Video-based biomechanical analysis captures disease-specific movement signatures of different neuromuscular diseases

Parker S. Ruth[1]*, Scott D. Uhlrich[2]*, Constance de Monts[3], Antoine Falisse[4], Julie Muccini[5], Sydney Covitz[4], Shelby Vogt-Domke[3], John Day[3], Tina Duong[3†], Scott L. Delp[4,6,7†]

*Contributed equally
†Contributed equally
[1]Department of Computer Science, Stanford University,
[2]Department of Mechanical Engineering, University of Utah,
[3]Department of Neurology, Stanford Medicine
[4]Department of Bioengineering, Stanford University
[5]Department of Radiology, Stanford Medicine
[6]Department of Mechanical Engineering, Stanford University
[7]Department of Orthopaedic Surgery, Stanford Medicine

ABSTRACT

Assessing human movement is essential for diagnosing and monitoring movement-related conditions. Timed function tests (TFTs) are among the most widespread assessments due to their speed and simplicity. However, TFTs cannot capture disease-specific movement patterns. Recent advances in smartphone video-based biomechanical analysis allow detailed movement quantification with the ease and speed required for clinical settings. To compare video-based analysis against TFTs, we collected data from 129 individuals: 28 with facioscapulohumeral muscular dystrophy, 58 with myotonic dystrophy, and 43 with typical movement. Video-based movement recordings took a median time of 16 minutes per participant. In addition to reproducing TFTs ($r > 0.98$), video metrics can outperform TFTs at disease classification ($p < 0.001$). Video-based analysis also identifies differences in gait kinematics and arm range of motion not evident in TFTs. We conclude that video-based biomechanical analysis can complement existing functional movement assessments and capture more sensitive, disease-specific outcomes from human movement.

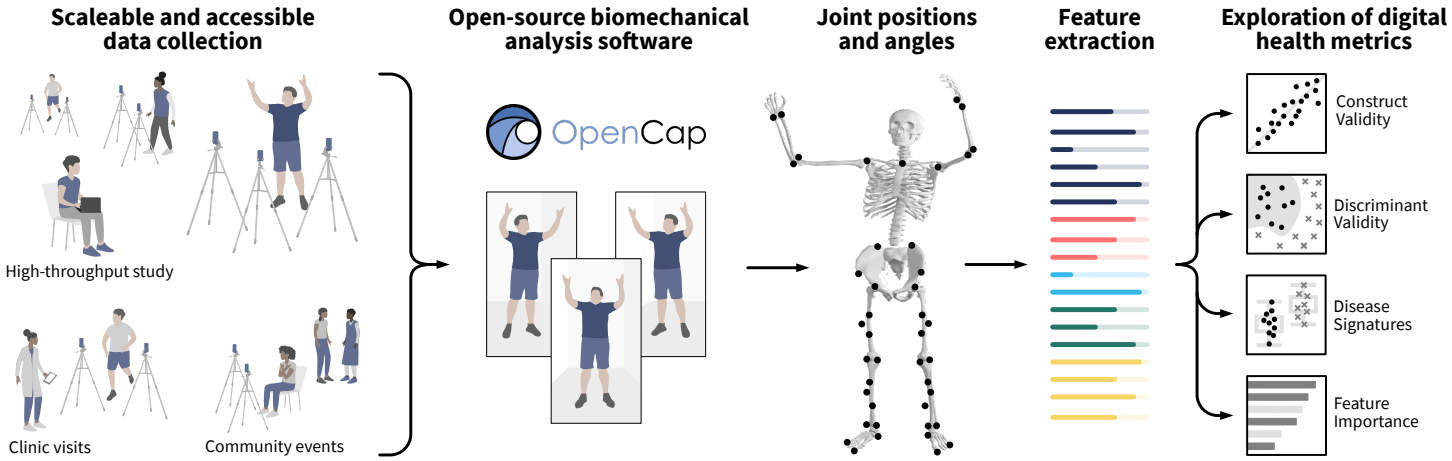

Fig. 1: Video-based biomechanical analysis. Video-based biomechanical analysis enables scalable and portable data collection for studying human movement. We used the open-source OpenCap application[1] to convert smartphone videos to three-dimensional anatomical marker positions and joint angles. Leveraging clinical domain knowledge, we engineered a set of video-based metrics describing aspects of human motion known to have salience for the neuromuscular diseases we study. These features are then used to explore the design of novel digital health metrics.

INTRODUCTION

Quantitative 1 measurements of human movement can improve the diagnosis, monitoring, and understanding of movement-related disorders. The state of the art for measuring human movement involves marker-based motion capture and force plates in specialized laboratories, which can cost hundreds of thousands of dollars. Collecting motion data typically requires hours of expert labor.[2] While these laboratory-based methods are the gold standard in human movement analysis,[3] practical applications in real-world clinical settings require more speed and convenience.

Timed function tests (TFTs), which assess gross motor function by the time taken to complete an activity, are widely used in clinical practice and clinical trials due to their speed and simplicity. Despite their ubiquity, TFTs are less informative than laboratory-based analysis tools employed in biomechanics research. Reducing human movement to a single time score overlooks movement quality factors like muscle activation, imbalances, compensatory gait changes, or limited joint range of motion, which may better reflect changes over time. Thus, there is a pragmatic gap between the sophisticated methods used in biomechanics research and those currently employed in clinical research and practice.

With the approval of disease modifying therapies and treatment options for neuromuscular diseases, improved mobility and function measurements are crucial for informed clinical decision-making and disease management.[4,5] The sensitivity and accuracy of the clinical outcome assessments used as clinical endpoints significantly impact trial design and therapy advancement, particularly for rare diseases. For slower-progressing diseases, TFTs may lack the sensitivity to detect subtle changes in muscle function.These coarse metrics may fail to detect both natural disease progression and the effect of therapy on function. For example, since a task like walking comprises a wide range of possible gait patterns, individuals with early-stage disease may retain a normal walking speed (and hence an unchanged TFT time) but display compensations or decreased movement quality due to weakness indicating lower muscle control or activation. There is a need for functional assessments sensitive to the unique impairments that may provide a phenotypical signature of muscle weakness across neuromuscular diseases of varying severity.[6]

Here, we focus on two rare and relatively slow progressing neuromuscular diseases: facioscapulohumeral muscular dystrophy (FSHD) and myotonic dystrophy (DM). Both conditions impact gross motor function (e.g., reduce walking speed), but they affect different muscle groups, leading to biomechanical differences in task performance. FSHD often leads to weakness in the facial, scapulothoracic, trunk and ankle dorsiflexors resulting in impairments in locomotion and overhead reaching.[7,8] DM primarily causes weakness in the lower trunk and distal muscles of the ankle, hand, trunk, knee, and hip, leading to difficulty rising from a chair, squatting, and activities of daily living.[9]

Tools for quantifying human movement from video have the potential to improve sensitivity of measurement by providing richer metrics of movement with minimal added burden.[10] For example, machine learning models can predict clinical scores of Parkinson's severity from videos of gait[11] and sit-to-stand transitions.[12] The speed and ease of video capture can facilitate scalable data collection within the clinic and remotely. For example, at-home videos of sit-to-stand can predict osteoarthritis and health outcomes.[13] Video capture is also rich and flexible enough to tailor metrics for individual diseases. For example, "reachable workspace" is a video-derived metric quantifying upper-body functional volume developed for FSHD.[14] Reachable workspace has been validated against existing upper-limb metrics,[15] reflects longitudinal trends in progression,[16] correlates with patient reported outcomes, and is used as clinical outcome assessments in clinical trials.[17]

We previously developed OpenCap, an open-source software tool that automatically estimates the kinematics of human movement (i.e., joint positions and angles) from smartphone videos.[1] Since no on-body sensors or markers are used, setup time is typically under three minutes, and data can be collected as quickly as the participant can perform the activity. By eliminating the need for time-consuming analysis in specialized motion capture labs, this approach significantly expands the scale and accessibility of obtaining high-quality movement data in diverse settings. This technology is a promising complement to TFTs. For example, beyond merely measuring walking speed, OpenCap can estimate more in-depth quantities such as ankle dorsiflexion angle, indicating steppage gait or hip abduction angle, indicating circumduction gait.[18] Nonetheless, the potential of video-based biomechanical analysis to overcome the limitations of TFTs has remained untested until now.

The purpose of our study was to compare video-based biomechanical analysis (**Fig. 1**) to TFTs for quantifying human movement among individuals with FSHD and DM. Specifically, we investigated the following research questions: (1) Can video-based metrics accurately reproduce clinician-graded TFTs? (2) Can video-based metrics capture additional disease-specific movement patterns? To answer these questions, we collected a dataset of simultaneous OpenCap video recordings of clinician-graded TFTs, extracted a set of video-based movement features, and explored the potential for video-based digital assessments of human movement.

RESULTS

Study participants and data collection

The dataset includes 129 individuals (**Table 2**), including 28 with FSHD, 58 with DM (49 DM1, 9 DM2), and 43 without a diagnosed neuromuscular condition (i.e., typical movement group). Among these participants, 31 repeated the data collection more than once. Data collection was conducted in three settings. 98 individuals participated in a high-throughput data collection, primarily conducted in a university gymnasium using multiple parallel OpenCap motion capture stations. 29 participants were recruited on-site at community events such as local support group meetings and national conferences. Five participants were recruited during regular clinic visits. Three individuals participated at multiple locations. More detailed participant demographics are available in Supplementary Table 1. Participants performed a set of nine activities (10-meter run, 10-meter walk, timed up-and-go, five-time sit-to-stand, arm abduction, shoulder range of motion, elbow flexion, calf raise, and jump), the first four of which are used in existing TFTs (**Table 1**). Since TFTs do not assess upper-body movement, we also measured the Brooke Upper Extremity Scale, which ordi-

Activity	Description	Clinical Assessment
Run	As fast as able, run or walk 10 meters from standing cold start	10m walk/run test (TFT)
Walk	At self-selected comfortable pace, walk 10 meters from standing cold start	10m walk test (TFT)
Timed up-and-go	As fast as able, stand up, walk 3 meters, turn around a cone, walk back, and sit down	Timed up-and-go (TFT)
5-time sit-to-stand	As fast as able, rise from chair and sit down again five times with arms crossed over chest	5-time sit-to-stand (TFT)
Arm abduction	Abduct arms overhead with straight elbows	Brooke Upper Extremity Scale
Arm RoM	Raise arms with straight elbows as high as able in four directions (0º flexion, 45º scaption, 90º abduction, 180º extension); horizontally sweep straight arms at waist and eye levels	N/A
Elbow flexion	Flex elbows as far as able, starting from full extension	N/A
Calf raise	Rise to toes for 30s or as long as able	N/A
Jump	Jump as high as able; if unable, rise from a squat as fast as able	N/A

Table 1: Physical activities. (RoM = range of motion, TFT = timed function test)

	Total	FSHD	DM	Typical
Participants	129	28	58	43
Location				
High-throughput study	98	26	29	43
Community events	29	2	27	2
Neuromuscular clinic visits	5	1	4	0
Sex				
Female (%)	57 (44%)	4 (14%)	31 (53%)	22 (51%)
Age in years				
Mean (SD)	43 (16)	48 (16)	40 (14)	42 (19)
BMI				
Mean (SD)	25.2 (5.5)	26.2 (4.9)	25.4 (6.5)	24.3 (4.2)

Table 2: Study participant characteristics. (FSHD = facioscapulohumeral muscular dystrophy, DM = myotonic dystrophy)

nally scores functional ability of overhead arm abduction, hand-to-mouth movement, and distal hand function.[19,20] For each activity, the OpenCap video-based biomechanical analysis system was used to compute three-dimensional anatomical marker positions and skeletal joint angles from inverse kinematics. We leveraged expertise of clinicians on our team and previously-reported movement patterns in DM and FSHD to compute a set of 34 interpretable, activity-specific features from the kinematic data (Supplementary Table 3).

Video-based reproduction of TFTs

We tested the ability of video metrics to reproduce the clinician-graded TFT times. Automated video metrics can accurately reproduce TFT times with Pearson r > 0.98 for all four TFT activities (**Fig. 2**). To contextualize these results, we compare our mean absolute error (MAE) against the reported values of distribution-based minimum clinically important difference (MCID) for each metric among reported neuromuscular MCIDs. Literature values for the MCID in individuals with DM are available for 10-meter walk and timed up-and-go tasks. In both cases, our MAE is lower than the MCID (10-meter walk test: MCID = 0.69s, MAE = 0.30s; timed up-and-go: MCID = 0.92s, MAE = 0.22s).[22] For the 10-meter run test, our MAE of 0.25s falls under the literature-reported MCID for Duchenne muscular dystrophy (2.3s).[23] MCID for the 5-time sit-to-stand activity in neuromuscular disease are not available in existing literature.

To measure test-retest reliability, 13 participants with DM repeated the study protocol on two sequential days. The video-based predictions have similar test-retest reliability as TFT times and the Brooke score. (**Fig. 3**). Other video features exhibit a range of reliability. 70% of the video metrics have moderate reliability or better (Supplementary Fig. 1).

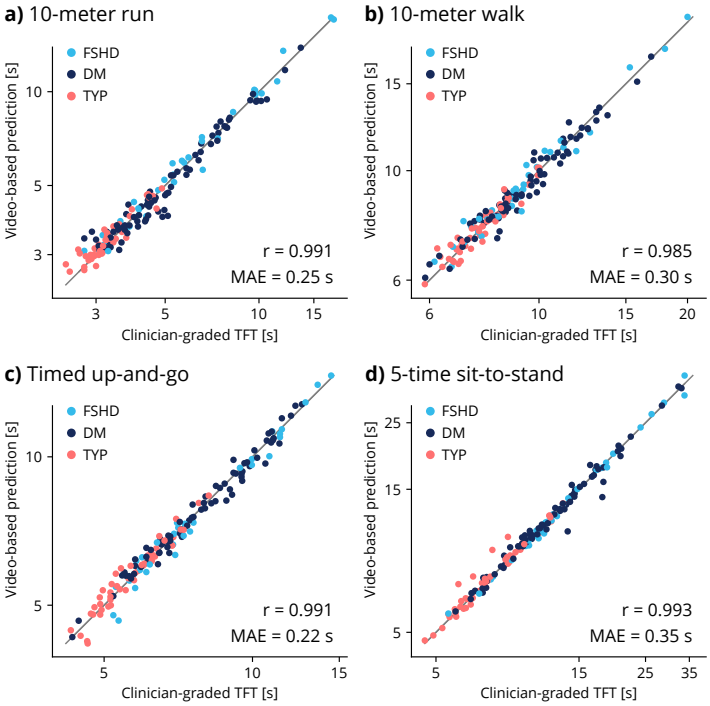

Fig. 2: Reproduction of TFTs from video. To estimate clinician-graded timed function tests (TFT) from video-based biomechanical data, we computed time-based features (e.g., walking speed for the ten-meter walk test) and trained a linear regressor to predict the TFT time based on the video-based features to account for systematic differences between the scores. Here, we show the predicted TFT times using a linear regressor fit with leave-one-subject-out cross validation for **a)** 10-meter run, **b)** 10-meter walk, **c)** timed up-and-go (TUG), and **d)** 5-time sit-to-stand (5xSTS) activities. (r = Pearson correlation, MAE = mean absolute error,
FSHD = facioscapulohumeral muscular dystrophy, DM = myotonic dystrophy, TYP = typical movement)

Disease classification

To evaluate whether video-based biomechanical metrics or TFT times can better differentiate between movement patterns in FSHD and DM, we compared the performance of binary disease classification models trained on either video metrics or TFT times (**Fig. 4**). A binary classification model trained on video movement metrics derived from all nine activities performed better (balanced accuracy = 86%) than one trained on TFT times (balanced accuracy = 50%, p = 0.0001) and one trained on TFT times and Brooke score (balanced accuracy = 74%, p = 0.049). To investigate the potential for features from a single activity to discriminate between diseases, we trained classification models using only video metrics or TFT times from each movement in the dataset. Video models

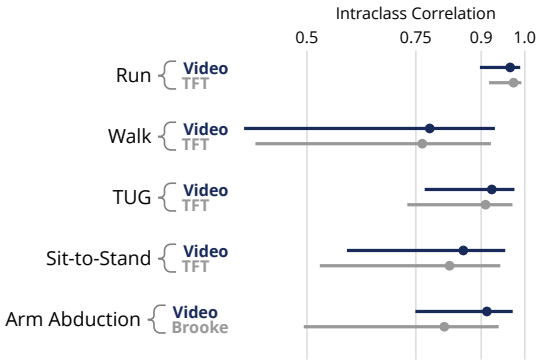

Fig. 3: Test-retest reliability. Video-based metrics (run speed, walk speed, timed up-and-go [TUG] time, sit-to-stand time, and arm abduction shoulder angle) have test-retest reliability similar to clinical tests (timed function tests [TFTs] and the Brooke Upper Extremity Scale). Bars indicate 95% confidence intervals. ICC > 0.5 indicates "moderate" reliability, ICC > 0.75 indicates "good" reliability, and ICC > 0.9 indicates "excellent" reliability.[21]

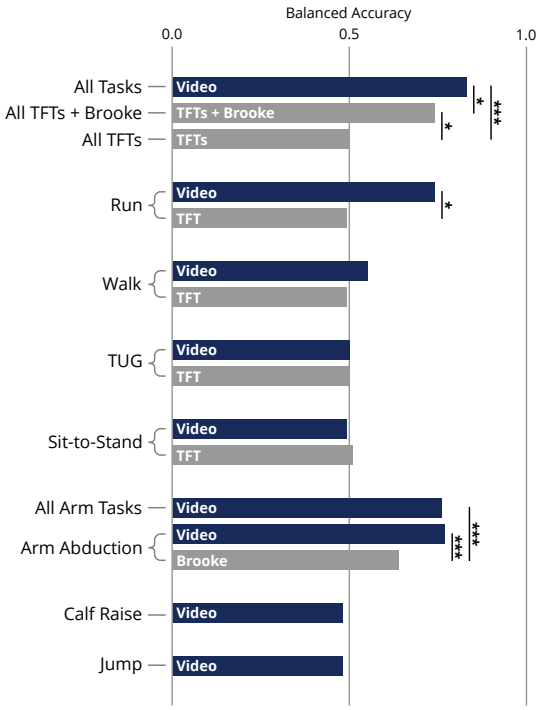

Fig. 4: Disease classification performance. Binary classification models were trained to distinguish between participants with facioscapulohumeral muscular dystrophy (FSHD) and myotonic dystrophy (DM). A model using video-based biomechanical metrics from nine activities is more accurate than models using timed function test tests (TFTs) and the Brooke Upper Extremity Scale. Video models using running and arm range of motion tasks also outperformed clinical scores. Classification model performance is measured by balanced accuracy, which corrects for unequal class sizes; 0.5 corresponds to always predicting the more common disease. Statistical differences in performance are tested using the McNemar test (*$p < 0.05$, ***$p < 0.001$). (TUG = timed up-and-go)

perform as well or better than TFT models across all activities. The video model for running performed better than the TFT model for running ($p = 0.029$). This is because video-based biomechanical analysis captures disease-specific movement qualities that TFT times do not capture. For example, video-based biomechanical analysis detected differences in stride length ($p = 0.009$) and stride time ($p = 0.011$) between FSHD and DM participants, while running speed (analogous to the TFT time) was not significantly different between groups (Supplementary Fig. 2). The clinician-graded Brooke score is outperformed by both the video model using upper-limb metrics ($p = 0.0003$) and the video model using only arm abduction metrics ($p = 0.00006$).

Disease-specific movement signatures

To better understand why the video models outperformed TFT models, we investigated group differences in video-based biomechanical metrics. For example, compared to the DM group, the FSHD group had shorter stride lengths ($p = 0.003$) and higher ankle heights ($p = 0.037$) during the swing phase of walking (**Fig. 5**). This is consistent with the presentation of foot drop in FSHD due to weakness in muscles that dorsiflex the ankle.[7] There was no significant difference in walking speed between FSHD and DM, consistent with the inability of the walking TFT model to discriminate between diseases (**Fig. 4**). Additional disease differences emerged in running gait stride length and stride time (Supplementary Fig. 2). FSHD affects the proximal arm muscles, limiting upper-limb mobility, while these muscles are less affected in DM. We found significant between-group differences in both Brooke scores and video-based biomechanical metrics of upper-limb tasks (Supplementary Fig. 2). Ninety participants received a Brooke score of 6, indicating an ability to abduct hands overhead without elbow flexion; 23 received a score of 5 (able to raise hands overhead with some elbow flexion), and 21 received a score of 4 (unable to raise hands overhead). Our video metric comprising a composite of shoulder and elbow angles (**Fig. 6**) correlates with clinician-graded Brook scores (Spearman's rank correlation $r = 0.795$). Despite this correlation, there is high variance in the video metric among individuals with the same Brooke score, demonstrating the relative insensitivity of ordinal scores like Brooke. Additionally, the video metric detected differences between individuals with different diseases who received a Brooke score of 4 ($p = 0.027$) and among those who received a Brooke score of 5 ($p = 0.039$), showing how the video metric detects disease-specific features of movement quality (**Fig. 6**). Together, these results highlight the strength of a continuous video-based metric to characterize upper-limb function compared to a subjective ordinal score.

Disease-specific feature importance

To identify the video metrics most salient to individual diseases, we computed the relative importance of each video-based biomechanical metric in models that classify individuals with each neuromuscular disease from those with typical movement. The most important features for detecting FSHD were upper-limb activities representing shoulder and elbow range of motion (**Fig. 7**). By contrast, the most important features for detecting DM are from lower-limb activities such as running, walking, and jumping. These observations are consistent with the known early clinical presentations of upper arm and shoulder girdle weakness in FSHD and distal lower-extremity weakness in DM.

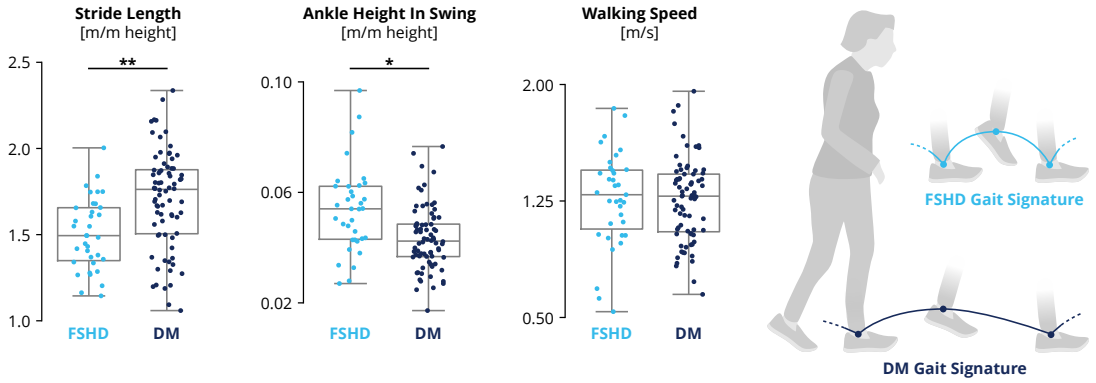

Fig. 5: Disease-specific walking gait signatures. Patterns in walking gait differ between individuals with facioscapulohumeral muscular dystrophy (FSHD) and those with myotonic dystrophy (DM). Despite similar walking speeds, the FSHD group walked with shorter stride lengths (p = 0.003) and higher ankle height during swing (p = 0.037) than did the DM group. Statistical significance is measured with a two-sided Kolmogorov–Smirnov test. ("m/m height" indicates units of meters normalized by participant height in meters, i.e. dimensionless)

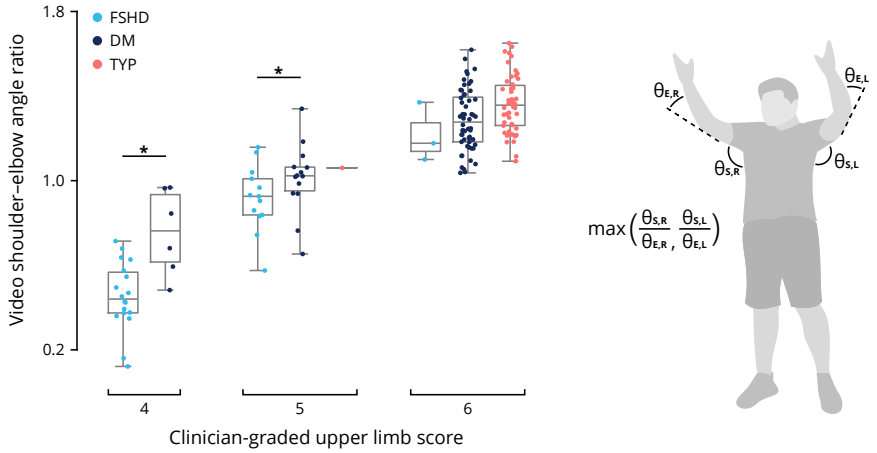

Fig. 6: Clinical and video-based upper-limb metrics. Video-based biomechanical analysis of upper-body function agrees with the clinician-rated Brooke Upper Extremity Scale, which ordinally scores the ability to raise arms overhead with straight elbows. To compute our analogous video-based metric, we compute the ratio of shoulder abduction angle (θ_S) to elbow flexion angle (θ_E) and take the maximum ratio between right and left sides. The mean values of the video-based metric are ordered in correspondence with the Brooke score. Furthermore, the video metric reveals differences between diseases among participants with the same Brooke score. Statistical significance is measured with a one-sided Kolmogorov–Smirnov test. (FSHD = facioscapulohumeral muscular dystrophy, DM = myotonic dystrophy, TYP = typical movement)

DISCUSSION

Video-based biomechanical analysis can reproduce clinician-graded TFT assessments of mobility and function while also encoding more disease-specific information about human movement. Since we can accurately predict TFT scores from video-based metrics (**Fig. 2**), it is unsurprising that video features have similar test-retest reliability (**Fig. 3**) and perform no worse than TFTs at disease classification (**Fig. 4**). Unlike TFTs, video-based analysis can also differentiate the movement patterns of FSHD versus those of DM (**Fig. 4**). TFT times alone cannot classify these diseases because TFTs are integrative assessments of function agnostic to the mechanisms of impairment. This agnosticism benefits TFTs as general-purpose assessments of movement health but limits their utility for more sensitive assessment of disease-specific progression and loss of function.

Like TFTs, video-based analysis can extract general time-based assessments (e.g., walking speed), while also being sensitive to disease-specific motor impairments caused by dysfunction in specific muscle groups. Unlike TFTs, for running gait, video-based analysis captures enough information to distinguish between diseases while the TFT times alone cannot (**Fig. 4**). For walking gait, video-based analysis reveals differences in gait parameters between the two diseases (**Fig. 5**). This implies that video metrics capture additional salient information about disease signatures that TFT times do not. Furthermore, video-based metrics can distinguish differences between upper-limb range of motion in FSHD and DM, even among individuals who receive the same clinician-graded Brooke Upper Extremity Scale (**Fig. 6**). This suggests that video-based biomechanical analysis may enable more sensitive upper-limb functional assessments.

One reason models using video metrics outperforms TFT times for discriminating between FSHD and DM is that video features include additional measures of upper-limb mobility such as shoulder and elbow range of motion. Analysis of video-based feature importance reveals that, unlike DM, the FSHD group differed from the typical movement group mainly in upper limb movement (**Fig. 7**). This finding is consistent with the clinical presentation of these diseases – the deltoids, which abduct the shoulder, are among the

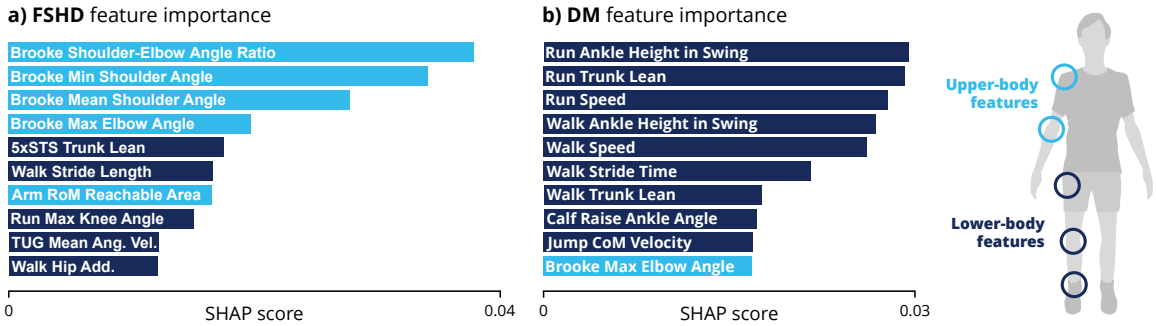

Fig. 7: Disease-specific feature importance. Binary classification models were trained to distinguish participants with typical movement versus those with a) facioscapulohumeral dystrophy (FSHD) or b) myotonic dystrophy (DM). Shapley additive explanations (SHAP) values indicate the contribution of each video-based biomechanical metric toward predictive accuracy. The top ten features are shown here; all feature importance scores are in Supplementary Fig. 3. The classification models are support vector machines with radial basis function kernels. The FSHD model relies mainly on upper-limb features, while the DM model relies mainly on lower-limb features. Descriptions of video-based features are in Supplementary Table 3. (5xSTS = 5-time sit-to-stand, RoM = range of motion, TUG = timed up-and-go, CoM = center of mass)

most affected muscles in FSHD[8] whereas they are less affected in DM[8]. However, video-based analysis of the Brooke task outperforms the Brooke score alone, and even when Brooke scores are combined with TFTs, video-based analysis still achieves significantly better classification accuracy (**Fig. 4**). Thus, the advantage of video-based analysis goes beyond the inclusion of upper-limb activities.

The video-based biomechanical metrics extracted and analyzed here are not comprehensive. A different set of features may demonstrate different per-task or all-task performance for disease classification. For example, our features extracted from the timed up-and-go features exclude gait parameters to avoid redundancy with the walk and run tasks. An alternative to this manual feature creation process based on pathophysiological, clinical, and biomechanical domain-knowledge is to derive data-driven features using supervised or unsupervised dimensionality reduction.[24] It is possible that such methods may lead to improvement in performance but at the expense of clinical interpretability.

The ultimate goal is to create digital metrics for use as clinical outcome measures that are more sensitive and reliable then traditional TFTs. Biomechanical analysis shows promise for achieving greater sensitivity than existing clinical outcome metrics.[25] Some of the video-based biomechanical metrics have low reliability scores among the test-retest participants. With the present dataset, we cannot determine what proportion of this variance is attributable to day-to-day variability in the biomechanics, error in the video-based analysis pipeline, or human error in the case of clinician-graded TFT times. Greater variance in an outcome measure decreases the sensitivity and increases the sample size required for clinical trials. Unwanted variance could be reduced by altering physical task instructions to control for confounding factors, engineering different movement metrics, or averaging repeated measurements.

The advantages of video-based biomechanical analysis must be balanced with the additional burden of data collection and analysis. In this study, the median time required to record all nine activities using OpenCap was 16 minutes, and kinematics data were automatically computed on OpenCap's cloud servers. For comparison, acquiring the same data using typical motion capture laboratory methods would take three or more hours including calibration, equipment setup, marker placement, activity recording, and data post-processing. OpenCap achieves near-lab-quality kinematics with much lower burden.[1] Since OpenCap requires only ordinary smartphones and tripods to operate, it enables data collection in clinics and at larger community events such as conferences. This is especially significant for studying rare diseases for which a single study location may have a limited reachable recruitment pool. For example, our cohort of 58 individuals with DM is nearly four times larger than the largest existing study of 3D kinematics in DM.[26] By collecting data in five locations, including national conferences, we enrolled participants from 14 different states across the US. Compared with traditional motion capture equipment, OpenCap is inexpensive to scale. In our high-throughput data collection, we operated six simultaneous recording stations. Taken together, the speed, portability, and scalability of OpenCap enabled us to acquire a participant group that is seven times larger (n=129) than the average laboratory-based biomechanics study (n=18) and the largest studies of biomechanics in FSHD or DM.[27,28]

It is important to acknowledge the limitations of this study. First, the video-based metrics salient to our classification tasks may not generalize to other movement-related diseases or mobility levels outside this study's inclusion criteria. However, our findings imply that video-based biomechanical analysis generally provides more information than traditional do functional outcomes, like TFTs, and has the potential to improve the creation of movement biomarkers and functional outcome metrics for other conditions. Second, while the video-based biomechanical analysis was much faster, less expensive, and easier than traditional lab-based methods, the multi-camera setup we used here complicates at-home measurement and continuous monitoring of function. Inertial measurement units could better capture long-term patterns in passive movement and behavior,[29–32] but we believe video is better suited for periodic, rapid assessments of whole-body biomechanics. Biomechanical analysis from a single video will likely be possible in the near future,[33–35] which would open the possibility of at-home measurements of movement. Finally, we acknowledge that the present analysis is a purely cross-sectional assessment of construct and discriminant validity. Further development of digital biomarkers and functional outcome measures should determine

repeatability, correlation with functional and patient-reported outcomes, minimum clinically important differences, and sensitivity to natural progression and therapeutic interventions.

In this work, we show that video-based biomechanical analysis can reproduce existing clinician-scored functional assessments. However, unlike functional assessments, video analysis captures movement signatures that distinguish between different neuromuscular diseases. This study was enabled by the recent advancements in collecting near-lab-quality biomechanics data at scale using smartphones. Our study results and the accessibility of video-based biomechanical analysis holds promise for translational impact in the study and treatment of movement-related conditions. Specifically, video-based biomechanical analysis could enable digital health solutions for assessing and monitoring motor function, supporting diagnosis of rare movement-related conditions, and complementing traditional clinical outcome measures to enhance care, management, and clinical trial design.

METHODS

Participant screening and recruitment

The inclusion criteria for all study participants were the ability to independently rise from a chair and ambulate without assistance. Participants were eligible to enroll in the typical movement group if they had no neuromuscular condition, movement disorder, or injury affecting movement in the last 6 months. Participants with movement disorders were eligible if they had confirmed diagnosis of myotonic dystrophy (DM1 or DM2) or facioscapulohumeral muscular dystrophy (FSHD). Participants in the high-throughput study were recruited from a database of patients who consented to be contacted about studies in the Stanford Medicine Neuromuscular Research Clinic. At local and remote community events, participants were reached through flyers and on-site verbal recruitment. All participants provided informed consent to a protocol approved by the Stanford University Institutional Review Board.

Selection of physical activities

In consultation with the clinical domain experts on the study team, we selected a set of physical activities that captured both isolated muscle groups and whole-body function. We selected the activities listed in **Table 1**. We chose run, walk, timed up-and-go, and sit-to-stand because they are common TFTs. In addition to these lower-body functional tasks, we added a 30-second calf raise (i.e., ankle plantarflexion), designed to test balance and plantarflexor strength, and a jump task, designed to test maximal force of lower-extremity extensor muscles. Finally, we added three upper-body activities: shoulder abduction, elbow flexion, and arm range-of-motion. Among these, the arm abduction task is used in the existing ordinal Brooke Upper Extremity Scale. Together, all nine activities represent a wide spectrum of human movement accessible to video-based biomechanical analysis. Task descriptions and scripts are available in the Supplement.

Domain expert-informed feature engineering

To design a set of video-based biomechanical metrics, our interdisciplinary clinical and engineering team examined representative videos and skeletal kinematics from the dataset. Clinicians visually identified phenotypes and compensations, and engineers designed algorithms to quantify the clinical observations. The full list of features is included in Supplementary Table 3. All video-based metrics are continuous variables computed automatically from marker positions and joint kinematics. Some participants were unable to complete the ankle-based activities (calf raise and jump), these missing feature values were filled with zeros, since inability to perform an activity corresponds to a known encoding of the feature variables.

Scalable video-based human movement assessments

To record human movement scalably and efficiently, we used OpenCap, a free and open-source software tool enabling smartphone video-based biomechanical analysis.[1] We selected the OpenCap settings that use HRNet pose estimation algorithm[36] and a deep learning model to augment the sparse 3D keypoints into a richer anatomical marker set.[37] Within OpenCap, joint angles are computed using the OpenSim[3] Inverse Kinematics tool and a biomechanically constrained skeletal model.[38] All videos were recorded with at least 60 frames per second. The capability to record videos at 120 frames per second was developed mid-way into our data collection, so some videos of gait activities were collected at 120 frames per second. Trained evaluators measured timed function tests using stopwatches simultaneously with OpenCap recordings of the activities. Dimensions and layout of the camera setup are available in the supplementary materials.

Analysis of video-based reproduction of TFTs

Time scores analogous to TFTs can be extracted from videos by measuring the duration of time the participant takes to complete a task. However, video-derived times do not exactly match clinician-graded TFT times. In the case of gait trials, this is because the full ten meters for the TFT are not all visible in the camera field of view. In the case of sit-to-stand and timed up-and-go trials, discrepancies can arise because clinicians look for the moment a person makes contact with the chair seat, while our video algorithm is based on the center of mass elevation. Furthermore, all TFTs are subject to human error in correctly measuring timing of heel strikes or chair contact. However, to the extent that these discrepancies are systematic, they can be corrected by a linear adjustment. We measured the Pearson correlation and mean absolute error between TFT times and video-based predictions using univariate linear regression and leave-one-subject-out cross validation.

To measure test-retest reliability of video-based biomechanical analysis and TFTs, 13 individuals with DM repeated the data collection on two sequential days. Reliability was measured using intraclass correlation (ICC) with a two-way mixed effects model for absolute agreement of single measurements.[21]

Analysis of disease classification

To study whether video-based biomechanical metrics can detect disease-specific movement patterns that TFTs do not, we compared the performance of binary classification models trained on either video-based metrics or TFT times. We used a support vector classifier with a nonlinear kernel (radial basis function) because some movement features were hypothesized to be nonlinearly related to the labels. For example, abnormal gait may manifest as low hip

flexion due to shortened strides or high hip flexion due to steppage gait. The nonlinear kernel permits decision boundaries that accommodate the nonlinearities of such features. Input features were standardized (zero mean, unit standard deviation) prior to model fitting.

The classification models were trained with leave-one-subject-out cross validation. The support vector classifier includes a regularization parameter, which is selected in each training fold with an inner cross validation over an array of parameter values. Model performance was quantified by balanced accuracy of the predictions. This metric represents the percent of correct predictions, correcting for imbalance in class sizes. We used the McNemar test of proportions to compare the performance of different models.[39] This non-parametric test considers, in the cases where the predictions of two models disagree, whether one is statistically more likely to be correct.

Analysis of disease-specific gait signatures

To compare the distributions of each of the 34 video metrics between participants with FSHD versus those with DM, we performed two-sided Kolmogorov-Smirnov tests with Bonferroni corrections for multiple hypothesis testing. To further examine the sensitivity of video-based biomechanical metrics to upper-body range of motion impairment in FSHD, we performed one-sided Kolmogorov-Smirnov tests to compare FSHD versus DM within each ordinal category of clinician-assigned Brooke scores.

Analysis of feature importance

To analyze the relative importance of the video-based biomechanical metrics for each disease, we first trained binary classification models to distinguish FSHD from typical movement and to distinguish DM from typical movement. Feature importance values are calculated using Shapley additive explanations (SHAP) scores,[40] which represent the average contribution of each feature towards model predictions. The SHAP scores were computed for models trained on the full dataset without held-out participants. As in the analysis of disease discrimination, the models trained were support vector classifiers with radial basis function kernels.

DATA AVAILABILITY

The datasets generated and analyzed for this study are available in the "Kinematics and timed function tests of facioscapulohumeral muscular dystrophy and myotonic dystrophy" data repository, available at https://doi.org/10.5281/zenodo.13788592.

CODE AVAILABILITY

The code for feature extraction, figure generation, and statistics is available in the following GitHub repository:

https://github.com/stanfordnmbl/opencap-fshd-dm-analysis.

ACKNOWLEDGEMENTS

We thank the 42 volunteers who helped facilitate recruitment, data collection, and data curation, especially Lin Karman, Sarah Ismail, Paxton Ataide, and Carmichael Ong. We are grateful to Trevor Hastie for consultation on statistical methods. Thank you to the Myotonic Dystrophy Foundation, and the FSHD Society for their support of our study. We especially wish to thank the 129 study participants without whom this research would not have been possible. This study was funded by the Wu Tsai Human Performance Alliance and the Mobilize Center under NIH Grant P41EB027060. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

AUTHOR CONTRIBUTIONS

P.S.R., S.D.U., J.D., T.D., and S.L.D. conceptualized and designed the study. S.D.U., C.M., A.F., J.M., and T.D. developed the methods used to perform the experiments. P.S.R., S.D.U., C.M., A.F., J.M., S.C., S.V., and T.D. conducted the experiments. P.S.R. synthesized results and figures. P.S.R. and S.D.U. drafted the manuscript. C.M., A.F., J.M., S.C., S.V. J.D., T.D., and S.L.D. revised the manuscript.

COMPETING INTERESTS

A.F. and S.D.U. are cofounders of Model Health, Inc., which supports the non-academic, commercial use of the open-source software OpenCap used for data collection in this study.

REFERENCES

[1] Uhlrich, S. D. *et al.* OpenCap: Human movement dynamics from smartphone videos. *PLOS Computational Biology* **19**, e1011462 (2023).

[2] Uhlrich, S. D., Uchida, T. K., Lee, M. R. & Delp, S. L. Ten steps to becoming a musculoskeletal simulation expert: A half-century of progress and outlook for the future. *Journal of Biomechanics* **154**, 111623 (2023).

[3] Seth, A. *et al.* OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLOS Computational Biology* **14**, e1006223 (2018).

[4] Pascual-Gilabert, M., López-Castel, A. & Artero, R. Myotonic dystrophy type 1 drug development: A pipeline toward the market. *Drug Discovery Today* **26**, 1765–1772 (2021).

[5] Tihaya, M. S. *et al.* Facioscapulohumeral muscular dystrophy: The road to targeted therapies. *Nature Reviews Neurology* **19**, 91–108 (2023).

[6] Ghasemi, M., Emerson, C. P. & Hayward, L. J. Outcome Measures in Facioscapulohumeral Muscular Dystrophy Clinical Trials. *Cells* **11**, 687 (2022).

[7] Gambelli, C. N. *et al.* The effect of tibialis anterior weakness on foot drop and toe clearance in patients with facioscapulohumeral dystrophy. *Clinical Biomechanics* **102**, 105899 (2023).

[8] Dijkstra, J. N. *et al.* Natural History of Facioscapulohumeral Dystrophy in Children: A 2-Year Follow-up. *Neurology* **97** (2021).

[9] Bouchard, J.-P., Cossette, L., Bassez, G. & Puymirat, J. Natural history of skeletal muscle involvement in myotonic dystrophy type 1: A retrospective study in 204 cases. *Journal of Neurology* **262**, 285–293 (2015).

[10] Stenum, J. *et al.* Applications of Pose Estimation in Human Health and Performance across the Lifespan. *Sensors (Basel, Switzerland)* **21**, 7315 (2021).

[11] Lu, M. *et al.* Quantifying Parkinson's disease motor severity

under uncertainty using MDS-UPDRS videos. *Medical Image Analysis* **73**, 102179 (2021).

[12] Mehta, D. *et al.* Towards Automated and Marker-less Parkinson Disease Assessment: Predicting UPDRS Scores using Sit-stand videos. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3836–3844 (IEEE, Nashville, TN, USA, 2021).

[13] Boswell, M. A. *et al.* Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study. *npj Digital Medicine* **6**, 1–7 (2023).

[14] Kurillo, G., Chen, A., Bajcsy, R. & Han, J. J. Evaluation of upper extremity reachable workspace using Kinect camera. *Technology and Health Care* **21**, 641–656 (2013).

[15] Han, J. J. *et al.* Reachable workspace and performance of upper limb (PUL) in duchenne muscular dystrophy. *Muscle & Nerve* **53**, 545–554 (2016).

[16] Hatch, M. N. *et al.* Longitudinal study of upper extremity reachable workspace in fascioscapulohumeral muscular dystrophy. *Neuromuscular Disorders* **29**, 503–513 (2019).

[17] Hatch, M. N., Kurillo, G., Chan, V. & Han, J. J. Motion sensor-acquired reachable workspace correlates with patient-reported upper extremity activities of daily living (ADL) function in facioscapulohumeral dystrophy. *Muscle & Nerve* **63**, 250–257 (2021).

[18] Horsak, B. *et al.* Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait. *Journal of Biomechanics* **159**, 111801 (2023).

[19] Brooke, M. H. *et al.* Clinical trial in duchenne dystrophy. I. The design of the protocol. *Muscle & Nerve* **4**, 186–197 (1981).

[20] Lue, Y.-J., Lin, R.-F., Chen, S.-S. & Lu, Y.-M. Measurement of the Functional Status of Patients with Different Types of Muscular Dystrophy. *The Kaohsiung Journal of Medical Sciences* **25**, 325–333 (2009).

[21] Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* **15**, 155–163 (2016).

[22] Knak, K. L., Sheikh, A. M., Andersen, H., Witting, N. & Vissing, J. Intrarater reliability and validity of outcome measures in myotonic dystrophy type 1. *Neurology* **94**, e2508–e2520 (2020).

[23] McDonald, C. M. *et al.* The 6-minute walk test and other clinical endpoints in duchenne muscular dystrophy: Reliability, concurrent validity, and minimal clinically important differences from a multicenter study. *Muscle & Nerve* **48**, 357–368 (2013).

[24] Winner, T. S. *et al.* Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. *PLOS Computational Biology* **19**, e1011556 (2023).

[25] Statland, J. M., Karanevich, A., Bruetsch, A. & Huisinga, J. A pilot study of the responsiveness of wireless motion analysis in facioscapulohumeral muscular dystrophy. *Muscle & Nerve* **60**, 590–594 (2019).

[26] Tiffreau, V. *et al.* Gait among Patients with Myotonic Dystrophy Type 1: A Three-Dimensional Motion Analysis Study. *Journal of the International Society of Physical and Rehabilitation Medicine* **1**, 65 (2018-10/2018-12).

[27] Knudson, D. Twenty-Year Trends of Authorship and Sampling in Applied Biomechanics Research. *Perceptual and Motor Skills* **114**, 16–20 (2012).

[28] Knudson, D. Confidence crisis of results in biomechanics research. *Sports Biomechanics* **16**, 425–433 (2017).

[29] Knak, K. L., Sheikh, A. M., Witting, N. & Vissing, J. Physical activity in myotonic dystrophy type 1. *Journal of Neurology* **267**, 1679–1686 (2020).

[30] Zhuparris, A. *et al.* Smartphone and Wearable Sensors for the Estimation of Facioscapulohumeral Muscular Dystrophy Disease Severity: Cross-sectional Study. *JMIR Formative Research* **7**, e41178 (2023).

[31] Ricotti, V. *et al.* Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. *Nature Medicine* **29**, 95–103 (2023).

[32] Servais, L. *et al.* First regulatory qualification of a digital primary endpoint to measure treatment efficacy in DMD. *Nature Medicine* **29**, 2391–2392 (2023).

[33] Dong, Y. *et al.* MD-Vibe: Physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, UbiComp-ISWC '20, 525–531 (Association for Computing Machinery, New York, NY, USA, 2020).

[34] Wei, W.-L., Lin, J.-C., Liu, T.-L. & Liao, H.-Y. M. Capturing Humans in Motion: Temporal-Attentive 3D Human Pose and Shape Estimation from Monocular Video (2022). 10.48550/arXiv.2203.08534.

[35] Shin, S., Kim, J., Halilaj, E. & Black, M. J. WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2070–2080 (2024).

[36] Sun, K., Xiao, B., Liu, D. & Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5686–5696 (2019).

[37] Falisse, A., Uhlrich, S. D., Chaudhari, A. S., Hicks, J. L. & Delp, S. L. Marker Data Enhancement For Markerless Motion Capture (2024). 10.1101/2024.07.13.603382.

[38] Rajagopal, A. *et al.* Full-Body Musculoskeletal Model for Muscle-Driven Simulation of Human Gait. *IEEE Transactions on Biomedical Engineering* **63**, 2068–2079 (2016).

[39] Kim, S. & Lee, W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? *Statistical Methods in Medical Research* **26**, 142–154 (2017).

[40] Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).