Requirements on training data quantity for robust human arm posture prediction form pressure sensors

Sebastian Helmstetter, Jacob Sembritzki, Simone Arnold and Sven Matthiesen*

Abstract— Objective: **This study investigates whether machine-learning models can be efficiently trained to predict human arm motion during horizontal drilling using pressure distribution data from a power-tool handle while maintaining high precision and robustness against varation of pressure distribution.** ***Methods:*** **Three participants perform horizontal wood drilling. Handle-integrated thin-film pressure sensors record pressure distribution. Arm joint motion (radial abduction, elbow flexion, shoulder flexion) is captured as ground truth using an IMU-based motion capture system (Xsens). For the arm posture prediction, Deep neural networks (DNN) and sparse Gaussian process regression (SGPR) models are trained. The influence of training data quantity is analyzed by varying the number of trials and datapoints per trial. Robustness is evaluated by the scatting of the prediction error across ten grouped train–test splits.** ***Results:*** **Accurate arm motion prediction was achieved for all joint angles. DNN models reached average RMSE values of 2.51° (radial abduction), 8.57° (elbow flexion), and 9.72° (shoulder flexion), with standard deviations below 1.6°. Relative errors remained within 7–12% of the motion range. Increasing the number of independent trials significantly reduced RMSE, whereas datapoints per trial showed minimal influence beyond ~100 samples. SGPR models showed consistently lower precision and robustness than DNN models.** ***Conclusion:*** **The study demonstrates that sufficient quantity of training data enables robust and precise prediction of human arm posture during horizontal drilling from pressure sensor data. Maximizing the number of independent trials represents the key factor to ensure prediction robustness against varation of pressure distribution under constrained data volume.**

Index Terms— **Human-machine interaction, Motion analysis, Smart power tools, Machine Learning**

I. INTRODUCTION

A healthy workforce is essential to economic success and sustainable productivity in a modern society [1]. Disorders that compromise workers' performance burden healthcare systems and reduce the operational efficiency of national economy. Musculoskeletal disorders (MSDs) and injuries of connective tissues are representing the leading causes of workplace absenteeism. During 2021–2022, workers in the European Union were abstinent for 2,246,900 workdays due to MSDs [1]. 9,230 days - about 25 % - attribute specifically to MSDs involving the upper extremities [1]. This economic harm underscores the substantial relevance of the protection against MSDs.

Due to their essential function in manual tasks, upper extremities are major prone to biomechanical stress and a high risk for MSDs. Key risk factors as repetitive movements, awkward postures, and high physical loads tissues and promote chronic impairments [2].

To mitigate the risks of work-related MSD, real-time monitoring systems allow an early detection and warn workers about hazardous postures before they cause long-term damage [3]. Therefore, scalable and cost-efficient methods are needed to capture working postures and human motion. However, most established technologies for posture and motion capturing are designed for laboratory studies and require an high effort of equipment and installation processes, that makes them impractical for large-scale workplace applications [3].

Modern machine learning techniques enable low-cost and more flexible alternatives to the established motion capture systems [4]. A major class of methods leverages computer vision algorithms to estimate human posture by detecting anatomical landmarks in RGB video streams [5]. Without the requirement of physical markers attached to the human body, it reduces the effort and limitations in preparation of the person significantly compared to established infrared marker-based motion capturing.

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. For example, “This work was supported in part by the U.S. National Science Foundation under Grant BS123456”.

S. Helmstetter and J. Sembritzki are with the Institute of Product Engineering (IPEK), Karlsruhe Institute of Technology (KIT)

S. Arnold is an Independent Researcher, Karlsruhe, Germany.

*S. Matthiesen is with the Institute of Product Engineering (IPEK), Karlsruhe Institute of Technology (KIT), Kaiserstrasse 10, 76131 Karlsruhe, Germany (correspondence e-mail: sven.matthiesen@kit.edu).

An alternative sensing approach when no camera application is available is the prediction of working postures by data from wearable pressure sensors [6–8]. Choffin *et al.* [9]and Alemayoh *et al.* [10] report high model performance also using pressure map data for estimating lower limb joint angles in gait analysis. Helmstetter *et al.* [8] first predicted human arm postures for manual power tool tasks by using pressure sensors embedded in the handle. This approach enables direct integration into conventional working devices while maintaining high mobility of the sensing system. In the proof-of-concept study, Helmstetter *et al.* [8] demonstrated the prediction of arm postures with a mean absolute error (MAE) up to 7.3° and a root-mean-square error (RMSE) up to 10.7°, even with a small amount of training data for each person. However, the prediction performance varies substantially across participants, indicating limited model robustness against varation of pressure distribution and suggesting that the training data may not be homogeneous enough. Similarly, to other applications of machine learning, the method proposed by Helmstetter *et al.* [8] appears to depend on a larger training dataset to ensure robust and accurate predictions.

To build robust models with low prediction error while keeping the training dataset as small as possible, this study investigates the influence of training data quantity on model accuracy and robustness against varation of pressure distribution by the following research question:

How do the following aspects of training data quantity influence the accuracy and robustness of predicting the arm joint angles from pressure sensors at a power tool handle during wood drilling?

- ***number of drilling trials***
- ***amount of measurement data per drilling trial***

To answer the research question **parametric and non-parametric models** are trained varying the influence parameters. The robustness against varation of pressure distribution of the models is evaluated by using different **splits of training and test data** and comparing the model accuracy between these different sets.

II. Methods

To investigate the effect of the influences on parametric and non-parametric machine learning models, two model architectures — *Deep Neural Networks (DNN)* and *Sparse Gaussian Process Regression (SGPR)* — are trained with variations of the training data set. The overall training data set is acquired by an experimental study on wood drilling with 3 participants. The pressure distribution at the power tool handle is measured by a matrix of thin-film pressure sensors. The ground truth data of arm posture is captured by the established motion capture system Xsens Awinda. Details of the data acquisition, data processing, model training and variation of the influence parameter are given in the following.

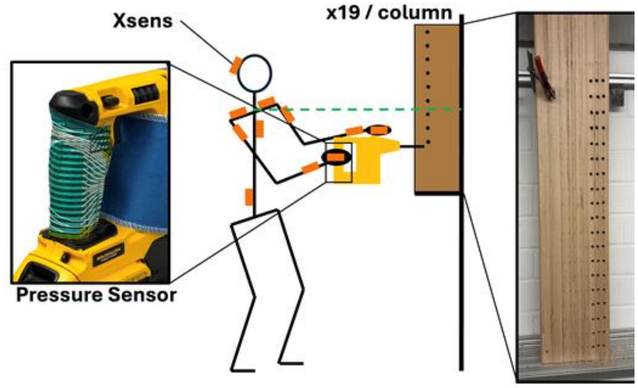

Figure 1: *Experimental setup for the training and test data acquisition: horizontal drilling in a wood beam; green line: the center height of beam is aligned to participant's shoulder height; left hand side: piezoresistive Tekscan E3000 sensor matrix at the power tool handle; orange boxes: position of MU-tracker of the motion capture system Xsens Awinda.*

A. Data Acquisition

The training and validation data are acquired in an experimental study with three right-handed, male participants performing horizontal drilling in a beam of laminated beech wood. The experimental setup represents a simplified version of a typical drilling operation on a construction side. This configuration encompasses multiple disturbances and influencing factors that are also present during real-world use of power tools. **Figure 1** illustrates the setup.

The center height of the wooden beam is aligned with the individual shoulder height of each participant. The exact body posture is not defined to avoid influencing the participant's natural behavior. However, the definitions of the drilling height ensure a systematic and comparable variation in arm posture across all participants. Each participant performs nine drilling operations above, nine below, and one at this reference height, thus spanning a total vertical distance of 90 cm. The procedure is repeated 14 times per participant. The task is performed using a cordless drill (DCH 273NT, Dewalt, Stanley Black and Decker, Inc., Towson, Maryland, USA) with no impact mode and a 10 mm diameter drill bit for wood.

The pressure data is measured by a piezoresistive thin-film pressure sensor Type 3000E (Tekscan, Inc., Norwood, MA, USA), which contains a matrix of single sensing elements spaced 5.1 mm apart, in combination with the VersaTek measurement system. The sensor placed on the handle of the power tool as shown in **Figure 1** captures the pressure distribution during each drilling operation.

The IMU-based Xsens Awinda motion capture system (Movella Inc., Henderson, NV, USA, Software: Xsens MVN 2018, model: upper body with hands, calibrated in N-pose + walking) is used to measure the ground truth posture of the right, dominating arm. The synchronization of the two measurement systems is achieved through an analog trigger signal, with the pressure measurement system acting as the master. At the beginning of each drilling operation, the supervisor starts the measurement manually and stops it when the target drilling depth of 5 cm is reached. All data are recorded at a sampling rate of 60 Hz. Due to varying material resistances during drilling and different levels of push force applied by the participants, the duration of the measurement, and the amount of data recorded per trial, varies.

The three male participants are non-professional power tool users, with body heights ranging from 183 cm to 190 cm. At the time of the study, none of the participants report any musculoskeletal disorders affecting the upper limbs. Prior to the experiment, the participants were informed about the storage and processing of the recorded data. The study is approved by the ethics committee of Karlsruhe Institute of Technology by the (EVV ID: 737). The results of the data acquisition are synchronous sets of pressure distribution data at the power tool handle and corresponding arm postures.

B. Data Processing and Model Training

During the data processing, the measurement data is transferred into a matrix with training and test data for each participant. Based on this matrix we train and test the parametric DNN and non-parametric SGPR models to predict the arm posture during drilling. The data processing and model training is performed in Python using the scikit library [11].

Each matrix combines all measurements of one participant. The measurements are splitted by its frames and the data of one frame is transformed into one row of the matrix. The columns correspond to features of each pressure sensor elements, additional features calculated during the data processing, and the joint angles of the right arm (radial abduction at the wrist, elbow flexion, and shoulder flexion). These angles are used as targets and ground truth from the Xsens system. A trial ID is added in the first column to distinguish the different trials to avoid data leakage when splitting the data into training, validation and, test datasets. The three additional features of the pressure sensor data are the average load, total load and, the number of activated sensor elements. All features from the sensor element as well as the target data from Xsens are filtered with a low pass filter with a cutoff frequency of 1 Hz and edge padding. Sensor elements that are not activated over all trials are removed from the data. Additionally, features that do not contribute significant information are filtered out by the "VarianceThreshold" function from sklearn. It reduces the dimensionality of the data.

For model training the data is split into training sets (85 %) and test sets (15 %). The split is done by the trail IDs. The training data is fitted and transformed with the "MinMaxScaler" function from sklearn, such that all features are within the range 0 and 1. The test data is only transformed with the scaling factor of the training data. 5-fold cross validation is used for each training to assess generalization.

The parametric deep neural network (DNN) has multiple hidden layers. Dropout layers are added after every hidden layer as well as LASSO (least absolute shrinkage and selection operator) regularization is performed. The configuration for the prediction of the different angles can be seen in **Table 1**. Both measures help preventing the DNN from overfitting the training data. The non-parametric sparse gaussian process regression model (SGPR) is a variation of a gaussian process regression model (GPR) that uses inducing points to handle the resource intensive structure of GPR with large datasets. Optimizable inducing points are set to every 75th data point in the training data for initialization. A kernel of Matern52 is used.

Table 1: Configuration of the deep neural network to predict the radial abduction, elbow flexion, and shoulder flexion

configuration	radial abduction	elbow flexion	shoulder flexion
layers	3	4	5
neutrons	64	128	128

The final meta model is built by ensemble learning with a combination of the 5-fold models with optimal parameter for the validation data set. The prediction by the meta model is calculated by the mean of the prediction of the 5-fold models to reduce the risk of over fitting.

C. Evaluation of Model Performance

The evaluation of the model performance is divided into error metrics and an analysis of the model robustness against varation of pressure distribution. To evaluate the average error of posture prediction for each model the mean absolute error (MAE) and the root means square error (RMSE) are calculated. The RMSE is used in the following to present and discuss the results of the study due to its higher sensitivity to large errors which are more relevant in case of work posture monitoring.

The robustness against varation of pressure distribution of the model is evaluated by varying training-test splits. Ten training-test splits are generated by using the sklearn GroupShuffleSplit function with ten different random states. The scattering of the RMSE for the ten train-test splits is displayed in the results as box plots. A low scattering of the error metrics over the ten train-test splits shows a good generalizability and high robustness of the model. An ID number for each random state is used to ensure reproducibility of the train test split.

D. Model Variations

To analyze the effect of the influencing parameters on the model performance and to evaluate what quantity of training data is needed for acuate and robust models, the training data are systematically varied. **Table 2** gives an overview over the variations and the respective step sizes.

Table 2: Variations of training data quantity

parameter	minimum	maximum	step size
number of trials	10 %	100 %	10 %
amount of data per trial	10 %	100 %	10 %

1) Variation of the Number of Drilling Trials

To investigate the required quantity of training data, the numbers of drilling trials is reduced. Starting with the whole dataset of 266 trails per participants random trails are excluded stepwise from 0 % to 90 % by a step width of 10 %. It provides a comprehensive overview of how the models react to the number of trials used for training and testing.

2) Variation of Amount of Data per Trial

Further, the quantity of training data is reduced by decreasing the amount of data per trilling trial. The influences of this variation are investigated by using only the initial part of the data of each trial. Due to the control of the drilling depth and individual drilling speed for each hole, the amount of data per trial is not equal over all trials. To start with the same amount of data for each trial, the data of each trial is resampled to the average amount of data per trial individually for each participant. For the variation, the total resampled dataset is reduction in steps until only 10% is left. This variation provides a comprehensive overview of how the models react to different amount of data per trial, that is indirectly defining the duration of the data acquisition.

3) Evaluation of the data variation

First, the overall performance of the joint angle prediction is analyzed to the improvement by the large set of training data. Therefore, the mean RSME of the ten test splits and he standard deviation (std) is evaluated for redial abduction, elbow flexion, shoulder flexion. Due to a better comparability, the relative mean RSME is calculated by dividing the mean RSME by the range of joint angle motion. It eliminates the influence of significantly different range of motion for the different joint angles.

Thereafter, the influence of the training data quantity is analyzed for the elbow flexion prediction by plotting the RSME values of the prediction over increasing the number of trials and amount of data per trial. The median RSME shows the accuracy of the prediction. The scattering of the RSME values is showing the robustness against varation of pressure distribution. Low median RSME and low scattering is showing a good performance of the prediction. Further, the trend of the median RSME is analyzed to see the effects of parameter variation and to decide when a sufficient accuracy can be reached.

III. Results

The following section presents the results categorized by variations in the influence parameters. First, an overview of the experimental results and the model performance trained on the full dataset is provided. The models achieve robust arm-posture predictions, with average RMSE values ranging from 4.09° to 12.73°. Across all evaluated variations, the number of drilling trials is the dominant factor influencing model performance. The model performance increases continuously when increasing the trial number. The amount of measurement data per drilling trial has only a minor effect on the model performance and robustness against varation of pressure distribution. The results are presented and discussed with examples of selected plots for the elbow flexion.

A. Experimental Data and General Model Performance

To evaluate the general model performance and robustness against varation of pressure distribution on large datasets, this section first presents the results without dataset variation. **Table 3** provides an overview of the recorded data for all three participants. The amount of data and the average time per trial vary depending on factors such as feed force, material resistance, and tool wear. Since feed force was not directly measured and further influencing factors were not examined, no clear causal relationship can be derived from the observed differences in average trial duration. Nevertheless, the variation may reflect underlying differences in feed force and may partly relate to the performance differences reported in the following section.

The overall model performance trained on the full dataset is shown in **Table 4**. When trained on larger datasets, the parametric models achieve better performance than the non-parametric SGPR models. The lower mean RMSE across the ten test splits for each participant and joint angle indicates better model performance. This advantage can be also recognized for the better robustness of the prediction with the DNN models visible by a lower overall standard deviation of the RMSE values. Focusing on the RSME values for the different joint angles we see an increase of the nominal RSME from the radial abduction to the elbow flexion and to shoulder flexion. This pattern suggests increasing RMSE with greater distance between the joint angle and the sensor location used as the data source. This phenomenon is visible in Figure 2. The nominal RSME values are varying from 2.51° (radial abduction, DNN) to 12.73° (shoulder flexion SGPR). By comparing the nominal RSME for the different joint angles with the range of motion of each joint angle during the data recording as visualized in

Table 3: Overview of the amount of recorded data for all 3 participants and the average time per trial

participant	number of trials	total data points	average time per trial
# 1	266	65,300	4.1 s
# 2	266	53,136	3.3 s
# 3	266	103,613	6.5 s

Figure 3, we observe a constant relative RSME at about 7 % to 12 % when dividing the nominal RSME by the range of joint angle motion. The relative performance values are closely aligned, indicating consistently stable prediction across joint angles.

B. Robust and precise human arm posture prediction with large training data sets

To validate the prediction approach for wood drilling, the dataset was expanded to 266 drilling trials, resulting in more than 50,000 data points per participant for three participants. Models trained on this dataset achieve an RMSE between 6° and 15° for elbow flexion across all test splits and participants. This performance is comparable to the RMSE reported by Helmstetter *et al.* [8], thereby reinforcing the findings of Helmstetter *et al.* [8].

The total number of available training samples per participant varies by up to 100 %. This variation results from differences in drilling duration, as summarized in **Table 3** reflecting the time required by each participant to drill a hole of predefined depth. Drilling duration and feed rate are influenced by multiple factors. While material resistance and tool sharpness are not directly controlled by the participant, the feed force depends on the force applied by the user to press the power tool against the wooden beam. This force directly affects the recorded pressure distribution at the handle and, consequently, the structure of the training dataset. To reduce the influence of feed force, a predefined average feed force could be specified for the participants and displayed via an optical feedback signal, comparable to the setup presented by Lindenmann *et al.* [12]. However, prescribing the feed force

Table 4: Performance of the prediction models trained with the total data sets for all participants and joint angles of their right arm

participant	Root Mean Square Error (RMSE) of the Deep Neural Network (DNN) model								
	radial abduction			elbow flexion			shoulder flexion		
	mean (nom.) [DEG]	std [DEG]	mean* (rel.)	mean (nom.) [DEG]	std [DEG]	mean* (rel.)	mean (nom.) [DEG]	std [DEG]	mean* (rel.)
# 1	3.15	0.24	8%	9.33	1.63	10%	10.88	1.78	7%
# 2	1.85	0.24	4%	7.39	0.84	9%	7.87	0.76	7%
# 3	2.55	0.84	8%	8.99	1.08	9%	10.41	2.23	8%
average	2.51	0.44	7%	8.57	1.18	9%	9.72	1.59	7%

participant	Root Mean Square Error (RMSE) of the Spares Gaussian Process Regression (SGPR) model								
	radial abduction			elbow flexion			shoulder flexion		
	mean (nom.) [DEG]	std [DEG]	mean* (rel.)	mean (nom.) [DEG]	std [DEG]	mean* (rel.)	mean (nom.) [DEG]	std [DEG]	mean* (rel.)
# 1	4.30	0.47	11%	12.15	1.08	13%	14.23	1.43	9%
# 2	3.32	0.57	7%	10.55	1.98	13%	11.03	2.51	9%
# 3	4.63	0.52	15%	11.17	1.38	11%	12.92	1.76	10%
average	4.09	0.52	11%	11.29	1.48	12%	12.73	1.90	9%

*RMSE relative to the range of joint angle motion from lower upper whiskers end

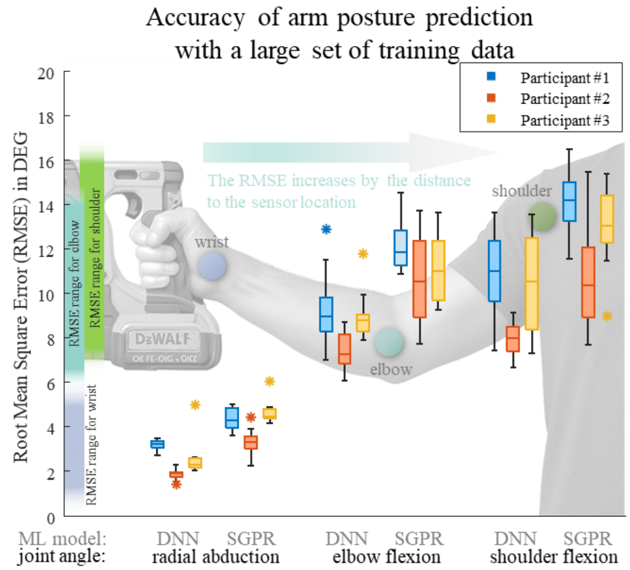

Figure 2: *Overview of RMSE scattering for the joint angle prediction trained with the total data.The boxplots summarize the RMSE distribution across ten test splits (median, interquartile range (IQR) between the 25th and 75th percentiles, whiskers at 1.5×IQR). The boxes are grouped by the joint angeles and by the training models. The background displays the corresponing joint angeles of the right human arm. The colored bars along the Y axis show the range of RSME for the various joint angles.*

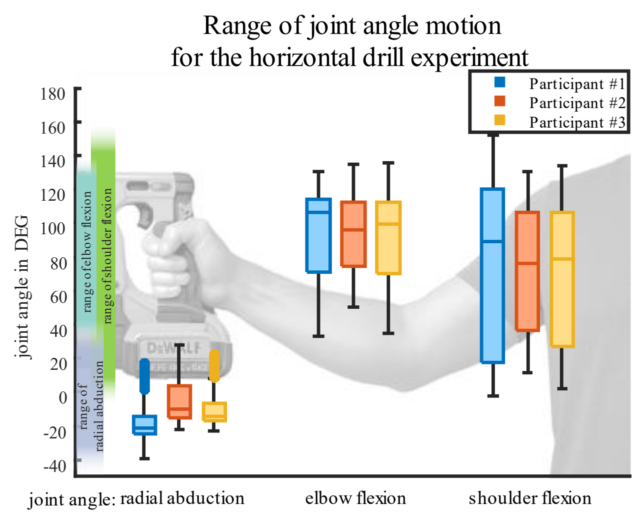

Figure 3: *Overview of the joint angle ranges measured with Xsens during the drilling tasks. Boxplots (median, interquartile range (IQR) between the 25th and 75th percentiles, whiskers at 1.5×IQR). grouped by joint angle illustrate the distribution of all measurement data across the three participants. Colored bars along the y-axis summarize the overall range of each joint angle.*

may alter the participants' natural behavior and thereby limit the transferability of the results to real-world working conditions.

Despite these differences in dataset size, no significant influence on prediction accuracy or robustness can be observed. Although participant #2, who contributed the smallest dataset, achieved the lowest RMSE values, participant #1—who provided a comparable number of samples—exhibited the

highest RMSE. However, the absolute differences between participants are small and do not indicate a disruptive influence of dataset size within the investigated range.

C. *Variation of the Number of Drilling Trials*

To evaluate the influence of data quantity, we analyze model performance using the median RMSE and robustness against varation of pressure distribution using the RMSE deviation across the varied datasets. Varying the number of drilling trials strongly affects model performance. A high number of trials leads to an improved performance for all models and participants. Figure 5 a) and b) illustrate this effect using boxplots of the RMSE for elbow flexion across increasing trial numbers, separated into DNN and SGPR models. Additional curves connecting median values across variation steps per participant visualize how trial number affects model RMSE. Both figures show similar trends for the parametric and non-parametric models, although the DNN generally achieves slightly lower RMSE values than the SGPR. With only a small number of trials, all models exhibit high error and low robustness, indicated by large box sizes and strong scattering across data splits. Increasing the number of trials reduces the median RMSE since it stabilizes itself at around 10° to 15° for the models trained with the total data sets.

D. *Variation of the Amount of Data per Trial*

The variation of the amount of data per trial indicates only a minor influence on the prediction performance. **Figure 4 a) and b)** show the effects on the RSME of predicting the elbow flexion when reducing the amount of data per trial. As mentioned in the methods, the variance of the data per trial in the raw data is aligned to get a comparable result. Each box represents a variation of the amount of data of 10 % of the average amount of data per trial for each participant.

However, decreasing the amount of data per trial shows less influence on the RMSE than decreasing the number of trials. For more than 100 data points per trial, no effect of the amount of data on the performance and robustness against varation of pressure distribution is visible. The DNN performs again better than the SGPR. However, the effect of a very small number of data – less than 100 data points per trial – effects the DNN models stronger.

IV. Discussion

The aim of this study was to evaluate the performance and robustness against varation of pressure distribution of human arm posture prediction from pressure distribution of gripping the power tool handle during usage. Helmstetter *et al.* [8] presented an approach how to train machine learning regression models for arm posture prediction. Similar approaches are used by Choffin *et al.* [9] and Alemayoh *et al.* [10] to predict gait poses from the pressure distribution at an insole. However, Helmstetter *et al.* [8] observed a high variance in the prediction accuracy of their person-specific trained models. By the

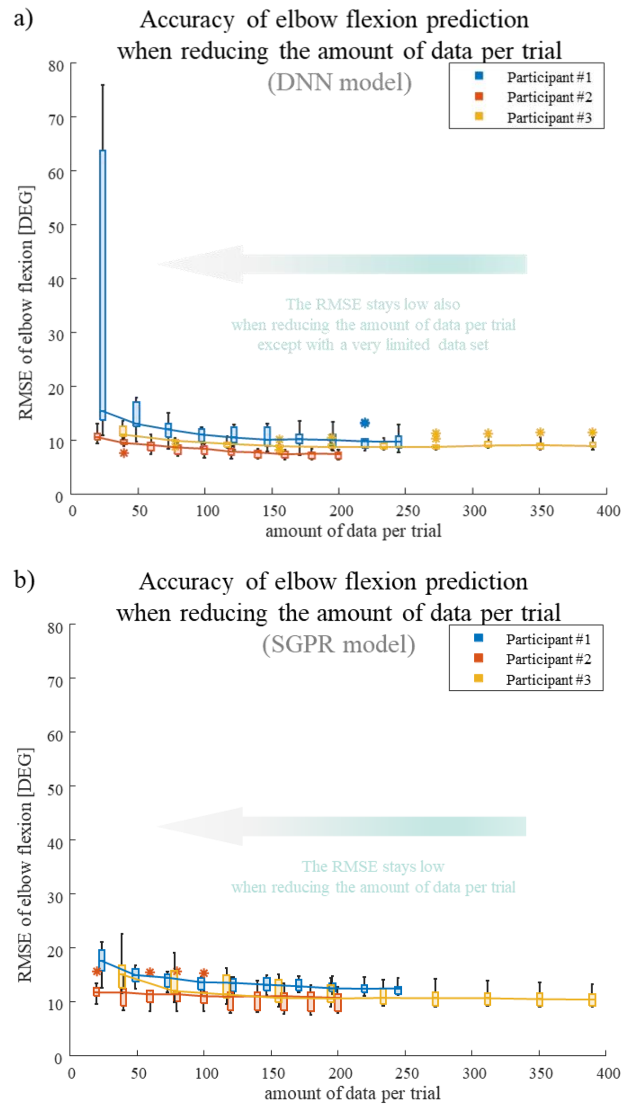

Figure 4: *Test RMSE for elbow flexion prediction as a function of the relative training data per trial. (a) Parametric DNN models. (b) Non-parametric SGPR models. Training data were reduced in 10% increments per participant. Due to participant-specific drilling speeds, absolute step sizes differ. To account for varying trial durations, all trials were resampled to the participant-specific mean drilling duration. Boxplots summarize the RMSE distribution across ten test splits (median, interquartile range (IQR) between the 25th and 75th percentiles, whiskers at 1.5×IQR). Values beyond this range are plotted as outliers. The participant-specific medians are linearly connected to visualize the trend with increasing data per trial.*

hypothesis, that the low robustness of the prediction approach by Helmstetter *et al.* [8] results from a too small data set and too low variance of the training data set, we tried to improve the arm posture prediction approach by Helmstetter *et al.* [8] with a large set of training data. We tried find out what are the main influence to find a training data set which is various enough for a precise and robust prediction and also as small as possible to reduce the effort for the data acquisition.

A more detailed analysis of joint-specific prediction performance reveals that, for all participants, the nominal RMSE for radial abduction is substantially lower than for elbow flexion and shoulder flexion. However, radial abduction also exhibits a considerably smaller range of motion during drilling. Consequently, the prediction task for this joint involved a narrower angular range. When normalizing the RMSE by the respective range of motion, comparable relative errors between 10% and 20% are obtained across all joints. These findings indicate that, with increased training data volume and enhanced variability across trials, the prediction approach exhibits substantially improved robustness compared to Helmstetter *et al.* [8]. This conclusion is further supported by the dedicated train–test split analysis. Across ten different splits, the standard deviation remains below 2°, as illustrated in **Table 4**, indicating stable and reproducible model performance.

As already discussed by Helmstetter *et al.* [8], absolute accuracy is critical when applying machine learning models for arm posture prediction from handle-integrated sensor data for ergonomic monitoring during power tool usage with ergonomic standards such as ISO 11226 [13].

When interpreting prediction accuracy, the measurement accuracy of the Xsens motion capture system must be considered. Xsens serves as ground truth in this study and is widely applied in biomechanics and ergonomic analyses. While it provides consistent relative measurements, systematic absolute errors have been reported for elbow flexion/extension. A systematic offset of 15.4° with a standard deviation of 2.1° has been validated against Vicon [14]. Consequently, the prediction accuracy achieved by our models is within the measurement uncertainty of the reference system. Compared to gait pose prediction and IR marker-based optical motion capturing, the accuracy of arm posture prediction is moderate but application-relevant [9, 10, 15].

Xsens accuracy is influenced by sensor placement, calibration quality, and the environment [14]. Furthermore, the integration-based measurement principle can lead to drift and error accumulation during long recordings. The influence of Xsens measurement uncertainty on model prediction performance must therefore be considered. In particular, larger deviations observed for wrist angle prediction may partly reflect uncertainty in the reference measurement. Comparable magnitudes of error between Xsens and the prediction models may lead to both underestimation and overestimation of true model error due to compensatory effects.

Using a reference system with lower measurement uncertainty for model training could reduce this limitation. Systematic measurement errors in Xsens do not directly bias model evaluation, as they are consistently present in both training and test datasets. However, statistical uncertainty in the reference system directly affects the achievable robustness of the prediction model. Even if, statistical noise can be partially averaged out during model training.

For future studies, training and validation using high-precision optical IR-marker-based motion capture systems such

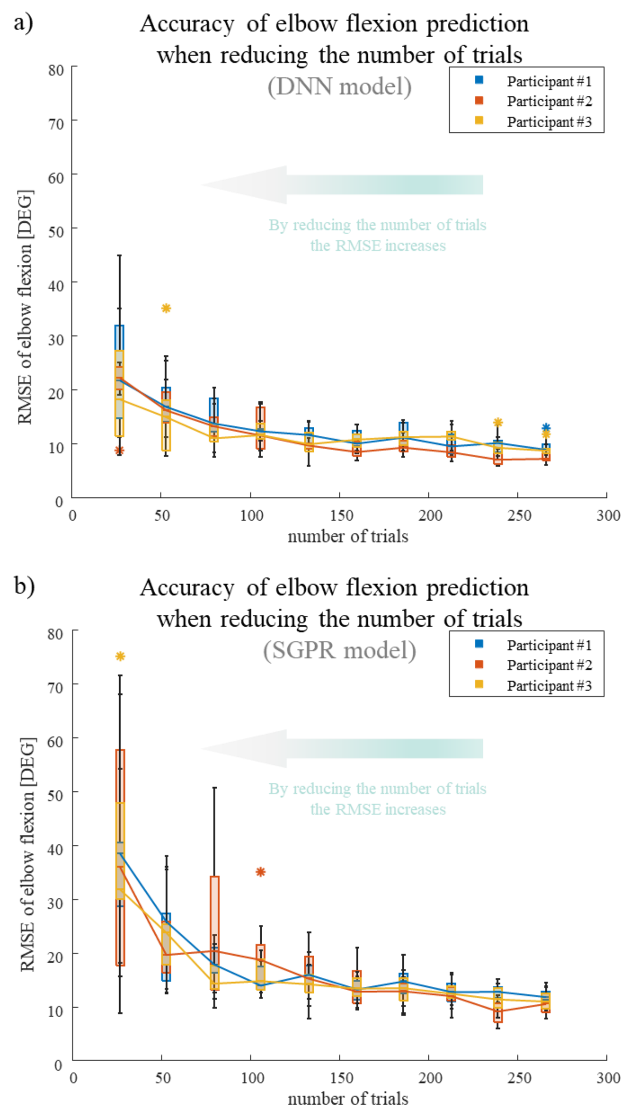

Figure 5: *Test RMSE for elbow flexion prediction as a function of the relative number of trials. (a) Parametric DNN models. (b) Non-parametric SGPR models. The number of trails are varied in ten steps with step sizes of 27 which represents 10 % of the overall number of trails for each participant. Each box (median, interquartile range (IQR) between the 25th and 75th percentiles, whiskers at 1.5×IQR) and the two whiskers contain the scatting results for ten variants of splitting the test and training data for the same configuration. Outliers are marked separately. The participant specific medians are linearly connected to visualize the trend with increasing number of trials.*

as Vicon are recommended. The superior precision of such systems would enable a more accurate quantification of the true achievable prediction performance of handle-integrated pressure sensing for human arm posture prediction.

A. Model performance is mainly influenced by the number of trials in the training data

To determine how the training dataset should be designed in order to achieve precise and robust models with minimal data acquisition effort, the influence of data quantity on prediction performance was analyzed by varying both the number of trials and the amount of data per trial.

The number of drilling trials exhibits the strongest impact on model performance. Starting from a high variability and a high mean RMSE when only few trials are used, both precision and robustness against varation of pressure distribution increase substantially with an increasing number of trials. The RMSE approaches asymptotically a value of approximately 10°. As the number of trials increases, the marginal benefit of additional trials decreases. As illustrated in Figure 5, the 255 trials used in this study appear to mark a saturation region beyond which the RMSE decreases only negligibly. At this point, the dispersion of the results has also stabilized. This effect is observed independently of whether a parametric DNN or a non-parametric SGPR model is used for training. However, a small number of trials appears to affect the prediction performance of the SGPR models more strongly than that of the DNN models.

Influence of the amount of data per trial: The results indicate that good model performance can already be achieved with a comparatively small number of datapoints per trial. This may be attributed to the limited variation in body posture and pressure distribution at the power-tool handle within a single trial. In the present experimental setup, 100 datapoints per trial correspond to 1.66 s at a recording frequency of 60 Hz. This duration is less than one third of the average drilling time of participant #2, who exhibited the shortest mean trial duration. Consequently, resampling the data per trial for evaluation purposes has no material impact on the reported metrics within the investigated range and does not substantially affect the results.

Due to varying drilling durations and the resulting differences in the number of recorded datapoints, a normalization strategy was required. Although resampling provides a practical solution, it may introduce distortions in the training dataset. In particular, interpolating small datasets to match larger ones can be problematic, as the generated datapoints closely resemble the original data and may not adequately reflect the natural variability. This limitation becomes particularly relevant when drawing conclusions about prediction quality in specific body postures, since the number of available datapoints and thus the effective training quality may depend strongly on the respective posture.

However, since all datasets were consistently evaluated jointly within the same training and test splits, the interpolation-induced bias affects all configurations in a comparable manner and therefore does not systematically favor one condition over another. As a result, relative performance differences between configurations remain interpretable in the overall analysis. Otherwise, the smallest dataset would implicitly serve as the reference, potentially biasing the comparison.

Overall, the findings indicate that future training of prediction models for arm posture based on pressure distribution at the power-tool handle should prioritize a dataset comprising a large number of independent trials performed in comparable postures. Provided that a minimum of 100 datapoints per trial is ensured, the overall data volume can be kept relatively small, as a reduction in datapoints per trial which is visualized in **Figure 4** does not lead to a significant degradation in prediction performance.

B. Toward Generalizable and Application-Ready Posture Prediction Models

The results demonstrate a clear convergence behavior once a minimum amount of measurement data is reached, while the additional influence of further datapoints per trial remains limited. In contrast, the number of independent trials has a substantial impact on prediction performance.

A general limitation of this study is that only horizontal drilling into a wooden beam was investigated comparable to the study by Helmstetter *et al.* [8]. Consequently, the external validity of the findings is restricted, and generalizability to other tasks, tool types, and user populations cannot be assumed without further validation. The transferability of the findings to other power-tool applications is therefore limited. Body postures, grip configurations, weight distributions of the power tool, and both magnitude and direction of the applied user forces may vary considerably across different applications. Even within drilling tasks, the results cannot be directly transferred to other working positions, such as vertical overhead drilling. It remains unclear whether pressure distribution alone is sufficient to enable accurate posture prediction under such conditions. To enhance robustness across varying applications without substantially increasing the training effort, additional sensors integrated into the power tool such as an inertial measurement unit (IMU) for determining tool orientation could provide complementary information and support prediction performance.

Another promising approach to significantly reduce overall training effort is the development of person-unspecific models [9]. Although such an approach may require a higher initial data acquisition effort, the resulting trained prediction models could subsequently be applied to a large number of users. The present finding that a high number of trials is required for accurate prediction suggests that the force distribution at the handle varies considerably even within a single participant across trials. It is therefore plausible that, given a sufficiently large number of trials, variability caused by inter-individual differences in anatomy and behavioral patterns could also be captured within a generalized prediction model.

V. Conclusion

This study investigated whether machine-learning models for predicting human arm posture during horizontal drilling can be trained efficiently while maintaining high precision and robustness against varation of pressure distribution using pressure distribution data from a power-tool handle. The results demonstrate that accurate and robust posture prediction is achievable when a sufficiently large training dataset is provided. A systematic variation of the number of trials and the amount of data per trial revealed that the number of independent trials is the dominant factor influencing model performance. Increasing the number of trials leads to a continuous reduction of the root mean square error (RMSE) prediction until an asymptotic saturation level is reached. The lowest achieved RMSE across participants ranged between 7.39° and 12.15°. In contrast, increasing the number of datapoints per trial had only minor influence on prediction accuracy and robustness against varation of pressure distribution, provided that a minimum threshold of approximately 100 datapoints per trial was exceeded. The achieved prediction accuracy is within the measurement uncertainty range of the IMU-based Xsens motion capture system used as ground truth, indicating application-relevant performance.

The validity of the presented models is currently restricted to horizontal drilling without impact mode. Future investigations should extend the dataset to additional power-tool applications and more diverse working postures. For larger datasets, DNN models demonstrated superior precision and robustness against varation of pressure distribution compared to SGPR models. In combination with interaction forces derived from the same handle-integrated pressure sensors, the predicted postures provide a basis for future product-integrated ergonomic assessment without external motion capture systems.

VI. References

[1] Eurofound, "European Working Conditions Telephone Survey 2021," Dublin, 2023. Accessed: Dec. 3 2024. [Online]. Available: https://www.eurofound.europa.eu/en/data-catalogue/european-working-conditions-telephone-survey-2021-0

[2] J. N. Chopp, S. L. Fischer, and C. R. Dickerson, "The impact of work configuration, target angle and hand force direction on upper extremity muscle activity during sub-maximal overhead work," *Ergonomics*, vol. 53, no. 1, pp. 83–91, 2010, doi: 10.1080/00140130903323232.

[3] N. D. Nath, T. Chaspari, and A. H. Behzadan, "Automated ergonomic risk monitoring using body-mounted sensors and machine learning," *Advanced Engineering Informatics*, vol. 38, pp. 514–526, 2018, doi: 10.1016/j.aei.2018.08.020.

[4] Z. Yang, D. Song, J. Ning, and Z. Wu, "A Systematic Review: Advancing Ergonomic Posture Risk Assessment Through the Integration of Computer Vision and Machine Learning Techniques," *IEEE Access*, vol. 12, pp. 180481–180519, 2024, doi: 10.1109/ACCESS.2024.3509447.

[5] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," 2018. Accessed: Feb. 3 2020. [Online]. Available: https://arxiv.org/pdf/1812.08008.pd

[6] M. F. Antwi-Afari *et al.*, "Deep learning-based networks for automated recognition and classification of awkward working postures in construction using wearable insole sensor data," *Automation in Construction*, vol. 136, p. 104181, 2022, doi: 10.1016/j.autcon.2022.104181.

[7] Z. Choffin *et al.*, "Ankle Angle Prediction Using a Footwear Pressure Sensor and a Machine Learning Technique," *Sensors (Basel, Switzerland)*, vol. 21, no. 11, 2021, doi: 10.3390/s21113790.

[8] S. Helmstetter, L. Neumann, J. Sembritzki, A. Lindenmann, and S. Matthiesen, *Human Arm Posture Estimation using only Pressure Sensors Integrated into Handles - A proof of concept*, 2025.

[9] Z. Choffin, N. Jeong, M. Callihan, E. Sazonov, and S. Jeong, "Lower Body Joint Angle Prediction Using Machine Learning and Applied Biomechanical Inverse Dynamics," *Sensors (Basel, Switzerland)*, vol. 23, no. 1, 2022, doi: 10.3390/s23010228.

[10] T. T. Alemayoh, J. H. Lee, and S. Okamoto, "A Neural Network-Based Lower Extremity Joint Angle Estimation from Insole Data," in *2023 20th International Conference on Ubiquitous Robots (UR)*, Honolulu, HI, USA, 2023, pp. 787–791.

[11] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[12] A. Lindenmann, M. Uhl, T. Gwosch, and S. Matthiesen, "The influence of human interaction on the vibration of hand-held human-machine systems - The effect of body posture, feed force, and gripping forces on the vibration of hammer drills," *Applied ergonomics*, vol. 95, p. 103430, 2021, doi: 10.1016/j.apergo.2021.103430.

[13] Internationale Organisation für Normung, "Ergonomics - evaluation of static working: = Ergonomie - évaluation des postures de travail statiques," Geneva ISO 11226:2000(E), [2000].

[14] A. Bicchi and A. Colombo, "Improved Estimation of Elbow Flexion Angle from IMU Measurements Using Anatomical Constraints," *IRBM*, vol. 45, no. 1, p. 100820, 2024, doi: 10.1016/j.irbm.2024.100820.

[15] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A Study of Vicon System Positioning Performance," *Sensors*, vol. 17, no. 7, 2017, doi: 10.3390/s17071591.