Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Machine Learning Workflows for Motion Capture-driven Biomechanical Modelling

**Vikranth Harthikote Nagaraja**

vikrantmvit@gmail.com

University of Oxford, Institute of Biomedical Engineering, https://orcid.org/0000-0001-7491-8242

**Rahul Sharma**

Laboratory for Computation and Visualization in Mathematics and Mechanics, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland.

**Abhishek Dasgupta**

University of Oxford https://orcid.org/0000-0003-4420-0656

**Challenger Mishra**

Department of Computer Science & Technology, University of Cambridge, 15 J.J. Thomson Ave., Cambridge, CB3 0FD, United Kingdom.

**Physical Sciences - Article**

**Keywords:** Checklist, Carbon Footprint, Learning curve, Machine learning, Model Complexity, Motion capture, Musculoskeletal modelling, Neural network, Explainable AI, Transferability

**Posted Date:** May 22nd, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6588431/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** There is **NO** Competing Interest.

# Machine Learning Workflows for Motion Capture-driven Biomechanical Modelling

Vikranth Harthikote Nagaraja[1*†], Rahul Sharma[2*†], Abhishek Dasgupta[3] and Challenger Mishra[4]

[1*]Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, Oxford, OX3 7DQ, United Kingdom.
[2*]Laboratory for Computation and Visualization in Mathematics and Mechanics, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Lausanne, 1015, Switzerland.
[3]Doctoral Training Centre, University of Oxford, 1–4 Keble Road, Oxford, OX1 3NP, United Kingdom.
[4]Department of Computer Science & Technology, University of Cambridge, 15 J.J. Thomson Ave., Cambridge, CB3 0FD, United Kingdom.

*Corresponding author(s). E-mail(s): vikrantmvit@gmail.com; rahul.sharma@epfl.ch;
Contributing authors: abhishek.dasgupta@dtc.ox.ac.uk; cm2099@cam.ac.uk;
†These authors contributed equally to this work.

**Biomechanical models driven by motion capture (MoCap) offer unprecedented insight into musculoskeletal (MSK) function and aid clinical decision-making. However, traditional MSK models are computationally expensive, laborious to implement, and require meticulously curated inputs. Such models are being complemented by machine learning (ML) methods for user-friendly, real-time predictions, which, however, have often lacked rigorous implementation and assessment. Choosing a fit-for-purpose ML technique is fraught with trade-offs. Here, we show the comparative implementation of nine ML techniques on relatively understudied human upper-extremity MSK modelling from optical MoCap input data (non-invasive gold standard). We identified and investigated model selection and accuracy, generalisability, robustness (to instrumentation errors, soft-tissue artefacts, and anatomical landmark misplacement—inherent to optical MoCap systems), model complexity, transferability (from intact-limbed participants to 'mimicked' transradial prosthesis usage), and interpretability. We also undertook the first assessment of data sufficiency using learning curves and the carbon footprint of training/inference. We found convLSTM to be the optimal ML technique, which efficiently learns the spatial and temporal aspects of MoCap data, while random forest offers a computationally-efficient alternative with minimal accuracy trade-off. This novel holistic characterisation helps lay the methodological foundation towards better deployment (via increased interpretability and robustness) of ML pipelines in biomechanical studies. Finally, we provide best practices and a reporting guideline (LearnABLE) for systematic implementation *and* transparent reporting of ML techniques, aiding their development to better complement and improve traditional MoCap-driven biomechanical modelling.**

**Keywords:** Checklist, Carbon Footprint, Learning curve, Machine learning, Model Complexity, Motion capture, Musculoskeletal modelling, Neural network, Explainable AI, Transferability

Motion capture (MoCap) is the technique of recording the movement of people, animals, or objects. Developed initially for gait analysis in the movement sciences, it is now widely used in animation, entertainment, medical applications, military, robotics, sports, and virtual reality. Observing the structure and movement of the human body, as well as attempting to characterise the underlying structure and mechanism, has appealed to artists and scientists since antiquity [1]. Continuing on this enduring trajectory, motion analysis and biomechanical modelling have become pervasive and increasingly sophisticated in recent decades [2–7]. Applications range from human running biomechanics, movement analysis of extant mammals and birds [8, 9], as well as modelling of extinct species like 'Lucy' [10] and the dinosaurs [11]. Measuring the quality of movement and characterising musculoskeletal (MSK) function has been pivotal in developing fields like evolutionary biology, sensorimotor control, neuroscience, etc., [9, 12, 13].

Various MoCap technologies, developed over decades, offer unique advantages with accuracy and convenience: (i) stereophotogrammetry (Optical Motion Capture [OMC]) with passive retro-reflective markers minimally influence task execution and are regarded as the non-invasive 'gold standard' [2], but are lab-based; (ii) wearable sensors (e.g., Inertial Motion Capture [IMC]) enable data collection beyond the lab in real-world settings; and (iii) Deep-learning video analysis enables unobtrusive motion capture without encumbering subjects.

Three-dimensional motion analysis (particularly through OMC and IMC) helps numerically describe joint and segmental kinematics. Based on Newtonian mechanics, computational modelling and simulation of biomechanical systems—*Inverse dynamics* and *Forward dynamics* techniques—enable estimating in vivo loading information non-invasively [14, 15]. In vivo, joint and muscle loading (difficult to measure experimentally) can be directly estimated from MSK models (e.g., via AnyBody Modeling System [14] or OpenSim [15]) using experimental MoCap data. MSK modelling is increasingly used in personalised modelling and virtual prototyping in automotive, consumer products, exoskeletons, human-system interaction, orthopaedics, and rehabilitation fields. Besides, MSK modelling is adopted for objective clinical decision-making, restricted to some applications [3, 13]. However, the true potential of MSK modelling is yet to be realised in virtual prototyping and personalised treatment (unlike that offered by computer-aided engineering tools, e.g., Finite Element Analysis and Computational Fluid Dynamics).

MSK models are typically laborious to create, scale, and/or set up for each subject or trial individually, and are computationally expensive, restricting their routine use [3, 16]—shortcomings that machine learning (ML) techniques have attempted to remedy [5, 17–19]. Particularly, ML models can help establish a direct relationship between experimental MoCap input and intended biomechanical targets, e.g., joint/intersegmental kinematics and joint/muscle loading [17]. Recent applications include classifying pathological movement, predicting disease risk, quantifying intervention effects, and out-of-clinic patient monitoring [17]. ML methods amplify the benefits of MSK modelling by facilitating real-time predictions, population-level insights [16, 20], and their real-world transition.

Although ML applications are gaining traction in biomechanical modelling [17–35], no study has comprehensively characterised the intersection of biomechanical modelling and ML implementation. Most of these studies reported model accuracy whilst comparing various ML methods but overlooked crucial aspects, such as robustness to inherent issues with (marker-based) MoCap data [36–38], data sufficiency, model selection, generalisability to new unseen subjects from the same distribution or a completely different set (e.g., model trained on non-disabled participants while testing generalisability for people with limb loss/absence), and model explainability. A few studies, beyond review discussions [17, 18], have investigated some of these issues, including model's robustness to added noise [25, 35]), transferability [20, 24], and explainability [31–34].

Adherence to standardised implementation and reporting guidelines could significantly improve the quality of ML practices, reduce biases [39], and facilitate reproducibility [40]. However, there are no guidelines tailored for biomechanical modelling. This work comprehensively covers *all* the above-mentioned issues and offers best-practice guidelines, **LearnABLE** (Section 4), for systematic implementation and transparent reporting of ML applications in MoCap-driven biomechanical modelling. These guidelines, meticulously synthesised from relevant literature and our past studies [28, 29], are similar in spirit to other ML reporting guidelines in various scientific disciplines, e.g., CONSORT-AI, SPIRIT-AI, TRIPOD-ML, and TRIPOD-AI (https://www.equator-network.org/).

This work addresses the above-stated issues by applying ML in upper-extremity biomechanical modelling [24, 28–30]—an area considerably under-researched compared to the lower extremity [17, 20, 23, 25, 32]. We map OMC inputs to kinematic and kinetic biomechanical estimates from MSK modelling and evaluate the performance of nine widely used (neural network and decision tree-based) ML techniques 4. We employed learning curves to assess data sufficiency, grid-search for optimal hyperparameter search, developed separate models for subject-exposed and subject-naive settings to evaluate the models' generalisability, and checked the model transferability on an entirely different population (mimicked transradial prosthesis usage) from the participants used to train the model. Furthermore, we used explainable AI approaches to understand black-box model inference, and also quantified ML model robustness concerning inherent issues with MoCap-based measurements, such as (i) instrumental error [36], (ii) marker or anatomical landmark misplacement [37], and (iii) soft-tissue artefacts [38]. This study offers a critical evaluation and guidelines for ML implementation in this domain, paving the way for future explorations.

# Results and Discussion

Figure 1 shows the ML pipeline that uses the experimental OMC data to predict MSK biomechanical outputs, including joint angles, joint reaction forces, and joint moments (Materials and Methods). We also present a critical evaluation of implementation choices and propose the **LearnABLE** checklist for ML implementation in MoCap-driven biomechanical modelling.

## Model accuracy comparison

We evaluated the performance of various ML techniques (with linear model [LM] as baseline) under Subject-Exposed (SE) and Subject-Naive (SN) settings (Materials and Methods), based on whether model training incorporates some trials of test subjects. In both settings, LM performed poorly with Pearson correlation of $r_{SE,LM,avg} = 0.14 \pm 0.53$ and $r_{SN,LM,avg} = 0.11 \pm 0.56$ between MSK outputs and model predictions, justifying the use of ML techniques (Figure 2, SI-section 2).

For each ML technique, we fine-tuned hyperparameters based on average three-fold cross-validation accuracy and trained the best-fit model for optimal hyperparameters (SI-section 1). This crucial, but often overlooked, hyperparameter optimisation step [17], is guided by our previous work [28, 29]. We explored approximately 36,000 combinations and selected the simplest hyperparameters when validation accuracies were comparable. Consistently, ReLU activation and Adam optimizer were the optimal choices.

Agreeing with earlier observations [17, 25, 28, 29], SE models performed better than SN models ($r_{SE,convLSTM,avg} = 0.82 \pm 0.25$, $r_{SN,convLSTM,avg} = 0.71 \pm 0.40$) for all considered ML techniques and are suited for repeated studies on specific subjects, but do not generalise well to unseen subjects ($r_{SE,convLSTM,avg} = 0.62 \pm 0.52$) (Figure 2 and SI-section 2). For unseen subjects, SN models are considerably more effective than SE models.

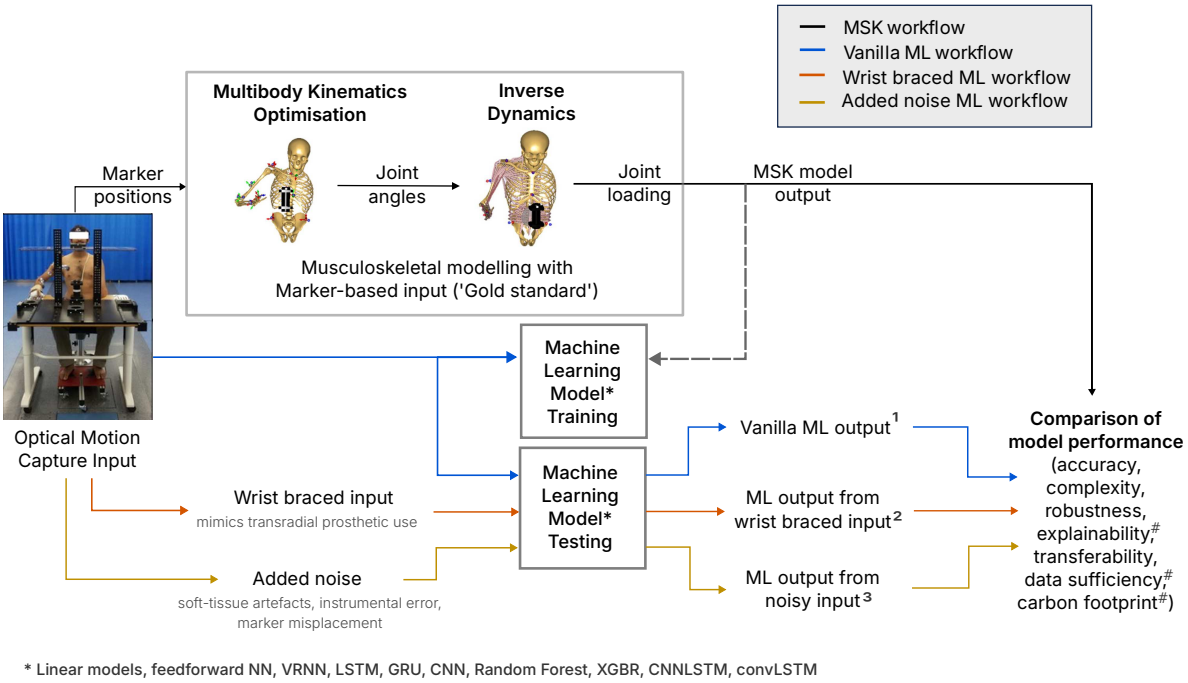
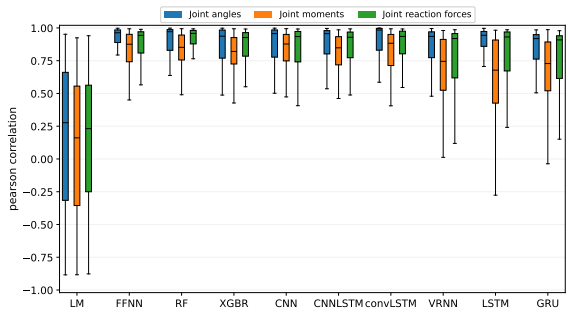


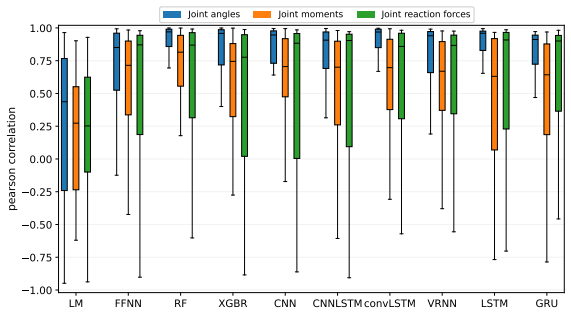

**Fig. 1**: The pipeline used to implement and assess machine learning techniques for mapping marker-based MoCap inputs to MSK-model-based biomechanical targets. This assessment is done independently under two distinct kinds of model training: Subject-Exposed and Subject-Naive (Materials and Methods). [1]Several works exist, which we advance by more comprehensive and rigorous comparison of **nine** widely-used ML methods (in Subject-Exposed and Subject-Naive conditions); [2]Implementation for transferability of ML models in simulated transradial amputation for upper extremity; [3]Robustness study for ML models for inherent errors in MoCap systems. [2,3,#]First-ever implementation of their kind in this domain.

The convLSTM model offered best performance with the lowest RMSE (SI-section 2 and Figure 2). Previous work also reported superior performance of convLSTM than conventional CNN and LSTM in estimating shoulder joint reaction forces [24]—this can be attributed to their capacity to *simultaneously* capture spatial and temporal aspects of MoCap data by integrating convolutional operations within LSTM cells. In contrast, CNNLSTM, a similar hybrid architecture, has poorer accuracy, possibly due to the *sequential* treatment of spatial and temporal aspects. Convolutional models (CNN, convLSTM, and CNNLSTM) generally outperformed recurrent models (VRNN, LSTM, and GRU), highlighting the importance of spatial information in MoCap-driven modelling. Detailed statistics for these models are presented in Figure 2 and SI-section 2. Furthermore, tree-based ML techniques are excellent alternatives due to their inherent interpretability, and the RF model achieved accuracy comparable to that of deep learning models, similar to previous findings [23, 26]. Lastly, the model accuracy varied for different MSK outputs. Generally, MSK outputs with less inherent variability (e.g., Trunk bending or rotation) are captured less accurately by ML, likely due to the dominant effects of noise (Figure 3e and SI-section 3).

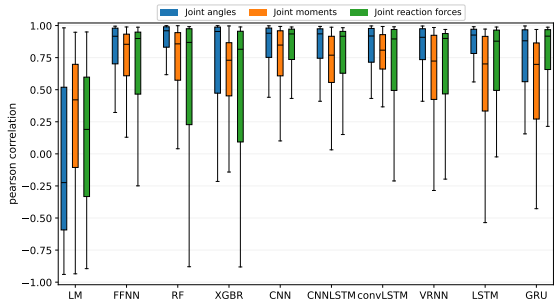
Finally, OMC, while considered the non-invasive 'gold standard', has limitations. High-fidelity alternatives [41], such as bone-pin markers, fluoroscopy, and magnetic resonance imaging, are cumbersome, invasive, require multidisciplinary expertise, and/or pose radiation risks—restricting their routine use. Using synchronously recorded OMC and IMC data, we mapped lower-fidelity IMC data to OMC-driven MSK model outputs [29], which yielded gait lab-quality MSK model estimates directly from IMC data. Along similar lines, ML models mapping skin-mounted, marker-based OMC data to high-fidelity inputs



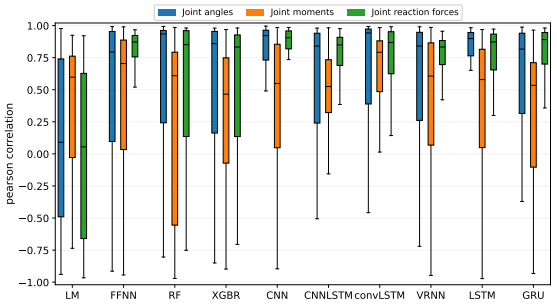

(a) Subject-exposed



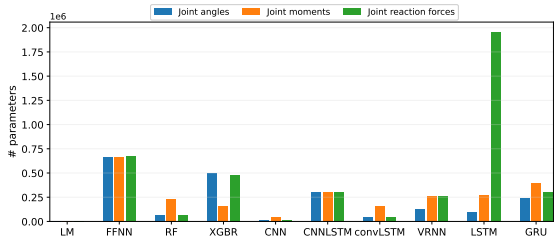

(b) Subject-exposed for unseen subjects



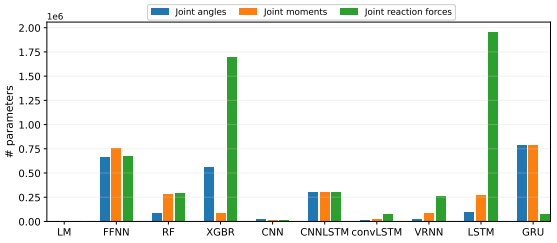

(c) Subject-naive



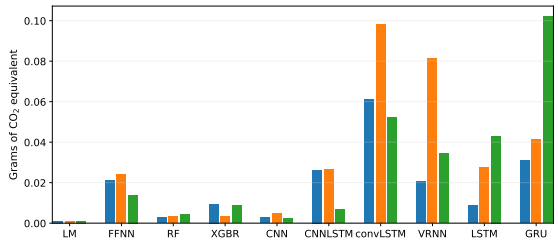

(d) Subject-naive model transferability to wrist braced



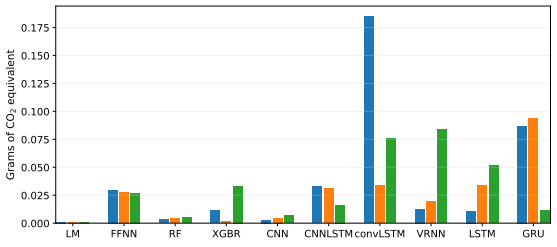

(e) Subject-exposed (# parameters)




(f) Subject-naive (# parameters)




(g) Subject-exposed (Carbon footprint)




(h) Subject-naive (Carbon footprint)

**Fig. 2**: This figure reports the performance and complexity of widely used supervised ML techniques. Accuracy is reported in terms of average Pearson correlation for held-out test data (consisting of multiple trials and/or subjects), with the error bar showing the standard deviation across trials and/or subjects. Note that the subject-naive models were not retrained for wrist-braced subjects and, therefore, are a measure of the transferability of the model. The analogous quantifications in terms of RMSE and NRMSE are reported in SI-section 2 (SI is Supplementary Information). (e, f) plot the number of parameters in the model. For RF and XGBR, the number of parameters is quantified as the total number of nodes in all decision trees. Note that, a direct comparison of the number of parameters in RF and XGBR with NN models might not be appropriate due to the fundamentally different nature of these models. (g, h) plot the carbon footprint (or computational cost) of training the best-fit model – this plot reflects that certain architectures require more computational resources than others despite fewer parameters. Plots with computational cost per epoch during training and per forward pass for predictions are in SI-section 6. All the models trained are computationally viable for real-time predictions with best-fit convLSTM approximately 100 times faster than the sampling frequency of the OMC measurements (generally 100 Hz).

driven MSK model outputs [41] could help address soft-tissue artefacts issues [37] in OMC-tracked tasks. For clinical applications, the MSK model itself can be improved by scaling it to closely match functional anatomy using high-resolution imaging data.

## Data sufficiency and Learning curves

Sufficient available data for ML training is crucial in data-scarce/expensive domains such as healthcare. A learning curve (i.e., model performance versus sample size) typically follows a logarithmic pattern that plateaus at an ML model's limiting accuracy, offering insights into data sufficiency [42].

Learning curves for the best-fit ML models (Figure 3, SI-section 4) exhibited a typical pattern with substantial improvement in validation accuracy as the number of training subjects increased, plateauing beyond six subjects for the joint angle model and five subjects for joint moment and joint reaction force models. This pattern can be explained as each subject comprises three trials with ∼400 frames/trial, which should be sufficient to capture task complexity, marker mappings, and inter-subject variability in this dataset. Furthermore, the reduced variability in model performance (trained on different subjects or for initial solutions) with the number of training subjects demonstrated their robustness. We found that training ML models with 5–6 participants is reasonable and common [17, 28, 34]. Notably, upper-extremity motions are complex, involving a greater range of motion than in lower extremities, where less training data may suffice. However, the datasets used for biomechanical studies are often not heterogeneous enough (also applicable in this study), and therefore, a larger and more diverse dataset is always desirable, and the results should be interpreted in this context.

In case of insufficient data, synthetic data offers an alternative, for instance, simulated IMU data from existing OMC data [22] or using Generative Adversarial Networks for synthetic MoCap datasets involving OMC marker trajectories and force plate-based ground reaction forces (GRFs) [27] or anatomical features of virtual subjects using Metropolis Monte-Carlo Markov Chain model [30]. However, models trained on synthetic data might not generalise well. Furthermore, marker dropouts or occlusions lead to gaps in OMC trajectories. For optimal use of such data, commercial MoCap software, like Vicon Nexus$^{TM}$ v2.5 [43], offers marker gap-filling options. Alternatives include Kalman filters, piecewise local linear models, and ML techniques like bidirectional LSTM and autoencoders [44], particularly for imputing markers absent for longer periods.

## Model robustness

Primary error sources in OMC include: (a) soft-tissue artefacts caused due the smooth, continuous motion of skin-mounted retro-reflective markers relative to underlying bone [37]; (b) inaccurate or inconsistent marker placement over anatomical landmarks [38]; and (c) instrumental errors [36]. These errors vary with subject anthropometry, task, investigator, marker location, and body segments, resulting in noise that conventional filtering methods struggle to eliminate. These errors affect joint kinematics and subsequent analysis [37, 45]. Earlier works examined the robustness/sensitivity of NN predictions due to noisy IMU signals [25] or of RNN predictions to Gaussian noise in mapping surface electromyography signals to joint motion [35].

We assessed the robustness of ML models to inherent errors in OMC systems [36–38] by introducing corresponding simulated noise (Figure 1 and Materials and Methods). In Figure 3 (d-f) and SI-section 3, we have presented the sensitivity of best-fit model accuracy to added simulated noise in terms of spread obtained from 500 trials (for each test subject). In general, the ML models are robust to added noise with a mean signal-to-noise ratio of 18.9 ± 9.6 decibels (for best-fit convLSTM); however, noise dominates output

variables with negligible variations during task execution, such as trunk internal/external rotation (Figure 3e). Arguably, such output variables can be removed from the training but often kept for completeness and in this case, to also examine the impact of noise.

## Model generalisability

We explored the generalisability of ML models across several scenarios. In this domain, a common kind of model trained is SE [17], often leading to over-fitting and performing well in studying the same subject. We have conclusively demonstrated that such models generalise poorly to unseen subjects, and SN models should be preferred for better generalisation over unseen subjects (Figure 2). Notably, SN models only have modest accuracy, as reported earlier [25, 28, 29], highlighting the scope of improvement with a larger and diverse dataset to better capture inter-subject variability. We also assessed the transferability of SN models to different datasets comprising participants wearing wrist braces to mimic transradial prosthesis usage (Figure 2, SI-section 2). The transferability of SN models for unseen wrist-braced participants is reasonable with ($r_{SN,convLSTM,avg} = 0.66 \pm 0.43$).

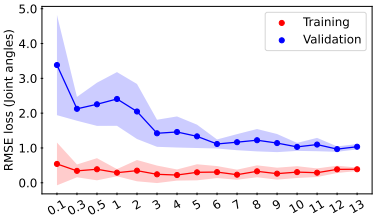
Previous works have also looked into generalisability on different tasks, for instance, generalisability on various shoulder activities in different planes [24] or across age and body morphologies [20]. These works demonstrated generalisability over different tasks/subjects but were limited by the similarity of tasks and the homogeneous demographics. In this direction, generative models may provide a viable solution, e.g., *GaitDynamics* [46] for human walking and running or *RelCon* [47] with motion data from wearable accelerometry sensors. Such models are currently not viable for upper-extremity or data-scarce applications as they require intensive training data. An extensive study is desirable on the generalisability of ML models to a variety of tasks and to considerably different populations, e.g., patient datasets (with pathological conditions or movement disorders).

## Model complexity

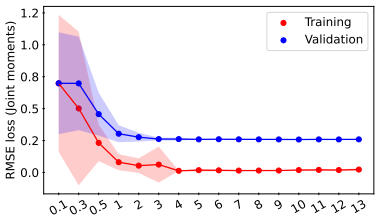
Limiting resource usage and the carbon footprint of ML applications has recently appealed to AI/ML ethics communities due to the high training cost, particularly for deep learning. Factors motivating this trend include the climate crisis and a proliferation of low-cost wearables, smartphones, and vision-based markerless systems (democratising access to health-related metrics, especially in resource-constrained settings and real-world monitoring) [7, 16]. It also facilitates using MSK modelling for personalised and population-level insights—providing resource-efficient techniques.

Figure 2 (e–h) and SI-section 6 provide a comprehensive benchmark for model complexity, training cost (carbon footprint), and inference time. For OMC data, spatial information is critical while estimating biomechanical variables, as evidenced by the outperformance of CNN and convLSTM while using fewer parameters, underscoring the importance of architecture choice. Furthermore, much simpler tree-based models, in particular, RF performed comparably to deep learning models—as also reported elsewhere [23, 26]. Though complexity in terms of the number of parameters does not directly equate to the cost of training or inference, convLSTM, being the most appropriate architecture for such spatiotemporal studies, is inherently costlier to train. These aspects might impede model exploration stages and continuous training on devices. This analysis identified RF to be the optimal model with considerably lower carbon footprint at a very small trade-off with accuracy. Finally, one of the key goals of approximating the computationally intensive MSK model with ML techniques is to enable real-time predictions [20], and we found that all the models trained are computationally viable for real-time predictions with best-fit

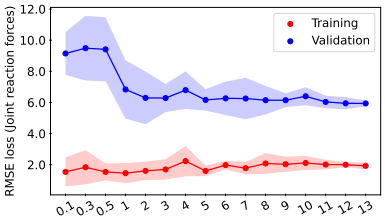
convLSTM and RF models being approximately 30 and 63 times faster than the sampling frequency for full-body MoCap via OMC systems (typically 100 Hz).



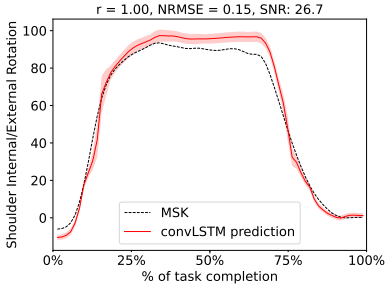

(a) Learning curve convLSTM (joint angles)



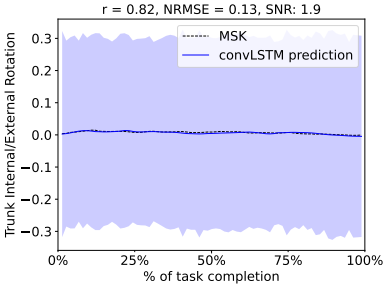

(b) Learning curve convLSTM (joint moments)



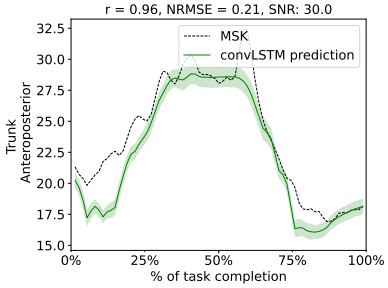

(c) Learning curve convLSTM (joint reaction forces)



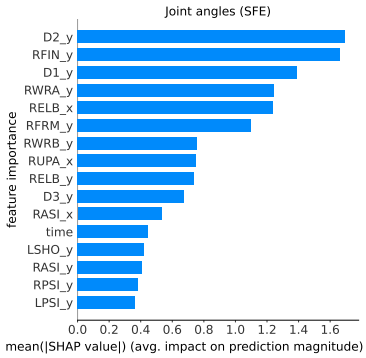

(d) convLSTM SE model (joint angles)



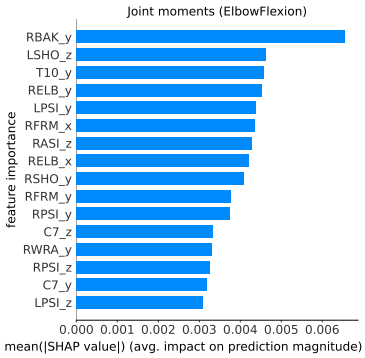

(e) convLSTM SN model (joint moment)



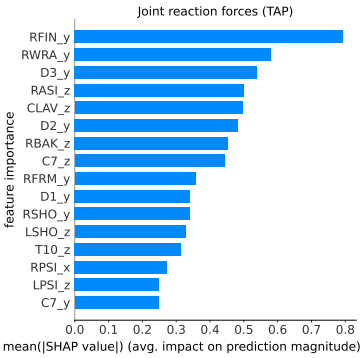

(f) convLSTM SN model (joint reaction forces)




(g) Feature importance (Shoulder Flexion/Extension joint angles)




(h) Feature importance (Elbow Flexion joint moments)




(i) Feature importance (Trunk Mediolateral joint reaction forces)

**Fig. 3**: (a–c) Learning curve plotting training and validation errors vs training data for convLSTM model (using optimal hyperparameters in SI-section 1). 0.1, 0.3, and 0.5 represent approximately 10, 30, and 50% data for a subject. RMSE loss represents the total loss across all output variables. To ensure robust statistics, we generated learning curves for ten different combinations of training subjects or multiple random model initialisation, and the plot shows the mean RMSE loss for ten independent learning curves, with the standard deviation represented as the spread around the mean. (d–e) Comparison of musculoskeletal (MSK) model outputs with convLSTM predictions for a few output variables of a trial in the held-out test data. In addition, the OMC input is subjected to simulated noise corresponding to STA, marker or anatomical landmark misplacement, and OMC system instrumental error and the robustness of the ML models is illustrated by the variability (range) in convLSTM predictions observed in 500 trials. Analogous plots and detailed statistics for all the models and output variables are in SI-sections 2 and 3. (g–i) shows the mean SHAP value, which tells the importance of a given feature (marker location, x, y, and z) in explaining the output. Marker abbreviations are detailed in Figure 5. Here, only three representative outputs are shown with complete results in SI-section 5.

## Model interpretability

As ML applications become mainstream, they must become transparent and accessible for clinicians. Inherently interpretable architectures (e.g., RF) and explainable AI-based approaches (e.g., Local Interpretable Model-Agnostic Explanations, Shapley Additive Explanations [SHAP]) aid model interpretability. ML model interpretation remains scarce, with recent studies exploring feature relevance using Layer-wise Relevance Propagation [31, 32] and examining the interpretability of input representations [33] in gait classification. Alternatively, incorporating physical constraints into model optimisation offers the flexibility of NNs while being more interpretable and generalisable, e.g., in mapping joint kinematics and muscle forces to surface electromyography [34].

Here, we modelled a goal-oriented, cyclic, Reach-to-grasp task in the forward direction (Y-axis: anteroposterior) performed in a seated position (with minimal movement along proximodistal [Z-axis] and mediolateral [X-axis] directions) to assess upper-limb function isolated from upper-extremity movements. The task involves negligible trunk markers' excursion compared to upper-limb markers, with the more distal the marker on the ipsilateral arm (e.g., RWRA, RWRB, RFIN), the larger the excursion, especially along the Y-axis. Details on task, marker placements, and MSK outputs are provided in Materials and Methods, Figure 5, and [48].

Figure 3 (g–i) and SI-section 5 show the feature (marker) importance in RF models for each output. For shoulder angles (Shoulder Flexion/Extension, Abduction/Adduction, and Internal/External Rotation), the ML predictions are primarily determined by marker movement in the Y-axis located on or near the wrist (RWRA, RWRB, RFRM, RFIN) and the correlated dumbbell markers (D1, D2, D3). In general, joint moments have less variability, which makes their interpretation difficult. For elbow joint moments, markers located on the elbow and wrist (RWRA, RELB, RFIN) and shoulder/back (RBAK, LSHO, T10) are dominant in ML prediction. Although trunk involvement in the task is limited, it remains necessary for task performance. The trunk joint reaction forces (e.g., Trunk AnterioPosterior) are explained by movement of torso markers (RASI, LSHO, RBAK, LPSI, CLAV) in Z-direction relative to dumbbell or wrist markers' Y-positions (D1, D2, D3, RFRM, RFIN, RWRA).

## LearnABLE: Machine learning best practices and reporting checklist for biomechanical modelling)

The issue of incomplete or selective reporting of ML implementations, especially in the clinical and healthcare field, is prevalent, leading to calls for greater transparency [49] and for introducing relevant reporting guidelines to ensure that ML research is reproducible. To address the lack of standardised reporting methodology and best practices while applying ML techniques to biomechanical modelling, and MSK modelling in particular, we propose the **Learn**ing m**A**chines for **B**iomechanica**L** mod**E**lling (**LearnABLE**) checklist (Table 1 and Figure 4). The recommendations in this checklist are informed by a comprehensive literature review—identifying gaps in current practices [17–19] and draws best practices from other fields [50–55].

This checklist considers potential future issues in the field, such as big data, responsible and ethical AI use (from conception to evaluation [52, 56, 57]) and the carbon footprint of ML models. While **LearnABLE** focuses on marker-based MSK modelling, most recommendations are suitable for other MoCap techniques (e.g., [7, 16]) and biomechanical modelling (e.g., [3]). Besides this checklist, it is helpful to consult the literature for data collection and quality assessment issues [54, 58] and for a glossary of relevant technical terms [17, 51, 52, 54].

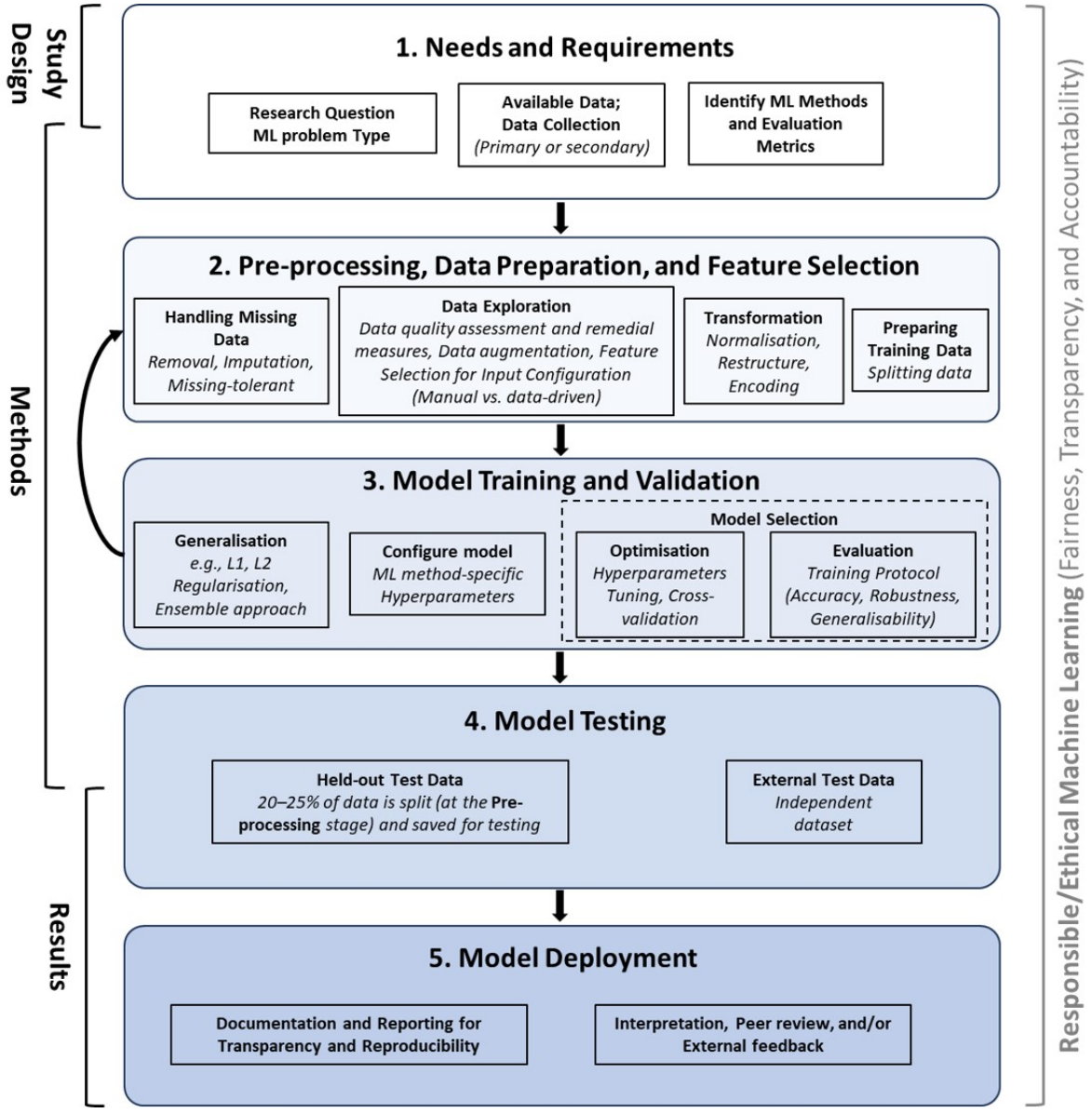



**Fig. 4**: Supervised ML implementation workflow (in Biomechanical modelling) proposed in the **Learn-ABLE** Checklist; corresponding best practices and recommendations are detailed in Table 1.

# Conclusion

Extensive use of personalised biomechanical modelling has been limited in clinical settings for want of a technology that is user-friendly, robust, fast, and accessible. Addressing this need, ML has been gaining traction in biomechanical applications in recent decades, though its implementation often remains inconsistent and cursory. Here, we have developed a holistic pipeline and a checklist for deploying and reporting ML models in biomechanical modelling.

We critically evaluated various implementation choices and compared numerous ML methods, resulting in the following novel insights: (i) 5–7 subjects' data (with three trials each) are sufficient to train robust ML models; (ii) convLSTM model, owing to its spatiotemporal architecture, offered the highest accuracy, which is also a simpler model (in terms of the number of parameters), and random forest also achieved

comparable accuracy at almost half the training and prediction cost (i.e., $CO_2$ emissions); (iii) models are generalisable for unseen trials of the same subject and for unseen subjects ($r_{SE,convLSTM,avg} = 0.82 \pm 0.25$, $r_{SN,convLSTM,avg} = 0.71 \pm 0.40$); (iv) models are transferable to unseen wrist-braced subjects (mimicked transradial prosthesis usage), albeit with decreased accuracy, $r_{SN,convLSTM,avg} = 0.66 \pm 0.43$; (v) models are robust to added artificial noise (corresponding to instrumentation errors, soft-tissue artefacts, and marker misplacement—typically associated with optical MoCap systems); and (vi) features/markers' SHAP values for various biomechanical MSK outputs revealed that the predictions were explained by a combination of the predominant markers' excursion, intuitively relevant for that degree of freedom.

Our characterisations offer an enhanced understanding of ML applications in biomechanical studies, aiding their development in complementing and improving traditional MoCap-driven biomechanical modelling. Furthermore, the interpretability of the model improved trust, especially when trained with limited data. Finally, we have provided best practices and a workflow (LearnABLE checklist) for implementing and transparent reporting of ML in biomechanical modelling. The near real-time prediction capability and accuracy of ML models, coupled with ubiquitous vision or markerless systems in the future, can help with democratising and transitioning motion analysis and biomechanical modelling *beyond the lab* and making their applications even more widespread.

# Materials and Methods

## Data

The anonymised datasets in this work are from previous studies [48, 59], which were approved by relevant Ethics Committees (Reference numbers: 16/SC/0051 and 14/LO/1975). The dataset includes 16 adult right-handed non-disabled (or intact-limbed) participants (four females and 12 males; Age: 25.25 ± 3.34 years; Height: 1.71 ± 0.09 m; Weight: 68.07 ± 12.76 kg; Body Mass Index (BMI): 23.16 ± 3.70 kg/m$^2$).

Each participant performed a Reach-to-Grasp (RGF) task in the 'Forward' direction three times at a self-selected pace. This was undertaken using a bespoke test rig (Figures 1 and 5) that was adapted to the anthropometric dimensions of different subjects and helped with the task execution whilst seated. The considered task required the volunteer reaching out to grasp a 'dumbbell'-shaped object (weighing 2 kg) and moving it between specific designated points as instructed. The designated *'Front'* point was positioned within 90% of the subject's arm length (measured from the acromion to the tip of the middle finger with the arm relaxed at the side) to reduce the contribution of trunk movement during the activity. The initial frame is defined when the hand (both in wrist unbraced and braced settings) is raised more than 10 mm from the starting position while moving in the 'Forward' direction (Figure 5a), whereas the final frame is marked when the hand reaches a height of 10 mm from the initial position while moving in the 'Return' direction during task execution [48].

The OMC data corresponding to task execution was recorded (sampling frequency: 100 Hz) through a 16-camera Vicon$^{TM}$ T40S Series System (Vicon Motion Systems, Oxford, UK). Passive retro-reflective markers (⌀ 9.5 mm) were positioned on the subject's skin using a double-sided tape (hypoallergenic) based on the Plug-in Gait marker set (Figure 5).

Before collecting the data, calibration was carried out according to the Vicon manufacturer's instructions. The cameras were calibrated to maintain a tolerance of $< 2$ mm, which was within the typical MoCap system error range of 1–5 mm [36].

Notably, RIC and LIC markers were included in this protocol to minimise the occlusion of LASI and/or RASI markers due to the table (Figure 1); this helps with reconstructing the Pelvis segment's

| Category | | Best practices and Recommendations |
|---|---|---|
| **1. Needs and Requirements** | | |
| Research question | 1a | Report the clinical condition studied, population or distribution (human or animal study; partial-body or whole-body study) about which the scientific claims are made. ‡ |
| | 1b | State the motivation for using ML methods in the study. * |
| Available data | 1c | Ensure that the dataset represents the population pertinent to the scientific claim. * |
| | 1d | Report the number of samples or subjects in the dataset and the data size. * |
| | 1e | Detail the input data source(s) (e.g., Inertial motion capture, Optical motion capture, Electromyography, Markerless technologies), justifying their relevance to the modelling task. Report inconsistencies in data collection. Ensure appropriate pre-processing of the input data. ‡ |
| | 1f | Detail the output or training label source(s) (e.g., MSK models, Electromyography, Force plates). When dealing with MSK model outputs, ensure model construction/adaptation, simplifications, and assumptions are fit-for-purpose. ‡ |
| Data collection | 1g | Consider synthetic data to mitigate limited data issues (often in biomechanical studies) $ |
| | 1h | Report potential biases in the source dataset and mitigation measures taken, if any. † |
| Identify ML methods | 1i | List ML tasks to be undertaken (e.g., classification, clustering, regression) and ML techniques used ‡. Consider using xAI (Explainable AI) techniques to enhance trust in ML predictions. |
| | 1j | Provide details of simple/trivial models (e.g., Linear models, Support Vector Machines, 1-Nearest Neighbour) for baseline comparisons to justify using ML models. † |
| Miscellaneous | 1k | Provide details of the ethics approval for the dataset used in the study. |
| | 1l | Adhere to responsible AI/ML practices, e.g., report data provenance, consider possible algorithmic bias (STANDING Together [56] and FUTURE-AI [57] recommendations), attribute/acknowledge dataset creators [52] |
| | 1m | Ensure accurate and transparent reporting of race and ethnicity, sex and gender-disaggregated data, etc., per relevant checklists (e.g., SAGER, GATHER, CONSIDER, REP–EQUITY checklists)£. |
| | 1n | Consult general reporting guidelines and best practices relevant to clinical and healthcare fields [40] (e.g., CONSORT-AI, SPIRIT-AI, TRIPOD-ML and TRIPOD-AI)£. |
| **2. Pre-processing, Data preparation, and Feature selection** | | |
| Handling missing data | 2a | Improve data quality by reducing noise, imputing missing data, and removing extensively corrupt data with a rationale. ‡ |
| Data exploration | 2b | Clearly and fully describe the data cleaning steps in text or as a code pipeline. † |
| | 2c | Include confounders as covariates in the model to analyse their contribution (e.g., regress them out) to help prevent the prediction of confounders instead of an outcome due to class imbalance. ‡ |
| | 2d | Report data summary and statistical analysis. |
| | 2e | Use feature engineering (feature fusion or dimensionality reduction via domain knowledge, PCA). |
| Transformation | 2f | Describe all dataset transformations, including treatment of missing data, normalisation or scaling (performed to eliminate inter-individual anthropometric differences). * |
| | 2g | Report instances of combining data from multiple sources and address potential issues. † |
| Preparing training data | 2h | Split data into different sets for training, validation (hyperparameter optimisation, model selection), and testing (final evaluation). Report data splitting method (e.g., random, cluster- or time-based splitting). Ensure test data is excluded from pre-processing and model selection. † |
| | 2i | Reduce or eliminate dependencies or duplicates between the training and test datasets unless such dependency is explicit, such as in subject-exposed scenarios. * |
| **3. Model Training and Validation** | | ⇳ **(Categories 2 & 3 can be updated iteratively)** |
| Generalisation | 3a | Use appropriate regularisation. ‡ |
| Configure model | 3b | Configure the model by selecting ML method-specific hyperparameters. |
| Model selection | 3c | Use learning curves to assess the sufficiency of data and/or subjects for ML model training. |
| | 3d | Do cross-validation with the validation set. Describe the total number of hyperparameter combinations and a justification for selecting this hyperparameter space. |
| | 3e | Report and justify all metrics used to assess and compare model performance (accuracy, Area Under Receiving Operating Characteristics [AUROC], Silhouette, etc.). |
| | 3f | Use simpler models (in terms of model complexity) if accuracy is not considerably compromised to help reduce overfitting, enhance interpretability, and lower carbon footprint. |
| **4. Model Testing** | | |
| Model testing | 4a | Check for the robustness of ML predictions (e.g., to added noise, instrumental errors, marker misplacement, and missing markers), where real-world input data is often noisy. |
| | 4b | Provide evidence of external validity (e.g., testing model on different or independent data)*. |
| **5. Model Deployment** | | |
| Reporting | 5a | Provide descriptive statistics of model(s') performance, including uncertainty estimates (e.g., standard deviations, confidence intervals) and implementation details. * |
| | 5b | Ensure reporting completion of the considerations from 'Category 1. Needs and Requirements.' |
| Documentation | 5c | Report model cards$^{\epsilon}$ with carbon footprint, FLOPS, bias, risks, and limitations. † |
| | 5d | Describe the infrastructure used for computing [51]. Provide scripts with a README and a DOI for reproducibility, and facilitate training and testing with new data. *† |
| | 5e | Provide the dataset along with a link or DOI. Alternately, for external database(s), provide the access date and/or version number. *† |
| Interpretation | 5f | Compare with the current state-of-the-art and highlight contexts where findings may not apply. *† |
| Peer review | 5g | Obtain peer review and external feedback for further improvements in the model. |
| Model deployment | 5h | Consider developing a suitable data pipeline for model deployment. $ |

**Table 1**: **LearnABLE** checklist for supervised ML implementation workflow (in Biomechanical modelling); *Source*: Halilaj et al. [17]‡, Artrith et al. [50]†, Kapoor et al. [58]*, Zhang et al. [55]$, https://www.equator-network.org£, https://huggingface.co/docs/hub/en/model-cards$^{\epsilon}$

orientation and position, required in the AnyBody™ model [14]). In the data pre-processing stage for ML inputs, three pelvis markers—RASI, LPSI, and RPSI—were selected from the six markers (RASI, LASI, RPSI, LPSI, RIC, and LIC) based on their common availability for all subjects and trials.

Additionally, three passive markers D1, D2, and D3 (⌀ 9.5 mm) were firmly attached to the dumbbell, which were pre-processed and labelled manually in an open-source software, Motion Kinematic & Kinetic Analyzer (MoKKA). Furthermore, for 11 of these participants, we have additional data for a reach-to-grasp (in the forward direction) task in 'wrist braced condition', i.e., with the brace mimicking movement restrictions akin to the lack of an active/powered wrist in a typical transradial prosthesis [48]. A detailed comparison of upper-extremity motion in wrist 'braced' and 'unbraced' conditions for intact-limbed participants with transradial prosthetic limb users can be found in Nagaraja [48]. We have utilised this data to quantify the transferability of various ML models trained on non-disabled participants to mimicked transradial prosthesis usage.

We used Vicon Nexus™ v2.5 [43] to pre-process and label OMC marker trajectories as inputs for the MSK model, which were exported as Coordinate 3D (*.C3D) files (Motion Lab Systems Inc., Baton Rouge, LA). A subject-specific upper-extremity MSK model was developed using AnyBody™ Modeling System (AnyBody™ Technology A/S, Aalborg, Denmark) based on their anthropometric dimensions for scaling [48]. Notably, a model of a right-hand body was developed, and movements of the head, left arm, and lower extremity were neglected [48]. The *'Simple Full-Body model'* in the AnyBody Managed Model Repository (AMMR) v.2.1.1 was modified to estimate the kinematic (e.g., joint angles) and kinetic (e.g., joint reaction forces and joint moments) variables. The load from the grasped object was considered to act on the hand segment once it was lifted from the custom-built test rig. During grasping movements, the force exerted at the hand was calculated based on the object's weight, velocity, and acceleration. Additional information regarding the interaction between the dumbbell and the human MSK model is available in [48].

The experimental data obtained from the OMC-driven MSK model estimations, including both kinematic and kinetic parameters, served as the training dataset for our ML efforts.

## MSK outputs

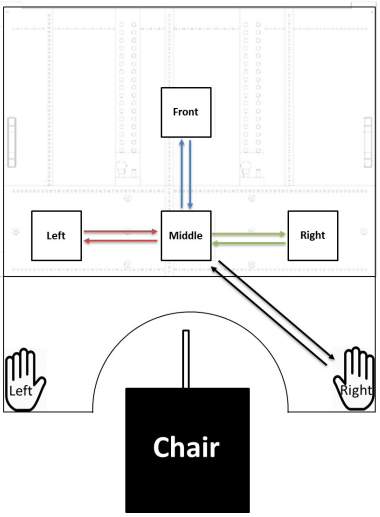
Here, we have mapped the OMC inputs (marker locations described in the previous section and Figure 5) to three biomechanical MSK model outputs, including joint angles, joint moments, and joint reaction forces. Each output consists of multiple features as listed:

(a) Joint angles: Trunk Forward/Backward Bending (TFE), Trunk Right/Left Bending (TAA), Trunk Internal/External Rotation (TIR), Shoulder Flexion/Extension (SFE), Shoulder Abduction/Adduction (SAA), Shoulder Internal/External Rotation (SIR), Elbow Flexion/Extension (EFE), Elbow Pronation/Supination (EPS), Wrist Flexion/Extension (WFE), Wrist Radial/Ulnar Deviation (WAA)

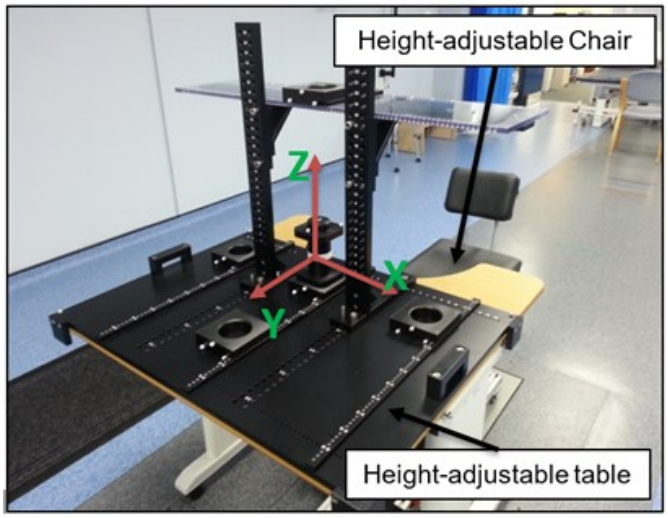
(b) Joint moments: SacrumPelvisFlexionExtensionMoment, SacrumPelvisAxialMoment, SacrumPelvisLateralMoment, GlenoHumeralFlexion, GlenoHumeralAbduction, GlenoHumeralExternalRotation, ElbowFlexion, ElbowPronation, WristFlexion, WristAbduction

(c) Joint reaction forces: Trunk Mediolateral (TML), Trunk Proximodistal (TPD), Trunk Anteroposterior (TAP), Shoulder Mediolateral (GML), Shoulder Proximodistal (GPD), Shoulder Anteroposterior (GAP), Elbow Mediolateral (EML), Elbow Proximodistal (EPD), Elbow Anteroposterior (EAP), Wrist Mediolateral (WML), Wrist Proximodistal (WPD), Wrist Anteroposterior (WAP)

In human motion analysis, anthropometric and demographic factors like age, weight, height, and gender can influence kinematic and kinetic estimates, potentially acting as confounding factors if left uncorrected.




(a) Schematic of test-rig

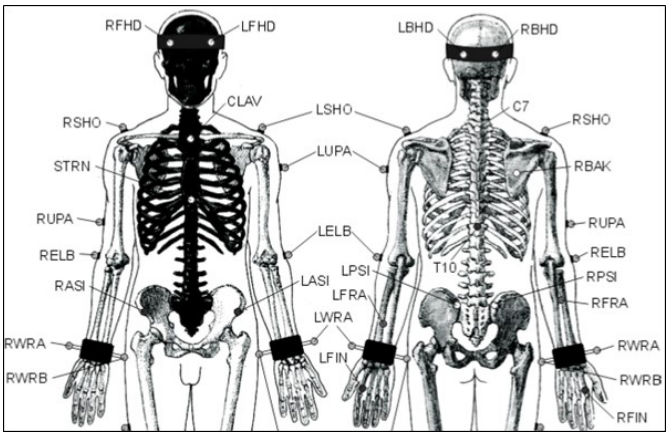



(b) Bespoke apparatus for functional task execution
(Note: D1, D2, D3 markers are on dumbbell)

| Marker name | Marker description |
|---|---|
| R/LFHD | Right / Left front head |
| R/LBHD | Right / Left front back head |
| C7 | Spinous process of the 7th cervical vertebrae |
| T10 | Spinous process of the 10th thoracic vertebrae |
| CLAV | Jugular notch - the clavicles meet the sternum |
| STRN | Xiphoid process of the sternum |
| RBAK | Right middle of the right scapula (asymmetrical) |
| R/LSHO | Right / Left acromio-clavicular joint |
| R/LUPA | Right / Left upper arm between the elbow and shoulder markers (Should be placed asymmetrically with LUPA) |
| R/LELB | Right / Left lateral epicondyle approximating elbow joint axis |
| R/LELBM | Right / Left medial epicondyle (static trial only) |
| R/LFRM | Placed on the lower arm between the wrist and elbow markers (Should be placed asymmetrically with LFRM) |
| R/LWRA | Right / Left wrist thumb side (radial styloid) |
| R/LWRB | Right / Left wrist pinkie side – on the pisiform (ulnar styloid) |
| R/LFIN | Right / Left dorsum of the metacarpophalangeal joint of the index finger |
| R/LFIN2 | Right / Left dorsum of the metacarpophalangeal joint of the ring finger |
| R/LFIN3 | Right / Left dorsum of the tip of the index finger |
| R/LFIN4 | Right / Left dorsum of the tip of the thumb |
| R/LCAP | Right / Left centre of the capitate bone |
| R/LASI | Right / Left anterior superior iliac spine |
| R/LPSI | Right / Left posterior superior iliac spine |
| R/LIC | Right / Left iliac crest |

**Note**: UPA and FRM markers are optional; however, using them improves marker tracking during dynamic trials.

(c) Marker locations description




(d) Plug-in Gait® marker placement scheme

**Fig. 5**: (a) Schematic showing the *Start* and *End* positions for participants' unbraced and braced intact hand (marked as 'Right' or 'Left' as appropriate); Image adapted from [48]. (b) Bespoke apparatus for executing the Reach-to-Grasp task; Image adapted from [48]. (c) Marker positions; Table adapted with written permission from (Vicon Plug-in Gait 2010); Image reproduced from [48]. (d) Plug-in Gait® marker placements; Image reproduced with permission from (Plug-in Gait Product Guide 2010).

Therefore, outcomes are reported per relevant International Society of Biomechanics guidelines [60], as this method effectively reduces gender-based differences. To enable inter-individual comparison: (i) joint reaction forces are normalised to the subject's body weight; and (ii) joint moments are normalised to body weight times height. Both outcome variables are represented as a percentage.

## Subject-naive and Subject-exposed models

One of the major concerns, particularly in data-scarce or data-expensive domains, such as healthcare, is model generalisation. In this domain, two ML model types are particularly interesting: Subject-naive (SN,

models with no prior exposure to the test subjects) and Subject-exposed (SE, models with some exposure to the test subjects) [20]. The latter category is useful when longitudinally studying the same or limited subjects. Such models are often highly accurate for the exposed subjects but inherently do not generalise well, particularly when trained on limited subjects. In contrast, the former category of the model (SN) performs better on unseen subjects (complete statistics are provided in SI-section 2). Here, we have mapped OMC input data to biomechanical MSK model outputs for upper-extremity reach-to-grasp tasks using various supervised ML architectures for both model categories (SE and SN) and analysed their performances.

## ML techniques

This article explores various ML architectures, including vanilla feed-forward Neural Networks (FFNNs), vanilla Recurrent Neural Networks (VRNNs), their advanced variants Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), and Convolutional Neural Networks (CNNs), as well as non-NN architectures, such as Random Forest (RF) and Extreme Gradient Boosting Regression (XGBR) for mapping OMC input data to biomechanical MSK model outputs. Furthermore, we have also used hybrid deep learning architectures, including CNNLSTM and convLSTM. At the onset, we used the linear model as the baseline to establish the need for more complex ML techniques.

## Cross validation for hyperparameter tuning

In total, we have data for 16 unbraced and 11 braced subjects performing the reach-to-grasp task (in the forward direction; three trials). To identify the optimal hyperparameters for a given ML architecture, we conducted a comprehensive grid search (informed by our previous works [28, 29]) exploring a range of hyperparameter combinations for both subject-naive (SN) and subject-exposed (SE) model categories. The best-fit model was selected based on its average validation accuracy obtained from 3-fold cross-validation while ensuring that the validation-train split was performed in accordance with the respective SN and SE model criteria. Additional details on the hyperparameters explored and best-fit hyperparameters for various NN architectures considered are provided at https://github.com/rahul2512/MoCap_ML_workflow/tree/main/hyperparameters and SI-section 1. Note that in the SE model training, a subset of subjects was excluded from the training data to assess the generalisation performance of the SE ML model on novel data. Furthermore, we have non-disabled participants in braced condition data to quantify the transferability of various ML models (SN category) trained on non-disabled participant data to mimicked transradial prosthesis usage.

## Robustness of models against the inherent noise in the OMC

To quantify the robustness against the inherent noise in OMC input (corresponding to STA [37], marker misplacement [38], and instrumental error [36]), we added simulated noise corresponding to these error sources to the inputs and investigated their effect on the model's predictions. Soft-tissue artefacts are approximated by continuous noise of the form $A\sin(\omega t + \phi)$, as per the methodology established by [61]. The parameter values were selected randomly $A \in [0, 10]$ mm, $\omega \in [0, 12\pi]$ Hz and $\phi \in [0, 2\pi]$ rad$^{-1}$s [61]. Furthermore, to account for the inherent error due to marker (mis)placement, an additional Gaussian noise (with mean = 0 and standard deviation = 2 mm) was added [62]. For a given trial, this noise is fixed. Finally, for the inherent error in the MoCap system, we have used Gaussian noise (with mean = 0 and standard deviation = 2.89 mm) corresponding to a maximum of 5 mm tolerance in the OMC measurements [36].

In SI-section 3 and Figure 3, we have plotted the model outputs with statistics (mean ± standard deviation) for 500 samples with added noise for one particular test trial. The tightly clustered outputs (almost indistinguishable in most cases) around the mean emphasise the model's robustness to the inherent noise present in the input data. To quantify the robustness of the model, we have used Signal-to-Noise Ratio (SNR) as a metric. SNR is defined as:

$$\text{SNR} = 10 \cdot \log\left(\frac{\|\mu\|^2}{\text{tr}(\Sigma)}\right)$$

where $\mu$ and $\Sigma$ are the mean and covariance of the samples. $\|.\|$ is the norm of a vector, and tr is the trace of a matrix.

### Model interpretability

Here, we have explored the interpretability of the best-fit random forest model using SHapley Additive exPlanations (SHAP, https://shap.readthedocs.io/). This analysis is done individually for joint angles, joint moments, and joint reaction forces, each containing 10, 10, and 12 biomechanical parameters. Using SHAP, we looked into the input marker location (features) importance for one parameter at a time and plotted it in SI-section 5.

### Model's performance and complexity

To quantify the model's performance, we primarily used NRMSE (Normalised RMSE; scaled by the variable's standard deviation) as the error metric, with OMC-driven MSK output as the ground truth. In addition, Pearson's correlation coefficients (r) between the OMC-driven MSK model outputs and the ML model-predicted outputs are used as a secondary measure. In the main paper, we have reported r as it is easily interpretable, while the corresponding results in terms of RMSE and NRMSE are in SI-section 2. We have approximated the complexity of ML models as the number of model parameters.

The carbon footprint calculation is performed following the guidelines by SCITAS, EPFL (https://scitas-doc.epfl.ch/user-guide/co2-estimation/). Specifically, we used the 'JED' cluster, which produces an emission of 1.02 grams of $CO_2$ equivalent for every CPU core hour used. JED has 2.4 GHz Intel(R) Xeon(R) Platinum 8360Y processors.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Data availability statement

The Python codes (Python 3.12, keras 3.1, and sklearn 1.4) used for implementing ML models and analysing data, as well as the Model card, are available at https://github.com/rahul2512/MoCap_ML_workflow. Due to ethical and privacy reasons, the complete datasets considered in this work are not publicly provided but can be obtained on request, although we have provided the test data.

# Author Contributions

VHN and RS initiated the study. VHN designed the biomechanical methodology and created the anonymised dataset. All authors jointly developed the ML methodology. Data processing and the implementation of the ML models were carried out by RS, AD, and VHN. The initial version of the LearnABLE Checklist was prepared by VHN and AD. RS led the development of the ML methodology and implemented the associated coding and analysis, with AD's input. All authors contributed to the analysis and interpretation of the final results and participated in drafting, reviewing, and revising the manuscript and have approved the final version.

# Acknowledgments

This project made use of data originally collected for a previous study funded by the *Research Councils UK (RCUK) Digital Economy Programme grant number EP/G036861/1* (Oxford Centre for Doctoral Training in Healthcare Innovation) and the *Wellcome Trust Affordable Healthcare in India Award (Grant number 103383/B/13/Z)*. CM's Fellowship is supported through the Accelerate Program for Scientific Discovery at the Computer Laboratory, University of Cambridge. We also acknowledge funding from the *Swiss National Science Foundation* (grant number 200020 182184, supporting RS and Article Processing Charges) and the computational resources provided by *SCITAS, EPFL*.

# References

[1] Cappozzo, A.: Observing and Revealing the Hidden Structure of the Human Form in Motion Throughout the Centuries, pp. 3–15. Springer, Cham (2018)

[2] Cappozzo, A., Della Croce, U., Leardini, A., Chiari, L.: Human movement analysis using stereophotogrammetry: Part 1: Theoretical background. Gait & Posture **21**(2), 186–196 (2005)

[3] Erdemir, A., McLean, S., Herzog, W., Bogert, A.J.: Model-based estimation of muscle forces exerted during movements. Clinical Biomechanics **22**(2), 131–154 (2007)

[4] Deng, B.: Motion studies: See how they run. Nature **525**(7567), 145–146 (2015)

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

[6] Seth, A., Hicks, J.L., Uchida, T.K., Habib, A., Dembia, C.L., Dunne, J.J., Ong, C.F., DeMers, M.S., Rajagopal, A., Millard, M., *et al.*: Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLoS Computational Biology **14**(7), 1006223 (2018)

[7] Mathis, A., Schneider, S., Lauer, J., Mathis, M.W.: A primer on motion capture with deep learning: Principles, pitfalls, and perspectives. Neuron **108**(1), 44–65 (2020)

[8] Catavitello, G., Ivanenko, Y., Lacquaniti, F.: A kinematic synergy for terrestrial locomotion shared by mammals and birds. eLife **7**, 38190 (2018)

[9] Clemente, C.J., De Groote, F., Dick, T.J.: Predictive musculoskeletal simulations reveal the mechanistic link between speed, posture and energetics among extant mammals. Nature Communications

**15**(1), 8594 (2024)

[10] Wiseman, A.L.: Three-dimensional volumetric muscle reconstruction of the *Australopithecus afarensis* pelvis and limb, with estimations of limb leverage. Royal Society Open Science **10**, 230356 (2023)

[11] Hutchinson, J.R., Anderson, F.C., Blemker, S.S., Delp, S.L.: Analysis of hindlimb muscle moment arms in tyrannosaurus rex using a three-dimensional musculoskeletal computer model: implications for stance, gait, and speed. Paleobiology **31**(4), 676–701 (2005)

[12] Wolpert, D.M., Diedrichsen, J., Flanagan, J.R.: Principles of sensorimotor learning. Nature Reviews Neuroscience **12**(12), 739–751 (2011)

[13] Ku, J.P., Hicks, J.L., Hastie, T., Leskovec, J., Ré, C., Delp, S.L.: The mobilize center: An NIH big data to knowledge center to advance human movement research and improve mobility. Journal of the American Medical Informatics Association **22**(6), 1120–1125 (2015)

[14] Damsgaard, M., Rasmussen, J., Christensen, S.T., Surma, E., De Zee, M.: Analysis of musculoskeletal systems in the Anybody Modeling System. Simulation Modelling Practice and Theory **14**(8), 1100–1111 (2006)

[15] Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E., Thelen, D.G.: OpenSim: Open-source software to create and analyze dynamic simulations of movement. IEEE Transactions on Biomedical Engineering **54**(11), 1940–1950 (2007)

[16] Uhlrich, S.D., Falisse, A., Kidziński, Ł., Muccini, J., Ko, M., Chaudhari, A.S., Hicks, J.L., Delp, S.L.: OpenCap: 3D human movement dynamics from smartphone videos. PLoS Computational Biology **19**(10), 1011462 (2023)

[17] Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J., Delp, S.L.: Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of Biomechanics **81**, 1–11 (2018)

[18] Saxby, D.J., Killen, B.A., Pizzolato, C., Carty, C., Diamond, L., Modenese, L., Fernandez, J., Davico, G., Barzan, M., Lenton, G., *et al.*: Machine learning methods to support personalized neuromusculoskeletal modelling. Biomechanics and Modeling in Mechanobiology **19**(4), 1169–1185 (2020)

[19] Mundt, M.: Bridging the lab-to-field gap using machine learning: A narrative review. Sports Biomechanics, 1–20 (2023)

[20] Rane, L., Ding, Z., McGregor, A.H., Bull, A.M.: Deep learning for musculoskeletal force prediction. Annals of Biomedical Engineering **47**(3), 778–789 (2019)

[21] Mundt, M., Johnson, W.R., Potthast, W., Markert, B., Mian, A., Alderson, J.: A comparison of three neural network approaches for estimating joint angles and moments from Inertial Measurement Units. Sensors **21**(13), 4535 (2021)

[22] Mundt, M., Koeppe, A., David, S., Witter, T., Bamer, F., Potthast, W., Markert, B.: Estimation of gait mechanics based on simulated and measured IMU data using an artificial neural network.

Frontiers in Bioengineering and Biotechnology **8** (2020)

[23] Sohane, A., Agarwal, R.: Knee muscle force estimating model using machine learning approach. The Computer Journal **65**(5), 1167–1177 (2022)

[24] Mubarrat, S., Chowdhury, S.: Convolutional LSTM: A deep learning approach to predict shoulder joint reaction forces. Computer Methods in Biomechanics and Biomedical Engineering **26**(1), 65–77 (2023)

[25] Giarmatzis, G., Zacharaki, E.I., Moustakas, K.: Real-time prediction of joint forces by motion capture and machine learning. Sensors **20**(23), 6933 (2020)

[26] Moghadam, S.M., Yeung, T., Choisne, J.: A comparison of machine learning models' accuracy in predicting lower-limb joints' kinematics, kinetics, and muscle forces from wearable sensors. Scientific Reports **13**(1), 5046 (2023)

[27] Bicer, M., Phillips, A.T., Melis, A., McGregor, A.H., Modenese, L.: Generative deep learning applied to biomechanics: A new augmentation technique for motion capture datasets. Journal of Biomechanics, 111301 (2022)

[28] Sharma, R., Dasgupta, A., Cheng, R., Mishra, C., Nagaraja, V.H.: Machine learning for musculoskeletal modeling of upper extremity. IEEE Sensors Journal **22**(19) (2022)

[29] Dasgupta, A., Sharma, R., Mishra, C., Nagaraja, V.H.: Machine learning for optical motion capture-driven musculoskeletal modelling from inertial motion capture data. Bioengineering **10**(5), 510 (2023)

[30] Eghbali, P., Becce, F., Goetti, P., Büchler, P., Pioletti, D.P., Terrier, A.: Glenohumeral joint force prediction with deep learning. Journal of Biomechanics **163**, 111952 (2024)

[31] Slijepcevic, D., Horst, F., Lapuschkin, S., Horsak, B., Raberger, A.-M., Kranzl, A., Samek, W., Breiteneder, C., Schöllhorn, W.I., Zeppelzauer, M.: Explaining machine learning models for clinical gait analysis. ACM Transactions on Computing for Healthcare (HEALTH) **3**(2), 1–27 (2021)

[32] Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R., Schöllhorn, W.I.: Explaining the unique nature of individual gait patterns with deep learning. Scientific Reports **9**(1), 2391 (2019)

[33] Dindorf, C., Ludwig, O., Simon, S., Becker, S., Fröhlich, M.: Machine learning and explainable artificial intelligence using counterfactual explanations for evaluating posture parameters. Bioengineering **10**(5), 511 (2023)

[34] Zhang, J., Zhao, Y., Shone, F., Li, Z., Frangi, A.F., Xie, S.Q., Zhang, Z.-Q.: Physics-informed deep learning for musculoskeletal modelling: Predicting muscle forces and joint kinematics from surface EMG. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2022)

[35] Taneja, K., He, X., He, Q., Chen, J.-S.: A multi-resolution physics-informed recurrent neural network: formulation and application to musculoskeletal systems. Computational Mechanics **73**(5), 1125–1145 (2024)

[36] Chiari, L., Della Croce, U., Leardini, A., Cappozzo, A.: Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors. Gait & Posture **21**(2), 197–211 (2005)

[37] Leardini, A., Chiari, L., Della Croce, U., Cappozzo, A.: Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation. Gait & Posture **21**(2), 212–225 (2005)

[38] Della Croce, U., Leardini, A., Chiari, L., Cappozzo, A.: Human movement analysis using stereophotogrammetry: Part 4: Assessment of anatomical landmark misplacement and its effects on joint kinematics. Gait & Posture **21**(2), 226–237 (2005)

[39] Singh, K., Beam, A.L., Nallamothu, B.K.: Machine learning in clinical journals: Moving from inscrutable to informative. Circulation: Cardiovascular Quality and Outcomes **13**(10), 007491 (2020)

[40] Flanagin, A., Pirracchio, R., Khera, R., Berkwits, M., Hswen, Y., Bibbins-Domingo, K.: Reporting Use of AI in Research and Scholarly Publication—JAMA Network Guidance. JAMA **331**(13), 1096–1098 (2024)

[41] Cereatti, A., Bonci, T., Akbarshahi, M., Aminian, K., Barré, A., Begon, M., Benoit, D.L., Charbonnier, C., Dal Maso, F., Fantozzi, S., *et al.*: Standardization proposal of soft tissue artefact description for data sharing in human motion measurements. Journal of Biomechanics **62**, 5–13 (2017)

[42] Viering, T., Loog, M.: The shape of learning curves: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(6), 7799–7819 (2022)

[43] Vicon Nexus: Vicon Nexus™ 2.5 Manual. 2016. What's new in Vicon Nexus 2.5 – Nexus 2.5 Documentation. Documentation docs.vicon.com (2016). (Accessed on 10/03/2025)

[44] Pavllo, D., Delahaye, M., Porssut, T., Herbelin, B., Boulic, R.: Real-time neural network prediction for handling two-hands mutual occlusions. Computers & Graphics: X **2**, 100011 (2019)

[45] Uchida, T.K., Seth, A.: Conclusion or illusion: Quantifying uncertainty in inverse analyses from marker-based motion capture due to errors in marker registration and model scaling. Frontiers in Bioengineering and Biotechnology **10**, 874725 (2022)

[46] Tan, T., Van Wouwe, T., Werling, K., Liu, C.K., Delp, S., Hicks, J., Chaudhari, A.: GaitDynamics: A generative foundation model for analyzing human walking and running. Research Square, 3 (2025)

[47] Xu, M.A., Narain, J., Darnell, G., Hallgrimsson, H., Jeong, H., Forde, D., Fineman, R., Raghuram, K.J., Rehg, J.M., Ren, S.: RelCon: Relative Contrastive Learning for a motion foundation model for wearable data. In: International Conference on Learning Representations (ICLR) 2025 (2025)

[48] Nagaraja, V.H.: Motion capture and musculoskeletal simulation tools to measure prosthetic arm functionality. PhD thesis, Department of Engineering Science, University of Oxford, Oxford, United Kingdom (2019). https://ora.ox.ac.uk/objects/uuid:70253345-b7a8-49ab-9bb8-2d247bae993b

[49] Haibe-Kains, B., Adam, G.A., Hosny, A., *et al.*: Transparency and reproducibility in artificial intelligence. Nature **586**(7829), 14–16 (2020)

[50] Artrith, N., Butler, K.T., Coudert, F.-X., Han, S., Isayev, O., Jain, A., Walsh, A.: Best practices in machine learning for chemistry. Nature Chemistry **13**(6), 505–508 (2021)

[51] Stevens, L.M., Mortazavi, B.J., Deo, R.C., Curtis, L., Kao, D.P.: Recommendations for reporting machine learning analyses in clinical research. Circulation: Cardiovascular Quality and Outcomes **13**(10), 006556 (2020)

[52] Scott, I., Carter, S., Coiera, E.: Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics **28**(1) (2021)

[53] Collins, G.S., Moons, K.G.: Reporting of artificial intelligence prediction models. The Lancet **393**(10181), 1577–1579 (2019)

[54] Lones, M.A.: Avoiding common machine learning pitfalls. Patterns **5**(10) (2024)

[55] Zhang, A., Xing, L., Zou, J., Wu, J.C.: Shifting machine learning for healthcare from development to deployment and from models to data. Nature Biomedical Engineering **6**(12), 1330–1345 (2022)

[56] Alderman, J.E., Palmer, J., Laws, E., McCradden, M.D., Ordish, J., Ghassemi, M., Pfohl, S.R., Rostamzadeh, N., Cole-Lewis, H., Glocker, B., *et al.*: Tackling algorithmic bias and promoting transparency in health datasets: the standing together consensus recommendations. The Lancet Digital Health **7**(1), 64–88 (2025)

[57] Lekadir, K., Frangi, A.F., Porras, A.R., et al.: FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ **388** (2025) https://doi.org/10.1136/bmj-2024-081554

[58] Kapoor, S., Cantrell, E.M., Peng, K., Pham, T.H., Bail, C.A., Gundersen, O.E., Hofman, J.M., Hullman, J., Lones, M.A., Malik, M.M., *et al.*: Reforms: Consensus-based recommendations for machine-learning-based science. Science Advances **10**(18), 3452 (2024)

[59] Nagaraja, V.H., Cheng, R., Kwong, E.M.T., Bergmann, J.H., Andersen, M.S., Thompson, M.S.: Marker-based vs. inertial-based motion capture: Musculoskeletal Modelling of Upper Extremity Kinetics. In: ISPO Trent International Prosthetics Symposium (TIPS) 2019, Manchester, United Kingdom, pp. 37–38 (2019). ISPO. https://www.researchgate.net/publication/344409362

[60] Derrick, T.R., Bogert, A.J., Cereatti, A., Dumas, R., Fantozzi, S., Leardini, A.: ISB recommendations on the reporting of intersegmental forces and moments during human motion analysis. Journal of Biomechanics **99**, 109533 (2020)

[61] Cheze, L., Fregly, B., Dimnet, J.: A solidification procedure to facilitate kinematic analyses based on video system data. Journal of Biomechanics **28**(7), 879–884 (1995)

[62] Groot, J.H.: The variability of shoulder motions recorded by means of palpation. Clinical Biomechanics **12**(7-8), 461–472 (1997)

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.pdf