Annals of Biomedical Engineering (2026) 54:284–302
https://doi.org/10.1007/s10439-025-03882-0

BMES BIOMEDICAL ENGINEERING SOCIETY

ORIGINAL ARTICLE

From Skin to Skeleton: A Statistical Shape Modelling Approach for Predicting Hand and Foot Bony Geometry

Kate Duquesne[1,2] · Adris Molnar[2] · Roel Huysentruyt[1,2] · Aline Van Oevelen[1,2,5] · Jing Li[1,2] · Jan Sijbers[3] · Wim Van Paepegem[4] · Emmanuel Audenaert[1,2,5,6]

Received: 2 July 2025 / Accepted: 9 October 2025 / Published online: 30 October 2025

Abstract

Purpose In clinics, musculoskeletal assessment mainly relies on conventional imaging which is expensive, involves radiation exposure, and lacks accessibility. Inferring skeletal morphology from 3D body surface scans shows potential as an alternative screening aid, although lacks detail in particular at extremities. This study leverages Statistical Shape Models (SSMs) to estimate position and orientation of the hand and foot bones from the skin surface.

Methods Two datasets (140 feet, 79 hands with diverse morphologies and poses) were collected. For each dataset, a coupled skin-bone SSM was created. A nested cross-validation approach was used to optimize hyperparameters and prevent overfitting while fitting the isolated skin model of the coupled SSM to unseen skin data to infer bone structure and position.

Results For the feet, the mean absolute error (MAE) was 1.68 mm, with the highest errors occurring at the hindfoot. Similarly, for the hands, the MAE reached 1.37 mm, with the largest deviations observed at carpal bones.

Conclusion This study demonstrates the feasibility of predicting bone morphology from skin surfaces using SSM-based shape completion, offering a potential non-invasive and accessible alternative to traditional imaging for musculoskeletal assessments.

Keywords Statistical shape modeling · Foot · Hand · Skeleton prediction

Introduction

The human musculoskeletal system, which is responsible for our ability to move and to perform complicated tasks, is a complex network of bones, muscles, tendons, and ligaments. Detailed anatomical assessment of this system is essential for numerous musculoskeletal and research applications. These range from medical diagnosis and treatment of musculoskeletal conditions like arthritis and sports injuries, to biomechanics research, development of customized footwear, and even forensic science [3]. Traditional imaging modalities such as X-ray imaging, computed tomography (CT), and magnetic resonance imaging (MRI) provide valuable internal anatomical information. However, each modality has significant limitations: they can be costly and time-consuming, involve exposure to ionizing radiation, offer limited resolution in large field-of-view imaging, and suffer from constrained accessibility in many healthcare settings [15]. As a result, researchers have started exploring new methodologies that derive clinical anthropometric measurements, body composition and even predict bone morphology

Associate Editor Joel Stitzel oversaw the review of this article.

✉ Kate Duquesne
kate.duquesne@ugent.be

1 Department of Orthopedic Surgery and Traumatology, Ghent University Hospital, Corneel Heymanslaan 10, 9000 Ghent, Belgium

2 Department of Human Structure and Repair, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium

3 Imec—Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerp, Belgium

4 Department of Materials, Textiles and Chemical Engineering, Faculty of Engineering and Architecture, Ghent University, Tech Lane Ghent Science 46, 9052 Ghent, Belgium

5 Department of Electromechanics, InViLab research group, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

6 Department of Trauma and Orthopedics, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, UK

directly from skin surface measurements utilizing technologies, such as 3D body scanners, as well as single- and multi-view cameras and smartphones [6, 9, 14, 20–22, 25, 26].

Early attempts to address this challenge originated in the field of computer graphics and animation. These approaches typically involved non-linearly deforming a template anatomical atlas to match an observed skin surface, effectively warping a generic skeletal model to fit individual body shapes [1, 19]. However, because these graphics-based techniques were not validated against 3D medical imaging data, their clinical application remained limited.

More recently, attention has shifted toward data-driven methods, with the aim to directly predict the full-body skeleton from skin surface data. For instance, Obtaining Skeletal Shape from Outside (OSSO) is trained on 2D Dual-energy X-ray Absorptiometry (DXA) scans of supine subjects and employs a three-step process for skeleton prediction and positioning [22]. However, its optimization-based positioning neglects soft tissue interactions, which can result in anatomically invalid poses, unrealistic joint dislocations, and mesh penetrations. To address some of these shortcomings, Skeletal Kinematics Enveloped by a Learned body model (SKEL) offers an alternative strategy. SKEL integrates an idealized biomechanical skeleton with joint constraints, derived from OpenSim to re-rig the Skinned Multi-Person Linear Model (SMPL) mesh [21]. While this approach allows for scaling and positioning of the 24-segment skeleton within SMPL, it imposes inherent limitations by oversimplifying joint kinematics—restricting them to idealized hinge or spherical joints—and by failing to capture the full spectrum of anatomical bone variability. In contrast, Human Implicit Tissues (HIT) directly addresses the latter limitation [20]. Trained on 3D MRI data, HIT performs a binary prediction of the combined long bones (Humerus, Ulna, Radius, Pelvis, Femur, Tibia, and Fibula) given the skin surface. Building on the foundation laid by HIT and SKEL, Dakri et al. further refined the approach by splitting HIT's combined long bone predictions into individual bone segments (humerus, radius-ulna, pelvis, femur, tibia-fibula) and by extending SKEL to SKEL-J [9]. SKEL-J allows for joint location adjustments and pose parameter refinement to better align the skeletal kinematic tree with segmented bone geometry. Despite these advancements, both HIT and the work of Dakri et al. remain limited as they focused exclusively on long bones.

Beyond the computer vision and animation techniques discussed above, statistical shape models (SSMs) offer an alternative promising approach. SSMs, as described in Sect. "Prediction of Bony Morphology from the Skin Surface," have proven to be a powerful tool in medical image analysis, aiding in tasks, such as segmentation, shape completion, pre-operative planning, and biomechanical analysis [3, 7]. Numerous individual and coupled bony anatomy SSMs have been created [5, 13, 29–31]. However, to the best of our knowledge, only one existing model, the Bones, Organs, and Skin Shape model (BOSS), combines the complete skeleton with the skin. BOSS is a pose-neutralized SSM of the combined skeleton, organs, and skin, learned from 300 (partial) CT scans [32]. Its bone model comprises 70 segments, including the skull, femur, humerus, forearm, lower leg, scapula, clavicle, sternum, hands, feet, vertebrae, ribs, and pelvis. Pose neutralization is achieved through a kinematic chain of 63 joints, connecting 63 segments. As a result of the pose neutralization, the prediction of bony morphology can only be performed in that specific position and the method can as such not be used to complete other scan positions.

A key limitation of the aforementioned methods is the absence of detailed hand and foot models, which are essential for understanding locomotion and dexterity. The aim of this study was to predict the skeletal morphology and position of the individual bones from the skin surface for the hands and feet. We address this by exploiting SSMs. This offers a less invasive and accessible alternative for musculoskeletal applications.

Materials and Methods

This section details the methodology for predicting bony anatomy directly from skin surface data, focusing on two distinct yet complex anatomical regions: the feet and the hands. First, the characteristics of the feet and hands datasets are described. Next, the shape completion methodology based on statistical shape modeling to predict the bony anatomy is outlined. Following this, an experiment involving the application of the feet bone prediction models to input data of reduced quality is discussed. Finally, the evaluation metrics are discussed.

Data Description and Preprocessing

The data gathering was conducted in compliance with the ethical standards established in the Declaration of Helsinki.

Dataset 1: Feet

A weightbearing Computed Tomography (WBCT) foot dataset, derived from standing CT scans and ethically approved by the Ghent University Hospital (B6702022000639), was gathered. Patients were positioned with the feet parallel and at shoulder width. The scans had a pixel size of 0.5 mm x 0.5 mm and a slice thickness of 0.5 mm. This foot dataset comprised 73 subjects featuring a diverse range of foot types (i.e., healthy, flat, and hollow feet). Feet with major defects or deformities—including osseous foot coalitions, clubfoot,

and charcot feet—were excluded. Additionally, any participant with a history of prior surgical interventions involving screws or implants at the foot or ankle was also excluded from the study. This filtering process resulted in a dataset of 140 individual feet. The participant cohort (40 females, 33 males) presented a mean age of 39.1 years (standard deviation (SD) 16.7 years), a mean height of 1.73 meters (SD 0.11 m), and a mean weight of 79.6 kilograms (SD 18.3 kg). Only patients aged 18 years or older were included. Figure 1 illustrates two perspectives of five representative subjects' skeletal structure and skin.

Dataset 2: Hands

The hand dataset was derived from CT scans obtained from the publicly available New Mexico Decedent Image Database (NMDID, https://nmdid.unm.edu/), in accordance with their ethical protocols. The scans had a pixel size of 1.34 mm x 1.34 mm and a slice thickness of 3 mm. Exclusion was based on two criteria: the presence of visible major hand deviations (e.g., amputation, fractures, prior surgery involving screws or implants, or polydactyly) or hands that were not fully contained within the CT scanner's field of view. Further, no subjects younger than 18 years were included. The final dataset comprised 41 right hands and 38 left hands in a wide range of poses. The dataset consisted of 45 subjects (24 males, 21 females) with a mean age of 44.0 years (SD: 13.7 years), an average height of 1.68 meters (SD: 0.12 m), and a mean weight of 80.0 kilograms (SD: 23.0 kg). Figure 2 illustrates two perspectives of five representative subjects' skeletal structure and skin.

Preprocessing

A combined manual and automated segmentation process was used to delineate bones and skin in both datasets. This involved initial manual segmentation of a subset of subjects, nnU-Net training, automated prediction, and manual refinement of the prediction [16, 17]. Separate nnU-Nets were trained for the segmentation of the foot bones, hand bones, and skin. The nnU-Nets were trained for 1000 epochs with a combined Dice and cross-entropy loss and a batch size of 2. For the foot bones, the model was trained on 45 WBCT images with a patch size of (112, 92, 96). For the hand bones and skins, the model was trained on 13 CT images with patch sizes of (64, 160, 128) and (64, 192, 160), respectively. Foot bone segmentations included 28 bones namely the talus, calcaneus, navicular bone, cuneiforms (medial, inter and lateral), cuboid, metatarsals (1–5), phalanges (proximal 1–5, middle 2–5, and distal 1–5), and sesamoids (medial and lateral). The included foot bones are depicted in Fig. 1 in distinct colors. Hand bone segmentations included 27 bones,

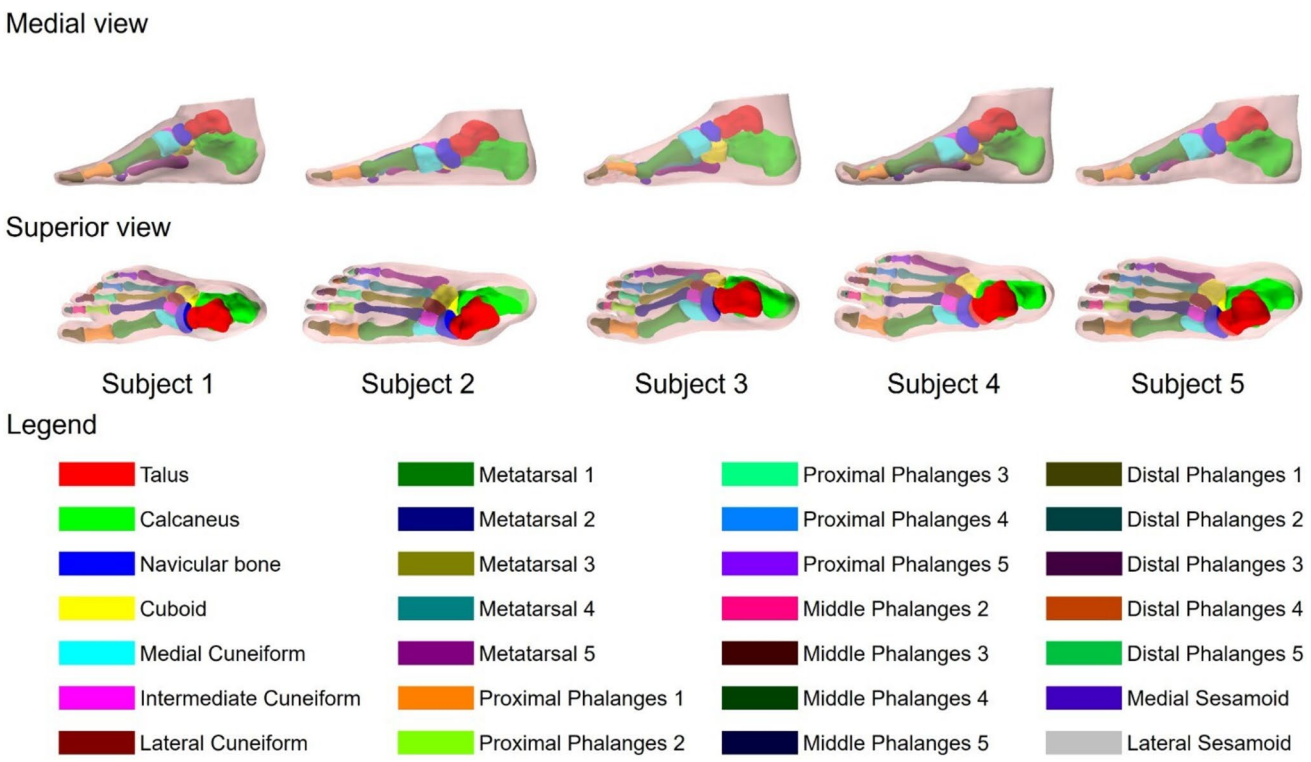

Fig. 1 Medial and superior views of the feet of five representative subjects from the WBCT dataset. Each individual modeled bone is indicated in a unique color

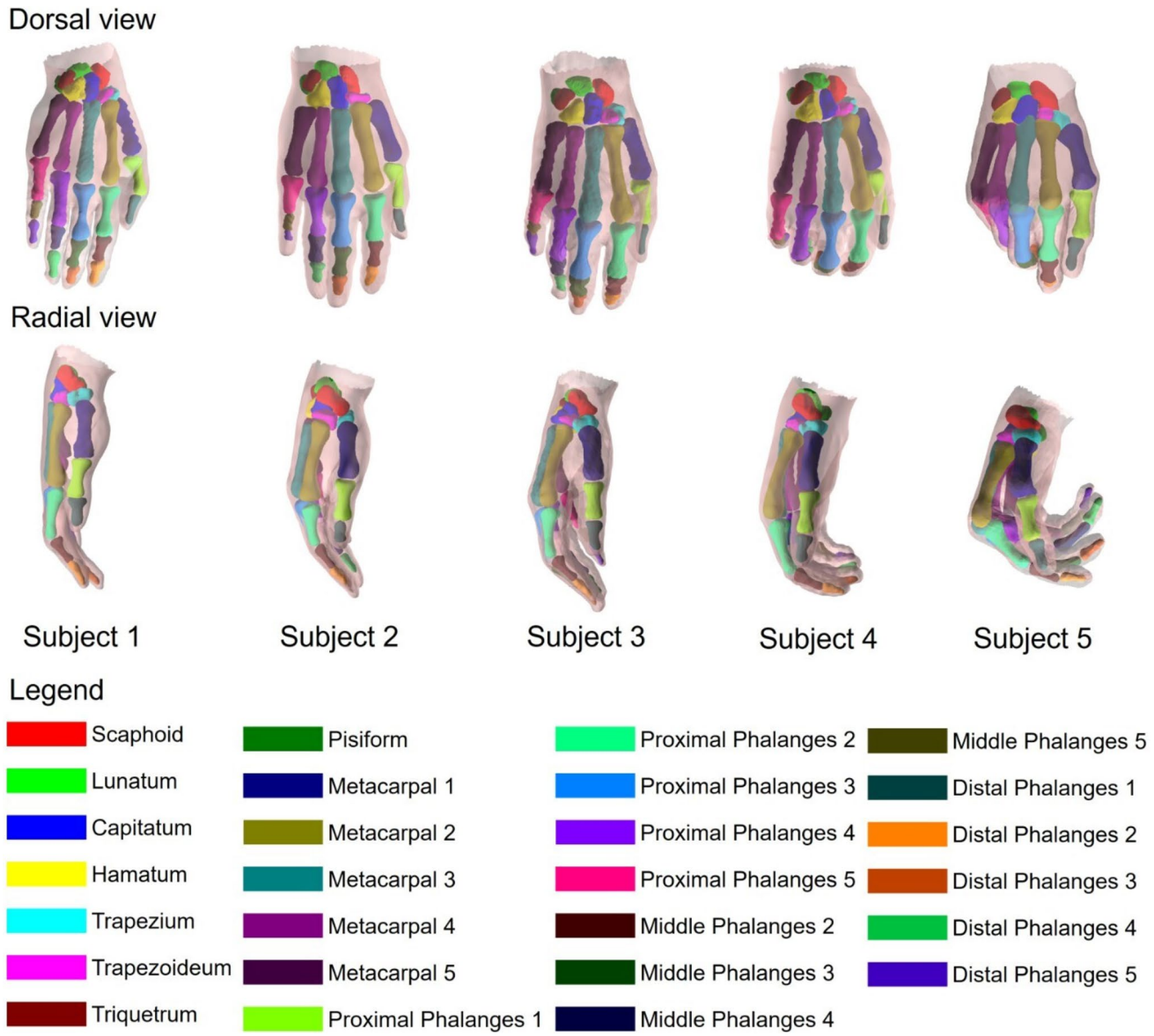

Fig. 2 Dorsal and radial views of the hands of five representative subjects from the hand dataset. Each individual modeled bone is indicated in a unique color

namely the carpal bones (i.e., scaphoid, lunate, trapezium, trapezoid, hamate, capitate, pisiform, triquetrum), metacarpals (1–5), and phalanges (proximal 1–5, middle 2–5, and distal 1–5). The included hand bones are depicted in Fig. 2 in distinct colors. Following segmentation, the method by Audenaert et al. was used to perform non-rigid registration of a template mesh [5]. This involved global registration followed by a local registration step, resulting in anatomical correspondence over the bones and skins with a consistent number of vertices. The complete set of foot bones, m_{bones}, was represented using 34362 vertices (edge length average: 1.38 mm, SD: 0.53 mm), while its corresponding skin surface, m_{skin}, had 65869 vertices (edge length average: 1.16 mm, SD: 0.34 mm). For the hand, the skeletal structure, m_{bones}, consisted of 22985 vertices (edge length average: 1.24 mm, SD: 0.47 mm), and the skin surface, m_{skin}, of 13475 vertices (edge length average: 1.98 mm, SD: 0.70 mm). The mesh resolutions were chosen to balance anatomical detail with computational efficiency and are consistent with previous work [5, 30].

Prediction of Bony Morphology from the Skin Surface

Bone morphology prediction was achieved through shape completion using SSMs (Fig. 3). To avoid overfitting and

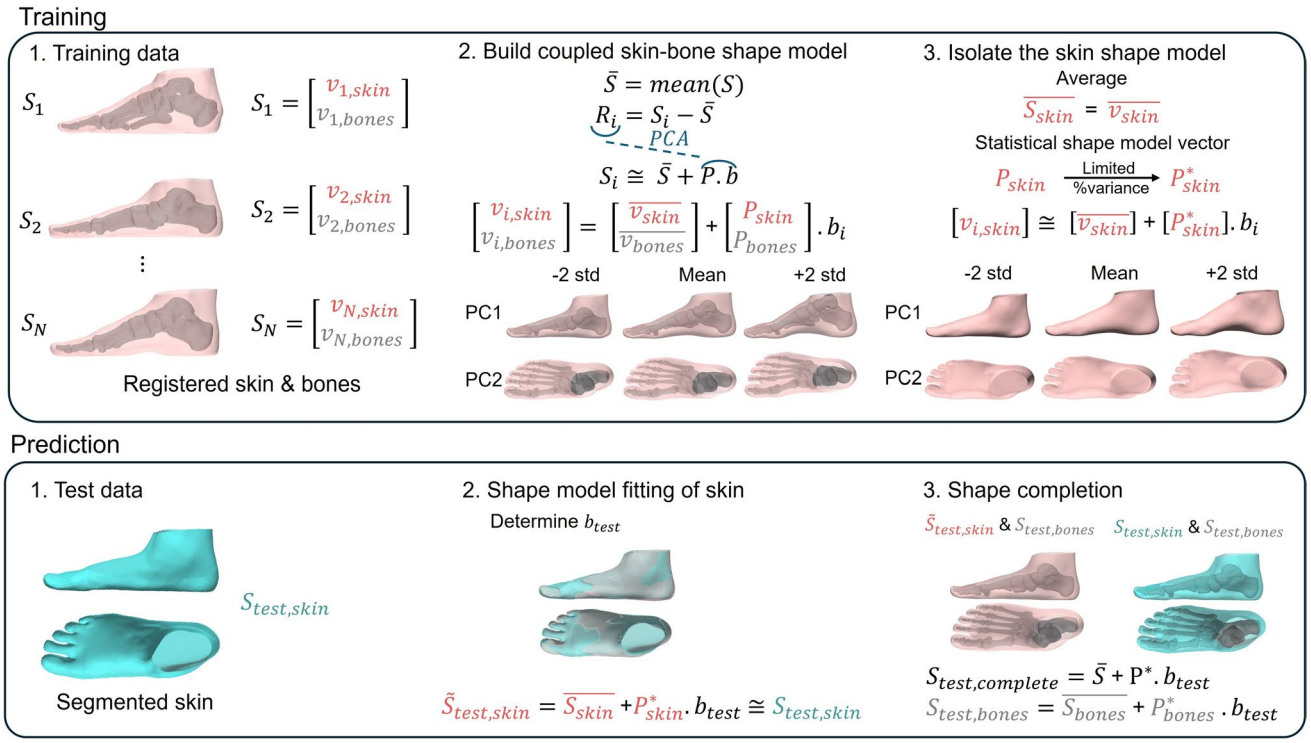

Fig. 3 Schematic overview illustrating the training and testing processes in an outer loop for bone prediction using shape model completion. The upper box illustrates the training process: (1) non-rigid registration of the skin and bones, (2) building a coupled statistical shape model of the skin and bones, and (3) isolating the skin portion of the statistical shape model. The lower box illustrates the testing process: (1) test data consists of the skin, (2) fitting the isolated statistical shape model of the skin to the segmented test skin, determining b_{test}, and (3) using the derived weight vector to complete the model with bones

optimize hyperparameters, a nested cross-validation procedure with 5 outer folds and 3 inner folds was used. Nested cross-validation uses two levels of data splitting: an outer loop and an inner loop. In the outer loop, the entire dataset is split in a number of outer folds (here: 5). Each outer fold serves once as an independent test set (containing approximately 28 samples for the feet dataset and 15–16 samples for the hands dataset), while the remaining folds are used in the inner loop. Concurrently, within each outer iteration, the remaining data are divided into a set number of inner folds (here: 3) to perform hyperparameter tuning. Each inner fold is rotated as the validation set while the others serve as the training set. This results in roughly 37 samples for validation and 74 for training in the feet dataset and about 21 for validation and 43 for training in the hands dataset. After hyperparameter optimization, the final training sets include approximately 112 samples for the feet and 64 for the hands. This method provides independent evaluation across 5 test sets while fine-tuning the hyperparameters in a separate inner loop. For the feet dataset, we kept paired left and right feet consistently within either the training, validation, or test set. Conversely, given the pronounced pose differences between the left and right hands, these were treated as separate entities in the hand dataset. Training samples ($n_{training}$) comprised registered skin and bone data as described in the previous section. To handle the extensive pose changes present in our dataset, a randomly selected case was used as a template within each training set instead of the traditional average. This method ensures an anatomically realistic foundation for the model, thereby avoiding smearing the features during initial alignment [11]. All skin data were then aligned to this template using Procrustes analysis, with the corresponding bone data being translated and rotated in the same manner. The skin and bone data are combined in one matrix (Eq. 1). A linear SSM encompassing both skin and bone was built from the aligned samples by (1) determining the mean shape (Eq. 2), (2) subtracting the mean shape from each sample (Eq. 3); and (3) applying principal component analysis (PCA) to the resulting residuals (Eq. 4). The linear SSM was constructed using the open-source Shape Model Builder MATLAB code (https://nl.mathworks.com/matlabcentral/fileexchange/49940-shape-model-builder) and presented in [5]. No eigenvalue threshold was used to reduce dimensionality, ensuring that all variance was initially captured in the model.

$$S_i = \begin{bmatrix} v_{i,skin} \\ v_{i,bones} \end{bmatrix} = \begin{bmatrix} X_{i,skin} \\ X_{i,bones} \\ Y_{i,skin} \\ Y_{i,bones} \\ Z_{i,skin} \\ Z_{i,bones} \end{bmatrix} \quad (1)$$

with S_i (size 3mx1, $m = m_{skin} + m_{bones}$) a sample from the training set combining the X, Y and Z coordinates (i.e., $X_{i,skin}$, $Y_{i,skin}$, $Z_{i,skin}$, $X_{i,bones}$, $Y_{i,bones}$, and $Z_{i,bones}$) of the skin vertices $v_{i,skin}$(size $3m_{skin}$x1) and the bone vertices $v_{i,bones}$. (size $3m_{bones}$x1), respectively.

$$\overline{S} = mean(S) = mean\left(\begin{bmatrix} v_{skin} \\ v_{bones} \end{bmatrix}\right) = \begin{bmatrix} \overline{v_{skin}} \\ \overline{v_{bones}} \end{bmatrix} \quad (2)$$

with $\overline{S}$ the mean shape of the training set (size 3mx1), consisting of the mean skin $\overline{v_{skin}}$ (size $3m_{skin}$x1) and the mean bones $\overline{v_{bones}}$. (size $3m_{bones}$x1)

$$R_i = S_i - \overline{S} \quad (3)$$

$$R_i = P.b_i \quad (4)$$

with R_i the residuals of a training set (size 3mx1), which can be represented by a product of the shape model vector P(size 3mx($n_{training}$-1)) and the weight vector b_i (size ($n_{training}$-1)x1). The weight vector consists of the coefficients transformed into the new coordinate system.

A sample from the training set can then be represented using the mean shape $\overline{S}$, the shape model vector P, and weight vector b_i (Eq. 5). The matrix P represents the fundamental directions of variation across our population. The weight vector b_i is a set of values that describe how each individual shape differs from the average shape or thus encodes the unique shape characteristics of an individual. When combining these P and b_i, the weight vector b_i explains the model how to adjust the average shape along these directions to create a precise representation of an individual's unique coupled skin-bone anatomy.

The shape model vector P can be decomposed into two components: one corresponding to the skin P_{skin} (size $3m_{skin}$ x($n_{training}$-1)) and the other to the bones P_{bones}(size $3m_{bones}$ x($n_{training}$-1)).

$$S_i = \begin{bmatrix} v_{i,skin} \\ v_{i,bones} \end{bmatrix} = \begin{bmatrix} \overline{v_{skin}} \\ \overline{v_{bones}} \end{bmatrix} + \begin{bmatrix} P_{skin} \\ P_{bones} \end{bmatrix}.b_i = \overline{S} + P.b_i \quad (5)$$

Next, the skin information is extracted from the combined shape model, yielding the mean skin shape $\overline{S_{skin}}$ (size $3m_{skin}$x1) and the corresponding model vector P_{skin} (size $3m_{skin}$x($n_{training}$-1)). The proportion of variance to be retained in the shape model is a controllable parameter which is optimized in the cross-validation by minimizing the overall Mean Absolute Error (MAE) between the predicted and ground truth bones. Varying this proportion results in a different number of modes ($nr_{ofshapemodes}$) being kept for subsequent prediction. The resulting shape model vector is denoted by P^*_{skin}(size $3m_{skin}$x $nr_{ofshapemodes}$).

For each sample in the validation or test set, the segmented skin was fitted to the reduced skin shape model to determine its weight vector (i.e., $b_{validation}$ or b_{test} (size $nr_{ofshapemodes}$x1)) (Eq. 6). Subsequently, this weight vector was applied to the combined skin-bone shape model to predict the morphology of the accompanying bones (Eqs. 7, 8).

$$\tilde{S}_{test,skin} = \overline{S_{skin}} + P^*_{skin}.b_{test} \cong S_{test,skin} \quad (6)$$

with $S_{test,skin}$ the skin of the validation or test case (size $3m_{skin}$ x1), b_{test} the weight vector of the skin, P^*_{skin} the shape model vector of the reduced shape model, $\overline{S_{skin}}$ the mean skin, and $\tilde{S}_{test,skin}$ the approximation of $S_{test,skin}$. Similarly, the complete approximated shape of the validation or test case can be written as

$$S_{test,complete} = \overline{S} + P^*.b_{test} \quad (7)$$

with $\overline{S}$ the mean shape (combined skin and bones), P^* the reduced shape model vector of the combined skin-bone shape model (size 3mx $nr_{ofshapemodes}$), and b_{test} the estimated weight vector of the validation/test case. Finally, the approximated bone morphology of the validation or test case is given by

$$S_{test,bones} = \overline{S_{bones}} + P^*_{bones}.b_{test} \quad (8)$$

with $\overline{S_{bones}}$ the mean morphology of the bones (size $3m_{bones}$x1), P^*_{bones} the reduced shape model vector of the bone shape model (size $3m_{bones}$x $nr_{ofshapemodes}$), and b_{test} the estimated weight vector of the validation/test case.

All was implemented in MATLAB v2023a (MathWorks, Natick, MA, USA).

Prediction of Bony Morphology from the Skin Surface Simulating Lower-Quality Input

To simulate real-world clinical data, such as that from video or LIDAR scans, two methods of data degradation were employed on the test sets of feet dataset. First, the number of vertices were reduced by 40%, which subsequently increased the average edge length from 1.16 mm to 1.73 mm. Second, Gaussian random noise was added such that 50% of the remaining vertices received a noise amplitude of ≥ 1 mm. This specific value was chosen because it aligns with the typical accuracy observed in clinical foot scans [2, 34]. The prediction models of the feet (from the corresponding cross-validation folds) were then tested against the reduced quality CT segmentations.

Evaluation Metrics

To evaluate the accuracy of the prediction method, bone shapes were predicted from the skin surfaces in the test set and compared against the ground truth. The resulting prediction errors were analyzed by calculating the point-to-surface distance between the predicted and ground truth bones (Absolute error) and were further partitioned into pose and anatomy errors. The pose error indicate the misalignment between the predicted bone and the ground truth. To avoid the influence of the shape discrepancies, the morphology of the predicted bone is replaced with its true morphology. To do so, Procrustes analysis is employed to align the true morphology with the predicted bone. Finally, leveraging the mesh correspondence, the pose error is quantified by averaging the pointwise distance per bone across all vertices between the ground truth bone and the true morphology at the predicted location. An additional analysis was conducted to quantify the contribution of translational and rotational errors to the overall pose estimation. This decomposition is described in full detail in the Appendix A. The anatomy error indicates the shape discrepancies. To calculate the anatomy error, the predicted bone is repositioned by applying Procrustes analysis to the location and orientation of the ground truth bone. Subsequently, the anatomy errors were assessed using point-to-surface distance between the correct morphology and the predicted morphology. In addition to evaluating bone prediction errors, the error in the reconstructed skin was quantified as the point-to-surface distance between the fitted and segmented skin.

Results

Feet

The nested cross-validation analysis demonstrated two optimal variance retention values (i.e., 90% and 95%) across the different outer folds. Detailed results for each fold, along with the computed mean across all folds, are shown in Fig. 4. Table 1 presents the average and standard deviation of evaluation metrics for using both the CT segmentations as well as the reduced quality CT segmentations as the input. As a visual example of our predictions,

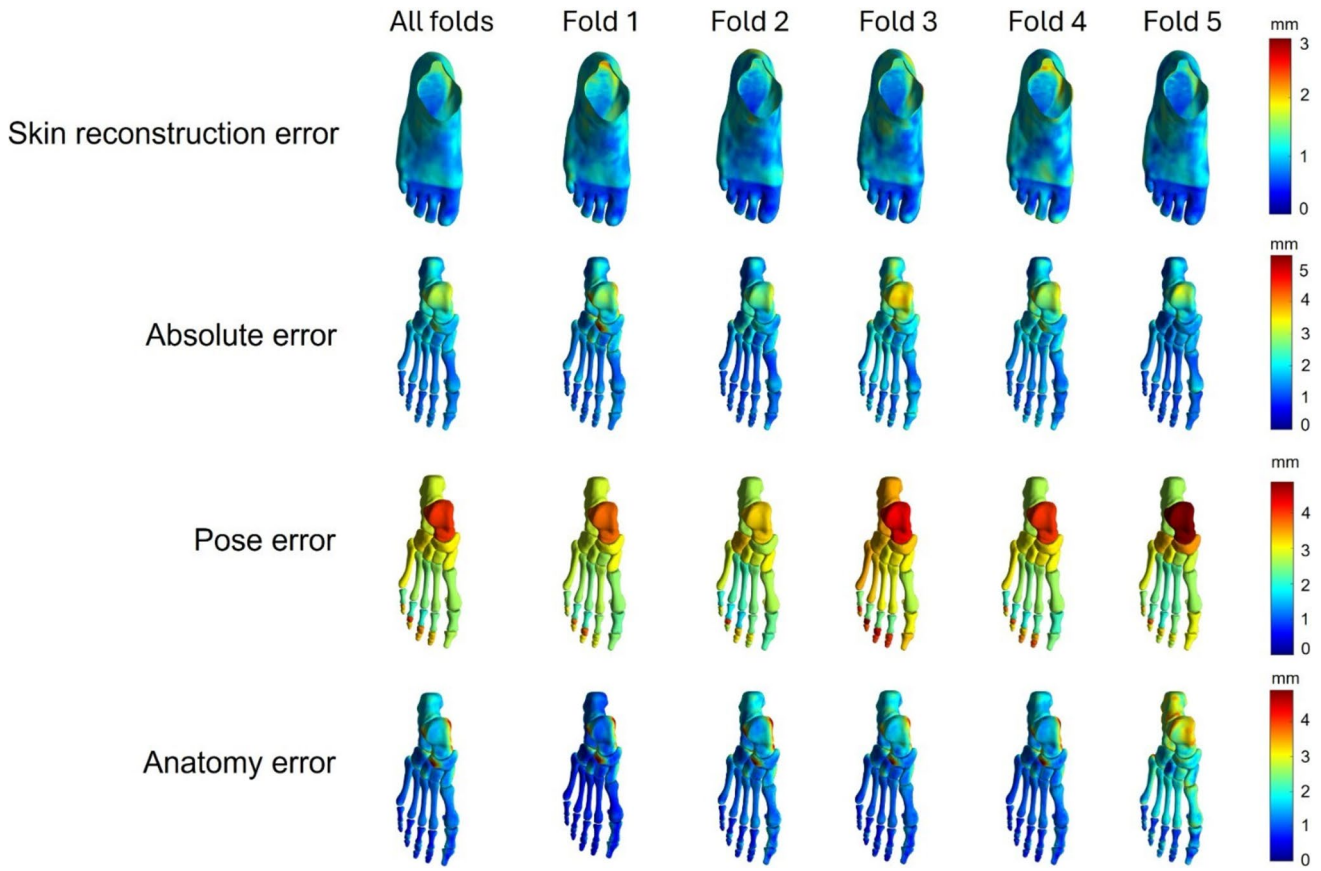

Fig. 4 The mean of the skin reconstruction error, the mean absolute error, the mean pose error, and the mean anatomy error over all folds and folds 1–5 for the feet dataset

Table 1 An overview of the average and standard deviation of the skin reconstruction error, mean absolute error, pose error, and anatomy error for the five different folds for the feet dataset using both CT segmentations and the reduced quality segmentations as an input

	Input	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Optimal variance retention (%)		90	90	95	90	90
Skin reconstruction error (mm)	CT segmentation	1.01 (0.78)	0.95 (0.69)	1.00 (0.82)	1.03 (0.75)	0.93 (0.69)
	Lower quality	1.01 (0.72)	0.98 (0.64)	1.00 (0.72)	1.03 (0.70)	0.95 (0.64)
MAE (mm)	CT segmentation	1.72 (1.39)	1.55 (1.10)	1.93 (1.62)	1.69 (1.35)	1.48 (1.04)
	Lower quality	1.80 (1.48)	1.64 (1.20)	2.09 (1.70)	1.77 (1.40)	1.63 (1.18)
Pose error (mm)	CT segmentation	2.82 (1.31)	2.71 (1.28)	3.37 (1.71)	2.90 (1.78)	2.89 (2.15)
	Lower quality	3.07 (1.45)	2.89 (1.35)	3.66 (1.85)	3.17 (1.86)	3.20 (2.21)
Anatomy error (mm)	CT segmentation	1.07 (1.02)	0.94 (0.71)	1.15 (1.02)	0.99 (0.77)	0.84 (0.60)
	Lower quality	1.10 (1.11)	1.02 (0.84)	1.23 (1.11)	1.00 (0.78)	0.87 (0.64)

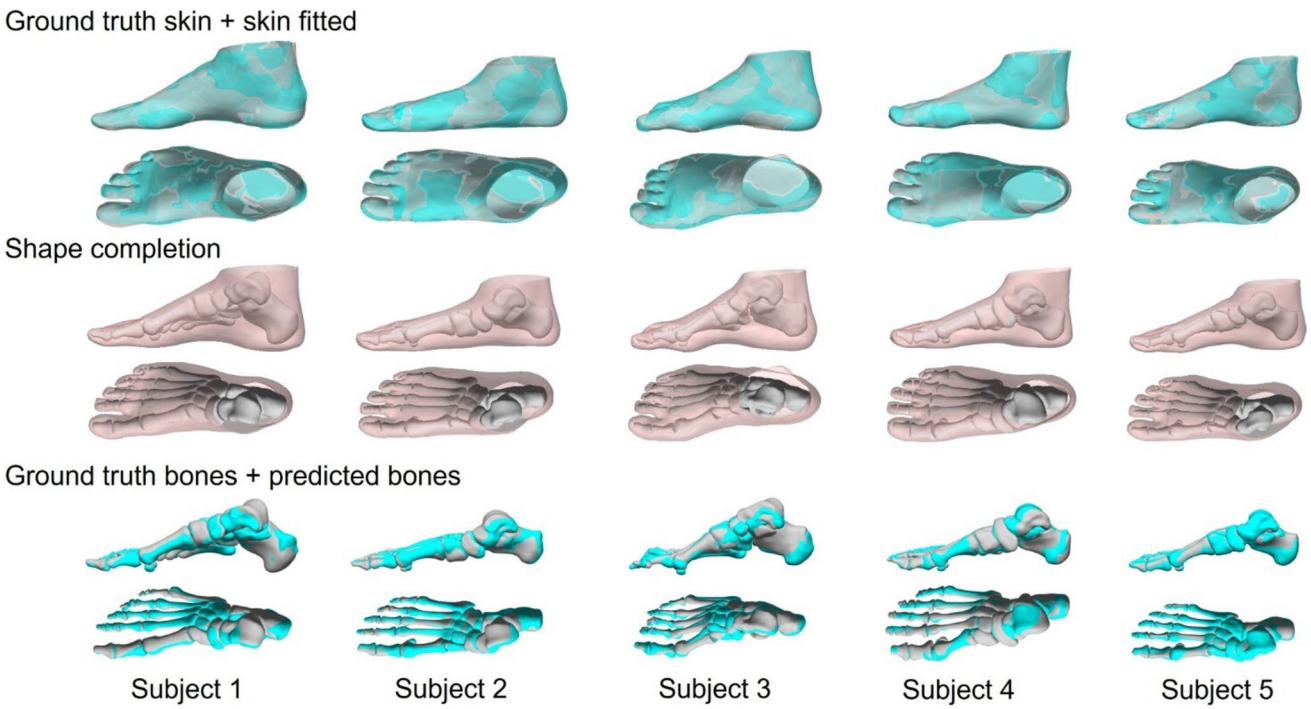

Fig. 5 Medial and superior views of the feet of five representative subjects from the WBCT dataset during the process of shape completion. The first two rows display the ground truth skin (cyan) and the fitted skin (pink). The third and fourth row show the completed shape including the skin fit (pink) and the predicted bones (gray). The bottom two rows present the ground truth and predicted bones, shown in cyan and gray, respectively

Fig. 5 illustrates the skin fit, the completed shape, and a comparison between the predicted and ground truth bones for representative subjects using the CT segmentations as an input. The skin reconstruction errors ranged between 0.01 mm and 12.34 mm, whereas the Mean Absolute Error (MAE) of the bones ranged from 1.07 mm to 3.01 mm with an average of 1.68 mm. The largest pose errors were observed in the hind foot and the middle and distal phalanges. The additional analysis (Appendix A) quantified the contributions of the translational and rotational components to the pose estimation separately. Both the translational and rotational error distributions were symmetrical. Meanwhile, the most noteworthy anatomical discrepancies were localized on (1) the calcaneum, more specifically on the joint surfaces and sustentaculum tali, (2) the talus, particularly on the medial and lateral flank, (3) the medial and lateral side of the navicular bone, and (4) the joint surfaces of the cuboid. Building the coupled skin-bone statistical shape model required around 4 seconds. Predicting the bony anatomy for a single test case required about 60 milliseconds.

Hands

The nested cross-validation analysis identified an optimal retained variance between 99.9% and 99.95% across all folds. Figure 6 and Table 2 provide a detailed breakdown of these results per fold, along with the overall mean. The skin reconstruction error exhibited a mean of 1.37 mm, with a range spanning from 0.02 mm to 22.75 mm. The greatest skin reconstruction errors, along with the largest absolute and pose errors, were observed at the fingertips and at the edges of the cut. The additional analysis (Appendix A) involved decomposing the pose estimation error into its translational and rotational components; the resulting individual error distributions were found to be symmetrical. The highest anatomy errors were noted at the top and the base of the metacarpals and phalanges. As a visual example of our predictions, Fig 7 illustrates the skin fit, the completed shape, and a comparison between the predicted and ground truth bones for the representative subjects using the CT segmentations as an input. Constructing the coupled skin-bone statistical shape model took roughly 1 second, while predicting the bony anatomy for one test case required about 10.

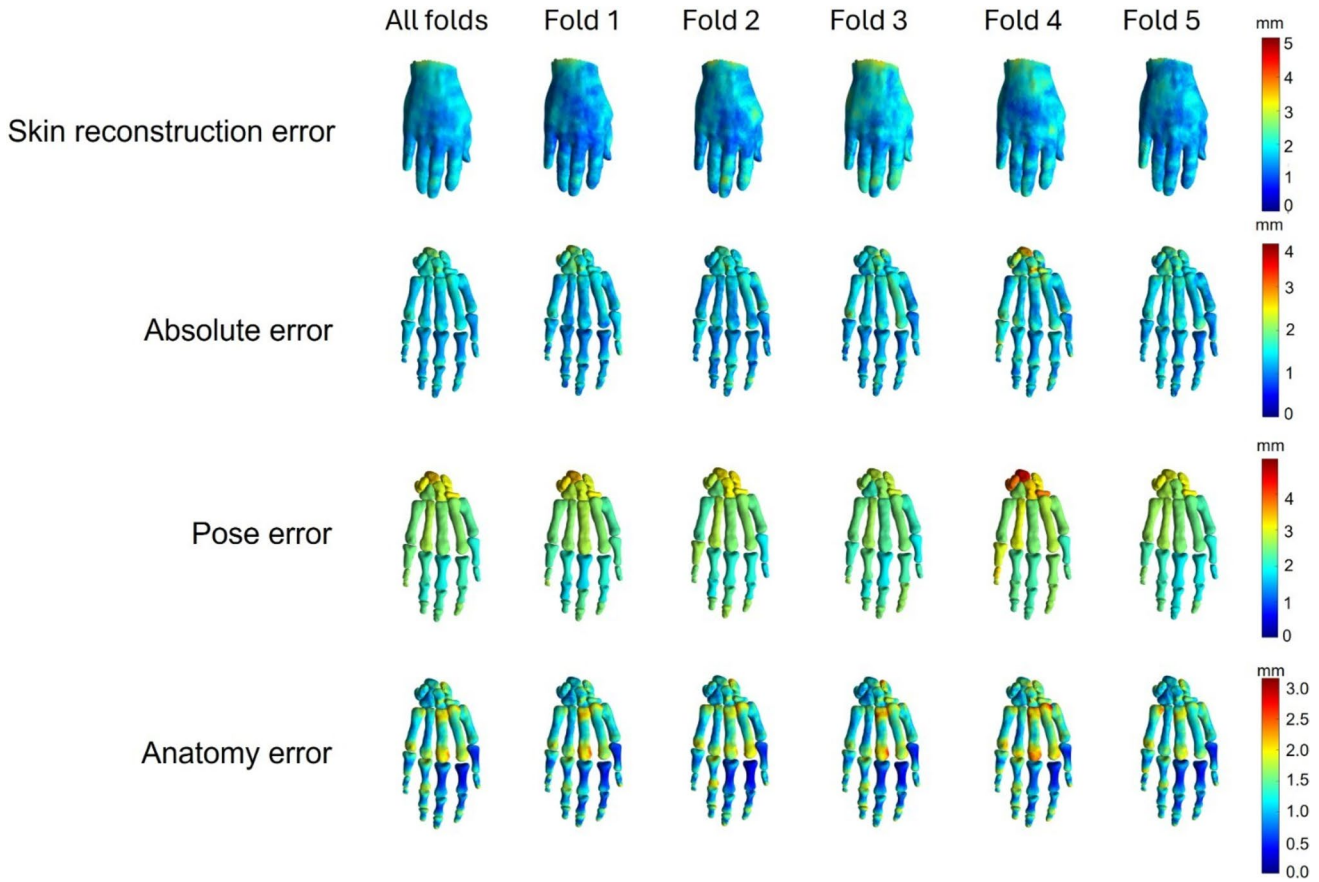

Fig. 6 The mean of the skin reconstruction error, the mean absolute error, the mean pose error, and the mean anatomy error over all folds and folds 1–5 for the hand dataset

Table 2 An overview of the average and standard deviation of the skin reconstruction error, mean absolute error, pose error, and anatomy error for the five different folds for the hand dataset

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Optimal variance retention (%)	99.95	99.95	99.9	99.95	99.95
Skin reconstruction error (mm)	1.53 (0.91)	1.55 (0.93)	1.78 (1.08)	1.69 (1.13)	1.51 (0.87)
MAE (mm)	1.36 (0.96)	1.34 (0.90)	1.33 (0.86)	1.50 (1.05)	1.34 (0.89)
Pose error (mm)	2.51 (1.09)	2.50 (1.08)	2.41 (0.89)	2.82 (1.32)	2.42 (1.03)
Anatomy error (mm)	0.88 (0.56)	0.86 (0.55)	0.88 (0.56)	0.95 (0.63)	0.84 (0.53)

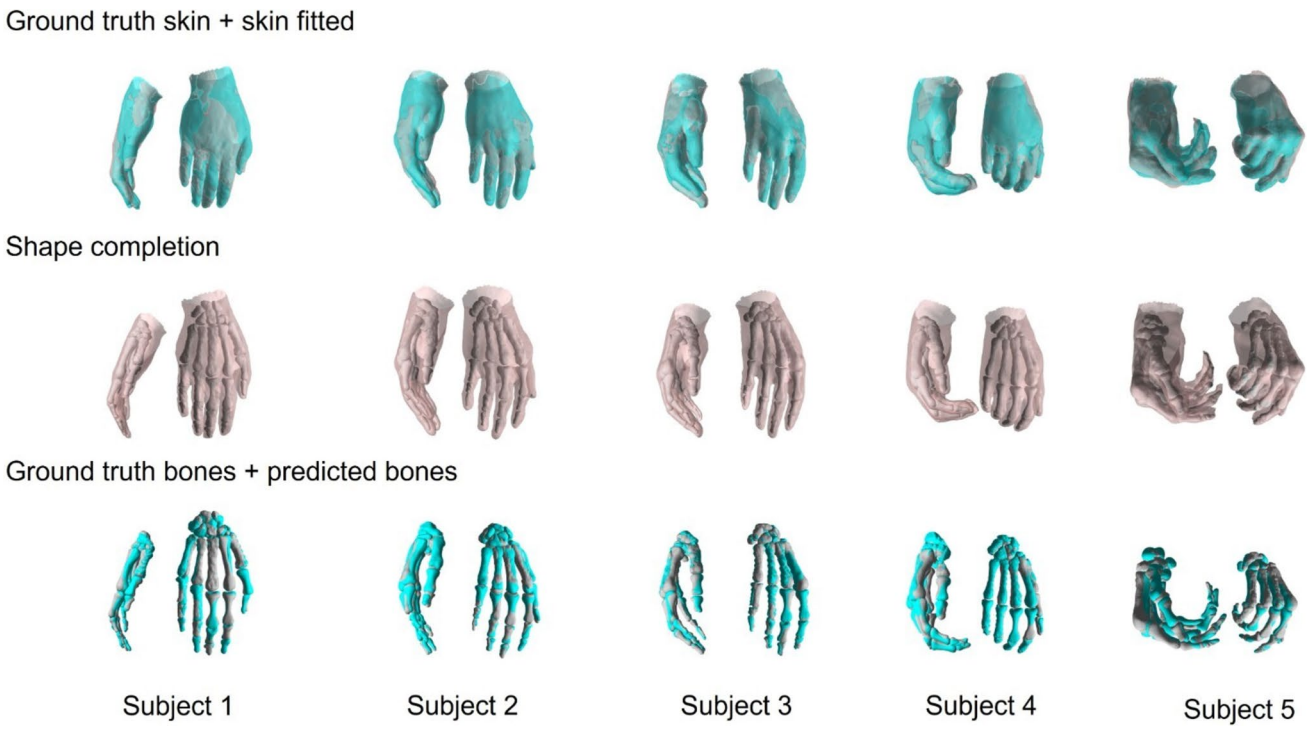

Fig. 7 Radial and dorsal views of the hands of five representative subjects from the hand dataset during the process of shape completion. The first row displays the ground truth skin (cyan) and the fitted skin (pink). The second row shows the completed shape including the skin fit (pink) and the predicted bones (gray). The bottom row presents the ground truth and predicted bones, shown in cyan and gray, respectively

Discussion

This study investigated the potential of shape completion within linear statistical shape modeling to predict bone morphology, especially for the hands and feet, from skin surface data. Achieving high accuracy, with MAEs, respectively, of 1.68 mm and 1.37 mm for the feet and hands, validates this method's feasibility. The presented approach therefore offers a potentially fast, non-invasive and accessible alternative to traditional imaging for musculoskeletal assessments, opening avenues for future development.

The proposed method predicts bone structures by treating them as a missing component in an SSM constructed from a combined skin-bone dataset. To achieve this, the skin component of the model is fitted to an unseen skin sample. The derived weight vector is then used to complete the model with bones. As a consequence, this approach is highly dependent on the fitting of the outer layer, i.e., the skin. The goodness of this skin fitting, in turn, is influenced by the amount of variance retained within the model, a hyperparameter optimized through nested cross-validation. Our results showed that the hand dataset required a higher retained variance than the feet dataset. Despite this, the skin reconstruction errors were slightly higher for hand skins than for the foot skins. The largest skin reconstruction errors in the hands were located at the fingertips and near the edges of the cut. The fingertip error is a direct consequence of the challenges in fitting the skin to diverse bent hand poses. For the feet, similar maximal errors were observed at the hindfoot near the cut edges. This edge error is likely caused by the abrupt boundary created during the template cutting process at a fixed height. Non-rigid registration is an ill-posed problem (i.e., it does not have a unique solution). This under-constrained nature is most pronounced at boundaries where there is a lack of geometric information to guide the deformation, leading to non-informative variance [18, 33]. As a consequence, the skin registrations near the edges are less reliable and therefore causes non-informative variance in the statistical shape model. This not only results in higher skin fitting errors but also in lower positional and anatomical accuracy. For both the feet and the hands, the translational and rotational error components of the pose error showed a symmetric distribution. This symmetrical pattern indicates that the observed pose inaccuracies are non-systematic and are not the result of an inherent model bias. Furthermore, reducing the quality of the feet dataset during the testing phase, only had a sub-millimetric effect on the evaluation metrics, suggesting a robust model.

To the best of our knowledge, no previous models have predicted detailed hand and/or feet bones from the skin

surface, making direct comparisons challenging. However, related research work exists. For instance, HIT performs a binary prediction of the combined long bones given the skin surface and reported median absolute point-to-surface errors of 10 mm (females) and 11 mm (males) for this prediction [20]. Dakri et al. used the HIT dataset and extended SKEL to predict the individual long bones (humerus, radius-ulna, pelvis, femur, tibia-fibula) from the skin surface [9]. They showed mean absolute point-to-surface errors ranging from 4 mm to 18 mm for individual long bone prediction. Shetty et al. introduced BOSS, an SSM encompassing bones, organs, and skin [32]. The model consisted of 70 segments, treating both hands and feet as a single unit. BOSS achieved a mean absolute error of 8.68 mm in predicting complete skeletal morphology from the skin surface. Unlike our approach, BOSS applied pose normalization to the skin, skeleton, and organs, which reduces articulation noise and enhances accuracy but restricts bone prediction to specific poses of the skin. While the results of our feet and hand models align with and even surpass the performance of above reported methods, a direct comparison is challenging. This is because Keller et al. and Dakri et al. focused their analysis on long bones and Shetty et al., while studying the entire skeleton, did so with a reduced set of segments and modeled the hands and feet each as a whole unit. Additionally, Shetty et al. did not elaborate extensively on their method for inferring bone structure from skin surfaces (i.e., number of shape modes, pose normalization during skin fitting, and the completeness of bone prediction).

This study offers valuable insights into predicting bone morphology and position using shape completion within statistical shape models; however, it is essential to acknowledge some inherent limitations. There are notable constraints in the datasets themselves. The feet data, while benefiting from standing posture capture, comes from an orthopedic population, not healthy individuals. Conversely, the hand data are constrained by its low resolution, affecting anatomical detail, and its cadaveric origin, although the effect of the hand muscle contraction is presumed to be limited. The models' ability to generalize is dependent on the specifics of the training data. Consequently, while potentially useful for healthy individuals, these kinds of models cannot predict conditions like broken bones or other clinical anomalies. By extending the training data to encompass several clinical pathological cases, these types of models would be valuable in (initial) screening and quantification of bone and joint deviations such as hallux valgus, pes planus and claw hands particularly through integration with wearable 3D scanners. To partially address real-world data issues, a sub study was performed to assess model robustness by degrading input data quality through increased edge length and added noise. Due to the scope of the current work, only one noise type was investigated. Subsequent research will broaden this analysis to include multiple noise modalities and simulate realistic clinical challenges such as occlusion between the toes, a common artifact in foot scanning. Furthermore, extending this work to build a regression model that correlates the statistical shape model of the bones with simple anthropometric measurements, like foot/hand length and width, would be a valuable direction for future study, particularly for low-cost clinical applications. Moreover, a notable characteristic of the current models is their reliance on static input data. Although such data is valuable—and our hand dataset effectively captures a wide range of poses—it does not fully represent the dynamic nature of the musculoskeletal system, a crucial element for clinical investigation. This constraint is, however, shared with traditional imaging modalities. To address this, future efforts will aim to integrate a strategy similar to Zhu et al., which optimizes the skeletal structure estimation over multiple poses and utilizes the anatomical model for tracking an input motion sequence [37]. Finally, the created models were developed exclusively on foot and hand data and, consequently, can only estimate the bones within those two regions. Future work will focus on scaling this validated framework into a full-body prediction method. The complete model will be released to support broad research applications.

Beyond the data limitations, the shape completion method itself also poses its own constraints. This work employs a linear SSM, built using PCA, to demonstrate the feasibility of predicting bone structures in hands and feet via shape completion. While this approach shows promise, it remains hampered by its linearity, especially with articulated shapes in diverse poses. A wide range of (nonlinear) statistical shape modeling techniques exist which could improve the handling of pose in the data [11, 12, 24, 28, 36]. Future work includes the exploration of these nonlinear statistical shape modeling techniques to predict the bones from the skin. Furthermore, pose normalization could be explored to improve the compactness of the SSMs. Pose normalization techniques have been applied to articulated bones, to skins and to skin-bone combinations [4, 10, 27, 32]. However, a significant challenge remains: to the best of our knowledge, there is currently no validated repositioning method for accurately reposing both bones and skin. In addition, fitting a pose normalized SSM to an unseen posed sample requires expensive optimization. Beyond statistical shape modeling techniques, alternative shape completion methods leveraging deep generative models including generative adversarial networks, and diffusion models have emerged and could be investigated for their ability to predict bones from the skin surface [8, 23, 35].

In conclusion, this paper presents a novel approach that directly predicts underlying skeletal structure from skin surfaces using shape completion within statistical shape models. The feasibility of this approach was demonstrated on a

foot and a hand dataset, achieving high accuracy in predicting both the shape and position of the individual bones. This research establishes a foundation for further development of fast and non-invasive diagnostic techniques that could provide a cost-effective and accessible alternative to conventional imaging.

Appendix

A Pose estimation errors were further analyzed by decomposing them into translational and rotational components along and around the anatomical axes. To facilitate this analysis, local coordinate systems were established for each bone of the foot and hand, as visualized in Figures A1 and A4, respectively. The pose error for a given bone was then defined as the rigid-body transformation that maps the local coordinate system of the ground truth bone to the coordinate system of the predicted bony location. This transformation was determined using the 'procrustes' in MATLAB v2023a (MathWorks, Natick, MA, USA). The resulting rotation matrix was then converted into Euler angles using the 'ZYX' rotation order. Boxplots of the absolute translation and rotation errors are presented in Figures A2 (foot translation), A3 (foot rotation), A5 (hand translation), and A6 (hand rotation). These plots show the distribution of errors along and around these axes across all folds of the cross-validation. For both the feet and the hands, the translational and rotational error components of the pose error showed a symmetric distribution. This symmetrical pattern indicates that the observed pose inaccuracies are non-systematic and are not the result of an inherent model bias.

See Figs. 8, 9, 10, 11, 12 and 13.

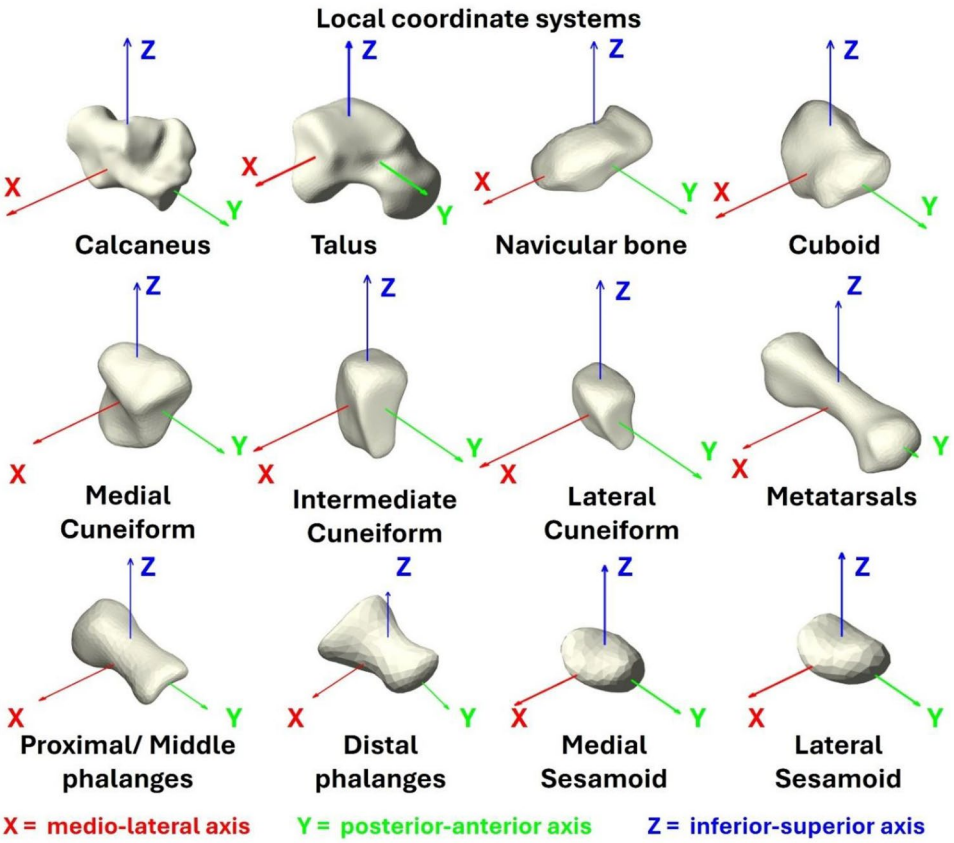

Fig. 8 Local coordinate systems for the different foot bones

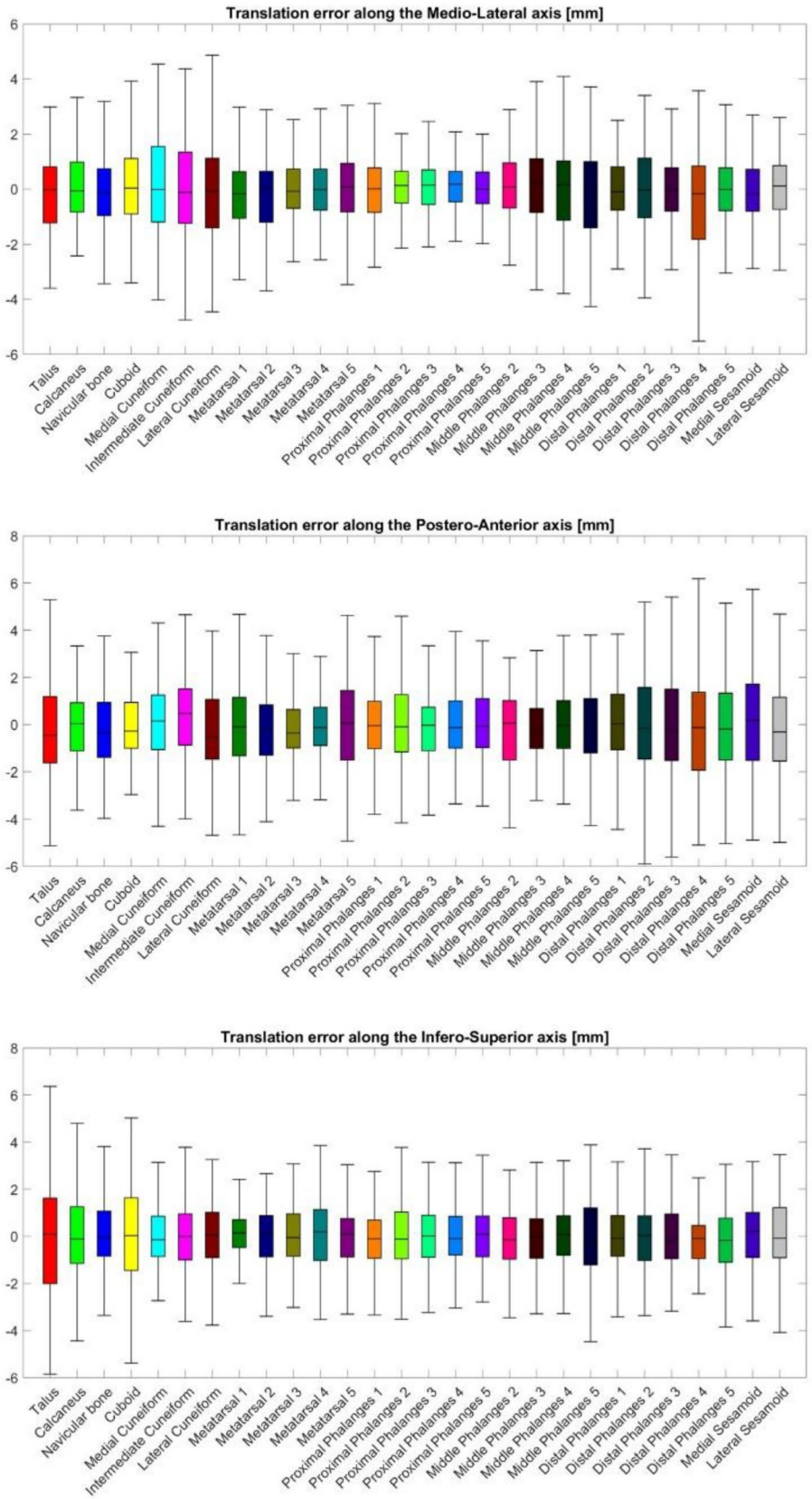

Fig. 9 Translation errors along the medio-lateral, postero-anterior, and infero-superior axis for the foot bone prediction

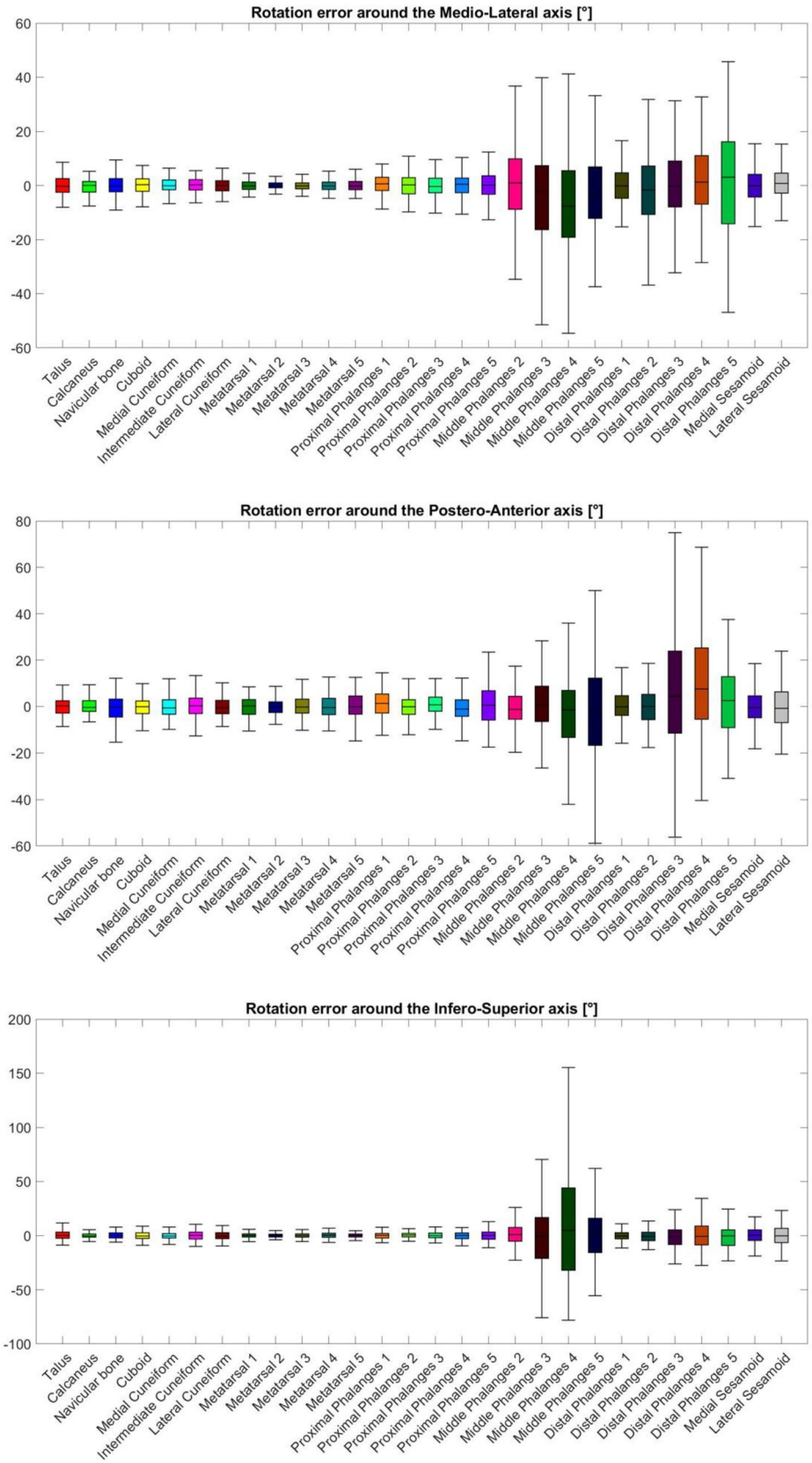

Fig. 10 Rotation errors around the medio-lateral, postero-anterior, and infero-superior axis for the foot bone prediction

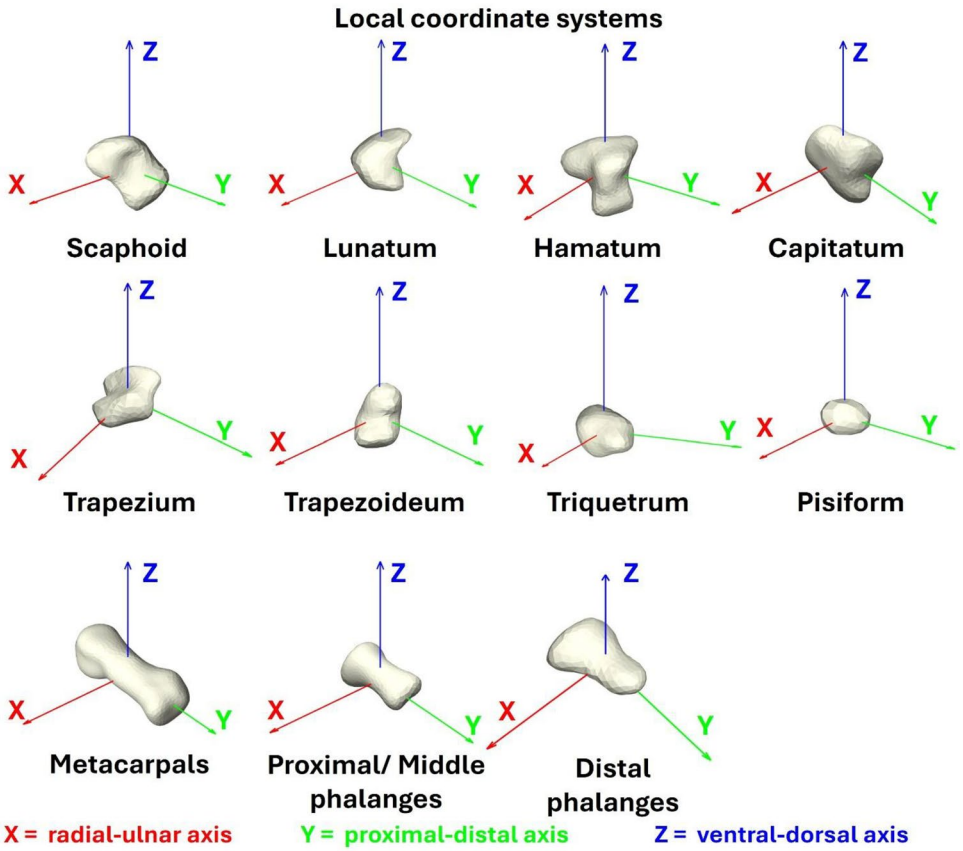

Fig. 11 Local coordinate systems for the different hand bones

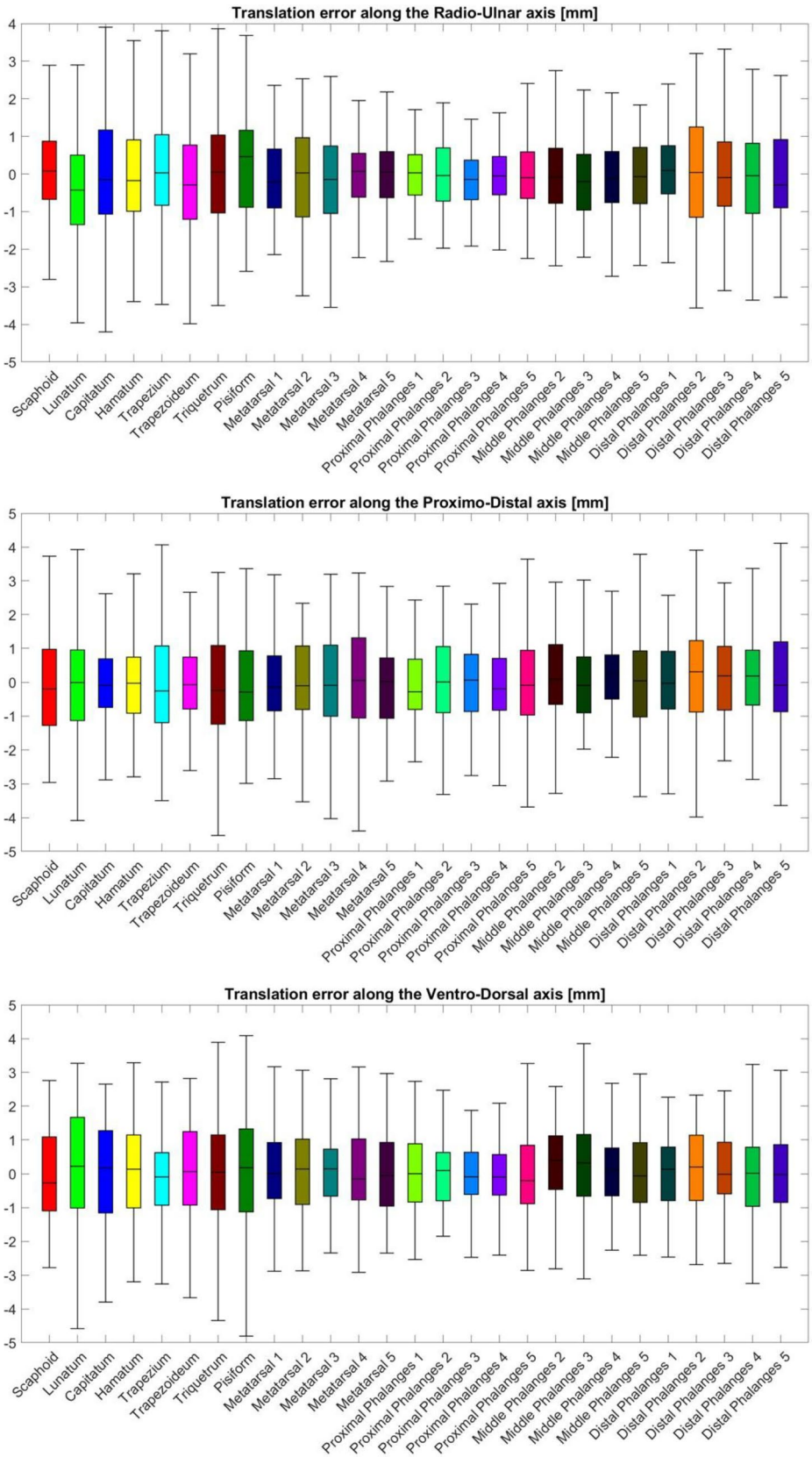

Fig. 12 Translation errors along the radio-ulnar, proximo-distal, and ventro-dorsal axis for the hand bone prediction

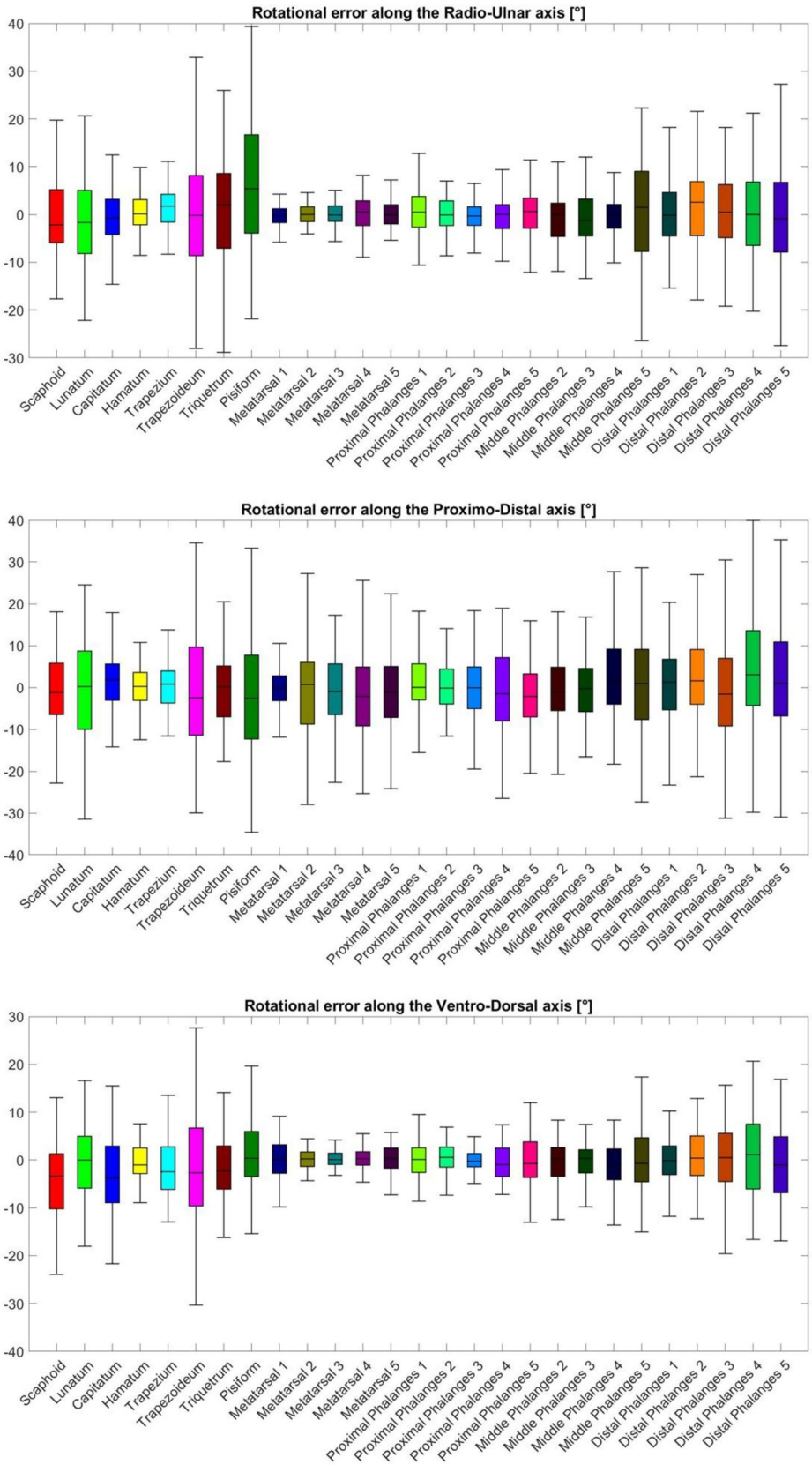

Fig. 13 Rotation errors around the radio-ulnar, proximo-distal, and ventro-dorsal axis for the hand bone prediction

Author Contributions All authors contributed to the study conception and design. Conceptualization was performed by Kate Duquesne, Emmanuel Audenaert, Jan Sijbers, and Wim Van Paepegem. Material preparation, data collection, and analysis were performed by Kate Duquesne, Adris Molnar, Jing Li, Aline Van Oevelen, and Roel Huysentruyt. The first draft of the manuscript was written by Kate Duquesne and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by two Aspirant Grants from the Research Foundation-Flanders (#1137723N, #1122821N, FWO) and by a Senior Clinical Investigator Fellowship Grant from the Research Foundation-Flanders (#1842619N, FWO).

Data Availability The models are available upon request to the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical Approval and Consent to Participate The study was conducted in accordance with the Declaration of Helsinki 2013, and approved by the University Hospital Ghent (B6702022000639).

Informed Consent An informed consent was obtained from all subjects involved in gathering the Weight-Bearing Computed Tomography foot data.

References

1. Ali-Hamadi, D., T. Liu, B. Gilles, et al. Anatomy transfer. *ACM Trans Graph*. 32:188, 2013. https://doi.org/10.1145/2508363.2508415.
2. Allan, J. J., S. E. Munteanu, D. R. Bonanno, et al. Methodological and statistical approaches for the assessment of foot shape using three-dimensional foot scanning: a scoping review. *J Foot Ankle Res*. 16:24, 2023.
3. Ambellan F, Lamecker H, von Tycowicz C, Zachow S (2019) Statistical Shape Models: Understanding and Mastering Variation in Anatomy. In: Biomedical Visualisation. pp 67–84
4. Audenaert, E. A., J. Van den Eynde, D. F. Almeida, et al. Separating positional noise from neutral alignment in multicomponent statistical shape models. *Bone Rep*.12:100243, 2020. https://doi.org/10.1016/j.bonr.2020.100243.
5. Audenaert, E. A., J. Van Houcke, D. F. Almeida, et al. Cascaded statistical shape model based segmentation of the full lower limb in CT. *Comput Methods Biomech Biomed Engin*. 22:644–657, 2019. https://doi.org/10.1080/10255842.2019.1577828.
6. Balasubramanian, M., and P. Sheykhmaleki. Emerging roles of 3D body scanning in human-centric applications. *Technologies (Basel)*. 13:126, 2025. https://doi.org/10.3390/technologies13040126.
7. Cai, X., Y. Wu, J. Huang, et al. Application of statistical shape models in orthopedics: a narrative review. *Intelligent Medicine*. 4:249–255, 2024. https://doi.org/10.1016/j.imed.2024.05.001.
8. Chu R, Xie E, Mo S, et al (2023) DiffComplete: Diffusion-based Generative 3D Shape Completion. In: NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. pp 75951–75966
9. Dakri A, Arora V, Challier L, et al (2024) On predicting 3D bone locations inside the human body. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp 336–346
10. Danckaers F, Huysmans T, Hallemans A, et al (2017) Full Body Statistical Shape Modeling with Posture Nor-malization. In: Advances in Human Factors in Simulation and Modeling . pp 437–448
11. Duquesne, K., N. Nauwelaers, P. Claes, and E. Audenaert. Principal polynomial shape analysis: a non-linear tool for statistical shape modeling. *Comput Methods Programs Biomed*.220:106812, 2022. https://doi.org/10.1016/J.CMPB.2022.106812.
12. Fouefack J-R, Borotikar B, Douglas TS, et al (2020) Dynamic multi-object Gaussian process models: A framework for data-driven functional modelling of human joints. In: Computer Vision and Pattern Recognition
13. Grant, T. M., L. E. Diamond, C. Pizzolato, et al. Development and validation of statistical shape models of the primary functional bone segments of the foot. *PeerJ*. 8:e8397, 2020. https://doi.org/10.7717/peerj.8397.
14. Guarnieri Lopez, M., K. L. Matthes, C. Sob, et al. Associations between 3D surface scanner derived anthropometric measurements and body composition in a cross-sectional study. *Eur J Clin Nutr*. 77:972–981, 2023. https://doi.org/10.1038/s41430-023-01309-4.
15. Hussain, S., I. Mubeen, N. Ullah, et al. Modern diagnostic imaging technique applications and risk factors in the medical field: a review. *Biomed Res Int*. 2022:5164970, 2022.
16. Huysentruyt, R., I. Van den Borre, S. Lazendić, et al. Sample size effect on musculoskeletal segmentation: how low can we go? *Electronics (Basel)*. 13:1870, 2024. https://doi.org/10.3390/electronics13101870.
17. Isensee, F., P. F. Jaeger, S. A. A. Kohl, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 18:203–211, 2021. https://doi.org/10.1038/s41592-020-01008-z.
18. Jiang Y, Edmiston E, Wang F, et al (2009) Shape comparison using perturbing shape registration. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009. IEEE Computer Society, pp 683–690
19. Kadleček, P., A. E. Ichim, T. Liu, et al. Reconstructing personalized anatomical models for physics-based body animation. *ACM Trans Graph*. 35:213, 2016. https://doi.org/10.1145/2980179.2982438.
20. Keller M, Arora V, Dakri A, et al (2024) HIT: Estimating Internal Human Implicit Tissues from the Body Surface. In: Conference: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp 3480–3490
21. Keller, M., K. Werling, S. Shin, et al. From skin to skeleton: towards biomechanically accurate 3D digital humans. *ACM Trans Graph*. 42:253, 2023. https://doi.org/10.1145/3618381.
22. Keller M, Zuffi S, Black MJ, Pujades S (2022) OSSO: Obtaining Skeletal Shape from Outside. In: Conference: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp 20460–20469
23. Khademi W, Fuxin L (2023) Diverse Shape Completion via Style Modulated Generative Adversarial Networks. In: NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems. pp 79279–79292
24. Khan, N., A. C. Peterson, B. Aubert, et al. Statistical multi-level shape models for scalable modeling of multi-organ anatomies. *Front Bioeng Biotechnol*. 2023. https://doi.org/10.3389/fbioe.2023.1089113.
25. Kushwaha, M., J. Choudhary, and D. P. Singh. 3DPMesh: an enhanced and novel approach for the reconstruction of 3D human meshes from a single 2D image. *Comput Graph*.119:103894, 2024. https://doi.org/10.1016/j.cag.2024.103894.

26. Li X, Meng M, Wu Z, et al (2024) Human Mesh Recovery from Arbitrary Multi-view Images. In: Computer Vision and Pattern Recognition
27. Loper M, Mahmood N, Romero J, et al (2015) SMPL: A Skinned Multi-Person Linear Model. ACM Trans Graph 34:248:1-248:16. https://doi.org/10.1145/2816795.2818013
28. Ma, J., A. Wang, F. Lin, et al. A novel robust kernel principal component analysis for nonlinear statistical shape modeling from erroneous data. *Computerized Medical Imaging and Graphics*.77:101638, 2019. https://doi.org/10.1016/j.compmedimag.2019.05.006.
29. Melinska, A. U., P. Romaszkiewicz, J. Wagel, et al. Statistical shape models of cuboid, navicular and talus bones. *J Foot Ankle Res*. 10:6, 2017. https://doi.org/10.1186/s13047-016-0178-x.
30. Munyebvu, T. A., C. D. Metcalf, C. B. Burson-Thomas, et al. OpenHands: an open-source statistical shape model of the finger bones. *Ann Biomed Eng*. 52:2975–2986, 2024. https://doi.org/10.1007/s10439-024-03560-7.
31. Rusli, W. M. R., and A. E. Kedgley. Statistical shape modelling of the first carpometacarpal joint reveals high variation in morphology. *Biomech Model Mechanobiol*. 19:1203–1210, 2020. https://doi.org/10.1007/s10237-019-01257-8.
32. Shetty, K., A. Birkhold, S. Jaganathan, et al. BOSS: Bones, organs and skin shape model. *Comput Biol Med*.165:107383, 2023. https://doi.org/10.1016/j.compbiomed.2023.107383.
33. Sotiras A, Davatzikos C, Paragios N (2012) Deformable Medical Image Registration: A Survey
34. Wan, F. K. W., K. L. Yick, and W. W. M. Yu. Validation of a 3D foot scanning system for evaluation of forefoot shape with elevated heels. *Measurement (Lond)*. 99:134–144, 2017. https://doi.org/10.1016/j.measurement.2016.12.005.
35. Wu R, Chen X, Zhuang Y, Chen B (2020) Multimodal Shape Completion via Conditional Generative Adversarial Networks. In: Computer Vision – ECCV 2020: 16th European Conference. pp 281–296
36. Wuhrer, S., C. Shu, and P. Xi. Posture-invariant statistical shape analysis using Laplace operator. In: Computers and graphics, Amsterdam: Elsevier Ltd, 2012, pp. 410–416.
37. Zhu, L., X. Hu, and L. Kavan. Adaptable anatomical models for realistic bone motion reconstruction. In: Computer graphics forum, Hoboken: Blackwell Publishing Ltd, 2015, pp. 459–471.