Investigating Electrodermal Activity for Trust Assessment in Industrial Human-Robot Collaboration

Campagna, Giulio; Chrysostomou, Dimitrios; Rehm, Matthias

Published in:
21st International Conference on Ubiquitous Robots (UR 2024)

Publication date:
2024

Document Version
Accepted author manuscript, peer reviewed version

Investigating Electrodermal Activity for Trust Assessment in Industrial Human-Robot Collaboration*

Giulio Campagna[1], Dimitrios Chrysostomou[2], Matthias Rehm[1]

Abstract— **In the Industry 5.0 framework, due to the close collaboration between humans and robots, providing a safe environment and balance workload becomes an essential requirement. In this context, evaluating the trustworthiness of robots from a human-centric perspective is essential as trust impacts the interaction in human-robot collaborations. Numerous researchers in the literature have delved into physiological responses as indicators of user trust in robots. In this research endeavor, multiple machine learning models were employed, leveraging skin conductance response (SCR) to classify the trust level of the human operator. A chemical industry scenario was developed, where a collaborative robot supported a human operator by handing over a beaker used for the pouring of chemicals. The machine learning models achieved a moderate accuracy rate of 68.99% and AUC of 0.73 for the hand-over task. Nonetheless, this study underscores the importance of sensor fusion techniques to improve the accuracy of trust assessment within the context of human-robot collaborations.**

I. Introduction

Industrial environments are transitioning to Industry 5.0 paradigms, focusing on resilience, sustainability and human-centered production. This transition requires seamless and safe human-robot collaboration (HRC) while maintaining appropriate trust levels between human operators and robots [1]. Trust is a pivotal component that enables productive collaboration between human coworkers and robotic assistants in industrial settings. When human operators demonstrate an appropriate level of trust in robots, it enables them to rely on the robots' capabilities for reliable execution of collaborative tasks, thus facilitating effective coordination and seamless teamwork dynamics [2]. However, excessive trust can lead to complacency and safety risks, while insufficient trust impedes collaboration and overburdens human operators [3]. Therefore, effectively evaluating trust levels during real-time human-robot interactions (HRIs) and maintaining appropriate trust calibration are crucial prerequisites for ensuring safe coordination in HRC across next-generation smart factories and Industry 5.0 production facilities [4].

Trust, as a multidimensional concept, has been extensively studied in human interactions with automated systems and intelligent robots. Seminal works by Muir and Moray [5] and Lee and See [6] established the theoretical foundations where trust is defined as an operator's willingness to act on an automated system's recommendations in situations involving vulnerability, uncertainty, and risk. Factors affecting trust in HRC can be categorized into human-related, robot-related, and environmental clusters [7], with specific factors such as transparency of robot intentions [8], robot appearance [9], robot speed [10], and task importance [11] critically influencing trust. Recent research has shown that factors such as proxemic distances and risks associated with unexpected robotic movements, can significantly influence user trust, reducing trust levels during collaborative activities [10], [12]. Many prior studies have relied on post-interaction, retrospective evaluations of trust levels using surveys, and questionnaires [13], [14]. Although these tools are well-established tools for assessing trust, they cannot capture the trust fluctuations during dynamic interactions.

Electrodermal Activity (EDA) and its phasic component, Skin Conductance Response (SCR) offer potential in evaluating emotional arousal, anxiety, cognitive load, and stress responses, which are key indicators of trust [15], [16]. While EDA has been extensively used in emotion recognition research [17], [18], its application to trust estimation has been relatively unexplored. Studies have investigated the correlation between EDA and trust in various contexts, such as text-based games [19], semi-automated robot operation [20], and autonomous driving scenarios [21], [22], achieving accuracies ranging from 73% to 81.6% in trust classification using EDA features, sometimes in combination with other measurements. However, these studies faced challenges in data labeling, relying solely on user responses to assess trust.

This research aims to explore the use of SCR data as an implicit, objective psycho-physiological indicator of trust levels during HRIs in industrial settings. The study employs a representative scenario from the chemical industry, involving close coordination and collaboration between humans and robots during the handling of hazardous chemicals inside a beaker, where unexpected robotic movements may signify a potential decline in trust from the human perspective.

This study integrates skin conductance measurements from wearable sensors and machine learning techniques to build a data-driven framework for categorizing trust levels based on the implicit psycho-physiological cues and reactions of the human operator. Monitoring of SCR data during collaborative tasks allows the system to infer trust dynamics and adapt the robotic controller's behavior to maintain calibrated trust levels, ensuring enhanced safety and balanced workload coordination during close collaboration in Industry 5.0 environments. The key technical contributions are:

- Devising a data-driven framework combining SCR data streams and machine learning algorithms to categorize

* The work described in this paper was funded by the Independent Research Fund Denmark, grant number 1032-00311B.

[1] are with the Technical Faculty of IT and Design, Aalborg University, 9000 Aalborg, Denmark. {gica, matthias}@create.aau.dk.

[2] is with the Faculty of Engineering and Natural Sciences, Aalborg University, 9220 Aalborg, Denmark. dimi@mp.aau.dk.

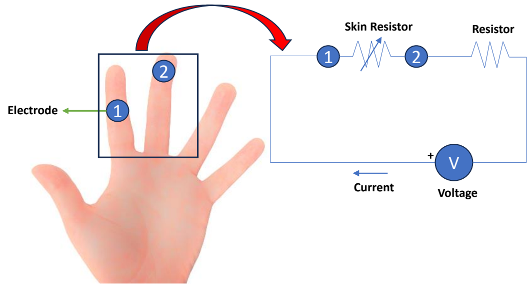


Fig. 1: The electric scheme for the Shimmer3 GSR+ Unit.

trust levels based on implicit psycho-physiological reactions and responses.
- Training and evaluating the proposed framework on SCR dataset systematically labeled through experimental conditions, to reliably map physiological cues to corresponding trust levels.

Research on wearable systems for trust inference is still limited, and most approaches combine EDA with other measurements, raising questions about the true effectiveness of EDA in this context. Additionally, most approaches use informal data labeling instead of well-established measurement tools. In the subsequent section, we introduce a data collection method for skin conductance signals that relies on automatic labeling derived from a reliable ground truth source [12].

II. Methodology

The research study delved into the analysis of SCR and its correlation with the trust levels exhibited by participants. This investigation took place within the context of a chemical scenario in which a collaborative robot assisted a human operator in tasks involving the handing of a beaker with chemicals inside.

In a previous research [12], it was demonstrated that the trust levels of human operators varied significantly based on the performance of the robot. More specifically, when the robot approached the human operator too closely, participants reported lower levels of trust. Conversely, when the robot exhibited high-performance capabilities, completing tasks without malfunctions and minimizing the risk of collisions or harm from chemicals, participants displayed higher levels of trust. Drawing from these findings, the current study sought to replicate the same scenario while manipulating the robot's performance, resulting in two distinct performance modalities: *high performance* and the other *low performance*. This deliberate manipulation allowed for the creation of two experimental conditions: *high trust* and *low trust*, related to the high and low performance modalities, respectively. This innovative approach facilitated the collection of SCR data with automatic labeling (i.e. high or low trust). Within the experimental framework, every participant was involved in two separate trials for each condition, resulting in a cumulative total of four task executions. In the course of each

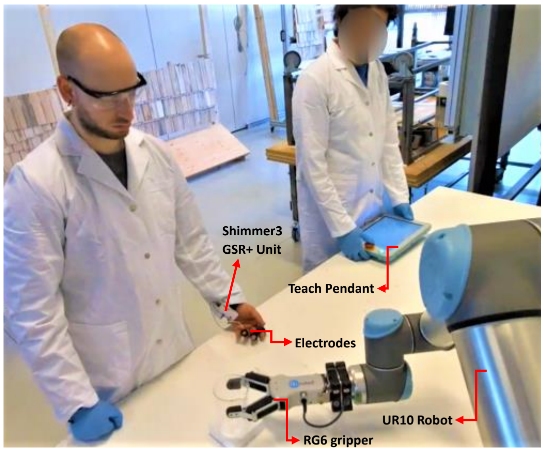


Fig. 2: The chemical industry environment and the Shimmer3 GSR+ Unit.

trial, the robot followed distinct trajectories for delivering the beaker to the user. This methodology facilitated a thorough investigation into the physiological responses linked to the diverse levels of trust exhibited by the user.

A. Experimental Setup

The main component in our experimental setup was the Universal Robots UR10 Robot[1] which was employed to assist the human in handling the beaker with the chemicals. The 6-DoF (degrees of freedom) UR10 robot is a six-axis manipulator arm with six rotational joints and it is designed to provide high reliability and precision in collaborative tasks. It was equipped with the RG6 gripper from OnRobot, a flexible 2-fingered gripper capable of providing up to 150 mm of stroke. Throughout the experiment, the path planning and grasping phases were pre-configured in advance. Nevertheless, participants were explicitly informed that the robot exhibited autonomous behavior, which introduced the possibility of malfunctions and an element of unpredictability. The assistant was provided with an emergency button to halt the robot in case of possible collisions with the human operator.

With the purpose of measuring skin conductance, the Shimmer3 GSR+ Unit[2] was utilized. The Shimmer3 GSR+ Unit is a wearable sensor equipped with advanced sensing technology, designed to capture variations in electrical conductance on the skin's surface. With reference to [23], one electrode should be placed on the palmar surface of the medial phalange (e.g. index finger) and the other on the palmar surface of the distal phalange (e.g. middle finger), as illustrated in Fig. 1. Shimmer is equipped with a 12-bit ADC, indicating that it can provide readings within the range of 0 to 4095. A calibration equation is employed to transform the ADC output into skin resistance, and subsequently, by taking the reciprocal, the skin conductance is derived. Skin conductance provides insights into shifts in an

[1]https://www.universal-robots.com/cb3/
[2]https://shimmersensing.com/product/shimmer3-gsr-unit/

individual's emotional or psychological condition. Typical skin resistance varies from 47kΩ to 1MΩ resistance (21μS to 1μS conductivity) [24]. Increased sweating results in higher skin conductance (lower skin resistance). Arousal, heightened alertness, discomfort, or sudden shock may induce individuals to perspire as their bodies prepare for potential exertion in response to a perceived threat, enabling temperature regulation through heightened sweat production.

To conclude, both participants and assistant were equipped with laboratory coats, gloves, and safety glasses for safety reasons. With reference to the chemicals utilized, the beaker held by the robot was filled with water. The true chemical composition was disclosed at the experiment's end, in line with the participants' prior awareness of the potential hazards linked to these substances. The experimental scenario is depicted in Fig. 2.

B. Procedure

The research encompassed a group of 20 participants, evenly split between 10 males and 10 females, showcasing a diverse range of ages (*M*=29.1, *SD*=7.54). The participants exhibited varying degrees of familiarity with robots, with 10 individuals possessing hands-on experience with robotic technology, while one had merely encountered robots in real-world scenarios, such as exhibitions. The remainder 9 participants had no previously experience with robots.

Following the Declaration of Helsinki and in accordance with ethical guidelines, the study was subjected to a rigorous review and received formal approval from the institutional review board, recognizing the importance of ethical standards in research involving human subjects. Prior to the commencement of the experiment, participants were provided with a detailed explanation of the study's purpose, the tasks involving interactions with the robot, the research methodology, and the potential risks. This comprehensive information was accompanied by a printed consent form, ensuring that participants were fully informed.

In addition, the assistant provided support to the participant by guiding them through the process of donning the necessary personal protective equipment. Lastly, the Shimmer3 GSR+ Unit was securely positioned beneath the non-dominant hand, which was not involved in the experiment due to the nature of the task. This placement was achieved using adjustable straps, guaranteeing comfortable and unobtrusive data collection. When using skin-contact electrodes with potential movement at their location, motion artifacts can affect the signals, typically appearing as high-frequency noise-like components in recordings. To mitigate this artifact, the electrodes were pressed tightly onto the skin to provide better contact and consequently higher conductance values. The subsequent sections will detail a cleaning phase designed to address and overcome this issue.

Each participant spent a total of 20 minutes, encompassing all the experiment stages, including the initial explanation. During this time, the participant performed the task four times (3 minutes for each trial), experiencing the robot's dual performance modes: high and low performance.

TABLE I: Extracted features for SCR.

Feature	Description
Mean	Average value of the amplitude of SCR.
Median	Center value of SCR.
Stand. Dev.	Dispersion of the SCR relative to its mean value.
Minimum	Lowest value of SCR.
Maximum	Highest value of SCR.

C. Data Collection

The controlled data collection was timed to capture the physiological responses of the participants for the handing task. The acquisition of raw skin conductance data was performed at a sampling rate of 60 Hz.

D. Data Pre-Processing

To begin with, the skin conductance raw data underwent a **cleaning phase**. A *Butterworth low-pass filter* with a *3 Hz cutoff frequency* was utilized on the data to eliminate high-frequency noise originating from factors such as motion artifacts and other sources of interference [25]. The filter permits all frequencies within the passband to pass with uniform gain, leading to minimal distortion. Skin conductance is divided in two components: *phasic* and *tonic* component. According to several studies found in the literature [26], [27], a commonly adopted approach involves distinguishing between the slowly-evolving *tonic* or baseline response, also referred to as Skin Conductance Level (SCL), which is primarily influenced by factors such as skin condition and temperature, and the rapidly-varying *phasic* response (SCR), which encompasses reactions to both specific and non-specific stimuli. The greater the variations in the phasic component of the signal, the more likely it is that the subject is in an increased state of arousal. Therefore, for the subsequent analysis, the **SCR component** was extracted using a *Butterworth high-pass filter* with a *cutoff frequency* of *0.05 Hz* [28].

In summary, the dataset for the handing task comprised 20275 samples. Every data sample received an automatic trust label, determined by the experimental condition in each participant's trial. A low-trust label was assigned when the robot's performance was poor, while a high-trust label was assigned when the robot operated with high-level performance. *Exploratory Data Analysis* was employed to inspect the dataset and identify outliers within the dataset. To eliminate outliers, the *Z-score method* was applied with a threshold set at 3. As a result, the Handing dataset was reduced to 19400 samples (a 4.32% decrease). Afterwards, the **extraction of the features** was conducted. With reference to [25], for each trial of the participant, the following features were extracted to analyse the physiological arousal derived from the interaction: *mean*, *median*, *standard deviation*, *minimum*, *maximum*. An overview of the features is presented in Table I. Following the feature extraction process, the dataset underwent *standardization* and *label encoding* processes. To conclude, the last phase focused on **feature**

TABLE II: The optimal hyperparameters of each model determined by Grid Search Cross-Validation with 5-fold cross-validation approach.

Model	Hyperparameter	Value
Random Forest		
	max_depth	10
	min_samples_leaf	4
	min_samples_split	10
	n_estimators	50
XGBoost		
	max_depth	3
	n_estimators	50
	learning_rate	0.05
LightGBM		
	max_depth	10
	n_estimators	50
	learning_rate	0.05
	num_leaves	63
	boosting_type	dart
Voting Classifier		
	voting	soft

TABLE III: Machine Learning models with related performance indicators for Handing task.

Model	Accuracy	AUC	Precision	Recall	F1-score
Random Forest	66.61%	0.67	0.66	0.67	0.66
XGBoost	68.99%	0.73	0.68	0.69	0.68
LightGBM	67.63%	0.68	0.67	0.68	0.67
Voting	67.97%	0.71	0.67	0.68	0.67

selection using a *tree-based algorithm*, as ensemble models such as Random Forest, XGBoost, and LightGBM will be employed in the machine learning analysis. *XGBoost* was utilized as the tree-based algorithm to select features due to its efficiency and robustness. The assessment of feature importance was conducted using the *gain* metric, which calculates the average performance improvement associated with each feature throughout the model's training process. A selection criterion was defined, with features considered significant if their importance score exceeded 5% of the maximum value. Furthermore, trial-and-error approach was used to fine-tune the final selection of the features. In conclusion, all features were considered as input for the machine learning models.

III. Experimental Results

A comprehensive analysis was conducted to examine the potential correlation between the trust levels exhibited by human operators, specifically categorized as either *high trust* or *low trust*, and their corresponding physiological responses, expressed as SCR values. These values serve as indicators of the operators' emotional states.

Machine learning algorithms, implemented on the Tensorflow platform, were applied to examine both the handing and pouring tasks. Specifically, the chosen algorithms encompassed **Random Forest**, **XGBoost**, and **LightGBM**. The selection relied upon the features of each algorithm. Random Forest enhances accuracy by constructing multiple decision trees and aggregating their predictions, effectively mitigating overfitting. XGBoost excels in predictive precision due to its gradient boosting technique. LightGBM, another gradient boosting algorithm, demonstrates high speed and efficiency, achieved through its histogram-based approach, which accelerates computations while maintaining accuracy. In conclusion, the **Voting Classifier** was employed to harness the advantages of combining the multiple classifiers for the final prediction of the class, thereby reducing classification errors and overfitting.

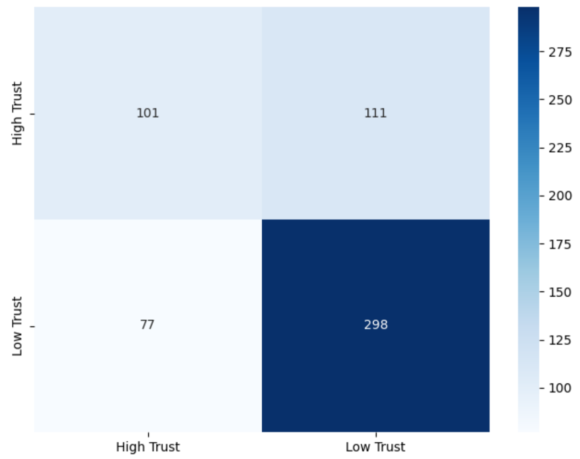
Before examining the **model's hyperparameters** tuning, it is important to highlight that the training set comprised 70% of the data, encompassing 14 participants, while the remaining 30%, involving 6 participants, constituted the test set. The utilization of this partitioning criterion guaranteed that the model underwent evaluation using entirely unseen data. In the following, the model's hyperparameters were tuned using *Grid Search Cross-Validation* with *5-fold cross-validation* approach.

In the **Random Forest** analysis, the optimization of the hyperparameters governing individual decision trees within the ensemble was carried out. These parameters included maximum depth, minimum samples in leaf nodes, minimum samples for node splitting, and the number of decision trees. For the Handing dataset, the best configuration was determined as *max depth* 10, *min samples leaf* 4, *min samples split* 10, and *n estimators* 50, resulting in an accuracy rate of 66.61% and 0.67 AUC score .

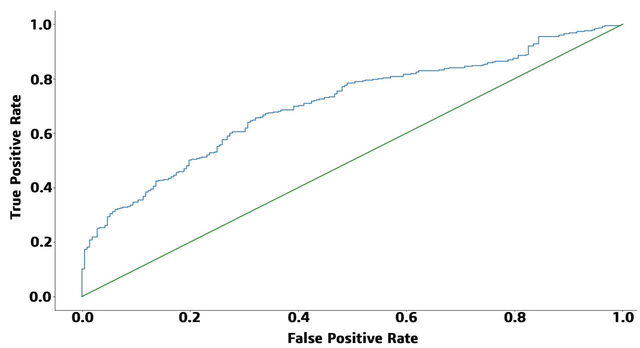
Considering **XGBoost**, the hyperparameters included maximum depth, number of decision trees and learning rate. The optimal hyperparameter configurations turned out to be *learning rate* 0.05, *max depth* of 3, and *n estimators* 50. Consequently, these parameter settings translated into final accuracy rates of 68.99% and an AUC score of 0.73.

With reference to **LightGBM**, the selected hyperparameters comprised boosting type, learning rate, maximum depth, number of decision trees, and the number of leaves in the decision tree. The optimal hyperparameters emerged as follows: 'dart' for *boosting type*, a *learning rate* of 0.05, *max depth* 10, *n estimators* 50, and *num leaves* 63. As a result, the achieved accuracy was 67.63% and AUC score was 0.68.

To conclude, as previously mentioned, the **Voting Classifier** was employed to mitigate overfitting and enhance the robustness of the analysis. In this case, the hyperparameter to fine-tune was the voting technique, as the Voting Classifier utilizes two distinct approaches: *hard* and *soft* voting. Under hard voting, the ultimate prediction hinges on the class that receives the most votes. In contrast, soft voting takes into account class probabilities, computing the average probability for each class. The final prediction is then based on the class with the highest average probability. The base classifiers, namely Random Forest, XGBoost, and LightGBM, were configured using the previously determined optimal hyperparameters. The *soft* voting method emerged as the optimal choice, achieving an accuracy rate of 67.97% and



(a) Confusion Matrix.



(b) AUC-ROC curve.

Fig. 3: Handing task and related analysis results using Voting Classifier.

0.71 as AUC score.

A comprehensive summary of the best hyperparameters of each model is reported in Table II. Table III presents the performance indicators results for the several algorithms.

IV. Discussion

In this analysis, the primary objective was to investigate the correlation between physiological responses, particularly SCR, and the trust levels exhibited by human operators. Given that trust plays a pivotal role in determining individuals' comfort and their inclination to engage in collaborative efforts with robots, SCR was examined for objectively measuring emotional reactions and the associated physiological arousal connected to trust. The three machine learning algorithms used for the analysis of SCR data and trust will be discussed, summarizing the key findings and performance metrics.

The study incorporated a variety of ensemble machine learning models, specifically *Random Forest*, *XGBoost*, and *LightGBM*. These algorithms yielded diverse outcomes when applied to the handing task. Table III presents the machine learning models and their associated performance indicators, including *classification accuracy* and *Area Under the Curve (AUC)*. XGBoost showcased superior performance by achieving a moderate classification accuracy of 68.99% and AUC score of 0.73. To mitigate the slight overfitting, the *Voting Classifier* was employed to enhance result robustness. This algorithm delivered an accuracy of 67.97% and 0.71 as AUC. The confusion matrix and Receiver Operating Characteristic (ROC) curve of Voting Classifier are reported in Fig. 3a and Fig. 3b, respectively.

In light of the results, the experimental scenario was able to elicit apprehension from the participants in relation to their proximity to the robot. This increased sensitivity to physical closeness played a role in the fluctuations observed in trust levels, which corresponded to variations in the SCR values. Furthermore, participant's physiological state could have been affected by the nature of the task, specifically, the transportation of a beaker containing hazardous chemicals. Nonetheless, there are various factors to consider for enhancing the experimental setup with the goal of improving the accuracy in classifying trust levels. To begin with, if the robot were to incorporate more dynamic or potentially hazardous movements during the handling task, it might increase the probability of eliciting a broader range of physiological responses from the participants. Moreover, it is important to consider the psychological factors at play during the interaction. Participants may have had preconceived notions or biases about robots, influencing their trust and physiological responses. Lastly, external factors, including the presence of researchers, the controlled laboratory environment, or the novelty of the robotic interaction, may have exerted an influence on participant responses, potentially obscuring any underlying correlations. To address this limitation, employing a more realistic scenario, such as an actual workplace environment within a company, may prove more effective in examining trust. In such settings, participants may experience a greater degree of real-world risk, providing a more authentic context for trust assessment. To conclude, the inclusion of additional sensors could improve the reliability and robustness of the machine learning framework to categorize trust accurately. Incorporating multi-modal data provides a comprehensive perspective by considering both physiological and behavioral cues (e.g., body posture, facial expressions), thereby refining the model's capacity to discern subtle variations in trust dynamics. Consequently, strategically deploying diverse sensors emerges as a promising approach to advance both the accuracy and comprehensiveness of trust classification within the experimental setup.

V. Conclusion

In this investigation, physiological responses were examined as an indicator of trust level towards the robot.

Specifically, SCR was selected for this purpose for its potential to reflect emotional engagement or discomfort, both integral aspects of trust. The scenario consisted of a chemical industry environment with a cobot working alongside with the human operator. The cobot's primary task was to hand a laboratory beaker with a chemical to the human operator. A machine learning analysis was subsequently carried out to explore potential correlations between user trust levels and the data derived from SCR measurements. The XGBoost algorithm demonstrated a moderate level of accuracy of 68.99% and AUC score 0.73.

The results highlight the need for sensor fusion to deliver a more reliable and unbiased evaluation of trustworthiness. Improved precision in classifying trust levels is essential for tailoring the robot's behavior to match the human's current trust level, thereby enhancing safety and optimizing the distribution of workload in the environment. With further refinement of the methodology and sensors, the use of physiological signals provide an effective approach for modeling human-robot trust. In the long term, the authors aim to develop a model capable of conducting a more detailed analysis of trust fluctuations. This endeavor seeks to establish a dynamic framework that can better capture and understand the nuanced changes in trust over time.

ACKNOWLEDGMENT

The authors want to extend their thanks to Mahed Dadgostar, a dedicated student from Aalborg University, for his commitment and assistance in this research endeavor.

REFERENCES

[1] 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.

[2] S. Matsumoto, A. Washburn, and L. D. Riek, "A framework to explore proximate human-robot coordination," *J. Hum.-Robot Interact.*, vol. 11, no. 3, 2022. [Online]. Available: https://doi.org/10.1145/3526101

[3] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa, "Planning with trust for human-robot collaboration," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: Association for Computing Machinery, 2018, p. 307–315. [Online]. Available: https://doi.org/10.1145/3171221.3171264

[4] A. Kramberger, A. Gams, B. Nemec, C. Schou, D. Chrysostomou, O. Madsen, and A. Ude, "Transfer of contact skills to new environmental conditions," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 668–675.

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

[7] 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.

[8] N. Wang, D. V. Pynadath, and S. G. Hill, "Trust calibration within a human-robot team: Comparing automatically generated explanations," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 109–116.

[9] A. S. Ghazali, J. Ham, E. I. Barakova, and P. Markopoulos, "Effects of robot facial characteristics and gender in persuasive human-robot interaction," *Frontiers in Robotics and AI*, vol. 5, p. 73, 2018.

[10] K. Hald, M. Rehmn, and T. B. Moeslund, "Human-robot trust assessment using motion tracking and galvanic skin response," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 6282–6287.

[11] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco, "Effects of changing reliability on trust of robot systems," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 73–80.

[12] G. Campagna and M. Rehm, "Analysis of proximity and risk for trust evaluation in human-robot collaboration," in *32nd IEEE International Conference on Robot and Human Interactive Communication*. IEEE, 2023.

[13] D. R. Billings, K. E. Schaefer, J. Y. Chen, and P. A. Hancock, "Human-robot interaction: developing trust in robots," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 109–110.

[14] M. Rueben, S. A. Elprama, D. Chrysostomou, and A. Jacobs, "Introduction to (re) using questionnaires in human-robot interaction research," *Human-Robot Interaction: Evaluation Methods and Their Standardization*, pp. 125–144, 2020.

[15] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, "Galvanic skin response (gsr) as an index of cognitive load," in *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 2651–2656. [Online]. Available: https://doi.org/10.1145/1240866.1241057

[16] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.

[17] H. Hamdi, P. Richard, A. Suteau, and P. Allain, "Emotion assessment for affective computing based on physiological responses," in *2012 IEEE International Conference on Fuzzy Systems*, 2012, pp. 1–8.

[18] R. B. Knapp, J. Kim, and E. André, "Physiological signals and their use in augmenting emotion recognition for human–machine interaction," in *Emotion-oriented systems: The Humaine handbook*. Springer, 2010, pp. 133–159.

[19] A. Khawaji, J. Zhou, F. Chen, and N. Marcus, "Using galvanic skin response (gsr) to measure trust and cognitive load in the text-chat environment," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1989–1994. [Online]. Available: https://doi.org/10.1145/2702613.2732766

[20] M. Lochner, A. Duenser, and S. Sarker, "Trust and cognitive load in semi-automated uav operation," in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*. New York, NY, USA: Association for Computing Machinery, 2020, p. 437–441. [Online]. Available: https://doi.org/10.1145/3369457.3369509

[21] K. Akash, W.-L. Hu, N. Jain, and T. Reid, "A classification model for sensing human trust in machines using eeg and gsr," *ACM Trans. Interact. Intell. Syst.*, vol. 8, no. 4, 2018. [Online]. Available: https://doi.org/10.1145/3132743

[22] J. Ayoub, L. Avetisian, X. J. Yang, and F. Zhou, "Real-time trust prediction in conditionally automated driving using physiological measures," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2023.

[23] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[24] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of psychophysiology*. Cambridge university press, 2007.

[25] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, and I. Molenaar, "Detecting emotions through electrodermal activity in learning contexts: A systematic review," *Sensors*, vol. 21, no. 23, p. 7869, 2021.

[26] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of neuroscience methods*, vol. 190, no. 1, pp. 80–91, 2010.

[27] D. R. Bach, G. Flandin, K. J. Friston, and R. J. Dolan, "Time-series analysis for rapid event-related skin conductance responses," *Journal of neuroscience methods*, vol. 184, no. 2, pp. 224–234, 2009.

[28] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE transactions on biomedical engineering*, vol. 63, no. 4, pp. 797–804, 2015.