Characterization of Human Trust in Robot through Multimodal Physical and Physiological Biometrics in Human-Robot Partnerships

Jesse Parron, Rui Li, Weitian Wang*, *Senior Member, IEEE,* and Mengchu Zhou*, Fellow, IEEE*

***Abstract*—Trust is an attribute that many people use daily, whether consciously thinking of it or not. Although commonly designated as a firm belief in reliability, trust is more complex than many think. It is not just physical, but rather an emotion, feeling, or choice that has many layers, and can be influenced in a variety of ways. As robotics and artificial intelligence grow, humans have to deliberate whether they trust working with these technical counterparts or not. In this work, we build computational models to quantitatively characterize and analyze humans' trust in robots using multimodal physical and physiological biometric data based on the TrustBase we have created through user studies in human-robot collaborative tasks. During human-robot collaborative processes, we have collected physical and physiological attribute data of human subjects as well as the users' trust levels for each interaction. This data is used to develop a database known as TrustBase. With the data from TrustBase, computational and analytical approaches are used to investigate the correlation between robot performance factors and humans' trust levels and to characterize humans' trust in robots during human-robot collaboration. Results and their analysis suggest the effectiveness of the developed models, providing new findings to the human factors and cognitive ergonomics in human-robot interaction. Future research directions are also discussed.**

I. Introduction

With the introduction and growth of ChatGPT and other Large Language Models (LLMs), robotics and artificial intelligence are pushing the boundaries of science and technology [1, 2]. Incorporating these technological advancements, industries can automate processes, expedite production, relieve personnel from tedious and dangerous work tasks [3, 4]. Industries are not the only ones able to take advantage of these new developments. With the progress that has been made, these technologies have become widely accessible to all people. Other industries such as education, healthcare, and transportation have also benefited from these improvements. For example, some researchers have explored the use of artificial intelligence in education. AI takes a form in web-based systems, humanoid robots, chatbots, and grading tools. These tools help both teachers and students achieve newfound heights [5]. With these synthetic intelligences being added into more industries, we can already see how great an influence they have generated.

While it is advantageous to have these artificial intelligent technologies assist in workloads and tasks, their widespread implementation is still in its early stages across industries. This may lead to a disconnect for some people, as they are not used to working with automated partners such as robots. Industries for decades have mainly had human-to-human collaboration. With the implementation of robotics and AI, we now have human-robot collaboration and human-AI collaboration. This raises the question of whether humans trust working with robots and AI. To dissect this question, trust must be defined. According to the Oxford Dictionary, the number one meaning of trust is the "firm belief in the reliability, truth, or ability of someone or something; confidence or faith in a person or thing, or in an attribute of a person or thing" [6]. From a psychological perspective, this makes sense as humans naturally can get a feeling about someone or something. Whether it is verbal queues or body language, humans can instinctively tell whether persons or things are good or bad and trust them or not. According to Buchan "trust development is dynamic and fluid, changing in nature from context to context, and is influenced by a multiplicity of economic, sociological, psychological, and political factors in the cultural environment of the trust relationship" [7]. Now, what happens when a person works with a robot counterpart? A robot does not express personable features a person may express. So, how does one find out if humans will be comfortable and trust working alongside these newfound advanced technologies?

To further anatomize if humans are comfortable and trust working alongside robotics, AI, and new technologies, in this study, we utilize the data from our developed database, TrustBase [8], to delve deep into this multifaceted term known as trust in human-robot partnerships. The data in TrustBase is composed of four types of physical and physiological biometric attributes, as well as real-time trust levels collected throughout the experiments where users work with a robot to complete collaborative tasks. Methodologies used to further analyze this data are time series analysis to find anomalies or outliers, as well as cross-correlation to incorporate feature analysis. This selects the most prominent features to be used for human trust modeling. Then the Support Vector Machine (SVM), TabPFN, and XGBoost are utilized to perform classification and prediction of human trust in human-robot collaboration. This will allow us to unveil if these physical and physiological attributes pair with the development of trust between a user and an inorganic counterpart.

II. TrustBase

A database known as TrustBase, developed in our previous work [8], is utilized to perform deep data analysis on the physical and physiological biometrics of participants and their trust levels for the robot. TrustBase is a multifaceted database management system (DBMS) with three layers: Trust-User, Trust-Logical, and Trust-Physical. TrustBase currently includes data from 100 user studies involving human-robot collaboration. The Trust-User layer is accompanied by a webpage that allows users to acquire information pertaining to the data stored on each user's experiment experience. The Trust-Logical layer is responsible for the logical structure and data management of the database. This layer maintains

J. Parron, R. Li, and W. Wang are with the School of Computing, Montclair State University, Montclair, NJ 07043 USA. (corresponding author: wangw@montclair.edu)
M. Zhou is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology Newark, NJ 07102, USA.

all the recorded data from users. The data obtained contains subject demographics, four datasets of physiological attributes, and the level of trust a user has with the robot through the experiment. The demographics of the subjects are compliance, gender, age, and education. The four datasets of physical and physiological features collected from each subject are electromyography, electroencephalography, electrocardiography, and ocular senses. Each dataset holds sub-attributes that contain specific data about the physiological feature. Users could utilize a 7-point Likert scale to record their trust rating. The concluding tier of the database is the Trust-Physical layer [8], tasked with the storage and retrieval of data. Aside from demographics and trust levels, stored as strings and integers, physiological data is preserved as character large objects (CLOBs) to accommodate the substantial volume of information recorded from a user.

III. Data Preprocessing

Utilizing the data of TrustBase, we can uncover patterns, errors, and potential correlations between physical features and trust. TrustBase consists of thousands of data points, some of which may have potential errors. For quality models and results, the data in the TrustBase first should be examined and cleaned for any uncertainties, errors, and outliers. Data cleaning and preprocessing methods used in this work are deletion, aggregation, and Synthetic Minority Oversampling Technique (SMOTE) [9].

An additional challenge during the experiments with subjects is the timing issue. When starting an experiment, all sensors are started sequentially. Although they all perform close in relation to time, there are time differences. Both the beginning and end of data series are to be trimmed to accurately reflect true values experienced through the experiment itself. To ensure the data being deleted is proper, each video must be analyzed to determine the exact time the experiment starts, and the exact time it ends. The data series is then trimmed to fit the duration of the experiment.

To analyze the extensive data stored in TrustBase and uncover patterns as well as modeling results later discussed in this study, the data must undergo multiple preprocessing and aggregation methods. The first aggregation method we realize to handle raw data is grouping. To accurately group the data to fit each series of human-robot interactions, time intervals are created. Since each experiment has the same amount of time, and the robot's performance factors are the same among experiments, static time intervals are employed. Utilizing these time intervals, the physical and physiological biometric data collected are grouped for each of the 27 interactions with the robot. This allows for easy implementation of data aggregation. The aggregation technique utilized for each group is taking the mean. Once the mean is calculated for each group, the data is then inserted into a data frame where each human subject has 44 columns of physical attributes and 27 rows of data, paired with the corresponding trust rating for the experiment.

Once all subjects' data are retrieved from TrustBase, grouped, aggregated, and fit into the same data frame, all data are further processed before trust characterization. The first form of preprocessing is normalization. Due to the data being collected in different scales and measurement units, normalization is critical for training the data. Normalization standardizes all data by scaling them to a similar range.

The normalization method used is Min-Max scaling [10].

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}. \quad (1)$$

After the data are normalized, an oversampling technique known as SMOTE [9] is applied to the data to balance underrepresented classes. The decision to apply SMOTE to the dataset is made from the structure of the data and the large difference between higher classes such as somewhat trustworthy to very trustworthy; versus the occurrence of lower-trust classes such as very untrustworthy to somewhat untrustworthy. This allows the models to better generalize the data, avoid bias, and produce better-trained models.

The final preprocessing technique incorporated is synthetically generating a timeline for the heartbeat data. For each subject, the dataset has a timeline, except for the heartbeat data. We create a proper timeline for heartbeat data by using the start and end time of EEG datasets, and then the time intervals of the human-robot object handover process are incorporated. A complete timeline is thus created and appended to the heartbeat dataset. This is validated by viewing and analyzing a time series graph of data, to be further discussed later.

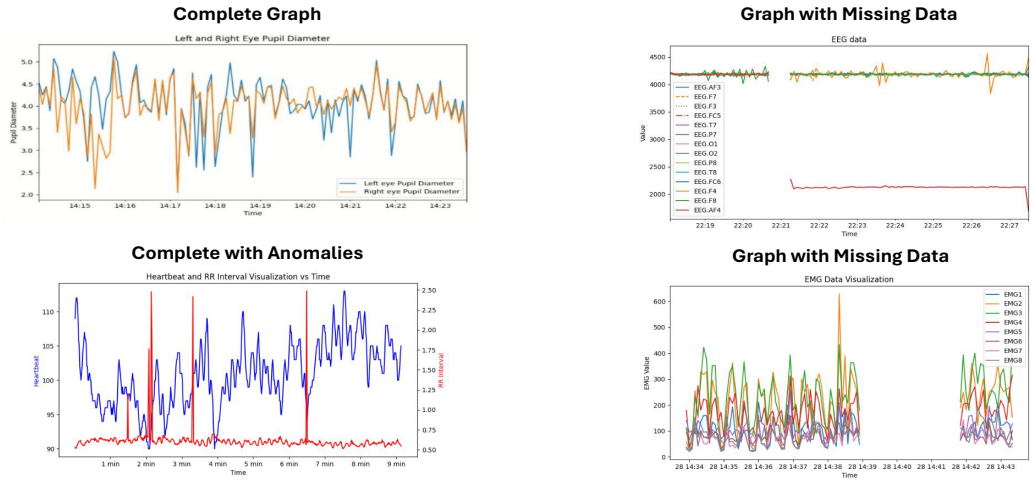

Fig. 1. Time series of features under three situations.

IV. Feature and Correlation Analysis

A. Analysis of Features in Time Series

To better understand the collected data, and visualize how each feature looks throughout the experiment, time series analysis graphs are created for every subject. Such graphs are used to visually compare the physical and physiological features with the trust ratings to see if any patterns exist. They are also used to identify patterns, anomalies, and outliers, and ensure the completeness of the data. Specifically, they are used to verify the absence of gaps and missing data in the recorded period. Fig. 1 portrays the difference between a subject with complete datasets, with potential anomalies, versus a participant who made some errors during the experiment. Each time series graph is made based on the complete period of the experiment. Each data point is a cluster of 5-second data. These clusters are grouped based on 5-second intervals and then aggregated by taking the mean. Each experiment takes approximately 9.5 minutes and each graph has about 114 data points.

B. Correlation Analysis

To further the analysis of the collected data, it is imperative to perform feature selection from each record in TrustBASE, which contains 44 features of physical and physiological data and 1 feature for trust ratings. Feature selection should be used to find the most prominent features that retain important information, while also reducing dimensionality by stripping the dataset of features that may not provide much influence when modeling. To ascertain the features to be used in training and human trust characterization, the first method is cross-correlation. According to [11], cross-correlation is a standard approach that can be utilized to evaluate the degree to which two series are correlated. Using this method, each feature is cross-correlated with one another. We create a heatmap of features that have a correlation compared to features that have no correlation. This heatmap can be seen in Fig. 2, with 44 physical and physiological attributes and 1 trust attribute employed in the correlation analysis.

In Fig. 2, the correlation heatmap's x-axis and y-axis contain each feature in the dataset which are paired with each other and outputs a correlation coefficient. This coefficient indicates the strength of the correlation between each pair of features. The closer the coefficient is to 1, the more intense the red becomes, representing a high correlation or similarity between the pair. On the other side, the less correlation between two features, a more intense blue appears. Less correlation is not necessarily bad. It does show that two features are less similar but is a key component helping one understand the negative correlation. As one feature may grow and have a stronger correlation, another feature may have less correlation and become negative, which is still very important in how each of these features affects each other during a human-robot collaboration process. Otherwise, grey areas have very little to no correlation. Overall, the graph gives a better understanding of how the data features react and pair with each other. After analyzing the features in time series and the correlation among features, we employ 38 of 44 physical and physiological attributes for later usage. The chosen physical and physiological attributes that the models are trained on are given in Table I.

TABLE I. The Chosen Physical and Physiological Attributes.

Sensors	Attributes
MYO	EMG1, EMG2, EMG3, EMG4, EMG5, EMG6, EMG7, EMG8
Emotiv	EEG.AF3, EEG.F7, EEG.F3, EEG.FC5, EEG.T7, EEG.P7, EEG.O1, EEG.O2, EEG.P8, EEG.T8, EEG.FC6, EEG.F4, EEG.F8, EEG.AF4
Vivi	Left Eye Openness, Right Eye Openness, Left Eye Pupil Diameter, Right Eye Pupil Diameter, Left Eye Normalized Gaze X, Left Eye Normalized Gaze Y, Left Eye Normalized Gaze Z, Right Eye Normalized Gaze X, Right Eye Normalized Gaze Y, Right Eye Normalized Gaze Z, Right Eye Pupil Position (Left), Right Eye Pupil Position (Right), Left Eye Pupil Position (Left), Left Eye Pupil Position (Right)
HeartStrap	Heartbeat, RR-Interval

V. Methodologies for Human Trust Characterization

A. Support Vector Machine

The first type of modeling architecture employed is Support Vector Machine (SVM). The SVM algorithm is a form of supervised learning, which can be used to solve nonlinear pattern recognition, classification, and regression problems [12]. It can handle high dimensional, highly complex data. Additionally, SVMs include a large class of neural nets, radial basis function (RBF) nets, and polynomial classifiers as special cases. They utilize a variety of kernels to perform all computations in input space. These computations lead the algorithm to find the optimal hyperplane among classes. These hyperplanes can mathematically be defined as:

$$(w * x) + b = 0 \; w \in R^N, b \in R. \tag{2}$$

Based on decision functions:

$$f(x) = sign(\textstyle\sum_i v_i \, (x * x_i) + b). \tag{3}$$

In the training of our model, we incorporate the use of GridSearchCV which is an exhaustive search to find the best combination of the input of specific parameters. In this study, we search different combinations of SVM parameters including C, Kernel, and Gamma. The best one is C=100, Gamma=10, and Kernel is RBF. C is a regularization parameter used to regulate misclassification. Gamma regulates the amount of influence a data sample applies. The RBF kernel can be mathematically defined as:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right). \tag{4}$$

With this combination of parameters, we are able to obtain a higher accuracy than other combinations. Results and analysis of the support vector machine model utilized in this study are further discussed later.

B. TabPFN

The next modeling algorithm used in this study is TabPFN that is a relatively new architecture. According to Hollmann et al. "TabPFN is a transformer that can do supervised classification for small tabular datasets in less than a second, and needs no hyperparameter or tuning" [13]. TabPFN fits our dataset well since the processed data barely reaches 2k rows of data points. TabPFN is pragmatic for our dataset, as splitting the training and test sets appropriately allows for fast training and obtaining excellent results. It is to be noted that for TabPFN, the data needs no normalization prior to training as its architecture handles all normalization. TabPFN preprocesses all inputs using a z-score normalization for each feature and log-scales outliers heuristically.

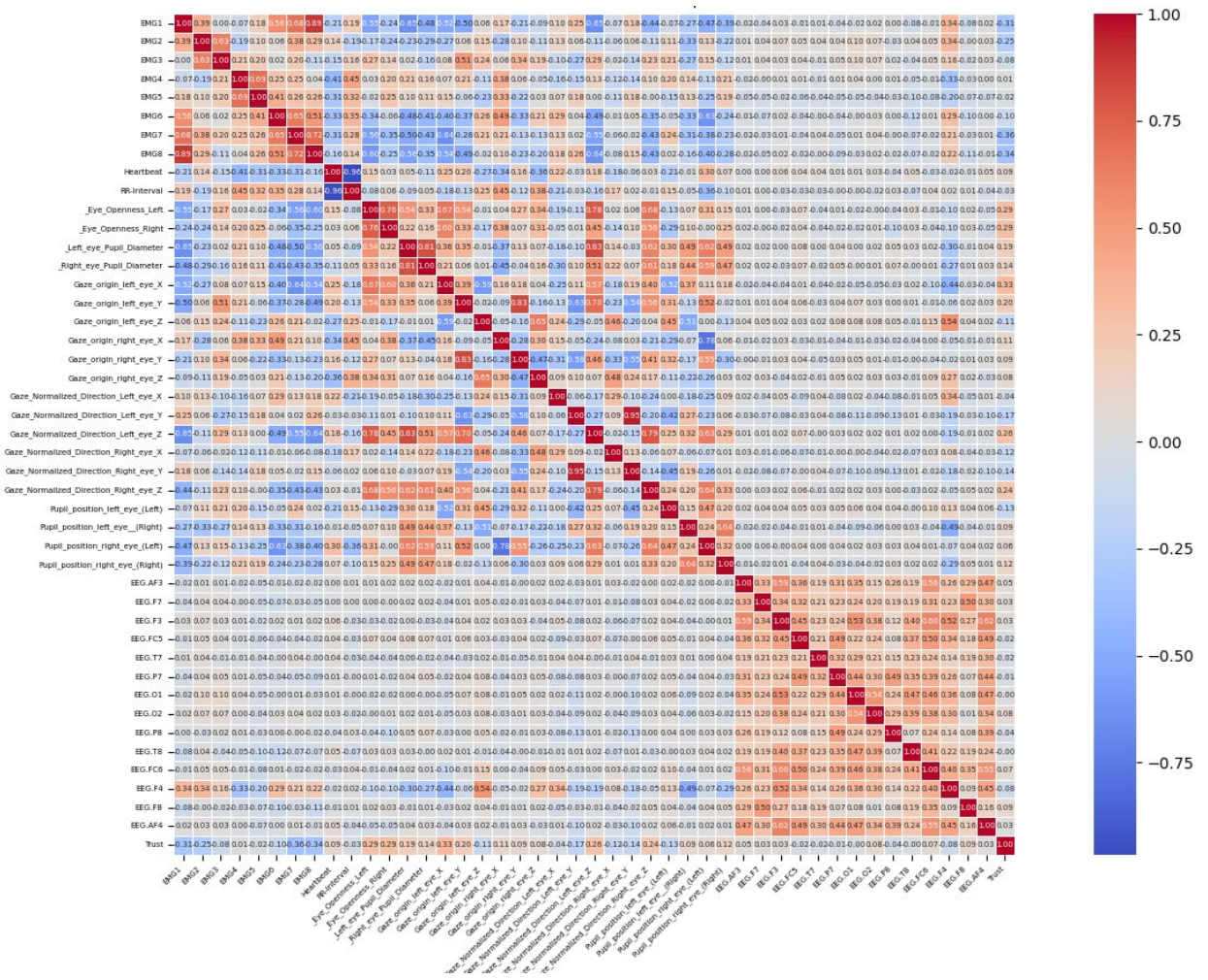

Fig. 2. Correlation analysis of all features.

Due to TabPFN being a transformer, it tokenizes the features and labels, and through a single forward pass, can return the estimate of the posterior predictive distribution [13]. The posterior predictive distribution (PPD) provides a way to make predictions for new data points. The PPD can be expressed as:

$$p(y|x, D) \propto \int_{\phi} p(y|x, \phi)p(D|\phi)p(\phi)d\phi. \quad (5)$$

In this study, we use the same PPD defined in [13] to characterize trust based on subjects' physical and physiological data. TabPFN's PPD can help predict new data points and have the model create decision boundaries. This enables us to gain a better understanding of how physical and physiological data can affect trust during human-robot interactions, to be discussed.

C. XGBoost

In this work, eXtreme Gradient Boosting (XGBoost) is used to model and characterize human trust through subjects' physical and physiological information. XGBoost is a well-known supervised machine learning model. It utilizes distributed gradient-boosted decision trees (GBDT). This approach incorporates a regularized model to prevent overfitting. Its tree ensemble model can be used to predict the output [14]:

$$\hat{y}_i = \varphi(x_i) = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F, \quad (6)$$

where is the F space of regression trees.

To learn the functions of this model, the following regulation function is minimized:

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad (7)$$

where $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ is the penalty function to penalize the complexity of the model, T represents the number of leaves in the tree, γ is a constant representing the penalty degree, w denotes leaf weight set, and λ is a constant for controlling the strength of regularization. l denotes a differentiable convex loss function used to measure the difference between the prediction $\hat{y}_i$ and the objective y_i.

In this study, we utilize XGBoost for its optimal learning rate and its ability to prevent overfitting for our dataset. It can also handle both classification and regression. Through this model, we use a tree method known as the exact greedy algorithm. We also incorporate 5 early stopping rounds to additionally prevent overfitting. The model sorts the data in order to accumulate the gradient statistic for the structure score in:

$$L_{split} = \frac{1}{2}\left[\frac{(\Sigma_{i\in I_L} g_i)^2}{\Sigma_{i\in I_L} h_i+\lambda} + \frac{(\Sigma_{i\in I_R} g_i)^2}{\Sigma_{i\in I_R} h_i+\lambda} - \frac{(\Sigma_{i\in I} g_i)^2}{\Sigma_{i\in I} h_i+\lambda}\right] - \gamma, \quad (8)$$

where L_{split} is the loss reduction after the tree split. I_L and I_R are the instance sets of left and right nodes of the tree branches after the split. g_i and h_i are the first and second order gradient statistics on the loss function. Utilizing XGBoost we are able to get new insights on how human trust is characterized, to be discussed next.

VI. Results and Analysis

A. Experimental Setup

Each participant in the study is configured with the sensors that best fit his/her body to collect the data. Prior to setup, each is asked if there are any reasons as to why he/she cannot wear a sensor. Some participants have religious needs and other reasons so they cannot wear a specific sensor, e.g. the Emotiv Epoc+ headset. If consent is given to wear all four sensors, participants are then configured with the same setting across the board. They first wear the MYO armband on their dominant arm, right below the elbow on the forearm. The polar H10 heartbeat strap is configured on their sternum, right below their chest. The heartbeat strap is tightened to fit snuggly against their chest for accurate heartbeat detection. As shown in Fig. 3, the Emotiv headset is placed on the participants' head, following a diagram on the headset application. This diagram shows where each node should be placed, and notifies when sufficient contact with the scalp has been made. The final sensor is a Vive headset, which is carefully placed over the participants' eyes, then over the Emotiv headset. Finally, users undergo a calibration procedure with the Vive headset, to ensure the eye data is accurate.

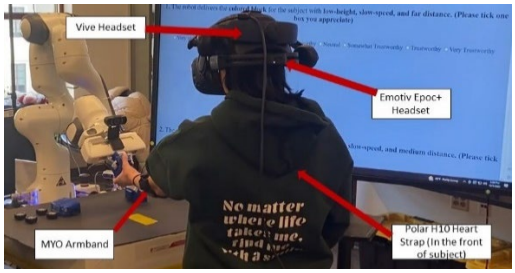

Fig. 3. Participant configuration.

B. Results and Analysis of SVM

Leveraging the robust architecture of SVM, the resulting model is able to successfully achieve high accuracy in training. Using the features mentioned in Section V.B, and the predictor data and labels associated with the trust class, the data go through multiple training attempts with different methods. Initially, we obtain low performance, as the model is underperforming when encountering classes 0-3. To improve it, we decide to remap the classes as we find that the margins of the data are too close that can cause errors during the training. We remap all classes related to low trust (classes 0-2) to a fused class known as untrustworthy (class 0), we keep neutral (class 3) as class 1, and all classes representing higher trust (classes 4-6) are remapped to trustworthy (class 2). The resampling effectively solves this issue, as it neutralizes those margins to a point such that the model can better understand. Comparatively, we find that using SVM with the remapped classes and the GridSearchCV algorithm to find the best combination of hyperparameters, as well as with the SMOTE balancing the underrepresented classes in our dataset, we can achieve the highest accuracy and performance. The SVM model comparisons are presented in Table II. It can be observed that, with the remapped classes, SMOTE, and GridSearch [15], we are able to get a higher accuracy of human trust characterization based on the collected physical and physiological information from participants than other SVMs.

TABLE II. SVM Model Comparisons.

Approach	Accuracy
SVM_7_classes_No_Smote_No_GridSearch	45.85%
SVM_7_classes_Smote_No_GridSearch	39.41%
SVM_7_classes_Smote_GridSearch Hyperparameters (C:10, Gamma:10, and Kernel: RBF).	49.90%
SVM_3_classes_Smote_No_GridSearch	63.7%
SVM_3_classes_Smote_GridSearch Hyperparameters (C:100, Gamma:10, and Kernel: RBF).	78.6%

C. Results and Analysis of TabPFN

TabPFN is chosen for its impressive architecture in handling complex data of small tabular datasets. Unlike SVM, the data need not be normalized prior to their usage for model training. TabPFN has a restriction of the dataset being larger than 1024 data points, so SMOTE is not needed since it may increase the dataset size making TabPFN not work efficiently. With TabPFN, we perform two training rounds, once with all 7 classes, and once with the remapped classes defined prior. We also train it on the full dataset, and in doing so, we have to split the data appropriately to fit the approach. To do this, we have applied a function that allows us to split the training and test data. For the data to be split appropriately, the test size is found to be 44% of the data. With this realization, we conduct two additional training methods, where we split the data into two training sessions. In the first session, the model is trained on 50% of the data with the split consisting of 80% training data and 20% testing data. The model is then reloaded and continues to be trained on the other half of the data with the same split. This method is executed twice, once with all 7 features, and once with the features being resampled. The accuracy comparisons of TabPFN-based approaches are shown in Table III.

TABLE III. TabPFN Model Comparisons.

Approach	Prediction Time (s)	Accuracy
TabPFN_7_Classes	~4.22	48.86%
TabPFN_3 Classes	~4.23	82.03%
TabPFN_Split_Training_7_Classes	~1.51	55.92%
TabPFN_Split_Training_3_Classes	~1.68	87.5%

From Table III we can see when remapping the 7 classes to 3 classes, the prediction accuracy significantly increases, with minimal effect on computation and prediction time. Furthermore, by splitting the data in half and training the model twice, the prediction time decreased by more than half, and accuracy increased by a significant amount.

D. Results and Analysis of XGBoost

This study capitalizes on the XGBoost algorithm's proficiency in applying gradient boost, regularization, and tree pruning. Due to the nature of the data being complex and

having a high dimensionality, XGBoost is selected since it can well handle such data and produce interesting results. Like SVM, we perform several training sessions in different ways. As shown in Table IV, we can see the comparison results.

TABLE IV. XGBOOST MODEL COMPARISONS.

Approach	Average mLogLoss	Accuracy
XGB_7_Classes_No_Smote_Exact_Method	~1.33	51.87%
XGB_7_Classes_Smote_Exact_Method	~1.34	49%
XGB_3_Classes_No_Smote_Exact_Method	~0.57	82.11%
XGB_3_Classes_Smote_Exact_Method	~0.65	76.15%

Dissecting the data in Table IV, the methods not using SMOTE perform better. This is due to the underrepresentation of the lower trust classes. The data is dominated by the higher trust classes. This ultimately creates a bias in the model, leading to higher prediction accuracy due to the overwhelming amount of data in the high trust classes. The other two methods using SMOTE do not completely solve the issue because they balance the high-trust and low-trust datasets more than the other two models without SMOTE. When they balance the dataset and increase the significance of the lower trust classes, the loss and accuracy do change. Although the accuracy is lower and the loss is slightly higher, the results are exceptional and indicate that participants' biometric data and their trust ratings have a correlation. The results also suggest that human trust can be characterized and predicted in human-robot collaboration.

E. Discussion

Three computational models (SVM, TabPFN, and XGBoost) are employed for the characterization of human trust based on the participants' physical and physiological information collected during the human-robot collaborative tasks. Table V showcases the top accuracy of each model. It can be observed that the best result is achieved by TabPFN. Due to its ability to handle small tabular datasets, high complexity handling, and internal preprocessing techniques, additionally being incrementally trained, it is proven to produce highly accurate results. SVM and XGBoost are proven to generate good results that can be further improved. Overall, we can conclude that human trust can be well characterized and predicted via the participant's physical and physiological information. The results also indicate that the biometric status of a human that is caused by the robot's performance in human-robot collaborative contexts may influence his/her trust in the robot.

TABLE V. COMPARISONS OF THE TOP RESULTS OF ALL APPROACHES

Approach	Accuracy
SVM_3_classes_Smote_GridSearch Hyperparameters (C:100, Gamma:10, and Kernel: RBF).	78.6%
TabPFN_Split_Training_3_Classes	87.5%
XGB_3_Classes_Smote_Exact_Method	76.15%

VII. CONCLUSION AND FUTURE WORK

In this study, we have utilized the data of TrustBase to conduct deep analysis as well as characterization and prediction of participants' trust levels based on their physical and physiological information in the context of human-robot collaboration. The data in TrustBase consists of 100 user studies, with datasets of participants' four types of biometric information and their trust ratings during a human-robot collaboration process. The participants' biometric information contains 44 physical and physiological attributes. After conducting feature analyses, 38 of 44 attributes are found to be prominent in the characterization and analysis of human trust. The data is preprocessed accordingly and then used for model training. The three computational architectures utilized in analyzing the data are SVM, TabPFN, and XGBoost. After fitting the data and training each algorithm, we have found that all models can obtain desired results. These results indicate that human trust can successfully be characterized by using human physical and physiological information.

Subsequent endeavors should further process and analyze the data in TrustBase. We hope to find more insights into how robot performance factors can affect users' trust and how this trust can be better parametrized by their information when working with their robot counterparts.

ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation under Grants CNS-2104742, CMMI-2338767, and CMMI-2301678.

REFERENCES

[1] I. Singh *et al.*, "Progprompt: Generating situated robot task plans using large language models," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023: IEEE, pp. 11523-11530.

[2] K. Mahadevan *et al.*, "Generative expressive robot behaviors using large language models," *arXiv preprint arXiv:2401.14673,* 2024.

[3] W. Wang, R. Li, Y. Chen, Y. Sun, and Y. Jia, "Predicting Human Intentions in Human-Robot Hand-Over Tasks Through Multimodal Learning," *IEEE Transactions on Automation Science and Engineering,* vol. 19, no. 3, pp. 2339-2353, 2022.

[4] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling Object Hand-Over in Human-Robot Collaboration Via Natural Wearable Sensing," *IEEE Transactions on Human-Machine Systems,* vol. 49, no. 1, pp. 59-71, 2019.

[5] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access,* vol. 8, pp. 75264-75278, 2020.

[6] O. E. Dictionary, in *Oxford English Dictionary*, ed: Oxford University Press, 2023.

[7] N. Buchan, "The complexity of trust: Cultural environments, trust, and trust development," *Cambridge handbook of culture, organizations, and work,* pp. 373-417, 2009.

[8] J. Parron, T. T. Nguyen, and W. Wang, "Development of A Multimodal Trust Database in Human-Robot Collaborative Contexts," in 2023 *IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2023, pp. 0601-0605.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[10] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462,* 2015.

[11] P. Bourke, "Cross correlation," *Cross Correlation", Auto Correlation—2D Pattern Identification,* 1996.

[12] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications,* vol. 13, no. 4, pp. 18-28, 1998.

[13] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, "Tabpfn: A transformer that solves small tabular classification problems in a second," *arXiv preprint arXiv:2207.01848,* 2022.

[14] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[15] P. Zhang, S. Shu, and M. Zhou, "An Online Fault Detection Method based on SVM-Grid in Clouds," *IEEE/CAA J. of Automatica Sinica*, 5(2), pp. 445-456, 2018.