Detecting Functional Performance Deterioration in Activities of Daily Living Using Conditional Preference Networks and Sequential Pattern Mining

by

Marian Enam Esi Kushigbor

Previous Degrees
Bachelor of Science in Computer Engineering, University of Ghana, 2023

A Thesis Submitted in Partial Fulfilment of
the Requirements for the Degree of

Master of Computer Science

In the Graduate Academic Unit of Computer Science

Supervisors: Michael Fleming, Ph.D., Computer Science
Scott Buffett, Ph.D., Computer Science

Examining Board: Aaron Tabor, Ph.D., Computer Science, Chair
Suprio Ray, Ph.D., Computer Science
Erik Scheme, Ph.D., Electrical and Computer Engineering

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

June, 2026

© Marian Enam Esi Kushigbor, 2026

Abstract

Aging in place depends on maintaining the ability to perform Activities of Daily Living (ADLs), yet clinicians can miss early functional decline because assessments are infrequent, subjective, and rarely capture performance in natural settings. This thesis examines whether a skeleton-based sensing pipeline can detect subtle deterioration within an individual by comparing ADL execution under baseline and aging-simulated conditions. Two Orbbec Femto depth cameras recorded ADL trials and produced 3D skeletons with 32 joints. Temporal encoding converted joint trajectories into activity sequences for two evaluators: (1) Sequential Pattern Mining in the SCOUT codebase, which outputs confidence scores that quantify similarity to baseline, and (2) a Conditional Preference Network model of baseline movement preferences, which uses ordering queries to compute an ADL-fit score. Across no-, low-, medium-, and high-gear conditions and multiple ADLs, both evaluators reflected expected degradation with increasing gear while revealing participant- and activity-specific variability. However, frame loss and limited per-participant training data reduced sensitivity.

Dedication

To Gloria Doegah, whose warmth, laughter, and spirit live on in the hearts of all who knew her. This work is dedicated to your memory.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Prof. Michael Fleming and Prof. Scott Buffett, for their guidance, patience, and invaluable feedback throughout this journey. Their support and expertise were instrumental in shaping this work.

I am deeply grateful to my parents and siblings for their unwavering encouragement and support. Their belief in me made all the difference, especially during the more challenging moments of this process.

Finally, I would like to thank my friends for their company, encouragement, and for always being there when I needed them most.

Table of Contents

Abstract ii

Dedication iii

Acknowledgments iv

Table of Contents v

List of Tables x

List of Figures xiii

Abbreviations xv

1 Introduction 1

1.1 Introduction 1

1.2 Research Questions 3

1.3 Main Contributions 4

1.4 Declaration of Generative AI Use 5

1.5 Thesis Outline 5

2 Background and Related Work 7

2.1 Activities of Daily Living 8

2.1.1 Age Simulation Suits and Their Role in ADL Assessment . . . 9

2.2 Human Activity Recognition And Sensor Methods 11

2.2.1 HAR Using Wearable Sensors for Monitoring Older Populations 11
2.2.2 Deep Learning Models in Health-Oriented HAR 12
2.2.3 Vision-Based and Multimodal HAR in Assisted Settings . . . 13
2.3 Behavioural Drift Detection and Smart Home Analytics 14
2.3.1 Mining-Based Approaches for Detecting Behavioural Change . 15
2.3.2 Clustering and Trend-based Monitoring 15
2.3.3 Context-Aware and Anomaly-Oriented Frameworks 16
2.4 Sequential Pattern Mining 17
2.4.1 Introduction 17
2.4.2 SPM for Smart Home Activity modeling and Monitoring . . . 19
2.4.3 Recognizing Complex and Interleaved Activities 19
2.4.4 Context-Aware and Targeted SPM Approaches 20
2.5 Conditional Preference Networks (CP-nets): Extensions and Theory . 22

3 Methodology **27**
3.1 Overview 27
3.2 Feature Extraction and Representation 31
3.2.1 Sequence Formation and Preprocessing 31
3.2.2 Discretisation 32
3.3 Sequential Pattern Mining (SPM) 34
3.3.1 SPM (Unconditional) 34
3.3.2 Conditional Sequential Pattern Mining (CSPM) 37
3.4 CP-net Modelling and Evaluation 42
3.4.1 Construction 42
3.4.2 Pairwise Comparison of Outcomes 47
3.4.3 Example: Evaluating test Sequences (Outcomes) with the CP-net 47

3.4.3.1 CP-net Traversal for Outcome Comparison Using Ordering Queries Theorem 50
3.4.3.2 Outcome Comparison Using Weighted Hybrid Voting 52
3.4.3.3 Outcome Comparison Using Global Heuristic Scoring 53
3.4.4 Dominance Testing Using Improving Flips 55
3.5 Cross-Dataset Comparison and Win-rate Computation 56

4 Results and Discussion 58

4.1 Introduction to the Results 58
4.2 Activities and Experimental Conditions 60
4.2.1 Data Preparation and Parameter Settings 61
4.3 RQ1: Evaluating ADL Fit Using SCOUT (SPM Classifier) 62
4.3.1 Understanding the Class 0 Confidence Scores 63
4.3.2 Trends Across Participants and Activities For SCOUT'S SPM 65
4.3.2.1 Walking 66
4.3.2.2 Vacuuming 66
4.3.2.3 Laundry 67
4.3.2.4 Sit-To-Stand (STS) 67
4.3.3 Statistical analysis of gear-level effects on SCOUT's confidence 68
4.3.3.1 Walking: gear-level effects on SCOUT Class 0 confidence 71
4.3.3.2 Vacuuming: gear-level effects on SCOUT Class 0 confidence 72
4.3.3.3 Laundry: gear-level effects on SCOUT Class 0 confidence 73
4.3.3.4 Sit-to-stand (STS): gear-level effects on SCOUT Class 0 confidence 75
4.4 RQ2: Evaluating ADL Fit Using CP-nets 77

4.4.1 Understanding Win Rates in CP-net Evaluation 79
4.4.2 Trends Across Participants and Activities For CP-nets 79
4.4.2.1 Walking 81
4.4.2.2 Vacuuming 81
4.4.2.3 Laundry 82
4.4.2.4 Sit-to-Stand (STS) 83
4.4.3 Statistical analysis of gear-level effects for CP-nets 84
4.4.3.1 Walking: gear-level effects on CP-net win rates . . . 84
4.4.3.2 Vacuuming: gear-level effects on CP-net win rates . . 86
4.4.3.3 Laundry: gear-level effects on CP-net win rates . . . 87
4.4.3.4 Sit-to-stand (STS): gear-level effects on CP-net win rates 88
4.4.4 Relationship Between CP-net Structure and Result Quality . . 90
4.4.4.1 CP-net Connectivity and Results Quality 91
4.4.4.2 Number of Conditional Nodes and Results Quality . 93
4.4.4.3 Average Node Degree and Results Quality 94
4.4.4.4 Summary of Structural Correlation Analyses 95
4.5 Discussion of Findings 96
4.5.1 Comparing SCOUT and CP-net Performance 97
4.5.2 Findings for RQ1: Evaluating ADL Fit with SCOUT 98
4.5.3 Findings for RQ2: Evaluating ADL Fit with CP-nets 99
4.6 Limitations of the Study 99
4.7 Conclusion of the Results Chapter 103

5 Conclusion and Future Work **106**

5.1 Key Findings 106
5.2 Limitations 109
5.3 Future Work 110

5.4 Final Remarks 112

A CP-net Solver: Pseudocode and Algorithms 122

A.1 Algorithm Descriptions 122

A.1.1 Pairwise Outcome Comparison (`determineIsBetterFit`) . . . 122

A.1.2 Cross-Dataset Win-Rate Computation 125

B Post Hoc Pairwise Tests: SCOUT Confidence Scores 126

B.1 Walking 126

B.2 Vacuuming 127

B.3 Laundry 127

B.4 Sit-to-Stand (STS) 128

C Post Hoc Pairwise Tests: CP-net Win Rates 129

C.1 Walking 129

C.2 Vacuuming 130

C.3 Laundry 130

C.4 Sit-to-Stand (STS) 131

Bibliography 122

Glossary 132

Vita

List of Tables

3.2.1 Excerpt from a preprocessed skeletal CSV file for the laundry-loading activity. Each row corresponds to one frame, and each column stores a 3D coordinate for a specific joint. 32

3.4.1 Feature presence vector for Outcome 0 ($S_0 = \langle A, B, D, E \rangle$). 48

3.4.2 Feature presence vector for Outcome 1 ($S_1 = \langle A, B, D, C, E \rangle$). 49

4.3.1 Class 0 confidence scores for all participants across the four activities and gear levels. Higher scores indicate closer similarity between a test sequence and the baseline (no-gear) activity model learned by SCOUT. These values form the basis for identifying declines in ADL fit as movement restriction increases. 64

4.3.2 Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for walking by gear level. 71

4.3.3 Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for vacuuming by gear level. 72

4.3.4 Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for laundry by gear level. 74

4.3.5 Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for sit-to-stand (STS) by gear level. 75

4.4.1 CP-net win rates for all participants across the four activities and gear levels. Win rates represent the proportion of pairwise comparisons in which a test sequence is deemed a better fit over others according to the learned CP-net, indicating its relative similarity to baseline movement patterns. 80
4.4.2 Descriptive statistics and mean ranks for CP-net win rates for walking by gear level. 85
4.4.3 Descriptive statistics and mean ranks for CP-net win rates for vacuuming by gear level. 86
4.4.4 Descriptive statistics and mean ranks for CP-net win rates for laundry by gear level. 87
4.4.5 Descriptive statistics and mean ranks for CP-net win rates for sit-to-stand (STS) by gear level. 89

B.1.1 Dunn–Bonferroni pairwise comparisons for walking SCOUT Class 0 confidence scores. 126
B.2.1 Dunn–Bonferroni pairwise comparisons for vacuuming SCOUT Class 0 confidence scores. 127
B.3.1 Dunn–Bonferroni pairwise comparisons for laundry SCOUT Class 0 confidence scores. 127
B.4.1 Dunn–Bonferroni pairwise comparisons for sit-to-stand (STS) SCOUT Class 0 confidence scores. 128

C.1.1 Dunn–Bonferroni pairwise comparisons for CP-net win rates for walking. 129
C.2.1 Dunn–Bonferroni pairwise comparisons for CP-net win rates for vacuuming. 130
C.3.1 Dunn–Bonferroni pairwise comparisons for CP-net win rates for laundry. 130

C.4.1Dunn–Bonferroni pairwise comparisons for CP-net win rates for sit-to-stand (STS). 131

List of Figures

2.1 CP-net for Cup preference example 22

2.2 Induced preference graph 24

3.1 Methodology workflow: from multi-condition data collection to CP-net evaluation and dataset ranking. 30

3.2 Example CP-net constructed from eight features. Directed edges encode conditional dependencies. CPT statements next to each node show the corresponding likelihood rules. 46

4.1 Participant wearing each of the four gear conditions: (a) no gear, (b) low gear, (c) medium gear, and (d) high gear. 61

4.2 Mean SCOUT Class 0 confidence scores across gear levels for all four activities. 77

4.3 Mean CP-net win rates across gear levels for all four activities. 89

4.4 CP-net connectivity versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression. 92

4.5 Number of conditional nodes versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression. 94

4.6 Average node degree versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression. 95

List of Abbreviations

ADL	Activity of Daily Living
BDD	Behavioral Drift Detection
BFS	Breadth-First Search
CP-nets	Conditional Preference Networks
CPT	Conditional Preference Table
CPU	Central Processing Unit
CSV	Comma-Separated Values
CSPM	Conditional Sequential Pattern Mining
GPU	Graphics Processing Unit
HAR	Human Activity Recognition
IADL	Instrumental Activity of Daily Living
NRC	National Research Council of Canada
SPM	Sequential Pattern Mining
STS	Sit-to-Stand

Chapter 1

Introduction

1.1 Introduction

The world's aging population is growing, and many older adults want to age in their own homes longer. To achieve this, they need to maintain the ability to perform Activities of Daily Living (ADLs). ADLs are basic tasks an individual needs to be able to perform in order to live independently. Subtle declines in how people perform ADLs can signal early changes in functional health, even before they escalate into dramatic events such as falls or hospitalisations. Clinicians usually measure ADL ability with questionnaires, interviews, or performance-based tests in clinics. However, these assessments are infrequent, subjective, and performed outside the person's real home context. As a result, small, gradual changes in how someone performs an activity often go unnoticed until more serious problems occur.

Recent research in Human Activity Recognition (HAR) uses sensors or cameras to automatically recognise and analyse activities. However, many HAR systems focus on classifying which activity is being performed rather than how well it is being performed. Some work looks at detecting anomalies or routine changes, but it still

often relies on wearable sensors or devices on the body. To address limitations like this, the National Research Council of Canada (NRC) is working on an Aging in Place initiative, which is a broader effort to design technologies that support older adults in their own homes. The aim of the initiative is to detect and track age-related changes without invasive body-worn sensors, using environment-based sensing such as depth cameras that can observe movement without attaching devices to anyone. The long-term goal is a lightweight, unobtrusive monitoring system that can flag meaningful changes in how a person performs everyday tasks, thus providing early warning of functional decline before it escalates into a fall, hospitalisation, or loss of independence and helping older adults live safely in their own homes for as long as possible. This thesis represents a first step toward that goal, investigating the feasibility of the approach under controlled laboratory conditions. The broader initiative encompasses multiple research streams, including complementary work on fall prevention and clinical assessment. This thesis focuses specifically on the home-monitoring strand, investigating whether pattern mining and preference-based reasoning can detect subtle changes in ADL performance from depth-camera skeletal data.

In this thesis, depth cameras such as Orbbec cameras are used to capture 3D skeletal data in a simulated apartment while participants perform common at-home activities. To approximate the effects of aging, participants wear different levels of aging-simulation gear. This approach allows controlled, repeatable conditions across participants, something that would be difficult to achieve with real aging trajectories where decline varies widely between individuals and unfolds over years rather than a single session. Participants perform selected ADLs under no-gear (baseline), low-gear, medium-gear, and high-gear conditions. This provides controlled levels of performance degradation to be analysed.

Most HAR systems seek to answer which activity is being performed. In contrast, this thesis seeks to find out how close a given performance is to a healthy baseline performance. This is termed ADL fit. The thesis explores whether this can be achieved with pattern mining (Sequential Pattern Mining, SPM) and preference reasoning (Conditional Preference Networks, CP-nets) to capture subtle differences in performance across gear levels. The SPM and Conditional SPM components are implemented within SCOUT, an existing Java codebase developed at the National Research Council of Canada. As part of this thesis, a CP-net evaluation module was developed and integrated into SCOUT to enable preference-based ADL fit comparisons. The main aim of this thesis is to investigate whether a hybrid framework that combines SPM and CP-nets can detect and quantify subtle changes in ADL performance across simulated aging conditions using non-invasive depth-camera motion data.

1.2 Research Questions

This thesis investigates whether a hybrid framework that combines Sequential Pattern Mining (SPM) and Conditional Preference Networks (CP-nets) can be used to measure ADL fit from non-invasive depth-camera motion data collected under simulated aging conditions. The work is guided by the following research questions:

RQ1: Can SCOUT's SPM classifier detect declines in ADL fit as participants perform everyday activities under higher levels of simulated aging gear?

RQ2: Can CP-net ordering queries identify performance deterioration by ranking lower-gear sequences as better fits to baseline movement than sequences recorded under higher-gear levels?

1.3 Main Contributions

This thesis makes the following main contributions:

- **Definition of ADL fit and hybrid evaluation framework.** The thesis defines ADL fit as the degree to which a test activity sequence resembles baseline (no-gear) performance and proposes a hybrid framework that combines pattern-based motion analysis (Sequential Pattern Mining, SPM) [45] with preference-based reasoning (Conditional Preference Networks, CP-nets) [9] to quantify this fit from depth-camera skeletal data.

- **Integration of CP-net evaluation into the SCOUT framework.** This thesis developed and integrated a CP-net evaluation module into the existing SCOUT codebase, introducing a novel use of CP-nets in which patterns learned from baseline (no-gear) and maximally restricted (high-gear) recordings are used in pairwise ordering queries to determine which of two test sequences is closer to the baseline movement profile. Four outcome-comparison methods were implemented within the solver: CP-net traversal using ordering queries, weighted hybrid voting, global heuristic scoring, and dominance testing via improving flips.

- **Experimental framework under simulated aging conditions.** The thesis presents an experimental setup in a simulated apartment, in which Orbbec depth cameras record 3D skeletal data while participants perform common at-home activities under progressively restrictive age-simulation gear (no-gear, low-gear, medium-gear, and high-gear). This setup provides controlled levels of performance degradation for evaluating ADL fit.

- **Empirical evaluation and structural analysis.** Through multiple experiments, the thesis evaluates how well the SPM-based and CP-net-based

evaluators distinguish between gear levels and how closely they match the expected decline in ADL fit. In addition, the structural properties of the learned CP-nets, namely overall connectivity, number of conditional nodes, and average node degree, are analysed in relation to the quality of the gear-level ordering results. The analysis highlights both promising trends and important limitations, and identifies concrete directions for future work, including improvements to the sensing pipeline, the cold-start labelling problem, and the path toward a deployable prototype.

1.4 Declaration of Generative AI Use

Generative AI tools were used during the writing of this thesis to assist with improving the clarity and readability of selected passages. All AI-assisted text was reviewed, revised, and verified by the author to ensure accuracy and that it faithfully represents the original research and intent. The research design, data collection, analysis, implementation, and conclusions are entirely the author's own work.

1.5 Thesis Outline

The remainder of this thesis is organised as follows.

Chapter 2 presents the background and related work. It reviews key concepts on aging and Activities of Daily Living (ADLs), sensor-based Human Activity Recognition (HAR), Sequential Pattern Mining (SPM), and Conditional Preference Networks (CP-nets). The chapter also discusses previous approaches to detecting behavioural drift and monitoring functional decline and positions the present work within this literature.

Chapter 3 describes the methodology used in this thesis. The chapter explains how 3D skeletal data are collected with Orbbec depth cameras, preprocessed, and transformed into symbolic sequences. It then outlines the SPM and CSPM pipeline within SCOUT, the procedure for learning CP-nets from the activity data, and the way ADL fit is defined and evaluated across gear levels.

Chapter 4 presents the experimental results and their analysis. It reports the performance of the SPM-based evaluator and the CP-net-based evaluator for each activity and gear level, and examines whether the expected decline in ADL fit appears as the simulated aging conditions become more restrictive. The chapter discusses both consistent and irregular patterns in the results, relates them to the underlying data characteristics, and uses these findings to address the research questions.

Chapter 5 concludes the thesis. It summarises the key findings and main contributions, interprets the results in light of the study limitations, and discusses the path toward a deployable system that can support older adults in living independently at home. The chapter outlines concrete directions for future work, including improvements to the sensing pipeline, participatory design with older adults, addressing the cold-start labelling problem, and the incremental retraining approach that would allow the system to adapt to individual users over time.

Chapter 2

Background and Related Work

This chapter reviews existing research on detecting and monitoring Activities of Daily Living (ADLs), focusing on the various methods used in care and health support systems for older individuals to promote independent living and aging in place. This involves checking in on how people manage everyday tasks. Small changes in how someone handles these activities can be a sign, signalling potential health issues like early dementia, arthritis, or increasing frailty. Spotting these changes early on means doctors and caregivers can better manage these conditions, improving the person's quality of life and slowing down further decline. The review explores key areas influencing ADL recognition and detection, including traditional clinical assessment tools, sensor-based Human Activity Recognition (HAR) systems, behavioural drift detection methods, and Sequential Pattern Mining (SPM) for activity analysis. Finally, the chapter introduces Conditional Preference Networks (CP-nets), a preference modeling framework that can be adapted for ADL recognition and ADL decline.

2.1 Activities of Daily Living

Activities of Daily Living (ADLs) refer to all the daily tasks a person should be able to perform for independent living. Basic ADLs refer to self-care activities such as bathing, dressing, and eating. In contrast, instrumental Activities of Daily Living (IADLs) include more complex tasks like managing finances, preparing meals, and maintaining a household [15]. ADL assessments are really important for healthcare decisions, figuring out care plans, and deciding who gets long-term support. How well someone can do these everyday tasks tells us a lot about how independent they are.

Pashmdarfar *et al.* [41] reviewed over twenty ADL assessment tools, including widely used instruments such as the Barthel Index, Katz Index, and Lawton IADL Scale. While these tools are foundational in clinical and caregiving environments, they often lack strong psychometric validation and are not designed for real-time or digital monitoring. As smart technologies become more common in healthcare, it is clear that we need better ways to evaluate them so they can adapt and respond to changing needs. The integration of smart technologies is transforming healthcare, making it more efficient, convenient, and personalized.

Edemekong *et al.* [15] emphasize that ADL assessments, while essential, often rely on self-reporting and infrequent evaluations. This makes it difficult to track gradual or subtle changes in the person's ability to function independently. The limitations of such approaches underline the need for continuous and objective monitoring systems that reflect changes over time.

Kang *et al.* [27] present a smart system called the Smart Plug Hub (SPH), which uses ambient sensor fusion and edge-level machine learning to identify ADLs in home settings. While the system effectively recognizes activities, it does not yet account for drift in ADL performance over time. These factors are becoming more relevant in

the development of truly adaptive care systems.

In all, these studies show that while traditional and emerging ADL assessment tools offer valuable insights, there remains a need for systems that not only monitor activity accurately but also adapt to the personal routines and evolving behaviours of the individuals being supported.

2.1.1 Age Simulation Suits and Their Role in ADL Assessment

Age simulation suits give people the chance to experience some of the challenges that come with getting older, like changes in how they walk, move, see and hear. These suits include things like weighted clothing and goggles that blur vision. People can get a better understanding of how aging affects everyday tasks and physical abilities. When worn together, they mimic age-related changes in gait, dexterity, posture, and perception. As such, they offer a controlled means of studying functional decline and its impact on Activities of Daily Living (ADLs).

Several studies have shown that simulation suits can induce measurable impairments in physical performance, particularly in tasks relevant to daily functioning. For example, Timm *et al.* [47] used the GERT suit to simulate age-related impairments in posture, strength, and motor control, reporting notable declines in walking speed, coordination, and fine motor tasks such as writing. Similarly, Tian *et al.* [46] and Lauenroth *et al.* [32] observed significant changes in gait mechanics, including reductions in stride length, arm swing, and gait velocity, metrics that closely mirror those found in actual older populations. These findings help validate the use of simulation suits as realistic proxies for aging in experimental research.

Beyond isolated movement metrics, simulation suits have been used to assess how

aging affects the execution of full ADL sequences. Studies by Lee and Teh [33] and Nelson *et al.* [40] highlighted how participants, when tasked with common self-care activities such as hair brushing, dressing, or walking with objects, exhibited longer completion times and required more effort to maintain balance and coordination. Such behavioural changes align with real-world patterns of functional decline and offer valuable insight into how seemingly simple tasks become fragmented or delayed with age.

The psychological effects of wearing these suits are also well documented. Researchers, including Coelho *et al.* [10] and Gerhardy *et al.* [19], found that users reported increased empathy and awareness of limitations faced by older populations after performing basic ADLs in simulation gear. These studies, primarily conducted in healthcare training environments, have shown that simulation can foster attitudinal change, especially in future caregivers. However, while most studies focus on the experiential or educational value, fewer address objective tracking of task degradation, leaving room for data-driven extensions.

Some works have attempted to bridge this gap by combining observational and quantitative performance measures. Gerhardy *et al.* [18] directly compared the movement of older adults to that of young participants in simulation suits, confirming that performance indicators such as postural sway, joint mobility, and response time decline under both conditions. These validation efforts support the use of suits as experimental surrogates for studying age-related behavioural drift.

In the context of this thesis, the use of age simulation suits serves two primary roles: first, as a method to systematically simulate and capture functional decline across multiple ADL scenarios; and second, as a foundation for comparing variations in task performance under different levels of simulated aging. This enables the study to examine not only whether an activity is performed correctly, but how its

execution deteriorates or adapts across physical constraints, an important step toward understanding behavioural drift and ADL fit in older adults.

2.2 Human Activity Recognition And Sensor Methods

Human Activity Recognition (HAR) is a foundational component of pervasive health monitoring systems, especially in supporting older adults through ambient assisted living, rehabilitation, and safety-oriented interventions. Recent advances have led to HAR systems that leverage wearable sensors, vision-based tracking, and machine learning to detect common Activities of Daily Living (ADLs). However, while many studies demonstrate high accuracy in classifying activity types, few consider the quality or consistency of activity execution, which is a critical aspect in detecting performance deterioration associated with aging or physical decline. This section reviews recent HAR research relevant to care and health monitoring for older people, with emphasis on models, sensing modalities, and system goals.

2.2.1 HAR Using Wearable Sensors for Monitoring Older Populations

A number of studies focus on using wearable sensor data to develop lightweight, yet effective HAR systems for older individuals. For instance, Deeptha *et al.* [14] designed a mobile and wearable sensor-based HAR framework that integrates data from accelerometers, gyroscopes, and magnetometers to classify daily activities. The authors prioritized model simplicity and real-time inference, enabling their system to operate efficiently on smartphones and embedded platforms, an important

consideration for eldercare applications. Similarly, Tian *et al.* [46] developed a single-accelerometer HAR framework using wavelet energy spectrum features and ensemble feature selection. Their system achieved high recognition accuracy with minimal sensor load, demonstrating suitability for long-term monitoring applications in aging populations.

While these approaches are promising for continuous monitoring, they primarily focus on detecting activity labels rather than changes in performance over time. They also rely on predefined activity templates and often lack mechanisms to identify subtle intra-individual drift or physical strain, which may be indicative of decline in functional ability.

2.2.2 Deep Learning Models in Health-Oriented HAR

Deep learning has significantly improved the performance of HAR systems, especially in handling noisy sensor data and complex activity transitions. Uddin and Soylu [48] developed a neural-structured HAR model using wearable sensor data, combining kernel discriminant analysis with LSTM networks. Their system achieved high recall and incorporated model interpretability tools, enhancing transparency in health monitoring contexts.

Likewise, Hayat *et al.* [24] compared CNN, LSTM, and Random Forest classifiers using smartphone sensors in monitoring scenarios for the older population. LSTM models performed best in recognizing temporal dependencies, though they struggled with ambiguity during activity transitions, a key limitation when assessing unstable or compensatory motion in aging populations.

Zeng *et al.* [55] proposed a deep residual CNN model for multichannel wearable data, improving recognition across variable user conditions. Although their model

generalized well to diverse users, it lacked personalized calibration mechanisms that could detect intra-individual decline across time.

In general, deep learning models show strong performance in activity classification, but they rarely address the issue of how activities are performed, a critical gap for systems aiming to monitor progressive changes in movement quality.

2.2.3 Vision-Based and Multimodal HAR in Assisted Settings

Beyond wearable sensors, some researchers have explored vision-based and multi-sensor approaches for HAR in smart home or care settings. The PriMA-Care dataset project [7] introduced a privacy-preserving multimodal dataset combining RGB-D video, audio, and sensor data for older activity analysis. While valuable for training HAR models in realistic environments, vision-based methods often face deployment challenges due to privacy concerns, occlusion, and environmental sensitivity.

Khan and Sohn [28] developed a deep learning framework for abnormal activity recognition in older participants using visual input and joint tracking. Their system effectively detected acute incidents such as falls, but did not assess routine activity performance or progression of physical decline.

Multimodal HAR approaches like these offer enhanced context awareness, yet they remain primarily geared toward classification and safety alerts rather than longitudinal health monitoring or performance drift detection.

Recent HAR research has made substantial progress in activity classification accuracy, sensor efficiency, and real-time processing. However, most systems remain focused on identifying which activity is performed, rather than evaluating how well it is performed over time, especially under changing physical conditions like aging. While

some studies address interpretability, personalization, or health monitoring, few explicitly model intra-individual performance degradation.

This thesis addresses this gap by using a symbolic HAR framework to track subtle performance shifts in ADLs under simulated aging conditions. Unlike conventional HAR systems, it evaluates how activity execution varies across gear levels, enabling detection of functional decline through sequence-based modeling and comparative preference structures.

2.3 Behavioural Drift Detection and Smart Home Analytics

As the aging population grows, long-term monitoring of behavioural patterns has become an essential component of smart home systems aimed at supporting independent living. Behavioural drift detection, in particular, focuses on identifying subtle or progressive changes in activity routines that may indicate cognitive or physical decline. This includes both changes in the composition of daily activity routines, such as shifts in the frequency or sequencing of activities, and declines in the quality or performance of individual activities, such as slower execution or increased difficulty completing tasks. In practice, these two forms of drift are often interrelated, as deteriorating activity performance may gradually reshape an individual's overall routine. Recent studies have approached this problem using various techniques, including activity pattern mining, clustering, anomaly detection, and context-aware modeling. This section reviews relevant work that contributes to the development of intelligent systems for monitoring and detecting behavioural drift.

2.3.1 Mining-Based Approaches for Detecting Behavioural Change

Several studies employ activity pattern mining to model and monitor behaviour over time. Veronese *et al.* [49] present a framework that mines frequent activity sequences to model normal routines and tracks deviations from these patterns to identify behavioural drift. Their method captures both structural and statistical changes in behaviour, enabling the system to detect early signs of decline more effectively than traditional anomaly detection approaches.

Similarly, Azefack *et al.* [6] propose a layered behavioural drift detection method that clusters daily activity distributions and applies statistical change detection to monitor long-term routine shifts. Their method demonstrates how clustering combined with drift metrics can uncover gradual changes in user behaviour that may not appear as immediate anomalies but signal longer-term health trends.

Shahid *et al.* [44] take a more threshold-based approach, modeling expected routines and flagging deviations, such as missed meals or extended periods of inactivity, as potential concerns. Although simpler in its detection mechanism, the study highlights the real-world applicability of smart homes for identifying early signs of behavioural deterioration in older individuals.

2.3.2 Clustering and Trend-based Monitoring

Some researchers have explored clustering and trend analysis to uncover behaviour drift. Zekri *et al.* [54] propose a framework that combines activity classification with unsupervised clustering to monitor aging in place. By tracking significant shifts in activity clusters over time, their system identifies changes in routine that may reflect transitions in health status.

Sepesy Maučec and Donaj [43] contribute a similar trend-based approach, transforming raw sensor sequences into structured activity sequences and applying Partitioning Around Medoids (PAM) for clustering. They measure behavioural consistency using generalized distance metrics and demonstrate how their method can differentiate between stable and changing activity patterns across multiple users.

These clustering-based approaches complement pattern mining techniques by providing an aggregated view of behaviour over time, making them especially useful for recognizing gradual and personalized forms of behavioural drift.

2.3.3 Context-Aware and Anomaly-Oriented Frameworks

Other studies incorporate contextual information and anomaly detection frameworks to improve sensitivity and interpretability. Zerkouk and Chikhaoui [56] introduce a hierarchical anomaly detection system that combines low-level sensor data with high-level activity context. By modeling both expected sensor events and their activity implications, the system reduces false positives and improves anomaly classification in aging-in-place scenarios.

Enshaeifar *et al.* [16] apply a hybrid machine learning framework to detect anomalies in the behaviour of individuals with dementia. Their system uses a combination of decision trees and clustering to assess routine stability, demonstrating the value of adaptive systems in early detection of cognitive decline.

Czabke *et al.* [12] focus on practical system design, developing a modular telemedicine platform that integrates ADL monitoring with vital sign tracking. Although not strictly an anomaly detection system, their platform emphasizes ease of use for older populations and supports early detection through the integration of diverse behavioural data streams

2.4 Sequential Pattern Mining

2.4.1 Introduction

Sequential Pattern Mining (SPM) is a data mining technique for finding statistically significant sequences of events or actions that occur in an ordered manner within a dataset. It was introduced by Agrawal and Srikant in 1996 [45] for analyzing customer transaction data and has since been applied to various fields such as web page usage analysis, manufacturing process monitoring, and Human Activity Recognition (HAR) [39]. In HAR, each sequence represents a series of symbolic actions or sensor events ordered over time.

SPM relies on several key concepts. A sequence database refers to a collection of *sequences*, where each sequence is an ordered list of events or actions. It is the primary input to SPM [45, 17]. An *item* is the smallest unit of observation, while an *itemset* is a set of items that occur together in a single time step. *Support* measures how often a sequence/pattern occurs across the sequence database. A pattern is considered *frequent* if it meets or exceeds the minimum support threshold of the SPM [45]. To limit the search space and ensure temporal relevance, temporal constraints such as maximum gap and maximum window restrict the spacing or total span of pattern elements.

Several algorithms offer different approaches to SPM. Early methods like the Generalized Sequential Pattern (GSP) algorithm follow an Apriori-style, level-wise search. GSP generates candidate patterns sequentially and prunes those that fail the support threshold [25]. While conceptually simple, candidate generation can become computationally expensive for large datasets.

Pattern-growth methods avoid candidate explosion by projecting the database based on discovered prefixes. PrefixSpan is a prominent algorithm of this type; it recursively

explores relevant subsequences while avoiding full candidate generation [22]. Vertical-format algorithms like SPADE represent the database as sets of transaction identifiers per item, which allows efficient operations on dense datasets [53].

The SPAM algorithm [5], used in this research, combines a vertical bitmap representation with depth-first search. Each sequence is encoded as a bitmap to enable efficient, bitwise support counting. SPAM extends patterns through two operations: S-step, appending a new item in a later time step, and I-step, appending a new item within the current time step. This combination allows efficient mining of long, dense sequences.

To reduce redundancy, some mining approaches focus on closed and maximal patterns. Closed patterns lack a super-pattern with the same support, while maximal patterns cannot be extended without falling below the support threshold [51, 52]. In classification-oriented applications, such as this thesis, discriminative sequential patterns are also important. These patterns occur at significantly different frequencies across classes and are identified using statistical measures such as chi-square testing. This allows the system to identify patterns that best separate different classes of sequences, such as those collected from baseline trials and trials with simulated aging gear.

In this study, SPM mines both frequent and discriminative patterns from baseline activity sequences, which serve as a performance reference. Sequences collected under different levels of aging-simulation gear are compared to this baseline. Changes in pattern coverage and class separation indicate performance drift, enabling the detection of subtle declines in execution, even when the overall activity category remains consistent.

2.4.2 SPM for Smart Home Activity modeling and Monitoring

Several studies have explored how SPM can model daily routines in smart environments, particularly in support of independent living among older adults. For instance, Rashidi *et al.* [42] introduced a method that segments raw sensor data into activity sessions, mines frequent sequences using PrefixSpan, and compares current routines against historical patterns via a Monitoring Pattern Tree (MPT). This allows the system to identify deviations in routine, such as prolonged inactivity or skipped actions, which may indicate early signs of cognitive or physical decline. Similarly, Virone *et al.* [50] proposed monitoring activity patterns and trends of older adults by applying SPM and trend-aware comparison techniques to detect abrupt or gradual changes in behaviour. These approaches demonstrate how SPM extends beyond activity recognition into behavioural monitoring, aligning closely with the thesis goal of detecting performance drift over time.

In a related direction, the work of Aizan *et al.* [1] applies SPM enriched with temporal features to improve ADL classification. By transforming raw sensor data into temporal sequences, extracting frequent patterns, and combining these with temporal features in a Random Forest classifier, their system significantly improved recognition accuracy. This reinforces the benefit of combining symbolic sequence mining with additional contextual signals, such as time duration and ordering, for modeling smart home activities.

2.4.3 Recognizing Complex and Interleaved Activities

In real-world scenarios, individuals often perform activities that are not neatly segmented or sequential, making complex activity recognition a critical area of

research. A key example is the Action2Activity framework [37], which models high-level activities as sequences of atomic actions. Rather than treating activities as flat labels, the system uses SPM to discover frequent action patterns per class and then applies probabilistic inference to identify the most likely activity from the observed action sequence. While it doesn't explicitly model interleaved activities, the layered representation improves the system's ability to distinguish complex behaviours composed of shared sub-actions.

Building on this idea, emerging patterns (EP) have been leveraged for activity recognition in environments where activities are interleaved or concurrent. Gu *et al.* [20] applied EP mining to discover patterns that occur more frequently in one activity class than others, encoding them into feature vectors used by standard classifiers. While this method still predicts a single activity per window, another study by Gu [21], Emerging Patterns based Approach to Sequential Interleaved and Concurrent Activity Recognition (epSICAR), extends this by using binary classifiers per activity, allowing the system to label multiple overlapping activities in the same time window. This makes epSICAR more robust for real-world smart home scenarios where activities often overlap, offering a more flexible solution than traditional multi-class approaches.

2.4.4 Context-Aware and Targeted SPM Approaches

Recent work has also explored enhancing SPM through context awareness and targeted mining. SAAP (Situation-Aware Activity Projection) [13], for example, integrates situational context into the SPM process to predict what activity is most likely to occur next. The framework includes a context-aware SPM module and a probabilistic projection model that considers both historical behaviour and the current state. This enables more personalized and responsive activity modeling, especially in dynamic

environments.

Complementing this, targeted SPM approaches focus on mining patterns that satisfy specific relevance constraints. For example, Liang *et al.* [35] propose mining only those patterns that are contextually linked to a specific outcome, incorporating prior knowledge to narrow the search space. Huang *et al.* [26] introduce time-interval constraints to capture the spacing between events, producing more interpretable and domain-relevant patterns. While these methods are not always designed specifically for HAR, their adaptability and focus on meaningful pattern filtering make them highly applicable for monitoring behavioural changes or flagging specific functional markers.

Finally, the work by Hassani *et al.* [23] offers a broader reflection on the utility of SPM primitives in process discovery. While more conceptual in nature, this work highlights the potential for adapting SPM techniques to model real-time behaviours and generate online process representations, a direction particularly relevant to systems that track behavioural drift continuously.

Most systems track routines across multiple people or focus on activity types. Few of these approaches detect changes in how a task is done. None of them evaluate decline within a single person over time. They miss slower movement, instability, or compensation.

This thesis addresses those gaps. SCOUT is an existing Java codebase developed at the National Research Council of Canada that uses symbolic SPM and CSPM to extract discriminative movement patterns and classify activity sequences. As part of this thesis, a CP-net evaluation module was added to SCOUT to enable preference-based comparison of activity sequences, extending the system's ability to quantify ADL fit.

2.5 Conditional Preference Networks (CP-nets): Extensions and Theory

Conditional Preference Networks (CP-nets) are a graphical framework used to represent and reason about conditional preferences [9]. Formally, a CP-net consists of a finite set of variables $V = \{X_1, \ldots, X_n\}$ represented as nodes in a directed graph whose edges denote the conditional dependencies among the variables.

Each node X_i in the network is associated with a Conditional Preference Table (CPT), which specifies a strict preference ordering over the possible values of X_i for each combination of values of its parent variables $Pa(X_i)$.

Consider a simple example as shown in Figure 2.1, which depicts the selection of cups based on two features: Material M (Ceramic or Plastic) and Size S (Small or large). The individual strictly prefers a ceramic cup (M_c) over a plastic cup (M_p), while their preference for the size, whether large (S_l) or small (S_s), depends on the Material of the cup: They prefer a large cup if the Material is ceramic, and if the material is plastic, a small cup is preferred.

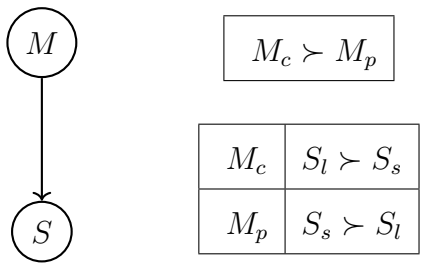


Figure 2.1: CP-net for Cup preference example

Figure 2.2 illustrates the outcome preference graph resulting from the CP-net example. A directed edge from outcome O_i to outcome O_j indicates that O_j is strictly preferred over O_i. For example, the preference of $M_p \wedge S_s$ over $M_p \wedge S_l$ can be directly inferred

from the Conditional Preference Table (CPT). The structure of the graph arranges outcomes in order of preference, with the least desirable outcomes placed at the top and the most desirable at the bottom.

Once a CP-net has been constructed, it can be used to compare pairs of outcomes through ordering queries. An ordering query asks whether a given outcome o can be consistently ordered above another outcome o', that is, whether there exists at least one preference ranking that satisfies the CP-net in which o is preferred to o'. Formally, this corresponds to checking whether the CP-net does not entail that o' is preferred to o. For acyclic CP-nets, ordering queries can be answered in time linear in the number of variables by examining whether any variable exists for which the two outcomes assign the same values to all ancestors but o assigns a more preferred value than o'. This efficiency makes ordering queries particularly practical for ranking sequences of activity performance, as they allow a consistent preference ordering over multiple outcomes to be established without exhaustive comparison.

A stronger form of outcome comparison is the dominance query, which asks whether a CP-net entails that outcome o is preferred to outcome o' across every preference ranking that satisfies the network. Dominance is established by searching for an improving flipping sequence, a chain of intermediate outcomes each obtained from the previous by substituting a variable's value for a more preferred one given its parent assignment, leading from o' to o. If such a sequence exists, o dominates o' with respect to the CP-net. While dominance queries are more informative than ordering queries, they are also computationally harder: for binary-valued tree-structured CP-nets the complexity is quadratic in the number of variables, and for more connected structures the problem becomes NP-complete. In this study, ordering queries are used rather than dominance testing, as they are sufficient for ranking gear-level sequences and are tractable given the size and structure of the learned CP-nets.

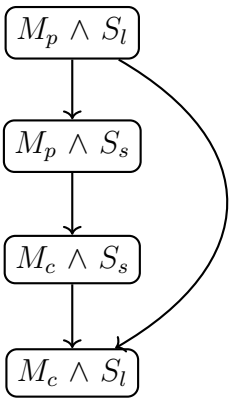


Figure 2.2: Induced preference graph

Beyond foundational contributions, recent studies have advanced the use of CP-nets in several key directions. Koriche and Zanuttini [30] developed algorithms for learning acyclic CP-nets from labelled datasets using membership and equivalence queries. Their work helped formalize the conditions under which exact learning is tractable. Building on this, Alanazi *et al.* [2] introduced a complexity-theoretic analysis of learning acyclic CP-nets from swap examples. They calculated tight bounds for information complexity measures like VC-dimension and teaching dimension, which offer theoretical guidance for designing efficient learning algorithms. Several researchers have proposed methods for structure learning. Bigot *et al.* [8] designed active learning strategies to minimize user burden during CP-net elicitation. Similarly, Liu *et al.* [36] applied minimum redundancy maximum correlation (mRMCR) to guide structure discovery from observed preferences, while Zhu *et al.* [57] combined constraint satisfaction and score-based search for learning CP-net topologies.

More recent studies have explored the application of CP-nets in adaptive, user-focused environments. Li *et al.* [34], for instance, employed CP-nets in smart home settings to reflect user preferences in context-aware automation. Laing *et al.* [31] demonstrated how CP-nets can be incorporated into decision-support systems, particularly in situations where explainability and personalization are essential.

Other works have expanded CP-net expressiveness. Alkhiri and Mouhoub [3] introduced Weighted Constrained CP-nets (WCCP-nets), which combine soft constraints with conditional preferences for more nuanced decision modeling. Cornelio *et al.* [11] proposed PCP-nets, which allow for partial conditional preferences, increasing flexibility in uncertain or sparse settings. Furthermore, Koriche [29] explored relational CP-nets to handle preference dependencies across related objects, such as product bundles or activity sequences. Finally, aggregation and comparison of CP-nets across users or contexts has become an active research area. Lukasiewicz *et al.* [38] analyzed the computational complexity of aggregating CP-nets under Pareto and majority voting rules. Alkhiri *et al.* [4] examined CP-net similarity metrics under constrained preference scenarios, supporting applications such as user profiling and preference matching.

These extensions highlight CP-nets' growing role in adaptive, context-aware systems. Together, they illustrate how preference learning has shifted toward more interactive, expressive, and practical models. Although CP-nets have traditionally been applied in recommendation and planning systems, their potential extends to behavioural modeling and activity monitoring. Their conditional structure makes them well-suited to modeling context-specific routines, while dominance testing provides a framework for comparing alternative activity sequences.

This thesis uses CP-nets to compare how people perform daily activities under different physical conditions. These conditions are simulated by wearing progressive levels of aging gear. The system builds CP-nets from symbolic activity sequences collected during each condition. It then uses ordering queries to compare how the same person performs the same activity across gear levels. This helps identify subtle changes in preference or execution. By modeling conditional structure and using query-based comparison, we explore whether CP-nets make it possible to detect fine-grained drift

in functional performance that other statistical or label-based methods may miss. Unlike typical HAR systems that focus on labeling activities or detecting broad shifts in behaviour, this approach focuses on how performance quality changes over time or with increased strain. Our hypothesis is that CP-nets capture dependencies between motion features and context, such as posture or sequence timing, making them suitable for tracking nuanced decline. This method is especially useful when you want to detect early signs of physical limitations before they appear in routine classification results.

Chapter 3

Methodology

3.1 Overview

This chapter describes the methodology used to build and evaluate the ADL-fit framework from skeletal motion data. The chapter follows the processing pipeline in Figure 3.1. First, CSV files containing 3D coordinates for 32 skeletal joints are collected under different gear conditions and converted into symbolic sequences. Next, Sequential Pattern Mining (SPM) and Conditional Sequential Pattern Mining (CSPM) are used to extract movement features from these sequences. Finally, a Conditional Preference Network (CP-net) is constructed from the mined features to compare activity performance across gear levels. The primary objective of this study is to investigate whether ADL performance can be quantified in a way that is sensitive to subtle deterioration caused by aging, and more generally whether ADL fit can be quantified at all. In this thesis, ADL fit refers to how closely a person's execution of an Activity of Daily Living matches a baseline, healthy pattern of movement.

SPM is used to identify symbolic movement patterns that occur frequently within each activity and therefore characterize the typical performance of that activity.

CSPM refines these results by detecting class-specific conditional relationships; in simple terms, it finds situations where, in one class, the presence of one pattern makes another pattern much more likely than in the contrasting class. In our experiments, the classes correspond to different gear conditions for the same activity (for example, no gear versus high gear). The CP-net encodes these patterns and conditional dependencies as a structured model that can be used to decide, for any pair of activity sequences, which one more closely resembles the learned baseline performance and thus has better ADL fit.

A multi-stage workflow was designed to process and analyze CSV recordings of participants performing common ADLs under four gear conditions: no gear; low gear (which includes ankle weights); medium gear (which includes ankle weights, knee braces, unstable shoes, and wrist weights); and high gear, which includes all the medium-gear components plus a weighted vest and gloves to inhibit mobility.

The first phase of the workflow is data collection and preprocessing. Skeletal data are captured using synchronized depth cameras and converted into temporal joint-coordinate sequences with 32 landmarks. Each dataset corresponds to approximately 45 seconds of a participant's activity performance under a given gear condition (no gear, low gear, medium gear, or high gear). One dataset for each level of gear is collected, with a second dataset for no-gear and high-gear used for training the classifier on non-aged versus aged patterns, with the remaining datasets reserved for evaluation.

In the SCOUT implementation, class labels are defined via a parameter file: the directories of the training data files are explicitly listed under `trainingFolders`, and the class IDs (0, 1) are directly mapped to each listed directory. Datasets reserved for evaluation are listed under `classificationDataFolder` in the parameter file.

Next, the Java implementation of the SCOUT codebase is applied to the training data

to extract a pool of frequent and discriminative patterns. These mined patterns form the features of the model, that is, the independent patterns that occur consistently within each class. In the subsequent stage of Conditional Sequential Pattern Mining (CSPM), SCOUT analyses these mined patterns to determine which are conditionally dependent on others, producing the conditional features that define the parent–child relationships of the CP-net.

The features mined by SCOUT in the SPM and CSPM stages are encoded as CP-net nodes and edges. The CP-net solver then evaluates each pair of test sequences by mapping their feature vectors onto the network and employing an algorithm to judge the better fit. Aggregating these pairwise wins yields dataset-level rankings across gear conditions, from best performing to worst performing. Figure 3.1 illustrates the workflow of information from the sensor data to the final CP-net-based evaluation and outlines the key mechanisms operating at each stage.

The subsequent sections detail each phase of the workflow. Section 3.2 describes the data collection protocol and preprocessing pipeline, including the capture environment, camera configuration, and the transformation of raw joint coordinates into symbolic sequences. Section 3.3 presents feature extraction using Sequential Pattern Mining (SPM) and Conditional Sequential Pattern Mining (CSPM), and introduces the unconditional and conditional feature sets derived from the training data. Section 3.4 then describes how these features are encoded in a CP-net, how the network structure is constructed and made acyclic, and how the solver uses node-level preference rules to compare pairs of test subsequences. Finally, Section 3.5 explains how the results of these pairwise comparisons are aggregated into dataset-level win-rates and how these win-rates are used to rank gear conditions in terms of ADL fit.

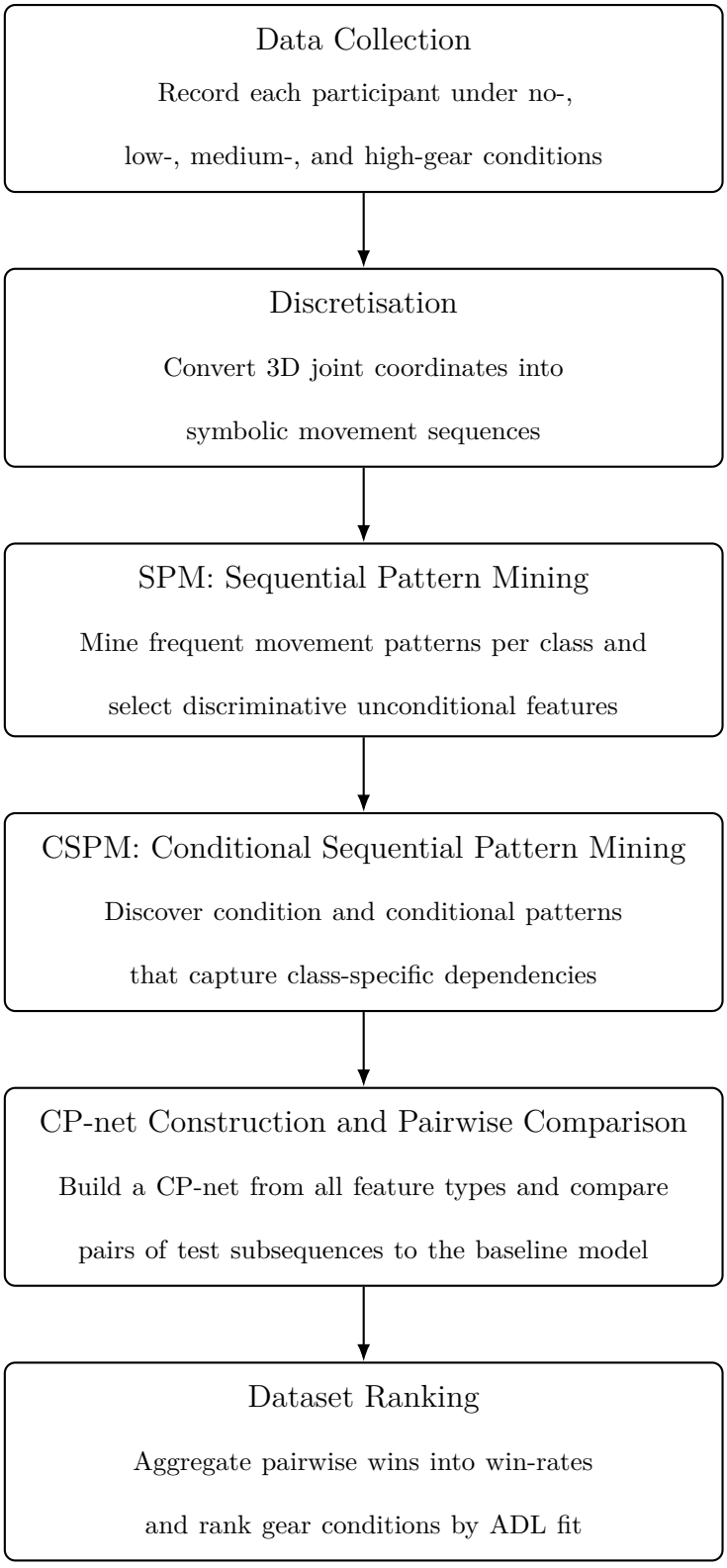


Figure 3.1: Methodology workflow: from multi-condition data collection to CP-net evaluation and dataset ranking.

3.2 Feature Extraction and Representation

The SCOUT framework operates under a parameter file that defines all experimental configurations. The parameter file is divided into several sections that define how SCOUT reads, processes, and analyzes the motion data. This design allows SCOUT to operate flexibly across different datasets and experimental setups. By adjusting the parameter file, users can control which data are used for training or testing, define sequence segmentation rules, and tune the thresholds for mining discriminative patterns without altering the underlying source code. Once the parameters are loaded, SCOUT begins by reading the skeletal motion data from the specified directories and segmenting them into short temporal sequences. It then converts the continuous joint coordinates into symbolic representations through discretisation. Sequential Pattern Mining (SPM) identifies frequent movement patterns in the training datasets for different classes of activity, and Conditional Sequential Pattern Mining (CSPM) extends this by detecting conditional dependencies between patterns. The Naive Bayes classifier then uses these features to estimate the likelihood that a new test sequence belongs to each class. The class with the higher confidence score signifies a closer match.

3.2.1 Sequence Formation and Preprocessing

Each activity recording is stored as a CSV file containing three-dimensional joint coordinates captured at 30 frames per second. Every row in the CSV file corresponds to a single captured frame, while the columns represent the x, y, and z coordinates and confidence values of the 32 skeletal joints (e.g., 0_x, 0_y, 0_z, 0_conf, . . . , 31_x, 31_y, 31_z, 31_conf).

During preprocessing, the parameter file specifies which columns in the data files

are used by SCOUT for pattern mining and classification. Confidence columns are excluded because they do not contain spatial–temporal information relevant to movement representation.

Table 3.2.1: Excerpt from a preprocessed skeletal CSV file for the laundry-loading activity. Each row corresponds to one frame, and each column stores a 3D coordinate for a specific joint.

Frame	0_x	0_y	0_z	1_x	1_y	...	31_x	31_y	31_z
1	0.3043	-0.0655	2.0172	0.0979	-0.2432	...	0.0800	-0.7091	1.9435
2	0.3063	-0.0675	2.0201	0.0979	-0.2434	...	0.0810	-0.7092	1.9427
3	0.3042	-0.0654	2.0256	0.0979	-0.2435	...	0.0811	-0.7087	1.9435
4	0.3041	-0.0665	2.0303	0.0979	-0.2462	...	0.0813	-0.7034	1.9445
5	0.3053	-0.0663	2.0351	0.0979	-0.2456	...	0.0823	-0.7056	1.9433
.	.	.	.	.	.	...	.	.	.
.	.	.	.	.	.	...	.	.	.
.	.	.	.	.	.	...	.	.	.
200	0.3143	-0.0655	2.0352	0.0979	-0.2421	...	0.0845	-0.7048	1.9421

3.2.2 Discretisation

After sequence formation, each subsequence consists of raw 3D joint coordinates, which must be converted into symbolic representations that describe how the joints change over time rather than their exact numeric values. This conversion process is called discretisation and is handled by SCOUT's discretisation module.

The discretiser applies a sliding window of length `windowSize` $= 3$ with a step of 1 to each numeric sequence. For every window and every column (joint coordinate), the three values in that window are collected into a 3-value vector. This vector is then normalised by subtracting its mean from each entry, so that only the short-term shape of the movement (for example, upward, downward, or almost flat) is retained. Across all subsequences, the normalised 3-value vectors are grouped into clusters separately for each column. Each cluster centre becomes a centroid that represents a typical short-term trend in that column, and all vectors assigned to the same centroid

share the same trend pattern.

The discretisation process transforms raw coordinate data into symbolic sequences in two phases. In the first phase, SCOUT learns the centroids by clustering normalised windows from the training data for each column; this step builds a dictionary of typical movement trends per coordinate. In the second phase, these learned centroids are used to encode sequences symbolically. The sliding window is applied again, each normalised 3-value vector is matched to the closest centroid for its column, and the symbol associated with that centroid is recorded. At each window position, the symbols from all columns are collected into an itemset, and the ordered list of itemsets forms the symbolic sequence for that subsequence, for example:

$$\langle \{s_{10}, s_{24}, \ldots\}, \{s_{10}, s_{25}, \ldots\}, \{s_{11}, s_{27}, \ldots\}, \ldots \rangle.$$

To interpret this sequence in real-world terms, each symbol s_i describes the short-term trend of a single joint coordinate over three frames. For example, if the symbolic sequence is derived from raw 3D joint coordinates for a walking activity, the symbol s_{10} might represent a short-term trend where the y-coordinate of the knee moves upward. An itemset such as $\{s_{10}, s_{24}, \ldots\}$ therefore summarises how multiple joints move together during that brief time window. As we move along the sequence of itemsets, each itemset represents the state of the body at a given moment, tracking how the body posture evolves over time and providing a compact description of the activity. These symbolic sequences are then used as input to the SPM and CSPM stages.

3.3 Sequential Pattern Mining (SPM)

3.3.1 SPM (Unconditional)

Sequential Pattern Mining (SPM) identifies ordered patterns that occur frequently across a collection of sequences. Before describing how SPM is applied to the skeletal data in this study, it is helpful to consider a small illustrative example. Suppose we have a set of items

$$I = \{a, b, c, d\}$$

and four short sequences:

$$S_1 = \langle a, b, c \rangle, \quad S_2 = \langle b, c, d \rangle, \quad S_3 = \langle a, c \rangle, \quad S_4 = \langle b, d \rangle.$$

A sequential pattern is an ordered subsequence that may be contained within a longer sequence as long as the order of items is preserved. For example, the pattern $\langle b, c \rangle$ is contained in S_1 and S_2 because both sequences contain a b that appears before a c. It is not contained in S_3 or S_4.

The *support* of a pattern is the fraction of sequences in which it appears. In this example, $\langle b, c \rangle$ appears in 2 out of the 4 sequences, so its support is $2/4 = 0.5$. A minimum support threshold is then used to decide which patterns are frequent. If the minimum support is set to 0.5, $\langle b, c \rangle$ would be considered a frequent sequential pattern, whereas any pattern that appears in fewer than half of the sequences would be discarded as too rare.

In the ADL-fit framework, the symbolic sequences produced by the discretisation step provide the input for pattern mining. After discretisation, each training trial is represented as a symbolic sequence of itemsets. The resulting symbolic sequence for

the training data might look like this:

$$\langle I_1, I_2, I_3, ..., I_L \rangle$$

where each itemset (I_t) contains one symbol per joint dimension for a given time window, for example

$$\langle S_{10}, S_{24}, ... \rangle$$

SCOUT then applies the `splitSequencesBySize` parameter to divide each long symbolic sequence into shorter subsequences of fixed length, measured in numbers of time windows. Each resulting subsequence is treated as an individual sequence for Sequential Pattern Mining (SPM).

For illustration, suppose that after splitting the symbolic sequence for Class 0 (for example, the no-gear condition) and collapsing each itemset to a single representative symbol for simplicity, the sequences are:

$$\text{Seq 1: } \langle A, B, C, D \rangle, \quad \text{Seq 2: } \langle A, B, C, E \rangle, \quad \text{Seq 3: } \langle A, C, E, F \rangle,$$

$$\text{Seq 4: } \langle A, F, C, E \rangle, \quad \text{Seq 5: } \langle A, D, B, E \rangle, \quad \text{Seq 6: } \langle A, B, D, E \rangle,$$

Similarly, for Class 1 (for example, the high-gear condition), we might obtain:

$$\text{Seq 7: } \langle A, D, C, E \rangle, \quad \text{Seq 8: } \langle A, D, E, C \rangle, \quad \text{Seq 9: } \langle A, E, F, E \rangle,$$

$$\text{Seq 10: } \langle A, D, B, C \rangle, \quad \text{Seq 11: } \langle A, C, D, B \rangle, \quad \text{Seq 12: } \langle A, C, D, E \rangle.$$

Here, each letter (A, B, C, D, E, F) represents a typical short-term movement pattern that arose from the discretisation step. In the actual implementation, each position is an itemset of symbols, rather than a single symbol, but the mining logic is the same.

SPM is applied separately within each class to identify frequent sequential patterns. A sequential pattern is an ordered sequence of symbols that may appear as a subsequence within a longer symbolic sequence, provided that the temporal order is preserved and the gap between consecutive symbols does not exceed `maxGapForMining`. SCOUT begins with all length-1 candidate patterns (for example, $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$, $\langle D\rangle$, $\langle E\rangle$, $\langle F\rangle$ for Class 0) and computes their support, defined as the proportion of subsequences in the class in which the pattern occurs at least once. Patterns whose support is lower than the minimum threshold `minSequenceSupport` are discarded, and the remaining frequent patterns are extended to form longer candidates. For example, if the length-1 patterns $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$ and $\langle D\rangle$ are frequent in Class 0, SPM may generate length-2 candidates such as $\langle A, B\rangle$, $\langle A, C\rangle$, $\langle B, D\rangle$ and so on, and again retain only those whose support meets the minimum requirement. This process continues iteratively, extending frequent patterns to length-3, length-4, and so on, up to the maximum pattern length specified by `maxSequenceSize`. At the end of this stage, each class is characterised by a set of frequent patterns that summarise short recurring movement fragments in that condition.

Next, patterns that are frequent and discriminative are identified. Intuitively, a pattern is discriminative if it has relatively high support in one class and relatively low support in the other. For each candidate pattern P, SCOUT constructs a 2×2 contingency table that records how often P is present or absent in the subsequences of Class 0 and Class 1. For example, in the simplified example above, the pattern $\langle B, E\rangle$ appears in 3 of the 6 Class 0 sequences and in none of the Class 1 sequences, which suggests that $\langle B, E\rangle$ is strongly associated with Class 0. Conversely, the pattern $\langle A, D\rangle$ appears in 3 of the 6 Class 0 sequences but in 5 of the sequences in Class 1, indicating that it is much more typical of Class 1. From this same logic, the algorithm finds $\langle D, C\rangle$ to have a higher support in Class 1 compared to Class 0.

From each contingency table, SCOUT's SPM algorithm computes a chi-square statistic that measures how strongly the presence of the pattern is associated with one class rather than the other. If the chi-square value for pattern P exceeds the user-defined threshold `minChiSquare`, then P is retained as a discriminative feature. Patterns that occur at similar rates in both classes have small chi-square values and are discarded, whereas patterns that are heavily skewed toward one class yield large chi-square values and are likely to pass the threshold. Finally, for every discriminative pattern P, SCOUT assigns a preferred class label `prefClass` based on which class has the higher support for that pattern (for example, `prefClass` $= 0$ for $\langle B, E \rangle$ and `prefClass` $= 1$ for $\langle A, D \rangle$ in this example). The resulting set of discriminative frequent patterns constitutes the unconditional feature set. These features are used both by the Naïve Bayes classifier and as nodes (with their associated preferred classes) when constructing the CP-net in the later stages of the framework.

3.3.2 Conditional Sequential Pattern Mining (CSPM)

Conditional Sequential Pattern Mining (CSPM) extends the standard SPM process by examining conditional relationships between patterns, rather than only their individual frequency. While SPM identifies sequences that frequently appear in a class, CSPM investigates whether the presence of one sequence tends to be accompanied by the presence of another, and whether this relationship differs between classes. In other words, SPM answers the question "what happens often", whereas CSPM asks "when pattern A appears, does pattern B tend to appear as well, and is this effect stronger in one class than in the other".

A simple way to think about this is to imagine two activities, such as walking and sweeping. Let w represent a leg movement pattern and v represent an arm movement pattern. Both w and v may appear in sequences from both activities. However, it may

happen that in sweeping sequences, whenever w occurs, v also appears in the same short time window most of the time, while in walking sequences this co-occurrence is rare. In that case, the conditional pattern $(w \Rightarrow v)$ is strongly associated with sweeping but not with walking. CSPM formalises this idea by measuring, for each class, how often the conditional pattern appears when the condition pattern is present.

In SCOUT, the conditional relationship between two patterns is written as $P(\text{Sequence}_2 \mid \text{Sequence}_1)$, where Sequence_1 is the condition sequence and Sequence_2 is the target (conditional) sequence. The algorithm computes this conditional probability separately for each class to determine whether the relationship holds similarly in both classes or is much stronger in one class than in the other.

To give an illustrative example, assume that SPM has already been run for a single activity (for example, walking) with two classes: Class 0 representing the no-gear condition and Class 1 representing the high-gear condition. Using the same simplified symbolic sequences introduced in Section 3.3, the Class 0 sequences are:

$$\text{Seq 1: } \langle A, B, C, D\rangle, \quad \text{Seq 2: } \langle A, B, C, E\rangle, \quad \text{Seq 3: } \langle A, C, E, F\rangle,$$

$$\text{Seq 4: } \langle A, F, C, E\rangle, \quad \text{Seq 5: } \langle A, D, B, E\rangle, \quad \text{Seq 6: } \langle A, B, D, E\rangle,$$

and the Class 1 sequences are:

$$\text{Seq 7: } \langle A, D, C, E\rangle, \quad \text{Seq 8: } \langle A, D, E, C\rangle, \quad \text{Seq 9: } \langle A, E, F, E\rangle,$$

$$\text{Seq 10: } \langle A, D, B, C\rangle, \quad \text{Seq 11: } \langle A, C, D, B\rangle, \quad \text{Seq 12: } \langle A, C, D, E\rangle.$$

Suppose that SPM has identified the following patterns as frequent, unconditional patterns, which are class discriminative patterns that appear without reference to

any other pattern:

$$P_1 = \langle A, D \rangle, \quad P_2 = \langle D, C \rangle, \quad P_3 = \langle B, E \rangle.$$

These patterns are selected because their supports differ strongly between classes:

$$\text{supp}_0(\langle B, E \rangle) = \frac{3}{6}, \quad \text{supp}_1(\langle B, E \rangle) = \frac{0}{6} \quad \Rightarrow \quad \langle B, E \rangle \text{ is associated with Class 0.}$$

$$\text{supp}_0(\langle A, D \rangle) = \frac{3}{6}, \quad \text{supp}_1(\langle A, D \rangle) = \frac{5}{6} \quad \Rightarrow \quad \langle A, D \rangle \text{ is associated with Class 1.}$$

$$\text{supp}_0(\langle D, C \rangle) = \frac{0}{6}, \quad \text{supp}_1(\langle D, C \rangle) = \frac{3}{6} \quad \Rightarrow \quad \langle D, C \rangle \text{ is associated with Class 1.}$$

CSPM now examines whether the appearance of one pattern, treated as a condition, affects the behaviour of another pattern, treated as a conditional pattern, and whether this conditional relationship differs between classes. Typically, the appearance of the conditional pattern without any context is more representative of one class, for example, Class 0, but when the context (the condition) is added, it flips to be more representative of the other Class, in this case Class 1. For example, let the condition pattern be

$$X = \langle A, B \rangle$$

and the conditional pattern be

$$Y = \langle D, E \rangle.$$

In CSPM we ask: when X appears in a sequence, how often does Y also appear, and does this behaviour differ between Class 0 and Class 1?

First, we examine the unconditional supports of $Y = \langle D, E \rangle$:

$$\text{supp}_0(\langle D, E \rangle) = \frac{2}{6}, \quad \text{supp}_1(\langle D, E \rangle) = \frac{3}{6}.$$

Given those, $\langle D, E\rangle$ seems to be more likely a Class 1 pattern. Next, we examine the support of $\langle D, E\rangle$ under the condition of $\langle A, B\rangle$'s presence.

Consider first the Class 0 sequences. The pattern $X = \langle A, B\rangle$ appears in 4 of Class 0 sequences, thus

$$\text{conditionCount}_0(X) = 4.$$

Among the sequences in Class 0, where $X = \langle A, B\rangle$ appears, $Y = \langle D, E\rangle$ appears twice. Thus, the conditional support of Y given the presence of X is:

$$\text{CondSupp}_0(X \to Y) = \frac{2}{4}$$

Next, consider the Class 1 sequences. Here X appears in 2 sequences:

$$\text{conditionCount}_1(X) = 2.$$

Among the sequences where X appears, Y appears in 0 of the 2. Thus:

$$\text{CondSupp}_1(X \to Y) = \frac{0}{2}.$$

Clearly, Y now appears to belong to Class 0 under the condition X ($\frac{2}{4}$ versus $\frac{0}{2}$), and thus the condition X has altered Y's class membership.

To quantify the difference between classes, CSPM constructs a contingency table around the conditional relationship and computes a conditional chi-square value. The algorithm tests whether the conditional relationship between X and Y differs significantly between classes by comparing the conditional chi-square statistic to the threshold `minCondChiSquare`. If the chi-square for the pair (X, Y) exceeds this threshold, (X, Y) is retained as a conditional dependency. In this case, X is treated as a condition feature (a parent pattern) and Y as a conditional feature (a child pattern),

and the class in which the relationship is stronger is recorded as the conditional preferred class `condPrefClass` (conditionally preferred class) for Y in the context where X is present.

In the example above, the conditional rule "if $\langle A, B\rangle$ then $\langle D, E\rangle$" makes sense as a description of Class 0, because it was determined that whenever these two patterns appear together the conditional pattern, which has a higher support in Class 1 without the condition, now has a higher support in Class 0 after the context of the condition $\langle A, B\rangle$ is applied.

Repeating these steps for the example given above, other conditional relationships to be found for Class 0 and Class 1 are $\langle A, E\rangle \rightarrow \langle D, B\rangle$ and $\langle A, D\rangle \rightarrow \langle C, E\rangle$ respectively. To briefly demonstrate why these relationships are conditional, consider first $\langle A, E\rangle \rightarrow \langle D, B\rangle$. Unconditionally, $\langle D, B\rangle$ is more likely a Class 1 feature:

$$\text{supp}_0(\langle D, B\rangle) = \frac{1}{6}, \quad \text{supp}_1(\langle D, B\rangle) = \frac{2}{6}.$$

However, under the condition $\langle A, E\rangle$, this flips to favour Class 0:

$$\text{CondSupp}_0(\langle A, E\rangle \rightarrow \langle D, B\rangle) = \frac{1}{5}, \quad \text{CondSupp}_1(\langle A, E\rangle \rightarrow \langle D, B\rangle) = \frac{0}{4}.$$

Similarly, for $\langle A, D\rangle \rightarrow \langle C, E\rangle$, unconditionally $\langle C, E\rangle$ favours Class 0:

$$\text{supp}_0(\langle C, E\rangle) = \frac{3}{6}, \quad \text{supp}_1(\langle C, E\rangle) = \frac{2}{6}.$$

Under the condition $\langle A, D\rangle$, this flips to more likely be a Class 1 feature:

$$\text{CondSupp}_0(\langle A, D\rangle \rightarrow \langle C, E\rangle) = \frac{0}{3}, \quad \text{CondSupp}_1(\langle A, D\rangle \rightarrow \langle C, E\rangle) = \frac{2}{5}.$$

At the end of the CSPM stage, the algorithm identifies a set of condition patterns (patterns that provide informative contexts), a set of conditional features (patterns whose class-specific behaviour changes depending on the context), a conditional preferred class for each conditional feature, and a conditional chi-square value that quantifies the strength of each dependency. Together with the unconditional features obtained from SPM, these condition and conditional features are used to construct the CP-net. In the CP-net structure, edges $X \rightarrow Y$ correspond to conditional dependencies discovered by CSPM, and the conditional preference information stored in `condPrefClass` helps define the local rules at each node

3.4 CP-net Modelling and Evaluation

3.4.1 Construction

Conditional Preference Networks (CP-nets), introduced in Section 2.5, are a graphical framework used to reason about a user's conditional preferences over multi-attribute outcomes. A CP-net models how the preferred value of each variable may depend on the values of a small set of parent variables. Traditionally, CP-nets are used to answer comparison or dominance queries: given two outcomes, the CP-net determines which one is preferred according to the conditional preference rules. In practice, CP-nets have been applied in many tasks such as building personalised recommendation systems and decision support or planning systems.

In this thesis, we apply CP-nets in a slightly different way. Instead of building a CP-net from an explicitly elicited preference database, we construct CP-nets from symbolic activity sequences derived from a user's class 0 (no-gear) and class 1 (high-gear) motion data. The learned CP-nets encode conditional patterns that characterise baseline versus high-strain performance. During evaluation, we treat pairs of test

activity sequences as alternative outcomes and use ordering queries on the learned CP-net to decide which sequence is closer to the class 0 performance profile.

The Conditional Preference Network is constructed from the two feature types mined in the SPM and CSPM stages: unconditional features and conditional features. Unconditional features are patterns that are typical of one class on their own. Conditional features are patterns whose class-specific behaviour depends on one or more conditions. The CP-net builder iterates through the feature sets and, for each distinct pattern that appears in any of them, creates a single node in the network. Each node is tagged according to the roles it plays (condition, conditional feature, and/or unconditional feature) and is added to the corresponding internal lists (for example, `conditionNodes`, `conditionalFeatureNodes`, and `unconditionalNodes`). In addition, each node stores its associated pattern, its lists of parents and children, the unconditional preferred class (`prefClass`), the conditional preferred class (`condPrefClass`), if any, and the conditional chi-square values on its incoming edges.

Edges between parent and child nodes are defined using the CSPM output. For each conditional feature, the builder examines which condition features act as its contexts and adds a directed edge from each condition node (parent) to the corresponding conditional node (child). Symmetrically, the parent's list of children is updated so that the resulting structure forms a directed graph that links condition features to the conditional features whose behaviour they influence. Because CSPM may introduce cycles, a cycle-removal step iteratively deletes the weakest edges, as measured by their chi-square values, from any directed cycles until the graph becomes a directed acyclic graph. The final CP-net therefore encodes both independent and context-dependent relationships among the mined patterns in a form that can be used for outcome comparison.

Let us continue with the examples presented in Sections 3.3.1 and 3.3.2. To give

a brief recap, the class discriminative patterns found were $\langle B, E\rangle$ for Class 0 and for Class 1, $\langle A, D\rangle$ and $\langle D, C\rangle$. The conditional relationship between patterns for Class 0 are $\langle A, B\rangle \rightarrow \langle D, E\rangle$ and $\langle A, E\rangle \rightarrow \langle D, B\rangle$. For Class 1 the conditional relationship established between patterns is $\langle A, D\rangle \rightarrow \langle C, E\rangle$.

The class discriminative patterns go on to become unconditional nodes, and the conditional patterns become conditional nodes whose relationship between their conditions is represented by a directed edge from the parent node (the condition) to the child node (the conditional node). Each node stores the `prefClass` and the `condPrefClass` which encode the local preference and conditional preference rules, respectively. The local conditional rules are stored in the `Conditional Preference Table` (CPT). In our ADL-fit framework, the likelihood of one outcome matching an activity better than another is evaluated using the conditional rules encoded in the Conditional Preference Table (CPT). The aim is to find which outcome is closest to the baseline (Class 0). Accordingly, the CPT encodes the preference statement that is most consistent with Class 0.

To explain further, let X denote the condition (parent) and Y the conditional (child), where a value is either present (X, Y) or absent ($\overline{X}$, $\overline{Y}$).

For an unconditional node, the preference rule is straightforward. If `prefClass`$= 0$, the pattern is more typical of Class 0, and so its presence indicates a higher likelihood of an outcome matching Class 0:

$$X \succ \overline{X}.$$

If `prefClass`$= 1$, the pattern is more typical of Class 1, and since we are scoring closeness to Class 0, its absence indicates a higher likelihood of an outcome matching Class 0:

$$\overline{X} \succ X.$$

If `condPrefClass`=0, this means that the appearance or absence of both the parent and child feature is more characteristic of Class 0. An outcome therefore has a higher likelihood of matching Class 0 when the child matches the parent (presence matches presence, absence matches absence):

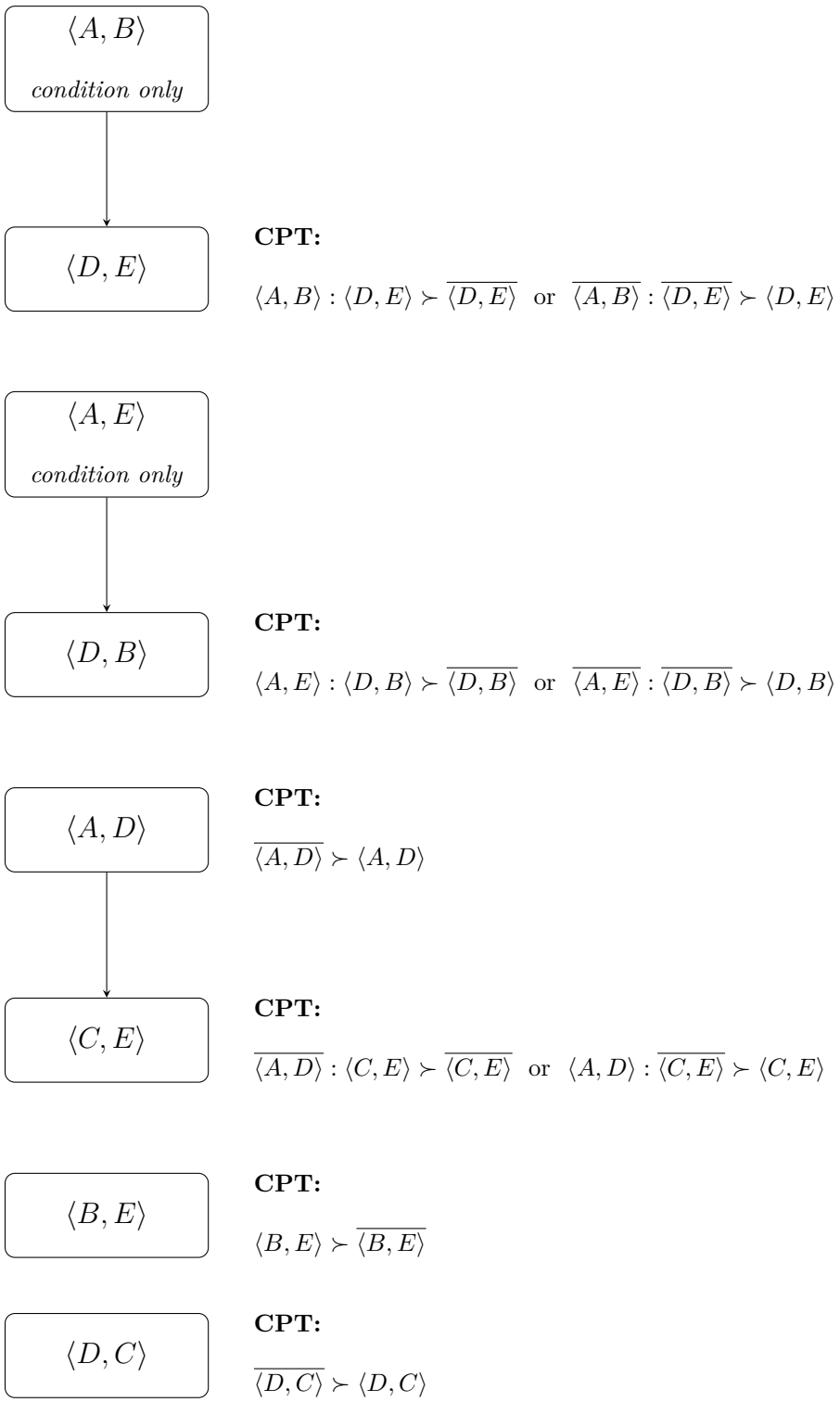
$$X : Y \succ \overline{Y} \qquad \text{and} \qquad \overline{X} : \overline{Y} \succ Y.$$

Finally, if `condPrefClass`=1, this means that the appearance or absence of both the parent and child feature is more characteristic of Class 1. Since we are scoring closeness to Class 0, an outcome has a higher likelihood of matching Class 0 when the child feature behaves oppositely to the parent (presence matches absence, and vice versa):

$$X : \overline{Y} \succ Y \qquad \text{and} \qquad \overline{X} : Y \succ \overline{Y}.$$

Figure 3.2 illustrates the CP-net constructed from the six features identified in the examples above, where directed edges represent the conditional dependencies discovered by CSPM.

Figure 3.2: Example CP-net constructed from eight features. Directed edges encode conditional dependencies. CPT statements next to each node show the corresponding likelihood rules.



3.4.2 Pairwise Comparison of Outcomes

After constructing the CP-net, pairs of test sequences can be compared to determine which is more likely a better match for Class 0. This is done by encoding each sequence as a Boolean outcome over the CP-net nodes and then applying one of several outcome-comparison strategies. The first strategy, **CP-net traversal using ordering queries**, performs a graph traversal and aggregates node-level evidence while enforcing a shared-context requirement: a conditional node is only used to compare two outcomes when all of its ancestor patterns have identical presence values in both outcomes. A second strategy, **weighted hybrid voting**, follows the same shared-context requirement but gives more influence to conditional nodes when more of their parent patterns are present, reflecting stronger contextual support for the conditional rule. To support large networks, a third strategy, **global heuristic scoring**, computes a single heuristic score for each outcome by summing node-wise contributions across the CP-net, without traversal or shared-context checks, trading CP-net semantic strictness for scalability. Finally, **dominance testing via improving flips** provides a stricter semantic check by determining whether one outcome can be reached from another through a sequence of locally preferred value changes under the learned conditional rules.

3.4.3 Example: Evaluating test Sequences (Outcomes) with the CP-net

For any pair of test sequences, denoted Outcome 0 and Outcome 1, the CP-net solver first checks which of the mined patterns are present in each sequence. This is done through a Boolean feature-encoding step: for every CP-net node, a Boolean flag `presence0` records whether the corresponding pattern appears at least once in

Outcome 0, and a second flag `presence1` records whether it appears in Outcome 1. If a pattern is found in a given test sequence, its presence flag is set to `true`; otherwise, it is set to `false`. These presence values provide a compact representation of how each outcome satisfies (or fails to satisfy) the learned patterns at every node.

To illustrate this process, consider again the CP-net from the previous subsection with eight nodes corresponding to the patterns $\langle A, B\rangle$, $\langle D, E\rangle$, $\langle A, E\rangle$, $\langle D, B\rangle$, $\langle A, D\rangle$, $\langle C, E\rangle$, $\langle B, E\rangle$ and $\langle D, C\rangle$. Suppose we wish to compare the following two test sequences:

$$\text{Outcome 0 } (S_0)\text{: } \langle A, B, D, E\rangle, \qquad \text{Outcome 1 } (S_1)\text{: } \langle A, B, D, C, E\rangle.$$

The solver examines the outcome to find the presence or absence of the features from the CP-net. Tables 3.4.1 and 3.4.2 demonstrate the presence and absence of the features in Outcome 0 and Outcome 1 respectively.

Table 3.4.1: Feature presence vector for Outcome 0 ($S_0 = \langle A, B, D, E\rangle$).

Feature	**Type**	**Present in S_0**
$\langle B, E\rangle$	Unconditional	✓
$\langle D, C\rangle$	Unconditional	×
$\langle A, B\rangle$	Condition	✓
$\langle A, E\rangle$	Condition	✓
$\langle A, D\rangle$	Condition + Unconditional	✓
$\langle D, E\rangle$	Conditional	✓
$\langle D, B\rangle$	Conditional	×
$\langle C, E\rangle$	Conditional	×

Table 3.4.2: Feature presence vector for Outcome 1 ($S_1 = \langle A, B, D, C, E \rangle$).

Feature	**Type**	**Present in S_1**
$\langle B, E \rangle$	Unconditional	✓
$\langle D, C \rangle$	Unconditional	✓
$\langle A, B \rangle$	Condition	✓
$\langle A, E \rangle$	Condition	✓
$\langle A, D \rangle$	Condition + Unconditional	✓
$\langle D, E \rangle$	Conditional	✓
$\langle D, B \rangle$	Conditional	×
$\langle C, E \rangle$	Conditional	✓

The presence information is then assigned to every node for both outcomes, and the CP-net solver performs a pairwise comparison of the two outcomes in both directions, that is, outcome 0 versus outcome 1 and vice versa. It traverses the CP-net and considers only those nodes where the presence flags differ between Outcome 0 and Outcome 1, since nodes with identical presences cannot distinguish between the two sequences. At nodes without parents (unconditional or purely condition nodes), the solver uses the unconditional preferred class `prefClass`: if a pattern is typical of Class 0, then an outcome in which that pattern is present is treated as more Class-0-like, whereas if a pattern is typical of Class 1, an outcome in which it is absent is treated as more Class-0-like. At nodes with parents (conditional nodes), the solver first checks that all ancestors have the same presence values in both outcomes. When the context defined by the parents is identical, it uses the conditional preferred class `condPrefClass` together with the nodes' presence flags to decide which outcome better satisfies the local conditional rule.

Each node where the outcomes differ can therefore cast a local "vote" for either

Outcome 0 or Outcome 1, indicating which sequence has a higher likelihood of matching Class 0 according to the node's (unconditional or conditional) preference rule. The solver sums these votes across all nodes in the CP-net. If Outcome 0 receives more votes than Outcome 1, it is judged to be a better fit to the baseline class; if Outcome 1 receives more votes, it is judged closer to the baseline class instead. In this way, the CP-net aggregates local preference rules into a global comparison between two activity sequences.

3.4.3.1 CP-net Traversal for Outcome Comparison Using Ordering Queries Theorem

To understand how the solver determines which outcome is closer to Class 0, it is useful to describe how the preference rules at each node are interpreted. Each CP-net node stores two labels learned from the training data. The field `prefClass` records the unconditional class association of the node. The field `condPrefClass` records the conditional class association of the node.

For nodes without parents (unconditional or pure condition nodes), the interpretation is straightforward. For unconditional nodes, if `prefClass` = 0, then an outcome where the feature is present is treated as more Class-0-like than an outcome where it is absent; if `prefClass` = 1, then an outcome where the feature is absent is treated as more Class-0-like, since the feature itself is typical of Class 1. For conditional nodes, the solver may use either `prefClass` or `condPrefClass`, depending on the parent context. When all parents are absent in both outcomes, the conditional context is effectively inactive and the node falls back to its unconditional preference `prefClass`. When at least one parent is present in both outcomes, the conditional context is active, and the node uses `condPrefClass`: if `condPrefClass` = 0, here the outcome that satisfies the feature under that parent context is treated as more Class-0-like;

if `condPrefClass` = 1, then an outcome that violates the feature in that context is treated as more Class-0-like. This way, each node contributes a local vote for which outcome is more likely to match Class 0, based on whether its unconditional or conditional rule is satisfied.

We continue from the outcome examples in Section 3.4.3:

$$\text{Outcome 0 } (S_0)\text{: } \langle A, B, D, E\rangle, \qquad \text{Outcome 1 } (S_1)\text{: } \langle A, B, D, C, E\rangle,$$

and the Boolean features encoded in Table 3.4.1 and 3.4.2.

The ordering query implementation starts with the initialisation of two vote counters, `outcome0count` and `outcome1count`, and sets both to zero. Next the implementation traverses the CP-net using a breadth-first (BFS) strategy: it first processes root nodes (nodes without parents), which are typically unconditional or pure condition nodes, and then moves downwards to their descendants.

Pure condition nodes are nodes that serve only as contexts for conditional features and do not themselves match a class and are therefore skipped.

From our example, the solver starts with the unconditional node $\langle D, C\rangle$. Nodes $\langle A, D\rangle$ and $\langle B, E\rangle$ are skipped since the presence values are the same for both outcomes. At $\langle D, C\rangle$ a vote is cast for Outcome 0 as it satisfies the CPT rule at that node. At the end of this step the counters for Outcome 0 and Outcome 1 are set at 1 and 0, respectively.

Once all unconditional nodes have been processed, the solver moves on to evaluate the conditional nodes. Here the evaluation is implemented using the rules in ordering queries, which states that both outcomes must agree on all the values of their ancestors [9]. This simply means that when determining which outcome is more likely to match Class 0 at a node, all the ancestors for outcome 0 and 1 must be the same. This

requirement is implemented using an ancestor check function that verifies that both outcomes have the same presence values for all their ancestors before going on to evaluate the preference at the current node, or skips the nodes when the ancestor values do not match.

Next, the solver processes conditional nodes whose ancestor presence values are identical across both outcomes, as this is required before a node can cast a vote. Here, only nodes whose presence values differ are evaluated; therefore, nodes $\langle D, B\rangle$ and $\langle D, E\rangle$ are skipped.

Finally, node $\langle C, E\rangle$ is evaluated. The conditional context at this node is active as both parent conditions are present. A vote is cast for Outcome 0 as it satisfies the CPT rule ($\langle A, D\rangle : \overline{\langle C, E\rangle} \succ \langle C, E\rangle$) at this node.

Once all unconditional and conditional nodes are satisfied, the solver function returns a boolean for $Outcome\ 0 > Outcome\ 1$. For this example, the function returns $True$, and this is recorded as a win for Outcome 0.

3.4.3.2 Outcome Comparison Using Weighted Hybrid Voting

This approach compares two outcomes by aggregating, across the CP-net, each node's local vote for which outcome is more likely to match Class 0. Each test sequence is first encoded using Boolean presence flags at every node, indicating whether the corresponding feature occurs in the sequence.

Nodes without parents contribute unconditional evidence using `prefClass`, following the same CPT rules described in Section 3.4.1. Pure condition nodes define context only and do not contribute votes.

Conditional nodes contribute only when the two outcomes agree on the values of all ancestors, ensuring that both outcomes are evaluated under the same context. When

this condition holds, the local CPT rule encoded by `condPrefClass` determines which outcome is more likely to match Class 0.

Conditional evidence is weighted by the strength of the observed parent context. For a conditional node Y with parent set $\mathrm{Pa}(Y)$, the number of present parents is counted (this value is identical for both outcomes because the ancestor agreement condition enforces identical ancestor values). If at least one parent is present, the conditional vote is multiplied by this parent-presence count; if all parents are absent, the node contributes a single unweighted vote. This weighting prioritises conditional rules supported by more observed conditions, which correspond to more specific contexts. Unlike the earlier CP-net traversal baseline, where each evaluated conditional node contributes one vote regardless of how many parents are active, this approach increases the influence of conditional rules when they are supported by a larger parent context.

Summing all unconditional votes and weighted conditional votes yields two totals, one for Outcome 0 and one for Outcome 1. The outcome with the larger total is judged closer to the baseline class (Class 0).

3.4.3.3 Outcome Comparison Using Global Heuristic Scoring

This method replaces CP-net traversal and node-by-node voting with a single global heuristic score computed independently for each outcome. The motivation is to obtain a fast ordering-query approximation that scales to large CP-nets by avoiding breadth-first traversal, ancestor-consistency checks, and repeated pairwise CPT evaluations. Instead, it evaluates how well each outcome matches the learned node labels and then compares the two resulting scores.

As in the other methods, each test sequence is first embedded as an outcome by setting Boolean presence flags for every CP-net node. Given these presence values, the method computes a heuristic score by iterating over all nodes in the CP-net

and accumulating signed contributions based on whether the outcome satisfies the node's Class-0-consistent interpretation. Unconditional evidence is handled using `prefClass`, following the same CPT rules described in Section 3.4.1.

Conditional evidence is incorporated using `condPrefClass` when a conditional context is active. For a node with parents, the solver checks whether at least one parent feature is present in the outcome. If so, the conditional context is treated as active and, provided `condPrefClass`$\neq -1$, the node contributes according to the conditional preference interpretation described earlier. When `condPrefClass`$= 0$, the outcome receives a positive contribution when the node's value matches the Class-0-consistent CPT behaviour in that context; when `condPrefClass`$= 1$, the outcome receives a positive contribution when it satisfies the inverted (Class-0-consistent) CPT behaviour. If the conditional context is inactive (all parents absent), the method falls back to the unconditional contribution based on `prefClass`.

After summing contributions across all nodes, the solver divides the total by the number of nodes that contributed a vote, yielding an average score that remains comparable across outcomes. This normalization ensures that outcomes are compared on the basis of average node-level agreement with Class 0, rather than the total number of active nodes, which may differ between outcomes. This heuristic score is then computed for Outcome 0 and Outcome 1 and returns `true` if Outcome 0 attains a higher score, indicating that it is judged closer to the baseline class. In contrast to the traversal-based methods, this method does not enforce the CP-net ancestor-agreement constraint and therefore trades CP-net semantic strictness for computational simplicity and robustness in large, sparse networks.

3.4.4 Dominance Testing Using Improving Flips

Using the CP-net constructed in Section 3.4, each node corresponds to a Boolean variable indicating whether the associated feature is present (`true`) or absent (`false`) in an outcome. Consider the two test sequences introduced in Section 3.4.2:

$$\text{Outcome 0 } (S_0) : \langle A, B, D, E \rangle,$$
$$\text{Outcome 1 } (S_1) : \langle A, B, D, C, E \rangle.$$

The presence encoding, as reported in Tables 3.4.1 and 3.4.2, yields the following differences between the two outcomes. S_0 and S_1 share the same presence values for $\langle B, E \rangle$, $\langle A, B \rangle$, $\langle A, E \rangle$, $\langle A, D \rangle$, $\langle D, E \rangle$, and $\langle D, B \rangle$. They differ only on two variables: $\langle D, C \rangle$, which is absent in S_0 and present in S_1, and $\langle C, E \rangle$, which is absent in S_0 and present in S_1.

An improving flip is determined by the local preference rule at the flipped node. For unconditional nodes, the preferred value is given by `prefClass`: if `prefClass` $= 0$, then the presence is more likely to match Class 0, and if `prefClass` $= 1$, the absence is more likely to match Class 0 because the feature is typical of Class 1. For conditional nodes, the preferred value is determined from the CPT using `condPrefClass` and the current parent context, following the Class-0-consistent interpretation described in Section 3.4.

To test whether S_0 dominates S_1, the improving-flip procedure begins at S_1 and applies locally improving changes until either S_0 is reached or no improving flips remain. In this example, two improving flips are available from S_1:

- The node $\langle D, C \rangle$ is unconditional with `prefClass` $= 1$, meaning absence of

$\langle D, C\rangle$ is more likely to match Class 0 under the ADL-fit interpretation, since the pattern is typical of Class 1. Because S_1 has $\langle D, C\rangle = \texttt{true}$, flipping it to `false` is an improving flip.

- The node $\langle C, E\rangle$ has parent $\langle A, D\rangle$ and `condPrefClass` $= 1$. Since $\langle A, D\rangle$ is present in S_1, the Class-0-consistent CPT rule prefers $\langle C, E\rangle$ to be absent. Because S_1 has $\langle C, E\rangle = \texttt{true}$, flipping it to `false` is an improving flip.

Applying these flips produces an improving path from S_1 to S_0. One valid sequence is:

$$
\begin{aligned}
S_1 &: \Big(\langle D, C\rangle = \texttt{true},\ \langle C, E\rangle = \texttt{true}\Big) \\
&\to \Big(\langle D, C\rangle = \texttt{false},\ \langle C, E\rangle = \texttt{true}\Big) \quad [\text{flip } \langle D, C\rangle] \\
&\to S_0 : \Big(\langle D, C\rangle = \texttt{false},\ \langle C, E\rangle = \texttt{false}\Big) \quad [\text{flip } \langle C, E\rangle]
\end{aligned}
$$

Because S_0 is reachable from S_1 through a sequence of improving flips that each satisfy the local CP-net preference rules, S_0 is more likely to match Class 0 than S_1 under dominance testing.

3.5 Cross-Dataset Comparison and Win-rate Computation

One of the aims of this study is to investigate whether the CP-net framework can be used to quantify ADL fit and detect behavioural drift. However, the CP-net evaluation method operates at the level of pairs of outcomes rather than entire datasets. In addition, the SCOUT framework does not compare full CSV files directly.

Instead, each test CSV is discretised into a symbolic sequence and then split into shorter subsequences. Each subsequence is treated as a test sequence (outcome), and the CP-net evaluation method (`determineIsBetterFit`), applied in both directions for each pair, decides whether one subsequence is strictly closer to Class 0 than the other or whether the pair should be treated as a tie. Whenever a subsequence is judged strictly closer to Class 0 than its competitor, it receives a "win".

To obtain a dataset-level measure of ADL fit, the CP-net solver aggregates these sequence-level wins across all subsequences belonging to the same test dataset (gear level). The win rate of a dataset is defined as the ratio between (i) the total number of wins achieved by subsequences from that dataset and (ii) the total number of pairwise comparisons in which those subsequences participated against subsequences from all other datasets. Finally, the solver ranks the test datasets by win rate, from highest to lowest. A higher win rate indicates that the corresponding dataset behaves more like the baseline class and is therefore interpreted as exhibiting better ADL fit.

Chapter 4

Results and Discussion

4.1 Introduction to the Results

This chapter presents the empirical results of the experiments conducted to evaluate whether SCOUT's Sequential Pattern Mining (SPM) classifier and Conditional Preference Networks (CP-nets) can detect changes in the quality of Activity of Daily Living (ADL) performance under simulated aging conditions. The central aim is to examine how the features extracted from skeletal data, the resulting confidence scores, and the CP-net preference comparisons reflect changes in activity execution as participants perform the same tasks while wearing increasingly restrictive gear.

The results are organised around the two research questions introduced in Chapter 1:

- **RQ1:** Can SCOUT's SPM classifier detect decline in ADL fit as participants perform everyday activities under higher levels of simulated aging gear?
- **RQ2:** Can CP-net ordering queries identify performance deterioration by ranking lower gear sequences as better fits to baseline movement than sequences recorded under higher gear levels?

To provide context for these analyses, the chapter first describes the four activities studied (walking, vacuuming, laundry, and Sit-to-Stand) and summarises the experimental conditions under which they were recorded. It then outlines how the skeletal recordings were prepared for analysis and explains the key SCOUT parameter settings that control discretisation, sequential pattern mining, conditional sequential pattern mining, and model evaluation. These sections establish the structure of the datasets used for training and testing and clarify how the SPM and CP-net models were configured.

The core of the chapter addresses the two research questions in turn. For RQ1, SCOUT's Class 0 confidence scores are examined to determine whether similarity to baseline (no-gear) movement decreases as gear level increases. Descriptive patterns across participants and activities are followed by formal statistical tests that assess the effect of gear level on these confidence scores. For RQ2, the chapter analyses the CP-net win rates obtained from pairwise ordering queries. These analyses investigate whether no-gear sequences tend to achieve higher win rates than low-, medium-, and high-gear sequences, and whether the resulting gear-level orderings show a consistent decline in ADL fit.

Building on these main results, the chapter then examines how structural properties of the learned CP-nets relate to result quality. Correlation analyses link CP-net connectivity, the number of conditional nodes, and the average node degree to the distance between observed gear-level win distributions and an ideal decline pattern. Finally, the chapter summarises the findings by research question, discusses the study's limitations, and concludes with an integrated interpretation of how SCOUT and CP-nets behave under noisy, low-frame-rate skeletal data when used to assess ADL performance under simulated aging conditions.

4.2 Activities and Experimental Conditions

The experiments in this study were conducted in a simulated apartment setting designed to resemble a typical home environment. The space included a living area, kitchen, laundry section, and open floor space that allowed participants to perform common Activities of Daily Living (ADLs). Four activities were selected for analysis: walking, vacuuming, laundry, and the Sit-to-Stand (STS) task. The walking activity involved participants walking in a straight line across the open floor space and returning to the starting position, repeated continuously for the duration of the recording. The vacuuming activity involved participants pushing and pulling a vacuum cleaner back and forth across a designated floor area in a controlled sweeping motion. The laundry activity involved participants repeatedly transferring items between a laundry basket and an improvised washing machine, bending down to pick up items and placing them into the machine, then reversing the process for the duration of the recording. The STS task involved rising from a seated position and taking three steps forward, repeated continuously for the duration of the recording. These activities were chosen because they represent everyday tasks that require coordinated whole-body movement and are sensitive to physical changes associated with aging.

To study functional decline in a controlled manner, each participant performed the activities under four gear conditions that introduced increasing levels of physical restriction. The no-gear condition served as the baseline and reflected the participant's natural movement. The low-gear condition involved ankle weights, which added mild resistance. The medium-gear condition introduced additional constraints: knee braces, unstable shoes, and wrist weights. The high-gear condition combined all previous elements with a weighted vest and hand-mobility-restricting gloves. These progressive restrictions were intended to simulate age-related changes in strength, balance, and

mobility while still allowing controlled comparisons across the four conditions. These different gear levels are shown in Figure 4.1.

Each recording consisted of a 45-second sequence captured using two synchronised Orbbec Femto Bolt depth cameras positioned at right angles to provide overlapping coverage. Data was collected at 4 frames per second and produced 32-joint skeletal data. The reduction in effective frame rate from 30fps to 4fps was due to a technical limitation in the data collection pipeline. All data collection was carried out using prototype software developed by the National Research Council of Canada (NRC) as part of the Aging in Place research program.



(a) No gear (b) Low gear (c) Medium gear (d) High gear

Figure 4.1: Participant wearing each of the four gear conditions: (a) no gear, (b) low gear, (c) medium gear, and (d) high gear.

4.2.1 Data Preparation and Parameter Settings

The skeletal recordings collected from the depth cameras were stored as CSV files, with each file containing a 45-second performance of one activity under a specific gear condition. Each row represented a single frame, and the columns stored the three-dimensional coordinates and confidence values for the 32 tracked joints. Before training or testing, the datasets were organized through the SCOUT parameter file, which specifies all directory paths, data preprocessing rules, and model-building

configurations. This file served as the central control mechanism for every stage of the SPM and CP-net workflow.

Several key settings within the parameter file determined how SCOUT processed the data. The entries under trainingDataFolders defined the directories for Class 0 and Class 1 training sequences, while classificationDataFolder identified the folder containing the test sequences. The boolean controls ScoutTrainInterface, ScoutTestInterface, and UseCPNet activated the three major components of the system: training the SPM and CP-net models, evaluating test sequences with the trained SPM classifier, and running CP-net preference-based comparisons.

Beyond preprocessing, the parameter file defined all mining thresholds for Sequential Pattern Mining (SPM) and Conditional Sequential Pattern Mining (CSPM). These thresholds determined the minimum support required for a sequence to be considered frequent, the maximum pattern length allowed during mining, and the statistical criteria used to identify discriminative or conditionally dependent movement patterns. Because the quality of the extracted features directly influenced SCOUT and CP-net performance, many of these values were experimentally tuned. Several iterations were conducted to adjust the support levels, chi-square thresholds, maximum pattern sizes, and coverage parameters until a set of configurations was reached that produced stable and meaningful results.

4.3 RQ1: Evaluating ADL Fit Using SCOUT (SPM Classifier)

This section presents the results of the SCOUT Sequential Pattern Mining (SPM) classifier and examines whether the Class 0 confidence scores decrease as gear level increases. The goal is to determine whether SCOUT reliably identifies baseline

movements and detects deviations that reflect a decline in ADL fit.

Before selecting the final configuration for training the SPM model, several combinations of Class 0 and Class 1 datasets were explored. Each activity was paired with every other activity to determine whether the resulting confidence scores demonstrated the expected decline across gear levels. These exploratory trials included training on pairs such as walking versus vacuuming, vacuuming versus laundry, and laundry versus Sit-to-Stand. These pairings were explored for two reasons. First, it was hypothesised that activities sharing similar movement characteristics might serve as a proxy for one another in representing performance decline. Second, this formed part of a broader exploratory process to identify the most stable and interpretable training configuration. Many of these cross-activity combinations produced inconsistent or contradictory patterns that did not align with the expected trend of decreasing similarity to baseline movement. The most stable and interpretable results were obtained when the no-gear recordings were used as Class 0 and the high-gear recordings were used as Class 1 for each activity. This configuration produced a clear separation between natural and restricted movement and was therefore adopted for all SCOUT analyses presented in this chapter.

The subsections that follow examine the Class 0 confidence scores produced for each activity and participant, analyze trends across the four gear levels, and discuss the extent to which SCOUT can detect performance deterioration under simulated aging conditions, thereby addressing Research Question 1.

4.3.1 Understanding the Class 0 Confidence Scores

The SCOUT classification model outputs two confidence scores for each test sequence. The first score, Class 0, represents how closely the test sequence matches the baseline no-gear performance. The second score, Class 1, reflects how similar the sequence is to

Table 4.3.1: Class 0 confidence scores for all participants across the four activities and gear levels. Higher scores indicate closer similarity between a test sequence and the baseline (no-gear) activity model learned by SCOUT. These values form the basis for identifying declines in ADL fit as movement restriction increases.

Participant	**Gear Level**	**Walking**	**Vacuuming**	**Laundry**	**STS**
P1	No Gear	0.573	0.513	0.569	0.464
	Low Gear	0.495	0.440	0.513	0.484
	Medium Gear	0.439	0.479	0.508	0.464
	High Gear	0.472	0.480	0.415	0.427
P2	No Gear	0.528	0.568	0.639	0.641
	Low Gear	0.525	0.688	0.516	0.574
	Medium Gear	0.506	0.686	0.493	0.512
	High Gear	0.519	0.604	0.492	0.553
P3	No Gear	0.553	0.630	0.575	0.568
	Low Gear	0.520	0.633	0.522	0.565
	Medium Gear	0.509	0.478	0.524	0.511
	High Gear	0.477	0.310	0.325	0.464
P4	No Gear	0.527	0.615	0.569	0.558
	Low Gear	0.528	0.587	0.619	0.549
	Medium Gear	0.527	0.316	0.484	0.506
	High Gear	0.469	0.270	0.448	0.519
P5	No Gear	0.418	0.499	0.485	0.406
	Low Gear	0.511	0.467	0.483	0.485
	Medium Gear	0.441	0.487	0.461	0.457
	High Gear	0.498	0.459	0.445	0.536
P6	No Gear	0.508	0.509	0.509	0.626
	Low Gear	0.511	0.498	0.440	0.278
	Medium Gear	0.450	0.506	0.353	0.207
	High Gear	0.457	0.509	0.317	0.262
P7	No Gear	0.687	0.669	0.668	0.674
	Low Gear	0.485	0.541	0.519	0.631
	Medium Gear	0.471	0.539	0.483	0.619
	High Gear	0.388	0.542	0.300	0.607

the high-gear training data. The scores are normalised so they sum to 1. Each score represents the probability that the activity belongs to the corresponding class. In this study, the Class 0 confidence scores are of primary interest because they indicate the degree to which a given activity performance still resembles the participant's natural, unrestricted movement. Lower Class 0 scores therefore suggest a greater deviation from baseline and a potential decline in ADL fit.

Table 4.3.1 presents the Class 0 confidence scores for each participant, activity, and gear level. For every participant (P1–P7), the rows correspond to the four gear conditions (no gear, low gear, medium gear, and high gear), while the columns list the scores for walking, vacuuming, laundry, and Sit-to-Stand (STS). Each entry in the table can be interpreted as the strength of evidence, according to the SPM classifier, that the corresponding test sequence matches the baseline movement patterns learned from the no-gear training data.

These scores provide the foundation for evaluating ADL fit with SCOUT. If SCOUT successfully captures performance deterioration, the Class 0 confidence values should generally decrease as the gear level increases, reflecting reduced similarity to baseline movement under greater physical restriction. The next subsection examines these trends across participants and activities in more detail.

4.3.2 Trends Across Participants and Activities For SCOUT'S SPM

This subsection examines whether the Class 0 confidence scores decrease as the gear level increases for each activity. A downward trend indicates a loss of similarity to baseline movement and therefore a potential decline in ADL fit. The results are presented activity by activity to highlight common patterns and notable participant-

specific variations. Section 4.3.3 then summarises these findings by examining the average confidence scores across all participants.

4.3.2.1 Walking

Walking showed several cases where Class 0 confidence decreased with increased restriction, but the trend was not uniformly monotonic. Participants P3 and P7 demonstrated the clearest declines across all gear levels, with P7 displaying the strongest reduction from no-gear to high-gear. Participants P1, P2, and P6 also showed a general downward direction, although each exhibited slight fluctuations, typically with the high-gear score being marginally higher than the medium-gear score. P4's scores for no-, low-, and medium-gear were almost identical, with a more noticeable decline only at high-gear. P5 displayed an irregular pattern, with low-gear unexpectedly producing the highest score and no-gear the lowest. Overall, walking showed partial support for declining baseline similarity, with several participants exhibiting mild to moderate decreases as gear level increased.

4.3.2.2 Vacuuming

Vacuuming produced the most variable results of all four activities. Only P4 showed a clear and consistent decline from no-gear to high-gear. P3 also showed a general decrease between no-gear and high-gear, although low-gear produced slightly higher values than the baseline. For many participants (P1, P5, P6, P7), the confidence scores fluctuated with no stable downward pattern. P2 produced an inverted trend where low- and medium-gear scored higher than no-gear. This variability suggests that vacuuming may allow compensatory upper-body and torso strategies that preserve baseline-like patterning even under moderate restriction. As a result, SCOUT's sensitivity appears weaker for this activity, with only a few participants exhibiting

clear performance deterioration.

4.3.2.3 Laundry

Laundry demonstrated the most consistent downward pattern across participants. P1, P2, P5, P6, and P7 all showed progressively lower Class 0 confidence scores from no-gear to high-gear. P3 displayed a similar overall decline, although medium-gear scored slightly higher than low-gear. P4 differed from the general pattern, with low-gear producing the highest score, followed by no-gear, medium-gear, and finally high-gear. Despite this exception, laundry still produced the strongest and most monotonic trends, suggesting that SCOUT reliably detects deviations in this complex full-body task as movement restrictions increase.

4.3.2.4 Sit-To-Stand (STS)

The Sit-to-Stand task showed several declining patterns, but with more participant-specific irregularities than laundry or walking. Participants P3 and P7 showed consistent decreases across all gear levels, while P2, P4, and P6 displayed general downward trends interrupted by slight rebounds at high-gear. P1's pattern fluctuated, with low-gear producing the highest score and high-gear the lowest. P5 produced the most irregular trend, with high-gear scoring the highest and no-gear the lowest. Overall, STS shows modest evidence of decline, but its short, discrete movement structure and sensitivity to timing variability likely contributed to the fluctuations observed across several participants.

Across the four activities, laundry showed the clearest and most uniform decline in Class 0 confidence, followed by walking, which displayed moderate but consistent decreases for several participants. vacuuming and STS exhibited mixed patterns, likely due to the greater influence of compensatory strategies or variability in short-

duration motion. Despite these inconsistencies, many participants exhibited lower confidence scores at higher gear levels, indicating that SCOUT can detect reductions in baseline similarity for at least some activities, though the strength of this detection varies with task complexity and movement characteristics.

4.3.3 Statistical analysis of gear-level effects on SCOUT's confidence

The previous section presented the SPM classification results for all participants and activities. These descriptive results showed that, although some participants exhibited the expected decline in Class 0 confidence scores as gear level increased (particularly for walking and laundry), the trends were not consistently monotonic across all activities and participants. For vacuuming and STS, several participants showed irregular or flat patterns rather than a clear decrease from no gear to high gear. Because visual inspection alone cannot determine whether gear level has a systematic effect on the SPM confidence scores, this section applies a formal statistical test to evaluate the gear-level effect.

The statistical test chosen for this task is the Friedman test. The Friedman test is used to determine whether there are differences between three or more related groups when the data are ordinal or when the assumptions for a parametric repeated-measures ANOVA are not met, that is, when the data are non-parametric. In this context, the Friedman test is used to determine whether gear level affects SCOUT Class 0 confidence scores for each activity. Each participant provides a set of confidence scores under all four gear levels (conditions). For each participant, the scores are ranked from the lowest (1) to the highest (4); if two values are tied, they are each assigned the average of the ranks they would occupy. The ranks are then summed for each condition. If there were no real differences between conditions, these rank

sums would be roughly similar across gear levels.

The Friedman test statistic is then computed using the following formula:

$$\chi_F^2 = \frac{12}{n\,k\,(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1),$$

where

- n is the number of participants,
- k is the number of conditions (here, gear levels), and
- R_j is the rank sum for condition j.

If all gear levels behave similarly, the R_j values will be close together and χ_F^2 will be small. If some gear levels consistently receive higher or lower ranks, the R_j values will be more spread out and χ_F^2 will be large.

A p-value is then obtained by comparing χ_F^2 to the chi-square distribution. Under the null hypothesis that all conditions have the same distribution, χ_F^2 is approximately chi-square distributed with

$$\mathrm{df} = k - 1.$$

The p-value is given by

$$p = P\left(\chi^2 \geq \chi_F^2 \,\middle|\, H_0 \text{ is true}\right).$$

If $p < \alpha$, where in this context α is set to 0.05, the null hypothesis is rejected and it is concluded that there is a global difference among conditions. Otherwise, the null hypothesis is retained and there is no evidence of a difference among conditions.

Kendall's W is used as an effect size for the Friedman test:

$$W = \frac{\chi_F^2}{n\,(k-1)}.$$

If $W \approx 0$, there is little or no agreement and the effect is very weak. If W is close to 1, the effect is strong. Kendall's W therefore summarises how strong the gear-level effect is for a given activity.

Finally, if the Friedman test indicates that gear level has an effect on the confidence scores, post hoc pairwise tests are performed to identify which specific pairs of gear levels differ. In this thesis, Dunn–Bonferroni tests are used for this purpose. The Bonferroni correction adjusts for the increased risk of false positives that arises when multiple pairwise comparisons are performed simultaneously. It does so by dividing the significance threshold α by the number of comparisons m, yielding an adjusted threshold of $\alpha' = \alpha/m$. A pairwise comparison is considered statistically significant only if its p-value falls below this adjusted threshold. These tests consider every possible pairing of the different gear levels and, after applying the Bonferroni correction, indicate which pairs show statistically significant differences in confidence scores. In practical terms, the Friedman test is used to determine whether gear level has any overall effect on the scores, while the post hoc Dunn–Bonferroni comparisons identify which specific pairs of gear levels differ significantly. Non-significant pairwise results are interpreted as a lack of sufficient evidence for a difference, rather than proof that the gear levels have identical effects. Subsections 4.3.3.1 to 4.3.3.4 present the Friedman test results and post hoc pairwise comparisons for each activity (walking, vacuuming, laundry, and STS).

4.3.3.1 Walking: gear-level effects on SCOUT Class 0 confidence

This subsection examines whether SCOUT Class 0 confidence scores for walking differ across the four gear levels (no, low, medium, and high) for the same seven participants. Table 4.3.2 presents the descriptive statistics for each gear level. The means and medians show a gentle but monotonic decline from no gear (mean 0.54) through low (0.51) and medium (0.48) to high gear (0.47). The standard deviations are small and broadly similar across conditions, indicating relatively consistent scores within each gear level. These descriptive results suggest that walking sequences recorded under higher gear levels tend to receive slightly lower Class 0 confidence scores than those recorded at no or low gear.

Table 4.3.2: Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for walking by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.54	0.53	0.08	3.21
Low gear	7	0.51	0.51	0.02	3.43
Medium gear	7	0.48	0.47	0.04	1.64
High gear	7	0.47	0.47	0.04	1.71

The mean ranks reported in Table 4.3.2 provide a participant-wise view of this pattern. No and low gear achieve the highest mean ranks (3.21 and 3.43, respectively), while medium and high gear have notably lower mean ranks (1.64 and 1.71). In terms of ranked performance, confidence tends to be higher at no and low gear and lower at medium and high gear.

The Friedman test yielded $\chi^2(3) = 11.44$, $p = 0.01$, Kendall's $W \approx 0.54$. This indicates a statistically significant gear-level effect with a moderate-to-strong effect size.

Post hoc Dunn–Bonferroni pairwise comparisons were examined to identify which specific gear-level contrasts contributed to this effect (Table B.1.1 in Appendix B). The adjusted p-values showed that low gear confidence scores were significantly higher than medium gear scores after Bonferroni correction. Other pairwise differences followed the expected ordering numerically (with no and low gear tending to have higher confidence than medium and high gear), but none of these additional contrasts reached statistical significance once the correction for multiple comparisons was applied.

In summary, for walking, SCOUT Class 0 confidence scores decrease as gear level increases, and the Friedman test confirms a significant overall gear-level effect. The most robust pairwise separation is between low and medium gear, where low gear yields significantly higher confidence scores, while other differences remain descriptive rather than statistically confirmed in this sample.

4.3.3.2 Vacuuming: gear-level effects on SCOUT Class 0 confidence

Vacuuming showed a modest numerical decline in Class 0 confidence scores across gear levels, falling from a mean of 0.57 at no gear to 0.45 at high gear, though variability increased slightly at the higher gear levels.

Table 4.3.3: Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for vacuuming by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.57	0.57	0.07	3.36
Low gear	7	0.55	0.54	0.09	2.43
Medium gear	7	0.50	0.49	0.11	2.14
High gear	7	0.45	0.48	0.12	2.07

Table 4.3.3 presents the descriptive statistics and mean ranks for the four gear levels. In terms of ranked performance, no gear has the highest mean rank (3.36), followed by low gear (2.43), medium gear (2.14), and high gear (2.07). Thus, both the raw confidence scores and the ranks suggest that vacuuming sequences recorded at no gear tend to receive higher Class 0 confidence than those recorded under higher gear levels.

The Friedman test yielded $\chi^2(3) = 4.41$, $p = 0.22$, Kendall's $W \approx 0.21$, indicating a small and non-significant gear-level effect. The observed downward trend is not strong enough, given the sample size and variability, to constitute statistical evidence of a gear-level effect at the 5% level.

Post hoc Dunn–Bonferroni pairwise comparisons were also examined (Table B.2.1 in Appendix B). All adjusted p-values were substantially greater than 0.05, indicating that no specific pair of gear levels exhibited a statistically significant difference in vacuuming confidence scores after correction for multiple comparisons. For example, the contrast between no gear and high gear showed a relatively small unadjusted p-value (0.062), but this did not remain significant after Bonferroni adjustment.

In summary, although SCOUT Class 0 confidence scores for vacuuming show a numerical decline from no gear to high gear, the Friedman test does not indicate a statistically significant gear-level effect, and no pairwise differences remain significant after Bonferroni correction. These findings suggest that, in this small sample, the apparent decline in vacuuming confidence across gear levels may be attributable to random variation rather than a robust gear-level effect.

4.3.3.3 Laundry: gear-level effects on SCOUT Class 0 confidence

Laundry produced the clearest and most consistent decline in Class 0 confidence scores, with means and medians falling monotonically from no gear (0.57) through

low (0.52) and medium (0.47) to high gear (0.39). The standard deviations are similar across conditions, indicating comparable variability in confidence scores at each gear level. These descriptive results suggest that laundry sequences recorded under higher gear levels tend to receive lower Class 0 confidence scores than those recorded at no gear.

Table 4.3.4: Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for laundry by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.57	0.57	0.06	3.86
Low gear	7	0.52	0.52	0.05	3.00
Medium gear	7	0.47	0.48	0.06	2.14
High gear	7	0.39	0.42	0.08	1.00

The mean ranks in Table 4.3.4 provide a participant-wise view of this pattern. No gear has the highest mean rank (3.86), followed by low gear (3.00) and medium gear (2.14), while high gear has the lowest mean rank (1.00). In terms of ranked performance, no gear almost always receives the highest confidence scores, high gear almost always receives the lowest, and low and medium gear occupy intermediate positions. This ranking pattern is consistent with the descriptive statistics and reflects a strong ordering from no gear to high gear.

The Friedman test yielded $\chi^2(3) = 18.77$,$p < 0.001$, Kendall's $W \approx 0.89$, indicating a highly significant gear-level effect with a large effect size, confirming that laundry confidence scores are strongly influenced by the level of restriction.

Pairwise comparisons (Table B.3.1) showed that high gear confidence scores were significantly lower than both no gear and low gear after Bonferroni correction. The contrast between no gear and medium gear approached significance (adjusted $p = 0.052$) but did not meet the corrected $\alpha = 0.05$ threshold, and the differences between

low and medium gear, and between medium and high gear, were not statistically significant. These results indicate that the most robust separations occur between high gear and the lower gear levels, particularly no and low gear.

In summary, SCOUT Class 0 confidence scores for laundry decrease clearly as gear level increases, and the Friedman test confirms a strong overall gear-level effect with a large effect size. Post hoc comparisons show that high gear produces significantly lower confidence scores than both no gear and low gear, while differences among no, low, and medium gear do not consistently reach statistical significance after Bonferroni correction, despite following the expected ordering numerically.

4.3.3.4 Sit-to-stand (STS): gear-level effects on SCOUT Class 0 confidence

STS showed a partial decline in Class 0 confidence scores, with the mean dropping from 0.56 at no gear to 0.47 at medium gear, before levelling off at 0.48 at high gear. The standard deviations are comparable across conditions, suggesting similar variability in STS confidence scores under each gear level.

Table 4.3.5: Descriptive statistics and mean ranks for SCOUT Class 0 confidence scores for sit-to-stand (STS) by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.56	0.57	0.10	3.36
Low gear	7	0.51	0.55	0.11	3.14
Medium gear	7	0.47	0.51	0.13	1.64
High gear	7	0.48	0.52	0.11	1.86

The mean ranks in Table 4.3.5 provide a participant-wise view of this pattern. No gear has the highest mean rank (3.36), followed by low gear (3.14), while medium

and high gear have lower mean ranks (1.64 and 1.86, respectively). Thus, in terms of ranked performance, confidence scores tend to be higher under no and low gear and lower under medium and high gear, with medium and high forming a lower-ranked pair.

The Friedman test yielded $\chi^2(3) = 9.64$, $p = 0.022$, Kendall's $W \approx 0.46$, indicating a significant gear-level effect with a moderate effect size.

After Bonferroni correction (Table B.4.1), none of the pairwise differences reached the adjusted significance threshold. The contrast between no gear and medium gear showed the smallest adjusted p-value (approximately 0.052), suggesting a potential difference that did not quite meet the conservative criterion. All other adjusted p-values were larger and clearly non-significant.

In summary, for STS, SCOUT Class 0 confidence scores show a numerical decline from no gear to the higher gear levels, and the Friedman test confirms a significant overall gear-level effect with a moderate effect size. However, the post hoc comparisons do not identify any individual pair of gear levels as significantly different after Bonferroni correction, indicating that the observed pattern is more evident at the global level than in specific pairwise contrasts within this small sample.

Figure 4.2 summarises the mean Class 0 confidence scores across all four activities and gear levels.

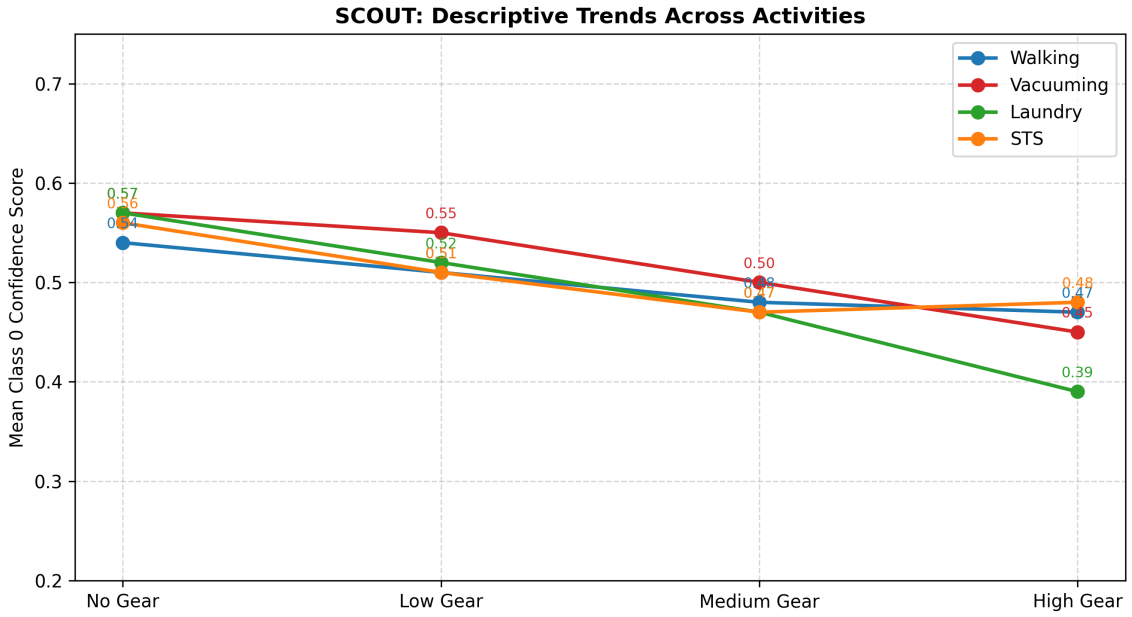


Figure 4.2: Mean SCOUT Class 0 confidence scores across gear levels for all four activities.

4.4 RQ2: Evaluating ADL Fit Using CP-nets

The objective of this research question is to determine whether or how well CP-nets can be used to determine if a subject's performance of an ADL is declining. While the assumption is that this decline is due to age, in our experiments we simulate decline by employing differing levels of wearable gear, as studied in the previous section. We assess this by using CP-nets to determine the "fit" of the subjects' activity to the accepted model. Rather than assigning direct numerical confidence scores to individual sequences, the CP-net model operates through pairwise comparisons. For each pair of test sequences belonging to the same participant and activity, the CP-net determines which sequence is a better fit according to the conditional preference rules derived from the baseline recordings.

The CP-net framework was evaluated using four outcome-comparison methods imple-

mented within the solver. Three of these methods are based on the ordering query theorem: (1) CP-net traversal using ordering queries, which performs a breadth-first graph traversal and aggregates node-level votes while enforcing a shared ancestor-context requirement; (2) weighted hybrid voting, which follows the same shared-context requirement but scales each conditional node's contribution by the number of active parent patterns, giving more weight to rules supported by stronger contextual evidence; and (3) global heuristic scoring, which computes a single score per outcome by summing node-wise contributions across the entire CP-net without traversal or ancestor-consistency checks, trading semantic strictness for scalability. The fourth method is dominance testing via improving flips, which applies the stricter CP-net dominance semantics by checking whether one outcome can be reached from another through a sequence of locally improving value changes.

All three ordering-query-based methods produced similar trends in terms of gear-level win-rate patterns across participants and activities. The results presented in this section are therefore based on the first method, CP-net traversal using ordering queries, as it most closely follows the standard CP-net ordering query semantics described in Section 3.4.1.1.

Dominance testing via improving flips could not be evaluated within the scope of this study due to computational constraints. In practice, the algorithm exhausted the available system memory before completing, with runs exceeding four hours before terminating due to memory allocation failure. This reflects the well-known computational complexity of dominance testing in CP-nets, which in the worst case requires exploring an exponential number of improving flip sequences. This limitation is discussed further in Section 4.6 and motivates the use of the ordering-query-based methods as practical approximations for large-scale ADL evaluation.

4.4.1 Understanding Win Rates in CP-net Evaluation

These pairwise outcomes are summarized using a *win rate*, which serves as the primary metric for interpreting the CP-net results. A win occurs when a given test sequence is judged by the CP-net to be a better fit than the sequence it is compared against, meaning that it better matches the learned baseline preference structure. The win rate for a sequence is calculated as the proportion of wins out of all pairwise comparisons in which it participates.

A higher win rate therefore indicates that a sequence more closely resembles the baseline no-gear performance and can be interpreted as reflecting better ADL fit. Conversely, a lower win rate suggests that the sequence deviates more from the baseline movement patterns and may reflect reduced functional performance under greater physical restriction. In this way, the win rate provides a direct and interpretable measure of how well each gear-level sequence aligns with the baseline preferences captured by the CP-net model.

Because the win rate captures how frequently a sequence is deemed a better fit over its alternatives, it offers a natural indicator for addressing Research Question 2, which examines whether CP-nets can identify performance deterioration as participants perform the same activities under progressively more restrictive gear conditions.

4.4.2 Trends Across Participants and Activities For CP-nets

Table 4.4.1 presents the CP-net win rates for each participant across the four activities. The win rate reflects how often the sequences from a given gear level were deemed to be better fits over those from other gear levels during the CP-net ordering evaluations. Higher win rates therefore indicate closer alignment with the baseline (no-gear) model.

Across most participants and activities, win rates generally declined as gear level

Table 4.4.1: CP-net win rates for all participants across the four activities and gear levels. Win rates represent the proportion of pairwise comparisons in which a test sequence is deemed a better fit over others according to the learned CP-net, indicating its relative similarity to baseline movement patterns.

Participant	**Gear Level**	**Walking**	**Vacuuming**	**Laundry**	**STS**
P1	No Gear	0.921	0.724	0.648	0.634
	Low Gear	0.507	0.396	0.407	0.375
	Medium Gear	0.369	0.346	0.476	0.524
	High Gear	0.081	0.381	0.283	0.309
P2	No Gear	0.532	0.679	0.910	0.600
	Low Gear	0.525	0.453	0.449	0.495
	Medium Gear	0.471	0.445	0.267	0.366
	High Gear	0.251	0.292	0.220	0.446
P3	No Gear	0.690	0.718	0.812	0.693
	Low Gear	0.445	0.649	0.436	0.509
	Medium Gear	0.454	0.350	0.454	0.443
	High Gear	0.248	0.112	0.093	0.236
P4	No Gear	0.471	0.406	0.424	0.677
	Low Gear	0.488	0.295	0.607	0.510
	Medium Gear	0.472	0.542	0.554	0.338
	High Gear	0.392	0.532	0.215	0.286
P5	No Gear	0.347	0.398	0.488	0.363
	Low Gear	0.613	0.525	0.518	0.418
	Medium Gear	0.298	0.454	0.505	0.613
	High Gear	0.505	0.416	0.350	0.408
P6	No Gear	0.765	0.426	0.453	0.854
	Low Gear	0.192	0.654	0.584	0.355
	Medium Gear	0.445	0.413	0.460	0.340
	High Gear	0.444	0.343	0.291	0.363
P7	No Gear	0.565	0.594	0.677	0.602
	Low Gear	0.470	0.350	0.215	0.486
	Medium Gear	0.415	0.471	0.538	0.628
	High Gear	0.430	0.487	0.511	0.246

increased, suggesting that higher levels of restriction produced movement sequences that deviated further from the baseline patterns. The following overview summarises the main patterns observed for each activity.

4.4.2.1 Walking

For walking, the CP-net win rates reveal a mix of expected and irregular patterns across participants. Participants P1 and P2 showed the clearest monotonic declines from no gear to high gear, with P1 dropping sharply from 0.921 to 0.081 and P2 decreasing more gradually from 0.532 to 0.251. These results align well with the anticipated reduction in baseline similarity as movement restrictions increase. Participants P3 and P7 also demonstrated an overall downward trend, although each displayed a minor mid-gear irregularity: P3's medium-gear win rate (0.454) was slightly higher than the low-gear value (0.445), while P7's high-gear score (0.430) exceeded the medium-gear score (0.415). In contrast, P4, P5, and P6 exhibited more inconsistent orderings. P4's win rates varied within a narrow range without forming a clear decline, P5 showed a pronounced reversal with low gear outperforming no gear, and P6 showed an unexpected increase at medium and high gear compared to the low-gear condition. Overall, although walking produced two strong monotonic cases and two additional partial declines, the remaining participants displayed fluctuations that highlight the CP-net's sensitivity to uneven feature distribution across segmented sequences.

4.4.2.2 Vacuuming

Vacuuming showed varied and often non-monotonic CP-net win-rate patterns across participants. Only two participants, P2 and P3, exhibited the expected decline from no gear to high gear. P1 showed a general downward trend, although the high-gear

score (0.381) exceeded the medium-gear score (0.346), interrupting monotonicity. The remaining participants displayed clear irregularities. For P4, medium and high gear produced the strongest win rates (both 0.542), surpassing no gear (0.406) and low gear (0.295). P5 also showed a pronounced reversal, with low gear achieving the highest win rate (0.525), followed by medium gear (0.454), while high gear (0.416) and no gear (0.398) scored lowest. P6 produced a similar pattern, with low gear again ranking highest (0.654), followed by no gear, medium gear, and high gear in descending order. P7 showed a partially inverted pattern: no gear remained the highest (0.594), but high gear (0.487) and medium gear (0.471) both scored higher than low gear (0.350). Overall, the vacuuming results highlight substantial fluctuations and mid-gear reversals across participants, underscoring the CP-net's sensitivity to inconsistent feature distribution in this activity.

4.4.2.3 Laundry

Laundry produced a mixture of expected and irregular CP-net win-rate patterns, with only one participant showing the anticipated monotonic decline from no gear to high gear. Participant P2 demonstrated the clearest downward trend, with win rates dropping steadily from 0.910 (no gear) to 0.220 (high gear). P1 and P3 also showed overall decreases from no gear to high gear, although both displayed mid-gear fluctuations: P1's medium-gear win rate (0.476) exceeded the low-gear value (0.436), and P3's medium-gear score (0.454) was slightly higher than the low-gear score (0.436). The remaining participants exhibited pronounced reversals. For P4 and P5, low gear produced the highest win rates (0.607 and 0.584, respectively), surpassing no gear and medium gear before declining at high gear. P6 showed a similar ordering, with low gear ranking highest (0.584), followed by medium gear (0.460), then no gear (0.453), and finally high gear (0.291). P7 also deviated from expectations: no gear

achieved the highest win rate (0.677), but medium gear (0.538) and high gear (0.511) both scored higher than low gear (0.217). Overall, the laundry results show limited consistency across participants, with only one clear no→high decline and several cases where mid-gear levels outperformed both baseline and higher restriction conditions.

4.4.2.4 Sit-to-Stand (STS)

The STS activity showed a mixture of expected declines and notable irregularities in CP-net win rates across participants. Participants P3 and P4 exhibited the clearest monotonic decreases, with win rates steadily falling from no gear to high gear (P3: 0.693 → 0.236; P4: 0.677 → 0.286). Several other participants showed partial downward trends but with mid-gear or high-gear reversals. For example, P2 displayed a general decline from 0.600 to 0.446, although the high-gear score exceeded the medium-gear value. P6 showed a similar pattern, with a high win rate for no gear (0.854), followed by high gear (0.363) and low gear (0.355), while medium gear had the lowest score (0.340). In contrast, P1 and P5 demonstrated clear irregularities: P1's medium-gear win rate (0.524) exceeded both low gear (0.375) and high gear (0.309), and P5 showed an inverted pattern with medium gear producing the highest win rate (0.613) and no gear the lowest (0.363). P7 also deviated from the expected ordering, with medium gear (0.680) narrowly surpassing no gear (0.602), while high gear (0.246) remained the lowest. Overall, the STS results reveal substantial variability, with only two participants showing a full no→high decline and several others displaying mid-gear reversals or elevated non-baseline scores.

Overall, the CP-net win rate results show that declines in baseline similarity did emerge across gear levels, but the strength of this pattern varied substantially across activities and participants. walking, vacuuming, and STS each produced two clear cases that followed the expected no→high ordering, while laundry showed the weakest alignment,

with only one participant exhibiting a monotonic decline. Across all activities, many participants showed mid-gear reversals or irregular increases at medium or high gear, indicating that CP-net evaluations were often sensitive to uneven feature distribution within segmented sequences. Although pockets of expected behaviour were present, the overall ordering across participants was inconsistent, highlighting the variability and volatility of CP-net outcomes in this study.

4.4.3 Statistical analysis of gear-level effects for CP-nets

The previous sections presented the CP-net win rates obtained for each participant, activity, and gear level. As with the SCOUT results, some participants exhibited the expected pattern of decreasing win rates as gear level increased for certain activities, while others showed irregular or non-monotonic trends. Visual inspection alone is therefore not sufficient to determine whether gear level has a systematic effect on CP-net win rates.

The same Friedman test and Dunn–Bonferroni post hoc procedure described in Section 4.3.3 were applied here, with CP-net win rates as the dependent variable. Sections 4.4.3.1 to 4.4.3.4 present the results for each activity.

4.4.3.1 Walking: gear-level effects on CP-net win rates

This subsection examines how CP-net win rates for walking vary across the four gear levels (no, low, medium, and high), and whether these differences are statistically significant. Table 4.4.2 presents the descriptive statistics for each gear level. The means and medians show a clear declining pattern from no gear to high gear, and the standard deviations are somewhat larger at the no, low, and high gear levels and smaller at medium gear, although all values remain within a reasonable range.

Overall, these descriptive results align with the expected ADL-fit pattern, with lower win rates at higher gear levels.

Table 4.4.2: Descriptive statistics and mean ranks for CP-net win rates for walking by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.61	0.56	0.19	3.43
Low gear	7	0.46	0.49	0.13	2.86
Medium gear	7	0.42	0.45	0.06	2.14
High gear	7	0.34	0.39	0.15	1.57

Table 4.4.2 summarises the mean ranks of the gear levels. The mean ranks indicate a strong ordering in terms of win-rate ranks, with no gear having the highest average rank, followed by low gear, medium gear, and high gear.

The Friedman test yielded $\chi^2(3) = 8.31$, $p = 0.04$, and Kendall's $W \approx 0.40$, a statistically significant result with a moderate effect size, confirming that CP-net win rates for walking decrease meaningfully as gear level increases.

Post hoc Dunn–Bonferroni pairwise comparisons (see Table C.1.1 in Appendix C) showed that the strongest and statistically confirmed contrast is between the no gear and high gear conditions. No gear win rates were significantly higher than high gear win rates after correction. Other pairwise differences followed the expected numerical ordering but were not strong enough, given the sample size ($n = 7$) and the Bonferroni correction, to be classified as statistically significant.

In summary, for walking, CP-net win rates decrease as simulated aging gear is added, and the Friedman test confirms that gear level has a significant overall effect. The most robust difference occurs between the no gear and high gear conditions, where no-gear sequences achieve significantly higher win rates than high-gear sequences,

consistent with a loss of similarity to the baseline CP-net as the gear level increases.

4.4.3.2 Vacuuming: gear-level effects on CP-net win rates

CP-net win rates for vacuuming declined numerically from no gear (mean 0.56) to high gear (0.37), with low and medium gear occupying intermediate positions, consistent with the expected pattern.

Table 4.4.3: Descriptive statistics and mean ranks for CP-net win rates for vacuuming by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.56	0.59	0.15	3.14
Low gear	7	0.47	0.45	0.14	2.71
Medium gear	7	0.43	0.45	0.07	2.29
High gear	7	0.37	0.38	0.14	1.86

The mean ranks in Table 4.4.3 tell a similar story. In terms of ranked performance, the no gear condition tends to achieve the highest ranks and the high gear condition the lowest, with the low and medium gear levels falling in between. Thus, both the raw win rates and the ranks suggest a monotonic decline in vacuuming performance with increasing gear level.

However, the Friedman test reported $\chi^2(3) = 3.86$ with $p = 0.277$, and Kendall's $W \approx 0.18$, which indicates a small effect size. These results imply that, given the sample size and variability, the observed downward trend in vacuuming win rates is not strong enough to conclude that gear level truly affects CP-net vacuuming win rates at the 5% significance level. The pattern could reasonably be explained by random variation. It is important to note that this does not prove that there is no effect, only that with a sample of seven participants, the evidence is not sufficient to

claim an effect statistically.

Although the Friedman test did not indicate a significant overall gear-level effect, post hoc Dunn–Bonferroni pairwise comparisons were still examined (Table C.2.1 in Appendix C). These comparisons showed that, for vacuuming, there was no specific pair of gear levels for which the difference in CP-net win rates could be declared statistically significant in this sample.

In summary, the CP-net results for vacuuming mirror the SCOUT findings in Section 4.3.3.2: despite a numerical decline in win rates from no gear (mean 0.56) to high gear (0.37), the Friedman test did not reach significance, and no pairwise contrasts survived Bonferroni correction. Taken together, neither evaluator found reliable statistical evidence of a gear-level effect for this activity.

4.4.3.3 Laundry: gear-level effects on CP-net win rates

Laundry showed a sharp decline in CP-net win rates, with the mean dropping from 0.63 at no gear to 0.28 at high gear, the largest absolute difference across gear levels of any activity. The standard deviations are comparable across all gear levels, indicating similar variability in win rates under each condition.

Table 4.4.4: Descriptive statistics and mean ranks for CP-net win rates for laundry by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.63	0.65	0.19	3.14
Low gear	7	0.46	0.45	0.13	2.86
Medium gear	7	0.46	0.48	0.10	2.86
High gear	7	0.28	0.28	0.13	1.14

Table 4.4.4 presents the mean ranks of the gear levels. No gear has the highest mean

rank (3.14), low and medium gear share the same mean rank (2.86), and high gear has the lowest mean rank (1.14). In terms of ranked performance, high gear is clearly the lowest, whereas the no, low, and medium gear conditions form a cluster with similar, higher ranks.

The Friedman test yielded $\chi^2(3) = 10.54$ with $p = 0.014$ and Kendall's $W \approx 0.50$. Subsequent Dunn–Bonferroni pairwise comparisons (Table C.3.1 in Appendix C) showed that high gear win rates are significantly lower than no gear win rates after correction. The high gear condition also has lower win rates than the low and medium gear conditions, but these differences do not quite exceed the Bonferroni-adjusted significance threshold ($\alpha = 0.05$), likely due to the combination of the correction and the small sample size.

In summary, CP-net win rates for laundry clearly decrease as gear level increases, and the Friedman test confirms a significant overall gear-level effect with a moderate-to-strong effect size. Post hoc comparisons show that high gear produces significantly lower win rates than no gear, while other pairwise differences do not reach statistical significance after Bonferroni correction, despite following the expected ordering numerically.

4.4.3.4 Sit-to-stand (STS): gear-level effects on CP-net win rates

This subsection investigates how CP-net win rates for STS differ across the four gear levels for the same seven participants. Table 4.4.5 presents the descriptive statistics for each gear level. The means show a clear drop in win rates from no gear to high gear. Low and medium gear have very similar mean win rates (0.45 and 0.46, respectively), while high gear clearly has the lowest mean win rate and median. The standard deviations are all within a reasonable range, with slightly higher variability for the no and medium gear conditions and lower variability for the low and high

gear conditions.

Table 4.4.5: Descriptive statistics and mean ranks for CP-net win rates for sit-to-stand (STS) by gear level.

Gear level	N	Mean	Median	Standard deviation	Mean rank
No gear	7	0.63	0.63	0.15	3.43
Low gear	7	0.45	0.49	0.07	2.57
Medium gear	7	0.46	0.44	0.13	2.43
High gear	7	0.33	0.31	0.08	1.57

Table 4.4.5 presents the mean ranks for the four gear levels. No gear has the highest mean rank and high gear the lowest, with low and medium gear again very close to each other. Overall, the ranking matches the expected pattern: no gear is ranked highest, high gear is ranked lowest, and low and medium gear occupy intermediate positions.

Figure 4.3 summarises the mean win rates across all four activities and gear levels.

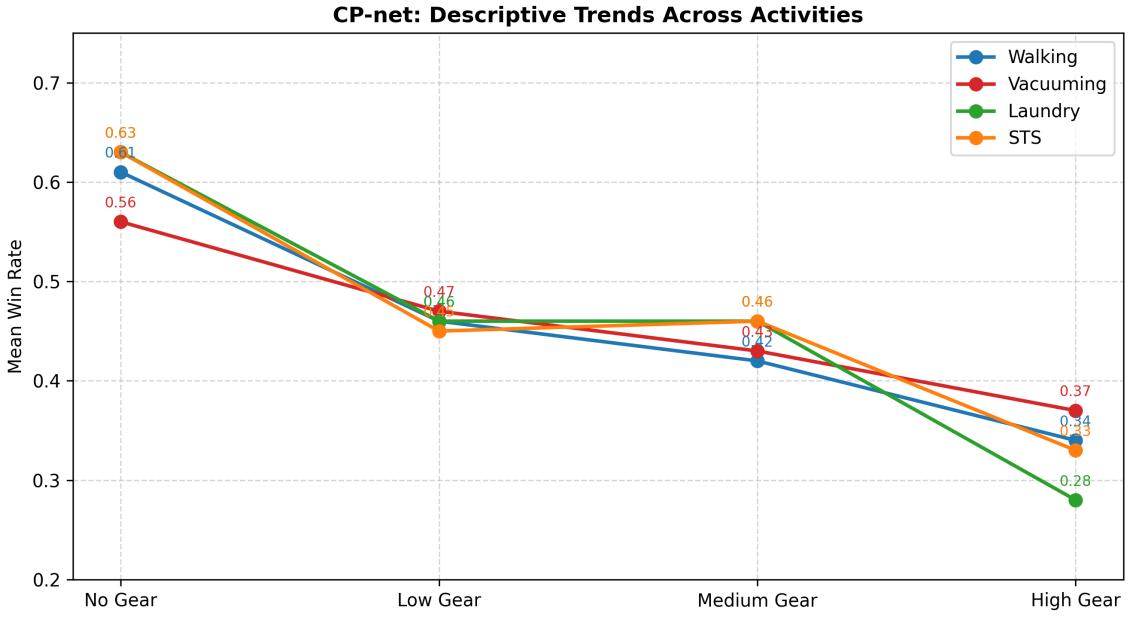


Figure 4.3: Mean CP-net win rates across gear levels for all four activities.

The Friedman test yielded $\chi^2(3) = 7.29$, $p = 0.063$, and Kendall's $W \approx 0.35$. The p-value of 0.063 is close to the conventional significance threshold of $\alpha = 0.05$, but since $p > \alpha$, the test does not provide sufficient evidence to claim an overall gear-level effect on STS CP-net win rates at the 5% level. In other words, the observed trend could plausibly be attributed to random variation in this small sample.

The Post hoc Dunn–Bonferroni pairwise comparisons were nevertheless examined (Table C.4.1 in Appendix C). These comparisons showed that no gear win rates are significantly higher than high gear win rates after Bonferroni correction, whereas all other pairwise differences were not statistically significant. Thus, although the overall evidence for a global gear-level effect is weak, there is a strong specific contrast between the no gear and high gear conditions.

In summary, for STS, CP-net win rates show a clear numerical decline from no gear to high gear, and the mean ranks indicate the expected ordering (no $>$ low $\approx$ medium $>$ high). However, the overall Friedman test does not reach the 5% significance level ($p = 0.063$), suggesting that the evidence for a global gear-level effect is borderline in this sample. Post hoc Dunn–Bonferroni comparisons nevertheless reveal a significant contrast between no gear and high gear, with higher win rates under no gear, while other pairwise differences are not statistically significant after correction.

4.4.4 Relationship Between CP-net Structure and Result Quality

This section investigates how characteristics of the learned CP-net structures relate to the quality of their ordering-query results. Three structural features are examined: overall CP-net connectivity (the proportion of nodes belonging to at least one connected component), the total number of conditional nodes (nodes with at least

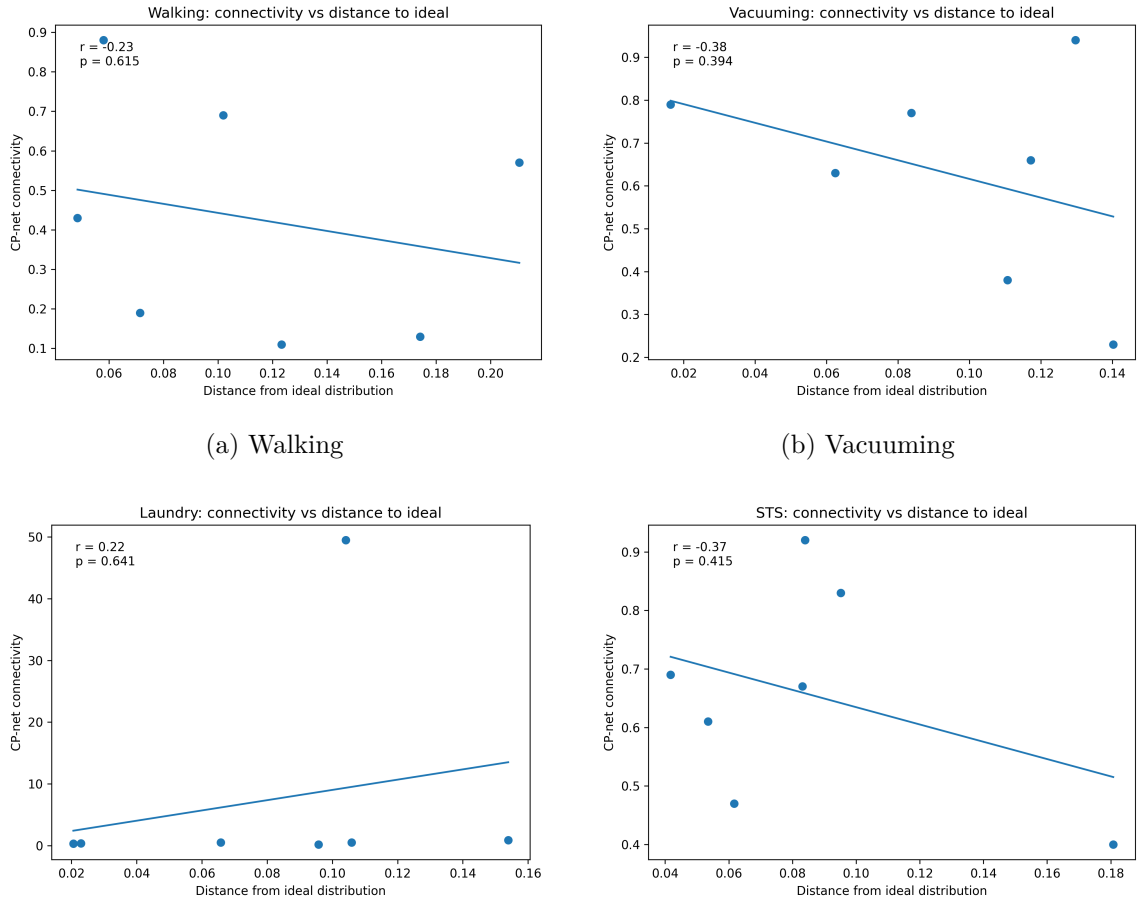
one parent), and the average node degree (the mean number of edges per node). For each participant and activity, these features were correlated against a distance measure quantifying how closely the observed gear-level win distribution matched the ideal decline pattern $q = (0.5,\ 0.333,\ 0.167,\ 0.0)$, with smaller distances indicating better alignment.

The central finding across all three metrics is that structural complexity did not reliably improve result quality, and in several cases was associated with poorer alignment to the ideal ordering. None of the metrics produced strong or statistically significant correlations with the distance measure for any activity. More strikingly, laundry, the activity with the clearest SCOUT results, consistently showed positive correlations across all three metrics, meaning that CP-nets with greater connectivity, more conditional nodes, and higher average node degree actually tended to produce worse gear-level orderings for this activity. This pattern suggests that additional conditional structure sometimes captured noise and spurious dependencies rather than stable movement regularities. The sub-analyses below detail the correlation results for each metric, with the figures providing scatter plots of each structural feature against the distance measure.

4.4.4.1 CP-net Connectivity and Results Quality

Figure 4.4 shows CP-net connectivity plotted against distance to the ideal distribution for each activity. For Walking ($r = -0.23$, $p = .62$), vacuuming ($r = -0.38$, $p = .39$), and STS ($r = -0.37$, $p = .42$), connectivity showed weak negative correlations in the expected direction, meaning that participants with more connected CP-nets tended to have slightly smaller distances from the ideal pattern. However, all three effects were small in magnitude and none reached statistical significance, so these trends should be treated as tentative rather than reliable.

laundry departed from this pattern, showing a small positive correlation ($r = 0.22$, $p = .64$), driven in part by one participant with unusually high connectivity whose CP-net was nonetheless associated with a large distance from the ideal ordering. Overall, connectivity showed the most consistent directional signal of the three metrics for walking, vacuuming, and STS, but the evidence remains too weak to treat it as a dependable predictor of result quality.



(a) Walking

(b) Vacuuming

(c) Laundry

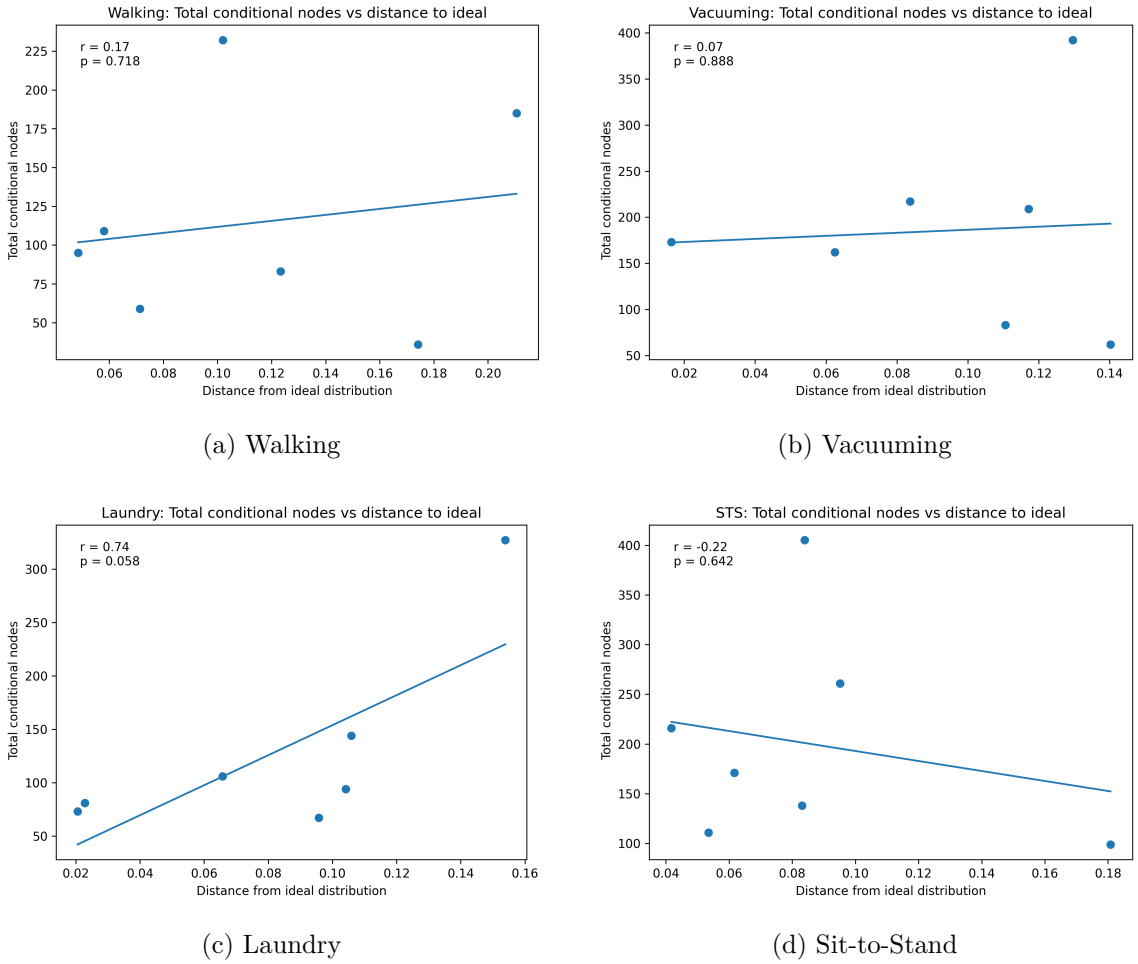
(d) Sit-to-Stand

Figure 4.4: CP-net connectivity versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression.

4.4.4.2 Number of Conditional Nodes and Results Quality

Figure 4.5 shows the total number of conditional nodes plotted against distance for each activity. For Walking ($r = 0.17$, $p = .72$) and vacuuming ($r = 0.07$, $p = .89$), the correlations were near zero, indicating that the absolute size of the conditional core had no meaningful relationship with how closely the win distributions followed the ideal decline. STS showed a weak negative trend ($r = -0.22$, $p = .64$), consistent with the direction seen for connectivity, but again far from significant.

laundry produced the strongest correlation of this sub-analysis ($r = 0.74$, $p = .06$), and it was positive, meaning participants with more conditional nodes tended to have larger distances from the ideal ordering. Although this does not reach the conventional $\alpha = 0.05$ threshold, the size and direction of the effect are consistent with the interpretation that a larger conditional core captured noise rather than informative structure for this activity.



(a) Walking

(b) Vacuuming

(c) Laundry

(d) Sit-to-Stand

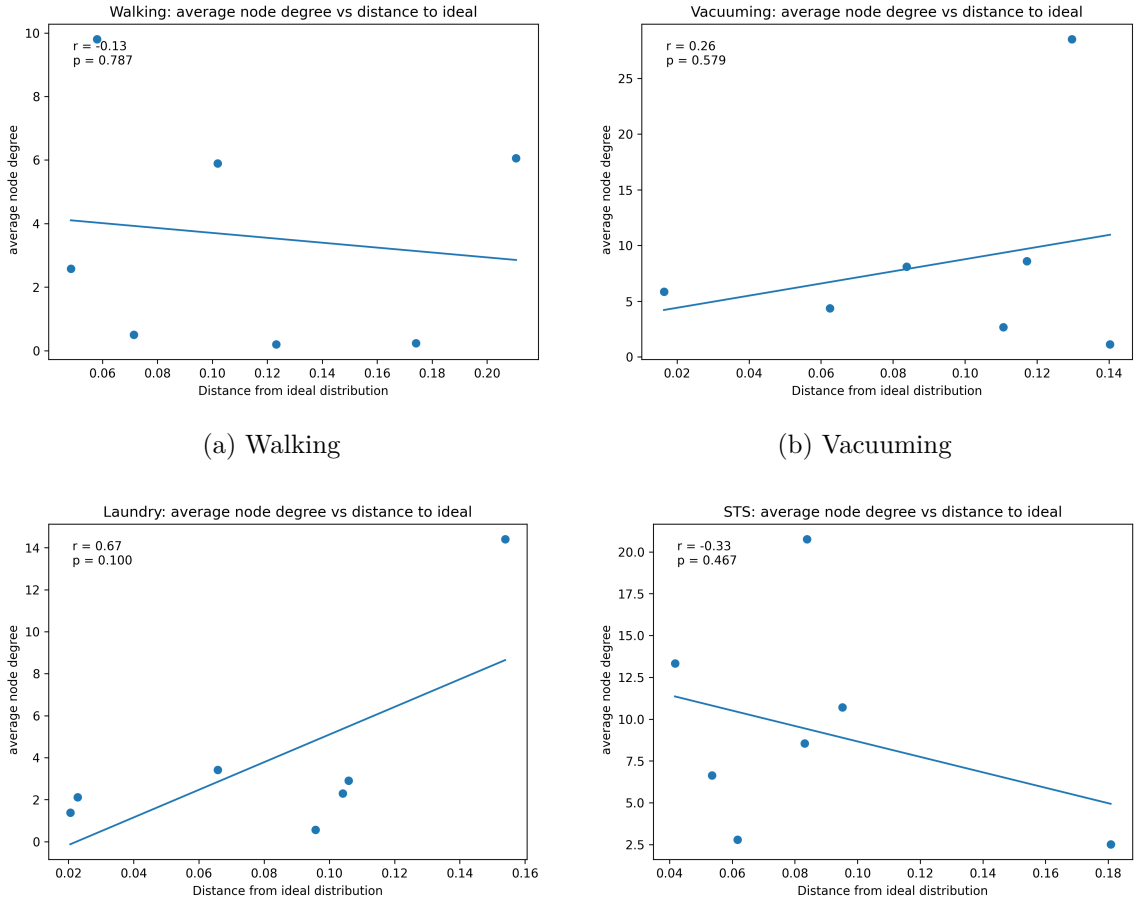
Figure 4.5: Number of conditional nodes versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression.

4.4.4.3 Average Node Degree and Results Quality

Figure 4.6 shows average node degree plotted against distance. The pattern mirrors the previous two analyses closely. walking ($r = -0.13$, $p = .79$) and vacuuming ($r = 0.26$, $p = .58$) showed near-zero correlations, and STS showed a weak negative association ($r = -0.33$, $p = .47$). As with the other metrics, none of these effects approached statistical significance.

Laundry again showed a positive correlation ($r = 0.67$, $p = .10$), with the participant

having the highest average node degree also having the largest distance from the ideal distribution. This reinforces the interpretation introduced at the start of this section: for laundry, denser conditional structure was not associated with better results and may have introduced additional instability into the ordering queries.



(a) Walking

(b) Vacuuming

(c) Laundry

(d) Sit-to-Stand

Figure 4.6: Average node degree versus distance to the ideal distribution for all four activities. Each point represents one participant, and the line shows the fitted linear regression.

4.4.4.4 Summary of Structural Correlation Analyses

Across all three metrics, the results tell a consistent story: in the tests conducted here, increased structural richness in CP-nets, whether measured by connectivity, the

total number of conditional nodes, or local node density, was not associated with improved gear-level ordering quality. For laundry, higher connectivity coincided with worse results; however, this correlation may suggest that structural connectivity had little positive effect on gear-level ordering, though it is equally possible that the connectivity arose from noisier data that independently produced poorer outcomes, or that without this connectivity, results may have been even worse. Without systematically controlling the number of conditional preferences, it is difficult to determine whether structural richness helped, hindered, or had no meaningful effect on ordering quality. STS is the one exception worth noting, as it is the only activity where all three metrics showed weak negative correlations with distance, hinting that structural richness may be modestly more beneficial for this task. For walking and vacuuming, the correlations were close to zero across all metrics."

These findings suggest that the usefulness of structural complexity in CP-nets depends on whether the mined dependencies capture stable, task-specific regularities rather than on structural richness alone. When input patterns contain tracking noise and occlusion artefacts, adding more conditional structure risks amplifying that noise rather than improving the model's ability to detect functional decline.

4.5 Discussion of Findings

This section interprets the results presented above in relation to the two research questions introduced in Chapter 1. RQ1 asked whether SCOUT's SPM classifier could detect declines in ADL fit under increasing levels of simulated aging gear. RQ2 asked whether CP-nets could identify performance deterioration by ranking lower-gear sequences as better fits to baseline movement. In both cases, the quality of the underlying skeletal data, which was affected by a low effective frame rate and

considerable measurement noise, is central to understanding what the results mean.

4.5.1 Comparing SCOUT and CP-net Performance

The most consistent finding across both evaluators is that performance deterioration was detectable for laundry and walking, but not reliably so for vacuuming and STS. For laundry and walking, SCOUT confidence scores and CP-net win rates both declined from no gear to high gear, and the Friedman tests confirmed statistically significant gear-level effects for both methods. This convergence suggests that when movement patterns are stable, whole-body, and sufficiently distinct between gear conditions, both the classifier-based and preference-based approaches can detect functional decline in a similar and mutually reinforcing way.

For vacuuming and STS, the two evaluators diverge. SCOUT showed numerical declines in confidence scores but without reaching statistical significance, while CP-nets displayed partial declines combined with frequent mid-gear reversals. Neither method consistently reinforced the other for these activities. This divergence most likely reflects how both evaluators respond to the same underlying data problem: both SCOUT and the CP-net builder operate on the same symbolic sequences derived from noisy skeletal input, so any instability in those sequences propagates into both the classifier's features and the CP-net's conditional dependencies. Vacuuming involves complex, variable arm and hand movements, while STS comprises brief discrete transitions sensitive to precise timing. Both characteristics amplify the impact of poor hand and foot tracking and low frame rates.

The structural analyses of the CP-nets further support this interpretation. None of the three structural metrics, namely connectivity, number of conditional nodes, and average node degree, showed reliable positive associations with result quality. Laundry specifically exhibited positive correlations where more structural complexity

coincided with poorer alignment to the ideal ordering, indicating that additional conditional structure sometimes encoded noise and spurious pattern dependencies rather than genuine movement regularities.

4.5.2 Findings for RQ1: Evaluating ADL Fit with SCOUT

SCOUT's sensitivity to gear-level changes was strongly shaped by the quality of the skeletal data. For activities with robust, full-body movement patterns, particularly laundry and walking, the SPM classifier captured declining similarity to the baseline as gear level increased, with significant Friedman test results and moderate to large effect sizes. For vacuuming and STS, the results were weaker and more irregular, a pattern most plausibly explained by the limitations of the data.

All four activities depend heavily on hand and foot movement, yet tracking quality for these extremities was noticeably poorer than for the trunk and major joints. Coupled with the low effective frame rate (4 fps), this made it difficult for the discretisation and pattern-mining stages to distinguish genuine gear-level changes from tracking jitter and occlusion artefacts. The prototype NRC software compounded this through calibration mismatches between cameras and intermittent desynchronisation, introducing brief unrealistic jumps in joint positions that propagated directly into the symbolic sequences SCOUT received as input.

Where movement was distinctive and stable enough to overcome this noise floor, as in laundry and walking, SCOUT detected meaningful decline. Where fine-grained timing or extremity movement played a larger role, as in vacuuming and STS, the signal was attenuated, resulting in the weak and inconsistent p-values observed in the statistical analyses.

4.5.3 Findings for RQ2: Evaluating ADL Fit with CP-nets

The CP-net analyses provide a complementary perspective based on pairwise preference comparisons. At the individual level, many win-rate profiles showed at least a partial decline from no gear to high gear, particularly for walking, vacuuming, and STS, but mid-gear reversals were common and some participants had low- or medium-gear sequences that outperformed their no-gear recordings. At the group level, significant gear-level effects were confirmed for walking and laundry, STS produced a borderline result with a significant no-versus-high contrast in post hoc comparisons, and vacuuming showed no significant effect at either level.

These patterns are consistent with the CP-net's dependence on a stable, informative feature set. When input sequences contain substantial noise from tracking errors and occlusion-related guessing, the pattern-mining stage produces unstable conditional dependencies, and small fluctuations in the data translate into volatile win-rate orderings. The frequent mid-gear reversals observed across activities reflect this instability, as the CP-net sometimes learned preferences that reflected measurement artefacts rather than systematic gear-level differences. Despite this, the significant no-versus-high contrasts for walking, laundry, and STS confirm that the preference structures retained useful information about gross changes in movement quality, even when finer intra-individual distinctions across middle gear levels remained out of reach.

4.6 Limitations of the Study

The findings presented in this chapter should be interpreted in light of several important limitations related to the quality of the motion data, the data collection procedure, the design of the simulated aging conditions, and the statistical properties

of the study.

A primary limitation concerns the quality of the skeletal data used as input to both SCOUT and the CP-net builder. Although the recording system was configured to operate at 15 frames per second, the effective frame rate was closer to 4 frames per second. This low temporal resolution restricted the amount of information available for capturing fast or subtle changes in movement, particularly during short transitions such as those that occur in the Sit-to-Stand task. With fewer frames per activity, the discretisation and windowing procedures had limited opportunity to capture fine-grained differences in timing and joint trajectories between gear levels.

The reliability of the skeletal tracking further constrained the data. The two depth cameras frequently tracked the trunk, hips, and major joints well, but they were considerably less accurate and stable when estimating the positions of the hands and feet. All four activities studied in this thesis depend heavily on the movement of these extremities. For example, vacuuming involves repeated arm and hand motions, laundry requires reaching and grasping, and walking and STS rely on precise foot placement and leg trajectories. At the same time, the differences between gear levels are subtle, because participants perform the same activity with and without added restrictions. When the tracking quality for hands and feet is poor, the symbolic sequences derived from these joints may fail to capture these subtle changes, or may capture them only weakly. This limitation directly affects the discriminative power of the features used by SCOUT and the patterns that form the CP-net nodes, and it provides a plausible explanation for the small and often non-significant correlations between CP-net structural measures and result quality.

Additional noise arose from the prototype NRC software used to record and merge the skeletal data from the master and sub cameras. In practice, the calibration between the two cameras was often imperfect, and the merged skeleton occasionally

displayed brief misalignments between the views. When one camera could not observe a particular joint due to occlusion, the system attempted to infer its position from the visible limb, which sometimes produced unrealistic jumps or jitter in the reconstructed joint locations. These artefacts propagate into the CSV files, where they appear as spurious fluctuations in the joint coordinates and, after discretisation, as irregular symbol transitions. As a result, SCOUT and the CP-net builder operated on data that contained a mixture of genuine movement differences and tracking artefacts, which likely reduced the clarity of gear-level trends and contributed to the weak p-values in both the SCOUT and CP-net analyses.

The study also has several more conventional limitations. The sample size was small, with only seven participants completing all activities under all four gear levels. This limited the statistical power of the Friedman tests and the Dunn–Bonferroni post hoc comparisons, especially when effect sizes were modest and inter-participant variability was substantial. In several cases, numerical trends and moderate Kendall's W values did not reach the conventional significance threshold after correction, which may reflect a lack of power rather than the absence of a true effect.

A further limitation concerns the computational feasibility of the dominance testing method implemented in the CP-net solver. As described in Section 3.4.2, dominance testing via improving flips determines whether one outcome is a better fit over another by checking whether a sequence of locally improving value changes can transform one outcome into the other. While this method provides a semantically stricter form of outcome comparison than the ordering-query-based approaches, it proved computationally intractable for the sequence lengths and CP-net sizes encountered in this study. In practice, evaluation runs consistently exceeded four hours without completing, ultimately terminating due to memory allocation failure as the available system memory was exhausted. This behaviour is consistent with the known worst-

case complexity of dominance testing in CP-nets, which in general requires exploring a number of improving flip sequences that grows exponentially with the number of differing variables between two outcomes. As a result, no quantitative results could be obtained for the dominance testing method, and the CP-net analyses presented in this chapter are based entirely on the ordering-query-based approaches. Improving the scalability of dominance testing, for example through pruning strategies or approximate methods, remains an open direction for future work.

A further practical limitation concerns the single-person constraint of the current sensing setup. The depth cameras and skeletal tracking software are designed to track one person at a time, which means the system cannot operate in multi-occupant households without modification. While this is not a barrier for older adults who live alone, which is in fact the primary target population for this kind of monitoring system, it does reduce the practical applicability of the current implementation in shared living environments.

Finally, the activity set and environment were constrained. Only four activities were analysed, and all recordings were collected in a controlled laboratory setting designed to resemble a home. Although this environment provided consistency across participants, it does not fully capture the variability and unpredictability of real homes and daily routines. The conclusions drawn here therefore apply most directly to this specific set of tasks and conditions.

A fundamental limitation concerns the real-world translatability of the current framework. All experiments were conducted in a controlled laboratory setting, and the models were trained using recordings that explicitly represented two clearly separated conditions: no-gear recordings as the healthy baseline class and high-gear recordings as the deterioration class. In a real deployment scenario, obtaining such clearly labelled data for a new user would require a controlled data collection session, which

may not always be feasible. There is no guarantee that movement data collected naturalistically in a home environment could be cleanly separated into baseline and worst-case categories without such a session. This means that the claims made in this thesis about detecting functional performance deterioration apply specifically to this controlled experimental setup, and should not be interpreted as evidence that the framework is ready for deployment in uncontrolled, real-world settings without further development.

Despite these limitations, the study offers a useful first step toward understanding how SPM-based classification and CP-net preference models behave under controlled manipulations of movement quality. The data-quality constraints and small sample size likely led to conservative estimates of the true effect sizes. They also highlight clear avenues for future work, including improvements to the sensing pipeline, more robust camera calibration, higher frame rates, and the collection of larger and more diverse datasets.

4.7 Conclusion of the Results Chapter

This chapter evaluated whether SCOUT's SPM classifier and CP-net preference models can detect changes in ADL performance under simulated aging conditions, and how CP-net structural properties relate to the quality of the resulting gear-level orderings. For RQ1, the analyses of SCOUT Class 0 confidence scores showed that laundry and walking exhibited clear and statistically supported declines in baseline similarity as gear level increased, while vacuuming and STS displayed weaker and more irregular trends. These results suggest that SCOUT can capture aspects of performance deterioration, particularly for activities with rich full-body movement patterns, but that its sensitivity is strongly influenced by task characteristics and by

the quality of the underlying skeletal data.

For RQ2, the CP-net win-rate analyses provided a complementary view based on pairwise preference comparisons. Walking and laundry again showed significant gear-level effects, with no-gear sequences achieving higher win rates than high-gear sequences and clear differences between these extremes. Vacuuming and STS exhibited numerical declines with smaller effect sizes and frequent mid-gear reversals, and the evidence for global gear-level effects was weaker. The structural analyses of the CP-nets further indicated that connectivity, number of conditional nodes, and average node degree were not strong or consistent predictors of result quality, and that in some cases more complex CP-net structures coincided with poorer alignment to the ideal gear-level pattern.

Across both research questions, the noisy and low-frame-rate skeletal data provide an important context for interpreting the results. The combination of reduced temporal resolution, imperfect tracking of hands and feet, camera calibration issues, and occlusion-related guessing introduced considerable variability into the symbolic sequences used by SCOUT and the CP-net builder. These factors likely attenuated true gear-level differences and contributed to the modest Pearson correlations and non-significant p-values observed in several analyses. The results presented in this chapter can therefore be viewed as conservative estimates of the extent to which pattern-based methods can detect subtle intra-individual changes in ADL performance when they operate on imperfect motion data.

Overall, the findings demonstrate that SCOUT and CP-nets can both capture meaningful aspects of performance decline under simulated aging, particularly for laundry and walking, but they also highlight the limitations of the current sensing and modelling pipeline. The chapter provides empirical support for the feasibility of using sequential pattern mining and preference-based evaluation to study ADL

fit, while at the same time motivating methodological refinements. The next chapter builds on these observations by situating the results within the broader literature on aging in place and movement assessment, discussing their implications for future monitoring systems, and outlining directions for improving data collection, feature extraction, and model design in subsequent work.

Chapter 5

Conclusion and Future Work

This chapter summarises the key findings of the study, discusses its limitations, outlines directions for future work, and closes with a reflection on the broader significance of the research.

5.1 Key Findings

This thesis evaluated a hybrid framework that combines SCOUT's Sequential Pattern Mining classifier with Conditional Preference Networks to detect changes in the quality of ADL performance under simulated aging conditions. The analysis addressed two research questions: whether SCOUT's Class 0 confidence scores decrease as physical restriction increases, and whether CP-net win-rate orderings reflect the expected decline in ADL fit from no gear to high gear. Overall, the findings suggest that the proposed framework can capture aspects of performance deterioration, but that these effects are modest, activity-dependent, and sensitive to data quality and modelling choices.

For Research Question 1, SCOUT showed partial support for the expected decline in

baseline similarity as gear level increased. At the descriptive level, Class 0 confidence scores often decreased from no gear to high gear, particularly for walking, laundry, and the Sit-to-Stand (STS) task. Friedman tests confirmed significant gear-level effects for walking and STS, with moderate effect sizes, and a strong and highly significant effect for laundry. Post hoc comparisons for laundry showed that high-gear confidence scores were significantly lower than those for no gear and low gear, indicating that SCOUT was able to separate the most restricted condition from baseline performance for this complex full-body activity. In contrast, vacuuming produced more variable and often non-monotonic patterns, and the Friedman test did not identify a significant overall gear-level effect. Taken together, these findings indicate that SCOUT can detect declines in ADL fit for some activities under controlled conditions, especially laundry and, to a lesser extent, walking and STS, but that the strength and consistency of the effect depend on the movement characteristics of the task.

For Research Question 2, CP-net win rates provided a complementary but more volatile view of performance change. At the descriptive level, many participants showed decreasing win rates from no gear to high gear for at least one activity, and the average win rates across participants often followed the expected ordering. Friedman tests identified significant gear-level effects for walking and laundry and a borderline effect for STS, while vacuuming did not show a statistically reliable effect. When gear-level differences were present, post hoc Dunn–Bonferroni tests most consistently identified a significant separation between no gear and high gear, indicating that CP-nets could distinguish the most restricted condition from baseline for some activities. However, many individual participants exhibited mid-gear reversals or irregular orderings, especially for vacuuming and laundry, which suggests that CP-net evaluations were highly sensitive to uneven feature distributions across segmented sequences and to variation in the learned conditional structure.

A third set of analyses examined how structural properties of the learned CP-nets related to the quality of their gear-level orderings. Pearson correlations between the distance to an ideal gear-level win distribution and three structural features (overall connectivity, number of conditional nodes, and average node degree) were small and non-significant for all activities. For walking, the correlations were close to zero, indicating no clear link between structural complexity and result quality. For STS, all three metrics showed modest negative correlations, hinting that richer structure might be mildly beneficial, while for laundry the correlations were positive and relatively large, suggesting that more complex CP-nets could even degrade result quality for that activity. These findings indicate that simply increasing CP-net complexity does not guarantee better alignment with the expected decline in ADL fit.

Overall, the results show that the proposed framework can detect some aspects of performance deterioration under simulated aging conditions, particularly at the extremes between no gear and high gear and especially for laundry and walking. However, the effects are often weak or inconsistent, and both SCOUT and CP-nets are strongly influenced by task-specific movement characteristics and by the quality and stability of the underlying skeletal data. The primary contribution of this thesis is a proof-of-concept demonstration that sequential pattern mining and preference-based evaluation can be combined into a coherent pipeline for assessing ADL fit under simulated aging conditions. A secondary contribution is the empirical analysis of how CP-net structural properties, namely connectivity, number of conditional nodes, and average node degree, relate to ordering quality, which to the best of our knowledge has not been previously examined in this context. Together, these contributions provide a concrete foundation and a set of identified bottlenecks for future research into movement-based functional health monitoring.

5.2 Limitations

The study's limitations fall into two broad categories, discussed in detail in Section 4.6. The first concerns data quality: the low effective frame rate (4 fps), unreliable hand and foot tracking, and camera calibration artefacts constrained what the models could learn and likely attenuated the gear-level effects observed across both SCOUT and CP-net analyses. The second concerns scope and generalisability: the small sample of seven participants, the use of simulated rather than real aging, the controlled laboratory setting, and the specific parameter configurations used for SCOUT and CP-net modelling all limit how far the findings can be extended to real-world monitoring scenarios.

A related modelling limitation concerns the labelling of training data. In this study, the SPM and CP-net models were trained using no-gear recordings as the baseline class and high-gear recordings as the deterioration class. In a real deployment, this labelling is not straightforward and there is no guarantee that a new user's data can be cleanly separated into these two categories without a controlled experiment. This represents a fundamental challenge for translating the current framework into a working prototype.

Additionally, the single-person constraint of the current camera setup limits deployment to single-occupant settings, which, while aligned with the core use case of supporting independently living older adults, reduces broader applicability.

Together, these constraints help explain why many observed effects were modest, why several statistical results did not reach conventional significance thresholds, and why CP-net structural complexity did not reliably predict result quality. They also point directly to the future work directions outlined in Section 5.3.

5.3 Future Work

Several avenues of future work follow naturally from the limitations and findings of this study and could strengthen the ability of SCOUT and CP-net based methods to detect deterioration in ADL performance.

A first priority is to improve the quality and stability of the sensor data. This includes deploying a more mature, GPU-accelerated data-collection pipeline that can maintain the intended frame rate, reduce dropped frames, and produce smoother skeletal trajectories. Future studies could also incorporate updated depth cameras or multimodal sensing, for example combining depth, inertial measurement units, or pressure sensors, to obtain more reliable estimates of hand and foot motion. More robust calibration procedures and automatic quality checks for the merged skeleton would further reduce jitter and misalignment between multiple camera views.

A second direction is to expand the scope and realism of the dataset. Collecting data from a larger and more diverse sample of participants, including older adults and individuals with early functional decline, would provide a stronger basis for evaluating whether the framework can generalise beyond simulated aging gear. Longitudinal data that track the same individuals over time would be particularly valuable for studying intra-individual change. Additional activities that cover a wider range of ADLs, such as cooking tasks with more object interactions or complex transfers, could also help test how the models perform under richer, more ecologically valid conditions.

Addressing the cold-start labelling problem is a key direction for future work. One practical approach would be to use simulated no-gear and high-gear recordings, similar to the protocol used in this thesis, to establish an initial training baseline for a new user. As the system accumulates real ADL recordings over time, sequences that score

closest to the simulated no-gear baseline could be used to incrementally retrain the Class 0 model, while sequences that score furthest from the baseline could be used to update the Class 1 model. This form of incremental self-supervised retraining would allow the system to adapt to an individual's movement patterns without requiring a controlled gear experiment for every new user.

A central goal for future work is to move the framework from a research prototype toward a deployable system that can genuinely support aging in place. This involves several practical steps. First, data should be collected through participatory design with older adults, engaging the actual target population in both the data collection process and the evaluation of system feedback. This would not only produce more ecologically valid training data but also ensure the system is designed around the real needs and routines of older adults rather than simulated conditions. Second, the cold-start labelling problem described above needs to be resolved before the system can be deployed for new users without a controlled gear experiment. The incremental retraining approach outlined above is one candidate solution. Third, as more real-world data accumulates from deployed systems, the models can be progressively refined to better capture individual variation in ADL performance and to detect more subtle early-stage changes in functional health. The long-term vision is a lightweight, non-invasive monitoring system that runs continuously in the background of an older adult's home, flags meaningful changes in ADL quality, and supports timely intervention, helping people live independently for as long as it is safely possible.

5.4 Final Remarks

This thesis set out to explore whether a hybrid framework based on SCOUT's Sequential Pattern Mining classifier and Conditional Preference Networks could detect changes in the quality of ADL performance under simulated aging conditions. The work combined a custom dual-camera data-collection system, age-simulation gear, symbolic sequence mining, and preference-based evaluation in order to study how movement patterns shift as physical restrictions increase.

The findings show that the framework can capture aspects of performance deterioration, especially when contrasting no gear with high gear and particularly for activities such as laundry and walking that place sustained demands on balance and coordination. At the same time, the results highlight how challenging it is to detect degradation in ADL fit using noisy skeletal data, a small sample, and complex movement patterns. Both SCOUT and CP-nets showed activity-specific strengths and weaknesses, and structural complexity in the CP-nets was not a simple predictor of result quality.

Despite these limitations, the thesis demonstrates that it is feasible to integrate sequential pattern mining and preference-based models into a coherent pipeline for analysing ADL performance. The work also identifies the key bottlenecks that must be addressed for such systems to provide reliable early indicators of functional decline, namely improved sensing, more stable features, and more robust modelling. By documenting both the promise and the difficulties of this approach, the thesis contributes to the broader effort to support aging in place through unobtrusive, movement-based monitoring of everyday activities.

In the future, refinements in sensing, modelling, and experimental design can build on these foundations to move closer to systems that can detect deterioration in ADL

performance before they manifest as falls, hospitalisations, or loss of independence. What this thesis demonstrates is that the building blocks for such a system, sequential pattern mining, model-based comparison, and structural analysis of learned CP-nets, can be assembled into a coherent evaluation pipeline and applied to real skeletal data under controlled aging simulation. The path from proof of concept to a deployed system that genuinely supports older adults in living independently is long, but this work maps one section of it.

Bibliography

[1] Josky Aïzan, Cina Motamed, and Eugène C Ezin, *Activity mining in a smart home from sequential and temporal databases*, Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020), SCITEPRESS, 2020, pp. 542–547.

[2] Eisa Alanazi, Malek Mouhoub, and Sandra Zilles, *The complexity of exact learning of acyclic conditional preference networks from swap examples*, Artificial Intelligence **278** (2020), 103182.

[3] Hassan Alkhiri and Malek Mouhoub, *Weighted constrained CP-nets: an extension of constrained CP-nets with weighted constraints*, 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2021, pp. 1685–1690.

[4] Hassan Alkhiri and Malek Mouhoub, *Constrained CP-nets similarity.*, ICAART (3), 2022, pp. 226–233.

[5] Jay Ayres, Jason Flannick, Johannes Gehrke, and Tomi Yiu, *Sequential pattern mining using a bitmap representation*, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 429–435.

[6] Cyriac Azefack, Raksmey Phan, Vincent Augusto, Guillaume Gardin, Claude Montuy Coquard, Remi Bouvier, and Xiaolan Xie, *An approach for*

behavioral drift detection in a smart home, 2019 IEEE 15th international conference on automation science and engineering (CASE), IEEE, 2019, pp. 727–732.

[7] Adel Baselizadeh, Md Zia Uddin, Weria Khaksar, Diana Saplacan Lindblom, and Jim Torresen, *Prima-care: Privacy-preserving multi-modal dataset for human activity recognition in care robots*, Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, 2024, pp. 233–237.

[8] Damien Bigot, Bruno Zanuttini, Hélene Fargier, and Jérôme Mengin, *Probabilistic conditional preference networks*, Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013), 2013, pp. 72–81.

[9] Craig Boutilier, Ronen I Brafman, Carmel Domshlak, Holger H Hoos, and David Poole, *CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements*, Journal of Artificial Intelligence Research **21** (2004), 135–191.

[10] Adriana Coelho, Vítor Parola, Daniela Cardoso, Susana Duarte, Maria Almeida, and João Apóstolo, *The use of the aged simulation suit in nursing students: a scoping review*, Revista de Enfermagem Referência **4** (2017), no. 14, 147–158.

[11] Cristina Cornelio, Judy Goldsmith, Umberto Grandi, Nicholas Mattei, Francesca Rossi, and K Brent Venable, *Reasoning with PCP-nets*, Journal of Artificial Intelligence Research **72** (2021), 1103–1161.

[12] Axel Czabke, J Loeschke, and Tim C Lueth, *Concept and modular telemedicine platform for measuring of vital signs, ADL and behavioral patterns of elderly in home settings*, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 3164–3167.

[13] Giuseppe D'Aniello, Roberto Falcone, Matteo Gaeta, Zia Ur Rehman, and Giancarlo Fortino, *A sequential pattern mining approach for situation-aware*

human activity projection, 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2024, pp. 2171–2176.

[14] R Deeptha, K Ramkumar, Sri Venkateswaran, Mohammad Mehedi Hassan, Md Rafiul Hassan, Farzan M Noori, and Md Zia Uddin, *Enhancing human activity recognition for the elderly and individuals with disabilities through optimized internet-of-things and artificial intelligence integration with advanced neural networks*, Frontiers in Neuroinformatics **18** (2024), 1454583.

[15] Peter F Edemekong, Deb Bomgaars, Sukesh Sukumaran, and Shoshana B Levy, *Activities of Daily Living*, StatPearls, StatPearls Publishing LLC, Treasure Island, FL, 2019.

[16] Shirin Enshaeifar, Ahmed Zoha, Andreas Markides, Severin Skillman, Sahr Thomas Acton, Tarek Elsaleh, Masoud Hassanpour, Alireza Ahrabian, Mark Kenny, Stuart Klein, et al., *Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques*, PLOS ONE **13** (2018), no. 5, e0195605.

[17] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Rage Uday Kiran, Yun Sing Koh, and Rincy Thomas, *A survey of sequential pattern mining*, Data Science and Pattern Recognition **1** (2017), no. 1, 54–77.

[18] Thomas Gerhardy, Laura Schmidt, Hans-Werner Wahl, Katja Mombaur, and Lizeth Sloot, *Do aging suits adequately simulate objective age-related decline in gait? a kinematic comparison of induced aging in young and middle-aged adults*, Educational Gerontology (2025).

[19] Thomas H Gerhardy, Anna Schlomann, Hans-Werner Wahl, and Laura I Schmidt, *Effects of age simulation suits on psychological and physical outcomes: a systematic review*, European journal of ageing **19** (2022), no. 4, 953–976.

[20] Tao Gu, Liang Wang, Zhanqing Wu, Xianping Tao, and Jian Lu, *A pattern mining approach to sensor-based human activity recognition*, IEEE Transactions on Knowledge and Data Engineering **23** (2010), no. 9, 1359–1372.

[21] Tao Gu, Zhanqing Wu, Xianping Tao, Hung Keng Pung, and Jian Lu, *epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition*, 2009 IEEE international conference on pervasive computing and communications, IEEE, 2009, pp. 1–9.

[22] Jiawei Han, Jian Pei, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal, and Meichun Hsu, *Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth*, proceedings of the 17th international conference on data engineering, IEEE Piscataway, NJ, USA, 2001, pp. 215–224.

[23] Marwan Hassani, Sebastiaan J van Zelst, and Wil MP van der Aalst, *On the application of sequential pattern mining primitives to process discovery: Overview, outlook and opportunity identification*, Wiley interdisciplinary reviews: data mining and knowledge discovery **9** (2019), no. 6, e1315.

[24] Ahatsham Hayat, Fernando Morgado-Dias, Bikram Pratim Bhuyan, and Ravi Tomar, *Human activity recognition for elderly people using machine and deep learning approaches*, Information **13** (2022), no. 6, 275.

[25] Yu Hirate and Hayato Yamana, *Generalized sequential pattern mining with item intervals.*, J. Comput. **1** (2006), no. 3, 51–60.

[26] Gengsen Huang, Wensheng Gan, and Philip S Yu, *TaSPM: Targeted sequential pattern mining*, ACM Transactions on Knowledge Discovery from Data **18** (2024), no. 5, 1–18.

[27] Homin Kang, Cheolhwan Lee, and Soon Ju Kang, *A smart device for non-invasive ADL estimation through multi-environmental sensor fusion*, Scientific Reports **13** (2023), no. 1, 17246.

[28] Zafar A Khan and Won Sohn, *Abnormal human activity recognition system based on r-transform and kernel discriminant technique for elderly home care*, IEEE Transactions on Consumer Electronics **57** (2012), no. 4, 1843–1850.

[29] Frédéric Koriche, *Relational networks of conditional preferences*, Machine learning **89** (2012), no. 3, 233–255.

[30] Frédéric Koriche and Bruno Zanuttini, *Learning conditional preference networks*, Artificial intelligence **174** (2010), no. 11, 685–703.

[31] Kathryn Laing, Peter Adam Thwaites, and John Paul Gosling, *Rank pruning for dominance queries in cp-nets*, Journal of Artificial Intelligence Research **64** (2019), 55–107.

[32] Andreas Lauenroth, Stephan Schulze, Anestis Ioannidis, Andreas Simm, and René Schwesig, *Effect of an age simulation suit on younger adults' gait performance compared to older adults' normal gait*, Research in Gerontological Nursing **10** (2017), 227–233.

[33] Shaun Wen Huey Lee and Pei-Lee Teh, *"suiting up" to enhance empathy toward aging: A randomized controlled study*, Frontiers in Public Health **8** (2020), 376.

[34] Minyi Li, Bao Quoc Vo, and Ryszard Kowalczyk, *Efficient heuristic approach to dominance testing in CP-nets*, Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), vol. 1, 2011, pp. 353–360.

[35] Shuang Liang, Lili Chen, Wensheng Gan, Philip S Yu, and Shengjie Zhao, *Targeted mining of time-interval related patterns*, IEEE Transactions on Neural Networks and Learning Systems (2025).

[36] Su Liu and Jinglei Liu, *CP-nets structure learning based on mrmcr principle*, IEEE Access **7** (2019), 121482–121492.

[37] Ye Liu, Liqiang Nie, Lei Han, Luming Zhang, and David S Rosenblum, *Action2Activity: Recognizing complex activities from sensor data*, Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015), 2015, pp. 1617–1623.

[38] Thomas Lukasiewicz and Enrico Malizia, *Complexity results for preference aggregation over (m) CP-nets: Max and rank voting*, Artificial Intelligence **303** (2022), 103636.

[39] Carl H Mooney and John F Roddick, *Sequential pattern mining–approaches and algorithms*, ACM Computing Surveys (CSUR) **45** (2013), no. 2, 1–39.

[40] Elaine EC Nelson, Andrew D Spence, and Gerard J Gormley, *Stepping into the shoes of older people: a scoping review of simulating ageing experiences for healthcare professional students*, Age and Ageing **52** (2023), no. 12, afad235.

[41] Marzieh Pashmdarfard and Akram Azad, *Assessment tools to evaluate activities of daily living (ADL) and instrumental activities of daily living (IADL) in older adults: A systematic review*, Medical journal of the Islamic Republic of Iran **34** (2020), 33.

[42] Parisa Rashidi and Diane J Cook, *Com: A method for mining and monitoring human activity patterns in home-based health monitoring systems*, ACM Transactions on Intelligent Systems and Technology (TIST) **4** (2013), no. 4, 1–20.

[43] Mirjam Sepesy Maučec and Gregor Donaj, *Discovering daily activity patterns from sensor data sequences and activity sequences*, Sensors **21** (2021), no. 20, 6920.

[44] Zahraa Khais Shahid, Saguna Saguna, and Christer Åhlund, *Detecting anomalies in daily activity routines of older persons in single resident smart homes: Proof-of-concept study*, JMIR aging **5** (2022), no. 2, e28260.

[45] Ramakrishnan Srikant and Rakesh Agrawal, *Mining sequential patterns: Generalizations and performance improvements*, International conference on extending database technology, Springer, 1996, pp. 1–17.

[46] Yiming Tian, Jie Zhang, Jie Wang, Yanli Geng, and Xitai Wang, *Robust human activity recognition using single accelerometer via wavelet energy spectrum features and ensemble feature selection*, Systems Science & Control Engineering **8** (2020), no. 1, 83–96.

[47] Ingo J. Timm, Heike Spaderna, Stephanie C. Rodermund, Christian Lohr, Ricardo Buettner, and Jan Ole Berndt, *Designing a randomized trial with an age simulation suit—representing people with health impairments*, Healthcare **9** (2021), no. 1.

[48] Md Zia Uddin and Ahmet Soylu, *Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning*, Scientific Reports **11** (2021), no. 1, 16455.

[49] Fabio Veronese, Andrea Masciadri, Sara Comai, Matteo Matteucci, and Fabio Salice, *Behavior drift detection based on anomalies identification in home living quantitative indicators*, Technologies **6** (2018), no. 1, 16.

[50] G Virone and AEIMABS Sixsmith, *Monitoring activity patterns and trends of older adults*, 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2008, pp. 2071–2074.

[51] Jianyong Wang, Jiawei Han, and Chun Li, *Frequent closed sequence mining without candidate maintenance*, IEEE Transactions on Knowledge and Data Engineering **19** (2007), no. 8, 1042–1056.

[52] Xifeng Yan, Jiawei Han, and Ramin Afshar, *Clospan: Mining: Closed sequential patterns in large datasets*, Proceedings of the 2003 SIAM international conference on data mining, SIAM, 2003, pp. 166–177.

[53] Mohammed J Zaki, *Spade: An efficient algorithm for mining frequent sequences*, Machine learning **42** (2001), no. 1, 31–60.

[54] Dorsaf Zekri, Ahmed Snoun, Thierry Delot, and Marie Thilliez, *A deep learning based system for a long-term elderly behavioral drift detection*, SN Computer Science **5** (2024), no. 7, 910.

[55] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang, *Convolutional neural networks for human activity recognition using mobile sensors*, 6th international conference on mobile computing, applications and services, IEEE, 2014, pp. 197–205.

[56] Meriem Zerkouk and Belkacem Chikhaoui, *Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models*, Sensors **20** (2020), no. 8, 2359.

[57] Yang Zhu, Zhaowei Liu, and Yuanqing Ma, *Structure learning of CP-nets based on constraint and scoring search*, 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2020, pp. 1103–1108.

Appendix A

CP-net Solver: Pseudocode and Algorithms

A.1 Algorithm Descriptions

A.1.1 Pairwise Outcome Comparison (`determineIsBetterFit`)

Algorithm 1: CP-net outcome comparison (`determineIsBetterFit`), Part 1: Initialisation and node traversal

Input: CP-net N with nodes V; two test subsequences S_0 and S_1 encoded as Boolean flags `presence0`(v) and `presence1`(v) for every $v \in V$

Output: `true` if S_0 is judged closer to Class 0 than S_1, `false` otherwise

```
1  foreach v ∈ V do
2      v.visited ← false
3  outcome0count ← 0
4  outcome1count ← 0
5  Initialise a queue Q with all nodes that have no parents
6  while Q is not empty do
7      remove the first node v from Q
8      if v.visited then
9          continue
10     v.visited ← true
11     foreach child c of v do
12         if all parents of c are marked as visited then
13             add c to Q
       // Skip pure condition nodes with no preference labels
14     if v.prefClass < 0 and v.condPrefClass < 0 then
15         continue
       // Skip nodes that cannot distinguish the outcomes
16     if presence0(v) = presence1(v) then
17         continue
       // Decide whether to use unconditional or conditional preference
18     if v has no parents then
19         label ← v.prefClass
20     else
21         if AncestorContextDiffers(v) then
               // Outcomes lie in different contexts at this node
22             continue
23         anyParentPresent ← false
24         foreach ancestor a of v do
25             if presence0(a) = true then
26                 anyParentPresent ← true
27         if anyParentPresent and v.condPrefClass ≥ 0 then
28             label ← v.condPrefClass
29         else
30             label ← v.prefClass
                                                  // (continued in Algorithm 2)
```

Algorithm 2: CP-net outcome comparison (`determineIsBetterFit`), Part 2: Voting and return

```
// For each node v processed in Algorithm 1, cast a local vote based
   on the active label
if label = 0 then
    if presence0(v) = true and presence1(v) = false then
        outcome0count ← outcome0count + 1
    if presence0(v) = false and presence1(v) = true then
        outcome1count ← outcome1count + 1
else if label = 1 then
    if presence0(v) = true and presence1(v) = false then
        outcome1count ← outcome1count + 1
    if presence0(v) = false and presence1(v) = true then
        outcome0count ← outcome0count + 1
return outcome0count > outcome1count
```

Algorithm 3: AncestorContextDiffers(v)

Input: Node v in CP-net N

Output: `true` if any ancestor of v has different presence in S_0 and S_1

```
foreach ancestor a of v do
    if presence0(a) ≠ presence1(a) then
        return true
return false
```

A.1.2 Cross-Dataset Win-Rate Computation

Algorithm 4: Cross-dataset comparison and win-rate computation, Part 1: Initialisation and pairwise comparison

Input: CP-net N; datasets $\mathcal{D} = \{D_1, \ldots, D_K\}$, each D_k a set of test subsequences
Output: Win rate `winRate`$[k]$ for each dataset D_k

for $k \leftarrow 1$ **to** K **do**
    `wins`$[k] \leftarrow 0$; `comparisons`$[k] \leftarrow 0$

`// Compare subsequences from different datasets in both directions`
for $i \leftarrow 1$ **to** K **do**
    for $j \leftarrow i+1$ **to** K **do**
        foreach $s \in D_i$, $t \in D_j$ **do**
            `// First direction: is` s `closer to Class 0 than` t`?`
            EncodePresence(N, s, t); $res_{ij} \leftarrow$ `determineIsBetterFit`(N)
            `// Second direction: is` t `closer to Class 0 than` s`?`
            EncodePresence(N, t, s); $res_{ji} \leftarrow$ `determineIsBetterFit`(N)
            `// Tally wins; equal results treated as a tie`
            if $res_{ij} =$ ***true*** ***and*** $res_{ji} =$ ***false*** **then**
                `wins`$[i] \leftarrow$ `wins`$[i] + 1$
            else if $res_{ij} =$ ***false*** ***and*** $res_{ji} =$ ***true*** **then**
                `wins`$[j] \leftarrow$ `wins`$[j] + 1$
            `comparisons`$[i] \leftarrow$ `comparisons`$[i] + 1$
            `comparisons`$[j] \leftarrow$ `comparisons`$[j] + 1$

`// (continued in Algorithm 5)`

Algorithm 5: Cross-dataset comparison and win-rate computation, Part 2: Win-rate calculation and ranking

`// Convert wins to win rates and rank (continued from Algorithm 4)`
for $k \leftarrow 1$ **to** K **do**
    if *comparisons*$[k] > 0$ **then**
        `winRate`$[k] \leftarrow$ `wins`$[k]$ / `comparisons`$[k]$
    else
        `winRate`$[k] \leftarrow 0$
Sort D_k in descending order of `winRate`$[k]$

Appendix B

Post Hoc Pairwise Tests: SCOUT Confidence Scores

B.1 Walking

Table B.1.1: Dunn–Bonferroni pairwise comparisons for walking SCOUT Class 0 confidence scores.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	p	p_{adj}
No gear – Low gear	−0.21	0.69	−0.31	0.756	1.000
No gear – Medium gear	1.57	0.69	2.28	0.023	0.091
No gear – High gear	1.50	0.69	2.17	0.030	0.119
Low gear – Medium gear	1.79	0.69	2.59	0.010	0.039
Low gear – High gear	1.71	0.69	2.48	0.013	0.052
Medium gear – High gear	−0.07	0.69	−0.10	0.918	1.000

In each row, the null hypothesis states that the two gear levels have the same distribution of SCOUT Class 0 confidence scores. The adjusted p-value (p_{adj}) is obtained by multiplying the unadjusted p-value by the number of pairwise tests (Bonferroni correction).

B.2 Vacuuming

Table B.2.1: Dunn–Bonferroni pairwise comparisons for vacuuming SCOUT Class 0 confidence scores.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	p	p_{adj}
No gear – Low gear	0.93	0.69	1.35	0.178	0.714
No gear – Medium gear	1.21	0.69	1.76	0.078	0.314
No gear – High gear	1.29	0.69	1.86	0.062	0.250
Low gear – Medium gear	0.29	0.69	0.41	0.679	1.000
Low gear – High gear	0.36	0.69	0.52	0.605	1.000
Medium gear – High gear	0.07	0.69	0.10	0.918	1.000

Null hypothesis and p_{adj} as in Table B.1.1.

B.3 Laundry

Table B.3.1: Dunn–Bonferroni pairwise comparisons for laundry SCOUT Class 0 confidence scores.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	p	p_{adj}
No gear – Low gear	0.86	0.69	1.24	0.214	0.857
No gear – Medium gear	1.71	0.69	2.48	0.013	0.052
No gear – High gear	2.86	0.69	4.14	<0.001	<0.001
Low gear – Medium gear	0.86	0.69	1.24	0.214	0.857
Low gear – High gear	2.00	0.69	2.90	0.004	0.015
Medium gear – High gear	1.14	0.69	1.66	0.098	0.391

Null hypothesis and p_{adj} as in Table B.1.1.

B.4 Sit-to-Stand (STS)

Table B.4.1: Dunn–Bonferroni pairwise comparisons for sit-to-stand (STS) SCOUT Class 0 confidence scores.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	p	p_{adj}
No gear – Low gear	0.21	0.69	0.31	0.756	1.000
No gear – Medium gear	1.71	0.69	2.48	0.013	0.052
No gear – High gear	1.50	0.69	2.17	0.030	0.119
Low gear – Medium gear	1.50	0.69	2.17	0.030	0.119
Low gear – High gear	1.29	0.69	1.86	0.062	0.250
Medium gear – High gear	−0.21	0.69	−0.31	0.756	1.000

Null hypothesis and p_{adj} as in Table B.1.1.

Appendix C

Post Hoc Pairwise Tests: CP-net Win Rates

C.1 Walking

Table C.1.1: Dunn–Bonferroni pairwise comparisons for CP-net win rates for walking.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	$\boldsymbol{p}$	$\boldsymbol{p}_{\text{adj}}$
No gear – Low gear	0.57	0.69	0.83	0.408	1.000
No gear – Medium gear	1.29	0.69	1.86	0.062	0.250
No gear – High gear	1.86	0.69	2.69	0.007	0.028
Low gear – Medium gear	0.71	0.69	1.04	0.301	1.000
Low gear – High gear	1.29	0.69	1.86	0.062	0.250
Medium gear – High gear	0.57	0.69	0.83	0.408	1.000

In each row, the null hypothesis states that the two gear levels have the same distribution of CP-net win rates. The adjusted p-value (p_{adj}) is obtained by multiplying the unadjusted p-value by the number of pairwise tests (Bonferroni correction).

C.2 Vacuuming

Table C.2.1: Dunn–Bonferroni pairwise comparisons for CP-net win rates for vacuuming.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	$\boldsymbol{p}$	$\boldsymbol{p}_{\text{adj}}$
No gear – Low gear	0.43	0.69	0.62	0.535	1.000
No gear – Medium gear	0.86	0.69	1.24	0.214	0.857
No gear – High gear	1.29	0.69	1.86	0.062	0.250
Low gear – Medium gear	0.43	0.69	0.62	0.535	1.000
Low gear – High gear	0.86	0.69	1.24	0.214	0.857
Medium gear – High gear	0.43	0.69	0.62	0.535	1.000

Null hypothesis and p_{adj} as in Table C.1.1.

C.3 Laundry

Table C.3.1: Dunn–Bonferroni pairwise comparisons for CP-net win rates for laundry.

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	$\boldsymbol{p}$	$\boldsymbol{p}_{\text{adj}}$
No gear – Low gear	0.29	0.69	0.41	0.679	1.000
No gear – Medium gear	0.29	0.69	0.41	0.679	1.000
No gear – High gear	2.00	0.69	2.90	0.004	0.015
Low gear – Medium gear	0.00	0.69	0.00	1.000	1.000
Low gear – High gear	1.71	0.69	2.48	0.013	0.052
Medium gear – High gear	1.71	0.69	2.48	0.013	0.052

Null hypothesis and p_{adj} as in Table C.1.1.

C.4 Sit-to-Stand (STS)

Table C.4.1: Dunn–Bonferroni pairwise comparisons for CP-net win rates for sit-to-stand (STS).

Comparison	**Test statistic**	**Standard error**	**Std. test statistic**	$\boldsymbol{p}$	$\boldsymbol{p}_{\text{adj}}$
No gear – Low gear	0.86	0.69	1.24	0.214	0.857
No gear – Medium gear	1.00	0.69	1.45	0.147	0.589
No gear – High gear	1.86	0.69	2.69	0.007	0.028
Low gear – Medium gear	0.14	0.69	0.21	0.836	1.000
Low gear – High gear	1.00	0.69	1.45	0.147	0.589
Medium gear – High gear	0.86	0.69	1.24	0.214	0.857

Null hypothesis and p_{adj} as in Table C.1.1.

Glossary

Activity of Daily Living (ADL)

A routine task that a person performs as part of everyday life, such as walking, preparing meals, doing laundry, or rising from a chair. The ability to perform ADLs independently is a key indicator of functional health, particularly in older adults.

ADL Fit A measure of how closely a person's execution of an Activity of Daily Living resembles a healthy baseline pattern of movement. Higher ADL fit indicates that the activity was performed in a manner consistent with unimpaired movement; lower ADL fit suggests deviation from the baseline, which may reflect functional decline.

Aging in Place

The ability of older adults to live safely and independently in their own homes for as long as possible, rather than moving to assisted living or care facilities. Supporting aging in place is the broader motivation for this thesis.

Baseline A reference standard representing healthy, unimpaired performance of an activity. In this thesis, the baseline corresponds to recordings made under the no-gear condition, in which participants performed activities without any physical restrictions.

Behavioral Drift

A gradual and often subtle change in the way a person performs everyday activities over time. In the context of this thesis, behavioral drift refers to a decline in the quality of ADL performance that may signal early-stage functional deterioration.

Class 0 / Class 1

Labels used to distinguish the two categories of training data. Class 0 refers to baseline (no-gear) recordings, representing healthy movement. Class 1 refers to high-gear recordings, representing movement under maximum physical restriction. The CP-net and SPM models are trained to distinguish between these two classes.

Conditional Preference Network (CP-net)

A structured model that encodes preferences over a set of features, where the preferred value of one feature may depend on the values of other related features. In this thesis, CP-nets are used to compare pairs of activity sequences and determine which one more closely resembles the learned baseline movement profile.

Conditional Preference Table (CPT)

A table associated with each node in a CP-net that stores the preference rules for that feature. For unconditional features, the CPT specifies whether the presence or absence of the feature is preferred. For conditional features, the CPT encodes rules that depend on the values of parent features.

Conditional Sequential Pattern Mining (CSPM)

An extension of Sequential Pattern Mining that identifies context-dependent relationships between patterns. Specifically, it discovers

cases where the presence of one pattern changes how another pattern is distributed across classes, revealing dependencies that standard SPM would not capture.

Depth Camera

A type of camera that captures both colour and depth information, allowing it to estimate the three-dimensional positions of objects and body joints in a scene. In this thesis, depth cameras are used to record skeletal data of participants performing ADLs without attaching any sensors to their bodies.

Discretisation

The process of converting continuous numerical data into a finite set of symbolic categories. In this thesis, raw 3D joint coordinates are discretised into symbolic movement tokens that represent typical short-term movement trends, making the data suitable for pattern mining.

Dominance Testing

A method for comparing two outcomes in a CP-net by determining whether one can be reached from the other through a sequence of locally improving changes. It provides a stricter form of preference comparison than ordering queries but is computationally expensive for large networks.

Gear Condition

One of four levels of physical restriction under which participants performed activities in this study. The conditions range from no gear (baseline, no restrictions) through low gear (ankle weights) and medium gear (ankle weights, knee braces, unstable shoes, and wrist weights) to high gear (all medium-gear components plus a weighted vest and

mobility-restricting gloves). Higher gear levels simulate increasing degrees of age-related physical decline.

Human Activity Recognition (HAR)

A field of research concerned with automatically identifying which physical activity a person is performing, using data from sensors or cameras. This thesis extends beyond traditional HAR by evaluating not just which activity is performed, but how well it is performed.

Instrumental Activity of Daily Living (IADL)

A more complex category of daily tasks that require coordination and planning, such as managing finances, cooking, or using transportation. IADLs build on basic ADLs and are important indicators of independent living capacity.

Ordering Query

A method for comparing two outcomes in a CP-net by aggregating local preference votes at each node. The outcome that receives more votes is judged to be closer to the preferred baseline. This thesis uses ordering queries as the primary mechanism for evaluating ADL fit.

Sequential Pattern Mining (SPM)

A data mining technique that identifies ordered sequences of events or symbols that occur frequently within a dataset. In this thesis, SPM is applied to symbolic movement sequences to extract recurring movement patterns that are characteristic of either baseline or restricted activity performance.

Simulated Aging Gear

A set of physical items worn by participants to approximate the bodily

restrictions associated with aging, including reduced strength, balance, and joint mobility. Items include ankle weights, knee braces, unstable shoes, wrist weights, a weighted vest, and mobility-restricting gloves.

Skeletal Data

A representation of the human body as a set of tracked joint positions in three-dimensional space. In this thesis, skeletal data are captured using depth cameras and consist of the coordinates of 32 body joints recorded over time.

Support In pattern mining, the proportion of sequences in a dataset in which a given pattern appears. A pattern with high support occurs frequently across the data; a pattern with low support is rare. Support is used to decide which patterns are frequent enough to be considered meaningful.

Symbolic Sequence

A representation of a movement recording as an ordered series of discrete symbols, where each symbol captures a typical short-term movement trend across multiple body joints at a given point in time. Symbolic sequences are produced by the discretisation step and serve as the input for pattern mining.

Win Rate The proportion of pairwise comparisons in which a given activity sequence is judged by the CP-net to be more similar to the baseline than the sequence it is compared against. A higher win rate indicates better ADL fit; a lower win rate suggests greater deviation from healthy baseline movement.

Vita

Candidate's full name: Marian Enam Esi Kushigbor

University attended:

- Master of Computer Science, University of New Brunswick, Fredericton, NB, Canada (2023–2026)
- Bachelor of Science in Computer Engineering, University of Ghana, Legon, Greater Accra, Ghana (2018–2023)