive preprocessing techniques applied, ensuring practical applicability. The development of HAR systems under authentic conditions reflects the variability, constraints, and dynamics of actual assembly workflows, and conditions that cannot be replicated in laboratory settings.
- **Cognitive Wearable Assistance** (linked to **RQ3**): A system designed to unobtrusively support workers to complete repetitive and tedious tasks, minimize errors, and ensure smooth execution of daily work operations. Additionally, it is expected to allow workers to keep their cognitive load at low levels while reducing rework time and costs for the industry, ultimately enhancing productivity and workflow efficiency.
- **Model Comparison** (addresses **RQ3**): Different deep learning models were tested to assess their effectiveness in detecting assembly activities from the collected dataset. Their capabilities were compared in terms of accuracy and real-time applicability with a CNN architecture found to be the most effective with a recognition accuracy of 91.19%.

All the findings regarding the activity classification are presented in detail in – see Chapter 5. In addition to the classification task, this work addresses also the counting task of assembly activities, which is essential for monitoring workflow efficiency and process optimization. The methodology and results related to activity counting are discussed in Chapter 5, where an LSTM model was developed and tested against an industrial dataset. The results indicate that this method can count instances of activities when it corresponds to a single type of activity and has shown promising results when training input data contains multiple types of activities. This is supported by the findings in the aforementioned cases, which state that the error for the sequences was always within ± 1 iteration.

- **Activity Counting Method - Weak Labels** (addresses **RQ4**): During this work a method was developed to identify the end of an activity based on raw calibrated acceleration data with weak labeling. Weak labels were used to reduce the expense of data annotation when fully labeled datasets were not available.
- **Variable-Length Data Sequences** (addresses **RQ4**): Instead of using a fixed window size, the generated sequences have variable sizes, allowing the system to adapt and capture activities of different durations.
- **LSTM-Based Regression Model** (addresses **RQ4**): An LSTM model for regression analysis was implemented with the counting ability embedded within the training loop for the learning task. It was evaluated by counting different classes of the Skoda dataset for human activity recognition (HAR).

Building on the structured taxonomy for human activities, this work explored AI-driven activity classification and counting to enhance workflow understanding. These approaches were applied to prototypes that were developed and implemented in industrial settings. The aim was to assess the feasibility and applicability of real-time activity recognition and cognitive augmentation in practical applications using only real captured data for each specific use case.

- **Demonstration in Industrial Settings** (integrates **RQ1–RQ4**): The feasibility of AI-driven activity recognition was demonstrated through five use cases in real-world industrial environments. The systems utilized various sensor modalities such as wearable and non-wearable IMUs, RGB cameras for activity sensing, and smartwatches, sound devices, and monitors as feedback actuators. While the prototypes were successfully implemented and tested, formal empirical validation of their impact on cognitive load and workflow efficiency remains an area for future research.

Another topic that is relevant and important to address is the scalability of the discussed approaches in this work. The presented work draws inspiration from real workflows, worker behavior, and the task complexity that is observed in production environments. While the systems and prototypes that are developed here are validated through specific use cases, the framework and the tools that are recommended, offer broader opportunities for applicability in other use cases. Those may require domain-specific adjustments and adaptation; however, the fundamental principles remain consistent.

The scalability of this work is supported by several choices in its design. First, the hierarchical taxonomy of activity abstraction levels guides the implementation of systems at varying levels of complexity based on the demands of the task and context. The use of low-cost devices that are commercially available facilitates that the sensor setup can be reproduced in other environments, assisting the human without major infrastructural changes. The deep learning models also follow the main principles using annotated data and supervised approaches or rely on variable-length sequences and weakly labeled data that enable their use in different settings. Similarly, the real-time feedback recommendations, offer personalized support and can be adapted to other cases that the human is at the center of the process.

While the focus of this thesis remains on the industrial assembly processes, the architectural principles, wearable sensor setup, and model design are transferable to other domains that require human-AI collaboration. Although the proposed hierarchical activity abstraction framework is developed for assembly use cases, the conceptual foundation and the AI design that it proposes may offer useful recommendations to support humans in other highly complex environments and tasks e.g., in logistics, maintenance, or healthcare.

7.1.1 Challenges

This section examines the challenges and limitations encountered during the development and implementation of this work about cognitive augmentation in manual assembly. These limitations concern the theoretical and practical constraints, hardware and software components, data processing efficiency, real-world implementation factors, and methodological considerations that may influence the system's effectiveness.

The taxonomy at the moment constitutes an initial method to identify more detailed differences in various stages of an assembly compared to the existing approaches. Additionally, it provides guidance, tools, and models to support the development of strategies for efficient task execution and cognitive augmentation. However, the variability of daily work tasks requires a high level of attention to detail, highlighting the need for adaptable and precise classification methods. This was an important reason to capture and use real data in this work and not rely on datasets that might simulate or include activities similar to assembly.

The existing literature and the development of taxonomy showed the way for the optimal selection and the optimal position of the sensors also based on the requirements of each use case. However, multiple issues regarding the synchronization of data due to hardware misalignment had to be addressed. Moreover, many times environmental factors such as variations in lighting, occlusion, and noise affected sensor data quality. Throughout this work, all the studies prioritized less invasive and privacy-friendly sensors to ensure compliance with ethical considerations and worker acceptance in system development.

Beyond the data acquisition, the challenges extended to data annotation for the classification task which was important for training and validating the models later on. The data imbalance and the inter-class similarity as a result of the real data collection was a very significant challenge that had to be addressed. The participants performed their daily assembly tasks without any specific order or instruction from the studies' creators, and this variability was reflected in the dataset. Furthermore, the activities of assembly were relatively limited compared to the numerous other tasks during a worker's shift. This also contributed to the data imbalance. To mitigate this, various strategies were employed, including data augmentation techniques to artificially expand the dataset and additional data recordings that were important to ensure a more balanced distribution.

Initially, the counting method was also tested with the collected dataset but the results were inconsistent due to the imbalance and the data variability. Consequently, to validate the LSTM approach with the variable size and the weak labels, a public dataset related to industrial tasks was selected. Additionally, a multitask approach similar to the one presented in [159] was tested for combining the classification and counting however it provided no significant improvement in the counting results. Furthermore, the CNN architecture that was used for the classification depends on a supervised approach with fixed data windows that require padding or other methods to work in combination with the flexible "windows" of the LSTM model. Apart from that the systems requirements in power consumption and overall costs must be considered when integrating systems into industrial processes. Moreover, real-time adaptation, worker acceptance, and the transferability of the methods to other domains remain critical aspects for ensuring practical implementation and broader applicability.

My work focused on the development and evaluation of methods for recognizing, detecting, classifying, and counting assembly activities, as well as exploring their potential for cognitive augmentation in industrial settings. However, while this work provides a solid basis for method development and evaluation, it does not extend to the full evaluation of industrial product deployment. This is also one of the difficulties of the use cases that are realized in non-laboratory settings where limited access to real industrial facilities often restricts long-term testing and comprehensive evaluation. Instead, the focus remained on implementing and analyzing individual systems and assessing their applicability and effectiveness rather than conducting large-scale evaluations within operational environments. Subsequently, the purpose was not to create complete end-user commercial products but rather to address key challenges in industrial assembly and propose solutions that support workers, enhance their capabilities, and reduce their cognitive load. The evaluation conducted in this work focuses on the performance of the developed models and methods, demonstrating their potential and effectiveness achieve cognitive augmentation for manual assembly.

7.1.2 Implications in Research – Industrial Environments

This work contributes to the development of an unobtrusive mobile and flexible assistive system to digitalize manufacturing assembly operations and enhance cognitive augmentation by offering guidance to the operators. It focuses on providing real-time feedback to workers on completed or incomplete worksteps within industrial processes and raising awareness of missing activities. In this respect, errors that occur during these processes can be detected while they happen by wearables and enable immediate corrective actions. The contributions of this thesis have many implications for the research community as well as for practitioners in the industrial field across various work levels.

The advancement in the activity recognition area is supported by the provided structured taxonomy framework that introduces the atomic-mega levels and can be extended to other tool-based manual industrial tasks or even broader manufacturing tasks. The recognition of those tasks relies on the descriptive dataset that was collected for the testing and evaluation of the models. The data that was used demonstrates the possibility of recognizing and classifying micro-level activities with IMU sensors as a more privacy-compliant, energy-efficient, and cost-effective alternative in terms of setup and operation.

Moreover, the variable-size windows that were used, provide the option to support and explore alternative methods for segmenting time series data while preserving the important information. When they are combined with the weak labels, this approach can reduce annotation time, which is a resource-intensive task. However, it is still important to focus on further exploration into real-time AI deployment, ethical considerations in worker monitoring, and the development of more balanced industrial datasets.

In the recent past, the IBM Institute for Business Value, in collaboration with Oxford Economics surveyed and interviewed 550 technology and operations executives to understand how machine learning is beginning to transform the way businesses organize their operations and benefit from technology investments [260]. Their findings show that intelligent automation augments employees' skills and expertise, increasing productivity and creativity, but it is not a plug-and-play solution. They highlight that successful implementation requires human involvement and adaptation. However, they also underscore the most prominent challenges in AI adoption, including a lack of skills and resources, difficulties in aligning strategies, and trust issues with automated decision-making.

From the industrial perspective, this work provides the standardization of assembly tasks and offers real-time tracking of the progress. With respect to their work, priority aspects, for every industrial practitioner include productivity and cost efficiency. The proposed methods suggest cognitive augmentation of the workers, that supports the reduction of manual assembly errors, and directly impacts the the product quality and the cost of reworks. Following that, it can lead to shorter cycle times and improve production speed and efficiency, while the system can be integrated with broader IoT tools or other manufacturing execution systems to help streamline production scheduling and workflow coordination. This could also influence the use of equipment and materials in different industries and in combination with predictive analytics models, optimize resource allocation for improved performance. However, it is essential to ensure that the proposed systems for cognitive augmentation are used for positive and ethical purposes with a focus on the acceptance, empowerment, and satisfaction of workers through AI, and address the concerns and the fear of replacing them. Finally, transparency, fairness, and responsible AI deployment should remain at the core of technological advancements in industrial settings.

7.2 Future Work

This research lays the foundation for several directions in future work. Building on the findings of this thesis, further studies can explore the presented topics or similar topics in related domains to enhance and expand the scope of this work.

Cognitive augmentation or assistance is important as it was demonstrated through this work for the workers in manual assembly tasks since humans remain a vital component of the process. While parts of the proposed approaches for cognitive augmentation systems were implemented and tested in an industrial setting, their impact on cognitive load and workflow efficiency was not empirically validated. Instead, this research focused on establishing the feasibility of complete AI-driven activity recognition systems, with future work needed to conduct evaluation experiments assessing its practical benefits. For instance, Younas et al. [261], developed a sensor-based learning analytics platform to classify cognitive abilities by analyzing behavioral differences between experts and novices while executing hand-written answers, mathematical calculations, and complex representations such as drawing figures and plots. Similar and adapted approaches could be employed in industrial cases, to assess the impact on cognitive load and to evaluate the practical feasibility and effectiveness of AI-assisted cognitive augmentation systems.

The landscape of human interaction with advanced technologies beyond the manufacturing sector should also be explored through the findings of this work. Addressing the complexity of activities within HAR while extending hierarchical analysis to diverse domains would be a priority to create a more general framework that could include multiple activities. Moreover, these categorizations of activities could support the research and development of activity datasets and models that specifically address the complexities of each level in the taxonomy and enlarge the database to extend the research in the field. Additionally, it is mandatory to recognize the importance of safety within industrial environments [262, 263], while improving the system's interoperability and the user's experience, making assembly activity recognition systems more human-responsive. This means creating systems that understand and respond to human needs and behaviors better. One example would be to develop technologies that detect when a worker needs help, is feeling stressed, or is feeling fatigued, and assists or adjusts the workload accordingly.

As industries increasingly demand flexible, adaptive, and personalized mechanisms, the methods that are proposed in this thesis offer a promising direction for sensor-based training support and skill enhancement to individual workers. Through the developed hierarchical taxonomy, which provides a decomposition of tasks, novices can be trained, following progressive learning of tasks in different levels of industrial processes. Moreover, the demonstrated real-time activity recognition and counting capabilities, create possibilities for future applications such as performance monitoring, skill assessment, and feedback-based learning, contributing to improved workforce development in dynamic industrial environments.

Furthermore, merging AI with technologies such as edge computing, IoT, and cloud computing can improve data analysis in real-time and decision-making based on the results. Additionally, optimizing the placement of the sensors can improve the accuracy of the model and reduce the need for unnecessary sensor data, which can be an obstacle in the learning process. Initial contributions have been made in [264], where a two-stage semantic-aware knowledge distillation (KD) method was proposed to optimize sensor modalities and model size while maintaining high recognition performance. Besides that, it is interesting to investigate the potential application of the proposed classification and counting method to other time-series data. However, it is important to consider individual differences in movement patterns that may be influenced by factors such as body size, gender, and age. One approach is to use a diverse dataset of individuals with varying body sizes, genders, and ages to train a deep-learning model that can generalize to new individuals and accurately estimate their movements.

Another important direction is to explore and ensure the safe use and effective integration of AI in the industrial setting but also in other domains. Ethical and safety guidelines must be considered and regulation frameworks to comply with AI standards for deployment as emphasized by the authors in [265, 266]. It is therefore important to investigate how individuals interact with assistive technology (e.g., smart products, collaborative robots) for assembly tasks. Investigation of this socio-technical aspect is already present, e.g., in [267, 268]. Expanding on this, future studies could explore the aspects of acceptance and adoption of those systems which highly correlates to the understanding and explainability of them as discussed by the authors in [269, 270]. Furthermore, it can be explored whether AI explanations have an impact on user trust, comprehension, and performance in industrial settings. By exploring these insights in future research and development, we can advance the capabilities of assembly activity recognition systems, improve communication between humans and machines, promote teamwork, and facilitate more efficient and effective operations in industrial environments, while manufacturing processes can become more efficient.

List of Figures

List of Tables

Bibliography

[1] L Minh Dang et al. "Sensor-based and vision-based human activity recognition: A comprehensive survey". In: *Pattern Recognition* 108 (2020), p. 107561 (cit. on pp. 2, 30, 33).

[2] Christian Engel, Philipp Ebel, and Jan Marco Leimeister. "Cognitive automation". In: *Electronic Markets* 32.1 (2022), pp. 339–350 (cit. on p. 2).

[3] Hui Liu, Hugo Gamboa, and Tanja Schultz. *Sensor-Based Human Activity and Behavior Research: Where Advanced Sensing and Recognition Technologies Meet*. 2023 (cit. on pp. 3, 23).

[4] Moshe Chai Barukh et al. "Cognitive augmentation in processes". In: *Next-Gen Digital Services. A Retrospective and Roadmap for Service Computing of the Future: Essays Dedicated to Michael Papazoglou on the Occasion of His 65th Birthday and His Retirement* (2021), pp. 123–137 (cit. on p. 3).

[5] Hoda ElMaraghy and Waguih ElMaraghy. "Adaptive cognitive manufacturing system (ACMS)–a new paradigm". In: *International Journal of Production Research* 60.24 (2022), pp. 7436–7449 (cit. on pp. 3, 19, 41, 42).

[6] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. "When combinations of humans and AI are useful: A systematic review and meta-analysis". In: *Nature Human Behaviour* (Oct. 2024), pp. 1–11. DOI: 10.1038/s41562-024-02024-1 (cit. on pp. 3, 4, 40).

[7] Ron Fulbright and Miranda Morrison. "Does Using ChatGPT Result in Human Cognitive Augmentation?" In: *International Conference on Human-Computer Interaction*. Springer. 2024, pp. 133–146 (cit. on p. 4).

[8] Mangolika Bhattacharya et al. "Human-in-Loop: A Review of Smart Manufacturing Deployments". In: *Systems* 11 (Jan. 2023), p. 35. DOI: 10.3390/systems11010035 (cit. on pp. 4, 5).

[9] Caterina Cinel, Davide Valeriani, and Riccardo Poli. "Neurotechnologies for human cognitive augmentation: current state of the art and future prospects". In: *Frontiers in human neuroscience* 13 (2019), p. 13 (cit. on p. 4).

[10] Alexandre Marois and Daniel Lafond. "Augmenting cognitive work: a review of cognitive enhancement methods and applications for operational domains". In: *Cognition, Technology & Work* 24.4 (2022), pp. 589–608 (cit. on pp. 4, 19, 149).

[11] Behrooz Azadi et al. "Alpine skiing activity recognition using smartphone's IMUs". In: *Sensors* 22.15 (2022), p. 5922 (cit. on pp. 5, 23, 31).

[12] Manju Rana and Vikas Mittal. "Wearable sensors for real-time kinematics analysis in sports: a review". In: *IEEE Sensors Journal* 21.2 (2020), pp. 1187–1207 (cit. on pp. 5, 23).

[13] Sen Qiu et al. "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges". In: *Information Fusion* 80 (2022), pp. 241–265 (cit. on pp. 5, 23, 33).

[14] Alessandro Filippeschi et al. "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion". In: *Sensors* 17.6 (2017), p. 1257 (cit. on pp. 5, 23).

[15] Benedikt G Mark, Erwin Rauch, and Dominik T Matt. "Worker assistance systems in manufacturing: A review of the state of the art and future directions". In: *Journal of Manufacturing Systems* 59 (2021), pp. 228–250 (cit. on pp. 5, 9, 19, 23, 61).

[16] Tara Capel and Margot Brereton. "What is Human-Centered about Human-Centered AI? A Map of the Research Landscape". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3580959. URL: https://doi.org/10.1145/3544548.3580959 (cit. on pp. 6, 7).

[17] Joerg Franke et al. "Electronic module assembly". In: *CIRP annals* 70.2 (2021), pp. 471–493 (cit. on p. 7).

[18] Muaaz Abdul Hadi et al. "Towards flexible and cognitive production—addressing the production challenges". In: *Applied Sciences* 12.17 (2022), p. 8696 (cit. on pp. 8, 21, 44, 124).

[19] Ann-Christine Falck et al. "Criteria for assessment of basic manual assembly complexity". In: *Procedia CIRP* 44 (2016), pp. 424–428 (cit. on p. 8).

[20] Fahad M. Alqahtani et al. "A New Model for Determining Factors Affecting Human Errors in Manual Assembly Processes Using Fuzzy Delphi and DEMATEL Methods". In: *Symmetry* 15.11 (2023). ISSN: 2073-8994. DOI: 10.3390/sym15111967. URL: https://www.mdpi.com/2073-8994/15/11/1967 (cit. on pp. 8, 9, 37, 40).

[21] Ahmet Kolus, Richard Wells, and Patrick Neumann. "Production quality and human factors engineering: A systematic review and theoretical framework". In: *Applied ergonomics* 73 (2018), pp. 55–89 (cit. on p. 9).

[22] Yaniel Torres, Sylvie Nadeau, and Kurt Landau. "Classification and quantification of human error in manufacturing: A case study in complex manual assembly". In: *Applied Sciences* 11.2 (2021), p. 749 (cit. on p. 9).

[23] James Male and Uriel Martinez-Hernandez. "Recognition of human activity and the state of an assembly task using vision and inertial sensor fusion methods". In: *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*. Vol. 1. IEEE. 2021, pp. 919–924 (cit. on p. 9).

[24] Jéssica de Assis Dornelles, Néstor F Ayala, and Alejandro G Frank. "Smart Working in Industry 4.0: How digital technologies enhance manufacturing workers' activities". In: *Computers & Industrial Engineering* 163 (2022), p. 107804 (cit. on p. 9).

[25] Adriana Ventura Carvalho et al. "Cognitive manufacturing in industry 4.0 toward cognitive load reduction: A conceptual framework". In: *Applied System Innovation* 3.4 (2020), p. 55 (cit. on pp. 16, 17).

[26] Fred Paas, Alexander Renkl, and John Sweller. "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture". In: *Instructional science* 32.1/2 (2004), pp. 1–8 (cit. on p. 17).

[27] Cecilio Angulo, Alejandro Chacón, and Pere Ponsa. "Towards a cognitive assistant supporting human operators in the Artificial Intelligence of Things". In: *Internet of Things* 21 (2023), p. 100673 (cit. on pp. 17, 20, 39, 40).

[28] Michael Haslgrübler, Benedikt Gollan, and Alois Ferscha. "A cognitive assistance framework for supporting human workers in industrial tasks". In: *IT Professional* 20.5 (2018), pp. 48–56 (cit. on pp. 17, 21).

[29] Marta Lagomarsino et al. "An online framework for cognitive load assessment in industrial tasks". In: *Robotics and Computer-Integrated Manufacturing* 78 (2022), p. 102380 (cit. on p. 17).

[30] Francesco N Biondi et al. “Overloaded and at work: investigating the effect of cognitive workload on assembly task performance”. In: *Human factors* 63.5 (2021), pp. 813–820 (cit. on p. 17).

[31] Cecilia Berlin et al. “A Systemic Overview of Factors Affecting the Cognitive Performance of Industrial Manual Assembly Workers”. In: May 2021, pp. 371–381. ISBN: 978-3-030-74607-0. DOI: 10.1007/978-3-030-74608-7_47 (cit. on p. 17).

[32] Sukhpal Singh Gill et al. “Modern computing: Vision and challenges”. In: *Telematics and Informatics Reports* (2024), p. 100116 (cit. on p. 18).

[33] Dominic Lindsay et al. “The evolution of distributed computing systems: from fundamental to new frontiers”. In: *Computing* 103.8 (2021), pp. 1859–1878 (cit. on p. 18).

[34] Ryozo Yamashita. “History of personal computers in Japan”. In: *International Journal of Parallel, Emergent and Distributed Systems* 35.2 (2020), pp. 143–169 (cit. on p. 18).

[35] Sukhpal Singh Gill et al. “AI for next generation computing: Emerging trends and future directions”. In: *Internet of Things* 19 (2022), p. 100514 (cit. on p. 18).

[36] Jayavardhana Gubbi et al. “Internet of Things (IoT): A vision, architectural elements, and future directions”. In: *Future generation computer systems* 29.7 (2013), pp. 1645–1660 (cit. on p. 18).

[37] Mahadev Satyanarayanan. “Pervasive computing: Vision and challenges”. In: *IEEE Personal communications* 8.4 (2001), pp. 10–17 (cit. on p. 18).

[38] Matthew T West. “Ubiquitous computing”. In: *Proceedings of the 39th annual ACM SIGUCCS conference on User services*. 2011, pp. 175–182 (cit. on p. 18).

[39] Alois Ferscha. “Pervasive computing”. In: *Hagenberg Research*. Springer, 2009, pp. 379–431 (cit. on p. 18).

[40] Debashis Saha et al. “Pervasive computing”. In: *Networking Infrastructure for Pervasive Computing: Enabling Technologies and Systems* (2003), pp. 1–37 (cit. on p. 18).

[41] Emna Baccour et al. “Pervasive AI for IoT applications: A survey on resource-efficient distributed artificial intelligence”. In: *IEEE Communications Surveys & Tutorials* 24.4 (2022), pp. 2366–2418 (cit. on p. 18).

[42] Clément Belletier et al. “Wearable cognitive assistants in a factory setting: a critical review of a promising way of enhancing cognitive performance and well-being”. In: *Cognition, Technology & Work* 23 (2021), pp. 103–116 (cit. on p. 19).

[43] Kincsö Izsak, Jari Romanainen, and Blagoy Stamenov. *Advanced Technologies for Industry: Policy Brief: Industrial Recovery and Technology Policy*. Publications Office of the European Union, 2021 (cit. on p. 19).

[44] Farouk Belkadi et al. "Intelligent assistant system as a context-aware decision-making support for the workers of the future". In: *Computers & Industrial Engineering* 139 (2020), p. 105732 (cit. on p. 19).

[45] Niels Van Berkel, Mikael B Skov, and Jesper Kjeldskov. "Human-AI interaction: intermittent, continuous, and proactive". In: *Interactions* 28.6 (2021), pp. 67–71 (cit. on p. 20).

[46] Benedikt G Mark, Erwin Rauch, and Dominik T Matt. "Systematic selection methodology for worker assistance systems in manufacturing". In: *Computers & Industrial Engineering* 166 (2022), p. 107982 (cit. on p. 20).

[47] Benedikt Gregor Mark et al. "Analysis of user groups for assistance systems in production 4.0". In: *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE. 2019, pp. 1260–1264 (cit. on pp. 20, 21).

[48] Bastian Pokorni, Daniela Popescu, and Carmen Constantinescu. "Design of cognitive assistance systems in manual assembly based on quality function deployment". In: *Applied Sciences* 12.8 (2022), p. 3887 (cit. on p. 21).

[49] Akin Avci et al. "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey". In: *23th International conference on architecture of computing systems 2010*. VDE. 2010, pp. 1–10 (cit. on p. 22).

[50] Xinyu Li, Yuan He, and Xiaojun Jing. "A survey of deep learning-based human activity recognition in radar". In: *Remote Sensing* 11.9 (2019), p. 1068 (cit. on p. 22).

[51] Rene Grzeszick et al. "Deep neural network based human activity recognition for the order picking process". In: *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*. 2017, pp. 1–6 (cit. on p. 22).

[52] Shoya Ishimaru et al. "Towards reading trackers in the wild: detecting reading activities by EOG glasses and deep neural networks". In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 2017, pp. 704–711 (cit. on p. 22).

[53] Sizhen Bian et al. "Using Human Body Capacitance Sensing to Monitor Leg Motion Dominated Activities with a Wrist Worn Device". In: *Sensor- and Video-Based Activity and Behavior Computing*. Ed. by Md Atiqur Rahman Ahad et al. Singapore: Springer Nature Singapore, 2022, pp. 81–94. ISBN: 978-981-19-0361-8 (cit. on p. 22).

[54] Sungho Suh et al. "Worker Activity Recognition in Manufacturing Line Using Near-Body Electric Field". In: *IEEE Internet of Things Journal* 11.7 (2024), pp. 11554–11565. DOI: `10.1109/JIOT.2023.3330372` (cit. on p. 22).

[55] Mengxi Liu et al. "Smart-Badge: A wearable badge with multi-modal sensors for kitchen activity recognition". In: *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 2022, pp. 356–363 (cit. on p. 22).

[56] Yan Wang, Shuang Cang, and Hongnian Yu. "A survey on wearable sensor modality centred human activity recognition in health care". In: *Expert Systems with Applications* 137 (2019), pp. 167–190 (cit. on pp. 22, 23).

[57] Sizhen Bian et al. "The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey". In: *Sensors* 22.12 (2022). ISSN: 1424-8220. DOI: `10.3390/s22124596`. URL: `https://www.mdpi.com/1424-8220/22/12/4596` (cit. on p. 22).

[58] Andreas Bulling, Ulf Blanke, and Bernt Schiele. "A tutorial on human activity recognition using body-worn inertial sensors". In: *ACM Computing Surveys (CSUR)* 46.3 (2014), pp. 1–33 (cit. on pp. 22, 100).

[59] Md Zahangir Alom et al. "A state-of-the-art survey on deep learning theory and architectures". In: *Electronics* 8.3 (2019), p. 292 (cit. on p. 22).

[60] Kaixuan Chen et al. "Deep learning for sensor-based human activity recognition: overview, challenges and opportunities". In: *arXiv preprint arXiv:2001.07416* (2020) (cit. on pp. 22, 84).

[61] Florenc Demrozi et al. "Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey". In: *IEEE access* 8 (2020), pp. 210816–210836 (cit. on p. 23).

[62] Hui Liu and Tanja Schultz. "How Long Are Various Types of Daily Activities? Statistical Analysis of a Multimodal Wearable Sensor-based Human Activity Dataset." In: *HEALTHINF*. 2022, pp. 680–688 (cit. on p. 23).

[63] Jindong Wang et al. “Deep learning for sensor-based activity recognition: A survey”. In: *Pattern Recognition Letters* 119 (2019), pp. 3–11 (cit. on pp. 23, 76).

[64] Yu Zhao et al. “Deep residual bidir-LSTM for human activity recognition using wearable sensors”. In: *Mathematical Problems in Engineering* 2018 (2018) (cit. on pp. 23, 82).

[65] Emilio Sansano, Raúl Montoliu, and Óscar Belmonte Fernández. “A study of deep neural networks for human activity recognition”. In: *Computational Intelligence* 36.3 (2020), pp. 1113–1139 (cit. on p. 23).

[66] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. “Deep, convolutional, and recurrent models for human activity recognition using wearables”. In: *arXiv preprint arXiv:1604.08880* (2016) (cit. on p. 23).

[67] Shrehal Bohra, Vinayak Naik, and Varun Yeligar. “Towards Putting Deep Learning on the Wrist for Accurate Human Activity Recognition”. In: *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 2021, pp. 605–610. DOI: `10.1109/PerComWorkshops51409.2021.9430979` (cit. on p. 23).

[68] Alejandro Guinea, Mehran Sarabchian, and Max Mühlhäuser. “Image-based Activity Recognition from IMU Data”. In: Mar. 2021, pp. 14–19. DOI: `10.1109/PerComWorkshops51409.2021.9430990` (cit. on p. 24).

[69] Bulat Khaertdinov, Esam Ghaleb, and Stylianos Asteriadis. “Deep Triplet Networks with Attention for Sensor-based Human Activity Recognition”. In: *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2021, pp. 1–10. DOI: `10.1109/PERCOM50583.2021.9439116` (cit. on p. 24).

[70] Vishwanath Bijalwan, Vijay Bhaskar Semwal, and Vishal Gupta. “Wearable sensor-based pattern mining for human activity recognition: Deep learning approach”. In: *Industrial Robot: the international journal of robotics research and application* 49.1 (2022), pp. 21–33 (cit. on p. 24).

[71] Saurabh Gupta. “Deep learning based human activity recognition (HAR) using wearable sensor data”. In: *International Journal of Information Management Data Insights* 1.2 (2021), p. 100046 (cit. on p. 24).

[72] Wenchao Jiang and Zhaozheng Yin. “Human activity recognition using wearable sensors by deep convolutional neural networks”. In: *Proceedings of the 23rd ACM international conference on Multimedia*. 2015, pp. 1307–1310 (cit. on p. 24).

[73] Davide Anguita et al. “A public domain dataset for human activity recognition using smartphones.” In: *Esann*. Vol. 3. 2013, p. 3 (cit. on p. 24).

[74] Mi Zhang and Alexander A Sawchuk. “USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors”. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 2012, pp. 1036–1043 (cit. on p. 24).

[75] Muhammad Shoaib et al. “Fusion of smartphone motion sensors for physical activity recognition”. In: *Sensors* 14.6 (2014), pp. 10146–10176 (cit. on p. 24).

[76] Jianbo Yang et al. “Deep convolutional neural networks on multichannel time series for human activity recognition.” In: *Ijcai*. Vol. 15. Buenos Aires, Argentina. 2015, pp. 3995–4001 (cit. on p. 24).

[77] Nitin Nair, Chinchu Thomas, and Dinesh Babu Jayagopi. “Human activity recognition using temporal convolutional network”. In: *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*. 2018, pp. 1–8 (cit. on p. 24).

[78] Francisco Javier Ordóñez and Daniel Roggen. “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition”. In: *Sensors* 16.1 (2016), p. 115 (cit. on pp. 24, 43).

[79] Fernando Moya Rueda et al. “Convolutional neural networks for human activity recognition using body-worn sensors”. In: *Informatics*. Vol. 5. 2. Multidisciplinary Digital Publishing Institute. 2018, p. 26 (cit. on p. 24).

[80] Sungho Suh, Vitor Fortes Rey, and Paul Lukowicz. “Tasked: transformer-based adversarial learning for human activity recognition using wearable sensors via self-knowledge distillation”. In: *Knowledge-Based Systems* 260 (2023), p. 110143 (cit. on p. 24).

[81] Guillermo Diaz et al. “Cross-Domain Human Activity Recognition Using Low-Resolution Infrared Sensors”. In: *Sensors* 24.19 (2024). ISSN: 1424-8220. DOI: `10.3390/s24196388`. URL: `https://www.mdpi.com/1424-8220/24/19/6388` (cit. on p. 24).

[82] Mengting Fang et al. "Can a recurrent neural network learn to count things?" In: *CogSci*. 2018 (cit. on pp. 25, 103).

[83] Yunhua Zhang, Ling Shao, and Cees GM Snoek. "Repetitive activity counting by sight and sound". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 14070–14079 (cit. on p. 25).

[84] David Strömbäck, Sangxia Huang, and Valentin Radu. "Mm-fit: Multimodal deep learning for automatic exercise logging across sensing devices". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.4 (2020), pp. 1–22 (cit. on p. 25).

[85] Kacper Skawinski et al. "Workout type recognition and repetition counting with CNNs from 3D acceleration sensed on the chest". In: *International Work-Conference on Artificial Neural Networks*. Springer. 2019, pp. 347–359 (cit. on p. 25).

[86] Bruno Ferreira et al. "Deep learning approaches for workout repetition counting and validation". In: *Pattern Recognition Letters* 151 (2021), pp. 259–266 (cit. on p. 25).

[87] Sizhen Bian et al. "Passive capacitive based approach for full body gym workout recognition and counting". In: *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE. 2019, pp. 1–10 (cit. on p. 25).

[88] Andrea Soro et al. "Recognition and repetition counting for complex physical exercises with deep learning". In: *Sensors* 19.3 (2019), p. 714 (cit. on p. 26).

[89] Kun Wang, Jun He, and Lei Zhang. "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors". In: *IEEE Sensors Journal* 19.17 (2019), pp. 7598–7604 (cit. on p. 26).

[90] Wenbin Gao et al. "DanHAR: Dual attention network for multimodal human activity recognition using wearable sensors". In: *Applied Soft Computing* 111 (2021), p. 107728 (cit. on p. 26).

[91] Chenguang Shen, Bo-Jhang Ho, and Mani Srivastava. "Milift: Efficient smartwatch-based workout tracking using automatic segmentation". In: *IEEE Transactions on Mobile Computing* 17.7 (2017), pp. 1609–1622 (cit. on p. 26).

[92] Ghanashyama Prabhu, Noel E O'connor, and Kieran Moran. "Recognition and repetition counting for local muscular endurance exercises in exercise-based rehabilitation: A comparative study using artificial intelligence models". In: *Sensors* 20.17 (2020), p. 4791 (cit. on p. 27).

[93] Juri Taborri et al. “Automatic identification and counting of repetitive actions related to an industrial worker”. In: *2019 II Workshop on Metrology for Industry 4.0 and IoT (MetroInd4. 0&IoT)*. IEEE. 2019, pp. 394–399 (cit. on p. 27).

[94] Shun Ishii et al. “Exersense: real-time physical exercise segmentation, classification, and counting algorithm using an imu sensor”. In: *Activity and Behavior Computing*. Springer, 2021, pp. 239–255 (cit. on p. 27).

[95] Yuuki Nishino, Takuya Maekawa, and Takahiro Hara. “Few-Shot and Weakly Supervised Repetition Counting With Body-Worn Accelerometers”. In: *Frontiers in Computer Science* 4 (2022) (cit. on p. 27).

[96] Thomas Stiefmeier et al. “Wearable activity tracking in car manufacturing”. In: *IEEE Pervasive Computing* 7.2 (2008), pp. 42–50 (cit. on pp. 27, 76).

[97] JuHyeong Ryu et al. “Automated action recognition using an accelerometer-embedded wristband-type activity tracker”. In: *Journal of construction engineering and management* 145.1 (2019), p. 04018114 (cit. on p. 27).

[98] Jonatan Fridolfsson et al. “Workplace activity classification from shoe-based movement sensors”. In: *BMC Biomedical Engineering* 2.1 (2020), pp. 1–8 (cit. on p. 28).

[99] Takuya Maekawa et al. “Toward practical factory activity recognition: unsupervised understanding of repetitive assembly work in a factory”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2016, pp. 1088–1099 (cit. on p. 28).

[100] Behrooz Azadi et al. “Feasibility analysis of unsupervised industrial activity recognition based on a frequent micro action”. In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2019, pp. 368–375 (cit. on pp. 28, 32, 45, 58).

[101] Wenjin Tao, Ming C Leu, and Zhaozheng Yin. “Multi-modal recognition of worker activity for human-centered intelligent manufacturing”. In: *Engineering Applications of Artificial Intelligence* 95 (2020), p. 103868 (cit. on p. 28).

[102] Qingxin Xia et al. “Robust Unsupervised Factory Activity Recognition with Body-worn Accelerometer Using Temporal Structure of Multiple Sensor Data Motifs”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4.3 (Sept. 2020). DOI: `10.1145/3411836`. URL: `https://doi.org/10.1145/3411836` (cit. on p. 28).

[103] Ali Ahmad Malik and Arne Bilberg. "Complexity-based task allocation in human-robot collaborative assembly". In: *Industrial Robot: the international journal of robotics research and application* 46.4 (2019), pp. 471–480 (cit. on pp. 29, 51, 52, 149).

[104] Md Milon Islam et al. "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects". In: *Computers in Biology and Medicine* (2022), p. 106060 (cit. on pp. 29, 30, 33, 50).

[105] Alina Roitberg et al. "Multimodal human activity recognition for industrial manufacturing processes in robotic workcells". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 2015, pp. 259–266 (cit. on p. 30).

[106] Haitao Wu et al. "A survey on teaching workplace skills to construction robots". In: *Expert Systems with Applications* 205 (2022), p. 117658 (cit. on p. 30).

[107] Niccolò Lucci et al. "Workflow modelling for human–robot collaborative assembly operations". In: *Robotics and Computer-Integrated Manufacturing* 78 (2022), p. 102384 (cit. on p. 30).

[108] Alyssa Kubota et al. "Activity recognition in manufacturing: The roles of motion capture and sEMG+ inertial wearables in detecting fine vs. gross motion". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 6533–6539 (cit. on pp. 30, 31).

[109] Suhan Park et al. "Robotic furniture assembly: task abstraction, motion planning, and control". In: *Intelligent Service Robotics* 15.4 (2022), pp. 441–457 (cit. on pp. 30, 31).

[110] Kaishu Xia et al. "A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence". In: *Journal of Manufacturing Systems* 58 (2021), pp. 210–230 (cit. on pp. 30, 32).

[111] Arturo Daniel Sosa-Ceron, Hugo Gustavo Gonzalez-Hernandez, and Jorge Antonio Reyes-Avendaño. "Learning from Demonstrations in Human–Robot Collaborative Scenarios: A Survey". In: *Robotics* 11.6 (2022), p. 126 (cit. on pp. 30, 32).

[112] Rim Slama et al. "An overview on human-centred technologies, measurements and optimisation in assembly systems". In: *International Journal of Production Research* (2023), pp. 1–23 (cit. on pp. 30, 34).

[113] Pengyue Guo et al. "A skill programming method based on assembly motion primitive for modular assembly system". In: *IEEE Access* 9 (2021), pp. 101369–101380 (cit. on pp. 30, 34).

[114] Yue Cao and CS Lee. "Robot Behavior-Tree-Based Task Generation with Large Language Models". In: *arXiv preprint arXiv:2302.12927* (2023) (cit. on pp. 30, 34).

[115] Xi Wang et al. "Automatic high-level motion sequencing methods for enabling multi-tasking construction robots". In: *Automation in Construction* 155 (2023), p. 105071 (cit. on pp. 30, 34).

[116] Francisco Suárez-Ruiz and Quang-Cuong Pham. "A framework for fine robotic assembly". In: *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2016, pp. 421–426 (cit. on pp. 30, 34).

[117] Anastasia Motrenko et al. "Continuous physical activity recognition for intelligent labour monitoring". In: *Multimedia Tools and Applications* 81.4 (2022), pp. 4877–4895 (cit. on pp. 30, 61).

[118] Sharath Chandra Akkaladevi et al. "Programming by interactive demonstration for a human robot collaborative assembly". In: *Procedia Manufacturing* 51 (2020), pp. 148–155 (cit. on pp. 30, 34).

[119] Jianjing Zhang, Peng Wang, and Robert X Gao. "Hybrid machine learning for human action recognition and prediction in assembly". In: *Robotics and Computer-Integrated Manufacturing* 72 (2021), p. 102184 (cit. on pp. 30, 61).

[120] Faliang He et al. "Recent progress in flexible microstructural pressure sensors toward human–machine interaction and healthcare applications". In: *Small Methods* 5.3 (2021), p. 2001041 (cit. on pp. 30, 34).

[121] Dinh C Nguyen et al. "Federated learning for smart healthcare: A survey". In: *ACM Computing Surveys (CSUR)* 55.3 (2022), pp. 1–37 (cit. on pp. 30, 34).

[122] Farman Ali et al. "An intelligent healthcare monitoring framework using wearable sensors and social networking data". In: *Future Generation Computer Systems* 114 (2021), pp. 23–43 (cit. on pp. 30, 34).

[123] Kristina Host and Marina Ivašić-Kos. "An overview of Human Action Recognition in sports based on Computer Vision". In: *Heliyon* (2022) (cit. on pp. 30, 33).

[124] Dian Shao et al. "Finegym: A hierarchical video dataset for fine-grained action understanding". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 2616–2625 (cit. on pp. 30, 34).

[125] Oresti Banos et al. "Window size impact in human activity recognition". In: *Sensors* 14.4 (2014), pp. 6474–6499 (cit. on pp. 30, 32).

[126] Bradley Schneider and Tanvi Banerjee. "Bridging the Gap between Atomic and Complex Activities in First Person Video". In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE. 2021, pp. 1–6 (cit. on pp. 30, 50).

[127] Ye Liu et al. "Action2Activity: recognizing complex activities from sensor data". In: *arXiv preprint arXiv:1611.01872* (2016) (cit. on p. 30).

[128] Karim Assi et al. "Complex Daily Activities, Country-Level Diversity, and Smartphone Sensing: A Study in Denmark, Italy, Mongolia, Paraguay, and UK". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–23 (cit. on pp. 30, 31, 64).

[129] Liangying Peng et al. "Aroma: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.2 (2018), pp. 1–16 (cit. on pp. 30, 61).

[130] Liangying Peng et al. "Hierarchical complex activity representation and recognition using topic model and classifier level fusion". In: *IEEE Transactions on Biomedical Engineering* 64.6 (2016), pp. 1369–1379 (cit. on pp. 30, 50).

[131] Ruohong Huan et al. "Human complex activity recognition with sensor data using multiple features". In: *IEEE Sensors Journal* 22.1 (2021), pp. 757–775 (cit. on pp. 30, 34).

[132] Pratool Bharti et al. "HuMAn: Complex activity recognition with multi-modal multi-positional body sensing". In: *IEEE Transactions on Mobile Computing* 18.4 (2018), pp. 857–870 (cit. on pp. 30, 31).

[133] Sakorn Mekruksavanich et al. "Resnet-se: Channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors". In: *IEEE Access* 10 (2022), pp. 51142–51154 (cit. on pp. 30, 31).

[134] Paula Lago et al. "A dataset for complex activity recognition withmicro and macro activities in a cooking scenario". In: *arXiv preprint arXiv:2006.10681* (2020) (cit. on pp. 30, 31).

[135] Marcus Rohrbach et al. "Recognizing fine-grained and composite activities using hand-centric features and script data". In: *International Journal of Computer Vision* 119 (2016), pp. 346–373 (cit. on pp. 30, 31, 33).

[136] Jake K Aggarwal and Michael S Ryoo. "Human activity analysis: A review". In: *Acm Computing Surveys (Csur)* 43.3 (2011), pp. 1–43 (cit. on pp. 30, 32).

[137] Kaixuan Chen et al. "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities". In: *ACM Computing Surveys (CSUR)* 54.4 (2021), pp. 1–40 (cit. on pp. 30, 50, 61).

[138] Stefan Dernbach et al. "Simple and complex activity recognition through smart phones". In: *2012 eighth international conference on intelligent environments*. IEEE. 2012, pp. 214–221 (cit. on pp. 30, 31).

[139] Luis Sanhudo et al. "Activity classification using accelerometers and machine learning for complex construction worker activities". In: *Journal of Building Engineering* 35 (2021), p. 102001 (cit. on pp. 30, 31).

[140] Ling Chen et al. "Deep learning based multimodal complex human activity recognition using wearable devices". In: *Applied Intelligence* 51 (2021), pp. 4029–4042 (cit. on pp. 30, 31).

[141] Narit Hnoohom et al. "Deep learning approach for complex activity recognition using heterogeneous sensors from wearable device". In: *2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C)*. IEEE. 2021, pp. 60–65 (cit. on pp. 30, 31).

[142] Leandro Miranda, José Viterbo, and Flávia Bernardini. "A survey on the use of machine learning methods in context-aware middlewares for human activity recognition". In: *Artificial Intelligence Review* (2022), pp. 1–32 (cit. on pp. 30, 33).

[143] Emma Bouton-Bessac, Lakmal Meegahapola, and Daniel Gatica-Perez. "Your Day in Your Pocket: Complex Activity Recognition from Smartphone Accelerometers". In: *International Conference on Pervasive Computing Technologies for Healthcare*. Springer. 2022, pp. 247–258 (cit. on p. 30).

[144] E Ramanujam, Thinagaran Perumal, and S Padmavathi. "Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review". In: *IEEE Sensors Journal* 21.12 (2021), pp. 13029–13040 (cit. on p. 30).

[145] Chengjun Chen et al. "Repetitive assembly action recognition based on object detection and pose estimation". In: *Journal of Manufacturing Systems* 55 (2020), pp. 325–333 (cit. on p. 30).

[146] Saguna Saguna, Arkady Zaslavsky, and Dipanjan Chakraborty. "Complex activity recognition using context-driven activity theory and activity signatures". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 20.6 (2013), pp. 1–34 (cit. on pp. 30, 33).

[147] Adebola Omolaja, Abayomi Otebolaku, and Ali Alfoudi. "Context-aware complex human activity recognition using hybrid deep learning models". In: *Applied Sciences* 12.18 (2022), p. 9305 (cit. on pp. 30, 31).

[148] Lourdes Martínez-Villaseñor and Hiram Ponce. "A concise review on sensor signal acquisition and transformation applied to human activity recognition and human–robot interaction". In: *International Journal of Distributed Sensor Networks* 15.6 (2019), p. 1550147719853987 (cit. on pp. 30, 31).

[149] Sakorn Mekruksavanich, Datchakorn Tancharoen, and Anuchit Jitpattanakul. "Human Activity Recognition in Logistics Using Wearable Sensors and Deep Residual Network". In: *TENCON 2023-2023 IEEE Region 10 Conference (TENCON)*. IEEE. 2023, pp. 194–198 (cit. on pp. 30, 31).

[150] Alexandru Iulian Alexan, Anca Roxana Alexan, and Stefan Oniga. "Real-Time Machine Learning for Human Activities Recognition Based on Wrist-Worn Wearable Devices". In: *Applied Sciences* 14.1 (2023), p. 329 (cit. on pp. 30, 31).

[151] Li Liu et al. "Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors". In: *Information Sciences* 340 (2016), pp. 41–57 (cit. on pp. 30, 31).

[152] Sravan Kumar Challa, Akhilesh Kumar, and Vijay Bhaskar Semwal. "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data". In: *The Visual Computer* 38.12 (2022), pp. 4095–4109 (cit. on pp. 30, 31).

[153] Jihoon Chung et al. "Haa500: Human-centric atomic action dataset with curated videos". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 13465–13474 (cit. on pp. 30, 31).

[154] Pushpajit Khaire and Praveen Kumar. "Deep learning and RGB-D based human action, human–human and human–object interaction recognition: A survey". In: *Journal of Visual Communication and Image Representation* 86 (2022), p. 103531 (cit. on pp. 30, 34).

[155] Farzana Kulsoom et al. "A review of machine learning-based human activity recognition for diverse applications". In: *Neural Computing and Applications* 34.21 (2022), pp. 18289–18324 (cit. on pp. 30, 33, 61).

[156] Md Golam Morshed et al. "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities". In: *Sensors* 23.4 (2023), p. 2182 (cit. on pp. 30, 33).

[157] Fatima Amjad et al. "A comparative study of feature selection approaches for human activity recognition using multimodal sensory data". In: *Sensors* 21.7 (2021), p. 2368 (cit. on p. 31).

[158] Ye Liu et al. "From action to activity: sensor-based activity recognition". In: *Neurocomputing* 181 (2016), pp. 108–115 (cit. on p. 31).

[159] Behrooz Azadi et al. "Robust Feature Representation Using Multi-Task Learning for Human Activity Recognition". In: *Sensors* 24.2 (2024), p. 681 (cit. on pp. 31, 34, 120, 154).

[160] Bernhard Anzengruber-Tanase et al. "Determining Best Hardware, Software and Data Structures for Worker Guidance during a Complex Assembly Task". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. 2022, pp. 63–72 (cit. on pp. 31, 45, 58, 61, 92, 132, 140).

[161] Markus Laube et al. "Analyzing Arc Welding Techniques improves Skill Level Assessment in Industrial Manufacturing Processes". In: *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*. 2023, pp. 177–186 (cit. on pp. 31, 132, 138, 139).

[162] Shibofin Zhang et al. "Deep learning in human activity recognition with wearable sensors: A review on advances". In: *Sensors* 22.4 (2022), p. 1476 (cit. on p. 31).

[163] Felix Mannhardt et al. "A taxonomy for combining activity recognition and process discovery in industrial environments". In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer. 2018, pp. 84–93 (cit. on p. 32).

[164] Abdelrahman Ahmad et al. "Privacy Preserving Workflow Detection for Manufacturing Using Neural Networks based Object Detection". In: *Proceedings of the 11th International Conference on the Internet of Things*. 2021, pp. 126–133 (cit. on pp. 32, 58, 132, 142).

[165] Kari Kuutti et al. "Activity theory as a potential framework for human-computer interaction research". In: *Context and consciousness: Activity theory and human-computer interaction* 1744 (1996), pp. 9–22 (cit. on p. 33).

[166] Regina Kyung-Jin Lee, Hao Zheng, and Yuqian Lu. "Human-Robot Shared Assembly Taxonomy: A step toward seamless human-robot knowledge transfer". In: *Robotics and Computer-Integrated Manufacturing* 86 (2024), p. 102686 (cit. on p. 34).

[167] Florian Garms et al. *Capturing value at scale in discrete manufacturing with industry 4.0*. Sept. 2019. URL: `https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/capturing-value-at-scale-in-discrete-manufacturing-with-industry-4-0` (cit. on p. 37).

[168] Daniel Küpper and Kyriakos Triantafyllidis. *Harnessing the AI revolution in industrial operations*. URL: `https://www3.weforum.org/docs/WEF_Harnessing_the_AI_Revolution_in_Industrial_Operations_2023.pdf` (cit. on p. 37).

[169] Tahera Kalsoom et al. "Advances in sensor technologies in the era of smart factory and industry 4.0". In: *Sensors* 20.23 (2020), p. 6783 (cit. on p. 37).

[170] Shibo Zhang et al. "Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances". In: *Sensors* 22.4 (2022). ISSN: 1424-8220. DOI: `10.3390/s22041476`. URL: `https://www.mdpi.com/1424-8220/22/4/1476` (cit. on pp. 37, 38, 40).

[171] Darius Nahavandi et al. "Application of artificial intelligence in wearable devices: Opportunities and challenges". In: *Computer Methods and Programs in Biomedicine* 213 (2022), p. 106541 (cit. on pp. 38, 40).

[172] Naghmeh Niknejad et al. "A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges". In: *Engineering Applications of Artificial Intelligence* 90 (2020), p. 103529 (cit. on pp. 38, 40).

[173] Fabio Sgarbossa et al. "Human factors in production and logistics systems of the future". In: *Annual Reviews in Control* 49 (2020), pp. 295–305 (cit. on pp. 38, 40).

[174] Sunita Kumari Chaurasia and SRN Reddy. "State-of-the-art survey on activity recognition and classification using smartphones and wearable sensors". In: *Multimedia Tools and Applications* 81.1 (2022), pp. 1077–1108 (cit. on pp. 38, 40).

[175] Muhammad Haseeb Arshad, Muhammad Bilal, and Abdullah Gani. "Human Activity Recognition: Review, Taxonomy and Open Challenges". In: *Sensors* 22.17 (2022). ISSN: 1424-8220. DOI: 10.3390/s22176463. URL: https://www.mdpi.com/1424-8220/22/17/6463 (cit. on pp. 38, 39, 42, 43).

[176] Ziqiu Kang, Cagatay Catal, and Bedir Tekinerdogan. "Machine learning applications in production lines: A systematic literature review". In: *Computers & Industrial Engineering* 149 (2020), p. 106773 (cit. on pp. 38, 40).

[177] Simon Fahle, Christopher Prinz, and Bernd Kuhlenkötter. "Systematic review on machine learning (ML) methods for manufacturing processes–Identifying artificial intelligence (AI) methods for field application". In: *Procedia CIRP* 93 (2020), pp. 413–418 (cit. on pp. 39, 40).

[178] João Gonçalo Pereira and Joaquim Gonçalves. "Human Activity Recognition: A review". In: *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*. 2022, pp. 1–5. DOI: 10.1109/ISDFS55398.2022.9800781 (cit. on p. 40).

[179] Xiang TR Kong et al. "Industrial wearable system: the human-centric empowering technology in Industry 4.0". In: *Journal of Intelligent Manufacturing* 30.8 (2019), pp. 2853–2869 (cit. on p. 40).

[180] Josef Wolfartsberger, Jean D Hallewell Haslwanter, and René Lindorfer. "Perspectives on assistive systems for manual assembly tasks in industry". In: *Technologies* 7.1 (2019), p. 12 (cit. on p. 40).

[181] Fadi El Kalach et al. "Cognitive manufacturing: definition and current trends". In: *Journal of Intelligent Manufacturing* (2024), pp. 1–21 (cit. on p. 42).

[182] Daniel Roggen et al. "Wearable computing". In: *IEEE Robotics & Automation Magazine* 18.2 (2011), pp. 83–95 (cit. on pp. 43, 44, 76).

[183] Georgios Sopidis et al. "System Design for Sensing in Manufacturing to Apply AI through Hierarchical Abstraction Levels". In: *Sensors* 24.14 (2024), p. 4508 (cit. on pp. 44, 45).

[184] Michaela Murauer et al. "A task-independent design and development process for cognitive products in industrial applications". In: *Proceedings of the 12th ACM international conference on pervasive technologies related to assistive environments*. 2019, pp. 358–367 (cit. on pp. 45, 124, 132, 136, 137).

[185] Georgios Sopidis et al. "Micro-activity recognition in industrial assembly process with IMU data and deep learning". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. 2022, pp. 103–112 (cit. on pp. 45, 52, 58, 61, 104, 149).

[186] Georgios Sopidis et al. "Micro Activities Recognition and Macro Worksteps Classification for Industrial IoT Processes". In: *Proceedings of the 11th International Conference on the Internet of Things*. 2021, pp. 185–188 (cit. on pp. 45, 58, 61, 132).

[187] Georgios Sopidis, Michael Haslgrübler, and Alois Ferscha. "Counting Activities Using Weakly Labeled Raw Acceleration Data: A Variable-Length Sequence Approach with Deep Learning to Maintain Event Duration Flexibility". In: *Sensors* 23.11 (2023), p. 5057 (cit. on p. 45).

[188] Ken Peffers et al. "A design science research methodology for information systems research". In: *Journal of management information systems* 24.3 (2007), pp. 45–77 (cit. on p. 47).

[189] Muhammad Ali Hassan et al. "Systematic Analysis of Risks in Industry 5.0 Architecture". In: *Applied Sciences* 14.4 (2024), p. 1466 (cit. on p. 50).

[190] Matteo Capponi et al. "Assessing perceived assembly complexity in human-robot collaboration processes: a proposal based on Thurstone's law of comparative judgement". In: *International Journal of Production Research* (2023), pp. 1–21 (cit. on p. 50).

[191] Fuqiang Gu et al. "A survey on deep learning for human activity recognition". In: *ACM Computing Surveys (CSUR)* 54.8 (2021), pp. 1–34 (cit. on pp. 50, 61).

[192] Cunbo Zhuang, Jingcheng Gong, and Jianhua Liu. "Digital twin-based assembly data management and process traceability for complex products". In: *Journal of manufacturing systems* 58 (2021), pp. 118–131 (cit. on pp. 52, 149).

[193] Panagiotis Kyratsis et al. "CAD based product design: A case study". In: *Int J Mod Manuf Technol* 11.3 (2019), pp. 110–5 (cit. on pp. 52, 149).

[194] Praveen Kumar Reddy Maddikunta et al. "Industry 5.0: A survey on enabling technologies and potential applications". In: *Journal of Industrial Information Integration* 26 (2022), p. 100257 (cit. on p. 52).

[195] Adrian Miqueo, Marta Torralba, and José A Yagüe-Fabra. "Lean manual assembly 4.0: A systematic review". In: *Applied Sciences* 10.23 (2020), p. 8555 (cit. on p. 52).

[196] Yong Yin, Kathryn E Stecke, and Dongni Li. "The evolution of production systems from Industry 2.0 through Industry 4.0". In: *International Journal of Production Research* 56.1-2 (2018), pp. 848–861 (cit. on p. 52).

[197] John Pandremenos et al. "Modularity concepts for the automotive industry: A critical review". In: *CIRP Journal of Manufacturing Science and Technology* 1.3 (2009), pp. 148–152 (cit. on p. 52).

[198] Melanie BUCHMÜLLER et al. "The Status of Modular Sourcing Compared to Other Procurement Strategies". In: () (cit. on p. 52).

[199] Friedrich Niemann et al. "Lara: Creating a dataset for human activity recognition in logistics using semantic attributes". In: *Sensors* 20.15 (2020), p. 4083 (cit. on pp. 52, 61).

[200] Lin Guo, Zongxing Lu, and Ligang Yao. "Human-machine interaction sensing technology based on hand gesture recognition: A review". In: *IEEE Transactions on Human-Machine Systems* 51.4 (2021), pp. 300–309 (cit. on p. 61).

[201] Marco Bortolini et al. "Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes". In: *Computers & Industrial Engineering* 139 (2020), p. 105485 (cit. on p. 61).

[202] Edwin Valarezo Añazco et al. "Hand gesture recognition using single patchable six-axis inertial measurement unit via recurrent neural networks". In: *Sensors* 21.4 (2021), p. 1404 (cit. on p. 61).

[203] Koenraad Vandevoorde et al. "Using Artificial Intelligence for Assistance Systems to Bring Motor Learning Principles into Real World Motor Tasks". In: *Sensors* 22.7 (2022), p. 2481 (cit. on p. 61).

[204] Elisa Digo, Stefano Pastorelli, and Laura Gastaldi. "A narrative review on wearable inertial sensors for human motion tracking in industrial scenarios". In: *Robotics* 11.6 (2022), p. 138 (cit. on p. 61).

[205] Marco Bortolini et al. “Adaptive automation assembly systems in the industry 4.0 era: A reference framework and full–scale prototype”. In: *Applied Sciences* 11.3 (2021), p. 1256 (cit. on p. 61).

[206] Michela Borghetti et al. “Validation of a modular and wearable system for tracking fingers movements”. In: *ACTA IMEKO* 9.4 (2020), pp. 157–164 (cit. on p. 61).

[207] A Riedel et al. “A deep learning-based worker assistance system for error prevention: Case study in a real-world manual assembly”. In: *Advances in Production Engineering & Management* 16.4 (2021), pp. 393–404 (cit. on p. 61).

[208] Takeshi Kurata et al. “Analyzing operations on a manufacturing line using geospatial intelligence technologies”. In: *IFIP International Conference on Advances in Production Management Systems*. Springer. 2022, pp. 69–76 (cit. on p. 61).

[209] Mejdi Dallel et al. “Inhard-industrial human action recognition dataset in the context of industrial collaborative robotics”. In: *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. IEEE. 2020, pp. 1–6 (cit. on p. 61).

[210] Fadime Sener et al. “Assembly101: A large-scale multi-view video dataset for understanding procedural activities”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 21096–21106 (cit. on p. 61).

[211] A Mastakouris et al. “Human worker activity recognition in a production floor environment through deep learning”. In: *Journal of Manufacturing Systems* 71 (2023), pp. 115–130 (cit. on p. 61).

[212] SJ Hayward et al. “A survey of indoor location technologies, techniques and applications in industry”. In: *Internet of Things* 20 (2022), p. 100608 (cit. on p. 61).

[213] Toufik Benmessabih et al. “Online human motion analysis in industrial context: A review”. In: *Engineering Applications of Artificial Intelligence* 131 (2024), p. 107850 (cit. on p. 61).

[214] Jingren Xu et al. “The principle, methods and recent progress in RFID positioning techniques: A review”. In: *IEEE Journal of Radio Frequency Identification* 7 (2023), pp. 50–63 (cit. on p. 61).

[215] Ouijdane Guiza et al. “Monitoring of human-intensive assembly processes based on incomplete and indirect shopfloor observations”. In: *2021 IEEE 19th International Conference on Industrial Informatics (INDIN)*. IEEE. 2021, pp. 1–8 (cit. on p. 61).

[216] Santosh Kumar Yadav et al. "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions". In: *Knowledge-Based Systems* 223 (2021), p. 106970 (cit. on pp. 61, 63).

[217] Lisa C Günther, Susann Kärcher, and Thomas Bauernhansl. "Activity recognition in manual manufacturing: Detecting screwing processes from sensor data". In: *Procedia CIRP* 81 (2019), pp. 1177–1182 (cit. on p. 61).

[218] Maurizio Faccio et al. "Human Factor Analyser for work measurement of manual manufacturing and assembly processes". In: *The International Journal of Advanced Manufacturing Technology* 103 (2019), pp. 861–877 (cit. on p. 61).

[219] Jefferson Hernandez et al. "Hierarchical Human Action Recognition to Measure the Performance of Manual Labor". In: *IEEE Access* 9 (2021), pp. 103110–103119 (cit. on p. 61).

[220] Roy Shilkrot13 et al. "WorkingHands: A hand-tool assembly dataset for image segmentation and activity mining". In: (2019) (cit. on p. 61).

[221] Lukas Büsch et al. "Towards recognition of human actions in collaborative tasks with robots: extending action recognition with tool recognition methods". In: *Sensors* 23.12 (2023), p. 5718 (cit. on p. 61).

[222] Christos Gkournelos et al. "Praxis: a framework for AI-driven human action recognition in assembly". In: *Journal of Intelligent Manufacturing* (2023), pp. 1–15 (cit. on p. 61).

[223] Christopher Reining et al. "Human activity recognition for production and logistics—a systematic literature review". In: *Information* 10.8 (2019), p. 245 (cit. on p. 61).

[224] Anna Ferrari et al. "Deep learning and model personalization in sensor-based human activity recognition". In: *Journal of Reliable Intelligent Environments* 9.1 (2023), pp. 27–39 (cit. on p. 61).

[225] Eduardo Luis Gomes et al. "Clustering and Hierarchical Classification for High-Precision RFID Indoor Location Systems". In: *IEEE Sensors Journal* 22.6 (2022), pp. 5141–5149. DOI: 10.1109/JSEN.2021.3103043 (cit. on p. 61).

[226] Md Al-Amin et al. "Action recognition in manufacturing assembly using multimodal sensor fusion". In: *Procedia Manufacturing* 39 (2019), pp. 158–167 (cit. on p. 61).

[227] Nikola Gjeldum et al. “Performance analysis of the RFID system for optimal design of the intelligent assembly line in the learning factory”. In: *Procedia Manufacturing* 23 (2018), pp. 63–68 (cit. on p. 61).

[228] Borja Bordel, Ramón Alcarria, and Tomás Robles. “Recognizing human activities in Industry 4.0 scenarios through an analysis-modeling-recognition algorithm and context labels”. In: *Integrated Computer-Aided Engineering* 29.1 (2022), pp. 83–103 (cit. on p. 61).

[229] Duarte Moutinho et al. “Deep learning-based human action recognition to leverage context awareness in collaborative assembly”. In: *Robotics and Computer-Integrated Manufacturing* 80 (2023), p. 102449 (cit. on p. 61).

[230] Hafiz Mughees Ahmad and Afshin Rahimi. “Deep learning methods for object detection in smart manufacturing: A survey”. In: *Journal of Manufacturing Systems* 64 (2022), pp. 181–196 (cit. on p. 61).

[231] Firgan Feradov, Valentina Markova, and Todor Ganchev. “Automated detection of improper sitting postures in computer users based on motion capture sensors”. In: *Computers* 11.7 (2022), p. 116 (cit. on p. 61).

[232] Longfei Zhou, Lin Zhang, and Nicholas Konz. “Computer vision techniques in manufacturing”. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53.1 (2022), pp. 105–117 (cit. on p. 61).

[233] Dawi Karomati Baroroh, Chih-Hsing Chu, and Lihui Wang. “Systematic literature review on augmented reality in smart manufacturing: Collaboration between human and computational intelligence”. In: *Journal of Manufacturing Systems* 61 (2021), pp. 696–711 (cit. on p. 61).

[234] Stephan Bollinger et al. “Evaluation of Potential Benefits of Augmented Reality for Industrial Services”. In: *IFIP International Conference on Advances in Production Management Systems*. Springer. 2022, pp. 135–144 (cit. on p. 61).

[235] Patrick Rueckert, Katrin Birgy, and Kirsten Tracht. “Image Based Classification of Methods-Time Measurement Operations in Assembly Using Recurrent Neuronal Networks”. In: *International Conference on System-Integrated Intelligence*. Springer. 2022, pp. 53–62 (cit. on p. 61).

[236] Vignesh Selvaraj et al. "Intelligent assembly operations monitoring with the ability to detect non-value-added activities as out-of-distribution (OOD) instances". In: *CIRP Annals* 72.1 (2023), pp. 413–416 (cit. on p. 61).

[237] Mohammed M Mabkhot et al. "Requirements of the smart factory system: A survey and perspective". In: *Machines* 6.2 (2018), p. 23 (cit. on p. 61).

[238] Heli Koskimäki et al. "Behavior modeling in industrial assembly lines using a wrist-worn inertial measurement unit". In: *Journal of Ambient Intelligence and Humanized Computing* 4.2 (2013), pp. 187–194 (cit. on p. 76).

[239] Realtime Technologies Ltd Shimmer Research. *Shimmer 9DoF Calibration Application User Manual*. Rev 2.10a. Available at: `http://www.shimmersensing.com`. Shimmer Research. 2017 (cit. on p. 77).

[240] Haojie Ma et al. "AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition." In: *IJCAI*. 2019, pp. 3109–3115 (cit. on p. 82).

[241] Carlos Avilés-Cruz et al. "Coarse-fine convolutional deep-learning strategy for human activity recognition". In: *Sensors* 19.7 (2019), p. 1556 (cit. on pp. 82, 83).

[242] Chunyong Yin et al. "Anomaly detection based on convolutional recurrent autoencoder for IoT time series". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52.1 (2020), pp. 112–122 (cit. on p. 82).

[243] Muhammad Shoaib et al. "Complex human activity recognition using smartphone and wrist-worn motion sensors". In: *Sensors* 16.4 (2016), p. 426 (cit. on p. 84).

[244] Anqi Mao, Mehryar Mohri, and Yutao Zhong. "Cross-entropy loss functions: theoretical analysis and applications". In: *Proceedings of the 40th International Conference on Machine Learning*. ICML'23. Honolulu, Hawaii, USA: JMLR.org, 2023 (cit. on p. 84).

[245] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *J. Mach. Learn. Res.* 15.1 (Jan. 2014), pp. 1929–1958. ISSN: 1532-4435 (cit. on p. 84).

[246] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: `https://www.tensorflow.org/` (cit. on pp. 84, 112).

[247] François Chollet et al. *Keras*. `https://keras.io`. 2015 (cit. on p. 84).

[248] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a (cit. on p. 84).

[249] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2 (cit. on p. 84).

[250] Terry Taewoong Um et al. "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks". In: *arXiv preprint arXiv:1706.00527* (2017) (cit. on pp. 86, 100).

[251] Anindya Das Antar, Masud Ahmed, and Md Atiqur Rahman Ahad. "Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: A review". In: *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE. 2019, pp. 134–139 (cit. on pp. 87, 104).

[252] Xia Cai and Zekun Ye. "Gaussian Weighting Reversion Strategy for Accurate Online Portfolio Selection". In: *IEEE Transactions on Signal Processing* 67.21 (2019), pp. 5558–5570. DOI: 10.1109/TSP.2019.2941067 (cit. on p. 87).

[253] Attila Reiss. *PAMAP2 Physical Activity Monitoring*. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5NW2H. 2012 (cit. on p. 98).

[254] Hongbo Gao et al. "Human motion segmentation based on structure constraint matrix factorization". In: *Science China Information Sciences* 65 (2022), pp. 1–2 (cit. on p. 103).

[255] Jason Z Kim et al. "Teaching recurrent neural networks to infer global temporal structure from local examples". In: *Nature Machine Intelligence* 3.4 (2021), pp. 316–323 (cit. on p. 104).

[256] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780 (cit. on p. 106).

[257] Piero Zappi et al. "Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection". In: *Wireless Sensor Networks: 5th European Conference, EWSN 2008, Bologna, Italy, January 30-February 1, 2008. Proceedings*. Springer. 2008, pp. 17–33 (cit. on p. 108).

[258] *MinMaxScaler Scikit learn*. `https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html` (cit. on p. 110).

[259] Allen Newell. *Unified theories of cognition*. Harvard University Press, 1994 (cit. on p. 124).

[260] Karen Butner and Grace Ho. "How the human-machine interchange will transform business operations". In: *Strategy & Leadership* 47.2 (2019), pp. 25–33 (cit. on p. 156).

[261] Junaid Younas and Paul Lukowicz. "Cognitive Ability Classification using On-body Sensors". In: *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*. 2022, pp. 317–320 (cit. on p. 157).

[262] Martin Schobesberger et al. "Designing Proactive Safety Systems for Industrial Workers Using Intelligent Mechanisms". In: *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*. 2022, pp. 480–485 (cit. on p. 157).

[263] Jaroslava Huber et al. "Addressing Worker Safety and Accident Prevention with AI". In: *Proceedings of the 11th International Conference on the Internet of Things*. 2021, pp. 150–157 (cit. on p. 157).

[264] Hymalai Bello et al. "TSAK: Two-Stage Semantic-Aware Knowledge Distillation for Efficient Wearable Modality and Model Optimization in Manufacturing Lines". In: *arXiv preprint arXiv:2408.14146* (2024) (cit. on p. 158).

[265] Jahandar Muradov. "Artificial intelligence in industry. Application of ISO 9001." In: *1st INTERNATIONAL CONFERENCE ON THE 4th INDUSTRIAL REVOLUTION AND INFORMATION TECHNOLOGY*. Vol. 1. 1. 2023, pp. 6–10 (cit. on p. 158).

[266] Jesús Oviedo et al. "ISO/IEC quality standards for AI engineering". In: *Computer Science Review* 54 (2024), p. 100681 (cit. on p. 158).

[267] Erin E Makarius et al. "Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization". In: *Journal of business research* 120 (2020), pp. 262–273 (cit. on p. 158).

[268] Michael Sony and Subhash Naik. "Industry 4.0 integration with socio-technical systems theory: A systematic review and proposed theoretical model". In: *Technology in society* 61 (2020), p. 101248 (cit. on p. 158).

[269] Benedikt Leichtmann et al. “Effects of Explainable Artificial Intelligence on trust and human behavior in a high-risk decision task”. In: *Computers in Human Behavior* 139 (2023), p. 107539 (cit. on p. 158).

[270] Kathrin Meyer et al. “How to explain AI in an understandable way? Citizen scientists, AI experts and artists find answers together”. In: (2023) (cit. on p. 158).