2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Automatic Trust Estimation From Movement Data in Industrial Human-Robot Collaboration Based on Deep Learning

Matthias Rehm[1], Ioannis Pontikis[1] and Kasper Hald[1]

***Abstract*— Trust in automation is usually assessed with post-interaction questionnaires. For human robot collaboration it would be beneficial to assess the trust level during the interaction to adjust the robot's collaboration behavior to the user expectations. In this paper we investigate if trust can be estimated from observable behavior like movements during the interaction with a large industrial manipulator. To this end, we report on a data collection for two tasks during collaborative draping, the transport of large cut pieces and the actual draping process in close proximity to the robot. The data is used to train and compare different deep learning models. Results show that automatic trust estimation is feasible, which opens up to using trust as a parameter for informing the interaction with robots.**

I. INTRODUCTION

Trust is an emergent factor in human robot collaboration research (e.g. [1], [2]). Whereas trust in automation is used as an evaluation criterion for system design (e.g. [3]), trust in human robot collaboration has the potential of becoming a parameter that allows real-time adaptation of robot behavior to the user's trust level. It has been shown that collaboration is challenged when subjective trust perception and actual trustworthiness of the system do not match [4], possibly leading to under-usage or wrong usage of the robot or even to safety concerns regarding equipment or health. For the envisioned purpose of real-time adaptation, it becomes indispensable to estimate trust levels during the interaction with the robot. To this end, we need to find observable behavior that can be linked to trust and can then be exploited in data-driven trust assessment. In the Drapebot project[1], we are investigating collaborative draping of carbon fibres, where a worker collaborates with a large industrial manipulator. The robot is supposed to assist the human worker during the transport of large material patches and during the draping process itself. Specific emphasis is put on trust and usability, due to the complexity of the task and the sizes of the involved robots (see Figure 3). In this paper, we present an approach for automatic trust assessment based on motion tracking data collected in the two Drapebot tasks, i.e. transport and draping. The goal of the trust assessment system is the utilization of body tracking to calculate the operator's trust in the robot in real time during the draping process. By tracking the trust level throughout the collaboration session, based on the baseline trust of the operator, we can adapt the robot behavior according to our strategy for maintaining trust. An example of the strategy could be recognizing signs of decreases in or disruptions of trust in the robot. When this occurs, we can send a signal to the robot's operating system that the robot behavior should change, and/or signal to the operator to be more careful around the robot. The information flow is exemplified in Figure 1. This paper gives details about the data collection, training, and evaluation of the first version of the trust assessment system.

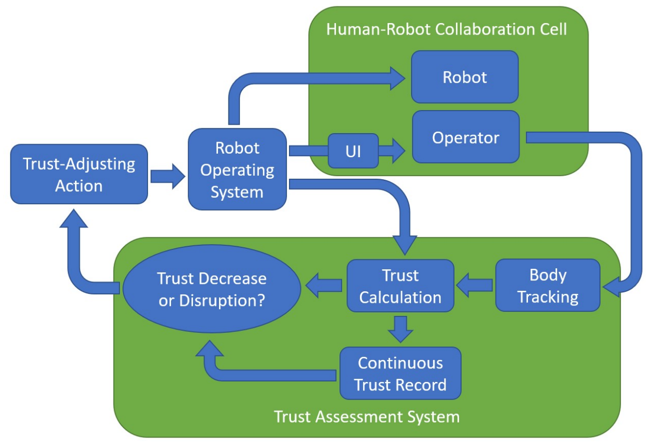

Fig. 1. The trust assessment system and how tracking data and trust calculations can be used to calibrate trust by communicating with the operator and/or by changing the robot's behaviour.

II. RELATED WORK

Trust definition: A general definition of trust is based on a trustor-trustee relationship that includes elements of risk and vulnerability on the side of the trustor [3]. A concrete example is how a worker (trustor) trusts a robot (trustee) to avoid collisions (risk) in a shared workspace and thus makes him/herself vulnerable if the robot is not able to do so. Trust is not a new topic in Robotics, but has mainly been looked at in relation to automation of processes. For instance, Muir and Moray [5] explain trust in an automated system to be mainly based on the human's perception of its reliability and performance. Thus, in automation we can see a strong focus on robot-related aspects that result in high or low trust towards the machine. With the advent of collaborative robots on the factory floor, focus is shifting towards trust in human robot interaction, and one of the first detailed analyses was presented by Hancock and colleagues [6] that distinguish between human-related (e.g. prior experience,

[1]Faculty of IT and Design, Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark matthias@create.aau.dk, cgcu@create.aau.dk

*This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101006732 (Drapebot).

[1]https://drapebot.eu

Fig. 2. Xsens tracking suit with IMUs

personality traits), robot-related (e.g. reliability, adaptability) and environment-related factors (e.g. communication, task complexity) that influence trust in the robot during human robot interaction.

Trust assessment tools: Over the last decade a number of trust assessment questionnaires have been developed to specifically assess trust in human robot interaction. For instance, Malle and Ullman [7] present the multidimensional model of trust that distinguishes between performance- and moral-based trust and is well suited for contexts with social robots. Charalambous and colleagues [8] focus instead on developing a trust scale for industrial contexts with collaborative robots and features for instance specific questions about reliability of grippers. Schaefer [9] presents a trust measurement scale that is based on Hancock et al.'s [6] analysis of factors that influence trust and has been broadly used in HRI.

Data driven trust assessment: First attempts have been reported for mapping sensor data to states of trust that exemplify the feasibility of this challenge. Hu and colleagues relate real-time sensing of physiological signals as well as EEG to human-machine trust [10]. Their trust measurement is unfortunately very shallow, asking participants merely if they trust the system or not. Rahman and colleagues [11] develop a trust measurement that is based on robot and human performance variables, where the real time trust measurement of the human is based on hand speed during part manipulation and handover tasks. Hand speed is measured through IMUs attached to the hand. The approach is evaluated by a single item Likert scale asking how much the user trusts the robot. Onnasch and Hildebrandt [12] investigate the robot-related trust factor anthropomorphism. They base their trust estimation in a handover task on the time that the user has both hands in the handover space, where shorter times indicate higher trust. The trust estimation was evaluated against Charalambous et al.'s [8] trust scale for collaborative robots. Shayesteh and colleagues [13] developed a masonry task in VR supported by a robot for testing the effect of different levels of autonomous movement of the robot on trust. Ground truth is collected through Schaefer's [9] Trust Perception Scale - HRI. This information is then used to label the EEG data obtained in the two different scenarios and subsequently to train machine learning algorithms for trust estimation. The low number of participants for the study makes it difficult to verify if the positive results are not due to overfitting. Previously, we have reported on relevant movement parameters that can be extracted from motion tracking data for the task of trust assessment [14]. Based on this work the data collection and feature extraction for the automatic trust assessment that is described in the next sections has been developed.

III. DATA COLLECTION

The training data for the trust assessment system was collected in four different human-robot collaboration scenarios that simulated the transport and draping task for carbon fibre draping: collaborative transport of large fabric with a slow (1) and a fast (2) robot and collaborative draping of fabric with a (3) slow and a (4) fast robot. This section describes the details of the tracking techniques, experimental procedures, and collaboration scenarios.

A. Tracking Hardware

For all sessions, body tracking was performed with the Xsens MVN Awinda tracking suit. It consists of a tight-fitting shirt, gloves, headband, and a series of straps used to attach 17 IMUs to the participant. The tracking suit with the IMUs is shown in Figure 2. After calibration, the system uses inverse kinematics to track and log the movements of the participant at a rate of 60 Hz. The measurements include linear and angular speed, velocity, and acceleration of every skeleton tracking point. For the most accurate tracking the participant's body dimensions are measured before the session. These measurements include height, foot length, shoulder height, shoulder width, elbow span, wrist span, arm span, hip height, hip width, knee height and ankle height. The system determines position by tracking feet and steps along a plane. This method is, however, vulnerable to position drift over time. The system allows for integration of SteamVR trackers and lighthouses, allowing for position-aiding by attaching a SteamVR-compatible tracker to the participant in addition to the 17 IMUs.

B. Experimental Procedures

After taking their measurements and preparing the tracking setup, the participants were introduced to the procedure of their first collaborative tasks. At the end of each session the participants filled out a trust questionnaire. The participants' trust towards the robot was assessed using Schaefer's [9] Trust Perception Scale - HRI, where the final score is calculated based on a series of questions regarding the operator's perception of the robot's capability, predictability, expected error rates and more. The full questionnaire consists of 40 questions, but we used the shortened version for repetitive tasks consisting of 14 questions. The final trust score is on

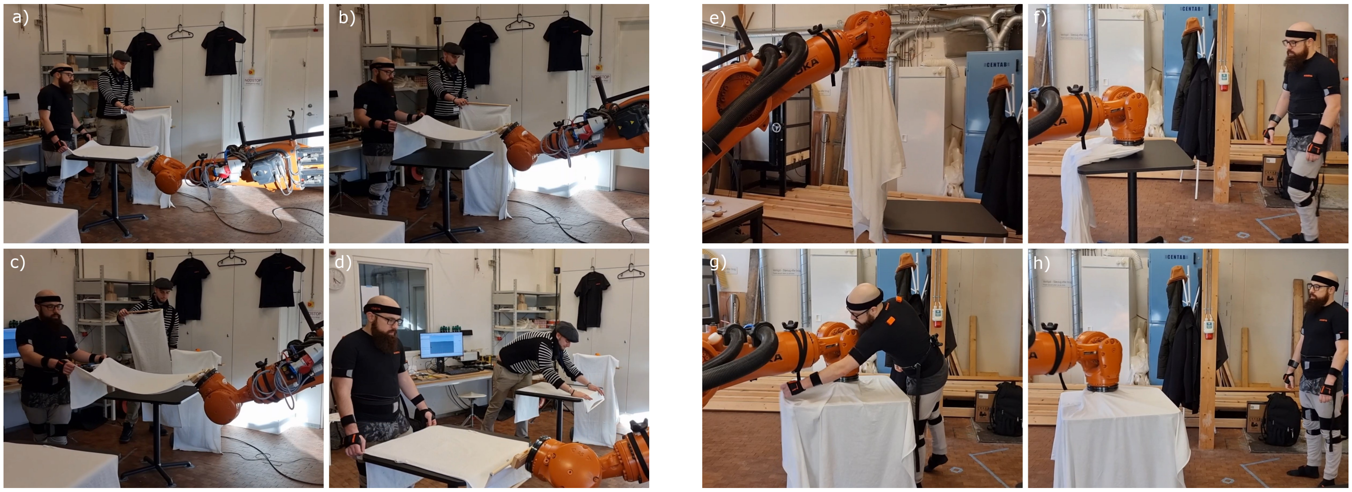

Fig. 3. The collaborative transportation (left) and draping (right) task that were performed. Left: Participant and robot take the cloth from the first table (a+b). The participant follows the robot for moving the cloth to the second table (c). Participant and robot position the cloth on the second table (d). Right: The robot positions the cloth on the table and keeps it in place (e+f). The participant moves to the table and begins draping the cloth on the table (g). Then the participant moves to the safe area before the robot is activated again (h).

a scale between 0 and 100, where the operator's agreements to statements suggesting trust in the robot are weighted positively, and statements suggesting expectations of errors are weighted negatively.

C. Robot Collaboration Tasks

For the tracking sessions, participants performed the two collaborative tasks repeatably for ten minutes each[2]. To counter bias from gaining experience with the robot over time, half the participants performed the collaborative transportation task first, while the other half started with the draping task.

1) Collaborative Transport: The transport task is set up with two tables two meters apart. The robot is on one side of the tables, the participant is on the other. The robot is at a distance from the tables where it can just reach the closest edge of both tables. The participant is instructed to always stay on their side of the tables, and a border line drawn on the floor. The distance is such that the participant must reach out slightly to reach the edge of the tables (see Figure 3 left). The transport task is done with towels, 120 cm long and 55 cm wide with a wooden dowel attached between the corners at one end. The test conductor lays the towels out on the starting table with the dowel hanging off the edge on the robot side. The robot is equipped with two hooks at the end effector, allowing it to scoop up the dowel and hold on to the towel by angling up the hooks slightly. The participant is instructed to follow the pace of the robot and grab the end of the towel as the robot does the same. The angling of the robot's hooks allows the participant to hold up the towel. Once the robot has picked up the towel, it starts moving the end-effector toward the goal table while keeping the same orientation until it reaches the end table where it points the end-effector downwards, letting go of the towel before returning to its starting position and repeating the task.

2) Fabric Draping: The draping task is performed at a table. The robot has a tablecloth attached to the end-effector using magnets. On the opposite side of the table a rectangular safe area is marked, allowing the participant to comfortably stand. They are instructed not to leave the marked area at any time while the robot is moving (see Figures 3 right). When activated by the test conductor, the robot moves so that the end-effector stops over the surface of the table, pointed downwards, allowing the cloth to cover the table. The participant is instructed to approach the table once the robot has come to a complete stop and lightly tug at the cloth and cover the table as evenly as possible before returning to the marked area. When the participant is in the safe area, the operator activates the robot's pre-programmed movement cycle, where it moves away from the table, removing the cloth, before returning it to the resting position above the table surface for another repetition.

TABLE I

SUMMARY OF DATA COLLECTION

	Transport		Draping		Total
	slow	fast	slow	fast	
Repetitions	189	290	241	231	951
Data points	265680	290880	261960	264660	1083180

3) Collected Data Samples: So far, data has been collected in the transport and draping tasks with both a slow and fast robot. Robot speed was used to induce different levels of trust. We have shown earlier that speed is a good predictor of trust levels [15]. Data was collected with the XSENS software with 60Hz, every task took on average 18-22 seconds. Data has been collected from 20 participants, 7 female and 13 male, average age 25 (SD = 4.0). Average

[2]A demo video of the data collection tasks is available: https://youtu.be/5pAXF7t1Us4

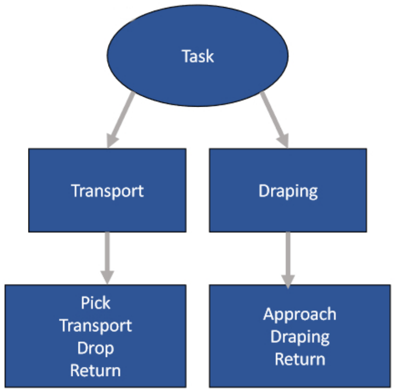

Fig. 4. Overview of the task used in data collection. Transport task consists of four phases: pick, transport, drop, return. Draping task consists of three phases: approach, drape, return.

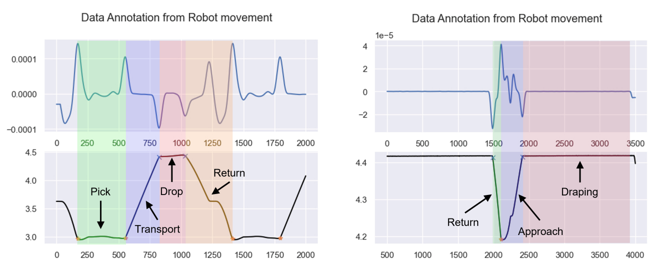

Fig. 5. Segmentation of trials and phases. Top row shows the robot movement on the x-axis, the bottom row the second derivation of this. The different phases are illustrated in the bottom row.

height was 1.74 meters (SD = 0.1). Participants either interacted with a slow or fast robot and the starting condition (transport or draping) was counterbalanced in each group. One session during the data collection consists of 24 trials on average for the transport and draping task resulting in 951 trials across all conditions. Table I summarizes the number of data that was collected. A detailed description of the data set can be found in [16][3]

IV. DATA SEGMENTATION

One session during the data collection consists of 24 trials on average for the transport and draping task which is recorded as a continuous data stream. This stream has to be segmented into the individual trials. Moreover, each trial consists of different phases that might prove relevant for a more fine-grained analysis (Figure 4). This segmentation can be done automatically making use of the XSENS IMU unit that was placed on the robot and captured the movement of the end-effector in relation to the user. Movement of this sensor on the x-axis was used as the indicator of the task (top row of Figure 5). Peaks and minimums of the robot movement on the x-axis were found with the help of the second derivation of the movement positions (bottom row of Figure 5). Transport was divided into pick, transport, drop and return, while draping phases are approach, draping and return. Figure 5 exemplifies this approach with trials from both transport and draping.

[3]The data set is available on Zenodo: https://doi.org/10.5281/zenodo.8224067.

TABLE II

SUMMARY OF LOW AND HIGH LEVEL MOVEMENT PARAMETERS.

Low-level	High-level
Kinematic/Dynamic descriptors	**Body descriptors**
Duration	Action presence
Velocity*	Center of mass displacement*
Acceleration*	Balance*
Jerk*	Support
Curvature*	
Quantity of movement*	
Geometric descriptors	**Space descriptors**
Bounding box*	Distance covered*
Bounding sphere	Area covered
Bounding ellipsoid*	Hip height
Convex hull	
Displacement*	
Rotation	
Center of mass	
Effort descriptors	**Shape descriptors**
Weight effort	Bounding volume*
Time effort*	Shape direction*
Space effort	Shaping
Flow effort*	Extensiveness*
	Arm shape
	Elbow shape
	Shoulder angle
	Hands relationship
	Feet relationship

Parameters used for training are marked by *

V. FEATURE EXTRACTION

Laban movement parameters serve as the basic features for trust assessment. Laban Movement Analysis[4] (LMA) was originally conceived as a tool for assessing expressiveness in dance (e.g. [17]). LMA's foundation is the interplay between the body, the surrounding space, the shape of the movements, the effort involved, the phrasing, or patterns, as well as the relationship with other participants. A number of quantifiable, computable parameters have been derived from LMA to describe and analyze motion. Based on a review by Larboulette and Gibet [18] and it's application by Hald and Rehm [14], we identified parameters that can be calculated from the collected body tracking data. From this set, we have selected 16 parameters that proved most relevant [14]. These are marked with * in table II. Features for these parameters are calculated for the different joints and in case of the low level parameters for different axes, e.g. shape directional (high level) for left upper arm and right upper arm or pelvis jerk (low-level) for x-, y-, and z-direction. The final set consists of 191 features. For illustrative purposes we show details here the spatial descriptor Distance covered. Distance covered was calculated on the center of mass as the cumulative sum of distances between the frames and is visualized in Figure 6 for both the transportation and draping task.

[4]https://www.laban-eurolab.org/lbms/theory/

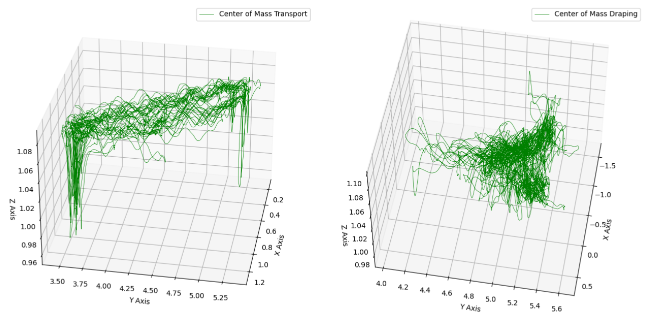

Fig. 6. Example of covered distance during transport and draping trial.

VI. DEEP LEARNING

1) Data preprocessing: To facilitate data preparation for deep learning models, a preprocessing methodology was devised. The primary objective was to maintain the data in a state as close to its raw form as possible subsequent to feature extraction. Consequently, no normalization or other transformation techniques were applied to the dataset. This approach was adopted to ensure that the inherent characteristics of the data were preserved for subsequent analysis.

Two distinct data structuring approaches were examined during the course of this study. The initial approach involved aggregating all available data into a unified dataset for classification purposes. For this, all the position, acceleration and velocity data for each sensor were concatenated in a single row in a data frame for each subject. The final dataset was then a unified matrix with all subjects concatenated into a single data frame, which was then randomly split into training, test, and validation sets. The same approach was taken for the LMA feature data, where the 191 features were calculated for each time frame and saved as a row in the data frame for each subject.

The second approach partitioned the data into discrete windows of 200 entries (ca. 3.3 seconds) with a 50% overlap between adjacent windows. This windowing strategy was employed to capture of temporal dependencies in the data, preserving its sequential nature. Each window corresponds to a fixed-length segment of the time series, enabling the model to discern and learn patterns and interrelationships within these defined temporal segments.

Two validation approaches were used, the first one trained the models on random 75% of the data and validated the models on the remaining 25%. With this approach, the validation set contained data from subjects that were also present in the training set. For a better generalizability, the second approach trained the models on 15 randomly selected subjects and validated with the data from the remaining five unseen subjects.

2) Classification: The data were labeled as a two-class problem, which corresponds to high trust and low trust. Given that the trust score range obtained from the trust questionnaires was standardized from 0 to 1, a systematic approach was implemented. The median value of all participants was computed for each task and served as the pivotal threshold. Instances with trust scores below the median were designated as representing low trust, while those surpassing the median were categorized as indicative of high trust. This methodology was employed to ensure a uniform and impartial classification process, grounded in the trust scores acquired from the trust questionnaires.

3) Neural Networks: In the deep learning approach employed for this study, two distinct model architectures were tested: Fully Connected Neural Network (FCNN) and Convolutional Neural Network (CNN), which is illustrated in Figure 7. For the FCNN we use 4 layers with 512, 256, 256, 128 nodes and ReLU as an activation function in each layer. The fifth last layer is the output with activation function Softmax and size 2. The CNN accordingly has 3 convolutional layers of size 32, 48, 64, a kernel of (1, 5) and also uses ReLU as activation function. After the convolution process we flatten the data in order to pass them through a dense output layer. The structural design of these models adheres to the fundamental architecture of their respective neural network types, with only minor hyperparameter adjustments introduced. It is essential to emphasize that this approach was conceptualized as a proof of concept rather than a comprehensive exploration of hyperparameter optimization on an established dataset. Consequently, only minimal hyperparameter tuning was undertaken during the model development phase.

4) Results: For the first approach with the FCNN model, the raw data were used. As can be seen in Table III, this resulted in a high accuracy. In this approach all the data from all participants were concatenated and a random 25% of the data was separated for validation purposes. The assessment of all the models was done by comparing standard metrics precision, recall, F1-score and the overall accuracy of the models.

TABLE III
RESULTS OF FCNN MODEL ON RAW DATASET.

Task	Class	Prec.	Recall	F1	Validation
Transp.	low	0.90	0.89	0.89	Random
	high	0.89	0.91	0.90	Random
	accuracy	0.90			
Draping	low	0.88	0.87	0.87	Random
	high	0.87	0.88	0.87	Random
	accuracy	0.87			

In the second approach, following the initial findings of high accuracy in the first approach, a critical assessment revealed the need for a more robust validation procedure. It became apparent that the dataset lacked a crucial element: validation with unseen data. The dataset had been originally amalgamated, and the validation subset thus contained data from the same participants as the training dataset, raising concerns about potential bias and overfitting. In response to these concerns, the second approach involved a refinement of the dataset organization. Specifically, the data were partitioned on a per-subject basis, ensuring a more rigorous

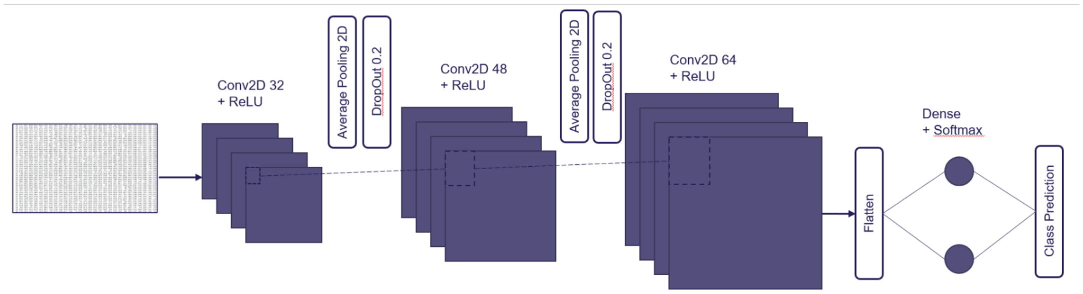

Fig. 7. Convolutional neural network structure.

evaluation process. Subsequently, the results obtained from the validation set were meticulously compared to those of the previous approach, contributing to a more comprehensive and reliable assessment of the model's performance. Lastly, to capture the dynamics of trust, the windowing approach was used in combination with a CNN. Windowed data with an overlap should be more suitable with CNNs for time-series data in order to generalise the task assigned. The results are shown in Table IV where the two validation methods are compared. We can see a huge drop in accuracy when the models are validated on the unseen data.

TABLE IV
RESULTS OF CNN MODEL WITH RAW WINDOWED DATASET.

Task	Class	Prec.	Recall	F1	Validation
Transp.	low	0.94	0.89	0.91	Random
	high	0.89	0.95	0.92	Random
	accuracy 0.92				
	low	0.41	0.64	0.50	unseen
	high	0.54	0.31	0.39	unseen
	accuracy 0.45				
Draping	low	0.88	0.92	0.90	Random
	high	0.91	0.88	0.89	Random
	accuracy 0.90				
	low	0.32	0.50	0.39	unseen
	high	0.73	0.56	0.64	unseen
	accuracy 0.55				

The third approach was based on the extracted LMA features, which were then compared with the outcomes obtained from the raw data models. Moreover, we took the dynamic development of trust during the data collection into account for data processing. We have shown earlier how trust evolves over time [15]. In line with this premise, we operated under the assumption that trust levels would begin low at the outset of each experimental trial and gradually increase as trust was build up throughout the experiment. Consequently, for the subjects who exhibited low trust scores in the trust assessment questionnaires, we selected the initial five iterations as representative of low trust. For subjects with high trust scores, we chose the final five iterations, which signified the culmination of trust development, to include in the training process. The results of this approach are presented in Table V, where we compare the model accuracy when it is trained on all the data, and when it is trained on the data that takes the dynamic trust development into account. It is worth noting that this categorization approach yielded higher accuracy and fostered more robust generalization. Accuracy is at 0.98 and 0.95 respectively for the two different tasks when validated on unseen data.

TABLE V
RESULTS OF CNN MODEL WITH LMA DATASET

Task	Class	Prec.	Recall	F1	Validation
Transp.	low	1.00	0.86	0.92	unseen
	high	0.98	1.00	0.99	unseen
	accuracy 0.98				
Draping	low	0.91	0.95	0.93	unseen
	high	0.97	0.96	0.97	unseen
	accuracy 0.95				

5) Future Work: Additional data has been collected in close-to-production settings at partners Profaktor and DLR. This data will be used to evaluate the generalizability of the current models and if necessary to modify and retrain them. It will also allow to compare the trust evaluation for naive users, which were used in this experiment, and trained operators, which participated in the new data collections.

VII. CONCLUSION

In this paper we presented an approach for automatic trust assessment from motion tracking data. We compared three approaches, two that trained on the raw data, one that was trained on features derived from Laban Movement analysis and that took the dynamics of trust development into account. The last approach delivers the best results also for unseen data with an accuracy of 0.98 and 0.95 respectively for the transport and draping tasks. In this paper we reported on our first proof of concept for classification of sensor data into discrete trust levels. For truly monitoring trust throughout the interaction, we need to move to architectures that capture the dynamics of trust more accurately and allow for estimating continuous trust levels during the interaction. The main conclusion from this paper is that automatic data-driven trust assessment from movements is feasible. Based on this result it becomes possible to investigate how trust can be used as a control parameter in human robot interaction.

REFERENCES

[1] M. Lewis, K. Sycara, and P. Walker, *The Role of Trust in Human-Robot Interaction*. Cham: Springer International Publishing, 2018, pp. 135–159.

[2] C. S. Nam and J. B. Lyons, Eds., *Trust in Human-Robot Interaction*. Academic Press, 2021.

[3] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human factors*, vol. 46, no. 1, pp. 50–80, 2004.

[4] E. J. De Visser, M. M. Peeters, M. F. Jung, S. Kohn, T. H. Shaw, R. Pak, and M. A. Neerincx, "Towards a theory of longitudinal trust calibration in human–robot teams," *International journal of social robotics*, vol. 12, no. 2, pp. 459–478, 2020.

[5] B. M. Muir and N. Moray, "Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation," *Ergonomics*, vol. 39, no. 3, pp. 429–460, 1996.

[6] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.

[7] B. F. Malle and D. Ullman, "A multidimensional conception and measure of human-robot trust," in *Trust in human-robot interaction*. Elsevier, 2021, pp. 3–25.

[8] G. Charalambous, S. Fletcher, and P. Webb, "The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 193–209, 2016.

[9] K. E. Schaefer, *Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"*. Boston, MA: Springer US, 2016, pp. 191–218. [Online]. Available: https://doi.org/10.1007/978-1-4899-7668-0_10

[10] W.-L. Hu, K. Akash, N. Jain, and T. Reid, "Real-time sensing of trust in human-machine interactions," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 48 – 53, 2016.

[11] S. M. Rahman, Y. Wang, I. D. Walker, L. Mears, R. Pak, and S. Remy, "Trust-based compliant robot-human handovers of payloads in collaborative assembly in flexible manufacturing," in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2016, pp. 355–360.

[12] L. Onnasch and C. L. Hildebrandt, "Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 11, no. 1, pp. 1–24, 2021.

[13] S. Shayesteh, A. Ojha, and H. Jebelli, "Workers' trust in collaborative construction robots: Eeg-based trust recognition in an immersive environment," *Automation and robotics in the architecture, engineering, and construction industry*, pp. 201–215, 2022.

[14] K. Hald and M. Rehm, "Determining movement measures for trust assessment in human-robot collaboration using imu-based motion tracking," in *Proceedings of ROMAN*, 2023.

[15] K. Hald, M. Rehm, and T. Moeslund, "Human-robot trust assessment using motion tracking & galvanic skin response," in *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

[16] M. Rehm, K. Hald, and I. Pontikis, "Benchmark movement data set for trust assessment in human robot collaboration," in *Proceedings of HRI 2024*. United States: IEEE, 2024.

[17] A. L. de Souza, *Laban Movement Analysis—Scaffolding Human Movement to Multiply Possibilities and Choices*. Cham: Springer International Publishing, 2016, pp. 283–297.

[18] C. Larboulette and S. Gibet, "A review of computable expressive descriptors of human motion," in *Proceedings of the 2nd International Workshop on Movement and Computing*, 2015, pp. 21–28.