Automated, Vision-Based Goniometry and Range of Motion Calculation in Individuals With Suspected Ehlers-Danlos Syndromes/Generalized Hypermobility Spectrum Disorders: A Comparison of Pose-Estimation Libraries to Goniometric Measurements

ANDREA SABO[1], NIMISH MITTAL[1,2,3,4], AMOL DESHPANDE[5], HANCE CLARKE[6,7,8],
AND BABAK TAATI[1,9,10], (Senior Member, IEEE)

[1]KITE Research Institute, Toronto Rehabilitation Institute–University Health Network, Toronto, ON M5G 2A2, Canada
[2]Division of Physical Medicine and Rehabilitation, Temerty Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A1, Canada
[3]Department of Anesthesia and Pain Medicine, University of Toronto, Toronto, ON M5S 1A1, Canada
[4]Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, ON M5S 1A1, Canada
[5]Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A1, Canada
[6]Department of Anesthesiology and Pain Medicine, University of Toronto, Toronto, ON M5S 1A1, Canada
[7]Canada Transitional Pain Service, Toronto General Hospital–University Health Network, Toronto, ON M5T 1V4, Canada
[8]Canada Transitional Pain Service, Toronto General Hospital, Toronto, ON M5G 2C4, Canada
[9]Department of Computer Science, University of Toronto, Toronto, ON M5S 1A1, Canada
[10]Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada
CORRESPONDING AUTHOR: B. TAATI (babak.taati@uhn.ca)

ABSTRACT Generalized joint hypermobility (GJH) often leads clinicians to suspect a diagnosis of Ehlers Danlos Syndrome (EDS), but it can be difficult to objectively assess. Video-based goniometry has been proposed to objectively estimate joint range of motion in hyperextended joints. As part of an exam of joint hypermobility at a specialized EDS clinic, a mobile phone was used to record short videos of 97 adults (89 female, 35.0 ± 9.9 years old) undergoing assessment of the elbows, knees, shoulders, ankles, and fifth fingers. Five body keypoint pose-estimation libraries (AlphaPose, Detectron, MediaPipe-Body, MoveNet – Thunder, OpenPose) and two hand keypoint pose-estimation libraries (AlphaPose, MediaPipe-Hands) were used to geometrically calculate the maximum angle of hyperextension or hyperflexion of each joint. A custom domain-specific model with a MobileNet-v2 backbone finetuned on data collected as part of this study was also evaluated for the fifth finger movement. Spearman's correlation was used to analyze the angles calculated from the tracked joint positions, the angles calculated from manually annotated keypoints, and the angles measured using a goniometer. Moderate correlations between the angles estimated using pose-tracked keypoints and the goniometer measurements were identified for the elbow (rho = .722; Detectron), knee (rho = .608; MoveNet – Thunder), shoulder (rho = .632; MoveNet – Thunder), and fifth finger (rho = .786; custom model) movements. The angles estimated from keypoints predicted by open-source libraries at the ankles were not significantly correlated with the goniometer measurements. Manually annotated angles at the elbows, knees, shoulders, and fifth fingers were moderately to strongly correlated to goniometer measurements but were weakly correlated for the ankles. There was not one pose-estimation library which performed best across all joints, so the library of choice must be selected separately for each joint of interest.

INDEX TERMS Generalized joint hypermobility, human pose-estimation, range of motion estimation, video-based goniometry.

Clinical and Translational Impact Statement— This work evaluates several pose-estimation models as part of a vision-based system for estimating joint angles in individuals with suspected joint hypermobility. Future applications of the proposed system could facilitate objective assessment and screening of individuals referred to specialized EDS clinics.

I. INTRODUCTION

JOINT hypermobility describes the ability of a joint to move beyond the typical range of motion [1]. Joint hypermobility is affected by age, sex, and ethnicity, can be localized to one or several joints, or be widespread throughout the body [1]. When joint hypermobility is present in multiple joints (usually five or more), an individual is said to have generalized joint hypermobility (GJH) [1]. Furthermore, GJH is one of the diagnostic criteria for most types of Ehlers-Danlos Syndromes (EDS), a collection of heritable connective tissue disorders with widespread systematic manifestations [1], [2]. Individuals who do not meet the strict diagnostic criteria for EDS but present with GJH and associated chronic pain are classified as having generalized hypermobility spectrum disorders (G-HSD) [1].

Therefore, the quantification of joint hypermobility is important to facilitate diagnosis of EDS and G-HSD. Currently, joint hypermobility is assessed in the clinic through the Beighton exam [1], [3]. This exam is scored out of 9, with a point assigned for each elbow and knee that can be actively hyperextended past 10°, each fifth finger that can be passively hyperextended past 90°, each thumb that can be touched to the forearm while the elbow is extended, and one point if the palms can be placed on the floor with the knees extended [1], [3]. The Beighton score has been criticized for its inclusion of limited joints and, while not part of the Beighton exam, dorsiflexion of the ankles past 15° and flexion at the shoulders past 180° are commonly reported and assessed to evaluate for hypermobility [4]. Although part of the diagnostic criteria, the Beighton score is also used as a screening and triage tool for specialized EDS clinics to facilitate care of appropriate patients. For example, a specialized EDS clinic may require a Beighton score (as assessed by a primary care physician) above a specified threshold before accepting a patient's referral to the clinic. It is therefore desirable to have an objective and reliable method of screening for joint hypermobility and aid in the determination of which patients should be evaluated in specialized clinics.

The Beighton exam has been assessed for reliability with generally positive inter- and intra-rater agreement [5]. However, most studies are well controlled, with clinicians discussing and standardizing their rating process prior to applying the Beighton scoring system on the study participants. This contrasts with general practice, where primary care physicians lack a standardized methodology for performing the Beighton exam with other primary care physicians. Furthermore, the Beighton exam has only been validated

when a goniometer is used, which is not the standard practice for all referring clinicians. Even when a goniometer is used, the largest inter- and intra-rater disagreement occurs for the elbow, knee, and fifth finger joints, which require measurement of precise angles (unlike the spine flexion or thumb-to-forearm movements) [6]. Therefore, in practice there is inconsistency in the assessment of hypermobility at these joints, and ultimately the referral of patients to specialized EDS clinics.

One way to standardize the estimation of joint angles and the subsequent process of screening for potential joint hypermobility is through an objective, algorithmic approach. Specifically, videos of individuals performing movements for assessing joint hypermobility in primary care or home settings can be analyzed using established computer vision techniques. Videos are unobtrusive and can be easily recorded in most settings using a webcam or mobile phone camera and human pose-estimation libraries that use deep learning models to predict the locations of key joints of the body in regular color videos are widely available [7], [8], [9], [10], [11], [12]. Using tracked joint positions, the maximum angle of flexion or extension can be calculated at each joint of interest and used to screen for potential joint hypermobility.

Related work on video-based goniometry has validated that angles calculated from manual annotations on still video frames are strongly correlated with goniometer measurements [13]. Furthermore, Fan and Gu et al. have shown that angles calculated from the joint locations tracked using the OpenPose pose-estimation library are moderately to strongly correlated to clinician annotations on the same videos for upper body joints [14]. Similarly, strong correlations have been found for movements of the hand when angles calculated from OpenPose joint locations were compared to measurements from a marker-based optical motion capture system [15].

While human pose-estimation libraries are a promising method of extracting meaningful information from video, it is unclear how well these libraries track hypermobile joints. This is because the underlying models are trained on datasets of individuals from the general population performing typical daily or sport activities [16]. It is unlikely that there are many examples of joints in hypermobile positions in the training data, so the generalization to more extreme joint positions is not well-understood.

Therefore, this study is the first to seek to understand whether current open-source human pose-estimation libraries are suitable for estimating the maximum flexion or extension

in joints of interest in a population of adults with suspected EDS/G-HSD. The joints which are assessed for hypermobility using a goniometer in the clinic (ie. the elbows, knees, fifth finger, ankles, and shoulders) will be examined and compared to clinical measurements taken by clinicians at a specialized EDS clinic. Differences between pose-estimation libraries will be explored and conclusions will be drawn with respect to which are best suited for use for each joint in this population.

II. METHODS

A. DATA COLLECTION

The data used for this study was collected as part of a larger observational study investigating the feasibility of vision-based range of motion estimation at the GoodHope EDS clinic at Toronto General Hospital [17]. The study protocol was approved by the Research Ethics Board of the University Health Network on 26 April 2022 (Study ID - 22-5073.2). A complete description of the protocol for this larger study is available [18].

A total of 100 adults with suspected EDS/G-HSD undergoing a physical examination of joint hypermobility at a specialized EDS clinic were recruited to participate in this study. Participants underwent an exam of joint hypermobility twice. In the first repetition, a clinician measured the active flexion of the ankles and shoulders, extension of the elbows, knees, and passive extension of the fifth fingers with a digital goniometer as per standard protocol of Beighton score assessment. The second repetition was recorded using a tripod-mounted mobile phone camera (Motorola Moto G Pure or Motorola E5 Play, 1080 × 1920 pixels, 30 Hz). Short videos were taken of each movement in which only the participant was visible in the frame to facilitate better pose-estimation. All movements were performed according to standard methods [18], [19]. The participants also performed the spine flexion and thumb-to-forearm opposition movements that are part of the Beighton exam, but these are not measured using a goniometer and were thus not analyzed further in this study.

In addition to the goniometer measurements and videos, the participants' age and sex, as well as an encoded identifier of the clinician performing the exam were recorded. The Research Ethics Board of the institute approved the study protocol and all participants provided written consent prior to participation.

1) CONSIDERATIONS FOR COLLECTION OF VIDEOS TO FACILITATE POSE-ESTIMATION

Prior to participant recruitment, preliminary experiments were conducted to inform the manner in which the videos should be recorded. The goal of this process was to select an angle and framing of the participant that captured the full range of motion for each movement while also facilitating high quality pose-tracking.

During this initial investigation, it was identified that the pose-estimation libraries generally struggled to track the keypoints for the ankle flexion movement when the full body was not visible in the video frame. Therefore, during data collection, the videos were framed such that the participants' entire body was visible for the ankle range of motion assessment.

Additionally, for the fifth finger extension movement, it was identified that the hand pose-estimation models failed to track keypoints when the hand was recorded from a strictly profile view where most of the fingers were hidden from view behind the fifth finger. To facilitate viable pose-tracking of this joint, the camera angle was adjusted to point downward, capturing the overall shape of the hand more clearly.

Figure 1 provides a visual summary of the approximate framing of the participants during each movement.

B. CALCULATION OF EXTREME JOINT ANGLES

In addition to the goniometer measurements taken in the clinic, the joint angles of interest were calculated from video using two approaches. The first approach was fully automated and involved the use of human pose-estimation libraries to extract the positions of joints of interest which were then used to calculate the joint angles. As a comparison, and to understand the effect of any errors in pose-estimation, keypoints were manually annotated and also used to calculate the joint angles.

1) AUTOMATED CALCULATION OF JOINT ANGLES USING HUMAN POSE-ESTIMATION

Multiple open-source human pose-estimation libraries were used to extract the positions of joints of interest in the recorded videos. For the movements involving large joints (ankles, shoulders, elbows, and knees) the AlphaPose, Detectron, MediaPipe-Pose, MoveNet – Thunder, and Open-Pose libraries were investigated, while the AlphaPose and MediaPipe-Hands pose-estimation libraries were used to analyze the videos of fifth-finger extension [7], [8], [9], [10], [11], [12]. These libraries were selected for the presence of complete documentation and published test performance, as well as for their ease of use and general support for developers. Furthermore, the selected libraries also represent a range of computational requirements: OpenPose, Alpha-Pose, and Detectron require access to graphics processing units (GPU) to facilitate inference in a reasonable time period and thus cannot be deployed on a mobile phone at a future stage of this project. Conversely, the MediaPipe and MoveNet–Thunder models are much smaller and less computationally intensive, leaving open the opportunity to deploy these models on mobile devices or servers without GPUs in the future.

One 3D body pose-estimation library [20] and two 3D hand pose-estimation libraries [21], [22] were also considered for inclusion in this study, but preliminary experimentation yielded visually poor results, so they were not pursued further.
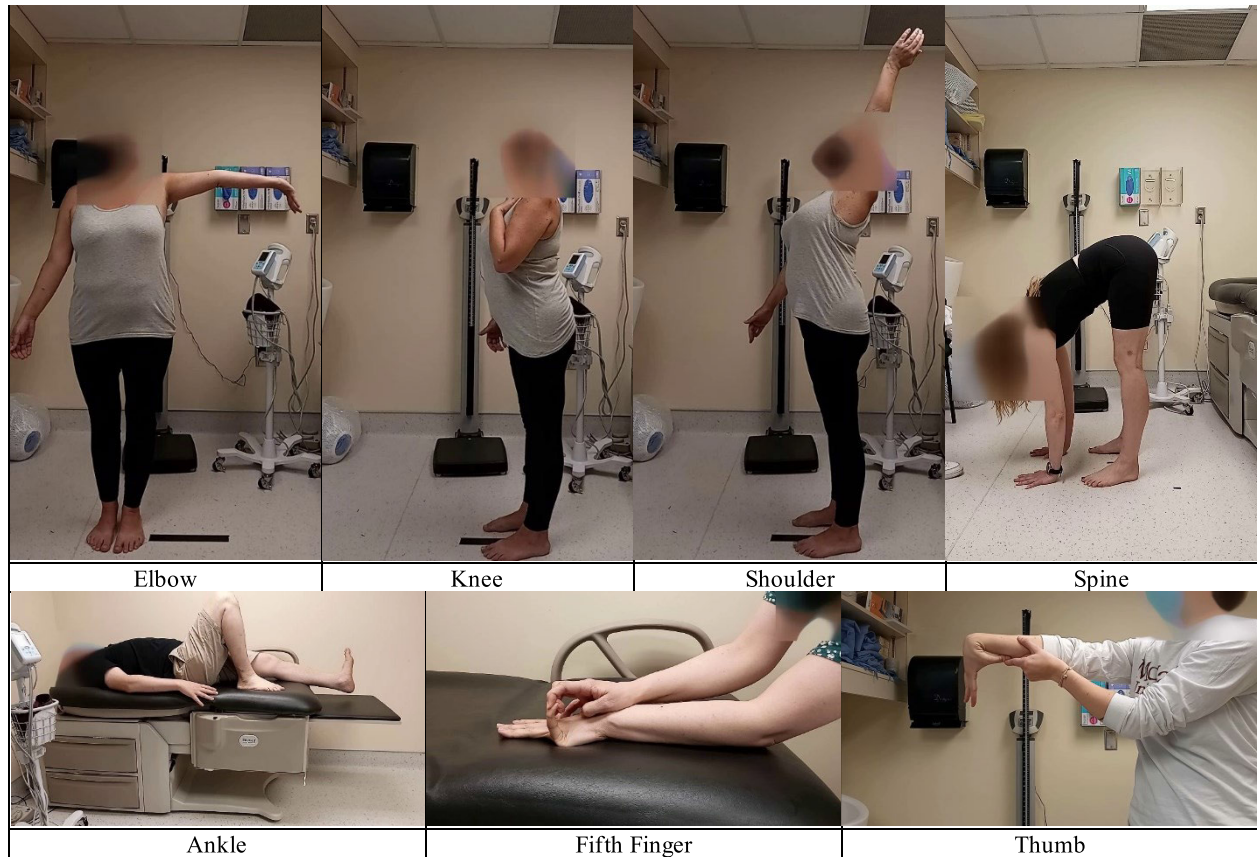
FIGURE 1. Framing of videos for hypermobility assessment by joint.

For each input frame of the recorded video, the 2D human pose-estimation models estimate the pixel positions (x and y) of each joint and a confidence score of the model's certainty of the prediction. The pose-estimation libraries use different deep learning models and may be trained on different datasets so they may output different predictions for a given joint on the same input image. However, these libraries also have different subsets of joints they are trained to predict. Specifically, the AlphaPose library predicts the 26 keypoints on the body and feet as part of the Halpe-Full Body keypoint format [8], the Detectron and MoveNet libraries predict the 17 COCO body keypoints [9], [12], the MediaPipe-Pose model predicts 33 keypoints in the body and feet [10], and the OpenPose library predicts the BODY_25 keypoints [7]. Of particular interest to this study, the Detectron and MoveNet libraries do not predict any keypoints in the feet below the ankles, so analysis of the ankle movement was not possible using the output from these models.

The most recent versions and models with the best published results available for each pose-estimation library as of February 2022 were used in this study.

a: FINETUNING OF A SPECIALIZED MODEL FOR TRACKING KEYPOINTS DURING FIFTH FINGER EXTENSION

In addition to the pre-trained open-source libraries for hand keypoint estimation, a domain-specific model trained for tracking keypoints of the hand was trained and evaluated on the collected dataset. Training was done using the DeepLab-Cut (DLC) library for Python [23], [24]. A model with an ImageNet pretrained MobileNet-v2 backbone [24], [25] was finetuned to predict six keypoints of interest in the hand: the wrist, pinky MCP, pinky PIP, pinky TIP, ring finger TIP, and ring finger MCP. The MobileNet-v2 model was selected for this application because of its small size and fast inference speed [24], [25], leaving the opportunity open for on-device deployment, or deployment on a server without a dedicated GPU in the future.

The keypoint locations used for training were manually labelled by two annotators as described in section B.4) of the Methods. Five-fold cross-validation was used to train and evaluate the model with 80% of the participants included in the train set and 20% included in the test set for each fold. The model was trained for 500, 000 epochs using a learning rate decaying from 0.005 to 0.001. Default values suggested by the DLC library were used for all other training parameters.

2) PREPARATION OF JOINT TRAJECTORIES

After obtaining the per-frame estimates of the joint positions, joint trajectories were formed by temporally combining the positions across the entire video. Joint positions at timesteps where the model was not confident in its prediction were removed and replaced by interpolating using the temporally

TABLE 1. Body keypoints used to calculate angle for each movement.

Movement	Keypoints used to Calculate Angle		
Shoulder flexion	Hip	Shoulder	Elbow
Elbow extension	Shoulder	Elbow	Wrist
Knee extension	Hip	Knee	Ankle
Ankle flexion	Knee	Heel	Large Toe
Fifth Finger extension	Wrist	Pinky MCP	Pinky PIP

adjacent position estimates where the model was confident. The interpolated trajectories were then processed with a zero-phase fourth-order low-pass Butterworth filter with a cut-off of 5 Hz.

3) ESTIMATION OF EXTREME JOINT ANGLES FROM POSE-TRACKED TRAJECTORIES

A deterministic method was used to estimate the joint angle at each timestep of the videos. The three keypoints used to calculate the angle for each movement are summarized in Table 1.

Given the three input keypoints (A, B, and C), the angle between them can be calculated as:

$$\theta = arctan2\left(A_y - B_y, A_x - B_x\right) \\ - arctan2\left(C_y - B_y, C_x - B_x\right) \quad (1)$$

where θ is the angle between the line segments AB and CB; and x and y are the horizontal and vertical pixel positions of each keypoint, respectively. The definitions of joints A and C were adjusted by according the side of the body being analyzed to ensure that the same angle of interest was being calculated. For example, when calculating the angle during right shoulder flexion, A was defined as the hip, B as the shoulder, and C as the elbow, whereas during left shoulder flexion, A was defined as the elbow, B as the shoulder, and C as the hip.

After calculating the angle of interest at each timestep, the most extreme position during the movement was calculated by computing the median in a sliding window of 30 frames and then selecting the largest median in all windows in the trajectory. Care was taken to select extrema where the participant was not moving and where the pose-estimation library was confident in its predictions of the joints used to calculate the angle of interest.

As appropriate, offsets were applied to ensure that the zero-point and direction of the angles were the same as those measured by the clinicians.

4) MANUAL ANNOTATIONS OF EXTREME JOINT ANGLES FROM VIDEO

Manual annotation was used to facilitate a comparison between the angles estimated using a completely automated system relying on pose-estimation libraries with the upper limit of performance from video assuming ideal pose-estimation. A subset of five frames from each video were manually annotated with the keypoints of interest in Table 1. The five frames to label were selected using random uniform sampling of 20 frames for each video and then manually selecting the frames in which the joint of interest was most hyperextended. Additionally, the top of the head and the chin were labeled in the videos of the elbow, knee, and shoulder flexion movements to allow for normalization of distances as a percentage of the size of the head. All image frames were manually annotated independently by two researchers. The researchers were trained by an experienced clinician (NM) and their annotation skills were validated by the clinician before labelling the dataset. For analyses where the proposed system was compared with the manual annotations, the mean angle calculated from the two raters was used.

The manually annotated keypoints were used to estimate the joint angle in each frame using (1). The most extreme angle across the five labelled frames was reported as the final estimated value for each video.

C. CORRELATION ANALYSIS

A correlation analysis was performed to evaluate the relationship between the joint angles measured using the goniometer and those calculated from video using automatic and manual pose-estimation. As a future goal of this work is to develop a screening tool rather than a diagnostic system or precise measurement system, absolute angles are not as important as the relative order of the measurements estimated from video and the goniometer.

The D'Agostino – Pearson test was used to determine if the measurements for each joint and angle estimation method were likely to have come from a normal distribution. As not all feature sets were normally distributed (full details in Table A of the Appendix), Spearman's rank correlation analysis was calculated as it does not make any assumptions about the data distribution.

Bonferroni correction was used to adjust for multiple comparisons per joint due to the pose-estimation methods used and the grouping of the analyses by left/right/both sides of the body. An alpha of 0.05 was used as the threshold for significance for all analyses. All analyses were performed in Python 3.8.10 with SciPy 1.9.1 [26].

III. RESULTS

A total of 100 individuals were recruited to participate in this study. Two individuals withdrew consent, while video data for one individual was not recorded during the data collection phase. Therefore, data from 97 individuals (89 female, 35.0 ± 9.9 years old)) was included in this study. The median Beighton score of the cohort was 3 (mean: 3.6, range: $0 - 8$), with 46 individuals having a Beighton score of 4 or more when assessed at the EDS clinic. Table 2 presents the number of recorded video and the prevalence of positive hypermobility findings by joint. Determinations of joint hypermobility were based on goniometer measurements with the following thresholds indicating positive findings: over 10° degrees extension of the elbow and knee, over 180° flexion of the

TABLE 2. Number of recorded videos and the prevalence of positive hypermobility findings by joint, as assessed by a clinician at the specialized eds clinic.

	Elbow		Knee		Shoulder		Ankle		Fifth Finger		Thumb		Spine
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	
Number of participants assessed in clinic (with goniometer measurement − if appropriate)	93	93	93	93	86	86	86	86	93	93	94	93	94
Number of positive with positive hypermobility finding	12	14	32	33	76	77	18	19	34	21	68	61	60
Percent positive	12.9%	15.1%	34.4%	35.5%	88.4%	89.5%	20.9%	22.1%	36.6%	22.6%	72.3%	65.6%	63.8%
Number of participants with recorded video of joint assessment	95	96	95	96	91	91	96	96	97	96	97	92	96

TABLE 3. Intra-class correlation coefficient (ICC) by joint between two human annotators.

	Side of Body								
	Left			Right			Both		
	ICC	95% CI	p	ICC	95% CI	p	ICC	95% CI	p
Elbow	.953	[.93 − .97]	<.001	.966	[.95 − .98]	<.001	.959	[.95 − .97]	<.001
Knee	.806	[.72 − .87]	<.001	.835	[.76 − .89]	<.001	.823	[.77 − .86]	<.001
Shoulder	.935	[.90 − .96]	<.001	.912	[.87 − .94]	<.001	.923	[.90 − .94]	<.001
Ankle	.975	[.96 − .98]	<.001	.962	[.94 − .97]	<.001	.968	[.96 − .98]	<.001
Fifth Finger	.942	[.92 − .96]	<.001	.967	[.95 − .98]	<.001	.947	[.93 − .96]	<.001

shoulder, over 15° flexion of the ankle, and over 90° extension of the fifth finger. The prevalence of positive findings for the thumb and spine joints, as determined by standard Beighton criteria, are also presented in Table 2, however these joints were not assessed using a goniometer and were thus not analyzed further in this work.

Some data points for the ankles and shoulders are missing as the recording of precise angle measurements for these joints in the patient chart was not done before this study, so they were omitted for the first participants in the study until this became standard practice. Furthermore, any joints where the participant was in pain and did not wish to proceed were not assessed, leading to discrepancies in the number of recorded videos by joint.

A. INTRA-CLASS CORRELATION ANALYSIS BETWEEN ANNOTATORS

An intra-class correlation (ICC) analysis was performed to assess whether the maximum joint angles as annotated by each researcher were consistent and reliable. Table 3 presents the ICC, 95% confidence interval, and p-value for each joint. The ICC values between the two annotators were significant across all joints. The magnitude of the ICC was greater than .9 for all joints other than the knee, indicating high agreement between the two raters. A weaker, but still moderate to strong agreement was noted for the knee joint (.823 when combining both left and right sides).

B. CORRELATION OF ANGLES CALCULATED FROM VIDEO WITH GONIOMETER MEASUREMENTS

Joint angles were calculated from video using the keypoints extracted using the pose-estimation libraries, as well as from the manually annotated keypoints. Table 4 presents

the Spearman's rank coefficients and Bonferroni-adjusted p-values for the correlation between the angles calculated from video and those measured with a goniometer for each joint.

As seen in Table 4, the correlation between the angles estimated from video and those measured using the goniometer were all statistically significant for the elbow, knee, and shoulder movements. In general, the correlations between the angles calculated from the manual annotations and the goniometer measurements are stronger than the angles calculated from the pose-tracked keypoints. There is a larger difference between the correlation when the manual annotations are used and the top-performing pose-estimation library for the knee joint than for the elbow or shoulder joints.

Conversely, the angles calculated from video for the ankle and fifth finger movements are not consistently significantly correlated when the pose-estimated keypoints from the open-source libraries were used. The correlations between the goniometer measurements and the angles calculated from the keypoints inferred from the domain-specific MobileNet-v2 model were strong and statistically significant. Furthermore, they were similar in magnitude to the correlations between the goniometer and manually annotated keypoints.

For the ankle movement, the angles calculated from the manually annotated keypoints were not significantly correlated when one side of the body was assessed at a time and only weakly correlated when data from both sides is combined.

C. CORRELATION OF ANGLES CALCULATED FROM MANUALLY ANNOTATED AND POSE-TRACKED KEYPOINTS

To control for the angles at which the videos were collected, the joint angles calculated using the pose-tracked keypoints

TABLE 4. Spearman's rank correlation coefficient and Bonferroni-Corrected right-tailed p-value for angles calculated from video and measured with a goniometer, grouped by joint and pose-estimation library.

| | | Side of Body | | | | | |
| | | Left | | Right | | Both | |
		rho	p	rho	p	rho	p
Elbow	Manual	.688	<.001	.796	<.001	.746	<.001
	AlphaPose	.707	<.001	.699	<.001	.699	<.001
	Detectron	.690	<.001	.755	<.001	.722	<.001
	MediaPipe-Pose	.594	<.001	.715	<.001	.653	<.001
	MoveNet-Thunder	.659	<.001	.637	<.001	.648	<.001
	OpenPose	.557	<.001	.661	<.001	.601	<.001
Knee	Manual	.711	<.001	.680	<.001	.693	<.001
	AlphaPose	.500	<.001	.565	<.001	.530	<.001
	Detectron	.526	<.001	.521	.002	.520	<.001
	MediaPipe-Pose	.563	<.001	.592	<.001	.577	<.001
	MoveNet-Thunder	.659	<.001	.562	<.001	.608	<.001
	OpenPose	.573	<.001	.627	<.001	.595	<.001
Shoulder	Manual	.649	<.001	.682	<.001	.661	<.001
	AlphaPose	.574	<.001	.640	<.001	.596	<.001
	Detectron	.507	<.001	.596	<.001	.555	<.001
	MediaPipe-Pose	.587	<.001	.629	<.001	.598	<.001
	MoveNet-Thunder	.591	<.001	.682	<.001	.632	<.001
	OpenPose	.559	<.001	.567	<.001	.563	<.001
Ankle	Manual	.290	.045	.278	.059	.290	<.001
	AlphaPose	.030	1.000	.227	.484	.131	1.000
	Detectron						
	MediaPipe-Pose	.345	.140	.198	.956	.274	.043
	MoveNet-Thunder						
	OpenPose	.230	.958	.184	1.000	.208	.356
Fifth Finger	Manual	.730	<.001	.783	<.001	.754	<.001
	AlphaPose	.126	1.000	.206	.688	.205	.150
	MediaPipe-Hands	.154	.890	.356	.003	.211	.026
	Finetuned MobileNet Model	.807	<.001	.759	<.001	.786	<.001

were correlated with the angles calculated from the manually annotated keypoints. Table 5 presents the Spearman's rank coefficient and Bonferroni-adjusted p-values for the correlation between the angles calculated from video using the manually annotated and pose-tracked keypoints.

From Table 5, it can be seen that the angles calculated at the elbow, knee, and shoulder using the pose-tracked keypoints are significantly correlated with the angles calculated using the manually annotated keypoints from the same videos. The correlations are moderate to strong for the elbows and shoulders, and moderate for the knees. For the ankle movement, only the angles calculated from the Open-Pose keypoints are significantly and moderately correlated with the manual annotations. Finally, there was a significant but weak correlation between the angles calculated from the MediaPipe-Hands and AlphaPose keypoints and manual annotations for the fifth finger movement when combining both sides of the body. There was a strong correlation between the angles calculated from the fifth-finger manual annotations and those predicted by the finetuned MobileNet-v2 model. All other correlations were not significant.

Figure 2 presents a visualization of the strongest correlations between the angles calculated from video using pose-estimation compared to the goniometer measurements (left column); and angles calculated using pose-estimation compared to those calculated from the manually annotated keypoints for the same videos (right column).

IV. DISCUSSION
Angles of suspected hypermobile joints calculated from video were correlated with those measured using a goniometer. As seen in Table 4, the strongest correlations were achieved when the positions of the keypoints were manually annotated in videos; however, significant correlations with the goniometer measurements were also achieved when keypoints were automatically extracted using open-source or finetuned domain-specific pose-estimation libraries. The strength of correlation between the angles calculated from manual annotations and the goniometer measurements indicate that it is feasible to assess joint hypermobility from video.

TABLE 5. Spearman's rank correlation coefficient and Bonferroni-Corrected right-tailed p-value for angles calculated from video using manually annotated and pose-tracked keypoints, grouped by joint and pose-estimation library.

		Side of Body					
		Left		Right		Both	
		rho	*p*	rho	*p*	rho	*p*
Elbow	AlphaPose	.854	<.001	.812	<.001	.821	<.001
	Detectron	.914	<.001	.887	<.001	.904	<.001
	MediaPipe-Pose	.869	<.001	.803	<.001	.832	<.001
	MoveNet-Thunder	.857	<.001	.753	<.001	.804	<.001
	OpenPose	.819	<.001	.789	<.001	.806	<.001
Knee	AlphaPose	.572	<.001	.690	<.001	.621	<.001
	Detectron	.765	<.001	.718	<.001	.739	<.001
	MediaPipe-Pose	.678	<.001	.685	<.001	.672	<.001
	MoveNet-Thunder	.771	<.001	.696	<.001	.731	<.001
	OpenPose	.644	<.001	.786	<.001	.701	<.001
Shoulder	AlphaPose	.819	<.001	.810	<.001	.813	<.001
	Detectron	.696	<.001	.693	<.001	.699	<.001
	MediaPipe-Pose	.891	<.001	.858	<.001	.877	<.001
	MoveNet-Thunder	.886	<.001	.901	<.001	.893	<.001
	OpenPose	.842	<.001	.828	<.001	.836	<.001
Ankle	AlphaPose	.200	.615	.169	.749	.185	.172
	Detectron						
	MediaPipe-Pose	.173	1.000	.119	1.000	.148	.549
	MoveNet-Thunder						
	OpenPose	.505	.002	.533	<.001	.505	<.001
Fifth Fing	AlphaPose	.285	.112	.162	.926	.252	.021
	MediaPipe-Hands	.122	1.000	.386	<.001	.225	.009
	Finetuned MobileNet Model	.941	<.001	.920	<.001	.936	<.001

A. CORRELATIONS OF THE ELBOW, KNEE, AND SHOULDER ANGLES

Overall, when using open-source pose-estimation libraries, the angles at the larger joints (elbow, knee, shoulder) were more strongly and significantly correlated to the goniometer measurements. In general, the correlations for the elbow and shoulder were stronger than those for the knee. This is consistent with the results noted as part of the ICC analysis (Table 3), where agreement between human annotators was lower for the knee joint than for the others.

Excluding the manual annotations, the Detectron library achieved the strongest correlation (Spearman's rho = .722, $p < .001$) across both sides of the body for the elbow extension movement. Angles calculated using MoveNet – Thunder keypoints achieved the strongest correlation to the goniometer measurements (Spearman's rho = .608, $p < .001$) on the knee extension movement, while those calculated from MoveNet – Thunder were most strongly correlated for shoulder flexion assessment (Spearman's rho = .632, $p < .001$). As seen in Table 4, the difference in the strength of the correlation to goniometer measurements across the top two pose-estimation libraries is minor for most joints, so there are several strong candidates when selecting which library to use.

Another consideration when selecting a pose-estimation library for a joint is the number of videos that were successfully analyzed. As presented in Table A of the Appendix, the angles could not be estimated for all videos using all pose-estimation libraries due to the low confidence of the

tracked keypoints. This was particularly evident for the knee extension movement for which the keypoints extracted using the Detectron library were only sufficiently confident to calculate the angles for 111 of the 191 available videos (58%). Angles were able to be estimated for a much larger proportion of the videos of the knee movement using the other four body pose-estimation libraries (MediaPipe – Pose: 100%, MoveNet – Thunder: 100%, OpenPose: 100%, AlphaPose: 97%). This trend was not observed for the elbow and shoulder where the number of successfully calculated angles was similar across all five body pose-estimation libraries. Of note, the Detectron and MoveNet libraries did not track the keypoints below the knee. It is possible that the lack of training data and lack of model focus on the lower body for the Detectron pose-estimation library contributed to the lower confidence in the keypoint predictions for the knee movement but not the others. This trend was not observed for the MoveNet – Thunder model.

Additionally, the angles calculated from the videos using pose-estimation were correlated with those calculated from the manually annotated keypoints. As presented in Table 5 and the right column of Figure 2, the correlations at the elbow, knee, and shoulder joints were moderate to strong across all pose-estimation libraries, indicating that pose-estimation libraries track the major joints of the upper body similarly to a human annotator. The Spearman's rho coefficient for the top-performing pose-estimation library per joint was .904 for the elbows (Detectron), .739 for the knees (Detectron), and
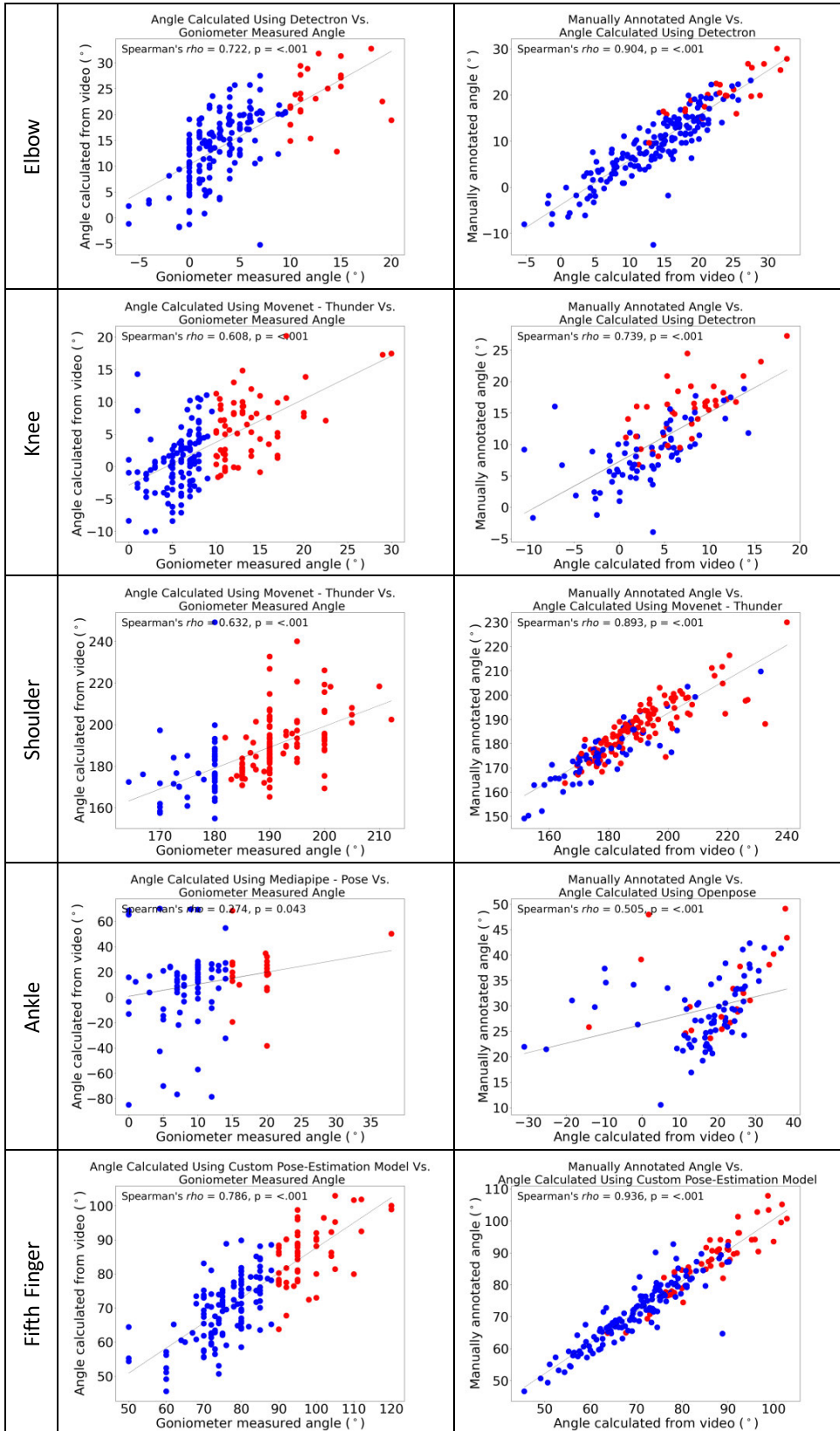
FIGURE 2. Scatterplots of angles calculated from video using pose-estimation libraries compared to goniometer measurements (left) and angles calculated from manual annotations (right). Red dots represent measurements that met the threshold for hypermobility at each joint, while blue dots did not.

.893 for the shoulders (MoveNet – Thunder), respectively. These findings are consistent with work by Fan, Gu et al. who described moderate Pearson's correlation coefficients (r) of .71 and .45 for elbow extension and shoulder flexion, respectively, for the relationship between angles automatically estimated using OpenPose and those calculated from manual annotations [14].

A further analysis evaluating the difference in the keypoint locations detected by each pose-estimation library and the manually annotated ones is presented in Figure A of the Appendix. While the pose-estimation library with the smallest difference in manually annotated and predicted joint locations varies by joint, the largest difference is generally observed for the AlphaPose library.

B. CORRELATIONS OF THE ANKLE AND FIFTH FINGER ANGLES

With the open-source pose-estimation models and data collection methodology evaluated in this study, the angles measured with a goniometer and those calculated from video for the ankles and fifth fingers were not consistently correlated. As presented in Table 4, a moderate to strong correlation between the manually annotated angles and the goniometer measurements was noted for the fifth fingers (.754).

The correlations between the angles measured with the goniometer and manually annotated angles were very weak for the ankle movement (Spearman's rho = .290, p < .001). These results suggest that the manner in which the videos were taken is not conducive to measuring the ankle dorsiflexion movement. As previously described, the videos of the ankle movement were framed to capture the whole body to facilitate better pose-estimation. From Table 5 and the right column of Figure 2, a moderate correlation (.505) is observed between the angles calculated from the manually annotated locations and the OpenPose library for the ankle movement. However, the lack of correlation with the goniometer measurements indicates that framing of the video does not appropriately capture the plane of the movement being assessed by the clinician.

This is in contrast to the trend observed for the fifth finger extension movement where the manually annotated angles and values measured using a goniometer were moderately correlated (Spearman's rho = .754, p < .001). However, when the open-source pose-estimation libraries were used to extract the positions of the hand keypoints, the angles subsequently calculated from the tracked keypoints were not significantly correlated to those calculated from the manual annotations or the goniometer measurements (Table 4). These results indicate that while the framing of the videos is appropriate to capture the movement of interest assuming ideal annotation of hand keypoints, the open-source pose-estimation models evaluated in this study are not able to track the joints accurately enough for this application. This is further confirmed by strong correlation (.786) between the goniometer measurements and the angles calculated from

video when the finetuned MobileNet-v2 model was used to extract keypoints (Table 4, Figure 2 – bottom row).

V. CONCLUSION AND FUTURE WORK

This study quantified the correlations between the angles measured with a goniometer and those estimated from video using various pose-estimation libraries across five movements used to assess joint hypermobility in a population of adults with EDS/G-HSD. Moderate to strong correlations between angles automatically calculated from video using open-source models and the goniometer measurements for the elbow, knee, and shoulder movements indicate that the proposed video-based method is suitable for screening for hypermobility in these joints. Conversely, the current video recording protocol paired with open-source pose-estimation methods is not suitable for screening for hypermobility in the ankle and fifth finger joints. Using a domain-specific keypoint extraction model for the fifth finger movement improved the correlation between the goniometer measurements and the angle estimates from video significantly.

Future work will focus on developing a more reliable means of recording the ankle movement to facilitate angle estimation from video. For example, rather than recording the participant's entire body in the video (which was required to improve accuracy of the general, open-source pose-estimation libraries), we will elect to only take videos of the lower leg. To enable more accurate angle estimation, this will likely also involve the development of a specialized pose-estimation models suitable for more accurately tracking of the keypoints of interest in the ankle. As was done for the fifth-finger model, existing models will be finetuned using data collected as part of this study to develop ankle keypoint tracking models that are optimized for individuals with more hypermobility.

After developing a methodology for reliably estimating the angle at all five joints presented in this study, cut-off thresholds will be selected to allow the system to screen out the individuals who do not exhibit evidence of hypermobility at each joint. The cut-off threshold will vary by joint and will be dependent on the prevalence of hypermobility at each joint. This system will be validated as a screening tool for generalized joint hypermobility in future work.

REFERENCES

[1] M. Castori, B. Tinkle, H. Levy, R. Grahame, F. Malfait, and A. Hakim, "A framework for the classification of joint hypermobility and related conditions," *Amer. J. Med. Genet. C, Seminars Med. Genet.*, vol. 175, no. 1, pp. 148–157, Mar. 2017.

[2] F. Malfait et al., "The 2017 international classification of the Ehlers–Danlos syndromes," *Amer. J. Med. Genet. C, Seminars Med. Genet.*, vol. 175, pp. 8–26, Mar. 2017.

[3] P. H. Beighton and F. Horan, "Orthopaedic aspects of the Ehlers–Danlos syndrome," *J. Bone Joint Surg. Brit.*, vol. 51, no. 3, pp. 444–453, 1969.

[4] C. Carter and J. Wilkinson, "Persistent joint laxity and congenital dislocation of the hip," *J. Bone Joint Surg. Brit.*, vols. 46, no. 1, pp. 40–45, Feb. 1964.

[5] L. N. Bockhorn, A. M. Vera, D. Dong, D. A. Delgado, K. E. Varner, and J. D. Harris, "Interrater and intrarater reliability of the Beighton score: A systematic review," *Orthopaedic J. Sports Med.*, vol. 9, no. 1, Jan. 2021, Art. no. 232596712096809.

[6] A. Vallis, A. Wray, and T. Smith, "Inter- and intra-rater reliabilities of the Beighton score compared to the contompasis score to assess generalised joint hypermobility," *Myopain*, vol. 23, nos. 1–2, pp. 21–27, Apr. 2015.

[7] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[8] H.-S. Fang et al., "AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time," 2022, *arXiv:2211.03375*.

[9] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. (2019). *Detectron2*. [Online]. Available: https://github.com/facebookresearch/detectron2

[10] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.

[11] F. Zhang et al., "MediaPipe hands: On-device real-time hand tracking," 2020, *arXiv:2006.10214*.

[12] Google Research. *Next-Generation Pose Detection with MoveNet and TensorFlow.js*. Accessed: Sep. 29, 2023. [Online]. Available: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

[13] A. B. Cunha et al., "Assessing the validity and reliability of a new video goniometer app for measuring joint angles in adults and children," *Arch. Phys. Med. Rehabil.*, vol. 101, no. 2, pp. 275–282, Feb. 2020.

[14] J. Fan et al., "Reliability of a human pose tracking algorithm for measuring upper limb joints: Comparison with photography-based goniometry," *BMC Musculoskeletal Disorders*, vol. 23, no. 1, p. 877, Sep. 2022, doi: 10.1186/s12891-022-05826-4.

[15] L. Gionfrida, W. M. R. Rusli, A. A. Bharath, and A. E. Kedgley, "Validation of two-dimensional video-based inference of finger kinematics with pose estimation," *PLoS ONE*, vol. 17, no. 11, Nov. 2022, Art. no. e0276799.

[16] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," 2014, *arXiv:1405.0312*.

[17] N. Mittal et al., "The GoodHope ehlers danlos syndrome clinic: Development and implementation of the first interdisciplinary program for multi-system issues in connective tissue disorders at the Toronto general hospital," *Orphanet J. Rare Diseases*, vol. 16, no. 1, pp. 1–9, Dec. 2021.

[18] N. Mittal, A. Sabo, A. Deshpande, H. Clarke, and B. Taati, "Feasibility of video-based joint hypermobility assessment in individuals with suspected ehlers-danlos syndromes/generalised hypermobility spectrum disorders: A single-site observational study protocol," *BMJ Open*, vol. 12, no. 12, Dec. 2022, Art. no. e068098, doi: 10.1136/bmjopen-2022-068098.

[19] C. C. Norkin and D. J. White, *Measurement of Joint Motion: A Guide to Goniometry*, 5th ed. Philadelphia, PA, USA: F. A. Davis Company, 2016.

[20] D. Rempe, T. Birdal, A. Hertzmann, J. Yang, S. Sridhar, and L. J. Guibas, "HuMoR: 3D human motion model for robust pose estimation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, vol. 2021, Oct. 2021, pp. 11468–11479.

[21] M. Li et al., "Interacting attention graph for single image two-hand reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2022, Jun. 2022, pp. 2751–2760.

[22] L. Ge et al., "3D hand shape and pose estimation from a single RGB image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2019, Jun. 2019, pp. 10825–10834.

[23] A. Mathis et al., "DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning," *Nature Neurosci.*, vol. 21, no. 9, pp. 1281–1289, Sep. 2018, doi: 10.1038/s41593-018-0209-y.

[24] A. Mathis et al., "Pretraining boosts out-of-domain robustness for pose estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, vol. 2021, Jan. 2021, pp. 1858–1867.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[26] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.