Vrije Universiteit Amsterdam **TNO**

Master Thesis

Markerless Pose-Based Classification of Duchenne Muscular Dystrophy from Single-Camera Video of Everyday Motor Tasks

---

Author: Georgios Xanthopoulos (2755088)

1st supervisor:	Mark Hoogendorn	
daily supervisor:	Sadegh Shahmohammadi	TNO
2nd reader:	Sandjai Bhulai	

A thesis submitted in fulfillment of the requirements for
the VU Master of Science degree in Business Analytics

May 10, 2026

Abstract

Duchenne muscular dystrophy (DMD) is a severe, progressive neuromuscular disease affecting approximately 1 in 3,500–5,000 boys, in which movement differences can appear years before diagnosis. Early detection from everyday video is appealing but difficult because existing approaches often require controlled environments, specialised sensors, or clinical expertise. This thesis investigates whether pose-derived movement features from standard single-camera RGB video can distinguish children with DMD from typically developing peers. Features were extracted using MediaPipe Pose Landmarker Heavy, a markerless pose-estimation model that tracks body joint positions without physical markers or specialist equipment. The approach was evaluated on walking and the rising-from-the-floor task (Gowers manoeuvre), using three interpretable classifiers and leave-one-out cross-validation on publicly sourced video. For walking, the best-performing model achieved a ROC-AUC of 0.825. The most predictive features were consistent with known DMD movement patterns, including altered hip movement, reduced stride length, and limited lower-limb mobility. Results for the Gowers task trended in the expected direction but were inconclusive, limited by the small number of available control videos. Overall, the results support markerless single-camera video as a feasible direction for interpretable DMD-related movement analysis, while remaining limited by public-video data quality and pose-estimation uncertainty.

Contents

List of Figures **v**

List of Tables **vii**

1 Introduction **1**

1.1 Problem Context and Motivation 1

1.2 Research Gap and Contribution 1

1.3 Research Objectives 2

1.4 Research Questions 2

1.5 Thesis Outline 2

2 Background **4**

2.1 Duchenne Muscular Dystrophy as a Movement-Analysis Problem 4

2.1.1 Why Early Movement Differences Matter 4

2.1.2 Current Assessment Practices and Their Limits 5

2.2 Markerless Pose Estimation for Child Movement Videos 6

2.2.1 What Single-Camera Pose Estimation Can Measure 6

2.2.2 Known Failure Modes in Child Movement Videos 6

2.2.3 Practical Measurement Scope for This Thesis 7

2.3 Movement Patterns Used for Feature Design 8

2.3.1 Walking Patterns 8

2.3.2 Rising-from-the-Floor Patterns 9

2.3.3 Mapping Movement Patterns to Candidate Features 10

3 Related Work **12**

3.1 Video-Based Machine Learning for Movement Assessment 12

3.1.1 Single-Camera Video as a Practical Data Source 12

3.1.2 Pose Estimation Methods in Related Video Studies 13

3.1.3 DMD Classification from Pose-Based Features . . . 13
3.1.4 Closest Baseline: Ali et al. (2024) . . . 14
3.1.5 Gaps in Child Movement and Non-Walking Tasks . . . 15
3.1.6 Explainable Models for Small Clinical Datasets . . . 16
3.1.7 Summary of Literature Gaps . . . 17

4 Methodology **18**
4.1 Dataset and Study Setup . . . 18
4.1.1 Data Source and Collection Protocol . . . 18
4.1.2 Inclusion and Exclusion Criteria . . . 19
4.1.3 Dataset Composition . . . 19
4.1.4 Video Characteristics . . . 20
4.1.5 Ethical Considerations and Privacy Guardrailing . . . 20
4.2 Task Definition and Clip Selection . . . 20
4.2.1 Motor Tasks Considered . . . 20
4.2.2 Task Boundary Definitions . . . 21
4.2.3 Dataset Composition by Task . . . 21
4.3 Pipeline Overview . . . 22
4.4 Pose Extraction and Coordinate Representation . . . 23
4.4.1 MediaPipe Pose Landmarker Heavy . . . 23
4.4.2 Coordinate Representation and Body-Frame Transformation . . . 23
4.5 Signal Preparation and Quality Control . . . 25
4.5.1 Landmark Filtering . . . 25
4.5.2 Temporal Smoothing . . . 26
4.5.3 Landmark Quality Control . . . 26
4.6 Movement Feature Engineering . . . 27
4.6.1 Notation for Feature Computation . . . 27
4.6.2 Joint Angle Features . . . 27
4.6.3 Segment Inclination Angles . . . 28
4.6.4 Continuous Relative Phase . . . 29
4.6.5 Bilateral Symmetry . . . 29
4.6.6 Gait Stride Segmentation . . . 29
4.6.7 Spatiotemporal Gait Features . . . 30
4.6.8 Heel Speed . . . 31
4.6.9 Frequency-Based Pose Features . . . 31

4.6.10 Toe-Walking Proxy . . . 33
4.6.11 GDI-Inspired Walking Deviation Proxy . . . 33
4.6.12 Gowers Features . . . 33
4.6.13 Feature Set Summary . . . 35
4.7 Classification Setup . . . 37
4.7.1 Classifier Selection . . . 37
4.7.2 Feature-Matrix Construction . . . 37
4.7.3 Hyperparameter Search Space . . . 38
4.8 Evaluation Design . . . 39
4.8.1 Cross-Validation Design . . . 39
4.8.2 Metrics and Baseline . . . 39
4.8.3 Univariate Analysis . . . 40
4.8.4 Attribution Analysis . . . 40

5 Results **42**
5.1 Overview of Results . . . 42
5.2 Landmark Quality Control . . . 42
5.3 Univariate Results . . . 43
5.3.1 Walking . . . 43
5.3.2 Gowers . . . 44
5.4 Classification Performance . . . 44
5.4.1 Walking . . . 46
5.4.1.1 Model Comparison . . . 46
5.4.1.2 Selected Model Performance . . . 47
5.4.1.3 Confusion Matrix . . . 47
5.4.1.4 ROC Curve, Feature Distributions, and Gait-Cycle Profiles 47
5.4.2 Gowers . . . 48
5.4.2.1 Model Comparison . . . 48
5.4.2.2 Selected Model Performance and Sample Size Limitations . 49
5.4.2.3 Confusion Matrix and ROC Curve . . . 50
5.5 Feature Attribution . . . 50
5.5.1 Walking . . . 50
5.5.2 Gowers . . . 52

6 Discussion **53**
6.1 Summary of Findings . . . 53
6.2 Walking Classification . . . 54
6.2.1 Performance in Context . . . 54
6.2.2 Interpreting the Movement Signal . . . 55
6.2.3 Reversed Direction Features . . . 55
6.2.4 Features That Did Not Discriminate . . . 56
6.3 Gowers Classification . . . 56
6.4 Contribution of the Two-Task Pipeline . . . 56
6.5 Limitations . . . 57
6.5.1 Dataset . . . 57
6.5.2 Recording Heterogeneity . . . 57
6.5.3 Pose Estimation . . . 58
6.5.4 Stride Segmentation . . . 59
6.5.5 Evaluation . . . 59
6.6 Directions for Future Work . . . 59

7 Conclusion **61**

A Threats to Validity and Supplementary Results **62**
A.1 Threats to Validity . . . 62
A.1.1 Internal Validity . . . 62
A.1.2 External Validity . . . 63
A.1.3 Construct Validity . . . 63
A.1.4 Conclusion Validity . . . 63
A.2 Detailed Feature Source References . . . 63
A.3 Univariate Feature Statistics (Low-Power Tasks) . . . 63
A.3.1 Gowers Univariate Statistics . . . 65
A.4 Low-Power Classification Visualisations . . . 65
A.4.1 Gowers Visualisations . . . 65

List of Figures

4.1 End-to-end pipeline architecture. Single-camera RGB video is manually segmented into task-specific clips, processed with MediaPipe pose extraction, filtered and transformed into the working coordinate representation, and converted into task-specific movement features. Candidate classifiers are trained and evaluated on the resulting feature matrices, with SHAP used after training for post-hoc feature attribution. 22

4.2 MediaPipe Pose Landmarker topology (33 full-body landmarks). Adapted from the Google AI Edge pose landmark documentation [46]. 24

4.3 ISB body-frame coordinate system with origin at hip centre. +X is forward, +Y is up, and +Z is sideways to subject-right (right-hand rule). 25

5.1 Raw vs smoothed traces for dmd_walking_07. Red shows original raw coordinates and blue shows prepared coordinates after filling and smoothing, plotted on the same frame axis for direct comparison. 43

5.2 Confusion matrix for walking classification under leave-one-video-out cross-validation. 47

5.3 Walking classification feature distributions associated with DMD-control separation. Top six features are ranked by statistical separation. Box plots show median, quartiles, and distribution spread. 48

5.4 ROC curve for walking classification. Dashed line indicates the random classifier baseline. 49

5.5 Walking gait-cycle profiles (0–100% cycle) comparing DMD and control groups for key stride-normalized signals. Curves show median trajectories and shaded bands show interquartile ranges across videos. 50

5.6 SHAP beeswarm plot for walking classification. Each point represents one video; horizontal position shows SHAP value (left = pushed toward control, right = pushed toward DMD); colour shows feature value (red = high, blue = low). Features ranked by mean absolute SHAP value (importance). . . . 51

A.1 Confusion matrix for Gowers classification under LOO-CV. Numbers indicate video-level predictions in this low-power feasibility setting. 65

A.2 ROC curve for Gowers classification in the low-power feasibility setting. Dashed line indicates random classifier baseline. 66

List of Tables

2.1 Mapping of DMD-related movement patterns to feature-design motivation. . 11

3.1 Comparison with Ali et al. (2024), the closest pose-based DMD walking-classification baseline. 15

4.1 Framerate (fps) and clip duration (seconds) statistics by diagnostic group and task. 20

4.2 Dataset composition by task and diagnostic group following manual segmentation and quality screening. Differences from initial collection reflect stricter framerate and clip-quality filtering applied uniformly across groups to reduce avoidable processing bias. 21

4.3 Comparison of MediaPipe world-landmark and body-centered coordinate systems. 24

4.4 Summary of pose-derived movement feature groups and source types. "Literature-based" features derive from published movement-analysis references. "Pose-derived approximation" features adapt clinical or gait concepts to single-camera landmark data. "Pose quality" refers only to landmark-visibility checks. 36

4.5 Hyperparameter search ranges for Random Forest and XGBoost. Settings marked RF only or XGB only are specific to that model; all others apply to both. 38

5.1 Top walking features associated with DMD-control separation. Separation is the signed rank-biserial correlation from a two-sided Mann–Whitney U test on the current walking feature set after metadata and coordinate columns are excluded by the pipeline configuration. 44

LIST OF TABLES

5.2 Top Gowers features associated with DMD-control separation. Separation is the signed rank-biserial correlation from a two-sided Mann–Whitney U test on the current Gowers clip feature table. 45
5.3 Model comparison by task. 45
5.4 Operating-point metrics for the selected model in each task. Metrics were computed from pooled leave-one-video-out predictions after applying the Youden-selected threshold within each training fold. 46

A.1 Detailed source references for the feature groups summarised in Table 4.4. "Literature-based" features derive from published movement-analysis references. "Pose-derived approximation" features adapt clinical or gait concepts to single-camera landmark data. 64
A.2 Top 15 Gowers features by absolute rank-biserial effect size at clip level. Limited sample size; reported as descriptive characterisation only. 67

1

Introduction

1.1 Problem Context and Motivation

Duchenne muscular dystrophy (DMD) is an inherited childhood muscle disease that causes progressive weakness and delayed movement development. Early detection is clinically important because developmental and motor delays can appear within the first months of life [1], [2], yet diagnosis commonly occurs around age 4.5–5 years [1]. This delay reduces the time available for coordinated care and treatments that can slow disease progression [1].

Current clinical assessment relies on infrequent, specialist-administered instruments that require clinic visits and trained raters. Recent advances in markerless pose estimation suggest a lower-burden alternative: movement analysis from ordinary smartphone video under controlled capture conditions, without specialist laboratory equipment [3]–[5].

1.2 Research Gap and Contribution

Recent work suggests that markerless pose estimation can recover useful movement information from ordinary video, but the DMD literature remains limited in task coverage, patient-level evaluation, and interpretability. Prior studies often focus on walking pattern (gait) alone, use validation designs where repeated subject information may inflate performance, or report model outputs without clearly linking predictions to movement patterns described in the DMD literature.

This thesis addresses those gaps with a single-camera pipeline that extracts interpretable pose-derived movement features from walking and the rising-from-the-floor task (Gowers manoeuvre), the characteristic sequence in which children with DMD use their arms to

push themselves upright. The pipeline evaluates classification under leave-one-out cross-validation and uses SHapley Additive exPlanations (SHAP) to show which features contribute to each prediction.

1.3 Research Objectives

The objectives of this thesis are:

1. Develop a single-camera markerless pose estimation pipeline extracting whole-body movement from walking and the Gowers manoeuvre.
2. Engineer task-specific pose-derived movement features grounded in known DMD movement patterns, distinguishing established movement measures from approximate video-based proxies.
3. Evaluate classification performance under leave-one-out cross-validation, reporting the receiver operating characteristic area under the curve (ROC-AUC) and the precision-recall area under the curve (PR-AUC), with SHAP-based feature attribution.

1.4 Research Questions

This thesis is guided by the following research questions:

1. To what extent can single-camera RGB video separate DMD from typically developing (TD) children in walking and the Gowers manoeuvre?
2. Which pose-derived movement features contribute most to that separation?
3. How well do those features align with movement patterns described in the DMD literature?

1.5 Thesis Outline

The remainder of this thesis is structured as follows. Chapter 2 provides the disease, movement, and pose-estimation background needed to understand the thesis. Chapter 3 reviews related video-based machine-learning work and identifies the literature gaps. Chapter 4 describes the dataset, pose extraction, feature engineering, classification setup, and evaluation design. Chapter 5 presents classification results for each task alongside SHAP feature importance analysis. Chapter 6 discusses the implications of the findings, limitations of

the pipeline and dataset, and directions for future work. Chapter 7 summarises the main findings and provides answers to the research questions.

2

Background

This chapter provides the disease, movement, and pose-estimation background needed to understand the thesis design and results. It covers: (1) DMD as a movement-analysis problem and current assessment practices, (2) what single-camera pose estimation can and cannot measure in child movement videos, and (3) the movement patterns used to motivate feature design. Prior machine-learning methods and related video-based studies are reviewed in Chapter 3.

2.1 Duchenne Muscular Dystrophy as a Movement-Analysis Problem

2.1.1 Why Early Movement Differences Matter

Duchenne muscular dystrophy (DMD) is a fatal, progressive childhood muscle disease caused by mutations in the dystrophin gene on the X chromosome. Because of its X-linked inheritance, it occurs almost exclusively in boys, with an estimated prevalence of approximately 1 in 3,500–5,000 male births [6]. Dystrophin helps stabilise muscle fibres during contraction. Without it, muscle cells are more easily damaged during movement; over time, damaged tissue is replaced by fat and scar tissue, reducing strength and movement ability. The complications of DMD eventually affect cardiac and respiratory function and reduce survival [1], [7].

Movement differences can appear long before many children receive a confirmed diagnosis. Developmental and motor delays may be visible within the first months of life, yet diagnosis commonly occurs around 4–5 years of age [1], [2], [8]. This delay matters because earlier recognition can increase the time available for genetic counselling, coordinated

multidisciplinary care, cardiac and respiratory monitoring, family support, treatment planning, and access to therapies that may slow disease progression [1]. Earlier detection can therefore improve care planning and reduce family burden, while not removing the progressive nature of the disease [1]. In practice, early signs may be mistaken for normal developmental variation or unrelated medical problems, such as abnormal blood enzyme levels [1], [2].

For this thesis, the important point is that DMD produces observable movement differences during everyday tasks before and during the early walking years. These differences motivate a video-based movement-analysis approach: if DMD-related movement patterns can be captured from ordinary video, they may provide useful feasibility evidence for lower-burden assessment outside specialist motion-analysis settings.

2.1.2 Current Assessment Practices and Their Limits

Current assessment practices begin with developmental surveillance in primary care. Milestone checks track age-specific movement, language, and behavioural development, but early DMD signs can be difficult to distinguish from normal developmental variation. In the Netherlands, routine youth health care uses developmental monitoring tools, but these are not DMD-specific and therefore depend on early clinical suspicion before neuromuscular testing is initiated [1], [2]. General developmental tools therefore have limited positive predictive value for DMD: many children flagged for delay will not have DMD, while some boys with early DMD signs may not be referred immediately for neuromuscular testing [2].

When DMD is suspected, serum creatine kinase (CK) testing is the standard first-line screening step. CK is an enzyme released when muscle is damaged; in DMD, CK levels are typically markedly elevated [1]. However, CK testing is only performed once suspicion has been raised. This makes early recognition of movement signs a bottleneck in the diagnostic pathway. Genetic testing is then required to confirm the diagnosis and identify the underlying dystrophin mutation [1].

After diagnosis, movement progression is monitored using structured functional assessments. These include the six-minute walk test (6MWT), the North Star Ambulatory Assessment (NSAA), and timed rising from the floor, which corresponds to the Gowers manoeuvre [1]. These assessments are clinically established and directly related to the movement tasks considered in this thesis: walking and the Gowers manoeuvre.

Despite their clinical value, these assessments have practical limitations. They require scheduled visits, trained personnel, and standardised testing conditions, and they are usually performed only at discrete time points. They also rely partly on human scoring or

task-level summaries, which can miss gradual changes between clinic visits. These limitations motivate complementary video-based movement analysis: not as a replacement for clinical diagnosis, but as a lower-burden feasibility direction for capturing DMD-related movement patterns from everyday video.

2.2 Markerless Pose Estimation for Child Movement Videos

Having outlined the disease background and current assessment practices, this section introduces the technical background for markerless pose estimation in child movement videos. The aim is to clarify what single-camera pose estimation can reasonably support, where it becomes unreliable, and how these limits shape the feature design used in this thesis.

2.2.1 What Single-Camera Pose Estimation Can Measure

Markerless pose estimation automatically detects body landmarks from ordinary video frames. In a single-camera setting, this allows visible body movement to be tracked over time without reflective markers, calibrated multi-camera systems, or specialist laboratory equipment. From these landmark trajectories, it is possible to extract movement features such as step timing, approximate limb angles, posture, and movement rhythm.

However, single-camera pose estimation is not equivalent to clinical 3D gait analysis. Although markerless systems can approach marker-based accuracy in controlled adult locomotion studies [9], these validation results do not directly transfer to child movement videos or uncontrolled public footage. Single-camera RGB video must infer body position from one viewpoint, making it more sensitive to camera angle, occlusion, image quality, clothing, and body-size differences.

For this reason, this thesis focuses on large, visible movement patterns rather than fine-grained clinical joint measurements. Pose-derived features are treated as interpretable movement proxies for classification, especially when they describe walking timing, trunk orientation, stride-related changes, and Gowers behaviour.

2.2.2 Known Failure Modes in Child Movement Videos

When applied to children, markerless pose estimation becomes less reliable because child body proportions differ from the adult-dominated datasets on which many models are trained. Children are physically smaller, so their bodies and limbs occupy fewer pixels in the frame; as a result, small keypoint errors can have a larger effect on derived spatial

features [10]. Accuracy also decreases for atypical walking patterns, because children with movement disorders may move outside the range of poses represented in standard validation datasets [11].

For example, in children with cerebral palsy, markerless motion capture shows large errors for pelvic tilt and hip rotation when compared with marker-based systems [11]. Reported errors reach 21.5° for pelvic tilt and 45.3° for hip rotation, with additional offsets in trunk, pelvis, and hip angles [11]. These findings reinforce the restricted measurement scope defined above, especially for pelvic, rotational, and side-to-side movement variables.

Despite these limitations, prior work confirms that single-camera pose coordinates can yield useful predictive information about gait and function [12].

Not all pose-derived measurements are equally robust. Accuracy is generally stronger for gait timing and front-to-back movement than for side-to-side or rotational variables. In toddlers, timing and step-length proxies correlate strongly with instrumented references ($r > 0.79$) [10], and front-to-back hip and knee angles can reach good-to-excellent agreement in child movement-disorder gait (ICC > 0.8) [13], [14]. Side-to-side, rotational, and ankle or foot measurements are substantially less stable and more prone to occlusion [13]–[16].

Environmental conditions are another major source of error. Prior work motivates subject-centering, background suppression or segmentation, confidence thresholding, gap-filling for short dropouts, temporal smoothing, and spatial normalisation to reduce sensitivity to camera distance and recording conditions [17], [18].

2.2.3 Practical Measurement Scope for This Thesis

The practical measurement scope follows from the limits above. Single-camera pose outputs are better suited to large, visible movement patterns, such as walking timing, trunk orientation, stride-related changes, and Gowers behaviour, than to fine rotational or depth-dependent joint measurements.

For this reason, the literature supports a conservative interpretation of pose-derived variables in child movement videos. Features based on precise pelvic rotation, fine ankle or foot tracking, or monocular depth should be treated cautiously because they are more sensitive to viewpoint, occlusion, and pose-estimation error.

This scope informs the feature-engineering strategy used in Chapter 4: pose outputs are used as interpretable movement proxies for video-level classification, not as clinical-grade 3D biomechanical measurements.

2.3 Movement Patterns Used for Feature Design

This section summarises the DMD-related movement patterns that motivate the feature families used later in the thesis. The aim is not to define the computational features themselves, but to explain why walking and the Gowers manoeuvre are relevant movement tasks for DMD-related video analysis.

Two movement domains are central for ambulatory children with DMD: altered walking patterns and Gowers adaptations. Both are repeatedly described in the DMD literature and reflect weakness around the hips, upper legs, and lower limbs. These domains provide the movement-level rationale for the pose-derived proxies and video-level descriptors introduced in Chapter 4.

Table 2.1 maps the main movement patterns to supporting evidence and explains how each pattern motivates later feature design.

2.3.1 Walking Patterns

Walking differences are among the most studied movement changes in ambulatory children with DMD. Instrumented 3D gait analysis has shown that DMD-related walking changes can appear before standard clinical assessments detect clear functional decline [19]. Early descriptions report a recurring pattern of forward pelvic tilt, reduced foot lifting during swing, altered step rate, and shorter steps. In Sutherland's three-stage model, gait deterioration could be predicted using cadence, reduced dorsiflexion during swing, and anterior pelvic tilt [19].

Later studies refined these observations in younger and corticosteroid-treated cohorts. Doglio et al. [20] reported reduced upward foot motion and altered ankle-related function in young boys with DMD. Gaudreault et al. [21] showed that some DMD-related walking differences remain after controlling for walking speed, including shortened step length and increased side-to-side loading. These findings suggest that DMD walking differences are not only a consequence of slower walking, but also reflect altered movement strategies.

Several studies also describe compensatory patterns around the hips, pelvis, trunk, and feet. Reduced force generation around the hip can lead to shorter steps, wider base of support, increased trunk movement, and greater reliance on compensatory posture [22], [23]. Longitudinal data further show that forward pelvic tilt and step width increase progressively in ambulatory boys over time [24], reinforcing pelvic and base-of-support features as meaningful targets. Toe-walking is also frequently reported in ambulatory DMD and is linked to calf muscle contracture and reduced foot-lifting capacity [25]. Together,

these findings motivate walking features related to step timing, stride length, base of support, trunk or hip orientation, lower-limb mobility, and toe-walking behaviour.

Composite gait summaries such as the Gait Deviation Index (GDI), which quantifies how far a walking pattern deviates from a typically developing reference pattern, further support the idea that DMD walking patterns can be represented as deviation from typical gait [26]. In this thesis, this idea is used only as motivation for a GDI-inspired walking deviation proxy; the implemented proxy is not the clinical GDI itself.

2.3.2 Rising-from-the-Floor Patterns

The rising-from-the-floor task corresponds clinically to the Gowers manoeuvre, a characteristic movement pattern in which children with DMD may use their hands and arms to help push themselves upright. The movement is shaped by weakness around the hips and upper legs: when the hip and thigh muscles cannot generate sufficient upward force, the child may shift through a four-point position, use the hands on the thighs, widen the base of support, and increase lower-back arching to bring the centre of mass over the feet [27].

Chang et al. [27] describe a spectrum of Gowers manoeuvre severity, ranging from relatively mild single-hand support to slower multi-phase rises involving crawling, repeated hand pushes, and thigh climbing. This supports treating the Gowers manoeuvre as a sequence of observable movement phases rather than a single posture. For video-based analysis, the most relevant signals are therefore task duration, transition timing, trunk posture, hip rise, and hand-to-leg proximity.

Timed Rise from Floor (TRF) is also used in DMD assessment and has been linked to later functional change. Mendell et al. [28] showed that prednisone treatment reduced rise time compared with placebo, supporting rise time as a treatment-sensitive functional measure. Mazzone et al. [29] reported that TRF longer than 7 seconds predicted greater decline in 6-minute walk distance over 12 months, while McDonald et al. [30] linked inability to rise from the floor with increased risk of losing ambulation. These findings motivate Gowers features related to rise duration, hip vertical motion, trunk movement, and hand-support behaviour.

In this thesis, these quantities are treated as pose-derived approximations of the Gowers pattern, not as clinical Gowers grades. This distinction is important because public single-camera video cannot measure physical support forces, exact muscle strength, or clinician-rated movement quality.

2.3.3 Mapping Movement Patterns to Candidate Features

The walking and Gowers patterns reviewed above provide the movement-level rationale for the feature families used later in the thesis. Walking-related evidence motivates features describing step timing, stride behaviour, base of support, trunk and hip orientation, lower-limb mobility, toe-walking behaviour, and deviation from a typically developing reference pattern. Gowers evidence motivates features describing rise duration, hip vertical motion, trunk movement, hand-support proximity, and task-phase transitions.

Table 2.1 summarises this mapping from DMD-related movement patterns to feature-design motivation. The table does not define the computational features themselves; those definitions are provided in Chapter 4. Instead, it clarifies why each feature family is relevant for DMD-related video movement analysis and whether the underlying movement pattern is directly observable or requires a pose-derived approximation.

Table 2.1: Mapping of DMD-related movement patterns to feature-design motivation.

Movement pattern	Movement interpretation	Supporting evidence	Feature-design motivation
Side-to-side walking movement	Side-to-side pelvic or trunk motion associated with hip and upper-leg weakness	[21], [24]	Motivates base-of-support, pelvis/trunk orientation, and side-to-side movement proxies.
Toe-walking behaviour	Reduced foot lifting and tendency toward toe-walking during walking	[20], [25]	Motivates empirical toe-walking and lower-limb mobility proxies.
Shorter, altered stepping pattern	Shorter steps and altered step timing associated with compensation for lower-limb weakness	[20], [21]	Motivates stride length, cadence, and step-timing features.
Deviation from typical walking	Overall walking pattern differs from typically developing reference gait	[26]	Motivates the GDI-inspired walking deviation proxy.
Rising-from-the-floor timing	Slow or multi-phase floor rise associated with weakness around the hips and upper legs	[27], [29]	Motivates rise duration, hip vertical motion, and task-phase transition features.
Hand-support behaviour	Use of hands and arms to assist rising from the floor	[27], [30]	Motivates hand-to-leg proximity and hand-support proxies.
Lower-back arching and trunk adaptation	Postural adaptation used to bring the centre of mass over the feet during rise or walking	[22], [23], [27]	Motivates trunk orientation and posture features.

3

Related Work

This chapter reviews prior work on video-based and pose-derived movement assessment relevant to DMD classification. It first places single-camera video within the broader landscape of movement-analysis technologies, then discusses pose-estimation methods used in related studies. The chapter then focuses on the closest DMD classification baseline, gaps in paediatric and non-walking task validation, and the role of interpretable models for small pose-derived datasets.

3.1 Video-Based Machine Learning for Movement Assessment

3.1.1 Single-Camera Video as a Practical Data Source

Movement analysis in neuromuscular disease can use inertial measurement units, multi-camera 3D motion capture, force plates, depth sensors, ultrasound-based inference, or video. These alternatives can provide richer biomechanical signals than monocular RGB video, but they also introduce equipment, calibration, cost, or environmental constraints. For lower-burden assessment outside specialist laboratories, single-camera RGB video is attractive because it can be captured with ordinary consumer devices and processed using markerless pose estimation [16], [31], [32].

The trade-off is reduced measurement precision. Single-camera markerless pose estimation cannot reproduce the full information available from calibrated 3D motion-capture systems, especially for depth-dependent, rotational, or occluded movements [13], [16]. Its value is therefore strongest when the goal is not clinical-grade kinematic measurement, but extraction of movement signals that may support classification, monitoring, or feasibility analysis.

This trade-off is directly relevant to DMD. A deployable video-based approach could make movement assessment less dependent on specialist equipment, but it must be interpreted cautiously because public or caregiver-recorded video introduces uncontrolled camera viewpoint, lighting, occlusion, and task-execution variability. Related work therefore supports single-camera video as a practical data source, while also motivating conservative feature interpretation and careful validation design.

3.1.2 Pose Estimation Methods in Related Video Studies

OpenPose and MediaPipe Pose are the most directly relevant 2D pose estimators for this thesis because they appear in nearby markerless-gait and clinical-video studies. OpenPose has been widely used in video-based gait analysis and has substantial validation evidence in controlled settings [16], [32]. MediaPipe Pose is lightweight and mobile-oriented, providing dense full-body landmarks suitable for scalable single-camera workflows [33], [34].

DMD-specific work also considers these pose-estimation backbones. Ali et al. [35] classified DMD gait from markerless video by comparing classical machine-learning models using clinical gait features with deep sequence models trained on pose time series. Their study found that pose-derived features can support DMD gait classification, while also showing that performance depends on the pose representation, extracted features, model type, and validation setup.

This literature does not establish a single best pose estimator for paediatric DMD movement analysis. Rather, it shows that OpenPose, MediaPipe, and monocular 3D pose approaches are relevant options within deployable video-based pipelines. None of these methods removes the core child-movement validity problem: paediatric body proportions, atypical walking patterns, occlusion, camera viewpoint, and non-walking movements can all reduce landmark reliability. For this reason, pose-estimator choice should be interpreted as one component of a broader pipeline, not as the sole determinant of validity.

3.1.3 DMD Classification from Pose-Based Features

The DMD-specific markerless-pose classification literature remains limited. The closest prior work is Ali et al. [35], who classified DMD gait from markerless video using pose-derived clinical gait features and compared classical and deep-learning approaches. Their study provides important evidence that markerless pose features can distinguish DMD from typically developing gait in publicly available video.

However, several limitations remain relevant for the present thesis. First, the task scope is mostly walking-focused, while DMD assessment also relies on other functional movements such as the Gowers manoeuvre. Second, public-video datasets introduce uncertainty in recording conditions, subject identity, and clinical verification. Third, validation protocols must be interpreted carefully when patient-level separation is not explicit, because repeated clips from the same subject or recording context can inflate performance estimates. Finally, strong sequence-model performance does not automatically provide feature-level explanations of which movement patterns drive the prediction.

Related work therefore supports the feasibility of pose-based DMD gait classification, but leaves open whether an interpretable, feature-attributed pipeline can be extended beyond walking to include the Gowers task.

3.1.4 Closest Baseline: Ali et al. (2024)

Ali et al. [35] is the closest baseline for pose-based DMD classification. The study is important because it directly evaluates DMD gait classification from markerless video, using both engineered clinical gait features and deep pose-sequence models. It therefore provides the most relevant point of comparison for the walking component of this thesis.

The present study differs in three main ways. First, it focuses on interpretable pose-derived feature engineering and SHAP-based attribution rather than prioritising end-to-end sequence modelling. Second, it evaluates leave-one-video-out predictions after retaining one clip per identified subject, reducing the risk of repeated-subject leakage within the available public-video setting. Third, it extends the task scope by considering both walking and the Gowers task.

Direct numerical comparison with Ali et al. [35] should remain cautious because the studies use different datasets, class balances, pose representations, metrics, and validation protocols. The comparison is therefore best understood as a contextual comparison rather than a leaderboard-style performance comparison. Table 3.1 summarises the main methodological differences.

Table 3.1: Comparison with Ali et al. (2024), the closest pose-based DMD walking-classification baseline.

Dimension	Ali et al. [35]	Present study
Task scope	Walking/gait classification	Walking plus Gowers task
Pose representation	2D and 3D pose-derived representations	MediaPipe pose-derived movement features
Model type	Classical ML on clinical gait features and deep sequence models	LR/RF/XGBoost comparison on engineered movement features
Primary metric	Accuracy and AUC where reported	ROC-AUC and PR-AUC, with PR-AUC included because of class imbalance
Validation design	Cross-validation and held-out evaluation; patient-level separation not explicit	Leave-one-video-out after retaining one clip per identified subject
Sample size	12 DMD clips for classification; matched TD subset from public benchmark	$n = 73$ walking clips, including 10 DMD and 63 control clips
Class balance	Not the main reporting focus	Class imbalance handled with balanced weighting and PR-AUC reporting
Interpretability	Clinical gait features for the classical ML path; limited attribution for the deep sequence model	SHAP-based attribution on engineered movement features
Main limitation for comparison	Different data, metrics, class balance, and validation setup	Feasibility-level public-video study with small DMD sample

3.1.5 Gaps in Child Movement and Non-Walking Tasks

Paediatric markerless-pose validation is stronger for walking than for non-walking functional tasks. Studies in child gait suggest that single-camera or markerless systems can capture some timing and sagittal-plane movement signals, but reliability decreases for ro-

tational, side-to-side, distal, or occluded joints [10], [13], [14], [16]. These limitations are especially relevant for children with movement disorders, whose movement patterns may differ from those represented in standard pose-estimation benchmarks [11], [36], [37].

The Gowers task introduces additional challenges. It includes floor-level postures, body-part overlap, hand support, changes in viewing angle, and transitions between movement phases. These properties make precise joint-angle estimation less reliable than in upright walking. For that reason, related work supports prioritising macro-spatial and temporal descriptors, such as time-to-stand, hand-support proximity, trunk posture, and phase transitions, rather than fine-grained non-walking joint kinematics [27], [31], [38].

This creates a clear gap for DMD-related video analysis. Prior pose-based DMD classification work mainly addresses walking, while the Gowers task remains underexplored in markerless video pipelines. A cautious two-task approach can therefore add value by testing whether interpretable pose-derived features capture useful movement signals beyond gait alone.

3.1.6 Explainable Models for Small Clinical Datasets

Small biomedical and movement-analysis datasets often favour models that can work with limited samples and support interpretation. For tabular datasets, tree-based ensembles often remain strong baselines, particularly when sample size is limited and the input consists of engineered features [39]. Neuromuscular video studies provide related precedent for using pose-derived features with interpretable or feature-attributed models [40], [41].

SHAP provides a model-interpretation framework for assigning feature-level contributions to predictions [42]. For tree ensembles, TreeSHAP offers an efficient way to explain how engineered features contribute to model outputs [43]. SHAP can also be applied to deep-learning models, but explanations for sequence models usually require additional choices about the input representation, background distribution, and temporal aggregation. In small public-video datasets, this can make feature-level interpretation less direct.

This does not mean that deep models are unsuitable for DMD video analysis. Rather, the literature motivates feature-attributed tabular models as a practical and interpretable baseline when the dataset is small, labels are uncertain, and the research question concerns which movement features contribute to DMD-control separation. This aligns with the emphasis of the present thesis on feasibility, cautious interpretation, and movement-pattern grounding.

3.1.7 Summary of Literature Gaps

The literature supports the feasibility of markerless video-based movement assessment, but several gaps remain for DMD classification. First, DMD-specific pose-classification evidence is sparse and mainly focused on walking [35]. Second, paediatric and non-walking markerless-pose validation remains limited, especially for tasks involving occlusion, floor-level postures, and body-part overlap [10], [11], [13], [14], [36], [37]. Third, publicly sourced or caregiver-recorded video introduces uncertainty in labels, recording conditions, and patient-level separation [31], [38]. Fourth, strong classification performance is often not accompanied by movement-level feature attribution, despite the relevance of explainability in small clinical or biomedical datasets [39]–[41].

These gaps motivate the present thesis: a cautious, interpretable, two-task feasibility pipeline for DMD-related movement classification from single-camera video. The next chapter describes the dataset, pose-extraction pipeline, feature engineering, classifier setup, and evaluation design used to examine this gap.

4

Methodology

This chapter describes the full methodology. Section 4.1 describes the dataset construction and video collection protocol. Section 4.2 defines the motor tasks and clip boundaries. Section 4.3 gives an overview of the analysis pipeline. Sections 4.4–4.5 cover pose extraction, coordinate representation, and signal quality control. Section 4.6 describes the movement feature engineering. Section 4.7 defines the classification setup. Section 4.8 describes the evaluation design.

4.1 Dataset and Study Setup

4.1.1 Data Source and Collection Protocol

Videos were collected from YouTube through manual search. DMD videos were identified using search terms such as "Duchenne muscular dystrophy walking", "Duchenne gait child", and "child rising from floor Duchenne". Control videos were identified through separate searches targeting typical child movement and physical activity, such as "child walking" and "child getting up from floor". Control clips were therefore treated as publicly sourced examples of typical child movement rather than as a clinically matched control cohort.

No data were collected from clinical settings, and no personally identifiable information beyond what was publicly visible in the source videos was accessed or stored. All videos were screened for inclusion eligibility prior to processing. Each retained clip was assigned a unique identifier. Source URLs, platform metadata, and clip start–end timestamps were recorded in a canonical metadata index available in the study repository to support reproducibility of the dataset construction pipeline.

4.1.2 Inclusion and Exclusion Criteria

Videos were included if the following criteria were met:

1. Frame rate of at least 20 fps. Videos below this threshold lack sufficient temporal resolution for fast Fourier transform (FFT)-based step rate estimation and velocity feature extraction.

2. Video resolution of at least 360p with no sustained motion blur that would render landmark localisation unreliable. Video quality must be sufficient for pose estimation from single-camera RGB frames.

3. The subject appeared to be between 4 and 8 years of age, as judged from video metadata, channel descriptions, accompanying captions, or visible developmental characteristics. Exact ages were not available for the majority of subjects; the age range of 4–8 years represents the target population defined a priori.

4. For DMD videos, a diagnosis of Duchenne muscular dystrophy was indicated by the video title, channel description, or accompanying caption, with preference given to sources explicitly referencing clinical diagnosis or genetic confirmation. Videos where the diagnosis was ambiguous or unspecified were excluded.

5. The subject was walking independently and performed at least one of the two target motor tasks (walking or the Gowers task) without continuous physical assistance.

6. Sufficient body visibility was present for landmark extraction: the full lower body and at least the lower trunk were visible throughout the task duration.

Videos were excluded if the subject was non-walking, if the recording angle rendered the task uninterpretable, or if the clip duration was insufficient to capture a complete task sequence.

4.1.3 Dataset Composition

The final dataset comprised 94 video clips (one per identified subject). DMD subjects were all male, consistent with X-linked inheritance; control subjects were mixed sex. The control group represents a broad typically developing reference rather than a sex-matched clinical cohort. Per-task counts and class imbalance are detailed in Section 4.2.

4.1.4 Video Characteristics

Source videos exhibited variable framerates across both groups and tasks (Table 4.1).

Table 4.1: Framerate (fps) and clip duration (seconds) statistics by diagnostic group and task.

		Framerate (fps)			Duration (s)		
Group	**Task**	Mean	Median	SD	Mean	Median	SD
DMD	Rise	34.4	30.0	11.8	12.5	9.6	11.3
	Walking	28.1	29.6	2.3	9.1	8.1	6.9
Control	Rise	26.9	25.0	3.0	6.6	5.0	4.8
	Walking	27.4	27.5	2.8	14.6	12.1	8.9

4.1.5 Ethical Considerations and Privacy Guardrailing

This study used exclusively publicly available video data and involved no patient contact, no clinical data access, and no collection of personally identifiable information. Under the conditions of this thesis project, formal institutional ethics review was not required.

Notwithstanding the public availability of the source material, privacy guardrailing was applied given that the subjects are minors. Prior to pose extraction, facial regions were blurred in all video clips using an automated face detection and blurring pipeline [44]. Head and face landmarks produced by MediaPipe were excluded from all subsequent feature computation; only landmarks from the shoulders distally and the trunk were retained. Raw video files were not stored beyond the processing stage. No attempt was made to re-identify any subject.

4.2 Task Definition and Clip Selection

4.2.1 Motor Tasks Considered

Two motor tasks were evaluated: overground walking and the Gowers manoeuvre. Each included video was manually segmented from the source material before pose extraction, yielding one task-specific clip for each retained subject. Manual segmentation was used because the YouTube source videos varied in recording angle, scene context, clip structure,

and task timing, making reliable automated boundary detection impractical. Manual clipping is common in small movement-video datasets where ecological validity is prioritised over recording standardisation [31].

4.2.2 Task Boundary Definitions

Task boundaries were defined prospectively and applied consistently across all videos before feature extraction or model training.

Overground walking. The walking clip was defined from the first observable foot contact in the direction of travel to the last foot contact before the subject decelerated, stopped, or exited the frame. Clips containing fewer than two complete stride cycles were excluded. Overground walking is a primary functional movement in DMD assessment and is closely related to established walking outcomes such as the six-minute walk test [31].

Rising from the floor. The rising-from-the-floor clip was defined from the first frame in which the subject initiated movement from a floor-level position—lying on the back, lying face-down, or seated—to the first frame in which stable upright standing was achieved. Only complete, independently executed sequences were included; attempts requiring physical assistance were excluded. This definition follows the Gowers manoeuvre used in DMD functional assessment [27], [31].

4.2.3 Dataset Composition by Task

Following segmentation and quality screening, the final dataset comprised 94 video clips, each from an identified distinct subject, distributed across two primary tasks. The 73 walking subjects and the 21 Gowers subjects are distinct identified individuals: no subject contributed video to both tasks. Task-level composition is summarised in Table 4.2.

Table 4.2: Dataset composition by task and diagnostic group following manual segmentation and quality screening. Differences from initial collection reflect stricter framerate and clip-quality filtering applied uniformly across groups to reduce avoidable processing bias.

Task	DMD	Control	Total
Overground walking	10	63	73
Gowers	15	6	21
Total	**25**	**69**	**94**

The class imbalance varies substantially across tasks, with the walking subset showing the most severe imbalance (DMD:Control ≈ 1:6.3). This reflects the availability of suitable public video rather than a deliberate design choice. Class imbalance is addressed at the classification stage through balanced class weighting, and motivates the use of Precision–Recall AUC as the primary evaluation metric alongside ROC-AUC [45]. The combined effect of class imbalance and limited sample size makes this a challenging classification setting.

4.3 Pipeline Overview

Building on the movement patterns and markerless-pose limitations reviewed in Chapter 2, the pipeline converts each single-camera RGB video clip into a scalar feature vector for task-specific classification. The analysis proceeds through five stages: pose extraction, signal conditioning and quality control, movement feature engineering, classifier training, and evaluation.

First, MediaPipe Pose Landmarker Heavy is used to extract landmark trajectories from each manually segmented clip. Second, landmark trajectories are filtered, smoothed, and transformed into the coordinate representation used for feature computation. Third, task-specific movement features are computed for walking and Gowers clips. Fourth, candidate classifiers are trained on the resulting feature matrices. Finally, held-out predictions are evaluated under the cross-validation procedure described in Section 4.8, with SHAP used only for post-hoc feature attribution. The complete workflow is shown in Figure 4.1.

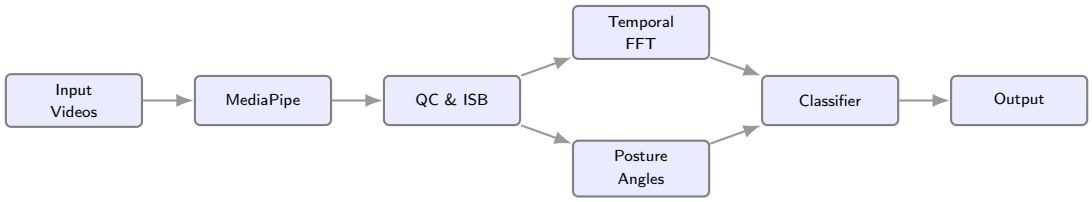

Figure 4.1: End-to-end pipeline architecture. Single-camera RGB video is manually segmented into task-specific clips, processed with MediaPipe pose extraction, filtered and transformed into the working coordinate representation, and converted into task-specific movement features. Candidate classifiers are trained and evaluated on the resulting feature matrices, with SHAP used after training for post-hoc feature attribution.

4.4 Pose Extraction and Coordinate Representation

4.4.1 MediaPipe Pose Landmarker Heavy

Skeletal landmark trajectories were extracted using the MediaPipe Pose Landmarker Heavy model [34]. The model was applied to each segmented RGB clip and its landmark outputs were used as the input to the coordinate transformation and feature extraction stages. Head and face landmarks were later excluded from feature computation as described in Section 4.5.

MediaPipe Pose Landmarker Heavy was selected as a practical, deployable pose-extraction backbone that provides the trunk, hip, knee, ankle, heel, and foot landmarks required by the engineered walking and Gowers features. The aim was not to benchmark pose estimators, but to evaluate whether a lightweight markerless-pose pipeline could support interpretable DMD-related movement classification from ordinary video.

MediaPipe Pose Landmarker Heavy uses a two-stage detector–landmark pipeline [34]. First, a detector identifies the person in the frame and estimates a body-centred region of interest using virtual body keypoints that describe the hip midpoint, approximate body scale, and torso orientation. The cropped person region is then passed to a landmark model, which predicts 33 full-body landmarks including the shoulders, elbows, wrists, hips, knees, ankles, heels, and feet (Figure 4.2).

The detector and landmark models are based on lightweight convolutional neural networks designed for real-time pose estimation. In video mode, MediaPipe tracks landmarks across frames and re-runs person detection only when tracking confidence falls below a threshold, making it efficient for continuous video clips.

The model outputs both normalised image landmarks and world landmarks. For normalised landmarks, x and y are image-normalised coordinates, while z is a relative depth coordinate with the hip midpoint as its origin; smaller z values indicate landmarks closer to the camera. The world-landmark output provides x, y, and z in metres, again with the origin at the hip midpoint.

4.4.2 Coordinate Representation and Body-Frame Transformation

MediaPipe world landmarks were used as the starting coordinate representation. The raw world-landmark vector for landmark i at frame t is denoted $\mathbf{p}_{i,t}^{\mathrm{mp}}$. From these coordinates, both MediaPipe world coordinates and body-centred ISB coordinates were retained downstream, with selective use depending on the feature type. The resulting feature pipeline used three coordinate representations: ISB/body-frame coordinates for joint-angle

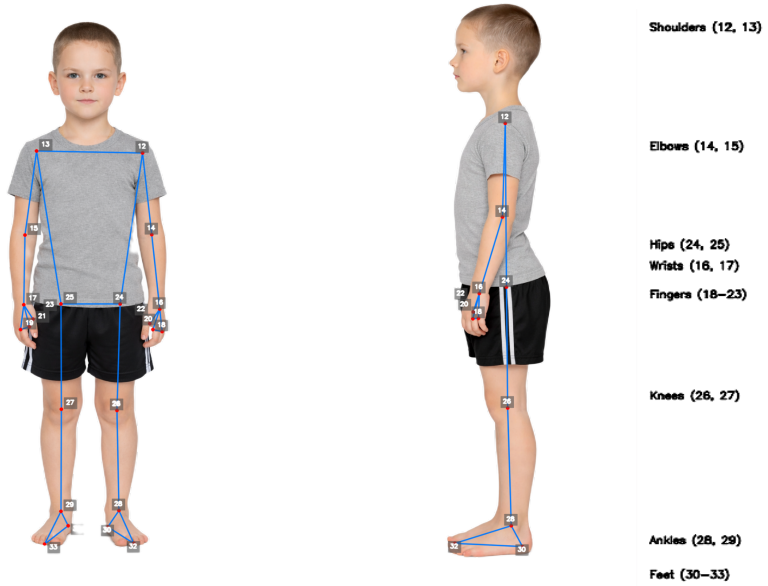

Figure 4.2: MediaPipe Pose Landmarker topology (33 full-body landmarks). Adapted from the Google AI Edge pose landmark documentation [46].

and body-relative gait features, world coordinates for gravity-referenced inclination and vertical-motion features, and image-space coordinates only for the Gowers hand-support proximity feature.

Landmarks were transformed to a person-centred ISB body frame [47], [48], with origin at the hip centre, +Y upward, +X forward (along walking direction), and +Z rightward (right-hand rule). Table 4.3 contrasts the two frames.

Table 4.3: Comparison of MediaPipe world-landmark and body-centered coordinate systems.

Axis	MediaPipe World Representation	Body Frame
X	Left–right (camera view)	Forward–backward (along walking direction)
Y	Up–down (image plane)	Up–down (vertical in world)
Z	Depth (away from camera)	Left–right (subject's lateral direction)

The transformation is a rigid rotation about the origin:

$$\mathbf{p}_{i,t}^{\text{body}} = R_t \left(\mathbf{p}_{i,t}^{\text{mp}} - \mathbf{o}_t \right),$$

where $\mathbf{o}_t$ is the hip centre and R_t is the rotation matrix defined by the three orthonormal axes at time t. This preserves distances and angles within the body frame while reduc-

ing camera-dependent variance. Depth-dependent features remain sensitive to monocular depth error; this limitation is discussed in Section 6.5.

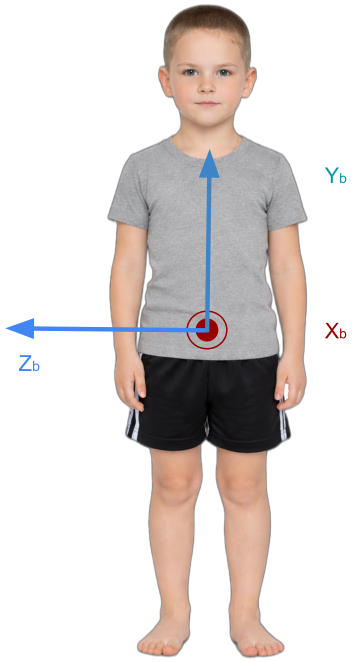

Figure 4.3: ISB body-frame coordinate system with origin at hip centre. +X is forward, +Y is up, and +Z is sideways to subject-right (right-hand rule).

4.5 Signal Preparation and Quality Control

Per-frame timestamps were derived from each video's reported framerate and frame index, producing a monotonic synthetic timeline for each clip. This avoided repeated or unstable container timestamps while preserving physical time units for velocity and frequency features.

4.5.1 Landmark Filtering

Prior to feature extraction, a subset of landmark columns was excluded from all downstream processing. Hand-digit landmarks (index finger, pinky, thumb) and facial landmarks were excluded. Hand-digit landmarks exhibited the lowest mean visibility scores across both cohorts and were not required by any extracted feature. Facial landmarks

were excluded as part of the privacy guardrailing described in Section 4.1. Image-space coordinate columns were discarded for both tasks, except in the Gowers pipeline, where image-space wrist and knee coordinates were retained for the hand-support proximity feature described in Section 4.6. The coordinate-frame policy used for feature computation is defined in Section 4.4.2.

4.5.2 Temporal Smoothing

Frame-to-frame landmark jitter arises from per-frame re-detection variability, compression artefacts in the source video, and stochastic neural network outputs. In this dataset, approximately 8–12% of frames were flagged by the quality-control pipeline before interpolation and smoothing were applied. Two sequential smoothing stages were applied.

Stage 1 — Smoothing during pose detection (internal). MediaPipe applies internal smoothing to landmark coordinates before returning them. This smoothing removes frame-to-frame noise while being smart about movement speed: during slow movements, it applies heavy smoothing to eliminate noise; during rapid movements, it reduces smoothing to preserve fast changes. Default parameters were retained, which may slightly reduce the peaks of very rapid movements in DMD gait but provide more stable overall trajectories.

Stage 2 — Additional smoothing (explicit, applied after pose detection). A smoothing filter (4th-order Butterworth with 9 Hz cutoff) is applied to remove high-frequency noise from pose landmarks. The 9 Hz cutoff was chosen because reported biomechanical content in gait signals lies below 6 Hz [49], [50]; the higher cutoff preserves walking motion while removing higher-frequency noise. The cutoff was derived from those frequency bounds rather than copied from their filter specifications, as neither study used the same 4th-order Butterworth design applied here. Clips with fewer than 30 finite samples in a coordinate column were not smoothed; such columns were retained as-is and excluded from downstream feature aggregation.

4.5.3 Landmark Quality Control

Two quality-control steps remove unreliable detections while keeping as much usable data as possible. First, landmarks with low confidence are marked as missing, and short gaps are linearly interpolated so brief occlusions do not break a clip. Second, landmarks that jitter between implausible positions across consecutive frames are flagged as unreliable. Walking clips with sustained both-sided knee tracking failure are removed, because they would

disrupt stride-based features. Gowers clips are kept more permissive because floor-level postures naturally hide some landmarks.

Additionally, clips shorter than 3 seconds after segmentation were excluded prior to feature extraction, reducing the walking set from 76 to 73 clips and the Gowers set from 22 to 21 clips.

4.6 Movement Feature Engineering

The feature set contains three types of variables. First, literature-motivated movement features approximate DMD-related walking and Gowers patterns described in Chapter 2. Second, pose-derived proxies translate movement concepts into quantities computable from single-camera landmarks, such as toe-walking fraction and hand-support proximity. Third, engineering descriptors summarise robust properties of the pose signals, such as rhythm and frequency content from high-visibility landmarks. These descriptors are included to improve robustness under noisy pose extraction and are not interpreted as established DMD-specific clinical biomarkers.

4.6.1 Notation for Feature Computation

Landmark i at frame t is denoted by $\mathbf{p}_{i,t}$, consistent with the world-landmark notation introduced in Section 4.4.2. In feature formulas, the index i is instantiated with anatomical names (e.g., $\mathbf{p}_{\text{heel},t}$, $\mathbf{p}_{\text{hip},t}$) for clarity. Each feature subsection states its coordinate frame once; within that subsection the superscript is omitted. Frame superscripts appear only in the coordinate-transformation definition (Section 4.4.2) and for image-space landmarks ($^{\text{img}}$), which form the explicit exception. Bold symbols denote vectors; matrices such as R_t are non-bold.

Pose-derived movement features were extracted from the smoothed landmark coordinate sequences. Feature engineering proceeded in two stages: a set of generic features used across all motor tasks, and task-specific feature families. All outputs are scalar summaries: for walking, features are first computed at stride level and then reduced to one value per video; for the Gowers task, features are computed directly at clip level. Head, facial, and hand-digit landmarks were excluded from feature computation.

4.6.2 Joint Angle Features

Joint flexion angles use the clinical convention rather than raw geometric angles, so that values are interpretable in clinical terms: 0° = full extension, and positive values increase

toward full flexion.

Joint flexion angles are computed as the angle subtended by two adjacent limb segment vectors at their shared joint. The raw geometric angle is computed using the standard three-point vector formulation:

$$\theta_{\text{raw}} = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}\right)$$

where $\mathbf{u}$ and $\mathbf{v}$ are the upper and lower segment vectors. The raw interior angle θ_{raw} was converted to a clinically conventional flexion angle as:

$$\theta_{\text{flex}} = 180^\circ - \theta_{\text{raw}}$$

such that 0° corresponds to full extension and values increase with flexion, following Winter [48]. This formulation was applied to both sides for the knee, hip, and elbow joints. Trunk tilt was decomposed into side-to-side tilt (side-to-side plane) and front-to-back tilt (front-to-back plane). Lateral pelvic tilt was computed as the signed vertical inclination of the inter-hip vector, and front-to-back pelvic tilt as the signed horizontal inclination of the same vector.

For each joint angle time series, the following summary statistics were extracted: mean, standard deviation, median, minimum, maximum, and robust range of motion (ROM_{p95-p5}), defined as the 95th minus 5th percentile range. This definition was preferred over the absolute maximum–minimum range to reduce sensitivity to transient outlier frames [32].

For the walking task, the ankle joint angle was computed as the interior angle at the ankle using the knee, ankle, and heel landmarks (knee→ankle→heel). This formulation does not correspond to the standard clinical ankle upward foot movement angle (shin bone-to-foot), which requires a foot-mounted goniometer (a clinical angle-measuring instrument); it is a geometric proxy whose values should be interpreted as relative indicators rather than anatomically precise measurements.

4.6.3 Segment Inclination Angles

Segment inclination angles are computed as the angle between each segment vector (thigh and shank, both sides) and the upward vertical direction, measured in gravity-referenced world coordinates. A value of 0° indicates a perfectly vertical segment; values increase with forward tilt. Per stride, mean and standard deviation of each inclination angle are extracted and aggregated to per-video medians and interquartile ranges.

Trunk tilt is similarly decomposed into front-to-back (forward lean) and side-to-side (lateral lean) components and computed relative to the upward vertical in world coordinates. While both ISB body-frame and world-frame versions are computed during preprocessing, only the world-frame trunk tilt features are retained for classification. This choice was made to prioritize robustness to monocular depth estimation artifacts, which can accumulate in body-frame transformations; world-frame representations remain stable across variable camera distances and angles characteristic of the heterogeneous YouTube source material.

4.6.4 Continuous Relative Phase

CRP quantifies hip–knee coordination timing rather than joint-angle magnitude [51], [52]. It is computed from the time-varying phase relationship between hip and knee motion via the Hilbert transform, which extracts instantaneous phase from angle signals:

$$\mathrm{CRP}(t) = \phi_{\mathrm{hip}}(t) - \phi_{\mathrm{knee}}(t)$$

where $\phi(\cdot)$ denotes the instantaneous phase of the zero-meaned angle signal obtained via the Hilbert transform, and the phase difference is unwrapped to remove 2π discontinuities. CRP mean and standard deviation were extracted per stride and reduced to per-video medians.

4.6.5 Bilateral Symmetry

Bilateral asymmetry is quantified using the Robinson Symmetry Index (SI) [53], which normalises left–right differences relative to the average magnitude:

$$\mathrm{SI} = \frac{|L - R|}{0.5(L + R)} \times 100$$

where L and R denote the left- and right-side median values of a given feature across strides. Values range from 0 (perfect symmetry) to 100+ (extreme asymmetry). The index was computed for peak knee angle, knee angle standard deviation, peak ankle angle, and stride length.

4.6.6 Gait Stride Segmentation

Walking features were normalised to stride-level summaries so clips with different frame rates and durations could be compared. A gait cycle is defined as one full stride of the same foot, from one contact event to the next contact event of that same foot.

Stride boundaries for walking were identified from peaks in the inter-leg angle signal. Each peak marks a point where the legs are spread apart, which is a stable way to segment the stride cycle in these videos. This is an approximation of gait timing, but it is robust enough for the noisy, unconstrained recordings used here.

Side assignment at each peak was determined by comparing the forward (+X-axis) displacement of each heel relative to the hip on the same side. Consecutive peaks from the same foot define a stride. Strides with durations exceeding 3.0 seconds were excluded as probable double-cycle detections. A minimum of four valid stride peaks per video was required; clips falling below this threshold were excluded from the walking analysis.

4.6.7 Spatiotemporal Gait Features

From the stride-segmented walking signal, the following spatiotemporal features were computed per stride and aggregated to per-video medians:

- **Stride frequency** f_{stride} — The dominant FFT frequency of the inter-leg angle signal within the physiological walking band (0.67–2.67 Hz), reported in Hz [54]. The FFT method is shared with the Step-Rate Consistency feature (Section 4.6.9); the distinction is that f_{stride} reports the frequency value itself, while Step-Rate Consistency reports how concentrated the power is at that frequency.

- **Geometric stride length** — Computed as $2L\sin(\theta/2)$, where L is median leg length (thigh bone + shin bone) and θ is inter-leg angle at peak spread [54].

- **Step width** — Computed as lateral distance between left and right heel positions, used as a proxy for base of support widening [19].

- **Double support fraction** — Computed as the fraction of frames where both heels are in contact with the ground, estimated from heel-motion signals in ISB coordinates [54].

- **Stride duration CV** — Computed as the coefficient of variation (CV; standard deviation divided by the mean, expressed as a percentage) of stride duration across strides [55].

- **Stride regularity and step regularity** — Computed from the autocorrelation of the inter-leg angle signal at stride and step lags [56].

- **Estimated walking speed** — Computed as $v_{\text{walk}} = f_{\text{stride}} \cdot \ell_{\text{stride}}$, where f_{stride} is stride frequency in Hz and ℓ_{stride} is the geometric stride-length proxy in metres, giving speed in m/s.
- **Trendelenburg proxy** — Computed as mean and standard deviation of hip drop during the weight-bearing phase [19].
- **Pelvic rotation robust range of motion** (ROM_{p95-p5}) — Computed as rotational-plane rotation angle range of inter-hip vector relative to mean walking direction [24].

4.6.8 Heel Speed

Heel velocity is computed from position trajectories to measure forward limb progression during walking. Direct differentiation amplifies pose-estimation noise, so velocity is computed via Savitzky–Golay filtering.

All coordinates in this subsection are in the ISB body frame. Let $\mathbf{p}_{\text{heel},t}$ denote the heel landmark. Velocity is computed via Savitzky–Golay differentiation:

$$\dot{\mathbf{p}}_{\text{heel},t} = \left.\frac{d\,\mathbf{p}_{\text{heel},t}}{dt}\right|_{\text{SG}}$$

and the speed (magnitude of velocity vector) is:

$$v_{\text{heel}}(t) = \|\dot{\mathbf{p}}_{\text{heel},t}\|$$

Per-frame instantaneous walking speed is then defined as the minimum of the left and right heel speeds:

$$v_{\text{walk}}(t) = \min(v_{\text{heel,L}}(t),\, v_{\text{heel,R}}(t))$$

This minimum captures the slower limb's forward progression during double-support phases and is more robust to monocular depth estimation noise than the faster (swing) limb's speed.

4.6.9 Frequency-Based Pose Features

Step rate, gait regularity, and smoothness were extracted from the inter-leg angle signal using FFT and autocorrelation [57]. These frequency-domain variables are treated as engineering descriptors from a high-visibility gait signal rather than as established DMD-specific clinical markers. Frequency-domain analysis is motivated by two complementary

reasons. First, gait is an approximately periodic task, so spectral and autocorrelation methods directly quantify stride rhythm and its consistency without requiring accurate event detection in noisy landmark sequences. Second, reduced gait regularity and increased stride-to-stride variability are reported features of DMD gait [55], [57], making step-rate concentration and autocorrelation peak strength natural candidate discriminators between DMD and typically developing children.

Three complementary metrics are extracted from FFT and autocorrelation. Let $X(f)$ denote the one-sided FFT coefficient at frequency f, and let $P(f) = |X(f)|^2$ denote spectral power.

Step-Rate Consistency (dominant peak-to-band ratio, PBR). Gait regularity is quantified as the concentration of power at the dominant gait frequency, and is reported as the Step-Rate Consistency index. The dominant FFT peak within the physiological gait band ($f \in [0.67, 2.67]$ Hz) is identified as the frequency with maximum power:

$$\text{PBR} = \frac{\max_{f\in[0.67,2.67]} P(f)}{\sum_{f\in[0.67,2.67]} P(f)}$$

High PBR values indicate concentrated periodic power; low values indicate diffuse power spread across frequencies. This index is used as the Step-Rate Consistency feature in the classifier.

Harmonic ratio (HR). Harmonic ratio is computed as:

$$\text{HR} = \frac{\sum_{k=1}^{n} |X(2k)|}{\sum_{k=1}^{n} |X(2k-1)|}$$

where $k \in \{1, \ldots, n\}$ indexes harmonics up to the n-th harmonic.

Autocorrelation peak (AC-peak). Periodic structure strength in the gait signal is computed via autocorrelation. The normalised autocorrelation function $\rho(\tau)$ of the inter-leg angle signal captures how well the signal repeats at different time lags. The peak height at the stride lag (corresponding to the dominant gait frequency) is extracted as:

$$\text{AC-peak} = \max_{\tau\in T_{\text{stride}}} \rho(\tau)$$

where T_{stride} denotes the lag interval corresponding to the stride-frequency band, approximately [1/2.67, 1/0.67] seconds (roughly 0.37–1.49 s), converted to samples for each

clip. This metric is independent of step-rate-based stride regularity indices and captures the overall periodicity strength of the gait signal.

4.6.10 Toe-Walking Proxy

Toe-walking was represented with a conservative frame-level proxy. A frame was flagged when two conditions were met: (1) no heel contact was detected from vertical heel motion, and (2) the ankle-angle proxy exceeded 95°. The threshold was used only as an empirical within-study flag for extreme ankle configurations, not as a clinical diagnostic cut-off.

The toe-walking fraction is computed as:

$$f_{\text{toe}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1} \left[\text{no\_heel}(t) \wedge \theta_{\text{ankle,proxy}}(t) \geq 95^{\circ} \right],$$

where $\mathbf{1}[\cdot]$ is an indicator function and T is the total number of frames.

4.6.11 GDI-Inspired Walking Deviation Proxy

This thesis computes a GDI-inspired walking deviation proxy (GDI*), not the clinical Gait Deviation Index, because the available single-camera landmarks do not provide the full marker-based 3D movement input required for standard GDI computation.

A normative reference was fitted on the 63 typically developing walking clips using 15 movement features: hip, knee, and ankle joint angles on both sides (median and interquartile range); walking speed; step width; and trunk and pelvis angles in the gravity-referenced world frame [26], [58]. For each subject, the proxy was computed as:

$$D_i = \|\mathbf{z}_i - \boldsymbol{\mu}_{\text{TD}}\|_2, \qquad \text{GDI}_i^* = 100 - 10D_i,$$

where $\mathbf{z}_i$ is the normalized walking-feature vector for video i, $\boldsymbol{\mu}_{\text{TD}}$ is the typically developing reference mean, D_i is the Euclidean deviation from the TD reference, and GDI_i^* is the GDI-inspired walking deviation proxy. A score of 100 corresponds to the typical mean and each 10-point decrement represents one standard deviation of deviation from it. This proxy was included as an additional scalar feature in the walking feature matrix.

4.6.12 Gowers Features

The Gowers task features were designed to capture rise duration, hip rise velocity, trunk movement, hand-support proximity, and joint-angle summaries. Unless noted otherwise, all landmark coordinates in this section are in the ISB body frame; image-space coordinates are marked explicitly with a $^{\text{img}}$ superscript.

Temporal features. Total rise duration is computed as the elapsed time from task onset to standing upright, operationally defined as the moment when the hip center reaches 90% of its maximum vertical height. The verticality index measures how high the hips have risen, normalized by leg length to account for body size differences [54]. Let $[\mathbf{p}_{\text{hip},t}]_y$ denote the vertical component of the hip-centre landmark:

$$V_{\text{index}} = \frac{\max_t [\mathbf{p}_{\text{hip},t}]_y - \min_t [\mathbf{p}_{\text{hip},t}]_y}{L_{\text{leg}}}$$

where L_{leg} is the estimated leg length (thigh + shank segment lengths).

Hip rise velocity is computed as the time derivative of the verticality index, restricted to frames where the hips are moving upward (positive derivative). This isolates the upward movement phase and excludes any downward motion. The mean hip rise velocity reflects the speed of the upward movement. Both metrics are extracted per clip.

Trunk flexion/extension. In the Gowers task, children start bent forward on the floor and straighten their trunk as they stand up. Two simple trunk signals are measured: how much the trunk moves and how upright it becomes during the rise.

Trunk elevation angle is computed as the angle between the hip-to-shoulder vector and the vertical. The trunk segment vector is defined as:

$$\mathbf{v}_{\text{trunk},t} = \mathbf{p}_{\text{shoulder},t} - \mathbf{p}_{\text{hip},t}$$

The elevation angle between this vector and the upward vertical $\hat{\mathbf{y}}$ is then:

$$\theta_{\text{trunk}}(t) = \arccos\left(\frac{\mathbf{v}_{\text{trunk},t} \cdot \hat{\mathbf{y}}}{\|\mathbf{v}_{\text{trunk},t}\|}\right)$$

Small angles (near 0°) indicate an upright trunk; large angles (near 90°) indicate a forward-bent or horizontal trunk. The robust range of motion (ROM_{p95-p5}) captures total trunk rotation throughout the rise, reflecting compensatory effort.

Additionally, an upright score is computed to measure how vertical the trunk becomes:

$$\text{Upright-score}(t) = \frac{|\mathbf{v}_{\text{trunk},t} \cdot \hat{\mathbf{y}}|}{\|\mathbf{v}_{\text{trunk},t}\|}$$

This ranges from 0 (fully horizontal) to 1 (fully vertical). Both metrics are extracted per frame and aggregated to per-clip medians.

Hand-support proximity. Hand-support proximity is measured by tracking the minimum same-side wrist-to-knee distance in image space during the rise:

$$d_{\text{hand-leg,min}} = \min_{t,\,s} \left\| \mathbf{p}^{\text{img}}_{\text{wrist}_s,t} - \mathbf{p}^{\text{img}}_{\text{knee}_s,t} \right\|_2$$

where $s \in \{L, R\}$ denotes body side, $\mathbf{p}^{\text{img}}_{\text{wrist}_s,t}$ is the image-space wrist position for side s at frame t, and $\mathbf{p}^{\text{img}}_{\text{knee}_s,t}$ is the image-space knee position for side s at frame t. Image-space coordinates use MediaPipe's normalised landmarks, with origin at the top-left corner of the frame. The minimum is taken across both sides and all frames of the full rise sequence. We use 2D image-space distances rather than world-frame distances because MediaPipe world landmarks are expressed relative to the hip centre, which is itself the centre of motion during the Gowers rise. As the hips move from floor to standing, the world-coordinate origin shifts continuously, making world-frame wrist-to-knee distances less stable as a proximity indicator. Camera motion during Gowers recordings is more stable than during walking, so image-space pixel distances provide a more stable measure of hand-to-leg proximity throughout the rise.

Joint angle features. Hip, knee, and elbow flexion angles are extracted over the full rise sequence. Hip and knee features capture the transition from floor-level flexion to final extension; elbow features capture upper-limb motion during push-up. All joint flexion angles use the clinical convention (0° = full extension, increasing toward flexion). Angles are extracted per frame and reported as minimum, maximum, and robust range of motion (ROM_{p95-p5}) across the full rise sequence per clip.

4.6.13 Feature Set Summary

The complete feature set is summarised in Table 4.4. Generic features were extracted for both tasks. Spatiotemporal and gait-specific features were extracted for the walking task only. Gowers features were extracted for the Gowers task only. The source labels separate features taken directly from movement-analysis literature from pose-derived approximations needed for single-camera video. A more detailed source table with the corresponding literature references is provided in Appendix A.2; that appendix table is the canonical reference for feature provenance.

Table 4.4: Summary of pose-derived movement feature groups and source types. "Literature-based" features derive from published movement-analysis references. "Pose-derived approximation" features adapt clinical or gait concepts to single-camera landmark data. "Pose quality" refers only to landmark-visibility checks.

Feature group	**Source type**
Generic (all tasks)	
Joint flexion angles (min, max, ROM_{p95-p5}, mean, std)	Literature-based
Trunk tilt (front-to-back, side-to-side)	Literature-based
Lateral pelvic tilt + front-to-back pelvic tilt	Literature-based
Segment tilt (thigh, shank)	Literature-based
Visibility confidence (regional)	Pose quality
Walking-specific	
Stride frequency (f_{stride}, Hz)	Literature-based
Step-rate consistency (PBR)	Pose-derived approximation
Harmonic ratio (HR)	Pose-derived approximation
Autocorrelation peak (AC-peak)	Pose-derived approximation
Stride length (geometric)	Literature-based
Double support fraction	Literature-based
Stride duration CV	Literature-based
Stride regularity (AC)	Pose-derived approximation
Robinson SI (knee, ankle, stride)	Literature-based
CRP hip–knee	Literature-based
Trendelenburg proxy	Pose-derived approximation
Pelvic rotation ROM_{p95-p5}	Literature-based
Step width, speed estimated	Pose-derived approximation
Stance frame fraction	Pose-derived approximation
GDI-inspired walking deviation proxy	Pose-derived approximation
Gowers-specific	
Rise duration, hip rise velocity	Pose-derived approximation
Trunk extension ROM_{p95-p5}	Literature-based
Min hand–leg distance (image-space)	Pose-derived approximation
Verticality index	Pose-derived approximation

4.7 Classification Setup

4.7.1 Classifier Selection

For each task (walking and Gowers), three tree-based models were evaluated: a single decision tree baseline (Tree), Random Forest (RF), and XGBoost (XGB). Tree-based methods are well-suited to the tabular feature matrices used here and are strong baselines in broad tabular-data benchmarks [39], [59]. These benchmarks are not specific to kinematic datasets, so they are used here as general evidence for tree-based baselines on small-to-medium tabular data rather than as direct evidence for movement-analysis data.

The Tree baseline provides a simple interpretable reference point. RF combines multiple trees trained on random subsets of features and samples, aggregating predictions via majority voting. XGB uses gradient boosting to sequentially add trees that correct errors from previous trees and provides a complementary boosted-tree baseline to RF. Class imbalance was handled through training-fold class weighting, and hyperparameters were kept deliberately conservative to reduce overfitting risk in the small-sample setting. RF was treated as the primary model because it supports tabular nonlinear interactions, does not require feature scaling, and is explainable via exact Tree SHAP; the rationale for preferring tree ensembles over neural networks for attribution is given in Section 4.8.4. Balanced class weighting was applied to Tree and RF; for XGB the positive-class weight was derived from the fold imbalance ratio.

4.7.2 Feature-Matrix Construction

For the walking task, per-stride features were aggregated to a single per-video feature vector by computing the median and interquartile range (IQR, the range between the 25th and 75th percentiles) across all valid strides. Median is robust to outliers inherent in noisy pose-estimation sequences (a key property for data derived from markerless pose estimation), while IQR captures stride-to-stride variability without assuming normality—important given small sample sizes and potentially skewed distributions of movement features. Both median and IQR summaries were included as candidate features; the final SHAP attribution reflects which of these ranked as most discriminative after training. Video-level features (step rate, double support fraction, stride regularity, stride duration CV, and symmetry indices) were already defined at the video level and required no further aggregation. For the Gowers task, features were computed directly over the full clip and summarized with median and IQR, since the Gowers motion does not repeat in identifiable strides. The resulting feature matrix contained one row per video clip and one column per

feature. Columns with more than 50% missing values were excluded prior to model fitting; remaining missing values were imputed with the training-set column median within each cross-validation fold.

Raw ISB coordinates and intermediate helper vectors were not passed to the classifier; only derived movement features with a clear gait interpretation were included. For the walking task, the feature set covers lower-limb angles, trunk lean, step width, stride features, step rate, and regularity metrics; upper-limb features were excluded because the walking feature set was designed around lower-limb and gait-rhythm patterns identified in the background literature. For the Gowers task, upper-limb (elbow) features were retained because they capture hand-support movement.

4.7.3 Hyperparameter Search Space

Hyperparameters were tuned within the cross-validation procedure, with model comparison based on PR-AUC. Ranges were kept conservative to reduce overfitting risk on a dataset of 73–94 samples, informed by tabular-model benchmarks [39]. The XGBoost positive-class weight was derived from the class imbalance within each training fold so that DMD errors carried more penalty when the fold was control-dominated. The search ranges are summarised in Table 4.5.

Table 4.5: Hyperparameter search ranges for Random Forest and XGBoost. Settings marked RF only or XGB only are specific to that model; all others apply to both.

Setting	**RF range**	**XGB range**
Number of trees	$\{100, 300\}$	$\{100, 300\}$
Maximum depth	$\{10, 20\}$	$\{3, 6\}$
Minimum node size	$\{1, 5\}$	$\{1, 5\}$
Row sampling rate	$\{0.7, 1.0\}$	$\{0.7, 1.0\}$
Column sampling rate	$\{0.7, 1.0\}$	$\{0.7, 1.0\}$
Learning rate (XGB only)	—	$\{0.03, 0.10\}$
Positive-class weight (XGB only)	—	Derived per fold

4.8 Evaluation Design

4.8.1 Cross-Validation Design

All classifiers were evaluated with leave-one-out cross-validation (LOO-CV) grouped by video identifier. In each fold, the videos from one subject are held out as test data, while all remaining subjects' videos are used for training and hyperparameter tuning. Since each subject contributed exactly one video in this dataset, this is effectively subject-level leave-one-out evaluation. For walking, stride-level features were collapsed to one video-level row by taking the median across strides, so clip length does not change the final feature vector.

Hyperparameter tuning (via grid search on PR-AUC) was nested inside each fold using only the training subjects. Feature aggregation, missing-data imputation, class-weight derivation, threshold selection, and prediction were all performed without access to the held-out video.

4.8.2 Metrics and Baseline

The goal of the evaluation is to establish a patient-level cross-validated feasibility baseline: can pose-derived movement features separate DMD from typically developing children at all, and which features contribute most? The metrics are chosen to reflect the two tasks' very different class structures.

For the **walking** task (10 DMD, 63 control; ratio $\approx$ 1:6.3), DMD is the minority positive class. ROC-AUC measures how well the model ranks DMD videos above controls across all thresholds; it does not require a fixed threshold and supports comparison with prior work such as Ali et al. [35]. PR-AUC summarises the precision–recall curve and is the *primary* optimisation metric here because it is more sensitive to minority-class performance than ROC-AUC under strong class imbalance [45]. A prevalence-based no-skill baseline (PR-AUC $\approx 10/73 \approx 0.14$) is reported so results can be interpreted against chance.

For the **Gowers** task (15 DMD, 6 control), the imbalance is reversed and the control support is very small. With only six control clips, even a model that ranks most DMD videos above most controls will produce highly variable AUC estimates. All Gowers metrics are therefore treated as *exploratory descriptors* rather than reliable performance estimates; they indicate whether the expected direction of separation is present, not whether the classifier generalises.

Accuracy measures the fraction of all held-out clips classified correctly. Precision measures the fraction of predicted DMD cases that are truly DMD, while recall measures the fraction of true DMD cases that are detected. Threshold-dependent metrics were reported

at the operating point selected by the Youden criterion, which maximises sensitivity + specificity − 1 on the training portion of each fold and is then applied to the held-out fold. Sensitivity (recall) and specificity report the true-positive and true-negative rates. Balanced accuracy averages sensitivity and specificity and is robust to class imbalance. F1 combines precision and recall. Confusion matrices show true positives, false positives, true negatives, and false negatives at the selected operating point. ROC-AUC and PR-AUC are used for threshold-free model comparison, while accuracy, precision, recall, specificity, balanced accuracy, and F1 are reported as operating-point metrics for the selected model in each task. Metrics were computed after pooling all leave-one-out held-out predictions.

Statistical comparison of ROC-AUC. Pairwise DeLong tests were used to compare ROC-AUC values between classifiers on the pooled leave-one-out prediction scores [60]. The test is appropriate for paired ROC-AUC comparison because all classifiers were evaluated on the same held-out videos, producing correlated ROC curves. Given the small number of DMD videos, the resulting p-values were interpreted as descriptive evidence rather than definitive proof of classifier superiority. Final model selection therefore considered ROC-AUC, PR-AUC, class-imbalance behaviour, interpretability, and stability under the leave-one-out setting.

4.8.3 Univariate Analysis

Univariate comparisons between the DMD and typically developing control groups were conducted at clip level using Mann-Whitney U tests (one row per video clip; for walking, stride-level features were first aggregated to per-video medians). Effect sizes are reported as rank-biserial correlation r_{rb}. Because task subsets are small, univariate results are reported as descriptive characterisation rather than as definitive inferential claims. For the Gowers task, formal Mann-Whitney U tests were not conducted because the control group comprised only six clips, making test outcomes unreliable; separation is reported descriptively via rank-biserial correlations (r_{rb}) only.

4.8.4 Attribution Analysis

SHAP (SHapley Additive exPlanations) was used to attribute RF predictions to individual pose-derived movement features [42], [43]. For each video, SHAP assigns each feature a signed contribution indicating whether it pushed the prediction toward DMD or toward control. Exact Tree SHAP values were computed from the trained RF structure [43]; this method is preferred over model-agnostic SHAP variants, which rely on approximation

choices and background distributions, making it a more direct and consistent fit for tree ensembles [61]. SHAP values explain the fitted model rather than the underlying disease mechanism; feature-relevance attribution with Shapley values is not equivalent to causal inference. They are therefore interpreted as model-level evidence that certain pose-derived movement features contributed to DMD-control separation, not as causal clinical claims.

Per-feature mean absolute SHAP values were aggregated to produce a global feature importance ranking for each task. A final model was fitted on the full dataset for attribution purposes.

5

Results

5.1 Overview of Results

The results show a clear difference between the two tasks. Walking produced the more stable classification signal, whereas Gowers remained a low-power feasibility analysis because of the small and imbalanced control subset. The chapter first reports landmark quality-control output (Section 5.2) and univariate evidence (Section 5.3), then presents classification performance in Section 5.4 and feature attribution in Section 5.5.

5.2 Landmark Quality Control

Before feature-level statistics and model evaluation, quality-control output was inspected to verify that smoothing reduced frame-level jitter without removing the underlying movement pattern. Figure 5.1 shows a representative walking sample (`dmd_walking_07`), selected because it contains both raw noise typical of single-camera MediaPipe detections and sufficient consecutive frames to demonstrate the effect of smoothing: the prepared trace is visibly less noisy while preserving the stride-scale oscillation shape.

In results interpretation, outlier handling is therefore reported as a data reliability step, not as a model-performance contribution: clips with persistent tracking failure are filtered at preparation time, whereas short gaps are filled and smoothed. The classification and separation results reported below are computed on this post quality-control (QC) cohort.

Quantitatively, the QC stage flagged and filtered a substantial number of unreliable coordinates. Videos were first discarded when the segment in which posture was captured was shorter than 3 seconds, which reduced the walking set from 76 to 73 clips (3.9% removed) and the Gowers set from 22 to 21 clips (4.5% removed). Within the retained

cohorts, high-jitter landmarks and short-gap interpolation affected only small shares of the available coordinate values, while the final removal of frame rows with unfillable gaps remained a minority operation relative to the posture-captured frames. The net effect is that most task-relevant motion was retained after QC, with aggressive filtering restricted to clearly unreliable detections.

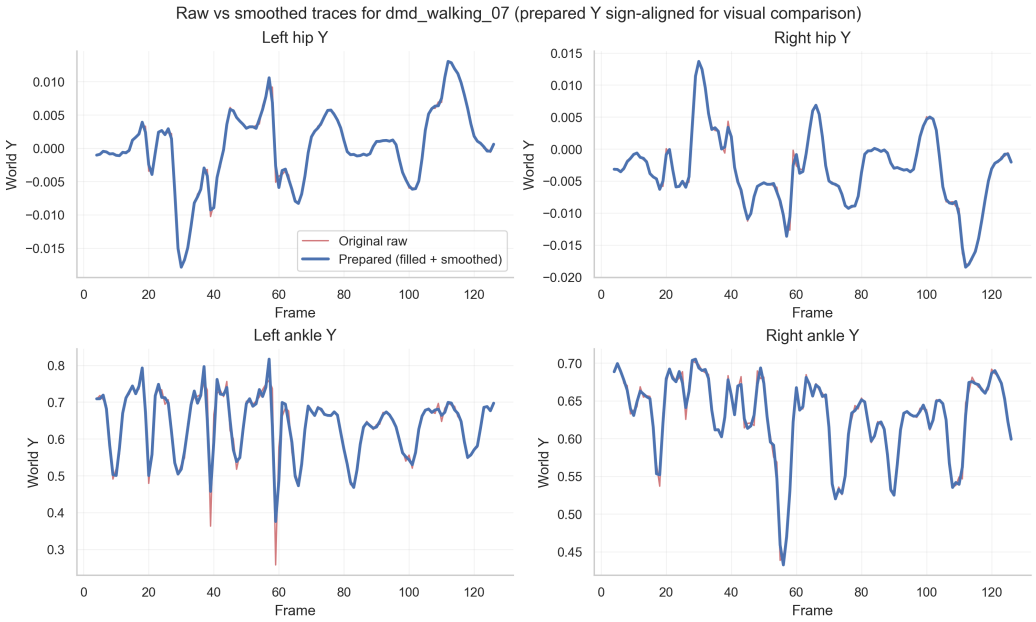

Figure 5.1: Raw vs smoothed traces for dmd_walking_07. Red shows original raw coordinates and blue shows prepared coordinates after filling and smoothing, plotted on the same frame axis for direct comparison.

5.3 Univariate Results

5.3.1 Walking

The walking feature set ($n = 73$ clips; DMD = 10, Control = 63) showed moderate univariate separation for a subset of lower-limb pose-derived movement features (Table 5.1). The strongest effects combine normalised hip-displacement variability with angle summaries, stride-length structure, and inter-leg angle statistics. This pattern is consistent with walking differences reported in DMD literature, including altered hip and upper-leg control and gait-rhythm changes, rather than pure speed differences. The convergence of joint-angle summaries with normalized displacement features suggests that both explicit angle measurements and inferred postural positioning contribute to separation, capturing related but

non-identical aspects of the movement pattern.

Table 5.1: Top walking features associated with DMD-control separation. Separation is the signed rank-biserial correlation from a two-sided Mann–Whitney U test on the current walking feature set after metadata and coordinate columns are excluded by the pipeline configuration.

Rank	Feature	Separation	p-value
1	Left Hip Vertical Variability (norm.)	-0.540	0.007
2	Right Hip Flexion Median (deg)	+0.441	0.026
3	Right Hip Vertical Variability (norm.)	-0.441	0.026
4	Stride Length P95 (norm.)	+0.413	0.038
5	Inter-Leg Angle Median (deg)	+0.400	0.044
6	Stride Frequency (Hz)	+0.387	0.051
7	Right Thigh Inclination Median (deg)	+0.375	0.059
8	Inter-Leg Angle Variability (deg)	+0.362	0.069
9	Left Thigh Inclination Median (deg)	+0.362	0.069
10	Step-Rate Consistency (index)	-0.337	0.091

5.3.2 Gowers

The Gowers feature set showed limited statistical power due to small sample size and class imbalance. Rank-biserial correlation was computed via a two-sided Mann–Whitney U test to rank features descriptively, but the resulting p-values are not interpreted as confirmatory given the very small control support ($n = 6$); they are reported for completeness only. The top ranked separation features are shown in Table 5.2.

The Gowers separation ranking is driven mainly by knee posture, trunk posture, and hand-to-leg support distance. The strongest signal is the right knee flexion median, followed by torso uprightness and the minimum hand-to-leg distance summary. The next features include left knee flexion, hip height, elbow flexion, and trunk forward-lean summaries, which are consistent with upper-body compensation during rise maneuvers. The p-values are moderate rather than very small, which is in line with the small sample size and limited control support.

5.4 Classification Performance

The hyperparameter settings selected within cross-validation were identical across both tasks. RF converged to the upper boundary of the search grid (300 trees, depth 20),

Table 5.2: Top Gowers features associated with DMD-control separation. Separation is the signed rank-biserial correlation from a two-sided Mann–Whitney U test on the current Gowers clip feature table.

Rank	Feature	Separation	p-value
1	Right Knee Flexion Median (deg)	+0.622	0.029
2	Torso Uprightness Median	-0.533	0.066
3	Hand-Leg Distance Median (norm.)	+0.533	0.066
4	Left Knee Flexion Median (deg)	+0.467	0.112
5	Hip Height Median (norm.)	-0.467	0.112
6	Hand-Leg Distance Variability (IQR, norm.)	-0.400	0.178
7	Left Elbow Flexion Median (deg)	-0.333	0.267
8	Trunk Forward-Lean Variability (IQR, deg)	-0.311	0.302
9	Right Elbow Flexion Median (deg)	+0.289	0.340
10	Trunk Forward-Lean Median (deg)	+0.289	0.340

suggesting that additional capacity was beneficial under balanced class weighting with this sample size. XGB converged to the most regularised configuration (100 trees, depth 3, learning rate 0.03), consistent with the small training sets. The search ranges are given in Table 4.5. Classification performance for all models and tasks is summarised in Table 5.3, which is the main threshold-free model-comparison reference for the chapter. Operating-point metrics for the selected model in each task are summarised in Table 5.4. Metric definitions and the evaluation protocol are given in Section 4.8. The following subsections interpret the walking and Gowers results separately.

Table 5.3: Model comparison by task. Best values within each task are shown in bold.[2]

Task	Model	ROC-AUC/AUC	PR-AUC	Base	Δ
Walking	Tree	0.602	0.365	0.137	+0.228
	RF	**0.825**	**0.511**	0.137	**+0.374**
	XGB	0.687	0.447	0.137	+0.310
Gowers	Tree	0.650	0.871	0.714	+0.157
	RF	0.622	0.743	0.714	+0.029
	XGB	**0.733**	**0.897**	0.714	**+0.183**

Table 5.4: Operating-point metrics for the selected model in each task. Metrics were computed from pooled leave-one-video-out predictions after applying the Youden-selected threshold within each training fold.

Task	Model	Accuracy	Precision	Recall	Specificity	Balanced accuracy	F1
Walking	RF	0.822	0.412	0.700	0.841	0.770	0.519
Gowers	XGB	0.714	0.909	0.667	0.833	0.750	0.769

5.4.1 Walking

5.4.1.1 Model Comparison

For walking, the best-performing model in Table 5.3 is the tuned RF, which achieved ROC-AUC = 0.825 and PR-AUC = 0.511 against a prevalence baseline of 0.137. The corresponding Tree and XGB baselines reached lower AUC values, so the main result is not just above random ranking but also above the alternative tree models under the same evaluation protocol. This level of performance still reflects a difficult small-sample feasibility setting rather than a model ready for real-world use.

AUC differences were tested using DeLong's test for correlated ROC curves [60], which compares ROC-AUC values directly on the continuous probability outputs without requiring a threshold. Given that only 10 DMD walking videos were available, the resulting p-values were interpreted as supporting evidence for model selection rather than definitive proof of classifier superiority. RF significantly outperformed the single Decision Tree (AUC 0.825 vs. 0.602; $z = 3.17$, $p = 0.002$). The RF–XGB gap (AUC 0.825 vs. 0.687) was borderline ($z = 1.89$, $p = 0.059$), and XGB did not significantly outperform Tree ($z = 1.02$, $p = 0.309$). RF was therefore selected as the primary walking model, supported by higher ROC-AUC and PR-AUC, stable leave-one-out performance, and direct TreeSHAP attribution.

The absence of a speed confound ($p = 0.910$, Section 5.3) strengthens the interpretation: the classifier discriminates on movement quality rather than movement pace. In any future low-prevalence setting where true DMD prevalence would be far lower than 13.7%, positive predictive value would degrade substantially, and the pipeline should be understood as a feasibility pipeline for identifying DMD-relevant movement-quality differences, not as a clinical decision tool.

5.4.1.2 Selected Model Performance

The operating-point metrics in Table 5.4 are consistent with the ranking result: the selected RF model captures most DMD clips while retaining moderate specificity in a control-dominated cohort.

5.4.1.3 Confusion Matrix

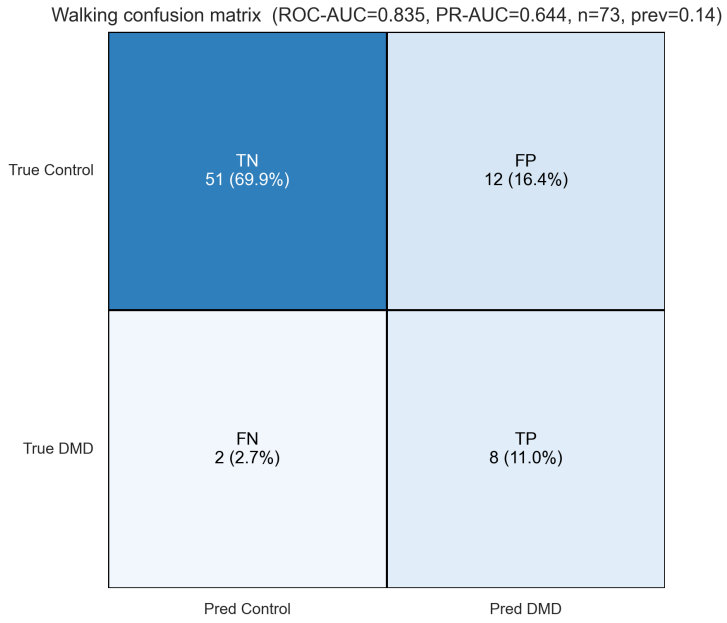

Figure 5.2: Confusion matrix for walking classification under leave-one-video-out cross-validation.

Of the 10 DMD clips, 7 were correctly identified (sensitivity = 0.700); of the 63 control clips, 53 were correctly identified (specificity = 0.841). The 10 false positives represent the main source of error in a control-dominated cohort.

5.4.1.4 ROC Curve, Feature Distributions, and Gait-Cycle Profiles

Figure 5.3 shows the top-six features by univariate separation; the same features dominate the SHAP ranking in Section 5.5, providing convergent evidence that the classifier operates on movement-level signals.

Figure 5.4 shows the ROC curve for the RF model; the area under the curve (0.825) reflects consistent ranking of DMD clips above controls across the full range of thresholds.

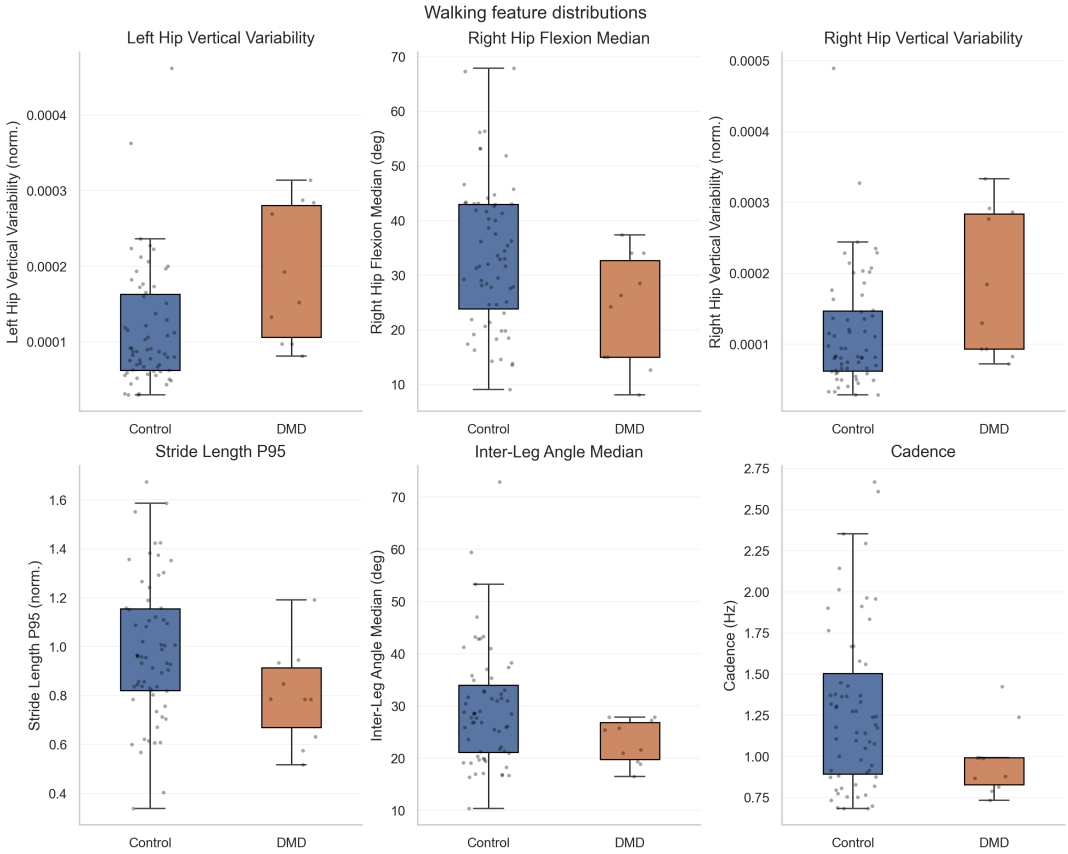

Figure 5.3: Walking classification feature distributions associated with DMD-control separation. Top six features are ranked by statistical separation. Box plots show median, quartiles, and distribution spread.

Figure 5.5 provides a time-normalized view of the same walking differences captured by the scalar features. In this cohort, the DMD median profiles appear shifted relative to controls across broad parts of the cycle, rather than at a single isolated point. This is consistent with a distributed walking-pattern (gait) difference, but the figure should be interpreted descriptively: overlap between IQR bands is substantial, and the sample size is limited. Accordingly, these curves are used as qualitative support for the feature-level and classifier results, not as a stand-alone basis for strong phase-specific claims.

5.4.2 Gowers

5.4.2.1 Model Comparison

For Gowers, the best-performing model in Table 5.3 is XGB, which achieved ROC-AUC = 0.733 and PR-AUC = 0.897 against a prevalence baseline of 0.714, yielding $\Delta = +0.183$.

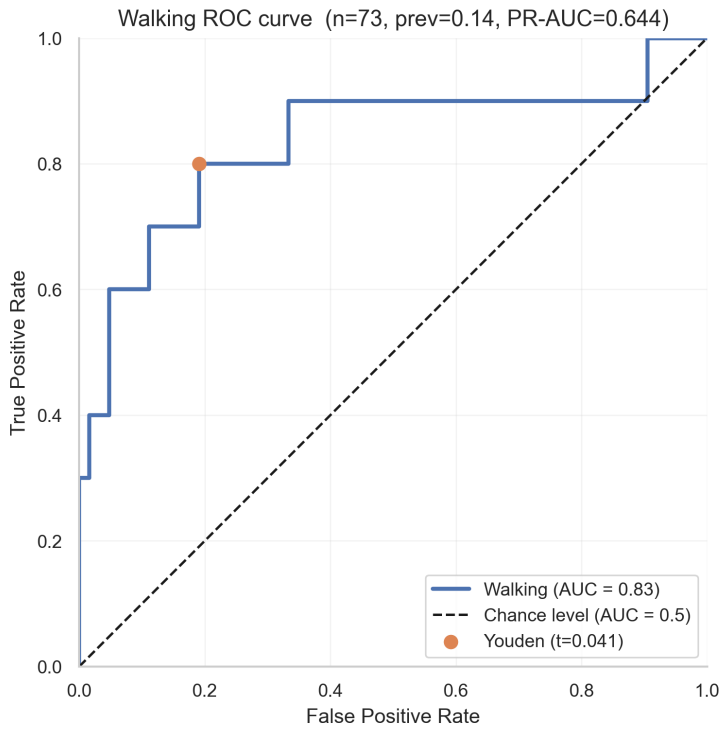

Figure 5.4: ROC curve for walking classification. Dashed line indicates the random classifier baseline.

Across models, Tree and RF performance was notably weaker (ROC-AUC = 0.650 and 0.622 respectively), whereas XGB achieved the highest ranking metric. DeLong's test for correlated ROC curves [60] found no statistically significant AUC differences between any classifier pair (XGB vs. RF: $z = 0.83$, $p = 0.406$; XGB vs. Tree: $z = 0.66$, $p = 0.508$; RF vs. Tree: $z = 0.17$, $p = 0.864$), consistent with the very limited control support ($n = 6$) which precludes reliable discrimination testing. The feature ranking is more informative than the point estimate itself, suggesting the learned features line up with the expected movement pattern. More reliable validation requires larger and better balanced samples.

5.4.2.2 Selected Model Performance and Sample Size Limitations

For the selected Gowers XGB model, the operating-point metrics in Table 5.4 should be interpreted as compact descriptive summaries of the chosen threshold rather than as stable real-world performance estimates. The small control group dominates the interpretation: the Gowers results should be understood primarily as evidence that the feature set captures DMD-relevant movement patterns, not as a validated classification benchmark.

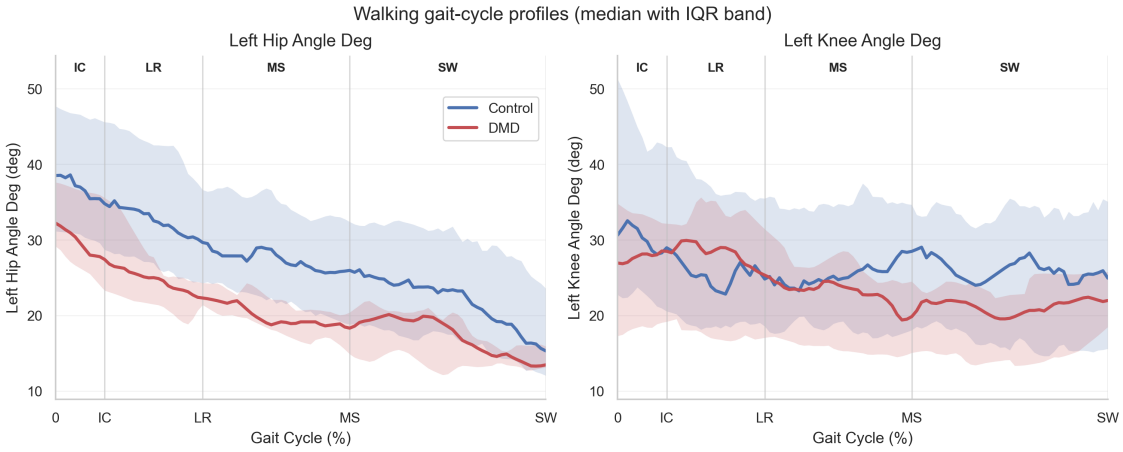

Figure 5.5: Walking gait-cycle profiles (0–100% cycle) comparing DMD and control groups for key stride-normalized signals. Curves show median trajectories and shaded bands show interquartile ranges across videos.

5.4.2.3 Confusion Matrix and ROC Curve

Associated visualisations are provided in the appendix (Figures A.1 and A.2).

5.5 Feature Attribution

The SHAP attribution methodology and its limitations are described in Section 4.8.4. The final model was fitted on the full dataset for attribution purposes. Per-feature mean absolute SHAP values for the walking task are shown in Figure 5.6.

5.5.1 Walking

The top SHAP features centre on inter-leg positioning, hip mobility, and gait rhythm, consistent with patterns reported in DMD walking literature. Inter-Leg Angle Median emerged as the dominant contributor, with lower median inter-leg spread pushing predictions toward DMD, consistent with the reduced stride range and hip repositioning associated with hip abductor compensation [24]. Step-Rate Consistency ranked second, capturing differences in gait regularity between groups. Right Hip Flexion Median ranked third, with reduced hip flexion pushing toward DMD, consistent with altered hip and upper-leg control. Hip Vertical Variability also featured prominently, capturing pelvic instability during single-leg stance. Stride Frequency, Left and Right Thigh Inclination Median, and Stride Length

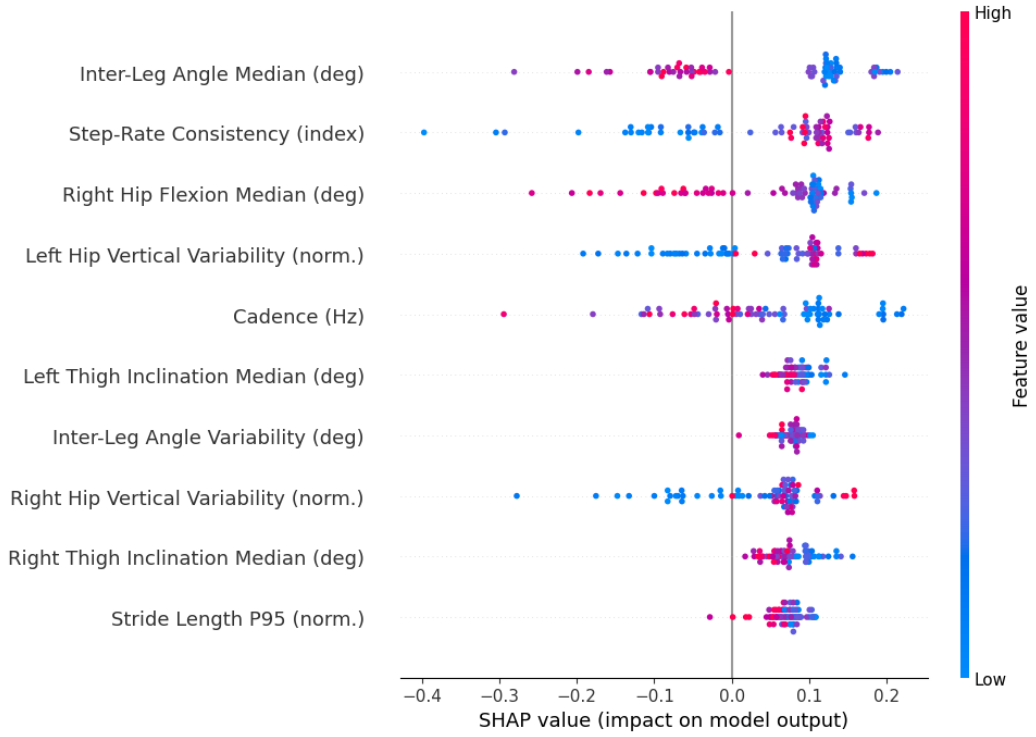

Figure 5.6: SHAP beeswarm plot for walking classification. Each point represents one video; horizontal position shows SHAP value (left = pushed toward control, right = pushed toward DMD); colour shows feature value (red = high, blue = low). Features ranked by mean absolute SHAP value (importance).

P95 rounded out the top features, together reflecting the compensatory gait pattern of reduced stride length, altered limb advancement, and rhythm adjustment documented in DMD walking studies [24].

The dominance of lower-limb features and gait rhythm metrics is consistent with DMD walking literature, where hip and knee weakness are associated with movement adaptations such as increased step rate, reduced stride length, and altered stance timing. The broad agreement between univariate separation (Table 5.1) and SHAP rankings suggests that classifier decisions are driven by pose-derived movement features rather than uninterpreted raw coordinate patterns alone. Raw landmark coordinates were explicitly excluded from the model to keep the interpretation clear.

5.5.2 Gowers

Right Knee Flexion Median emerged as the dominant SHAP contributor, with reduced knee flexion pushing toward DMD and reflecting the deep squat-like loading configuration during the transitional phase of impaired floor rise [27]. Left Hip Flexion Median ranked second, capturing hip mobility during the rising movement. Torso Uprightness Median ranked third, reflecting the trunk elevation phase of the manoeuvre. Hand-Leg Distance Median ranked fourth, capturing the proximity of the hands to the thighs that defines the Gowers sign [27]. These features map onto four movement components of the Gowers manoeuvre: lower-limb loading, hip mobility, trunk elevation, and upper-limb support.

Given the very small Gowers control support, the SHAP attribution mainly serves to test whether the learned signals are consistent with the published movement description of the Gowers manoeuvre. The fact that the top features do align with that description suggests the feature engineering captures the right constructs, even though the current sample is too small for a stable classification estimate.

6

Discussion

6.1 Summary of Findings

This thesis set out to determine whether interpretable pose-derived movement features extracted from single-camera RGB video could provide feasibility-level separation between walking children with Duchenne muscular dystrophy and typically developing peers under patient-level cross-validation, using walking and the Gowers manoeuvre as the two tasks. The practical motivation is that these movements may be analysed from ordinary video, which reduces dependence on specialist equipment and makes movement-analysis pipelines easier to scale. The pipeline also keeps the results interpretable through feature-level attribution, so the model output can be related back to movement patterns rather than treated as a black box.

Detailed results are presented in Chapter 5. The walking task provides the main discriminative result, while the Gowers manoeuvre is mainly informative at the level of movement-feature alignment and is limited by sample size and class imbalance. The walking result is the strongest outcome in the thesis and the most suitable point of contextual comparison with prior DMD markerless-video work [35]: the evaluation setup is stricter than that closest benchmark (patient-level LOO-CV versus within-dataset split), and the most important feature families match movement patterns reported in DMD literature. The combination of ROC-AUC, PR-AUC, and SHAP attribution supports this: the model separates DMD from typical development better than chance in this dataset, and the learned features are movement-level plausible rather than arbitrary.

Return to Research Questions. This thesis addressed three core research questions, which are now answered in light of the results:

1. *To what extent can single-camera RGB video separate DMD from TD in walking and the Gowers manoeuvre?* For walking, a tuned random forest achieves ROC-AUC = 0.825 under leave-one-out cross-validation, providing usable feasibility-level discriminative signal; the Gowers manoeuvre shows movement-plausible feature attribution but is limited by sample size and severe class imbalance.

2. *Which pose-derived movement features contribute most to that separation?* SHAP analysis identifies inter-leg angle median, step-rate consistency, right hip flexion median, and hip vertical variability as the top discriminative features for walking, while right knee flexion median, torso uprightness, and hand-leg distance median are the top-attributed Gowers features (exploratory only; $n = 6$ controls, severe class imbalance).

3. *How well do those learned features line up with movement patterns described in the DMD literature?* The feature attribution is consistent with reported DMD walking patterns: reduced inter-leg spread, restricted hip flexion, pelvic instability, and compensatory stride and rhythm adjustments are compatible with published evidence on hip abductor weakness, proximal muscle insufficiency, and reduced hip mobility [24], [62].

6.2 Walking Classification

6.2.1 Performance in Context

The walking ROC-AUC of 0.825 under LOO-CV is the primary quantitative result of this thesis. The closest DMD markerless-pose classification benchmark is Ali et al. [35], whose reported gait-classification results are noted as contextual prior-work reference in the footnote to Table 5.3. Ali et al. report higher headline performance, including accuracies of 0.962 for SVM with 3D HPE and 0.970 for their 3D HPE deep model, as well as AUC values for the deep models. These values should not be read as a direct leaderboard comparison, because the studies differ in dataset construction, pose representation, reported metrics, and validation setup. The result reported here is therefore best interpreted as a more conservative estimate under leave-one-video-out evaluation with one retained clip per identified subject.

The PR-AUC of 0.511 against a prevalence baseline of 0.137 shows clear improvement over a random classifier. Under the severe class imbalance of the walking subset (10 DMD vs 63 Control), a classifier that assigns the DMD label to every subject would achieve

ROC-AUC = 0.500 and PR-AUC = 0.137. In any future low-prevalence use context where true DMD prevalence would be far lower than 13.7%, positive predictive value would drop further, so this result should be read as an early feasibility signal rather than a standalone diagnostic result.

6.2.2 Interpreting the Movement Signal

A critical concern in heterogeneous video classification is whether the classifier detects DMD-relevant movement quality or movement pace. The walking univariate results show no walking speed difference between groups, which argues against pace as the main confound. This finding aligns with documented DMD compensation: children with DMD maintain walking speed through increased step rate despite reduced stride length [24]. The convergence between the univariate ranking and the SHAP feature ranking suggests that classifier decisions reflect movement-related signals rather than dataset artifacts.

Cross-fold stability of feature importance rankings provides additional support: the top features — inter-leg angle median, hip flexion, and hip vertical variability — ranked consistently across LOO-CV folds, suggesting the classifier learned reproducible movement patterns rather than subject-specific idiosyncrasies.

The prominence of features reflecting hip positioning and pelvic stability is notable. Hip Vertical Variability, captured in the top SHAP features, reflects pelvic instability during single-leg stance consistent with hip abductor weakness: under insufficient hip abductor force, the pelvis drops toward the swing side (Trendelenburg sign), producing greater vertical excursion of the hip landmark than is typical. Inter-Leg Angle Median and Thigh Inclination Medians complement this picture by capturing the restricted stride range and altered limb advancement associated with proximal muscle insufficiency. This interpretation requires validation in a larger clinically verified cohort.

6.2.3 Reversed Direction Features

Some features showed directions opposite to initial expectations from DMD movement literature: stride-duration variability metrics were lower in DMD than controls. Since walking speed was matched between groups, this reversal is not attributable to a speed confound. The most plausible explanation is a recording context effect: control clips sourced from YouTube contain more locomotion variability (children stopping, turning, and changing pace), whereas DMD clips consistently show slow, sustained, continuous walking due to effort limitation. This highlights a fundamental challenge of working with

heterogeneous video data: the statistical properties of a clip reflect not only the subject's movement but also the recording context. Controlled recording protocols are required to disambiguate these effects.

6.2.4 Features That Did Not Discriminate

Several features showed near-zero SHAP importance and weak univariate separation: step rate, double support fraction, autocorrelation-based stride regularity, CRP, and pelvic rotation range of motion (ROM). Two explanations are plausible. First, step rate and double support fraction are phase-locked to maximum inter-leg spread rather than heel contact, introducing systematic phase error (see Section 6.5, Stride Segmentation). Second, pelvic rotation ROM is estimated in body-centred world coordinates that are unreliable under mixed camera angles, suppressing this feature's discriminative validity despite its documented relevance to DMD progression [24] (see Section 6.5, Recording Heterogeneity).

6.3 Gowers Classification

The Gowers result (Chapter 5) is mainly useful as a movement-level check on the feature set rather than as a stable classifier estimate. SHAP attributions still line up with the expected movement pattern, so the feature engineering appears to capture relevant constructs. The main constraint is data availability: YouTube sourcing of the Gowers manoeuvre yielded very limited control support. Future work should prioritise controlled home-video collection with a target of 40–50 clips and balanced class distribution.

6.4 Contribution of the Two-Task Pipeline

A key contribution of this thesis is the establishment of an end-to-end, patient-level cross-validated classification pipeline across two DMD-relevant movement tasks from single-camera consumer video; to the author's knowledge, no prior work combines LOO-CV performance estimation with SHAP-based feature attribution for both walking and the Gowers manoeuvre from publicly sourced video. Prior work is limited in task scope: Ali et al. [35] address walking pattern (gait) only, and Ferrer-Mallol et al. [31] demonstrate feasibility across multiple tasks without performing classification. This thesis provides, for the first time, quantified LOO-CV performance estimates paired with explainable feature attribution (SHAP) for both overground walking and the Gowers manoeuvre from a single-camera pipeline operating on publicly sourced video.

The practical value of this contribution is twofold. First, it shows that walking carries feasibility-level discriminative signal. Second, it shows that SHAP-based feature attribution can make the predictions interpretable in this setting.

6.5 Limitations

6.5.1 Dataset

The dataset is sourced entirely from publicly available online video without clinical verification of diagnosis, age, or functional stage. Diagnostic labelling relies on self-report or community annotation in the source video metadata. Exact ages were not available for the majority of subjects; the 4–8 year range spans substantial motor development in typically developing children, introducing heterogeneity in the control group that may inflate apparent group separation. Sample sizes are the primary constraint on both tasks. The walking imbalance of 10 DMD vs 63 Control reflects the availability of suitable public video rather than a deliberate design choice; while balanced class weighting partially addresses this imbalance at the classifier level, the small positive class ($n = 10$) means that individual misclassified DMD clips have a disproportionate effect on sensitivity. The Gowers manoeuvre subset is more severely constrained: only six typically developing control clips were retained after quality filtering, making any classifier trained on this task exploratory by design and precluding reliable generalisation estimates. Quality-control filtering excluded clips with poor knee-tracking confidence, which may introduce a selection bias toward subjects and recording conditions where pose detection is more reliable, potentially underrepresenting children whose movement patterns or body configurations produce less stable landmark tracking.

6.5.2 Recording Heterogeneity

Source videos exhibit variable camera angles, distances, framerates, and scene contexts. Several features are sensitive to these recording conditions. Pelvic rotation ROM depends on an estimate of the walking direction from body-centred world coordinates, which is unreliable for non-front-to-back camera angles. Stride variability features showed reversed directions relative to clinical expectations, plausibly reflecting differences in clip content rather than disease-related movement. These recording confounds are irreducible when using heterogeneous online video and motivate the development of standardised home-recording protocols.

6.5.3 Pose Estimation

Landmark accuracy in children. No published study has validated MediaPipe Pose Landmarker Heavy accuracy in children aged 4–8 years. This study did not perform independent benchmarking against ground-truth motion capture in the study population; landmark error in this paediatric age range and task context therefore remains unquantified. Available validation evidence derives from adult populations: adult data (Stenum et al. [32]: hip mean absolute error (MAE) $\approx 4°$, knee MAE $\approx 5.6°$) cannot be assumed to transfer to this population given the body proportion mismatch and training data bias documented by Wishaupt et al. [11].

Floor-level activities. Floor-level activities represent a systematic failure mode for top-down pose estimators, as the reduced visible body area and non-standard subject orientation degrade both detection and landmark localisation [63]. The GHUM body model (a statistical 3D human body shape model used internally by MediaPipe to generate world-space landmark estimates) fitting that generates world landmarks is likely less reliable for the floor-level postures of the Gowers manoeuvre, where the visible body area is reduced and subject orientation is non-standard.

Camera angle and perspective error. Variable camera angles across the video source introduce perspective error (as documented in infant pose-estimation studies [37]); this analysis constrained features to landmark configurations with acceptable planar validity under mixed viewing angles.

Feature-level geometric limitations. The ankle joint angle computed from the knee-ankle-heel landmark chain is a geometric proxy that does not correspond to the standard clinical definition and should be interpreted as a relative indicator rather than an anatomically precise measurement. MediaPipe applies internal temporal smoothing across frames, which may attenuate the peak amplitudes of very rapid movements; this could slightly reduce measured range-of-motion values for features that depend on peak joint excursions in the DMD group. The hand-support proximity feature is expressed as a 2D image-space wrist-to-knee distance, which avoids unreliable monocular depth estimates but has no established validated reference range; it should be interpreted as a relative indicator rather than an absolute clinical measure.

6.5.4 Stride Segmentation

Strides were segmented at maximum inter-leg spread rather than at heel contact, introducing phase error into all phase-gated features. Heel-contact detection via ankle anterior-posterior position relative to the torso following Stenum et al. [32] was evaluated but found unreliable across the heterogeneous camera angles of the dataset, where body-centred MediaPipe world coordinates do not preserve a stable ground-plane reference. Cadence, double support fraction, and stance time fraction are the features most affected by this limitation.

6.5.5 Evaluation

LOO-CV at the video level prevents within-subject data leakage and is the most conservative evaluation strategy available. However, with $n = 73$ for walking, bootstrap confidence intervals remain wide, and the point estimate of ROC-AUC = 0.825 should not be interpreted as a stable population-level performance figure. The classifier was not evaluated on an independent held-out test set, as dataset size precluded a train/validation/test split. Generalisation to clinically verified, controlled-recording DMD video remains to be demonstrated.

6.6 Directions for Future Work

Clinically verified dataset. The most impactful improvement would be a clinically verified dataset with confirmed diagnoses, exact ages, and standardised recording protocols. A larger clinically verified dataset, for example 40–50 clips per task with controlled smartphone placement, would substantially stabilise LOO-CV estimates and enable developmental normalisation across the 4–8 year age range. For both tasks, targeted collection via a standardised home-video protocol, in which the child performs each task once in front of a smartphone camera, would produce clips suitable for the existing feature pipeline with minimal modification.

Heel-strike detection. Implementing reliable heel-strike detection would correct the phase error in all gait-cycle-dependent features. The ankle anterior-posterior relative to torso method of Stenum et al. [32] is the recommended approach for controlled front-to-back recordings and would enable proper step rate and double-support estimation.

Sequence models. Per-stride feature aggregation discards temporal structure across the walking bout. A shallow sequence model (1D convolutional neural network (1D-CNN) or bidirectional long short-term memory network (BiLSTM)) operating on the pose-derived movement time series directly would retain this structure and could be compared against the RF baseline under the same LOO-CV protocol. The walking timeseries parquet produced by this pipeline is ready for this extension.

Unsupervised phenotyping and similarity clustering. Rather than imposing a binary DMD vs. typically developing classification, an unsupervised approach could cluster subjects by their feature similarity, revealing whether video-derived movement patterns naturally group by disease severity, functional stage, or gait compensation pattern. This would support movement-presentation grouping of walking-stage DMD heterogeneity and could guide future monitoring research without relying on diagnostic labels.

Longitudinal monitoring. The pipeline as implemented produces a binary classification per clip. Extending it to track individual feature trajectories over repeated recordings — for example, monthly stride length symmetry index or Trendelenburg variability — would align more directly with the future use case of monitoring within-patient progression rather than cross-sectional group separation.

7

Conclusion

This thesis shows that interpretable pose-derived movement features extracted from single-camera RGB video can capture feasibility-level separation between ambulatory children with Duchenne muscular dystrophy and typically developing peers. For walking, the best-performing model achieved ROC-AUC = 0.825 and PR-AUC = 0.511 under strict leave-one-out cross-validation (baseline PR-AUC = 0.137; ΔPR-AUC = +0.374). The absence of a walking-speed confound in the univariate results ($p = 0.910$) is consistent with the classifier capturing DMD-related walking differences rather than pace alone. The strongest signals are consistent with movement patterns reported in DMD literature, especially restricted inter-leg spread, reduced hip flexion, pelvic instability, and altered gait rhythm.

The Gowers task contributes mainly at the level of movement-level plausibility rather than robust classification performance. Its leading features (right knee flexion, torso uprightness, and hand-leg distance) match the expected movement pattern of lower-limb loading, trunk elevation, and hand support, but reliable generalisation will require larger and better-balanced control cohorts. Taken together, these results provide a patient-level cross-validated feasibility baseline for two DMD-relevant movement tasks from single-camera consumer video. Two threats remain unresolved at this stage: the sex confound introduced by X-linked inheritance means that some walking-pattern separation may reflect sex rather than disease, and the absence of an independent held-out test set means the reported AUC figures are cross-validated point estimates rather than stable population-level performance figures. Because the pipeline operates without specialist equipment or physical contact, it offers a practical starting point for scalable, home-compatible movement-analysis pipelines. The next requirement for clinical translation is a clinically verified and standardised dataset.

Appendix A

Threats to Validity and Supplementary Results

A.1 Threats to Validity

This section summarises the main threats to validity, organised by the standard categories: internal, external, construct, and conclusion validity [64]. Each subsection points to the corresponding detailed treatment in Section 6.5.

A.1.1 Internal Validity

Internal validity concerns whether the observed classification behaviour is driven by the intended pose-derived movement constructs rather than confounds. Key risks include recording-context effects, pose estimation failure under occlusion and atypical paediatric postures, and stride phase error from the adopted segmentation strategy; these are discussed in detail in Section 6.5 (Recording Heterogeneity, Pose Estimation, and Stride Segmentation). A demographic confound is also present: DMD subjects are necessarily all male due to X-linked inheritance, whereas control subjects are mixed sex; the classifier may therefore partially capture sex-related walking-pattern differences rather than DMD-specific movement differences exclusively. No technical mitigation is available for the sex confound, which therefore remains an unresolved threat to internal validity. Data-leakage risk is addressed by video-level LOO-CV; reversed-direction features are reported explicitly in Section 6.2.

A.1.2 External Validity

External validity concerns generalisation beyond the sampled videos. The dataset is sourced from heterogeneous public online recordings without clinical verification of diagnosis, age, or functional stage, and with substantial variability in camera geometry and scene conditions (Section 6.5, Dataset). The results are therefore best read as a feasibility baseline rather than an estimate of clinical or real-world performance.

A.1.3 Construct Validity

Construct validity concerns whether the operationalised features measure the intended movement constructs. Several extracted quantities (e.g., ankle angles, step width proxies, and hand–leg distances) are geometric proxies that may not correspond exactly to standard movement-analysis definitions, particularly under the MediaPipe body-centred coordinate system and unknown camera viewpoints (Section 6.5, Pose Estimation). Interpretation is anchored to movement patterns reported in DMD literature; approximate features are distinguished from literature-based features in Appendix A.2.

A.1.4 Conclusion Validity

Conclusion validity concerns the reliability of the statistical conclusions. Sample size is the dominant threat, particularly for the Gowers task where the control support is too small for stable LOO-CV estimates (Section 6.5, Evaluation). Class imbalance and the absence of an independent held-out test set further limit the strength of inference. Conclusions are framed conservatively throughout, and the Gowers results remain provisional.

A.2 Detailed Feature Source References

A.3 Univariate Feature Statistics (Low-Power Tasks)

The following table presents univariate Mann–Whitney statistics for the Gowers task, reported for completeness and transparency. Due to limited sample size, these results are presented as descriptive characterisation and are not interpreted as definitive inferential claims (see Section 5).

Table A.1: Detailed source references for the feature groups summarised in Table 4.4. "Literature-based" features derive from published movement-analysis references. "Pose-derived approximation" features adapt clinical or gait concepts to single-camera landmark data.

Feature	Source	Reference
Joint flexion angles (knee, hip, elbow)	Literature-based	Winter (2009)
Trunk forward lean + side lean	Literature-based	Sutherland (1981); Vandekerckhove (2022)
Pelvic obliquity	Literature-based	Sutherland (1981)
Pelvic rotation ROM	Literature-based	Vandekerckhove (2022)
Segment angles relative to vertical (thigh, shank)	Literature-based	Eng & Winter (1995)
Double support fraction	Literature-based	Perry & Burnfield (2010)
Normalised stride length	Literature-based	Perry & Burnfield (2010)
Stride Frequency (FFT)	Literature-based	Perry & Burnfield (2010)
Stride duration CV	Literature-based	Hausdorff et al. (1997)
Stride regularity (autocorrelation)	Literature-based	Menz et al. (2003)
Robinson Symmetry Index	Literature-based	Robinson et al. (1987)
Hip–knee phase alignment	Literature-based	Hamill et al. (1999)
Trunk extension percentile range	Literature-based	Chang (2011)
Hip drop proxy	Pose-derived approximation	Sutherland (1981) adapted
Estimated walking speed ($f_{\text{stride}} \times \ell_{\text{stride}}$)	Pose-derived approximation	composite measure
Step width	Pose-derived approximation	pose-based proxy
Stance frame fraction	Pose-derived approximation	frame counting
Ankle–knee–heel angle	Pose-derived approximation	geometric proxy
Hip-to-heel vertical distance	Pose-derived approximation	pose-based proxy
Hip rise velocity	Pose-derived approximation	gradient measure
Minimum hand–leg distance (image plane)	Pose-derived approximation	2D image-space proxy
Rise duration	Pose-derived approximation	temporal measure
Rise phase fractions	Pose-derived approximation	quantile-partitioned
Toe-walking fraction	Pose-derived approximation	composite heuristic

A.3.1 Gowers Univariate Statistics

A.4 Low-Power Classification Visualisations

The following figures present classification plots for the Gowers task, retained for completeness and transparency despite insufficient sample size and unstable generalisation estimates. These results are not interpreted further (see Section 5.4).

A.4.1 Gowers Visualisations

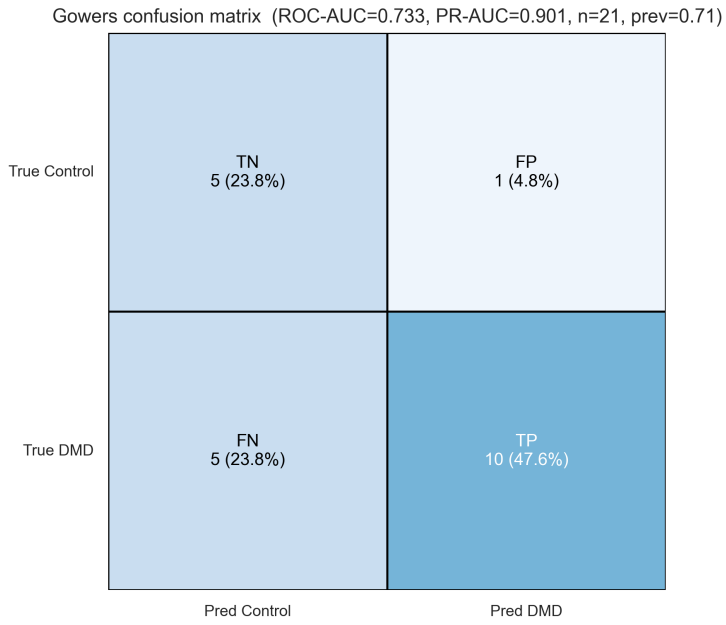

Figure A.1: Confusion matrix for Gowers classification under LOO-CV. Numbers indicate video-level predictions in this low-power feasibility setting.

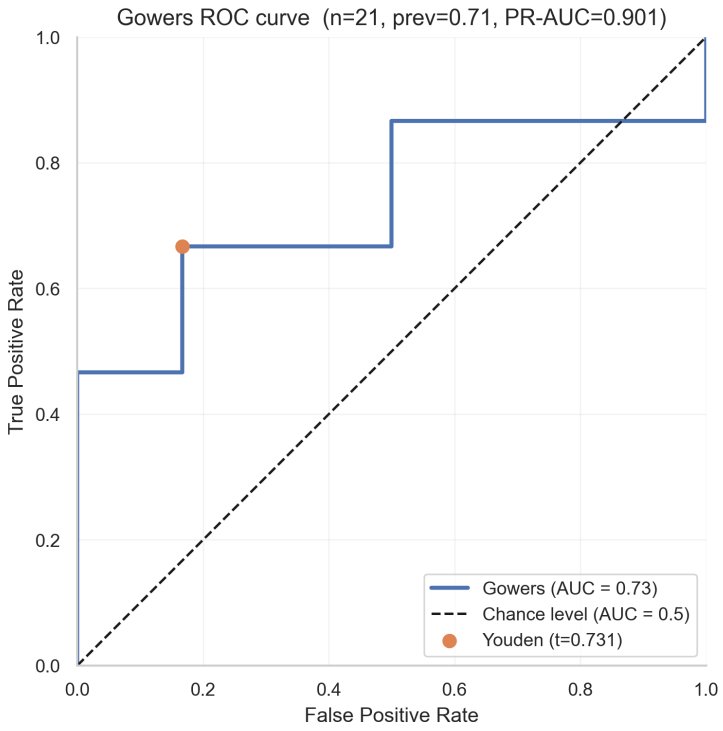

Figure A.2: ROC curve for Gowers classification in the low-power feasibility setting. Dashed line indicates random classifier baseline.

Table A.2: Top 15 Gowers features by absolute rank-biserial effect size at clip level. Limited sample size; reported as descriptive characterisation only.

Feature	r	p
Torso backward-bend variability (p05–p95)	0.643	0.010
Minimum hand–leg distance (image plane)	-0.643	0.010
Right hip flexion angle maximum	-0.786	0.018
Hip-to-heel vertical distance minimum	0.619	0.013
Right knee flexion angle minimum	-0.619	0.013
Right hip flexion angle minimum	-0.571	0.101
Left hip flexion variability (p05–p95)	0.571	0.156
Left hip flexion angle minimum	-0.476	0.244
Left knee flexion angle minimum	-0.476	0.244
Minimum hand–leg distance (body frame)	-0.467	0.185
Left knee flexion variability (p05–p95)	0.429	0.300
Right elbow flexion angle minimum	-0.417	0.262
Hip height variability	-0.250	0.071
Rise duration (s)	0.333	0.342
Left knee flexion angle maximum	-0.190	0.676

References

[1] E. Mercuri, M. Pane, G. Cicala, C. Brogna, and E. Ciafaloni, "Detecting early signs in Duchenne muscular dystrophy: Comprehensive review and diagnostic implications," *Frontiers in Pediatrics*, vol. 11, p. 1 276 144, Nov. 10, 2023, ISSN: 2296-2360. DOI: 10.3389/fped.2023.1276144. PMID: 38027286.

[2] P. Van Dommelen, O. Van Dijk, J. A. De Wilde, and P. H. Verkerk, "Early developmental milestones in Duchenne muscular dystrophy," *Developmental Medicine & Child Neurology*, vol. 62, no. 10, pp. 1198–1204, Oct. 2020, ISSN: 0012-1622, 1469-8749. DOI: 10.1111/dmcn.14623.

[3] L. Wade, L. Needham, P. McGuigan, and J. Bilzon, "Applications and limitations of current markerless motion capture methods for clinical gait biomechanics," *PeerJ*, vol. 10, e12995, Feb. 25, 2022, ISSN: 2167-8359. DOI: 10.7717/peerj.12995.

[4] J. Yang and K. Park, "Improving Gait Analysis Techniques with Markerless Pose Estimation Based on Smartphone Location," *Bioengineering*, vol. 11, no. 2, p. 141, Jan. 30, 2024, ISSN: 2306-5354. DOI: 10.3390/bioengineering11020141.

[5] B. Horsak, A. Eichmann, K. Lauer, *et al.*, "Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait," *Journal of Biomechanics*, vol. 159, p. 111 801, Oct. 2023, ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2023.111801.

[6] J. K. Mah, L. Korngut, J. Dykeman, L. Day, T. Pringsheim, and N. Jette, "A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy," *Neuromuscular Disorders*, vol. 24, no. 6, pp. 482–491, Jun. 2014, ISSN: 09608966. DOI: 10.1016/j.nmd.2014.03.008.

[7] J. Broomfield, M. Hill, M. Guglieri, M. Crowther, and K. Abrams, "Life Expectancy in Duchenne Muscular Dystrophy: Reproduced Individual Patient Data Meta-analysis," *Neurology*, vol. 97, no. 23, Dec. 7, 2021, ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.0000000000012910.

[8] P. van Dommelen, O. van Dijk, J. A. de Wilde, and P. H. Verkerk, "Short developmental milestone risk assessment tool to identify Duchenne muscular dystrophy in primary care," *Orphanet Journal of Rare Diseases*, vol. 19, p. 192, May 10, 2024, ISSN: 1750-1172. DOI: 10.1186/s13023-024-03208-8. PMID: 38730494.

[9] F. Roggio, B. Trovato, M. Sortino, and G. Musumeci, "A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review," *Heliyon*, vol. 10, no. 21, e39977, Nov. 2024, ISSN: 24058440. DOI: 10.1016/j.heliyon.2024.e39977.

[10] J. T. Anderson, J. Stenum, R. T. Roemmich, and R. B. Wilson, "Validation of markerless video-based gait analysis using pose estimation in toddlers with and without neurodevelopmental disorders," *Frontiers in Digital Health*, vol. 7, Feb. 25, 2025, ISSN: 2673-253X. DOI: 10.3389/fdgth.2025.1542012.

[11] K. Wishaupt, W. Schallig, M. H. Van Dorst, A. I. Buizer, and M. M. Van Der Krogt, "The applicability of markerless motion capture for clinical gait analysis in children with cerebral palsy," *Scientific Reports*, vol. 14, no. 1, p. 11 910, May 24, 2024, ISSN: 2045-2322. DOI: 10.1038/s41598-024-62119-7.

[12] Ł. Kidziński, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Communications*, vol. 11, no. 1, p. 4054, Aug. 13, 2020, ISSN: 2041-1723. DOI: 10.1038/s41467-020-17807-z.

[13] J. Lin, Y. Wang, J. Sha, *et al.*, "Clinical reliability and validity of a video-based markerless gait evaluation method," *Frontiers in Pediatrics*, vol. 11, p. 1 331 176, Dec. 22, 2023, ISSN: 2296-2360. DOI: 10.3389/fped.2023.1331176.

[14] L. E. Molteni and G. Andreoni, "Comparing the Accuracy of Markerless Motion Analysis and Optoelectronic System for Measuring Gait Kinematics of Lower Limb," *Bioengineering*, vol. 12, no. 4, p. 424, Apr. 16, 2025, ISSN: 2306-5354. DOI: 10.3390/bioengineering12040424.

[15] D. Menychtas, N. Petrou, I. Kansizoglou, *et al.*, "Gait analysis comparison between manual marking, 2D pose estimation algorithms, and 3D marker-based system," *Frontiers in Rehabilitation Sciences*, vol. 4, p. 1 238 134, Sep. 6, 2023, ISSN: 2673-6861. DOI: 10.3389/fresc.2023.1238134.

[16] E. P. Washabaugh, T. A. Shanmugam, R. Ranganathan, and C. Krishnan, "Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics," *Gait & Posture*, vol. 97, pp. 188–195, Sep. 2022, ISSN: 09666362. DOI: 10.1016/j.gaitpost.2022.08.008.

[17] H. Yoon, E. Jo, S. Ryu, J.-I. Yoo, M. Kim, and J. H. Kim, "Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders," *Scientific Reports*, vol. 15, no. 1, p. 42 074, Nov. 26, 2025, ISSN: 2045-2322. DOI: 10.1038/s41598-025-26169-9.

[18] S. Liu, A. Wong, S. Chen, P. J. Antonelli, and D. L. Guarín, "VisionMD-Gait: Scalable clinical gait assessment from smartphone videos," *Scientific Reports*, vol. 16, no. 1, p. 4711, Jan. 7, 2026, ISSN: 2045-2322. DOI: 10.1038/s41598-025-34912-5.

[19] D. H. Sutherland, R. Olshen, L. Cooper, *et al.*, "The pathomechanics of gait in Duchenne muscular dystrophy," *Developmental Medicine and Child Neurology*, vol. 23, no. 1, pp. 3–22, Feb. 1981, ISSN: 0012-1622. DOI: 10.1111/j.1469-8749.1981.tb08442.x. PMID: 7202868.

[20] L. Doglio, E. Pavan, I. Pernigotti, P. Petralia, C. Frigo, and C. Minetti, "Early signs of gait deviation in Duchenne muscular dystrophy," *European Journal of Physical and Rehabilitation Medicine*, vol. 47, no. 4, pp. 587–594, Dec. 2011, ISSN: 1973-9095. PMID: 21912365.

[21] N. Gaudreault, D. Gravel, S. Nadeau, S. Houde, and D. Gagnon, "Gait patterns comparison of children with Duchenne muscular dystrophy to those of control subjects considering the effect of gait velocity," *Gait & Posture*, vol. 32, no. 3, pp. 342–347, Jul. 2010, ISSN: 09666362. DOI: 10.1016/j.gaitpost.2010.06.003.

[22] K. Heberer, E. Fowler, L. Staudt, *et al.*, "Hip kinetics during gait are clinically meaningful outcomes in young boys with Duchenne muscular dystrophy," *Gait & Posture*, vol. 48, pp. 159–164, Jul. 2016, ISSN: 09666362. DOI: 10.1016/j.gaitpost.2016.05.013.

[23] J. Ropars, M. Lempereur, C. Vuillerot, *et al.*, "Muscle Activation during Gait in Children with Duchenne Muscular Dystrophy," *PLOS ONE*, vol. 11, no. 9, D. Fraidenraich, Ed., e0161938, Sep. 13, 2016, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0161938.

[24] I. Vandekerckhove, M. Van den Hauwe, N. De Beukelaer, *et al.*, "Longitudinal Alterations in Gait Features in Growing Children With Duchenne Muscular Dystrophy," *Frontiers in Human Neuroscience*, vol. 16, Jun. 2, 2022, ISSN: 1662-5161. DOI: 10.3389/fnhum.2022.861136.

[25] I. Vandekerckhove, E. Papageorgiou, B. Hanssen, *et al.*, "Gait classification for growing children with Duchenne muscular dystrophy," *Scientific Reports*, vol. 14, no. 1, p. 10 828, May 11, 2024, ISSN: 2045-2322. DOI: 10.1038/s41598-024-61231-y. PMID: 38734731.

[26] S. Sienko Thomas, C. E. Buckon, A. Nicorici, A. Bagley, C. M. McDonald, and M. D. Sussman, "Classification of the Gait Patterns of Boys With Duchenne Muscular Dystrophy and Their Relationship to Function," *Journal of Child Neurology*, vol. 25, no. 9, pp. 1103–1109, Sep. 2010, ISSN: 0883-0738, 1708-8283. DOI: 10.1177/0883073810371002.

[27] R. Chang and S. Mubarak, "Pathomechanics of Gowers' Sign: A Video Analysis of a Spectrum of Gowers' Maneuvers," *Clinical orthopaedics and related research*, vol. 470, pp. 1987–91, Dec. 28, 2011. DOI: 10.1007/s11999-011-2210-6.

[28] J. Mendell, R. Moxley, R. Griggs, *et al.*, "Randomized, Double-Blind Six-Month Trial of Prednisone in Duchenne's Muscular Dystrophy," *New England Journal of Medicine*, vol. 320, no. 24, pp. 1592–1597, Jun. 15, 1989, ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJM198906153202405.

[29] E. S. Mazzone, G. Coratti, M. P. Sormani, *et al.*, "Timed Rise from Floor as a Predictor of Disease Progression in Duchenne Muscular Dystrophy: An Observational Study," *PLOS ONE*, vol. 11, no. 3, M. Sampaolesi, Ed., e0151445, Mar. 16, 2016, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0151445.

[30] C. M. Mcdonald, E. K. Henricson, R. T. Abresch, *et al.*, "THE 6-minute walk test and other endpoints in Duchenne muscular dystrophy: Longitudinal natural history observations over 48 weeks from a multicenter study," *Muscle & Nerve*, vol. 48, no. 3, pp. 343–356, Jun. 26, 2013, ISSN: 0148-639X. DOI: 10.1002/mus.23902.

[31] E. Ferrer-Mallol, C. Matthews, M. Stoodley, *et al.*, "Patient-led development of digital endpoints and the use of computer vision analysis in assessment of motor function in rare diseases," *Frontiers in Pharmacology*, vol. 13, p. 916 714, Sep. 12, 2022, ISSN: 1663-9812. DOI: 10.3389/fphar.2022.916714.

[32] J. Stenum, C. Rossi, and R. T. Roemmich, "Two-dimensional video-based analysis of human gait using pose estimation," *PLOS Computational Biology*, vol. 17, no. 4, D. Schneidman-Duhovny, Ed., e1008935, Apr. 23, 2021, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008935.

[33] C. Lugaresi, J. Tang, H. Nash, *et al.* "MediaPipe: A Framework for Building Perception Pipelines." (2019), [Online]. Available: https://arxiv.org/abs/1906.08172 (visited on 03/06/2026), pre-published.

[34] I. Grishchenko, V. Bazarevsky, A. Zanfir, *et al.*, *BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation*, arXiv, Jun. 2022. DOI: 10.48550/arXiv.2206.11678.

[35] M. M. Ali, M. Medhat Hassan, and M. Zaki, "Human Pose Estimation for Clinical Analysis of Gait Pathologies," *Bioinformatics and Biology Insights*, vol. 18, p. 11 779 322 241 231 108, Sep. 2024, ISSN: 1177-9322, 1177-9322. DOI: 10.1177/11779322241231108.

[36] F. Gama, M. Mísař, L. Navara, S. T. Popescu, and M. Hoffmann, "Automatic infant 2D pose estimation from videos: Comparing seven deep neural network methods," *Behavior Research Methods*, vol. 57, no. 10, p. 280, Sep. 10, 2025, ISSN: 1554-3528. DOI: 10.3758/s13428-025-02816-x.

[37] L. Jahn, S. Flügge, D. Zhang, *et al.*, "Comparison of marker-less 2D image-based methods for infant pose estimation," *Scientific Reports*, vol. 15, no. 1, p. 12 148, Apr. 9, 2025, ISSN: 2045-2322. DOI: 10.1038/s41598-025-96206-0.

[38] M. G. Contesse, A. T. L. Sapp, S. D. Apkon, L. P. Lowes, L. Dalle Pazze, and M. G. Leffler, "Reliability and construct validity of the Duchenne Video Assessment," *Muscle & Nerve*, vol. 64, no. 2, pp. 180–189, Aug. 2021, ISSN: 0148-639X, 1097-4598. DOI: 10.1002/mus.27335.

[39] L. Grinsztajn, E. Oyallon, and G. Varoquaux. "Why do tree-based models still outperform deep learning on tabular data?" (2022), [Online]. Available: https://arxiv.org/abs/2207.08815 (visited on 02/26/2026), pre-published.

[40] A. Taleb, P. Rambaud, S. Diop, *et al.*, "Spinal Muscular Atrophy Hypotonia Detection Using Computer Vision and Artificial Intelligence," *JAMA Pediatrics*, vol. 178, no. 5, p. 500, May 1, 2024, ISSN: 2168-6203. DOI: 10.1001/jamapediatrics.2024.0030.

[41] P. Ruth, S. Uhlrich, C. de Monts, *et al.*, *Dataset Kinematics and Timed Function Tests of Facioscapulohumeral Muscular Dystrophy and Myotonic Dystrophy*, Zenodo, Sep. 2024. DOI: 10.5281/ZENODO.13788592.

[42] S. Lundberg and S.-I. Lee. "A Unified Approach to Interpreting Model Predictions." arXiv: 1705.07874 [cs]. (Nov. 25, 2017), [Online]. Available: http://arxiv.org/abs/1705.07874 (visited on 04/12/2026), pre-published.

[43] S. M. Lundberg, G. G. Erion, and S.-I. Lee. "Consistent Individualized Feature Attribution for Tree Ensembles." arXiv: 1802.03888. (Feb. 12, 2018), [Online]. Available: https://arxiv.org/abs/1802.03888, pre-published.

[44] G. Jocher and J. Qiu, *Ultralytics YOLO11*, version 11.0.0, 2024. DOI: 10.5281/zenodo.13941922.

[45] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, G. Brock, Ed., e0118432, Mar. 4, 2015, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0118432.

[46] "Pose landmark detection guide | Google AI Edge," Google AI for Developers. (), [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker (visited on 03/06/2026).

[47] G. Wu, S. Siegler, P. Allard, *et al.*, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: Ankle, hip, and spine," *Journal of Biomechanics*, vol. 35, no. 4, pp. 543–548, Apr. 2002, ISSN: 00219290. DOI: 10.1016/S0021-9290(01)00222-6.

[48] D. A. Winter, *Biomechanics and Motor Control of Human Movement*. Wiley, Sep. 17, 2009. DOI: 10.1002/9780470549148.

[49] K. Rácz and R. M. Kiss, "Marker displacement data filtering in gait analysis: A technical note," *Biomedical Signal Processing and Control*, vol. 70, p. 102 974, Sep. 2021, ISSN: 1746-8094. DOI: 10.1016/j.bspc.2021.102974.

[50] S. Wang, J. Pitts, R. Purohit, and H. Shah, "The Influence of Motion Data Low-Pass Filtering Methods in Machine-Learning Models," *Applied Sciences*, vol. 15, no. 4, p. 2177, Feb. 18, 2025, ISSN: 2076-3417. DOI: 10.3390/app15042177.

[51] J. Hamill, R. E. A. Van Emmerik, B. C. Heiderscheit, and L. Li. "A dynamical systems approach to lower extremity running injuries." (1999), [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/10521606/.

[52] P. F. Lamb and M. Stöckl, "On the use of continuous relative phase: Review of current approaches and outline for a new standard," *Clinical Biomechanics*, vol. 29, no. 5, pp. 484–493, May 2014, ISSN: 0268-0033. DOI: 10.1016/j.clinbiomech.2014.03.008.

[53] R. O. Robinson, W. Herzog, and B. M. Nigg, "Use of force platform variables to quantify the effects of chiropractic manipulation on gait symmetry," *Journal of Manipulative and Physiological Therapeutics*, vol. 10, no. 4, pp. 172–176, Aug. 1987, ISSN: 0161-4754. PMID: 2958572.

[54] J. Perry and J. M. Burnfield, *Gait Analysis: Normal and Pathological Function*, 2nd ed. Thorofare, NJ: SLACK, 2010, 551 pp., ISBN: 978-1-55642-766-4.

[55] J. M. Hausdorff, H. K. Edelberg, S. L. Mitchell, A. L. Goldberger, and J. Y. Wei, "Increased gait unsteadiness in community-dwelling elderly fallers," *Archives of Physical Medicine and Rehabilitation*, vol. 78, no. 3, pp. 278–283, Mar. 1997, ISSN: 00039993. DOI: 10.1016/S0003-9993(97)90034-4.

[56] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, "Acceleration patterns of the head and pelvis when walking on level and irregular surfaces," *Gait & Posture*, vol. 18, no. 1, pp. 35–46, Aug. 2003, ISSN: 09666362. DOI: 10.1016/S0966-6362(02)00159-5.

[57] E. K. Henricson and A. A. Ramli, "Harnessing Fast Fourier Transform for Rapid Community Travel Distance and Step Estimation in Children with Duchenne Muscular Dystrophy," *Sensors*, vol. 25, no. 10, p. 3234, May 21, 2025, ISSN: 1424-8220. DOI: 10.3390/s25103234.

[58] M. H. Schwartz and A. Rozumalski, "The gait deviation index: A new comprehensive index of gait pathology," *Gait & Posture*, vol. 28, no. 3, pp. 351–357, Oct. 2008, ISSN: 0966-6362. DOI: 10.1016/j.gaitpost.2008.05.001.

[59] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, May 2022, ISSN: 1566-2535. DOI: 10.1016/j.inffus.2021.11.011.

[60] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics*, vol. 44, no. 3, p. 837, Sep. 1988, ISSN: 0006341X. DOI: 10.2307/2531595. JSTOR: 2531595.

[61] D. Janzing, L. Minorics, and P. Blöbaum. "Feature relevance quantification in explainable AI: A causal problem." arXiv: 1910.13413. (Oct. 29, 2019), [Online]. Available: https://arxiv.org/abs/1910.13413, pre-published.

[62] C. M. McDonald, "Timed function tests have withstood the test of time as clinically meaningful and responsive endpoints in duchenne muscular dystrophy," *Muscle & Nerve*, vol. 58, no. 5, pp. 614–617, Nov. 2018, ISSN: 0148-639X, 1097-4598. DOI: 10.1002/mus.26334.

[63] S. Dill, A. Rösch, M. Rohr, *et al.*, "Accuracy Evaluation of 3D Pose Estimation with MediaPipe Pose for Physical Exercises," *Current Directions in Biomedical Engineering*, vol. 9, no. 1, pp. 563–566, Sep. 1, 2023, ISSN: 2364-5504. DOI: 10.1515/cdbme-2023-1141.

[64] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer Berlin Heidelberg, 2012. DOI: 10.1007/978-3-642-29044-2.