Predicting Team Well-Being Through Face Video Analysis With AI

Moritz Mueller , Ambre Dupuis , Tobias Zeulner , Ignacio Vazquez , Johann Hagerer , Peter A Gloor *

Posted Date: 20 December 2023

doi: 10.20944/preprints202312.1495.v1

Keywords: Individual well-being; machine learning; non-verbal communication; video analysis; Teamwork; PERMA

Article

Predicting Team Well-Being through Face Video Analysis with AI

Moritz Müller [1,‡], Ambre Dupuis [1,‡], Tobias Zeulner [1], Ignacio Vazquez [2], Johann Hagerer [3], and Peter A. Gloor [1,*]

1 Center for Collective Intelligence, Massachusetts Institute of Technology
2 System Design and Management, Massachusetts Institute of Technology
3 Technical University of Munich
* Correspondence: pgloor@mit.edu

Abstract: Well-being is one of the pillars of positive psychology, which is known to have positive effects not only on the personal and professional lives of individuals, but also on teams and organizations. Understanding and promoting individual well-being is essential for staff health and long-term success, but current tools for assessing subjective well-being rely on time-consuming surveys and questionnaires, which limit the possibility of providing the real-time feedback needed to raise awareness and change individual behavior. This paper proposes a framework for understanding the process of non-verbal communication in teamwork, using video data to identify significant predictors of individual well-being in teamwork. It relies on video acquisition technologies and state-of-the-art artificial intelligence tools to extract individual, relative, and environmental characteristics from panoramic video. Statistical analysis is applied to each time series, leading to the generation of a dataset of 125 features, which are then linked to PERMA (Positive Emotion, Engagement, Relationships, Meaning, and Accomplishments) surveys developed in the context of positive psychology. Each pillar of the PERMA model is evaluated as a regression or classification problem using machine learning algorithms. Our approach was applied to a case study, where 80 students collaborated in 20 teams for a week on a team task in a face-to-face setting. This enabled us to formulate several hypotheses identifying factors influencing individual well-being in teamwork. These promising results point to interesting avenues for research, for instance fusing different media for the analysis of individual well-being in teamwork.

Keywords: individual well-being; machine learning; non-verbal communication; video analysis; teamwork; PERMA

1. Introduction

Since the end of the 20th century, mental health and well-being have become the new driving forces of psychology. Positive psychology prefers, to the treatment of mental illnesses the exploration and nurturing of the elements that contribute to human fulfillment [1]. Indeed, research has shown that having a sense of well-being can lead to positive outcomes in life including improved health, flourishing relationships, better academic performance [2] but also in organizations to increase productivity, collaboration, customer satisfaction, and reduction of turnover[3,4]. Thus, understanding and promoting individual well-being is essential to the health of the workforce and the long-term success of an organization. However, despite these benefits, identifying individual well-being in the case of collaboration within a co-located team can prove challenging [5]. In addition, most current tools for assessing subjective well-being rely on time-consuming surveys and questionnaires, which limit the possibility of providing real-time feedback necessary to raise awareness and change individual behavior [6]. Since non-verbal communication, mostly visual cues [7,8], offers a precious and non-intrusive way to gather emotional and cognitive information on an individual's state of mind [9–11], the aim of this study is to understand the non-verbal communication process in teamwork,

 doi:10.20944/preprints202312.1495.v1

using video data to identify significant predictors of individual well-being in teamwork. We address the three following research questions :

- **RQ1 :** Which features of videos taken in a team setting will be predictive of individual and team well-being measured with PERMA (Positive Emotion, Engagement, Relationships, Meaning, and Accomplishments) surveys?
- **RQ2 :** How can the relevance of attributes for predicting individual well-being in a collaborative work context be measured?
- **RQ3 :** How can theories and hypotheses relevant to positive psychology be derived from AI-driven team video analysis?

Answering these questions will help experts from sociology and psychology to elaborate new theories and hypotheses based on large amounts of in-the-wild data representative of all the diversity of human behavior. Among other things, this information will be useful for organizing more effective and collaborative teamwork sessions. They could also help to promote policies that favor individual well-being, thereby increasing employee happiness and retention in companies.

In the following, a brief overview of the non-verbal communication and well-being data analysis research will be carried out in Section 2. The proposed framework to extract relevant features of non-verbal communication and well-being analysis will be presented in Section 3 while the experiment developed to test this framework will be presented in Section 4. The results obtained from a case study are introduced in Section 5 and will be discussed in Section 6. This will lead to the conclusion, in Section 7, about significant predictors of individual well-being in teamwork as well as on possible directions for future research.

2. Related work

2.1. PERMA and the notion of Well-being

The benefits of well-being as the overall state of an individual's happiness, health, and comfort [12] are widely recognized for individuals, organizations, and society as a whole [2–4]. Positive psychology is the branch of psychology concerned with the notion of well-being, as it explores and nurtures the elements that contribute to human flourishing [1]. Providing a holistic view of well-being, one of the leading figures of the positive psychology movement, Seligman [13] proposed the PERMA model. Based on the Well-being theory established by Forgeard et al. [14], the PERMA model decomposes well-being into five pillars described as the level of pleasant emotions such as happiness, joy, etc. experienced (**P**ositive emotions) [13], the level of absorption experienced during an activity (**E**ngagement), the degree of connection with other individuals (**R**elationships) [15], the degree to which the individual finds meaning in life (**M**eaning), and finally, the level of realization of one's full potential (**A**ccomplishment) [16]. Based on the model of Seligman [13], a number of PERMA measurement tools have been proposed for general assessments[15] or more work-related environments [16,17]. The PERMA+4 framework proposed by Donaldson et al. [17] represents a lean tool specifically tailored for the working environment allowing survey time to be reduced. The speed of data collection provided by this method is a considerable advantage over other methods since it simplifies the collection of a sufficient data set to enable data-based analysis of individual well-being in collaborative work.

2.2. Team collaboration and well-being data analysis

Various approaches for data collection in teamwork environments are widely available in the literature. Online settings have been used to measure emotional conditions or engagement in e-sports teams [18] and student groups [19,20] respectively. One of the advantages of the online setting is that it limits the need for data preparation since the records of each individual are already disentangled. Other studies focus on measuring team behavior in a co-located environment within surgical teams

 doi:10.20944/preprints202312.1495.v1

[21–23] and laboratory teams [24–27], working in highly controlled environments. While [24–27] used multimodal frameworks as Guerlain et al. [21] and Ivarsson and Åberg [22] which used audiovisual data, Stefanini et al. [23] used sociometric badges developed by Kim et al. [28] to extract behavioral features such as mutual gaze, interpersonal distance, and movement patterns.

All the research mentioned above uses data from highly controlled environments compared to *in-the-wild* data collected in real-world conditions, outside of a controlled environment, with multiple teams working in parallel.

While the examples listed above use sensors to measure interpersonal interaction, most teamwork is studied through surveys, which makes analyzing well-being in collaborative work all the more complex as surveys are generally time-consuming and intrusive [29].

3. Methods

To understand the non-verbal communication process in teams, we propose to use video data to identify significant predictors of individual well-being in teamwork. Towards this goal, a two-step facial-analysis-system (FAS), illustrated in Figure 1 and detailed below, has been developed. It leverages state-of-the-art deep learning technologies to combine a **multi-face tracking** approach and a **multi-task feature extraction**.

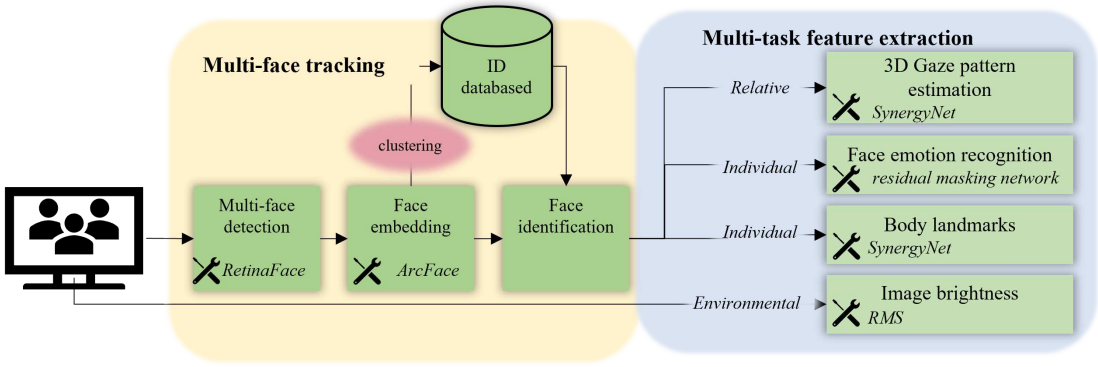


Figure 1. Two-step FAS proposed for video-feature extraction in a well-being analysis context

3.1. Data presentation

To test the proposed FAS, video data was collected. To do so, an experiment was conducted over three days with 20 co-located on-site teams, each composed of 4 master's students. During those teamwork sessions, participants were asked to work on a team project composed of different tasks such as project design and stakeholder analysis. The study only includes data from the 56 students who signed the informed consent form. Its purpose is to record non-verbal dynamics during collaborative teamwork in order to understand the non-verbal communication process, using video data to identify significant predictors of individual well-being in teamwork.

The experimental setup represented in Figure 2 has been replicated on each of the 20 team's tables.

As shown in Figure 2, the four participants in each team are placed on opposite sides of the table, in pairs, facing each other. A wide-angle camera [30] is placed in the exact center of the table (in both x and y directions) to record the 1.5 hours of daily teamwork. The camera is stacked on top of the mini-PC. The camera was connected via USB to minimize the size and intrusiveness of the measurement setup. Finally, to reduce visual background noise, whiteboards topped with folding partitions were placed between adjacent tables.

The acquisition of full panoramic scenes allows the analysis of non-verbal cues such as 3D gaze pattern estimation. The structure selected for recording is a stack of two 180-degree images. Participants on either side of the table are systematically observed on the top or bottom image respectively. This arrangement facilitates subsequent analysis of the video data by the FAS.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

Figure 2. Measurement setup to record a single team video data.

The final data collection and cleaning resulted in approximately 93 hours of video data stored as MP4 files for all 20 teams analyzed on the three days of observation. This resulted, on average, in 4.5 hours of video data per team and, thus, 1.5 hours per team per day and was taken as data source for the subsequent well-being analysis.

The video data collected had to be labeled with well-being attributes in order to be used to analyze participants' well-being. For this reason, participants were asked to complete a PERMA +4 questionnaire at the end of each work session, to assess their level of well-being according to the different pillars designated by the PERMA framework.

The PERMA data collected resulted in 104 data points from the 56 study participants over the three days.

These data points are used as ground truth for training the machine learning model with the video data collected with the proposed FAS detailed below.

3.2. Multi-face tracking

Each video is analyzed to determine the respective trajectory of each face present in the recording, using a **multi-face tracking** approach. All faces present in a single video frame are detected and embedded using the *RetinaFace* model [31] and the *ArcFace* model [32], respectively. The *RetinaFace* model detects a set of faces $F = \{F_1, F_2, \ldots, F_m\}$ in a given frame. Each $F_m \in F$ is transformed to a lower dimension face embedding $E = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m\}$ using *ArcFace* for greater computational efficiency. Finally, an ID database is generated by clustering a sample of frames from the video based on the number of individuals per team. It is then used to identify and track each individual in the video through face identification. The challenge of re-identification - the process of correctly identifying person identities across video frames - is tackled by calculating the cosine distances between preprocessed face templates $I = \{\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_n\}$ and the detected face embeddings E. Then the Hungarian algorithm [33] is used to solve the assignment problem. This approach allows an efficient tracking of multiple faces in a video stream. No tracking algorithm in the traditional sense is implemented, while the focus is on facial attributes.

3.3. Multi-task feature extraction

After the face of each member is identified, the second step of the proposed FAS, the **multi-task feature extraction**, is employed on the detected faces F to extract features for the subsequent well-being analysis. Four direct features are extracted.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

The Face emotion recognition (FER) is used to identify and classify human emotions based on facial expressions using the *residual masking network* [34], which performs state-of-the-art analysis on the FER2013 data set to estimate the six Ekman emotions [10] plus an added "neutral" emotion for increased machine learning accuracy. Face alignment is not explicitly employed in this methodology to prevent potential information loss or artifacts.

The body landmarks are based on the face-center position while the Gaze estimation evaluates who is looking at whom in a panoramic scene. The approach is based on 3D head pose and facial landmark estimations to identify where a person is looking. Specifically, *SynergyNet* [35] is used to estimate the full 3D facial geometry. The head poses, and facial landmarks are first spatially transformed to reconstruct the original 3D scene. Then, a visibility algorithm adapted from [36] is employed to detect gaze exchanges among individuals. To do so, the human field of view (FOV) angle for 3D gaze pattern estimation has to be set to a specific angle. The number of gaze exchanges is captured in a gaze matrix populated over the duration of the video stream and illustrated in Figure 3.

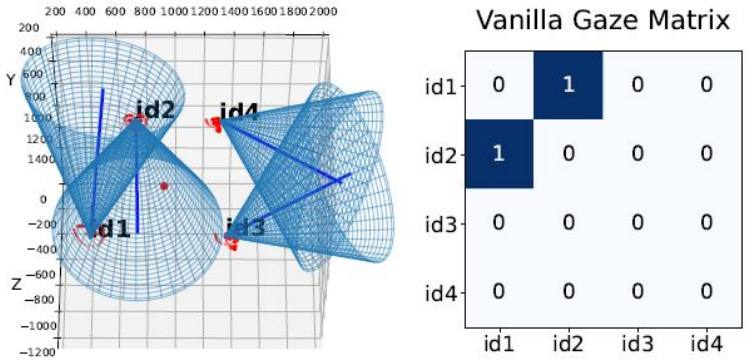


Figure 3. Sample team 3D gaze exchange and gaze matrix.

Finally, the brightness of the image is extracted directly from the video, reflecting an environmental characteristic. Each team member is assigned the perceived image brightness calculated across all images using the root mean square (RMS) described in Equation1. It weighs the contributions of the red (R), green (G), and blue (B) channels to take into account the heterogeneity of human perception [37].

$$\mathrm{b} = \sqrt{0.299 \cdot (R^2) + 0.587 \cdot (G^2) + 0.114 \cdot (B^2)} \tag{1}$$

While the face emotion recognition and body landmarks are specific to each individual, the gaze patterns are relative since they result from interactions between team members. Those direct features are used to extract derivative features valuable for the machine learning models and summarized in Table 1.

The emotion recognition data includes details about the emotional and affective states of every team member. The time series for each of Ekman's six basic emotions plus "neutral", alongside the distribution of each emotion (*Max Emotion*) and the frequency of changes in emotion (*Freq Emotion changes*), are extracted. The Body Landmarks data provides the position of the head centers of individuals using the standard deviation of the 2D kernel density data distributions in the X and Y direction. It expresses the spatial extent to which the individual moved during the analyzed video. From this data, the velocity of the head's movement is extracted as a time series by calculating the difference in position between two consecutive frames. Additionally, the Presence feature represents the percentage of frames an individual is identified in. The level of Brightness is directly extracted from the video as a time series. Finally, the 3D gaze pattern estimation is used to generate interaction matrices and extract social network metrics. The gaze matrix, illustrated in Figure 3, is computed by counting the number of times each individual looks at a team member. This asymmetrical matrix is combined into to symmetrical matrix, the gaze-difference matrix, and the mutual gaze matrix. The first

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

represents the difference between the total gazes emitted by person i to person j and the reciprocal, while the second only incorporates entries where two participants look at each other simultaneously. Features are extracted from those three matrices using 8 basic statistics Mean, Standard Deviation, Median, Max, Min, Slope, 75th percentile, and 25th percentile. Social network analysis of the gaze matrix allows us to extract in-degree and out-degree centrality for each individual.

Linear interpolation is used to fill in missing numerical data while a rolling average with a time-series-specific window is used to smooth noise.

The result of the proposed FAS is a dataset χ of 125 features generated using, once again, the 8 basic statistical features to describe each time series (Mean, Standard Deviation, Median, Max, Min, Slope, 75th percentile, and 25th percentile).

Table 1. Summary of the attributes extracted from the videos by the proposed FAS

Origin	Category	Feature	Type
Emotion Recognition	Emotional state	Neutral	Time Serie
		Happy	Time Serie
		Sad	Time Serie
		Disgust	Time Serie
		Surprise	Time Serie
		Angry	Time Serie
		Fear	Time Serie
		Max Emotion	%
		Freq Emotion changes	%
	Affective state	Valence	Time Serie
		Arousal	Time Serie
		Dominance	Time Serie
Body Landmarks	Head motion patterns	Velocity	Time Serie
		Presence	%
		Positional	[X,Y]
Image brightness		Brightness	Time Serie
3D Gaze pattern estimation	Gaze patterns	Gaze Social Network Analysis	SNA
		Gazes statistics	Statistics
		Gaze-difference statistics	Statistics
		Mutual Gaze statistics	Statistics

4. Experiments

4.1. Data Collection

In order to test the proposed framework, the following experiment is conducted. The experiment is based on the exploitation of panoramic video files of work teams and PERMA survey forms completed by each individual at the end of filmed work sessions. Based on the work of [38], audio and video data are collected simultaneously in distributed teams.

The results of each question of the PERMA+4 survey by Donaldson et al. [17] are averaged by pillar in order to get a dataset of 5 target variables *tar* representing the 5 pillars of the PERMA model for each individual in each video file.

Figure 4 resumes the experiment in which the proposed framework is implemented in order to understand the non-verbal communication process in teamwork, using video data to identify significant predictors of individual well-being in teamwork.

The panoramic video files are formatted and linked to the PERMA surveys in the **Data preparation** phase (green). Then, those data are used in regression and classification models in the **Data analysis** phase (yellow) in order to obtain a prediction and classification of individual well-being. The explainability of the prediction and classification by the identification of significant predictors is provided in the **Feature of importance** phase (blue) by the computation of SHAP values.

Each of these phases will now be described in detail.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

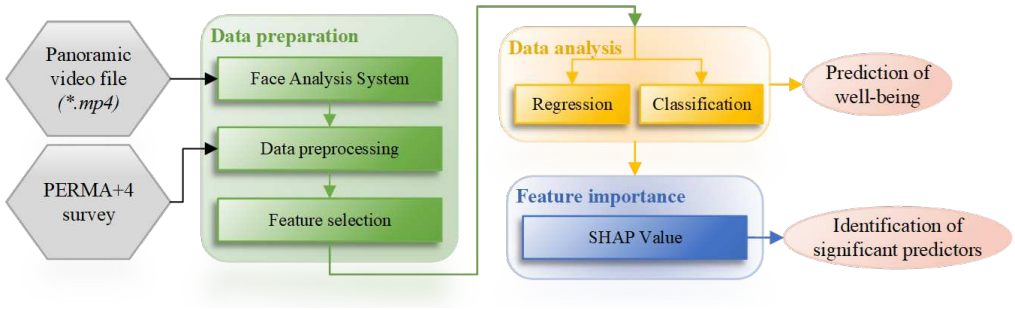


Figure 4. Experiment for individual well-being analysis using panoramic video data

4.2. Data preparation

The first phase is the **Data preparation**. The panoramic video files are preprocessed to extract pertinent information usable by the machine learning models.

First, the proposed FAS presented in Section 3 is used to generate the dataset of features related to each individual in each video. Then, each record is linked to the associated PERMA labels **Y**. The PERMA data in the **Y** dataset are preprocessed to handle missing values and outliers. Also, both the χ and the **Y** dataset are normalized to be used in the machine learning models. Thus, in the ***Data preprocessing*** step, all records linked to a missing value or to a constant value throughout all the pillars of the PERMA survey are removed. PERMA variables are normalized using a min-max normalization while the dataset features are normalized using a standard or robust scaling depending on their distribution [39].

The PERMA variables contained in the Y dataset are continuous variables. Regression is therefore the most straightforward data analysis model. However, it may also be useful to classify each variable into binary categories (High or Low level), as this aligns with the overall goal of the research. Classification metrics offer more intelligible performance scores than regression metrics [40]. Thus, a new dataset called Y_{bin} is generated by discretizing the Y dataset. The discretization is done by applying a median threshold to each dimension of Y for binary classification. In order to reduce the complexity of the methodology and provide interpretable results, each targeted variable *tar* present in Y and Y_{bin} is analyzed independently in univariate problems.

To further limit the complexity of the models and comply with Occam's razor principle, the features extracted in χ are then selected in the ***Feature selection*** step to generate the X dataset. The attribute selection method is preferred to the dimensionality reduction method for reasons of interpretability of the results [41]. To perform feature selection only within the training set to prevent data leakage, the **X**, the **Y**, and the Y_{bin} datasets are divided into a training set (X_{train}, y_{train}, and $y_{bin_{train}}$) and a test dataset (X_{test}, y_{test}, and $y_{bin_{test}}$) representing 80% and 20% of the total dataset respectively. Then, a voting strategy among filters presented in Table 2 is defined for feature selection. Those filters are chosen since they are relatively computationally efficient and model agnostic.

Table 2. Summary of filters used

Filter ID	Name	Reference
1	Univariate Linear Regression/ANOVA F-value	[42]
2	Mutual information	[43]
3	Variance thresholding	[42]
4	Percentile of the highest scores	[44]
5	False Positive Rate	[45]
6	False Discovery Rate	[45]
7	Family-wise error rate	[45]

Sets of features are evaluated for each target variable *tar* by the voting system using Equation 2.

$$S(\Phi) = \sum_{id=1}^{7} w_{id} \cdot S_{id}(\Phi) \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

Where Φ represents the set of features considered, *id* the filter ID, $S_{id}(\Phi)$ the ensemble scores from the filter *id* for all features in Φ, and finally, ω_{id} represents the weight given to the filter *id* based on the importance of the filter to the issue at hand [41]. The set of features with the highest $S(\Phi)$ score is chosen for the **Data Analysis** phase.

4.3. Data Analysis

To provide a classification and a regression analysis, different models are used. Each of those models has hyperparameters that have to be tuned for proper performance of the models. Table 3 provides a summary of the models used for the classification and the regression task respectively. It also summarizes the various hyperparameters tuned using grid search and cross-validation on the training dataset.

Table 3. Classification and regression models and associated hyperparameters used in the methodology

Classification	Regression	Hyperparameters
Gaussian Naive Bayes	-	var smoothing
K-Nearest Neighbors	K-Nearest Neighbors	n neighbors
Logistic Regression	-	C, penalty, solver, class weight
-	Linear Regression	-
Ridge Classifier	Ridge Regression	alpha, class weight
-	Lasso Regression	alpha
-	Elastic Net	alpha
Decision Tree	Decision Tree	max depth
Support Vector Machine	Support Vector Regression	kernel, C, shrinking, class weight, epsilon
-	Bayesian Ridge	alpha 1, alpha 2
Random Forest	Random Forest	n estimators, max depth, class weight
Extra Trees	Extra Trees	n estimators, max depth, class weight
AdaBoost	AdaBoost	n estimators, learning rate
Gradient Boosting	Gradient Boosting	n estimators, max depth, learning rate
CatBoost	CatBoost	iteration, depth, learning rate, auto class weights
XGBoost	XGBoost	iteration, depth, learning rate, scale pos weight

In red are hyperparameters used for classification models only while in in blue are hyperparameters used for regression models only.

For each target variable *tar* of the PERMA survey, the training set (X_{train}, y_{train}, and $y_{bin_{train}}$) is split in *k-folds* in order to find the best combination of hyperparameters. The chosen model is the one that has the lowest validation error or the highest performance metric, such as balanced accuracy for classification or MAE for regression. Finally, the models are trained using the training sets.

4.4. Feature of importance

Regression models can be used to analyze the coefficients associated with each attribute to determine its importance. Tree-based models can also provide insight into the importance of attributes by analyzing the mean decrease in the impurity (MDI). However, they don't really give any indication of the impact of attributes on prediction or classification [46]. For this purpose, the SHAP value can be used [47].

SHAP values are computed by averaging the influence of one feature over all possible combinations of features in the model [48]. In this way, the data from each of the models generated and trained during the **Data analysis** phase (subsection 4.3) are analyzed in order to extract the influence of features across multiple models allowing the comparison of the effects of each features and the identification of the most influential features for the prediction and classification of each PERMA pillar *tar*.

 doi:10.20944/preprints202312.1495.v1

5. Results

5.1. Data preparation

During the data preparation phase, the FAS extracts multiple initial features directly from the video stream in a time series structure as summarized in Table 1. The human field of view (FOV) angle for 3D gaze pattern estimation is set to 60°. A window size of 30 seconds is chosen for the rolling average on face emotion to reduce noise.

Since there is no contextual information that would allow one filter to be preferred to another in the proposed case study, the ω_{id} values of the voting system described by Equation 2 are set to 1.

5.2. Data analysis

As described in Section 4, the prediction of the PERMA scores is approached both as a regression and as a binary classification task (classification of the PERMA score level as high or low).

A 5-fold cross-validation on the training set is used to tune the models under consideration. Each pillar of the PERMA model is analyzed independently. Table 4 and Table 5 depict the regression and classification models respectively, as well as their hyperparameters offering the best performance on the validation set.

Table 4. Optimal hyperparameters of the regression models.

Dimension	Model	Best Hyperparameters
P	CatBoostRegressor	Iterations: 50, Learning Rate: 0.01, Depth: 4, Loss Function: RMSE
E	AdaBoostRegressor	Learning Rate: 0.1, N Estimators: 400
R	BayesianRidge	Alpha 1: 1.0, Alpha 2: 1.0
M	ElasticNet	Alpha: 0.01, L1 Ratio: 0.9
A	BayesianRidge	Alpha 1: 0.001, Alpha 2: 0.1

Table 5. Optimal hyperparameters of the binary classification models.

Dimension	Model	Best Hyperparameters
P	CatBoostClassifier	Iterations: 50, Learning Rate: 0.1, Depth: 3, Auto Class Weights: Balanced
E	CatBoostClassifier	Iterations: 50, Learning Rate: 0.01, Depth: 4, Auto Class Weights: SqrtBalanced
R	ExtraTreesClassifier	Class Weight: balanced, Max Depth: 2
M	CatBoostClassifier	Iterations: 50, Learning Rate: 0.1, Depth: 2, Auto Class Weights: Balanced
A	CatBoostClassifier	Iterations: 50, Learning Rate: 0.01, Depth: 2, Auto Class Weights: Balanced

The predominance of the CatBoostClassifier model in the classification task is evident in Table 5. This model is chosen for the classification of the level of four of PERMA's five pillars. There is no such evidence in the regression task since, as described in Table 4, each pillar is predicted by a different model, with the exception of pillars R and A, which are both predicted by the BayesianRidge model.

The best models and their associated hyperparameters are trained and tested using the training and the test set respectively.

The performance on the test set of the regression models is calculated using the MAE metric to measure the mean absolute error between predicted and actual values [44]. In Figure 5, the performance of each model is compared to a baseline where the proposed precision is the average pillar value observed over the test set.

The performance on the test set of the classification models is calculated using the balanced accuracy metric to encourage the model to correctly predict examples from all classes, regardless of their size [49]. This is done by averaging the percentage correctly predicted for each class individually. In the case of binary classification, the probability of predicting the right class when the data distribution is uniform is 50% [50]. Thus, a naive classifier with 50% balanced accuracy is used as the baseline. The comparison between the baseline and the performance of the classification models for each PERMA pillar is shown in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

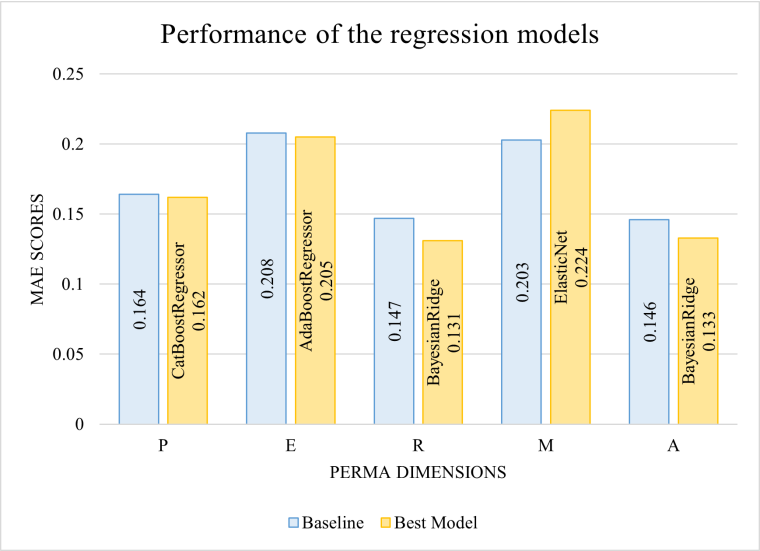


Figure 5. PERMA regression results (lower is better).

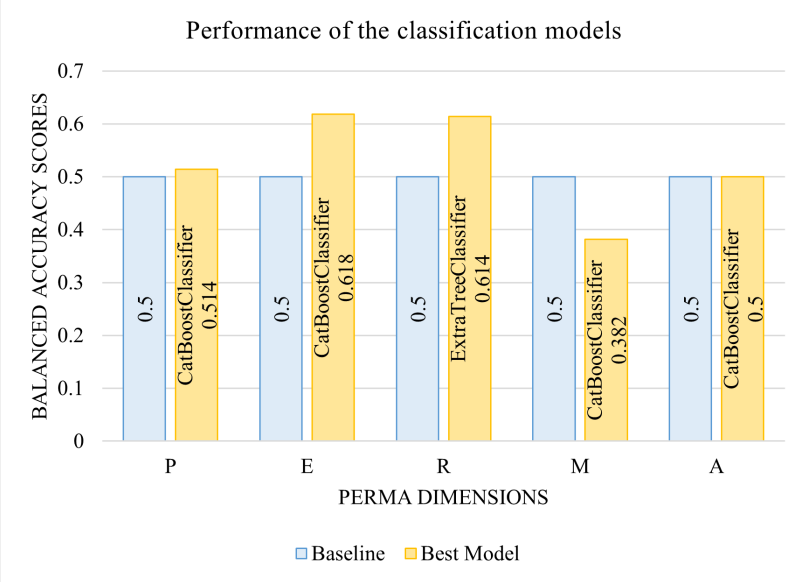


Figure 6. PERMA classification results (higher is better).

The results show that for most of the PERMA dimensions, with the exception of the *Meaning* dimension, the best-performing regression and binary classification models outperform the baseline.

The regression and binary classification models outperform the baseline, on average, by 1.5% and 5.6% respectively. This may be an indication of significant relationships discovered by the models in the data.

5.3. Feature importance

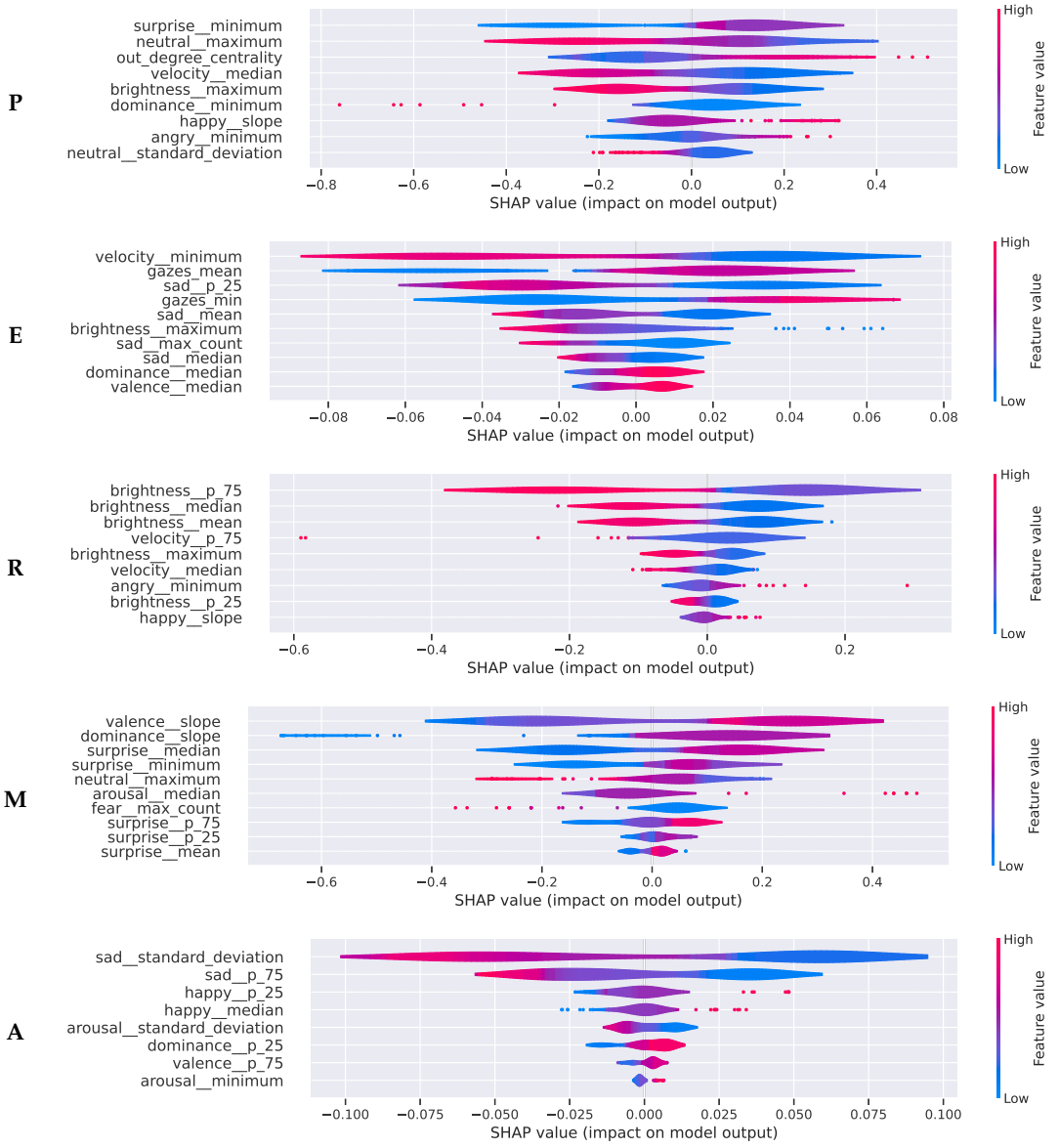
An analysis examining the Pearson correlation coefficient between each of the PERMA pillars and the individual features indicated at most weak correlations, with the highest being roughly 0.3.

To better understand the impact and dynamics of each feature on the final prediction and classification, a SHAP value analysis is undertaken. The SHAP analysis of the best binary classifier for the classification of each PERMA pillar is computed and the obtained results are proposed in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

Table 6. SHAP values for classification models across all PERMA-dimensions where red represents a high attribute value, while blue represents a low attribute value.

Binary Classification

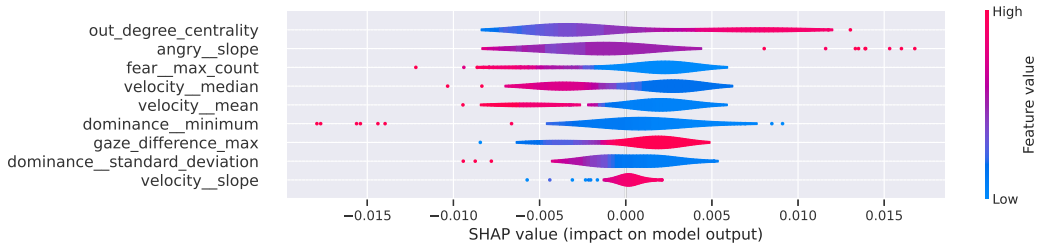
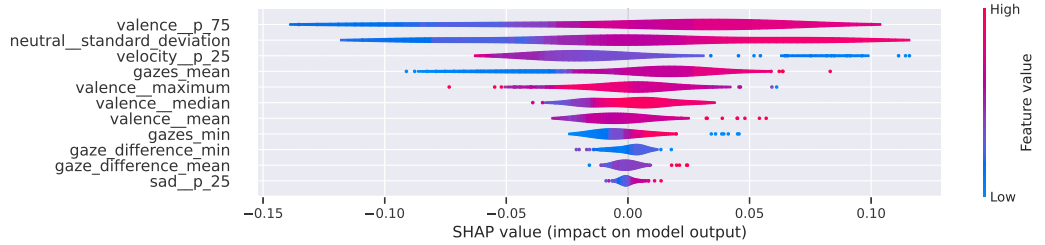
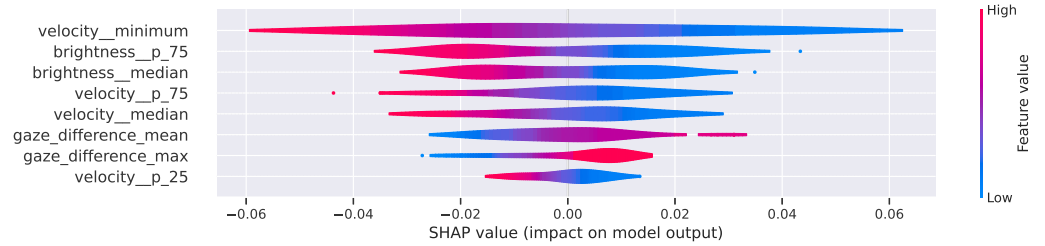
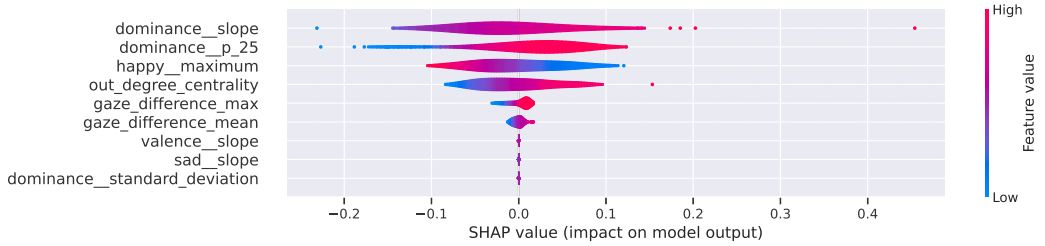
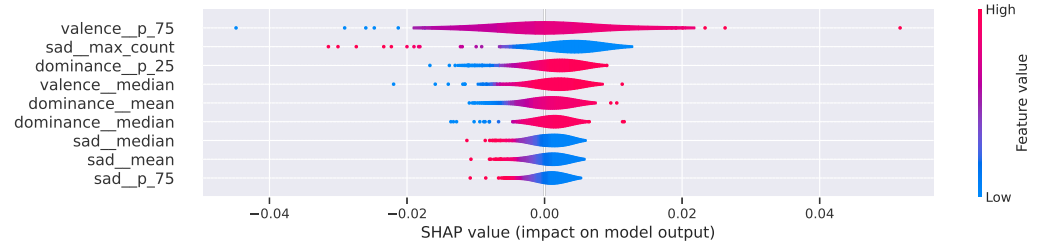


As presented in Figure 6, the attributes influencing classification vary greatly from pillar to pillar. The case study results indicate that the positive emotions (**P**), accomplishment (**A**) and meaning (**M**) pillars are largely influenced by the attributes derived from emotions. Based on Ekman's basic emotions, a high minimum level of surprise and a low maximum level of neutral emotion seem to positively influence pillar **P** while a low level of sadness standard deviation and third quartile seems to positively influence pillar **A**. This suggests that more stable emotional states are correlated with greater accomplishment. A high level of valence and dominance slope seems to be linked to the Meaning pillar (**M**) of the PERMA model. The engagement pillar (**E**) seems to be linked to head and gaze movements. A low level of minimum head velocity and a high average level of gaze exchange seem to have a positive impact on individuals' engagement in collaborative work. Finally, the relations pillar (**R**) seems to be linked to the environment in which the experiment takes place. Thus, attributes linked to luminosity have a strong impact on this pillar, with an advantage for low luminosity levels.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 December 2023 doi:10.20944/preprints202312.1495.v1

With the same objective of explicability, the SHAP analysis of the best regression model for the prediction of each PERMA pillar is computed and the obtained results are proposed in Figure 7.

Table 7. SHAP values for regression models across all PERMA-dimensions.

	Regression
P	
E	
R	
M	
A	

As for the binary classifier and as presented in Figure 7, the attributes influencing prediction and classification vary greatly from pillar to pillar. Once again, the case study results indicate that the accomplishment (**A**) and meaning (**M**) pillars are largely influenced by the attributes derived from emotions. However, the attributes used vary. The valence level as well as the number of times sadness is experienced by the participants seems to have an impact on the accomplishment pillar (**A**). For the meaning pillar (**M**), the dominance (slope and first quartile value) is once again influential with a positive correlation between meaning value and dominance attribute levels. Contrary to

 doi:10.20944/preprints202312.1495.v1

the binary classification model, the key element for the positive emotion pillar (**P**) in the regression task seems to be linked to the SNA metric of outdegree centrality. The more participants look at others, the more positive emotions they will experience. The commitment pillar (**E**) also seems to be linked to the participant's emotions, since the value of the third quartile of valence and the standard deviation observed for the neutral emotion are the most influential attributes for this pillar. Finally, it seems interesting that the relations pillar (**R**) seems, once again, to be linked to the brightness of the environment in which the experiment takes place but also to head movement. Similarly to the binary classifiers, attributes linked to brightness have a strong impact on this pillar, with an advantage for low brightness levels but contrary to the binary classifier, the minimum head velocity seems to have a positive impact on individuals' relationships.

6. Discussion

To recall, the aim of the proposed study was to understand the non-verbal communication process in teamwork, using video data and identify significant predictors of individual well-being in teamwork. The experiment conducted and the results obtained serve as a basis for discussion of the proposed research questions.

RQ1 : Which features of videos taken in a team setting will be predictive of individual and team well-being measured with PERMA surveys?

A framework combining state-of-the-art tools has been proposed in section 3 extracting from panoramic video data non-verbal cues, such as facial emotions, gaze patterns, and head motions as input for individual well-being analysis. An experiment presented in section 4 applies the proposed framework and links the extracted attributes to the results of PERMA+4 surveys evaluating the various pillars of well-being defined in positive psychology. This way, a data set of 125 features has been generated to predict the different pillars of the PERMA analysis. Machine Learning models were then trained for the regression and binary classification tasks to predict individual well-being scores, as defined by the PERMA framework.

When applied to a case study of collaboration within 20 co-located work teams, regression models outperform the baselines in four of the five PERMA dimensions, with a notable 1.5% improvement in MAE. Bayesian ridge regression was identified as particularly effective. In comparison, binary classification emerged as a more reliable approach, with models yielding a balanced accuracy improvement of 5.1%, also outperforming the baseline in four out of five PERMA dimensions. Ensemble models, specifically CatBoost, showed superior performance in this setting. Notably, the *Meaning* dimension of PERMA proved challenging in both prediction and classification settings, indicating difficulty in discerning a participant's sense of meaning purely from video cues.

RQ2 : How can the relevance of attributes for predicting individual well-being in a collaborative work context be measured?

SHAP values are used to interpret the impact of features on prediction and classification, independently of the machine learning model used. They also rank features according to their importance for the model under study [47]. Derived from cooperative game theory, SHAP values identify the importance of features for each data point, since they are decomposed into the sum of feature contributions. This provides a more transparent description of the model's behavior and therefore greater interpretability of the models[47]. Furthermore, this approach facilitates the identification of the most appropriate features for PERMA prediction by allowing the comparison of the influence of features across multiple models. [47].

RQ3 : How can theories and hypotheses relevant to positive psychology be derived from AI-driven team video analysis?

From the feature analysis with SHAP values, various theories, and hypotheses potentially relevant to experts in the field of positive psychology could be derived, for instance from the distribution of data points in the SHAP analyses.

 doi:10.20944/preprints202312.1495.v1

Based on the results of the case study, preliminary insights for team work could be gained: Paying attention to (i.e. looking at) team members appears instrumental in fostering happiness (**P**), calmer head movements seem to enhance engagement (**E**) and interpersonal relationships (**R**), the brightness of the environment (more light) may have an important impact on relationships (**R**), the sense of meaning (**M**) seems to be strongly tied to an increasing feeling of control, and finally results suggest that steady emotional states provide a greater sense of achievements (**A**).

Limitations

The results presented here are valid only for the discussed case study. Thus, although the methodology employed is generalizable, more similar case studies in different contexts and with different participants should be conducted to further investigate these conclusions in the field of cognitive sciences. These results show links but do not allow causalities to be determined. This is one of the limitations of the proposed methodology, but other factors should also be acknowledged. In data preparation, the FAS did not utilize explicit face alignment and treated each video frame in isolation, possibly overlooking the importance of temporal dynamics. These two factors could have a negative impact on the performance of the proposed model as they could, respectively, complicate emotion recognition and neglect temporal entanglements. Moreover, inherent assumptions in the employed algorithms, like using the Field of view (FOV)-cone model for gaze pattern estimation, can also introduce errors to the proposed findings. That is also true for the data preprocessing techniques employed, such as smoothing or linear interpolation, coupled with the dependence on specific feature selection strategies, which may introduce potential biases and uncertainties. Another limitation of the proposed study is the small number of data points available, which restricts an accurate exploration of the feature space. The relative scarcity of data points limits our predictive model's capacity to generalize beyond this study. While hyperparameter search space was leveraged by grid-search cross-validation, they might not capture the entirety of potential configurations. Also, the use of the SHAP-based feature analysis brings its own set of challenges. Finally, the modeling strategy relies on the fundamental assumption of relative independence among features, an ideal scenario that is challenging to achieve consistently. This assumption may mean that the model sometimes does not accurately capture interactions between features or possible non-linear effects.

7. Conclusion

Theories and hypotheses from sociology and psychology are necessary to better understand the behaviors and aspirations of the individuals and societies around us. However developing these theories and hypotheses is often difficult, as manual data collection for qualitative analysis by domain experts is time-consuming, limited, and prone to bias. To help experts develop theories based on a wider range of objective data, we propose a methodology to understand the non-verbal communication process in teamwork, using video data and identify significant predictors of individual well-being in teamwork.

Numerous studies analyze the well-being of individuals and teamwork, but thse studies are positioned in virtual or highly controlled environments (see Section 2). However, collaborative working generally takes place in uncontrolled, co-located environments.

To fill this gap, the proposed framework leverages video acquisition technologies and state-of-the-art artificial intelligence tools to extract from panoramic video individual, relative, and environmental features. Statistical analysis is applied to each time series, leading to the generation of a dataset of 125 features that is then linked to PERMA surveys.

A SHAP-based feature analysis unveils key indicators associated with the PERMA scores.

Applied to a case study, this method allows us to identify several hypotheses. For example, it seems that paying attention to team members is the key to happiness. It also appears that calm head movements promote individual commitment and interpersonal relations. Other hypotheses include the importance of the impact of the environment (brightness) on relationships, the close link between a sense of control and meaning, and the greater sense of achievement that stable emotional states bring.

However, those results should be nuanced since one case study is not enough to generalize these theories. The generalization of these results through the analysis of other case studies in various contexts is a promising line of research that will be interesting to study in the near future. In addition, practical improvements to the proposed FAS should be considered, such as explicit face alignment for better emotion recognition, taking into account the effects of temporal dynamics in image succession, or identifying and managing possible biases due to interpolation and line smoothing.

This study has identified some promising avenues of research. One lies in the fusion of different mediums for the analysis of individual well-being during teamwork. Indeed, the analysis of non-verbal communication could be combined with the analysis of verbal communication to have a holistic vision of communication patterns and develop an integrated framework for analysis of communication factors impacting individual well-being.

Author Contributions: Methodology, M.M., T.Z., P.G.; Software, M.M.; Formal analysis, M.M., T.Z.; Investigation, P.G., M.M, T.Z., I.V., J.H.; Data curation, M.M, T.Z.; Writing—original draft, M.M, A.D.;Writing—review & editing, P.G., J.H., I.V. All authors have read and agreed to the published version of the manuscript.

Funding: Moritz Mueller's stay at MIT was supported by the German Academic Exchange Service (DAAD).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. This study was approved by MIT COUHES under IRB 1701817083 dated 1/19/2023

Acknowledgments: We thank Bryan Moser for his invaluable support in the integration of our experiment during the Independent Activity Period at MIT. Moritz Müller acknowledges the financial support of the German Academic Exchange Service (DAAD) through a computer scientist (IFI) scholarship.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seligman, M.; Csikszentmihalyi, M. Positive Psychology: An Introduction. *The American psychologist* **2000**, *55*, 5–14. doi:10.1037/0003-066X.55.1.5.
2. Maccagnan, A.; Wren-Lewis, S.; Brown, H.; Taylor, T. Wellbeing and Society: Towards Quantification of the Co-benefits of Wellbeing. *Social Indicators Research* **2019**, p. 217–243. doi:https://doi.org/10.1007/s11205-017-1826-7.
3. Lyubomirsky, S.; King, L.; Diener, E. The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? **2005**. doi:10.1037/0033-2909.131.6.803.
4. Kompaso, S.; Sridevi, M. Employee Engagement: The Key to Improving Performance. *International Journal of Business and Management* **2010**, *5*. doi:10.5539/ijbm.v5n12p89.
5. Wright, T.; Cropanzano, R. Psychological well-being and job satisfaction as predictors of job performance. *Journal of occupational health psychology* **2000**, *5*, 84–94. doi:10.1037/1076-8998.5.1.84.
6. Gloor, P. Happimetrics; Edward Elgar Publishing: Cheltenham, UK, 2022; pp. 103–120. doi:10.4337/9781803924021.00015.
7. Mehrabian, A. *Silent messages.*; Wadsworth: Oxford, England, 1971; pp. 152, viii, 152–viii.
8. Birdwhistell, R.L. Kinesics and Context. *Kinesics and Context* **1971**. doi:10.9783/9780812201284.
9. Knapp, M.; Hall, J. Non-Verbal Communication in Human Interaction **2010**.
10. Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion, 1971. doi:10.1037/h0030377.
11. Pantic, M.; Rothkrantz, L.J.M. Toward an Affect-Sensitive Multimodal Human-Computer Interaction **2003**. doi:10.1109/JPROC.2003.817122.
12. Pinto, S.; Fumincelli, L.; Mazzo, A.; Caldeira, S.; Martins, J.C. Comfort, well-being and quality of life: Discussion of the differences and similarities among the concepts. *Porto Biomedical Journal* **2017**, *2*, 6–12. doi:https://doi.org/10.1016/j.pbj.2016.11.003.
13. Seligman, M.E.P. *Flourish: A visionary new understanding of happiness and well-being.*; Free Press: New York, NY, US, 2011; pp. 349, xii, 349–xii.
14. Forgeard, M.J.C.; Jayawickreme, E.; Kern, M.L.; Seligman, M.E.P. Doing the Right Thing: Measuring Well-Being for Public Policy. *International Journal of Wellbeing* **2011**, *1*. doi:10.5502/ijw.v1i1.15.
15. Butler, J.; Kern, M.L. The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing* **2016**, *6*, 1–48. doi:10.5502/ijw.v6i3.526.

16. Kun, A.; Balogh, P.; Gerákné Krasz, K. Development of the Work-Related Well-Being Questionnaire Based on Seligman's PERMA Model. *Periodica Polytechnica Social and Management Sciences* **2017**, *25*, 56–63. doi:10.3311/PPSO.9326.
17. Donaldson, S.I.; van Zyl, L.E.; Donaldson, S.I. PERMA+4: A Framework for Work-Related Wellbeing, Performance and Positive Organizational Psychology 2.0. *Frontiers in Psychology* **2021**, *12*, 817244. doi:10.3389/FPSYG.2021.817244.
18. Abramov, S.; Korotin, A.; Somov, A.; Burnaev, E.; Stepanov, A.; Nikolaev, D.; Titova, M.A. Analysis of Video Game Players' Emotions and Team Performance: An Esports Tournament Case Study. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26*, 3597–3606. doi:10.1109/JBHI.2021.3119202.
19. Nezami, O.M.; Dras, M.; Hamey, L.; Richards, D.; Wan, S.; Paris, C. Automatic Recognition of Student Engagement using Deep Learning and Facial Expression **2018**.
20. Savchenko, A.V.; Savchenko, L.V.; Makarov, I. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing* **2022**. doi:10.1109/TAFFC.2022.3188390.
21. Guerlain, S.; Shin, T.; Guo, H.; Adams, R.; Calland James, M.D. A Team Performance Data Collection and Analysis System. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **2002**, *46*. doi:10.1177/154193120204601608.
22. Ivarsson, J.; Åberg, M. Role of requests and communication breakdowns in the coordination of teamwork: a video-based observational study of hybrid operating rooms. *BMJ Open* **2020**, *10*, 35194. doi:10.1136/bmjopen-2019-035194.
23. Stefanini, A.; Aloini, D.; Gloor, P. Silence is golden: the role of team coordination in health operations **2020**. doi:10.1108/IJOPM-12-2019-0792.
24. Salvador Vazquez Rodarte, I. An experimental multi-modal approach to instrument the sensemaking process at the team-level. PhD thesis, 2022.
25. Koutsombogera, M.; Vogel, C. Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. Technical report, 2018.
26. Kontogiorgos, D.; Sibirtseva, E.; Pereira, A.; Skantze, G.; Gustafson, J. Multimodal reference resolution in collaborative assembly tasks. Proceedings of the 4th Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction, MA3HMI 2018 - In conjunction with ICMI 2018. Association for Computing Machinery, Inc, 2018, pp. 38–42. doi:10.1145/3279972.3279976.
27. Sanchez-Cortes, D.; Aran, O.; Mast, M.S.; Gatica-Perez, D. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia* **2012**, *14*, 816–832. doi:10.1109/TMM.2011.2181941.
28. Kim, T.; McFee, E.; Olguin, D.O.; Waber, B.; Pentland, A.S. Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior* **2012**, *33*, 412–427. doi:https://doi.org/10.1002/job.1776.
29. Kahneman, D.; Krueger, A.B.; Schkade, D.A.; Schwarz, N.; Stone, A.A. A survey method for characterizing daily life experience: The day reconstruction method. *Science* **2004**, *306*, 1776–1780. doi:10.1126/science.1103572.
30. available on https://en.j5create.com/products/jvcu360.
31. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild, 2020.
32. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition, 2019.
33. Sharma, A. MASTER OF COMPUTER APPLICATIONS. Technical report, 2002.
34. Pham, L.; Huynh Vu, T.; Anh Tran, T.; Chi Minh City, H.; Trung Ward, L.; Duc District, T. Facial Expression Recognition Using Residual Masking Network; Facial Expression Recognition Using Residual Masking Network **2020**. doi:10.1109/ICPR48806.2021.9411919.
35. Wu, C.Y.; Xu, Q.; Neumann, U. Synergy between 3DMM and 3D Landmarks for Accurate 3D Facial Geometry. *Proceedings - 2021 International Conference on 3D Vision, 3DV 2021* **2021**, pp. 453–463. doi:10.1109/3DV53792.2021.00055.
36. japreiss. How can I detect if a point is inside a cone or not, in 3D space?, 2023.
37. Smith, A.R. COLOR GAMUT TRANSFORM PAIRS. Technical report, 1978.

38. Törlind, P. A FRAMEWORK FOR DATA COLLECTION OF COLLABORATIVE DESIGN RESEARCH **2007**.
39. Raschka, S.; Mirjalili, V. *Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2*; 2017; p. 741.
40. James, G.G.M.; Witten, D.; Hastie, T.; Tibshirani, R. *An introduction to statistical learning : with applications in R*; 2013; p. 426.
41. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature selection: A data perspective. *dl.acm.org* **2017**, *50*. doi:10.1145/3136625.
42. Casella, G.; Berger, R.L.; Santana, D. Statistical inference-Solutions Manual. *Statistical Inference* **2002**, p. 195.
43. Cover, T.M.; Thomas, J.A. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing) (Hardcover). Technical report, 2012.
44. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning **2009**. doi:10.1007/978-0-387-84858-7.
45. Läuter, J. Multiple Testing Procedures with Applications to Genomics. S. Dudoit and M. J. van der Laan (2008). New York: Springer Science+Business Media, LLC. ISBN: 978-0-387-49316-9. *Biometrical Journal* **2010**, *52*, 699–699. doi:10.1002/BIMJ.201000174.
46. Molnar, C. Interpretable Machine Learning A Guide for Making Black Box Models Explainable. Technical report, 2019.
47. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems; Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds. Curran Associates, Inc., 2017, Vol. 30.
48. Scapin, D.; Cisotto, G.; Gindullina, E.; Badia, L. Shapley Value as an Aid to Biomedical Machine Learning: a Heart Disease Dataset Analysis **2022**.
49. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* **2014**, *7*, 1247–1250. doi:10.5194/gmd-7-1247-2014.
50. Kuhn, M.; Johnson, K. Applied Predictive Modeling **2013**.