A Classification Model for Sensing Human Trust in Machines Using EEG and GSR

KUMAR AKASH, WAN-LIN HU, NEERA JAIN, and TAHIRA REID, Purdue University, USA

Today, intelligent machines *interact and collaborate* with humans in a way that demands a greater level of trust between human and machine. A first step toward building intelligent machines that are capable of building and maintaining trust with humans is the design of a sensor that will enable machines to estimate human trust level in real time. In this article, two approaches for developing classifier-based empirical trust-sensor models are presented that specifically use electroencephalography and galvanic skin response measurements. Human subject data collected from 45 participants is used for feature extraction, feature selection, classifier training, and model validation. The first approach considers a general set of psychophysiological features across all participants as the input variables and trains a classifier-based model for each participant, resulting in a trust-sensor model based on the general feature set (i.e., a "general trust-sensor model"). The second approach considers a customized feature set for each individual and trains a classifier-based model using that feature set, resulting in improved mean accuracy but at the expense of an increase in training time. This work represents the first use of real-time psychophysiological measurements for the development of a human trust sensor. Implications of the work, in the context of trust management algorithm design for intelligent machines, are also discussed.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**; Empirical studies in HCI; • **Computing methodologies** → *Supervised learning by classification*; *Feature selection*;

Additional Key Words and Phrases: Trust in automation, human-machine interaction, intelligent system, classifiers, modeling, EEG, GSR, psychophysiological measurement

ACM Reference format:
Kumar Akash, Wan-Lin Hu, Neera Jain, and Tahira Reid. 2018. A Classification Model for Sensing Human Trust in Machines Using EEG and GSR. *ACM Trans. Interact. Intell. Syst.* 8, 4, Article 27 (November 2018), 20 pages.
https://doi.org/10.1145/3132743

1 INTRODUCTION

Intelligent machines and, more broadly, intelligent systems are becoming increasingly common in the everyday lives of humans. Nonetheless, despite significant advancements in automation, human supervision and intervention are still essential in almost all sectors, ranging from manufacturing and transportation to disaster-management and healthcare [43]. Therefore, we expect

This material is based upon work supported by the National Science Foundation under Award No. 1548616. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Authors' addresses: K. Akash, W.-L. Hu, N. Jain, and T. Reid, School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, Indiana 47907; emails: kakash@purdue.edu, wanlinhu@stanford.edu, {neerajain, tahira}@purdue.edu.

2160-6455/2018/11-ART27 $15.00
https://doi.org/10.1145/3132743

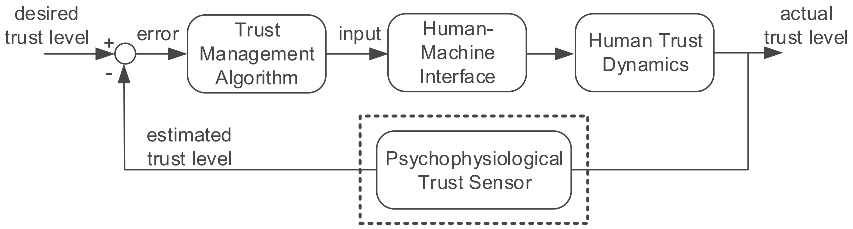


Fig. 1. A block diagram of a feedback control system for achieving trust management during human-machine interactions. The scope of this work includes psychophysiological trust-sensor modeling.

that the future will be built around *Human-Agent Collectives* [17] that will require efficient and successful coordination and collaboration between humans and machines.

It is well established that human *trust* is central to successful interactions between humans and machines [24, 32, 41]. In the context of autonomous systems, human trust can be classified into three categories: dispositional, situational, and learned [12]. Dispositional trust refers to the component of trust that is dependent on demographics such as gender and culture, whereas situational and learned trust depend on a given situation (e.g., task difficulty) and past experience (e.g., machine reliability), respectively. While all of these trust factors influence the way humans make decisions while interacting with intelligent machines, situational and learned trust factors "can change within the course of a single interaction" [12]. Therefore, we are interested in using feedback control principles to design machines that are capable of *responding to changes in human trust level in real time* to build and manage trust in the human-machine relationship as shown in Figure 1. However, to do this, we require a sensor for *estimating human trust level*, again in real time.

Researchers have attempted to predict human trust using dynamic models that rely on the experience and/or self-reported behavior of humans [18, 23]. However, it is not practical to retrieve human self-reported behavior continuously for use in a feedback control algorithm. An alternative is the use of psychophysiological signals to estimate trust level [39]. While these measurements have been correlated to human trust level [7, 27], they have not been studied in the context of real-time trust sensing.

In this article, we present a human trust-sensor model based upon real-time psychophysiological measurements, primarily galvanic skin response (GSR) and electroencephalography (EEG). The model is based upon data collected through a human subject study and the use of classification algorithms to estimate human trust level using psychophysiological data. The proposed methodology for real-time sensing of human trust level will enable the development of a machine algorithm aimed at improving interactions between humans and machines.

This article is organized as follows. In Section 2, we introduce related work in human-machine interaction, psychophysiological measurements, and their applications in trust sensing. We then describe the experimental study and data acquisition in Section 3. The data pre-processing technique for noise removal is presented in Section 4 along with EEG and GSR feature extraction. In Section 5, we demonstrate a two-step feature selection process to obtain a concise and optimal feature set. The selected features are then used for training Quadratic Discriminant Analysis classifiers in Section 6, followed by model validation and, finally, concluding statements.

2 BACKGROUND AND RELATED WORK

There are few psychophysiological measurements that have been studied in the context of human trust. We focus here on electroencephalography (EEG) and galvanic skin response (GSR), which are both noninvasive and whose measurements can be collected and processed in real time. EEG is an electrophysiological measurement technique that captures the cortical activity of the brain [10]. These brain activities exhibit changes in human thoughts, actions, and emotions. Brain-Computer Interface (BCI) technology utilizes EEG to design interfaces that enable a computer or an electronic device to understand a human's commands [34, 35]. The most extensive approach used to identify EEG patterns in BCI design includes feature selection and classification algorithms as they typically provide good accuracy [31].

Some researchers have studied trust via EEG measurements, but only with event-related potentials (ERPs). ERPs measure brain activity in response to a specific event. An ERP is determined by averaging repeated EEG responses over many trials to eliminate random brain activity [10]. Boudreau et al. found a difference in peak amplitudes of ERP components in human subjects while they participated in a coin toss experiment that stimulated trust and distrust [7]. Long et al. further studied ERP waveforms with feedback stimuli based on a modified form of the coin toss experiment [27]. The decision-making in the "trust game" [29] has been used to examine human-human trust level. Although ERPs can show how the brain functionally responds to a stimulus, they are event-triggered. It is difficult to identify triggers during the course of an actual human-machine interaction, thereby rendering ERPs impractical for real-time trust-level sensing.

GSR is a classical psychophysiological signal that captures arousal based upon the conductivity of the surface of the skin. It is not under conscious control but is instead modulated by the sympathetic nervous system. GSR has also been used in measuring stress, anxiety, and cognitive load [15, 33]. Researchers have examined GSR in correlation with human trust level. Khawaji et al. found that average GSR values, and average GSR peak values, are significantly affected by both trust and cognitive load in the text-chat environment [19]. However, the use of GSR for *estimating* trust has not been explored and was noted as an area worth studying [39]. With respect to both GSR and EEG, a fundamental gap remains in determining a static model that not only estimates human trust level using these psychophysiological signals but that is also suitable for real-time implementation.

3 METHODS AND PROCEDURES

In this section, we describe a human subject study that we conducted to identify psychophysiological features that are significantly correlated to human trust in intelligent systems and to build a trust-sensor model accordingly. The experiment consisted of a simple HMI context that could elicit human trust dynamics in a simulated autonomous system. Our study used a within-subjects design wherein both behavioral and psychophysiological data were collected and analyzed. We then used the data to build an empirical model of human trust through a process involving feature extraction, feature selection, and model training, which are described in Sections 4, 5, and 6, respectively. Figure 2 summarizes the modeling framework.

3.1 Participants

Participants were recruited using fliers and email lists. All participants were compensated at a rate of $15/h. The sample included 48 adults between 18 and 46 years of age (mean: 25.0 years old, standard deviation: 6.9 years old) from West Lafayette, Indiana (USA). Of the 48 adults, 16 were females and 32 were males. All participants were healthy and one was left-handed. The group of participants were diverse with respect to their age, professional field, and cultural background (i.e., nationality). The Institutional Review Board at Purdue University approved the study.

3.2 EEG and GSR Recording

EEG. The participant's brain waves were measured using a B-Alert X-10 9-channel EEG device (Advance Brain Monitoring, CA, USA), at a frequency of 256Hz from nine scalp sites (Fz, F3, F4,

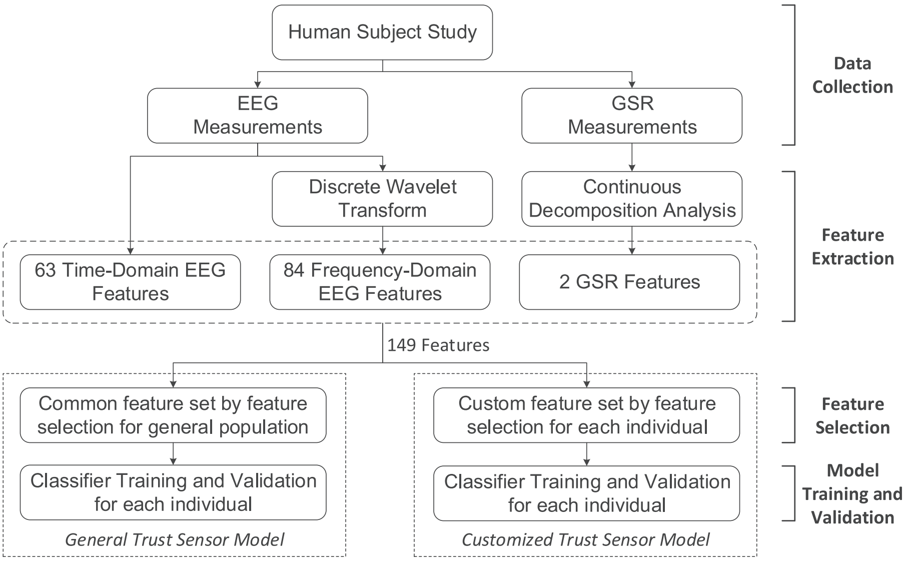

Human Subject Study
EEG Measurements
GSR Measurements
Data Collection
Discrete Wavelet Transform
Continuous Decomposition Analysis
63 Time-Domain EEG Features
84 Frequency-Domain EEG Features
2 GSR Features
Feature Extraction
149 Features
Common feature set by feature selection for general population
Custom feature set by feature selection for each individual
Feature Selection
Classifier Training and Validation for each individual
Classifier Training and Validation for each individual
Model Training and Validation
General Trust Sensor Model
Customized Trust Sensor Model

Fig. 2. The framework of the proposed study. The key steps include data collection from human subject studies, feature extraction, feature selection, model training, and model validation.

Cz, C3, C4, POz, P3, and P4, based on the 10–20 system). All EEG channels were referenced to the mean of the left and right mastoids. The surface of all sensor sites was cleaned with 70% isopropyl alcohol. Conductive electrode cream (Kustomer Kinetics, CA, USA) was then applied to each electrode including the reference. The contact impedance between electrodes and skin was kept to a value less than 40kΩ. The EEG signal was recorded via iMotions (iMotions, Inc., MA, USA) on a Windows 7 platform with Bluetooth connection.

GSR. The skin conductance was measured from the proximal phalanges of the index and the middle fingers of the non-dominant hand (i.e., on the left hand for 43 of 44 participants) at a frequency of 52Hz via the Shimmer3 GSR+ Unit (Shimmer, MA, USA). Locations for attaching Ag/AgCl electrodes (Lafayette Instrument, IN, USA) were prepared with 70% isopropyl alcohol. The participants were asked to keep their hands steady on the desk to minimize the influence of movement on the measured signals. The environment temperature was controlled at 72–74°F to minimize the effect of temperature. The GSR signal was also recorded via iMotions so that it would be synchronized with the recorded EEG signals using the common system-timestamps between these two signals.

3.3 Experimental Procedure

After the participants read and signed the informed consent, they were equipped with the EEG headset and the GSR sensor as shown in Figure 3. All participants finished a 9min EEG baseline task provided by Advanced Brain Monitoring and were then instructed to interact with our custom-designed computer-based simulation. Participants were told that they would be driving a car equipped with an image–based obstacle detection sensor. The sensor would detect obstacles on the road in front of the car, and the participant would need to repeatedly evaluate the algorithm report and choose to either trust or distrust the report based on their experience with the algorithm. Detailed instructions were delivered on the screen following four practice trials. Participants could have their questions answered while instructions were given and during the practice session.

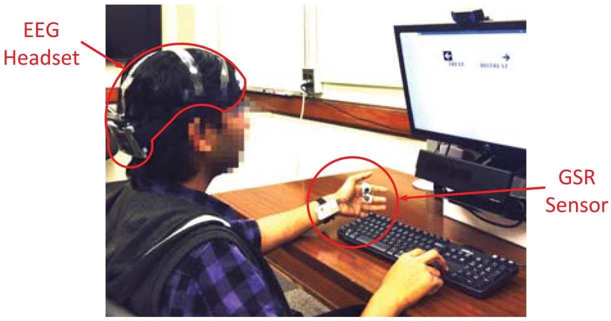


Fig. 3. Experimental setup with participant wearing EEG Headset and GSR Sensor.

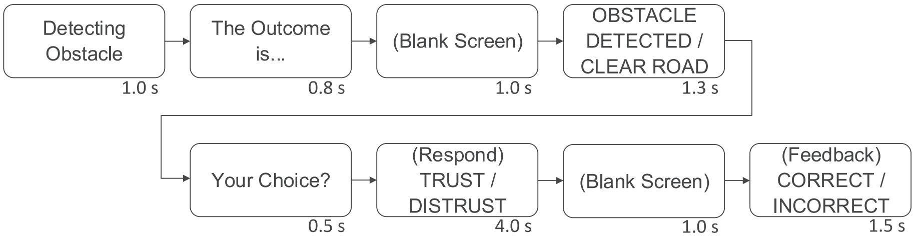
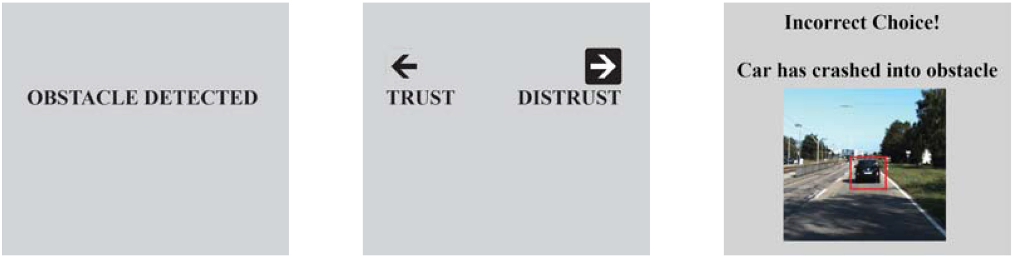


Fig. 4. Sequence of events in a single trial. The time length marked on the bottom right corner of each event indicates the time interval for which the information appeared on the computer screen.



OBSTACLE DETECTED

TRUST DISTRUST

Incorrect Choice!

Car has crashed into obstacle

(a) Stimuli (b) Response (c) Feedback

Fig. 5. Example screenshots of the interface of the experimental study. These screens correspond to three of the events shown in Figure 4: obstacle detected/clear road, trust/distrust, and correct/incorrect, respectively.

Each trial consisted of a stimulus (i.e., report on sensor functionality), the participant's response, and feedback to the participants on the correctness of their response. There were two stimuli, "obstacle detected" and "clear road," and both had a 50% probability of occurrence. Participants had the option to choose "trust" or "distrust" in response to the sensor report after which they received the feedback of "correct" or "incorrect." Figure 4 shows the sequence of events in a single trial, and Figure 5 shows example screenshots of the computer interface.

The independent variable was the participants' experience due to the sensor performance, and the dependent variable was their trust level. The sensor performance was varied to elicit the dynamic response in each participant's trust level. There were two categories of trials: *reliable* and *faulty*. In reliable trials, the sensor accurately identified the road condition with 100% probability; in faulty trials, there was only a 50% probability that the sensor correctly identified the road condition with sensor faults presented in a randomized order. We implemented the 50% accuracy for

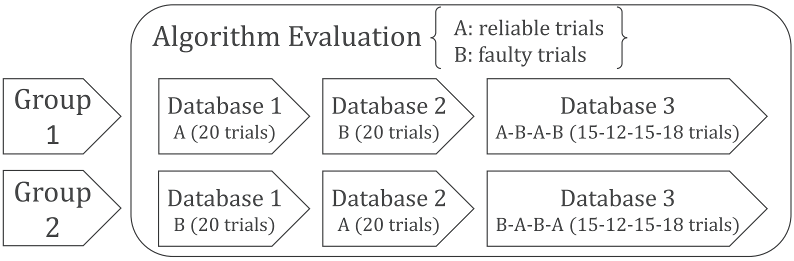


Fig. 6. Participants were randomly assigned to one of two groups. The ordering of the three experimental sections (databases), composed of reliable and faulty trials, were counterbalanced across Groups 1 and 2.

faulty trials, because pilot studies indicated that it would be perceived as a pure random chance by the participants. This should conceivably result in the lowest possible trust level that a human has in the simulated sensor. The participants received “correct” as feedback when they indicated trust in reliable trials, but there was a 50% probability that they received “incorrect” as feedback when they indicated trust in faulty trials.

Each participant completed 100 trials. The trials were divided into three phases, called “databases” in the study, as shown in Figure 6. Participants were randomly assigned to one of two groups for counterbalancing any possible ordering effects. Databases 1 and 2 consisted of either reliable (A) or faulty (B) trials (see details in Figure 6). The number of trials in each of these two databases was chosen so that the trust or distrust response of each human subject would approach a steady-state value [27]. Steady-state ensures that the trust level truly reaches the desired state (i.e., trust for reliable trials and distrust for faulty trials), which is essential for labeling the trials as trust or distrust. However, the accuracy of the algorithm was switched between reliable and faulty according to a pseudo-random binary sequence (PRBS) in Database 3. This was done to excite all possible dynamics of the participant’s trust response required for dynamic behavior modeling, which was the subject of related work by the authors [1]. Therefore, only the data from databases 1 and 2 (i.e., the first 40 trials) were analyzed.

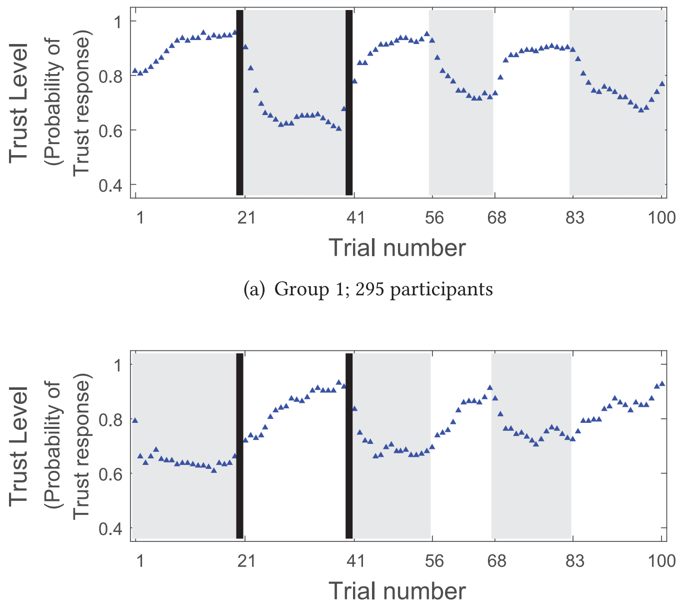
We collected psychophysiological measurements to identify any latent indicators of trust and distrust. In general, latent emotions are those that cannot be easily articulated. Latent distrust may inhibit the interactions between human and intelligent systems despite reported trust behaviors. We hypothesized that the trust level would be high in reliable trials and be low in faulty trials, and we validated this hypothesis using responses collected from 581 online participants (58 were outliers) via Amazon Mechanical Turk [2]. The experiment elicited expected trust responses based on the aggregated data as shown in Figure 7 [1]. Therefore, data from reliable trials were labeled as trust, and data from faulty trials were labeled as distrust. The data analysis and feature extraction methodologies will be discussed further in Section 4.

4 DATA ANALYSIS

In this section, we discuss the methods used to pre-process the data (collected during the human subject studies) to reduce noise and remove contaminated data. We then describe the process of feature extraction applied to the processed data.

4.1 Pre-Processing

We used the automatic decontaminated signals provided by the B-Alert EEG system for artifact removal. This decontamination process minimizes the effects of electromyography, electrooculography, spikes, saturation, and excursions. Before further processing the data, we manually examined



(a) Group 1; 295 participants

(b) Group 2; 228 participants

Fig. 7. The averaged response from online participants collected via Amazon Mechanical Turk. Faulty trials are highlighted in gray. Participants showed a high trust level in reliable trials and a low trust level in faulty trials regardless of the group they were in.

the spectral distribution of EEG data for each participant. We removed the participants having anomalous EEG spectra, possibly due to bad channels or dislocation of EEG electrodes during the study. This process resulted in 45 participants to analyze. Finally, EEG measurements from channel F3 and F4 were excluded from the data analysis due to contamination with eye movement and blinking [5]. For GSR measurements, we used adaptive Gaussian smoothing with a window of size 8 to reduce noise [6].

4.2 Feature Extraction

To estimate trust in real time, we require the ability to continuously extract and evaluate key psychophysiological measurements. This could be achieved by continuously considering short segments of signals for calculations. Levy suggests using short epoch lengths for identifying rapid changes in EEG patterns [25]. Therefore, we divided the entire duration of the study into multiple 1s epochs (periods) with 50% overlap between each consecutive epoch. Assuming that the decisive cognitive activity occurs when the participant sees the stimuli, we only considered the epochs lying completely between each successive stimulus (obstacle detected/clear road) and response (trust/distrust). Consequently, approximately 129 epochs were considered for each participant. We labeled each of these epochs as one of two classes, namely, *Distrust* or *Trust*, based on whether the epoch belonged to faulty or reliable trials, respectively. The number of epochs varied depending on the response time of the human subject for each trial.

EEG. Existing studies have shown the importance of both time-domain features and frequency-domain features for successfully classifying cognitive tasks [28]. To utilize the benefits of both, we extracted an exhaustive set of time- and frequency-domain features from EEG.

We extracted six time-domain features from all seven channels (Fz, C3, Cz, C4, P3, POz, and P4) for each epoch of length N. For this study in which EEG signals were sampled at 256Hz, each 1s epoch had a length of $N = 256$. Letting $k \in (1, n)$, where n is the total number of epochs and x_k represents the kth epoch of channel ch_x. These features were defined as

(1) mean $\mu_k(ch_x)$, where

$$\mu_k(ch_x) = \frac{1}{N}\sum_{i=1}^{N} x_{ki}, \tag{1}$$

(2) variance $\sigma_k^2(ch_x)$, where

$$\sigma_k^2(ch_x) = \frac{1}{N-1}\sum_{i=1}^{N} |x_{ki} - \mu_k|^2, \tag{2}$$

(3) peak-to-peak value $pp_k(ch_x)$, where

$$pp_k(ch_x) = \max_{1 \leq i \leq N} x_{ki} - \min_{1 \leq i \leq N} x_{ki}, \tag{3}$$

(4) mean frequency $\bar{f}_k(ch_x)$, defined as the estimate of the mean frequency from the power spectrum of x_k,

(5) root mean square value $rms_k(ch_x)$, where

$$rms_k(ch_x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |x_{ki}|^2}, \tag{4}$$

and

(6) signal energy $E_k(ch_x)$, where

$$E_k(ch_x) = \sum_{i=1}^{N} |x_{ki}|^2. \tag{5}$$

Therefore, we extracted 42 (6 features × 7 channels) time-domain features for each epoch. Moreover, the interaction between the different regions of the brain was also considered by calculating the correlation between pairs of channels for each epoch. The correlation coefficient between two channels (e.g., ch_x and ch_y) of the kth epoch $\rho_k(ch_x, ch_y)$ is defined as

$$\rho_k(ch_x, ch_y) = \frac{cov(x_k, y_k))}{\sqrt{var(x_k)var(y_k)}}, \tag{6}$$

where x_k and y_k are the kth epochs of channels ch_x and ch_y, respectively. The expressions $cov(.)$ and $var(.)$ are the covariance and variance functions, respectively. Therefore, 21 additional time-domain features were extracted (combinations of 2 out of 7 channels, C_2^7).

Next, we extracted features from four frequency bands across all seven channels for each epoch. Classically, EEG brain waves have been categorized into four bands based on frequency, namely, delta (0.5–4Hz), theta (4–8Hz), alpha (8–13Hz), and beta (13–30Hz). However, because of the non-stationary characteristics of EEG signals (i.e., their statistics vary in time), analyzing the variations in frequency components of EEG signal with time (i.e., time-frequency analysis) is more informative than analyzing the frequency content of the entire signal at a time. The Discrete Wavelet Transform (DWT) is an extensively used tool for time-frequency analysis of physiological signals, including EEG [3]. Therefore, we used DWT decomposition to extract the frequency-domain features from the EEG signals.

Table 1. Wavelet Decompositions and Their Frequency Range

Level	Wavelet coefficient	Frequency range	Classical band
3	D3	16–32Hz	Beta
4	D4	8–16Hz	Alpha
5	D5	4–8Hz	Theta
5	A5	0–4Hz	Delta

DWT uses scale-varying basis functions to achieve good time resolution of high frequencies and good frequency resolution for low frequencies. The DWT decomposition consists of successive high-pass and low-pass filtering of the signal with downsampling by a factor of 2 in each successive level [42]. The high-pass filter uses a discrete mother wavelet function, and the low-pass filter uses its mirror version. We used the mother wavelet function of the Daubechies wavelet (db5) for frequency decomposition of the EEG signal. The first low-pass and high-pass filter outputs are called approximation A1 and detailed coefficients D1, respectively. A1 is further decomposed, and the steps are repeated to achieve the desired level of decomposition. Since the highest frequency in our signal was 128Hz (sampling frequency $f_s = 256$Hz), each channels' signal was decomposed to the fifth level to achieve the decomposition corresponding to the classical bands as shown in Table 1.

Three features, namely, mean (Equation (1)), variance (Equation (2)), and energy (Equation (5)), were calculated from each of the four decomposed band decomposition coefficients shown in Table 1 for each channel's epoch. Therefore, 84 frequency-domain features were extracted (3 features × 4 bands × 7 channels).

GSR. GSR is a superposition of the tonic (slow-changing) and the phasic (fast-changing) components of the skin conductance response [4]. We used Continuous Decomposition Analysis from Ledalab to separate the tonic and phasic components of the signal [4]. Since the timescale of the study and the decision making tasks are, in general, much faster as compared to the tonic component, we only used the phasic component of the GSR. We calculated the *Maximum Phasic Component* and the *Net Phasic Component* for each epoch, thus extracting two features from GSR.

5 FEATURE SELECTION

Following the feature extraction described in Section 4, we next describe the process of feature selection. The selected features were considered to be potential input variables for the trust-sensor model, of which the output would be the *probability of trust response*. We define the probability of trust response as the probability of the human trusting the intelligent system at the next time instant. In this section, we discuss feature selection algorithms used for selecting optimal feature sets for two variations of our trust-sensor model, followed by a discussion of the significance of the features in each of the final feature sets.

5.1 Feature Selection Algorithms

The complete feature set consisted of 149 features (42 + 21 + 84 + 2) that were extracted for each epoch for every participant. These features were considered potential variables for predicting the *Trust* or *Distrust* classes. Out of this large feature set, it was necessary to downselect a smaller subset of features as predictors to avoid "the curse of dimensionality" (also called "Hughes phenomenon"), which occurs for high-dimensional feature spaces with a limited number of samples. Not doing feature selection leads to a reduction in the predictive power of learning algorithms [28].

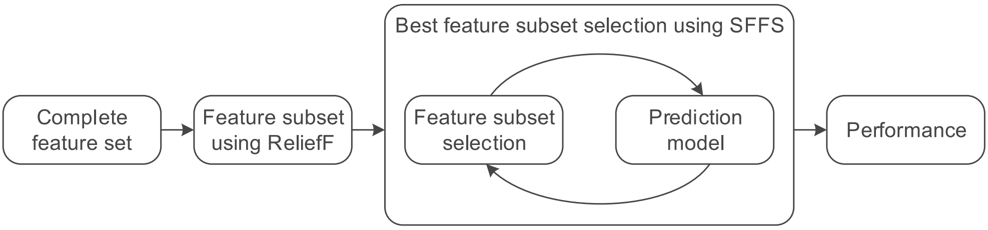


Fig. 8. A schematic depicting the feature selection approach used for reducing the dimension of the feature set. The ReliefF (filter method) was used for an initial shortlisting of the feature subset followed by SFFS (wrapper method) for the final feature subset selection.

Therefore, feature selection was achieved by removing irrelevant and redundant features from the feature set according to feature selection algorithms.

Feature selection algorithms are categorized into two groups: filter methods and wrapper methods. Filter methods depend on general data characteristics such as inter-class distance, results of significance tests, and mutual information, to select the feature subsets without involving any selected prediction model. Since filter methods do not involve any assumptions of a prediction model, they are useful in estimating the relationships between the features. Wrapper methods use the performance (e.g., accuracy) of a selected prediction model to evaluate possible feature subsets. When the performance of a particular type of model is of importance, wrapper methods result in a better fit for a selected model type; however, they are typically much slower than filter methods [21]. We used a combination of filter and wrapper methods for feature selection to manage the trade-off between training speed and model performance. We used a filter method called *ReliefF* for initially shortlisting features followed by a wrapper method called *Sequential Forward Floating Selection (SFFS)* for the final feature selection as shown in Figure 8.

5.1.1 *ReliefF.*

The basic idea of ReliefF is to estimate the quality of the features based on their ability to distinguish between samples that are near each other. Kononenko et al. proposed a number of improvements to existing work by Kira and Rendell and developed ReliefF [20, 22]. For a data set with n samples, the algorithm iterates n times for each feature. For our study, there were approximately 129 samples corresponding to each epoch as mentioned in Section 4.2. At each iteration for a two-class problem, the algorithm selects one of the samples and finds k nearest hits (same-class sample) and k nearest misses (different-class sample), where k is a parameter to be selected. Kononenko et al. suggested that k could be safely set to 10 for most purposes. We used $k = 10$ and calculated the ReliefF weights for all extracted features of each individual participant. The weight of any given feature is penalized for far-off near-hits and improved for far-off near-misses. Far-off near misses implies well-separated features, and far-off near-hits implies intermixed classes.

5.1.2 *Sequential Forward Floating Selection (SFFS).*

The SFFS is an enhancement of the Sequential Feature Selection algorithm for addressing the “nesting effect” [36]. The nesting effect means that a selected feature cannot be discarded when the forward method is implemented and the discarded feature cannot be re-selected when the backward method is implemented. To avoid this effect, SFFS builds the feature set with the best predictive power by continuously adding a dynamically changing number of features at each step to the existing subset of features. This operation occurs iteratively until no further increase in performance is observed. In this study, we defined the performance as the misclassification rate of the Quadratic Discriminant Analysis (QDA) classifier. We have examined that a QDA classifier achieved the highest accuracy for another data set

Table 2. Features to be Used as Input Variables for the General Trust-sensor Model

	Feature	Measurement	Domain
1	Mean Frequency—Fz	EEG	Time
2	Mean Frequency—C3	EEG	Time
3	Mean Frequency—C4	EEG	Time
4	Peak-to-peak—C3	EEG	Time
5	Energy of Theta Band—P3	EEG	Frequency
6	Variance of Alpha Band—P4	EEG	Frequency
7	Energy of Beta Band—C4	EEG	Frequency
8	Energy of Beta Band—P3	EEG	Frequency
9	Mean of Beta Band—C3	EEG	Frequency
10	Correlation—C3 & C4	EEG	Time
11	Correlation—Cz & C4	EEG	Time
12	Net Phasic Component	GSR	Time

based on the same experimental setup [13], and its output posterior probability is also suitable for interpreting trust. Therefore, we used the QDA classifier and calculated the misclassification rate using fivefold cross validation [11]. This validation technique randomly divides the data into five sets and predicts each set using a model trained for the remaining four sets.

5.2 Feature Selection for the Trust-sensor Model

The differences between humans could introduce differences in their trust behavior. This leads to two approaches for selecting features for sensing trust level: (1) to select a common set of features for a general population, which results in a *general trust-sensor model*; and (2) to select a different set of features for each individual, which results in *customized trust-sensor model* for each individual.

5.2.1 Feature Selection for the General Trust-sensor Model. A general trust-sensor model is desirable so that it can be used to reflect trust behavior in a general adult population. This model correlates significant psychophysiological features with human trust in intelligent systems based on data obtained from a broad range of adult human subjects. Since a general trust-sensor model requires a common list of features for all participants, we randomly divided the participants into two groups: the training-sample participants (33 out of 45 participants), which were used to identify the common list of features, and the validation-sample participants (12 out of 45 participants), which were used to validate the selected list of features. We calculated the median of the ReliefF weights across the training-sample participants for all features. The median was used instead of mean to avoid outliers [26]. Finally, we shortlisted features with the top 60 median weights and used SFFS for selecting the final set of features. For each training-sample participant's data, a separate classifier was trained and the average value of the misclassification rate for all training-sample participants was used as the predictive power for feature subsets for SFFS. We obtained a feature set with 12 features consisting of both time- and frequency-domain features of EEG along with net phasic components of GSR. Table 2 shows the final list of selected features for the general trust-sensor model using training-sample participants.

5.2.2 Feature Selection for the Customized Trust-sensor Model. We followed a similar approach to that used for feature selection in Section 5.2.1, but the list of features was selected individually for each of the 45 participants. We used ReliefF weights and shortlisted a separate set of features

Table 3. The Most Common Features that are Significant for at Least Four Participants

	Feature	Measurement	Domain
1	Mean Frequency—POz	EEG	Time
2	Mean Frequency—C4*	EEG	Time
3	Mean Frequency—P3	EEG	Time
4	Mean Frequency—Fz*	EEG	Time
5	Mean Frequency—C3*	EEG	Time
6	Peak-to-peak—C3*	EEG	Time
7	Variance of Beta Band—P3	EEG	Frequency
8	Mean of Beta Band—P3	EEG	Frequency
9	Correlation—Cz & C4*	EEG	Time
10	Net Phasic Component*	GSR	Time
11	Maximum Value of Phasic Activity	GSR	Time

Features marked with an asterisk (*) are also significant for the general trust-sensor model.

for each participant consisting of the top 60 weights. Then, for each participant, SFFS was used with the misclassification rate as determined by the quadratic discriminant classifier to select a final set of features from the shortlisted feature set. We obtained a relatively smaller feature set for each individual participant, with an average of 4.33 features in each participant's feature set, as compared to 12 features when all of the participants' data was aggregated into a single data set. Table 3 shows each of the features that are significant for at least four of the participants. We observed that there is great diversity in the significant features for each individual who supports the usage of a customized trust-sensor model. However, it is important to note that even within this diversity, more than half of the most common features (e.g., mean frequency at C4) are also significant for the general trust-sensor model.

5.3 Discussion on Significant Features in Trust Sensing

Several time-domain EEG features were found to be significant, especially the mean frequency of the EEG power distribution and the correlations between the signals from the central regions of the brain (C3, C4, Cz). Time-domain EEG features have been discovered to be significant in brain activities [28]. Moreover, our observation that activities at sites C3 and C4 play an important role in trust behaviors is supported by existing studies that have suggested that central regions of the brain are related to processes associated with problem complexity [16], anxiety in a sustained attention task [40], and mental workload [9].

Among the frequency domain EEG features, the measurements from the left parietal lobe, particularly in a high frequency range (i.e., the beta band), responded most strongly to the discrepancy between reliable and faulty stimuli. This is consistent with the finding that cognitive task demands have a significant interaction with hemisphere in the beta band for parietal areas [37]. The beta band is also an important feature that has been shown to be related to emotional states in the literature [14] and may represent the emotional component of human trust.

Finally, the results also showed that the phasic component of GSR was a significant predictor of trust levels for the general trust-sensor model as well as for several customized trust-sensor models. This aligns with the existing literature that shows that the GSR features could significantly improve the classification accuracy for mental workload detection [8] and could index difficulty levels of decision making [44]. The importance of phasic GSR to trust sensing was also supported by Khawaji's study in which the average of peak GSR values was affected by interpersonal trust [19].

6 MODEL TRAINING AND VALIDATION

The selected features discussed in Section 5 were considered as input variables for each of the trust-sensor models; the output variables were the categorical trust level, namely, the classes "Trust" and "Distrust." In this section, we introduce the training procedure of a quadratic discriminant classifier that was used to predict the categorical trust class using the psychophysiological features. We then present and discuss the results of the model validation.

6.1 Classifier Training

The quadratic discriminant classifier was implemented using the Statistics and Machine Learning Toolbox in MATLAB R2016a (The MathWorks, Inc., USA). The low training and prediction time of quadratic discriminant classifiers is advantageous for real-time implementation of the classifier [30]. Moreover, the posterior probability calculated by the classifier for the class "Trust" was used as the probability of trust response, thus resulting in a continuous output. The continuous output of probability of trust response would be particularly beneficial for implementation of a feedback control algorithm for managing human trust level in an intelligent system. To avoid large and sudden fluctuations in the trust level, the continuous output was smoothed using a median filter with a window of size 15. The general trust-sensor model and customized trust-sensor models were developed with the same training procedure but with different feature sets (i.e., input variables). The former was based on the common feature set, and the latter was based on customized feature sets, as described in Sections 5.2.1 and 5.2.2.

6.2 Model Validation Techniques

We used fivefold cross-validation to evaluate the performance of classifiers. The data, consisting of approximately 129 samples for each participant, was randomly divided into five sets. Each set was predicted using a model trained from the other four datasets. We used these predictions to evaluate the accuracy of the binary classification. Accuracy is defined as the proportion of correct predictions among the total number of samples and is given as

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total population}}. \tag{7}$$

Moreover, prediction performance of a classifier may be better evaluated by examining the confusion matrix shown in Figure 9. We calculated two statistical measures called sensitivity (true positive ratio) and specificity (true negative ratio) that are defined as follows.

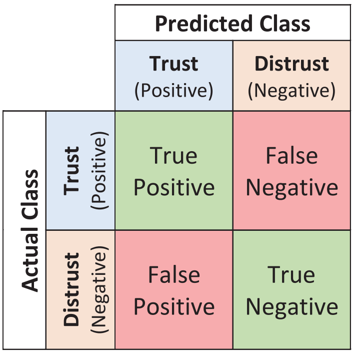
(1) Sensitivity: the proportion of actual trust (positives) that are correctly predicted as such, where

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}. \tag{8}$$

(2) Specificity: the proportion of actual distrust (negatives) that were correctly predicted as such, where

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}}. \tag{9}$$

To examine the robustness of the classifier to the variation in training data, we performed 10,000 iterations with a different random division of the five sets in each iteration and calculated the performance measures for each iteration. Tables 4 and 5 show the mean, maximum (Max), minimum



		Predicted Class	
		Trust (Positive)	Distrust (Negative)
Actual Class	Trust (Positive)	True Positive	False Negative
	Distrust (Negative)	False Positive	True Negative

Fig. 9. The actual class and the predicted class form a 2 × 2 confusion matrix. The outcomes are defined as true or false positive/negative.

Table 4. The Accuracy, Sensitivity, and Specificity (%) of the *General* Trust-sensor Model for Training-sample Participants with a 95% Confidence Interval

	Accuracy	Sensitivity	Specificity
Mean	70.52 ± 0.007	64.17 ± 0.010	75.49 ± 0.009
Max	93.72 ± 0.013	96.75 ± 0.020	96.38 ± 0.015
Min	54.67 ± 0.042	31.18 ± 0.040	44.92 ± 0.039
SD	11.29 ± 0.006	18.96 ± 0.009	14.35 ± 0.008

Table 5. The Accuracy, Sensitivity, and Specificity (%) of the *General* Trust-sensor Model for Validation-sample Participants with a 95% Confidence Interval

	Accuracy	Sensitivity	Specificity
Mean	73.13 ± 0.010	65.35 ± 0.015	79.49 ± 0.013
Max	99.89 ± 0.006	99.92 ± 0.006	99.85 ± 0.011
Min	59.29 ± 0.035	34.35 ± 0.081	57.04 ± 0.050
SD	10.91 ± 0.007	17.03 ± 0.016	12.26 ± 0.015

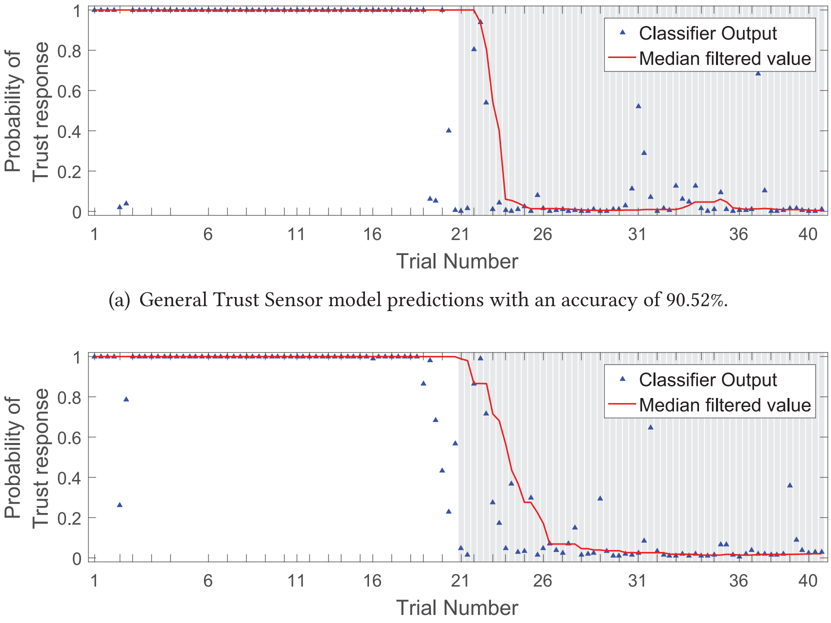
(Min), and standard deviation (SD) values for each of the performance measures for the *general trust-sensor model*. This is shown for both training-sample participants (Table 4) and validation-sample participants (Table 5) along with the 95% confidence interval (CI) obtained using the iterations. Table 6 shows the performance statistics of the *customized trust-sensor model* for all participants. The confidence intervals obtained for both models were very narrow, *indicating that models were robust to the selection of training data.*

6.3 Discussion on Performance of Classification Models

The mean accuracy was 70.52±0.007% for training-sample participants. Similarly, the mean accuracy for the *validation-sample* participants was 73.13±0.010%. The fact that the performance of the general trust model was consistent for both training-sample and validation-sample participants suggests that the identified list of features could estimate trust for a broad population of individuals. Moreover, the mean accuracy was 78.58±0.0005% for the customized trust-sensor models for all participants. Recall that the customized trust senor models were based on a customized

Table 6. The Accuracy, Sensitivity, and Specificity (%) of the *Customized* Trust-sensor Model for All Participants with a 95% Confidence Interval

	Accuracy	Sensitivity	Specificity
Mean	78.55 ± 0.005	72.83 ± 0.007	82.56 ± 0.007
Max	100.00 ± 0.000	100.00 ± 0.000	100.00 ± 0.000
Min	61.59 ± 0.041	34.77 ± 0.044	45.89 ± 0.040
SD	9.69 ± 0.005	17.02 ± 0.008	11.18 ± 0.007



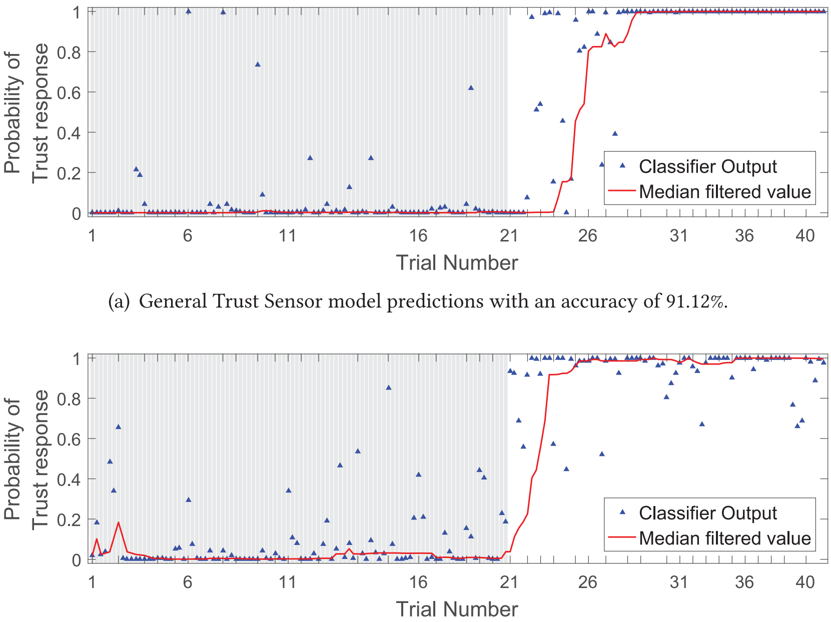
(a) General Trust Sensor model predictions with an accuracy of 90.52%.

(b) Customized Trust Sensor model predictions with an accuracy of 93.97%.

Fig. 10. Classifier predictions for participant 44 in group 1. Faulty trials are highlighted in gray. Trust-sensor models had a good accuracy for this participant. The classifier output of posterior probability was smoothed using a median filter with window of size 15.

feature set for each participant. There were 12 significant features to predict trust for the general trust-sensor models, while less than 5 features were needed for the customized trust-sensor models. These findings support the hypothesis that a customized trust-sensor model could enhance the prediction accuracy with a smaller feature set. For some individual participants, the mean accuracy increased to 100%.

Figures 10 and 11 are examples of good predictions for participants in groups 1 and 2, respectively. The customized trust-sensor models performed better for both participants, specifically at the transition state at the beginning of database 2. Figure 10(b) shows an example of a transition state at the beginning of database 2; it took five trials for this participant to establish a new trust level. The classification accuracy was low for some participants as shown in Figure 12. The classifier had difficulty correctly predicting trust (database 1), which may imply that this particular participant was not able to conclude whether or not to trust the sensor report, even in reliable trials. Another potential reason could be that trust variations of this participant did not result in



(a) General Trust Sensor model predictions with an accuracy of 91.12%.

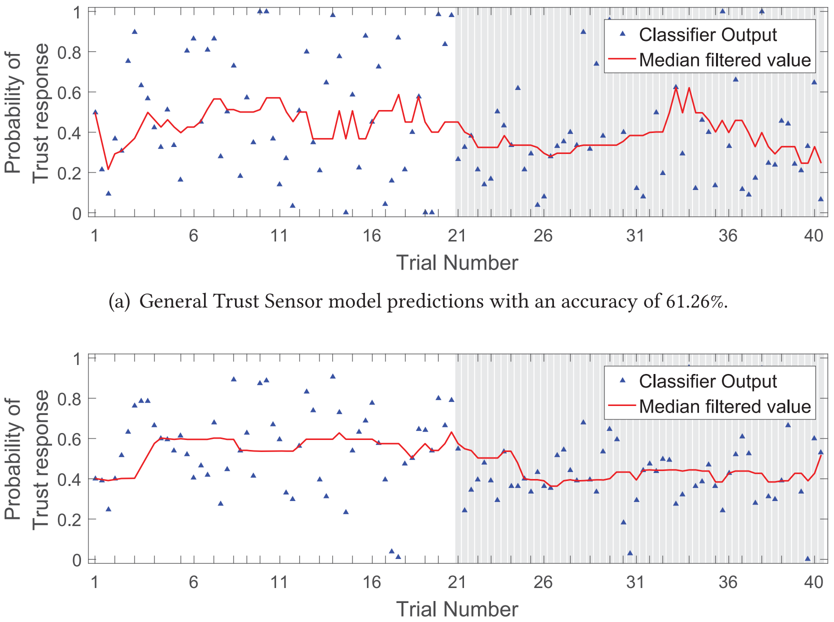
(b) Customized Trust Sensor model predictions with an accuracy of 96.45%.

Fig. 11. Classifier predictions for participant 10 in group 2. Faulty trials are highlighted in gray. Trust-sensor models had good accuracy for this participant. The classifier output of posterior probability was smoothed using a median filter with window of size 15.

significant changes in their physiological signals. Nevertheless, the customized trust-sensor model still showed a higher accuracy than the general trust-sensor model.

The general trust-sensor model resulted in mean *specificity* of 75.49±0.009% and 79.49±0.013% for training-sample and validation-sample participants, respectively. The customized trust-sensor model resulted in 82.56±0.007% for all participants. This indicates that the models are capable of correctly predicting distrust in humans. The models are less likely to predict a distrust response as trust (i.e., less false positives). The mean *sensitivity* was 64.17±0.010% and 65.35±0.015% for the general trust-sensor model for training-sample and validation-sample participants, respectively. The customized trust-sensor model resulted in 72.83±0.007% for all participants. Low sensitivity (more false negatives) occurs when the model often predicts trust as distrust. In the context of using this trust-sensor model to design an intelligent system that could be responsive to a human's trust level, low sensitivity would arguably not have an adverse effect, since the goal of the system would be to enhance trust.

There is a fundamental trade-off that exists between the general and customized models in terms of the time spent on model training and model performance as shown in Table 7. The results show that the selected feature set (Table 2) for the general trust-sensor models is applicable for a general adult population with a 71.22% mean accuracy (i.e., the mean accuracy calculated across all participants). Furthermore, by applying this common feature set, feature selection is not required while implementing the general model. This would reduce the model training time and potentially make the model adaptable to various scenarios. However, the common feature set for a general population is larger than feature sets optimized for each individual, because it attempts to accommodate an aggregated group of individuals. Therefore, in scenarios where the speed of the online prediction process is the priority, the customized trust-sensor model, with a smaller feature



(a) General Trust Sensor model predictions with an accuracy of 61.26%.

(b) Customized Trust Sensor model predictions with an accuracy of 72.07%.

Fig. 12. Classifier predictions for participant 8 in group 1. Faulty trials are highlighted in gray. Trust-sensor models did not have good accuracy for this participant. The classifier output of posterior probability was smoothed using a median filter with window of size 15.

Table 7. Comparison of General Trust-sensor Model and Customized Trust-sensor Model for Implementation

Model Characteristics	General Trust-sensor Model	Customized Trust-sensor Model
Required training time	Less	More
Size of final feature set	12	4.33 (Average)
Prediction Time	More	Less
Mean Prediction Accuracy	71.22%	78.55%

set, would be preferred. The customized trust-sensor model also enhances the prediction accuracy. Nonetheless, it is worth noting that implementing the customized trust-sensor model would still require extraction of a larger set of features initially for training followed by a smaller feature set extraction for real-time implementation. This would increase the time required for training the model as an additional feature selection step would need to be performed.

While we focused on situational and learned trust, dispositional trust factors, such as demographics, may have partially contributed to the observed lower accuracy of the general trust-sensor model due to individual differences in trust response behavior [1, 38]. Incorporating these additional factors and other psychophysiological signals may increase the trust estimation accuracy of the trust-sensor model, as the features included in the present model inherently represent only a subset of many non-verbal signals that correlate to trust level.

In summary, the proposed trust-sensor model could be used to enable intelligent systems to estimate human trust and in turn respond to, and collaborate with, humans in such a way that leads to successful and synergistic collaborations. Potential human-machine/robot collaboration contexts include robotic nurses that assist patients, aircrafts that exchange control authority with human operators, and numerous others [43].

7 CONCLUSION

As humans are increasingly required to interact with intelligent systems, trust becomes an important factor for synergistic interactions. The results presented in this article show that psychophysiological measurements can be used to estimate human trust in intelligent systems in real time. By doing so, intelligent systems will have the ability to respond to changes in human trust behavior.

We proposed two approaches for developing classifier-based empirical trust-sensor models that estimate human trust level using psychophysiological measurements. These models used human subject data collected from 45 participants. The first approach was to consider a common set of psychophysiological features as the input variables for any human and train a classifier-based model using this feature set, resulting in a general trust-sensor model with a mean accuracy of 71.22%. The second approach was to consider a customized feature set for each individual and train a classifier-based model using that feature set; this resulted in a mean accuracy of 78.55%. The primary trade-off between these two approaches was shown to be training time and performance (based on mean accuracy) of the classifier-based model. That is to say, while it is expected that using a feature set customized to a particular individual will outperform a model based upon the general feature set, the time needed for training such a model may be prohibitive in certain applications. Moreover, although the criteria used for feature selection and classifier training in this study was mean accuracy, a different criterion could be chosen to adapt to various applications. Finally, future work will involve increasing the sample size and augmenting the general trust-sensor model to account for dispositional trust factors to improve the prediction accuracy of the model. It will also be important to test the established framework in both simulated and immersive environments using, for example, driving or flight simulators and/or virtual reality, as well as in real-life settings.

REFERENCES

[1] Kumar Akash, Wan-Lin Hu, Tahira Reid, and Neera Jain. 2017. Dynamic modeling of trust in human-machine interactions. In *Proceedings of the American Control Conference*.

[2] Amazon. 2005. Amazon Mechanical Turk. Retrieved from https://www.mturk.com/.

[3] Hafeez Ullah Amin, Aamir Saeed Malik, Rana Fayyaz Ahmad, Nasreen Badruddin, Nidal Kamel, Muhammad Hussain, and Weng-Tink Chooi. 2015. Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Austral. Phys. Eng. Sci. Med.* 38, 1 (2015), 139–149.

[4] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 1 (2010), 80–91.

[5] Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven. 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78, 5 (2007), B231–B244.

[6] Herman Blinchikoff and Helen Krause. 1976. *Filtering in the Time and Frequency Domains.* Noble Publishing.

[7] Cheryl Boudreau, Mathew D. McCubbins, and Seana Coulson. 2008. Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Soc. Cogn. Affect. Neurosci.* 4, 1 (Nov. 2008), 23–34.

[8] Fang Chen, Natalie Ruiz, Eric Choi, Julien Epps, M. Asif Khawaja, Ronnie Taib, Bo Yin, and Yang Wang. 2012. Multimodal behavior and interaction as indicators of cognitive load. *ACM Trans. Interact. Intell. Syst.* 2, 4 (Dec. 2012), 1–36.

[9] Caroline Dussault, Jean-Claude Jouanin, Matthieu Philippe, and Charles-Yannick Guezennec. 2005. EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviat. Space Environ. Med.* 76, 4 (2005).

[10] Todd C. Handy. 2005. *Event-related Potentials: A Methods Handbook.* MIT Press.

[11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer, New York.

[12] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Fact.: J. Hum. Fact. Ergonom. Soc.* 57, 3 (2015), 407–434.

[13] Wan-Lin Hu, Kumar Akash, Neera Jain, and Tahira Reid. 2016. Real-time sensing of trust in human-machine interactions. In *Proceedings of the 1st IFAC Conference on Cyber-Physical & Human-Systems.*

[14] Toshiaki Isotani, Hideaki Tanaka, Dietrich Lehmann, Roberto D. Pascual-Marqui, Kieko Kochi, Naomi Saito, Takami Yagyu, Toshihiko Kinoshita, and Kyohei Sasada. 2001. Source localization of EEG activity during hypnotically induced anxiety and relaxation. *Int. J. Psychophysiol.* 41, 2 (2001), 143–153.

[15] Sue C. Jacobs, Richard Friedman, John D. Parker, Geoffrey H. Tofler, Alfredo H. Jimenez, James E. Muller, Herbert Benson, and Peter H. Stone. 1994. Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. *Amer. Heart J.* 128, 6 (1994), 1170–1177.

[16] Norbert Jaušovec and Ksenija Jaušovec. 2000. EEG activity during the performance of complex mental problems. *Int. J. Psychophysiol.* 36, 1 (2000), 73–88.

[17] Nicholas R. Jennings, Luc Moreau, David Nicholson, Sarvapali Ramchurn, Stephen Roberts, Tom Rodden, and Alex Rogers. 2014. Human-agent collectives. *Commun. ACM* 57, 12 (Nov. 2014), 80–88.

[18] Catholijn M. Jonker and Jan Treur. 1999. *Formal Analysis of Models for the Dynamics of Trust Based on Experiences.* Springer, Berlin, 221–231.

[19] Ahmad Khawaji, Jianlong Zhou, Fang Chen, and Nadine Marcus. 2015. 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.* ACM Press, 1989–1994.

[20] Kenji Kira and Larry A. Rendell. 1992. A practical approach to feature selection. In *Proceedings of the 9th International Workshop on Machine Learning.* 249–256.

[21] Ron Kohavi and George H. John. 1997. Wrappers for feature subset selection. *Artific. Intell.* 97, 1 (1997), 273–324.

[22] Igor Kononenko, Edvard Šimec, and Marko Robnik-Šikonja. 1997. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Appl. Intell.* 7, 1 (1997), 39–55.

[23] John Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 10 (1992), 1243–1270.

[24] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Hum. Fact.: J. Hum. Fact. Ergonom. Soc.* 46, 1 (2004), 50–80.

[25] W. J. Levy. 1987. Effect of epoch length on power spectrum analysis of the EEG. *Anesthesiology* 66, 4 (Apr. 1987), 489–495.

[26] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 4 (2013), 764–766.

[27] Yun Long, Xiaoming Jiang, and Xiaolin Zhou. 2012. To believe or not to believe: Trust choice modulates brain responses in outcome evaluation. *Neuroscience* 200 (2012), 50–58.

[28] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, 2 (2007), R1.

[29] Qingguo Ma, Liang Meng, and Qiang Shen. 2015. You have my word: Reciprocity expectation modulates feedback-related negativity in the trust game. *PLoS One* 10, 2 (Feb. 2015), 1–10.

[30] Mathworks. 2016. Statistics and Machine Learning Toolbox: User's Guide. Retrieved from https://www.mathworks.com/help/pdf_doc/stats/stats.pdf.

[31] Dennis J. McFarland, Charles W. Anderson, K. Muller, Alois Schlogl, and Dean J. Krusienski. 2006. BCI meeting 2005-workshop on BCI signal processing: Feature extraction and translation. *IEEE Trans. Neural Syst. Rehab. Eng.* 14, 2 (2006), 135.

[32] Bonnie M. Muir. 1987. Trust between humans and machines, and the design of decision aids. *Int. J. Man-Mach. Studies* 27, 5–6 (1987), 527–539.

[33] Reiner Nikula. 1991. Psychological correlates of nonspecific skin conductance responses. *Psychophysiology* 28, 1 (1991), 86–90.

[34] William D. Penny, Stephen J. Roberts, Eleanor A. Curran, and Maria J. Stokes. 2000. EEG-based communication: A pattern recognition approach. *IEEE Trans. Rehab. Eng.* 8, 2 (2000), 214–215.

[35] Gert Pfurtscheller, Doris Flotzinger, and Joachim Kalcher. 1993. Brain-computer interface—A new communication device for handicapped persons. *J. Microcomput. Appl.* 16, 3 (1993), 293–299.

[36] P. Pudil, J. Novovičová, and J. Kittler. 1994. Floating search methods in feature selection. *Pattern Recogn. Lett.* 15, 11 (1994), 1119–1125.

[37] William J. Ray and Harry W. Cole. 1985. EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* 228, 4700 (1985), 750–752.

[38] René Riedl, Marco Hubert, and Peter Kenning. 2010. Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *Manage. Info. Syst. Quart.* 34, 2 (2010), 397–428.

[39] René Riedl and Andrija Javor. 2012. The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging.*J. Neurosci. Psychol. Econ.* 5, 2 (2012), 63.

[40] Stefania Righi, Luciano Mecacci, and Maria P. Viggiano. 2009. Anxiety, cognitive self–evaluation and performance: ERP correlates. *J. Anxiety Disord.* 23, 8 (2009), 1132–1138.

[41] Thomas B. Sheridan and Raja Parasuraman. 2005. Human-automation interaction. *Rev. Hum. Fact. Ergonom.* 1, 1 (2005), 89–129.

[42] D. Sundararajan. 2016. *Discrete Wavelet Transform: A Signal Processing Approach.* Wiley.

[43] Yue Wang and Fumin Zhang. 2017. *Trends in Control and Decision-Making for Human–Robot Collaboration Systems.* Springer.

[44] Jianlong Zhou, Jinjun Sun, Fang Chen, Yang Wang, Ronnie Taib, Ahmad Khawaji, and Zhidong Li. 2015. Measurable decision making with GSR and pupillary analysis for intelligent user interface. *ACM Trans. Comput.-Hum. Interact.* 21, 6, Article 33 (Jan. 2015).

Received December 2016; revised June 2017; accepted July 2017