SHORT-PAPER

Towards Richer Insights in Human-Centred Observation Studies: Combining Thematic Analysis and Computer Vision

YUAN LIU, Queensland University of Technology, Brisbane, QLD, Australia

DR GLENDA AMAYO CALDWELL, Queensland University of Technology, Brisbane, QLD, Australia

MARKUS RITTENBRUCH, Queensland University of Technology, Brisbane, QLD, Australia

MATTHIAS GUERTLER, University of Technology Sydney, Sydney, NSW, Australia

MUGE FIALHO TEIXEIRA, Queensland University of Technology, Brisbane, QLD, Australia

ALAN G BURDEN, Queensland University of Technology, Brisbane, QLD, Australia

Open Access Support provided by:

University of Technology Sydney

Queensland University of Technology

Published: 29 November 2025

Citation in BibTeX format

OZCHI '25: 37th Australian Conference on Human-Computer Interaction
November 29 - December 3, 2025
Gadigal | Sydney, Australia

https://doi.org/10.1145/3764687.3769930
ISBN: 9798400720161

Towards Richer Insights in Human-Centred Observation Studies: Combining Thematic Analysis and Computer Vision

Yuan Liu
School of Architecture and Built Environment
Queensland University of Technology
Brisbane, QLD, Australia
Australian Cobotics Centre
Brisbane, QLD, Australia
yuan.liuy64@hdr.qut.edu.au

Glenda Amayo Caldwell
Construction and Architectural Robotics Lab (CARL), School of Architecture and Built Environment
Queensland University of Technology
Brisbane, QLD, Australia
Australian Cobotics Centre
Brisbane, QLD, Australia
g.caldwell@qut.edu.au

Markus Rittenbruch
School of Design
Queensland University of Technology
Brisbane, QLD, Australia
Australian Cobotics Centre
Brisbane, QLD, Australia
m.rittenbruch@qut.edu.au

Matthias Guertler
School of Mechanical and Mechatronic Engineering
University of Technology Sydney
Sydney, Australia
Australian Cobotics Centre
Sydney, Australia
matthias.guertler@uts.edu.au

Muge Teixeira
Construction and Architectural Robotics Lab (CARL), School of Architecture and Built Environment
Queensland University of Technology
Brisbane, QLD, Australia
Australian Cobotics Centre
Brisbane, QLD, Australia
muge.teixeira@qut.edu.au

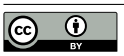
Alan G Burden
School of Architecture and Built Environment
Queensland University of Technology
Brisbane, QLD, Australia
Australian Cobotics Centre
Brisbane, QLD, Australia
alan.burden@qut.edu.au

Abstract

Due to the precision and efficiency that collaborative robots (cobots) offer, they are becoming increasingly vital to advanced manufacturing, particularly for dynamic and complex tasks that depend on human intelligence, such as decision making. To support cobot adoption, this paper presents a novel method for analyzing human decision-making and task complexity using video observation. Moving beyond conventional video analysis, the proposed approach combines thematic analysis with computer vision-based motion recognition to reveal behavioral patterns and decision-making processes. Through the application in a real-world manufacturing gasket room task, we demonstrate how the integration of interpretive coding and computational motion data can uncover insights into task structure, decision points, and potential cobot intervention zones. This method contributes a practical tool for aligning cobot functions with human needs in complex settings. It demonstrated how generative intelligence can augment human-centered research, and inform the design of future collaborative system that aligned with planetary sustainability.

CCS Concepts

• **Human centered computing** → **Human-computer interaction**; • **Human centered computing** → *Interaction design* ; • **Computing methodologies** → Artificial intelligence.



OZCHI '25, Sydney, Australia

ACM ISBN 979-8-4007-2016-1/25/11
https://doi.org/10.1145/3764687.3769930

Keywords

Human-Robot Collaboration, Human-Centered Design, Video Analysis, Thematic Analysis, Computer Vision, Task Complexity, Human Decision Making

ACM Reference Format:
Yuan Liu, Glenda Amayo Caldwell, Markus Rittenbruch, Matthias Guertler, Muge Teixeira, and Alan G Burden. 2025. Towards Richer Insights in Human-Centred Observation Studies: Combining Thematic Analysis and Computer Vision. In *37th Australian Conference on Human-Computer Interaction (HCI) (OZCHI '25), November 29–December 03, 2025, Sydney, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3764687.3769930

1 Introduction

As robots become increasingly accessible and affordable, research in the HCI community has expanded to include the growing field of human-robot interaction (HRI)[2, 20, 22, 23, 39]. Within HRI, a key area of focus is human-robot collaboration (HRC), where humans and robots work together in shared spaces to accomplish common tasks, with the human operator typically serving as the primary end user[6]. Collaborative robots, or cobots, are a category of robots specifically designed to support such collaborative systems[11, 31]. In manufacturing contexts, the effective design of cobots to interact seamlessly with operators is essential for enabling and supporting cobot-enabled tasks.

To design a meaningful human-centered HRC system in a real manufacturing setting, it is crucial to first understand the task itself [24, 25] and the human factors that shape its execution, including task complexity, physical workload, cognitive demands, and decision-making processes[18, 37]. Observation is one of the most widely used qualitative methods for understanding an existing workflow and human factors in real-world contexts [4, 10].

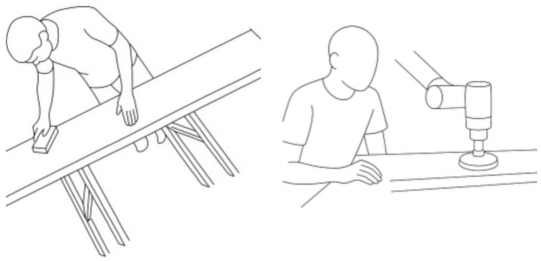


Figure 1: Current Manual Task (left) and Designed HRC Task (right)

Among observational tools, video recording offers non-intrusive, reviewable access to task performance [8]. Thematic annotation, a widely used approach in video analysis, allows researchers to identify behavioral patterns and decision points that shape task flow and complexity [5, 15, 17].

While thematic analysis provides rich interpretive insights, it is often time-consuming and limits the volume of video data that researchers can analyze in depth within a reasonable timeframe is onerous. Moreover, manual analysis increases the risk of overlooking subtle patterns in physical motion that could be important for understanding complex tasks. In parallel, computer vision technologies have advanced significantly, allowing for automated detection of motion trajectories and body pose directly from video [29, 30]. These technologies offer a practical way to track physical aspects of work, but they often lack the contextual sensitivity needed to explain the reason why particular behaviors occur. To address these limitations, this study integrates thematic annotation with computer vision-based motion analysis to address the research question: *How can we understand complex work tasks to inform the design of a human-robot collaboration system in manufacturing?*

This integrated approach is applied to a manufacturing task to uncover task workflow, patterns of human behavior, identify decision points, and inform future design of cobot systems for different contexts in the real world. It aims to demonstrate how generative intelligence such as computer vision can augment human-centered research, supporting the design of future collaborative system that are adaptive, efficient, aligned with goals of planetary sustainability.

2 Background and Related Work

Video-based observation offers a direct and minimally intrusive approach for understanding task workflows, human behavior, and decision making in manufacturing environments. This section reviews the relevant literature across three key domains that inform the proposed approach: observation through video recording as a foundation for studying human work in context, thematic annotation as a qualitative method for interpreting behavior from video, and computer vision-based approaches for extracting structured motion data. Together, these areas form the basis for developing a combined method capable of revealing patterns in human activity and supporting the design of HRC systems.

2.1 HRI and HRC Methods in HCI Literature

Recent research in human-robot interaction (HRI) and human-robot collaboration (HRC) has employed a range of methods, including experimental studies, video-based observation, interviews, and co-design workshops. Xu et al. [38] conducted laboratory experiments to examine whether engaging in caregiving behaviors toward a robot fosters stronger human-robot bonding, exploring both emotional and instrumental forms of care. Schneiders et al. [33] explored how entrainment affect human-robot collaboration through understanding human-human collaboration, where the research combined motion tracking, video recordings, and semi-structured interviews to capture both quantitative and qualitative insights. Hsu et al. [19] investigated robot facilitating reminiscence, using an experimental user study followed by co-design workshops to address instances where the robot responded inappropriately, and thematically analyzing the results of both studies. Together, these examples illustrate the diversity of methodological approaches in HCI, more specifically HRI/HRC research, and highlight the value of combining subjective and objective methods to capture complexity of human-robot interactions.

2.2 Thematic Analysis

Thematic analysis is one of the most popular qualitative methods for interpreting video or observational data to understand human behavior across diverse environments. It involves systematically identifying and analyzing patterns, or "themes," within qualitative data to gain insights [5] . For example, Fletcher and Gbadamosi [15] used thematic analysis to explore consumer decision-making in social media live streams. This approach has also informed the human-centered design of intensive care unit (ICU) spaces [12], helping designers gain insights into user experiences and needs. In medical robotics. Vermeulen et al. [36] applied thematic analysis to investigate human performance during robot-assisted surgeries, illuminating the dynamics between operators, clinicians and robotic systems. Within qualitative research, thematic analysis has proven highly effective in enabling researchers to gain in-depth, context-focused insights.

2.3 Computer Vision-based Approaches

Computer Vision (CV) is a field that employs algorithmic models to extract meaningful information from image and video data. In recent years, CV techniques have been increasingly applied across various domains to automatically analyze and interpret visual content. Within ergonomic research, for example, computer vision offers a promising approach for assessing human physical workload by enabling biomechanical pose estimation [13]. For example, CV-based video analysis has been used for biomechanical analysis of lifting tasks [8], ergonomic risk assessment [9], and the detection of risk factors associated with musculoskeletal disorders [26]. These CV-based approaches demonstrate the efficiency and accuracy of objective approach for interpreting real-world phenomena directly from pixel data. These areas form the basis for developing a combined method capable of revealing patterns in human activity and supporting the design of HRC systems. Such an approach integrates the depth of qualitative analysis with the scalability and precision of advanced techniques, enabling a more comprehensive

understanding of complex task environments. This dual perspective not only facilitates the identification of behavioral trends but also informs system design decisions that are grounded in both user needs and observable evidence.

3 Research Design: Case Study In a Gasket Room

To address the research question: *How can we understand complex work tasks to inform the design of a human-robot collaboration system in manufacturing?* , We employ a single case study [14]approach to this research. As part of a larger PhD research project within an industry focused research center, the Australian Cobotics Centre in Australia, this research has been developed in close collaboration with an industry partner providing access to the manufacturing facility as the case study. This section outlines the step-by-step process of the observational methods used and their application in a manufacturing gasket room to collect and analyze video data, with the aim of understanding task complexity and human decision-making relevant to cobot adoption. The gasket room task involves producing Formed-In-Place (FIP) gaskets[34, 35] using a semi-automated dispensing machine. Alongside the machine's operation, the task includes several manual steps, physically demanding, cognitively demanding operations, and real-time decision-making by human workers. This task presents clear opportunities for cobot integration, particularly in assisting with physically repetitive or ergonomically challenging elements, making it a strong candidate for exploring how cobots might be introduced into an existing workflow. Observation is a widely used method for understanding humans behavior and context from their perspectives [4], particularly in real-world environments such as manufacturing. To further minimize observer influence and maintain data richness, video recording was widely used. Video enables researchers to observe remotely and repeatedly, without intruding on the task space or altering worker behavior. It also can strengthen the credibility of the study by reducing observer bias and enabling the application of more systematic methods to support reliability [7]. Video recording represents one of the most effective methods for capturing and analyzing behavior in observational studies. In this study, a non-participant observation approach [10] was adopted to avoid disrupting the natural workflow of the task. This type of observation allows researchers to remain external to the activity while still capturing authentic human behavior. This case study was approved by the University Human Research Ethics Committee at Queensland University of Technology (Project ID 5545). All participants provided informed consent.

3.1 Data Collection

3.1.1 Video Recording. The tasks that we observed are carried out in a real manufacturing environment. Video data was collected using two cameras positioned at different locations within the workspace to minimize blind spots and reduce the likelihood of missing important actions. In the gasket room case study, two 360-degree cameras (GoPro Max) were positioned at distinct locations to provide comprehensive visual coverage of worker activity. Recordings were conducted on three separate workdays, each spaced one week apart, to capture variation in task flow and behavior. Each

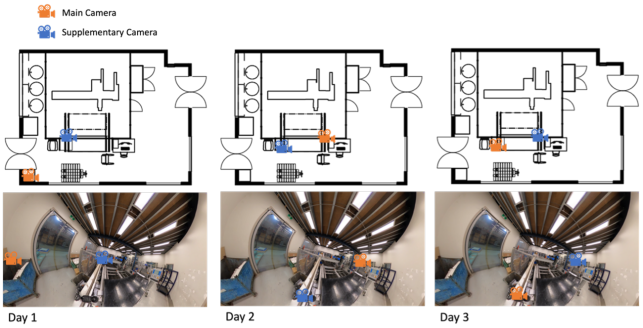


Figure 2: Camera setting in the gasket room

recording session lasted from 10:30 a.m. to 2:30 p.m., covering a complete work cycle as much as possible. Researchers were present solely to monitor the recording equipment and to replace camera batteries as needed, typically once per session to ensure minimal intrusion or interference with the tasks being performed.

3.1.2 Video Processing and Data Preparation. To prepare for data analysis, all video footage needs to be segmented into short clips. The recorded videos of gasket room were automatically segmented by the camera due to a file size limitation of 4GB, resulting in individual video segments of approximately eight minutes each. As a result, no manual segmentation was required. Following data collection, the footage was edited using GoPro Quik software and exported in 1080p MP4 format for analysis. The editing process involved adjusting the lens angle to ensure that both the operators and task-relevant activities remained consistently visible throughout each segment. These adjustments were applied uniformly across all recordings to maintain visual consistency and data quality for both thematic annotation and computer vision-based motion analysis. After the editing process, researchers reviewed all video recordings and selected footage from one camera as the primary source for analysis, with the second camera serving as a supplementary view. To minimize perspective bias, the designated primary camera varied across the three recording days. Video from the primary camera was analyzed in full, while the supplementary footage was reviewed alongside it to ensure that no relevant actions or behaviors were missed. In instances where a task-relevant event was captured only by the supplementary camera, that segment was included in the analysis to maintain completeness and accuracy.

3.2 Data Analysis

3.2.1 Hybrid Thematic Analysis. The thematic analysis followed a hybrid coding strategy, combining both deductive (pre-defined) and inductive (emergent) coding to capture the structure and complexity of human behavior during the task. The process was conducted in the following steps:

1. Development of Pre-Defined Codes. Based on prior knowledge, task documentation, and the study's research questions, an initial set of codes and code groups was created. These included categories related to the standard manufacturing process steps, physical and cognitive workload, decision-making types, and types of human-system interaction.

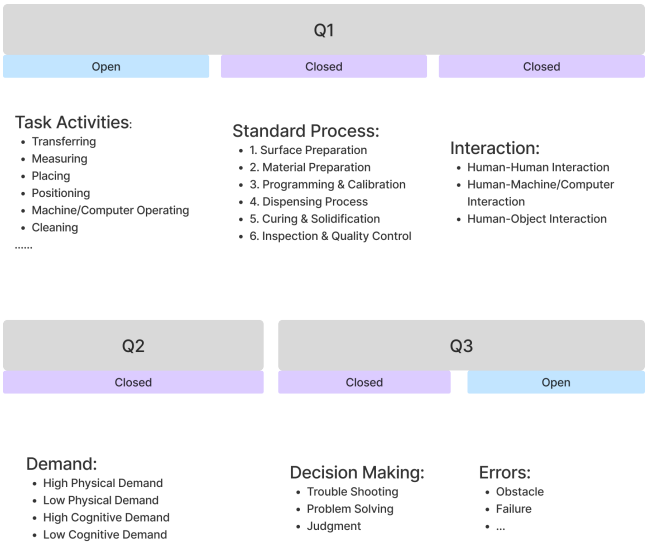


Figure 3: Code Groups

To guide the coding process, three questions were established:

Q1: What are the breakdown steps involved in the gasket room task?

Q2: Which steps present obvious physical demand and cognitive demand?

Q3: Among the cognitively demanding steps, which involve decision-making processes and factors trigger these processes?

A codebook was developed to support consistent annotation, with clear definitions and illustrative examples for each code.

2. Initial Video Coding and Emergent Code Identification. As video annotation progressed, the researchers remained open to new patterns or behaviors not captured by the initial codebook. Emergent codes, such as unexpected errors, task interruptions, and informal troubleshooting, were developed inductively during this phase. To ensure reliability, two researchers independently annotated selected video segments. Discrepancies were resolved through collaborative discussion and iterative refinement of the code definitions. A hybrid coding strategy was employed to capture both anticipated and emergent aspects of operator behavior. Deductive coding groups were derived from task documentation and domain expertise and included categories such as FIP Gasket Standard Process, Interaction, Demand, and Decision Making. One of the core closed code groups was based on the Standard Process of Formed-In-Place (FIP) gasketing, as defined by industry documentation and workplace procedures [3]. This process was broken down into six sequential stages:

(1) Surface Preparation - cleaning and preparing the surface area for gasket application;
(2) Material Preparation - loading and verifying the gasket material and equipment readiness;
(3) Programming and Calibration - setting up the machine parameters and calibrating the dispenser;
(4) Dispensing Process - executing the gasket application using the dispensing equipment;
(5) Curing and Solidification - allowing the applied material to cure and solidify as per process requirements;
(6) Inspection and Quality Control - visually inspecting the gasket and verifying adherence to quality standards.

The interaction group, demand group and decision making group are defined based on preliminary observations. Interaction includes human-machine(computer) interaction, human-object interaction and human-human interaction. Demand includes physical and cognitive demands and depends on the effort they are devised to high physical demand, low physical demand, high cognitive demand and low cognitive demand. Decision making includes judgment, troubleshooting and problem solving. These coding groups provided a structured lens for interpreting task execution and behavioral roles. In parallel, inductive coding was used to capture unexpected observations directly from the video data. Emergent codes included specific Task Activities within standard steps, as well as Errors, Interruptions, and nuanced Operator Behaviors not originally accounted for. This hybrid approach ensured the analysis remained both grounded in formal task structure and responsive to context-specific dynamics observed in the real-world setting.

3. Codebook Refinement. The codebook was iteratively updated to include both the closed and open codes. Definitions were clarified and examples added to support consistency across coders.

4. Reliability and Consensus Building. To enhance coding reliability, two researchers independently annotated selected video segments. Discrepancies were discussed and resolved collaboratively, leading to further refinement of the code definitions.

5. Theme Development. After all coding was complete, codes were grouped into higher-level themes through comparison, pattern identification, and alignment with the study's research questions. Both inductive and deductive codes contributed to final thematic categories, which reflected broader cognitive, behavioral, and procedural insights from the data. Annotation was performed using Atlas.ti [1], with the main camera footage serving as the primary source for coding. When relevant behaviors were only visible in the supplementary camera, those sequences were incorporated into the coding process. The resulting thematic annotations served as an interpretive layer that complemented the motion data extracted through computer vision, detailed in Step 4.

3.2.2 Computer Vision-based Video Analysis. After thematic annotation, some patterns may remain unobserved by the researchers. To address this, a computer vision-based method was used to detect additional patterns and ensure more comprehensive results. This method Realtime Collaborative Action Tracking (ReCAT) [28] combines MediaPipe [16], DeepSORT [32] and YOLO11 [21] to identify the most frequently repeated and physically demanding tasks, as well as to track operators' movement trajectories within the workspace. Our earlier version, which used only MediaPipe, performed well for detecting and tracking a single person in real time. However, in real-world scenarios, especially in collaborative environments, it is necessary to account for multiple individuals. Therefore, the current version of ReCAT combines YOLO11, which

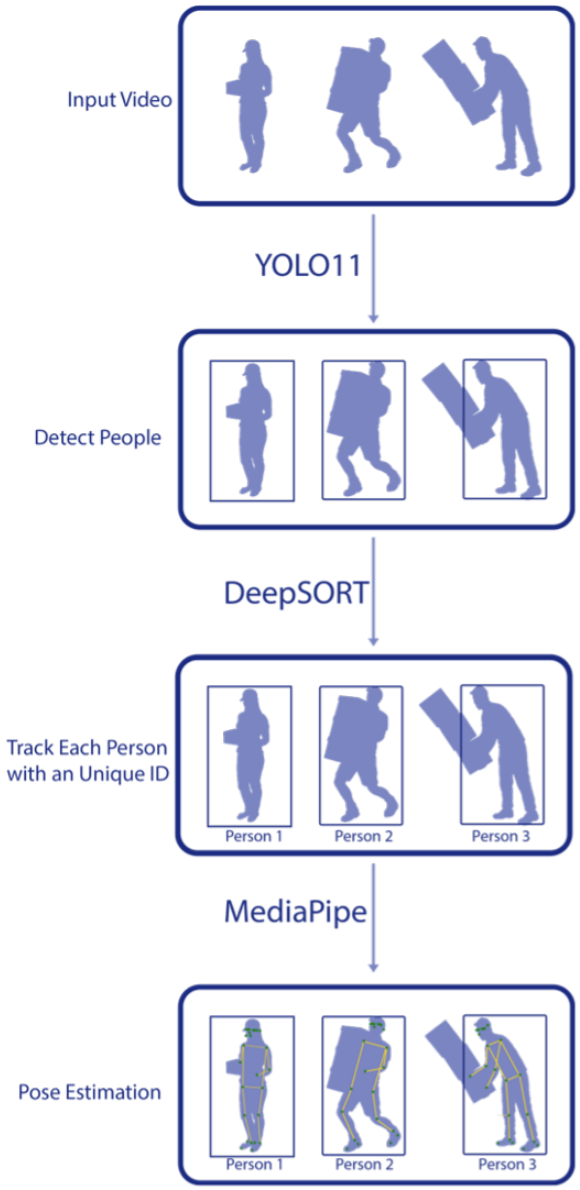


Figure 4: ReCAT Workflow

is effective for detecting multiple people, with MediaPipe, which provides highly accurate human pose estimation. Our new version of ReCAT detects people in each video frame using YOLO11, then assigns a unique ID to each detected person and tracks them across frames with DeepSORT. Finally, it estimates the pose of each person using MediaPipe as depicted in Figure 4.

1. People Detection. We employed the YOLO11-pose model, a state-of-the-art, real-time object detector, to identify and localize all human subjects in each video frame. YOLO11 was selected for its high detection accuracy and real-time inference capability, ensuring

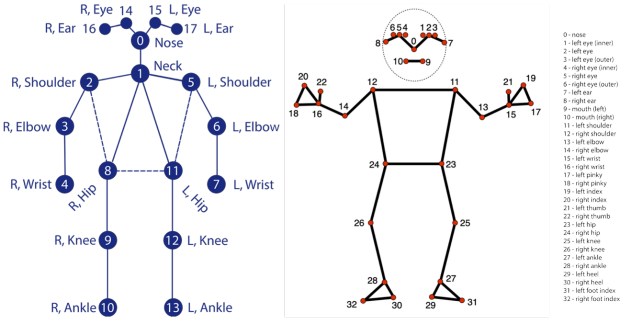


Figure 5: COCO format 18 key points, adapted from [27] (left), and MediaPipe's 33 key points landmarks [16] (right).

robust person detection even under challenging conditions such as partial occlusions and varying lighting.

2. Multi-People Tracking (Persistent Identity). To maintain consistent identities of individuals across frames, we integrated a DeepSORT tracker. DeepSORT combines bounding box motion prediction (via Kalman filtering) and deep appearance features, providing robust multi-person tracking with persistent unique IDs, even in crowded scenes or when temporary occlusions occur. Each detected person's bounding box and confidence score from YOLO were passed to DeepSORT, which outputs a track ID for every person in each frame.

3. Human Pose Estimation. Because YOLO Pose uses the COCO format, which defines 18 human body landmarks, we applied MediaPipe Pose, which provides 33 anatomical landmark coordinates per person per frame, capturing key body joints and segments in greater detail (Figure 5).

4. Physical Activity Recognition and Workload Metrics. We designed custom algorithms to recognize key physical activities and estimate workload metrics, including:

- Walking Distance: For each person, we computed the cumulative Euclidean distance of a central body point (nose or hip center) across consecutive frames, normalized by an estimated pixels-per-meter scaling derived from bounding box height and assumed real-world person height.
- Squat and Bend Detection: Squats were detected by monitoring the vertical displacement of the hip landmarks relative to a personalized standing baseline, using state machines to track up/down transitions and filter out noise. Bends were similarly detected using the average vertical position of the shoulder landmarks.
- Event Counting and Depth Measurement: For each tracked identity, the number of squats and bends, as well as the depth of each squat, were recorded over time.
- Result Visualization: The analysis pipeline overlays bounding boxes, landmark skeletons, person IDs, and live activity statistics directly on video frames for qualitative validation.

5. Data Output and Analysis. All extracted metrics (walking distance, squat and bend counts, event timings) were logged per frame

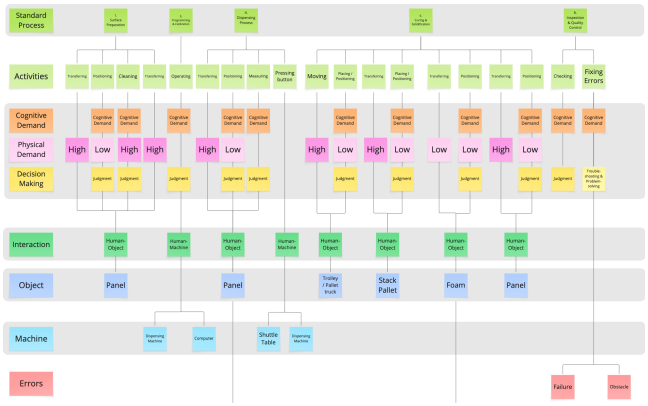


Figure 6: Code groups of Thematic Analysis

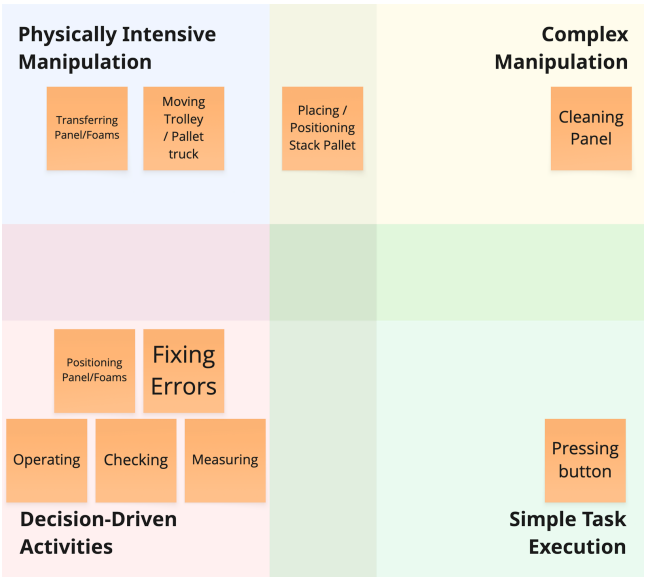


Figure 7: Themes from Thematic Analysis

and per person, and exported as a structured CSV for further statistical analysis. Processed annotated videos were saved for qualitative review. A key benefit of the computer vision-based approach is its high level of automation. However, automation does not mean the system cannot be customized. ReCAT can be adapted to suit specific needs. In this study, ReCAT is used to detect operators in the workspace and calculate their workload based on pose estimation and movement trajectories within the workspace. These analyses provide insights into the actual physical demands of specific tasks and reveal how operators move through the workspace. The code and supplementary materials are available at our GitHub repository: https://github.com/yuanliu233/ReCAT.

4 Preliminary Results

4.1 Results of Thematic Analysis

In the standard process group, all six procedures are observed in the videos. As these are consistent with common industrial standards, no modifications were necessary for this group in the *Codebook Refinement* step. Based on the analysis of the video recordings, the task activities were identified as follows:

- transferring the panel,
- positioning the panel,
- cleaning the panel,
- operating the dispensing machine,
- operating the computer,
- measuring panel dimensions,
- pressing the shuttle table button,
- pressing the dispensing machine button,
- moving the trolley or pallet truck,
- transferring pallet stack,
- placing and positioning pallet stack,
- transferring foams,
- positioning foams,
- checking quality,
- fixing errors.

It is difficult to accurately assess the level of cognitive effort exerted by operators based solely on video observation. Therefore, any activity requiring cognitive engagement was categorized as having cognitive demand. Examples include operating machines or computers, measuring, placing, positioning, checking quality, and fixing errors. Activities were classified as high physical demand if operators were required to move heavy loads, such as trolleys, pallet trucks, panels, or pallet stacks, or if they performed repetitive body movements within a short period (e.g., cleaning). Activities were classified as low physical demand if operators only made minor adjustments to heavy objects (e.g., positioning panels or pallet stack) or moved lightweight objects (e.g., transferring or positioning foams). Decision making was identified when operators exercised judgment and performed troubleshooting or problem-solving in response to errors. Examples of judgment include positioning objects for alignment and checking quality, while errors may involve obstacles or equipment failures. Interactions were categorized as human-object interaction, human-machine/computer interaction, and human-human interaction. These were all considered physical interactions, as they were identified through video observation. Human-object interaction was observed when operators handled objects such as panels, foam, trolleys, or other equipment. Human-machine/computer interaction was identified when operators operated machines or computers. Human-human interaction occurred when two or more operators worked together to handle the same item. Errors were observed when something disrupted the workflow. In this study, obstacles and failures were classified as errors.

From the codes, the following themes were identified: Physically Intensive Manipulation, Complex Manipulation, Decision-Driven Activities, and Simple Task Execution. Physically Intensive Manipulation refers to tasks primarily involving high physical demands and human-object interactions, such as transferring heavy panels or materials. Complex Manipulation involves tasks that combine

Figure 8: Output Video Clips from ReCAT

significant physical effort and cognitive demands during human-object interactions. Examples include tasks requiring careful cleaning, positioning, or precise adjustments to objects. Decision-Driven Activities represent tasks requiring significant cognitive engagement, including troubleshooting, measuring, positioning, and quality checking. Simple Task Execution describes tasks characterized by minimal cognitive and physical effort, typically straightforward actions such as pressing a button to operate machinery when all preparations are complete (Figure 7).

4.2 Results of ReCAT Analysis

To supplement the thematic analysis, we employed ReCAT to visually understand the movement of the operators in the gasket room. Both operators were successfully detected and assigned unique IDs throughout the ReCAT analysis. The system accurately identified key actions, including walking, squatting, and bending, for each operator. Using these detected actions, we calculated the estimated energy expenditure for each individual (Table 1). The results demonstrated that the computer vision approach provided effective physical workload estimations, validating the effectiveness of the method for ergonomic assessment.

These capabilities provide objective, granular data about when and where physical effort is concentrated within a task. When combined with thematic analysis, such data can reveal which task steps impose high physical demands or repetitive strain, offering clear targets for robotic assistance. In this way, the approach directly addresses our research question by helping to understand the structure, demands, and challenges of complex work tasks.

5 Discussion

The contributions of this work are threefold: (1) it introduces a video-based analytic framework for understanding human decision-making in task execution, (2) it introduces a computer vision tool based on OpenCV, MediaPipe and YOLO to track motion and pose data, (3) it demonstrates the application of this method in a real-world manufacturing case study, and (4) it generates actionable insights for designing human-robot collaboration by identifying decision-intensive moments and repetitive physical actions suitable for automation. Together, these contributions practically address the needs for identifying behavioral patterns and informing collaborative robot adoption in complex task environment.

This study demonstrates the value of combining thematic analysis with computer vision-based video analysis for understanding tasks and human behavior in real-world settings. Through computer vision, we were able to detect and uniquely identify operators and automatically recognize their physical actions, enabling an objective assessment of physical workload. In parallel, thematic analysis provided deeper insights into the context and subjective experiences behind these activities. Thematic analysis allowed us to identify connections among code groups, both predefined and those that emerged during the analysis. This approach offered insights into real-world task breakdowns, workload distribution, decision-making processes, interactions among humans, machines, and objects, as well as the types of errors that trigger troubleshooting and problem-solving. Such qualitative insights are essential for understanding task complexity and human decision making in future studies. Using ReCAT, our computer vision-based tool, we efficiently and objectively captured human physical activities such as walking, squatting, and bending by analyzing joint movement and angles. This enabled direct estimation of energy expenditure and, consequently, physical workload offering a level of objectivity not possible through thematic analysis alone. This integrated approach connects in-depth insights into human behavior with scalable, automated analysis. It not only provides insights derived from subjective analysis but also capture patterns objectively, thereby bridging the gap between the observer's perspective and actual task execution.

6 Conclusion

While the initial results are promising, several limitations must be acknowledged. First, the current case study was based on a specific manufacturing task scenario, which may limit the generalizability of the combined method. The accuracy of computer vision-based action recognition can be affected by factors such as video resolution, occlusions, and individual variations in movement. Additionally, thematic analysis relies on subjective coding and interpretation, which can introduce researcher bias and require significant time investment. Finally, while the integrated approach offers complementary perspectives, the process of systematically combining findings remains a challenge and is still in the early stages. Future work will focus on expanding the application of this combined method to larger and more diverse datasets. Incorporating human perspectives through interviews and other qualitative approaches will help to validate the method and strengthen the reliability of the results. Efforts will also be made to enhance the accuracy of the computer vision algorithms. Additionally, we aim to develop systematic approaches for integrating subjective and objective findings. Ultimately, ongoing refinement of this approach will contribute to the design of human-robot collaborative systems for the real world. In this study, we presented an integrated method that combines thematic analysis and computer vision-based video analysis, ReCAT, to better understand human tasks and decision making in real-world environments. By leveraging the strengths of both qualitative and quantitative methods, we were able to capture not only the objective measurement of physical workload through automated action recognition and energy estimation, but also the

Table 1: 60-Seconds Physical Workload Estimation from ReCAT

ID	Walk (m)	Squats	Bends	Work Walk (J)	Work Squat (J)	Work Bend (J)	Total Work (J)	Calories (kcal)
1	1.668	1	0	233.46	161.36	0.00	394.82	0.094
2	30.693	0	1	4297.01	0.00	172.22	4469.23	1.068

contextual and subjective aspects of task complexity and decision-making. This dual perspective provides a more comprehensive understanding of human work and lays the groundwork for designing more effective human-centered collaborative systems. Moving forward, refining this combined methodology and applying it to broader contexts holds promise for advancing both research and practice in human-robot collaboration and ergonomic assessment. Besides, this method offers qualitative researchers the opportunity to gather richer and more comprehensive data, while also serving as an effective preliminary approach for contextual and user understanding in the design process. Its utility extends beyond the design of human-robot collaboration to diverse domains of human-centered research. Our research aligns with the OzCHI 2025 theme of *Generative Intelligence, Planetary Futures* by integrating a robust qualitative method, thematic analysis, with an intelligent computer vision approach our ReCAT tool, to create a more comprehensive, flexible, and adaptable method for future human-centered research.

Acknowledgments

The authors would like to acknowledge the support received through the following funding schemes of the Australian Government: ARC Industrial Transformation Training Centre (ITTC) for Collaborative Robotics in Advanced Manufacturing under Grant IC200100001.

References

[1] 2025. The #1 Software for Qualitative Data Analysis. https://atlasti.com/. Accessed: 2025-8-9.
[2] Naoko Abe, David Rye, and Lian Loke. 2019. A microsociological approach to understanding the robot collaborative motion in human-robot interaction. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (Fremantle WA Australia). ACM, New York, NY, USA.
[3] Applications Engineering. 2022. Form-In-Place Gasketing. https://sealingdevices.com/blog/form-in-place-gasketing/. Accessed: 2025-4-22.
[4] Lynda Baker. 2006. Observation: A complex research method. *Libr. Trends* 55, 1 (2006), 171–189.
[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 2 (Jan. 2006), 77–101.
[6] Alan G Burden, Glenda Amayo Caldwell, and Matthias R Guertler. 2022. Towards human–robot collaboration in construction: current cobot trends and forecasts. *Construction Robotics* 6, 3 (Dec. 2022), 209–220.
[7] Kay Caldwell and Anita Atwal. 2005. Non-participant observation: using video tapes to collect data in nursing research. *Nurse Res.* 13, 2 (Oct. 2005), 42–54.
[8] Chien-Chi Chang, Simon Hsiang, Patrick G Dempsey, and Raymond W McGorry. 2003. A computerized video coding system for biomechanical analysis of lifting tasks. *Int. J. Ind. Ergon.* 32, 4 (Oct. 2003), 239–250.
[9] Theocharis Chatzis, Dimitrios Konstantinidis, and Kosmas Dimitropoulos. 2022. Automatic ergonomic risk assessment using a variational deep network architecture. *Sensors (Basel)* 22, 16 (Aug. 2022), 6051.
[10] Malgorzata Ciesielska, Katarzyna W Boström, and Magnus Öhlander. 2018. Observation Methods. In *Qualitative Methodologies in Organization Studies*. Springer International Publishing, Cham, 33–52.
[11] J Edward Colgate, Witaya Wannasuphoprasit, and Michael A Peshkin. 1996. Cobots: Robots for collaboration with human operators. In *Dynamic Systems and Control* (Atlanta, Georgia, USA). American Society of Mechanical Engineers, 433–439.
[12] Jody Ede, David Garry, Graham Barker, Owen Gustafson, Elizabeth King, Hannah Routley, Christopher Biggs, Cherry Lumley, Lyn Bennett, Stephanie Payne, Andrew Ellis, Clinton Green, Nathan Smith, Laura Vincent, Matthew Holdaway, and Peter Watkinson. 2023. Building a Covid-19 secure intensive care unit: A human-centred design approach. *J. Intensive Care Soc.* 24, 1 (Feb. 2023), 71–77.
[13] Darlington Egeonu and Bochen Jia. 2025. A systematic literature review of computer vision-based biomechanical models for physical workload estimation. *Ergonomics* 68, 2 (Feb. 2025), 139–162.
[14] Colin W Evers and Echo H Wu. 2006. On generalising from single case studies: Epistemological reflections. *J. Philos. Educ.* 40, 4 (Nov. 2006), 511–526.
[15] Kathy-Ann Fletcher and Ayantunji Gbadamosi. 2022. Examining social media live stream's influence on the consumer decision-making: a thematic analysis. *Electron. Commer. Res.* 24, 3 (Oct. 2022), 2175–2205.
[16] Google. 2023. MediaPipe.
[17] Greg Guest and Kathleen M MacQueen. 2025. Applied Thematic Analysis. https://uk.sagepub.com/en-gb/eur/applied-thematic-analysis/book233379. Accessed: 2025-7-29.
[18] Sarah Hopko, Jingkun Wang, and Ranjana Mehta. 2022. Human Factors Considerations and Metrics in Shared Space Human-Robot Collaboration: A Systematic Review. *Frontiers in Robotics and AI* 9 (Feb. 2022), 799522.
[19] Long-Jing Hsu, Manasi Swaminathan, Weslie Khoo, Kyrie Jig Amon, Hiroki Sato, Sathvika Dobbala, Kate Tsui, David Crandall, and Selma Sabanovic. 2025. Bittersweet snapshots of life: Designing to address complex emotions in a reminiscence interaction between older adults and a robot. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan). ACM, New York, NY, USA, 1–18.
[20] Toshihiko Isaka, Ryosuke Aoki, Naoki Ohshima, and Naoki Mukawa. 2018. Study of socially appropriate robot behaviors in human-robot conversation closure. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne Australia). ACM, New York, NY, USA.
[21] Glenn Jocher and Jing Qiu. 2024. Ultralytics YOLO11.
[22] Stine Johansen, Hashini Senaratne, Alan Burden, David Howard, Glenda Amayo Caldwell, Jared Donovan, Andreas Duenser, Matthias Guertler, Melanie Mcgrath, Cecile Paris, Markus Rittenbruch, and Jonathan Roberts. 2023. Empowering People in Human-Robot Collaboration: Bringing Together and Synthesising Perspectives. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction* (<conf-loc>, <city>Canberra</city>, <state>ACT</state>, <country>Australia</country>, </conf-loc>) *(OzCHI '22)*. Association for Computing Machinery, New York, NY, USA, 352–355.
[23] Stine S Johansen, Hashini Senaratne, Alan Burden, Melanie McGrath, Claire Mason, Glenda Caldwell, Jared Donovan, Andreas Duenser, Matthias Guertler, David Howard, Yanrang Jiang, Cecile Paris, Markus Rittenbruch, and Jonathan Roberts. 2024. Empowering People in Human-Robot Collaboration: Why, How, When, and for Whom. In *Proceedings of the 35th Australian Computer-Human Interaction Conference* (<conf-loc>, <city>Wellington</city>, <country>New Zealand</country>, </conf-loc>) *(OzCHI '23)*. Association for Computing Machinery, New York, NY, USA, 684–688.
[24] K. M. Rabby, M. Khan, A. Karimoddini, and S. X. Jiang. 2019. An effective model for human cognitive performance within a human-robot collaboration framework, Vol. 2019-October.
[25] Michail Karakikes and Dimitris Nathanael. 2023. The effect of cognitive workload on decision authority assignment in human–robot collaboration. *Cogn. Technol. Work* 25, 1 (Feb. 2023), 31–43.
[26] Li Li, Tara Martin, and Xu Xu. 2020. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Appl. Ergon.* 87, 103138 (Sept. 2020), 103138.
[27] Feng-Cheng Lin, Huu-Huy Ngo, Chyi-Ren Dow, Ka-Hou Lam, and Hung Linh Le. 2021. Student behavior recognition system for the classroom environment based on skeleton pose estimation and person detection. *Sensors (Basel)* 21, 16 (Aug. 2021), 5314.
[28] Yuan Liu. 2025. ReCAT: Real-time Collaborative Action Tracking.
[29] Tewodros Legesse Munea, Yalew Zelalem Jembre, Halefom Tekle Weldegebriel, Longbiao Chen, Chenxi Huang, and Chenhui Yang. 2020. The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation. *IEEE Access* 8 (2020), 133330–133348.
[30] Bruce Xiaohan Nie, Caiming Xiong, and Song-Chun Zhu. 2015. Joint action recognition and pose estimation from video. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, MA, USA). IEEE, 1293–1301.
[31] Michael Peshkin and J Edward Colgate. 1999. Cobots. *Ind. Rob.* 26, 5 (July 1999), 335–341.

[32] Abhijeet Pujara and Mamta Bhamare. 2022. DeepSORT: Real time & multi-object detection and tracking with YOLO and TensorFlow. In *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)* (Trichy, India). IEEE, 456–460.

[33] Eike Schneiders, Christopher Fourie, Stanley Celestin, Julie Shah, and Malte Jung. 2024. Understanding entrainment in human groups: Optimising human-robot collaboration from lessons learned during human-human collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu HI USA). ACM, New York, NY, USA.

[34] Joseph Tokarski. 1977. Formed-in-place gaskets: Concept vs reality. In *SAE Technical Paper Series*. SAE International, 400 Commonwealth Drive, Warrendale, PA, United States.

[35] Shingo Tsuno, Kiyotaka Sawa, Chiu-Sing Lin, and Masahiro Masujima. 2009. Next generation formed-in-place gasket (FIPG) liquid sealant for automotive intake manifold application. In *SAE Technical Paper Series*. SAE International, 400 Commonwealth Drive, Warrendale, PA, United States.

[36] Jasper Vermeulen, Alan Burden, Glenda Caldwell, Müge Belek Fialho Teixeira, and Matthias Guertler. 2025. Investigating human factors in Mako-assisted total knee arthroplasty surgeries. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Melbourne, Australia). IEEE, 1710–1715.

[37] Jasper Vermeulen, Glenda Caldwell, Müge Teixeira, Alan Burden, and Matthias Guertler. 2024. To safety and beyond! A scoping review of human factors enriching the design of human-Robot Collaboration. *Interact. Des. Archit.(S)* 61 (June 2024), 42–65.

[38] Jiaxin Xu, Chao Zhang, Raymond H Cuijpers, and Wijnand A IJsselsteijn. 2025. Does care lead to bonds? Exploring the relationship between human caregiving for robots and human-robot bonding. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan). ACM, New York, NY, USA, 1–15.

[39] Xinyan Yu, Yiyuan Wang, Tram Thi Minh Tran, Yi Zhao, Julie Stephany Berrio Perez, Marius Hoggenmüller, Justine Humphry, Lian Loke, Lynn Masuda, Callum Parker, Martin Tomitsch, and Stewart Worrall. 2023. Robots in the wild: Contextually-adaptive human-robot interactions in urban public environments. In *Proceedings of the 35th Australian Computer-Human Interaction Conference* (Wellington New Zealand). ACM, New York, NY, USA, 701–705.