Face the Needle: Predicting risk of fear and fainting during blood donation through video analysis

Judita Rudokaite[1,2] ; Itir Onal Ertugrul[3]; L.L. Sharon Ong[1]; Mart P. Janssen[2]; Elisabeth Huis in 't Veld[1,2]

[1] Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, the Netherlands
[2] Donor Medicine Research, Sanquin Research, Amsterdam, the Netherlands
[3] Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands

***Abstract*—There are physiological, hormonal and psychological markers that occur early in a procedure involving needles. These so-called vasovagal reactions range from feeling nauseous, dizzy, to completely passing out. In an early stage, they are difficult to measure and self-report before it is too late to prevent them. This study aims to explore different features from regular video and thermal facial video recordings of blood donors in the waiting room, prior to a blood donation procedure, in order to assess to what extent it is possible to predict whether a donor will experience a low or high level of vasovagal reaction later on during the blood donation. The results showed that the best performance was achieved using pre-trained ResNet152 models with GRU on a continuous video stream, achieving an F1 of 0.69, a PR-AUC score of 0.81, and an MCC score of 0.56. This model also achieved a precision of 0.52, recall of 0.94, F1 score of 0.67, and MCC score of 0.42 on new, previously unseen mobile video data. Although the model requires further improvement, it outperforms self-reported vasovagal reaction scores and shows the potential to predict who is at risk of experiencing vasovagal reactions using facial video data.**

I. Introduction

Needle-related procedures are essential part of medical treatment, however, people who are afraid of needles during such procedures often experience so-called vasovagal reactions (VVR) including sweating, breathing rapidly, vomiting or fainting. Experiencing VVR symptoms often results in some people avoiding medical procedures, for example, noncomplying with vaccinations, refusing blood tests or dental care [1-2]. In bloodbanking, needle fear is one of the main reasons given by young people for not donating blood [3] or after experiencing VVR – not returning to donate [2]. However, ensuring a sufficient number of blood donors is crucial for every blood bank given that only around 5% to 8% of eligible people donate blood in Western countries [4].

One of the issues in preventing VVRs is that they are difficult to self-report and occur suddenly. Some vasovagal reactions are the result of excessive arousal and anticipated negative emotions which triggers unconsciousness underlying processes of an autonomic nervous system, which manifests in increased heart rate [5-6], sweating, nausea, pupillary dilation, hyperventilation [7-8], or sudden drop in blood pressure [5]. Although there are multiple patient and donor characteristics that are related to who is at risk of experiencing VVR besides needle fear, such as younger age, being female, having a lower BMI [9-11], the main interventions that have been applied to date such as muscle tension techniques or water ingestion are geared towards a blood donation setting, targeting donors who experience VVRs due to loss of blood, but not due to arousal or negative emotions [1-2; 11-12]. However, previous studies [13-15] showed that there are physiological, hormonal and psychophysiological markers that already occur automatically and unconsciously prior to or at a very early stage of blood-related procedure, peaking around the time of needle insertion. It is unlikely that the current interventions address these affective physiological processes and furthermore, interventions for e.g. regular blood draws and injections are also scarce.

One of the solutions for early preventions of VVRs could be to continuously monitor physical and psychological states using psychophysiological techniques. Many features such as heart rate, heart rate variability, respiratory signals, skin conductance, or even brain wave measures with EMG or EEG could be useful targets for detecting early signs of VVRs and tracking them over time [16-18]. However, any needle-related procedures are short, and using any additional devices such as electrodes, EEG caps or respiratory vests are not feasible in real-life scenarios. On the other hand, video recordings are readily available, cheap, contactless, easy to record, and also contain rich information about facial expressions [15, 19-21], head movements, eye-gaze directions, changes in facial colour that can be extracted from the face [22-23], and even potentially can help to monitor physical changes real-time such as heart-rate or respiratory signal [25-27]. Additionally, a non-invasive technique called Infrared Thermal Imaging (ITI) allows the measurement of minute local changes in the human body temperature that can be influenced by changes in sympathetic and parasympathetic activities [28-30]. Previous studies showed that both regular RGB videos and thermal recordings performed well in automatically detecting stress [29-30], pain response [31], or recognizing emotions in real-time [32], for example, by using extracted facial action units using machine learning (ML) [33].

979-8-3503-9494-8/24/$31.00

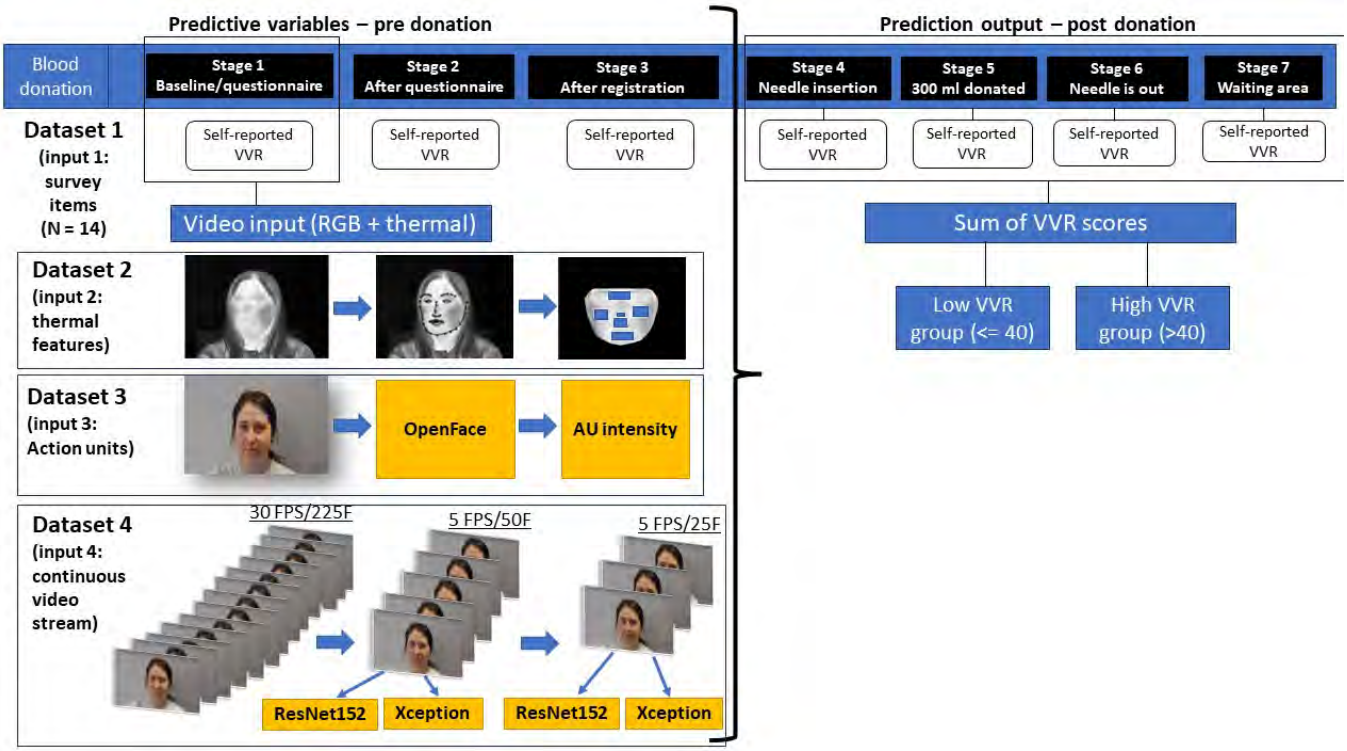

Fig. 1. Overview of the donation procedure, video and thermal data preprocessing steps.

Similarly, our study aims to apply similar ML techniques on various video streams extracted prior to needle-related procedure to assess whether it is possible to classify who is at risk of experiencing high or low VVR symptoms at the later stage of the donation. We aim to start from classical approach by extracting AU and other facial features, and then comparing to more advanced methods while incorporating temporal features (e.g. using LSTM or GRU). If successful, we aim to detect the earliest possible markers of VVRs, which could be implemented into biofeedback mechanism as a prevention strategy. So far, biofeedback training offered a promising avenue for treating anxiety and stress [34]. The main advantage is that such training based on visual facial information can be implemented in the mobile phone applications as shown in other examples [35, 36]. Therefore, this study is a first step to assess whether such solution is feasible using high-definition video data and whether it can be replicated using mobile video data.

II. Methods and Materials

A. Recruitment of participants

Participants were recruited from the regular blood donor

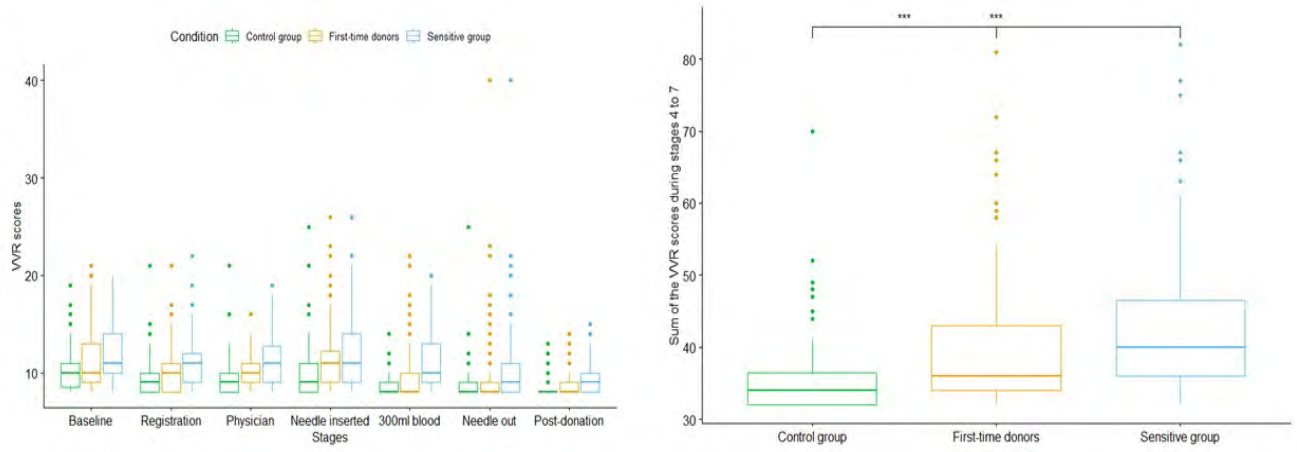

Fig. 2-A. Distribution of VVR ratings per donation stage and donor group. The dots above the box represent the outliers per group. The black line represents the mean VVR scores. **Fig. 2-B.** Variation in total VVR scores (during and after the blood donation, stages 4 - 7) per donor group. VVR symptoms consist of faintness, dizziness, weakness and lightheadedness, fear, stress, tension, and nervousness. The line in the box represents the mean of each donor group and the dots above the box represent the outliers per donor group.

pool from the not-for-profit organization responsible for the blood supply and distribution. The study was approved by the Research Ethics and Data Management Committee of the Tilburg university and the Ethical Advisory Board of the Sanquin. The study took place at four blood collection centers. All blood donors who fit into the following three groups were invited to participate: (1) control group; with between 5 to 10 previous donations, no previous experience of vasovagal reactions, (2) the sensitive group; with 5 to 10 previous donations but with the experience of a VVR at the previous donation, and (3) new donor group; first-time donors.

B. Procedure

Interested donors were contacted by the data manager for an appointment, and received information about the study, including ethical consent information. On arrival, participants completed a questionnaire containing items regarding needle fear and emotional states. This also served to stabilize the body temperature required for thermal imaging (T = 20-25 min). Then, the donors proceeded with the regular blood donation procedure, containing several distinct phases: registration, a health check at the donor physician, and blood donation. This resulted in seven distinct stages during which thermal, RGB videos and vasovagal reaction scores were recorded (fig. 1)

C. Materials and measures

Video and ITI recordings. The RGB video was recorded at 24 frames per second using the Nikon Coolpix AW130. Thermal video was recorded at 30 frames per second using a FLIR E95 camera with a thermal sensitivity of <40 mK at 30°C, an infrared resolution of 464 x 348 pixels. Both cameras were installed on a tripod at a distance of about 1m from the donor. Donors were free to behave as they normally would throughout the procedure.

Vasovagal reactions (based on the Blood Donation Reactions Inventory; [39]). At each of the seven stages, participants were asked to verbally rate to what extent they experienced physiological (faintness, dizziness, weakness, light-headedness) and emotional (fear, stress, tension, and nervousness) reactions, on the Likert scale from 1 (not at all) to 5 (extremely). The ratings of the last four stages (4-7) were summed, resulting in a score between 32 and 160. Then, we split the sample into a low VVR score (below the mean) and a high VVR score (above the mean). Since our sample is highly skewed, this his cut-off was selected to capture as many donors who may be at risk of experiencing VVR as possible.

D. Thermal video data preprocessing

The ITI data from the first stage prior to the blood donation (N frames = 1000) were preprocessed. For each frame, a visual representation of the face (.jpg) was exported as well as raw temperature values of each pixel. To estimate facial landmarks and track the face over time, we used the Face Alignment Network (FAN) [40]. The FAN received a thermal image file as input and produced the corresponding 2D landmarks and 2D projections of the 3D landmarks as outputs. Then the images were aligned in such a way that the features detected in one image would match the features in the following frame. Hence, all temperature values would be extracted from the same location. To achieve that, the frontal image was selected as a template and using the coordinates of the facial landmarks detected in each frame, a Warp Affine transformation technique was applied to warp each thermal image to fit this template. Next, we re-created each thermal image as a frontal one by pasting calculated triangles from our original image into our template image. The same procedure was completed for both a visual image and a raw temperature file. Finally, the following six regions of interest (ROI) were selected: nose, below the nose, cheeks, chin, and the area between the eyes from which the maximum temperature value at each frame for each participant was extracted (fig. 1).

The Tsfresh package [41] was used to extract the following 48 linear time series characteristics from each of the six facial areas such as the maximum, minimum, median, standard deviation, variance, mean, sum, and root mean square values.

E. Video data preprocessing

We used video data from the first stage, prior to the blood donation. Each video was shortened to 10 seconds and then 5 seconds from the beginning to test whether the length of the video would have an impact on the model performance. Redundant frames were eliminated by reducing the frame rate to 5 frames per second, with each frame having a resolution of 1920 x 1080 pixels (fig.1).

Each video frame served as an input that was passed to the pre-trained model, after it was resized to fit the default size of the models, specifically 299x299 for Xception [42] and 224x224 for ResNet152 [43]. Both pre-trained models returned vectors containing extracted features of size 2048, which were then used to train LSTM and GRU models.

F. Facial action unit extraction from video data

For extracting the intensity of the facial action units, we pre-processed video data from the first stage prior to the blood donation (N frames = 1000; fig.1). In particular, the intensity level of 17 action units at each frame were extracted using OpenFace [22]: AU1 (raised inner brow), AU2 (raised outer brow), AU4 (lowered brow), AU5 (raised upper lid), AU6 (raised cheeks), AU7 (tightened eye lids), AU9 (wrinkled nose), AU10 (raised upper lip), AU12 (pulled lip corner), AU14 (dimples formed), AU15 (lowered lip corners), AU17 (raised chin), AU20 (stretched lips), AU23 (tightened lips), AU25 (lips apart), AU26 (jaw drop), AU45 (blink). AU intensity shows how intense the activity of the AU is, ranging from a minimal value of 0 to a maximum value of 5.

Then, using the Tsfresh python package [41], 6 features from each of the AUs (sum, variance, standard deviation, maximum-, mean-, and mean root square values) were extracted, resulting in a total number of extracted features of 102.

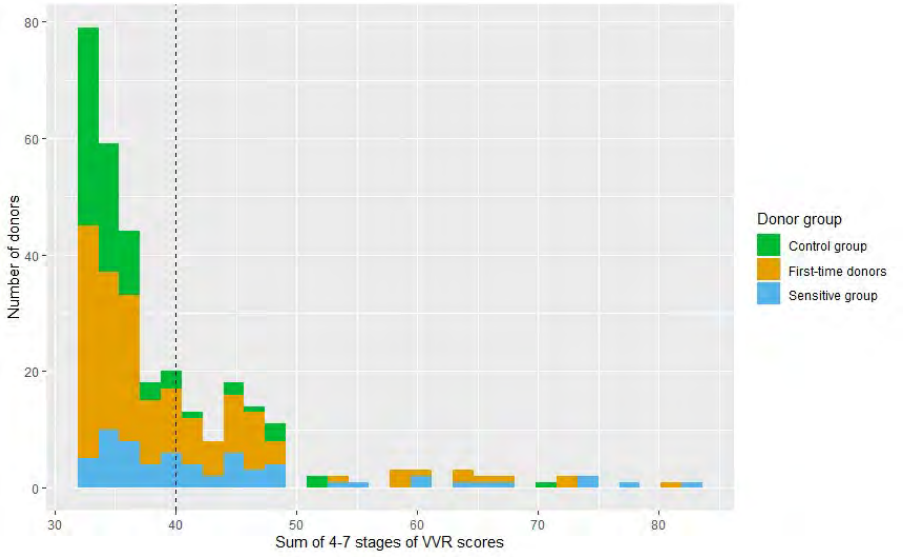

Fig. 3. Distribution of the VVR sumscores (over the post- donation stages 4-7). The black dashed line represents the mean-based cut-off of the sample on which the low vs high VVR groups were split.

G. Questionnaire data

In this study we accessed emotional and physiological states of the donors prior to the blood donation (fig. 1) using the following questionnaires. The scores were used as features for the ML models:

- Anxiety Sensitivity Index (ASI; [44]). a 16-item scale developed to measure 'fear of fear,' that is, the degree to which a person believes that physical symptoms of anxiety have negative consequences.
- Emotion Regulation Questionnaire (ERQ; [45]). A 10-item scale designed to measure respondents' tendency to regulate their emotions in two ways: (1) Cognitive Reappraisal and (2) Expressive Suppression. Cognitive reappraisal is a strategy that changes the emotional experiences and meaning of the situation without changing the situation objectively. Expressive Suppression is the strategy that intends to hide or reduce ongoing emotions and emotion-expressive behaviour.
- Somatosensory Amplification Scale [46], a 10-item scale designed to measure the tendency to detect somatic and visceral sensations and experience them as unusually intense or disturbing.
- Multidimensional Assessment of Interoceptive Awareness (MAIA (Version 2); [47]). A 37-item state-trait scale designed to measure multiple dimensions of interoception – the sense of the internal state of the body – by self-report. The MAIA consists of eight scales labelled as Noticing, Not-Distracting, Not-Worrying, Attention Regulation, Emotional Awareness, Self-Regulation, Body Listening, and Trust.

H. Model training, validation, and evaluation

Each dataset was split into a train (80%) and test (20%) set, on which the model performance was assessed.

For training machine learning models such as decision tree, random forest classifier, XGBoost [48] and artificial neural network, the input features were scaled using min-max normalization. Then, due to the class imbalance (high class = 28% in total dataset), Synthetic Minority Oversampling Technique (SMOTE) [49] was applied to the training set. We used a nested k-fold cross-validation with an outer k value of 10 and an inner k value of 3. The inner loop was used for feature selection using the Recursive Feature Elimination with cross-validation (RFECV; [50]) and hyperparameter tuning using GridSearchCV [51]. The outer loop was used for error estimation i.e. how well our classification algorithm performs.

Prior to training the LSTM and GRU models, the data was augmented to address the imbalance issue (high class = 28% in total) by applying horizontal flip and adding some noise to the minority class videos, in the training set only (Video Augmentation Library; [52]). After splitting the data into training and test sets, a validation split to automatically reserve the fraction of the training data was used to evaluate the loss and model metrics at the end of each epoch. Twenty percent of the computed data was selected by taking the last 20% of the samples of the video sequences received by the model. The architecture of both the GRU and LSTM consisted of two GRU or LSTM layers and a dropout layer as it previously showed yielding best results [53]. Adam was chosen as the optimizer, with a learning rate of 0.0001, as this is computationally efficient and suitable for a model with many parameters [54]. The selected activation function was a sigmoid that produces a number between zero and one, where everything below 0.5 is classified as negative and above as positive. The binary cross entropy was specified as a loss function where the target of predictions is zero or one and is using the sigmoid as the activation function for making these predictions [55]. The batch size and number of epochs were tested empirically.

As our baseline model, we used the self-reported pre-donation VVR scores from stage 1 and stage 2 as model input.

All described models were evaluated on the following metrics:

- Precision – a ratio of positive predictions (donors with high VVR) that are actually correct.
- Recall – the ratio of actual positives (donors with high VVR) that were predicted correctly.
- F1 score, which is the harmonic mean of precision and recall.
- AUC-PR score, which is the Area Under the Precision-Recall Curve that summarizes a precision-recall curve as the weighted mean of precisions over all recall values.
- Matthews correlation coefficient (MCC) that is a balanced method to measure whether there is a high agreement between predicted and actual values. The MCC ranges from -1 (total disagreement) to 1 (perfect agreement) with 0 showing that prediction is no better than chance.

Lastly, to evaluate which regions in the face are important for prediction of the best performing classification model, model performances after parts of the face (e.g. eye region) are occluded were assessed [56]. If model performance drops after the occlusion is applied, that region is considered to be important.

III. Results

A. Participants

The data was collected from N=310 blood donors of which n = 83 were in the control group, n = 63 in the sensitive group, and n = 164 were new donors. No significant gender ($F(2)$ = 2.33, p =.1) or location ($F(2)$ = 0.33, p =.6) differences were found between the groups.

B. Vasovagal reaction levels

The VVR score distribution was positively skewed with M = 39.40, SD = 9.31, median = 36, reflecting a higher proportion of blood donors who reported low VVR scores (see Fig. 2A, Fig. 3). The sample was split on the mean, into a low VVR score group (VVR score <= 40) and a high VVR score group (level > 40, see Fig. 3).

A one-way ANOVA showed a statistically significant main effect of the group on the total VVR symptoms during stages 4-7 ($F(2)$ = 15.08, p <.001). The control group experienced significantly lower levels of VVR levels than the first-time donors (p < .01), who in turn had lower VVR scores than the donors in the sensitive group (p <.001; see Fig. 2B).

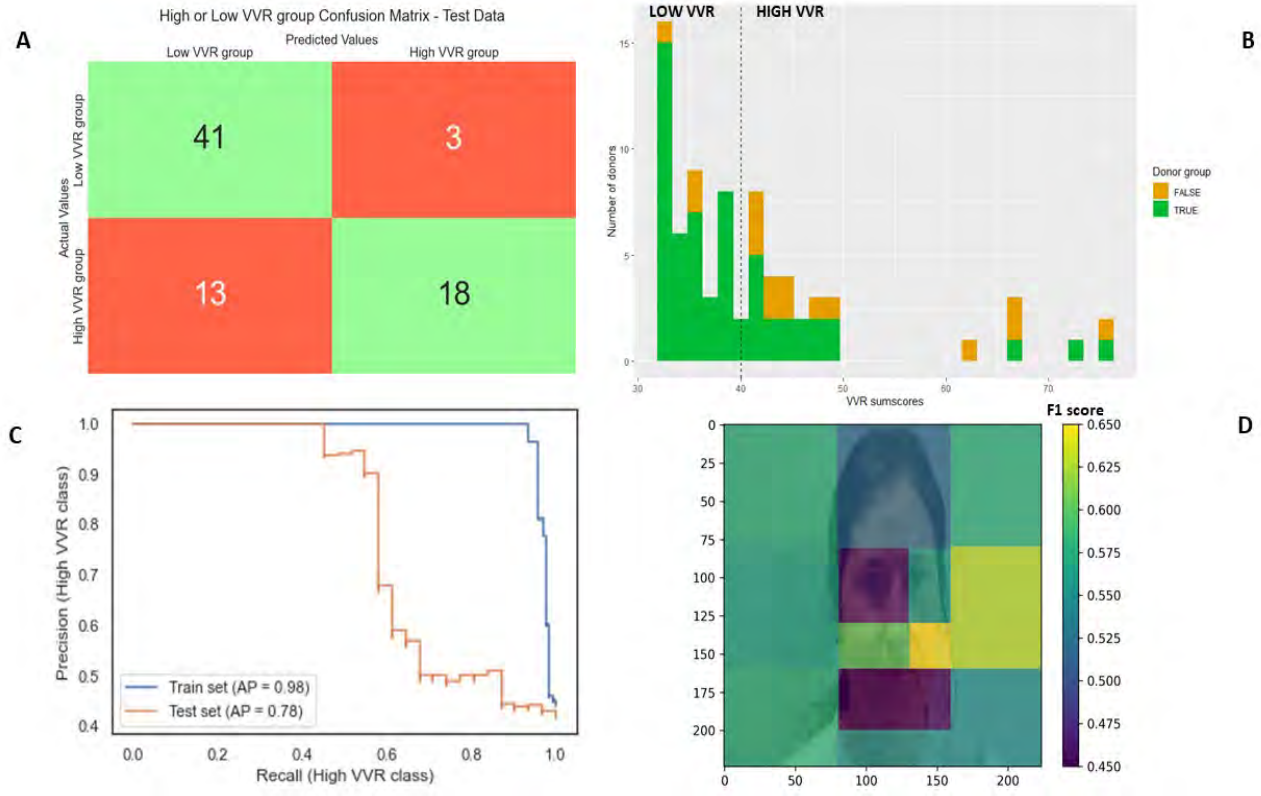

Fig. 4A. Confusion matrix showing correctly and incorrectly classified low and high VVR samples on the test set using a pre-trained model on ResNet152 with GRU on a 25 frames video sequence. **Fig. 4B**. Correctly and incorrectly classified low and high VVR samples on the test set using the pre-trained model on ResNet152 with GRU on a 25 frames video sequence on the original VVR sumscore scale. The dashed line separates low (on the left side) and high (on the right side) VVR groups. **Fig. 4C**. The precision-recall curve on train and test sets using the pre-trained ResNet152 model with GRU on 25 frames video sequence. **Fig. 4D.** To evaluate which parts of the image are important for classification, some regions were occluded in the image (see Ertugrul et al., 2020 [56]): the more F1 score drops (darker shades) after the occlusion, the more important the region is for the classification. We applied larger black rectangles (100x100) around the face and smaller rectangles (80x80 and 60x60) within the face region.

C. *Results of classifying low and high vasovagal reaction levels*

Table 1 shows the performance of the algorithms on each dataset. Using facial action units (MCC = 0.19; F1 = 0.52) and thermal imaging (MCC = 0.30; F1 = 0.50) resulted in the lowest performance, whereas the best performance was achieved by using continuous video data (MCC = 0.56; F1 = 0.69; see Fig. 4-C for the precision-recall curve).

We used the 2D-CNN pre trained on ResNet152 with GRU on 25 frames as our best-performing model to explore correctly and incorrectly classified samples (fig. 4A-B). Fig. 4A-B shows that the model correctly classified almost all samples in the low VVR class, but made the majority of errors in correctly predicting high VVR samples around the split. Furthermore, the results showed that the regions around the eyes, lips, chin, and forehead are important for the performance of the model (Fig. 4-D).

Table I. Machine learning performance values on the testing set for binary classification (high VVR group = positive class).

Input data type	Model (Number of features)	Precision	Recall	F1	PR-AUC	MCC
Pre-donation VVR scores (N low = 51, N high = 13)	XGboost (N = 2)	0.43	0.69	0.53	0.56	0.39
	Decision tree (N = 2)	0.47	0.69	0.56	0.57	0.44
	Random forest (N = 2)	0.47	0.69	0.56	0.67	0.44
	Artificial neural networks (N = 2)	0.47	0.69	0.56	0.68	0.44
Questionnaire data (N low = 46, N high = 18)	XGboost (N = 14)	0.53	0.50	0.51	0.49	0.33
	Decision tree (N = 14)	0.52	0.67	0.59	0.70	0.40
	Random forest (N = 14)	0.47	0.50	0.49	0.57	0.28
	Artificial neural networks (N = 14)	0.50	0.72	0.59	0.45	0.40
Action units (N low = 47, N high = 16)	XGboost (N = 102)	0.31	0.69	0.43	0.29	0.15
	Decision tree (N = 102)	0.31	0.50	0.38	0.33	0.10
	Random forest (N = 102)	0.30	0.62	0.41	0.31	0.11
	Artificial neural networks (N = 74)	0.47	0.44	0.45	0.30	0.27
Thermal files (N low = 47, N high = 16)	XGboost (N = 48)	0.39	0.56	0.46	0.43	0.24
	Decision tree (N = 48)	0.39	0.56	0.46	0.53	0.24
	Random forest (N = 48)	0.45	0.56	0.50	0.50	0.31
	Artificial neural networks (N = 48)	0.50	0.62	0.56	0.50	0.39
Continuous video data (N low = 44, N high = 31)	2D CNN (Xception) with GRU (Number of frames = 50)	0.66	0.61	0.63	0.71	0.39
	2D CNN (Xception) with LSTM (Number of frames = 50)	0.66	0.68	0.67	0.73	0.43
	2D CNN (ResNet152) with GRU (Number of frames = 50)	0.79	0.48	0.60	0.74	0.44
	2D CNN (ResNet152) with LSTM (Number of frames = 50)	0.75	0.58	0.65	0.75	0.47
	2D CNN (Xception) with GRU (Number of frames = 25)	1	0.39	0.56	0.82	0.52
	2D CNN (Xception) with LSTM (Number of frames = 25)	0.76	0.52	0.62	0.74	0.44
	2D CNN (ResNet152) with GRU (Number of frames = 25)	**0.86**	**0.58**	**0.69**	**0.81**	**0.56**
	2D CNN (ResNet152) with LSTM (Number of frames = 25)	0.68	0.61	0.64	0.73	0.42

D. *Cross-domain evaluation on videos from mobile phones*

The results show that it is possible to predict vasovagal reactions occurring during the blood donation from high resolution continuous video data collected prior to the blood donation. However, to test the cross-domain performance, we aimed to replicate our findings on a new, previously unseen dataset with different characteristics in terms of context and range of head pose. We used *mobile* video data obtained prior to a video-based blood donation experiment based on the rubber arm experiment [57]. In the experimental condition, the illusion of ownership of the arm seen on the screen was induced and in the control condition, it was not. For a description of this procedure, see [58; 59]). The sample consisted of N = 47 (n = 42 female, and after discarding of the recordings of n = 3 participants due to technical errors) from the university in return for course credit. The participants were informed about what would happen during the virtual donation and then asked to use a mobile phone for 5 minutes, that recorded their face in the background. There were no significant differences in demographic characteristics such as gender, age, or BMI between participants in the experimental and the control group nor between participants who self-reported suffering from needle fear versus those who did not.

After the experiment, participants rated the level of VVR they experienced during the virtual blood donation using the rating scale described before. Also in this experiment, VVR scores were positively skewed with a higher proportion of participants reporting low VVR scores (M = 15.87, SD = 8.17, median = 13.0; min = 8, max = 46; see figure 5). The videos were split into two groups on the mean score, representing video segments from participants who experienced low levels of VVR (N videos= 29, VVR score <=16 and high levels of VVR (N videos = 18, VVR level > 16). We included 25 frames of all videos obtained during this experiment in the test set and then applied the previously developed pre-trained ResNet152 and Xception with GRU and LSTM classification models. The results are reported in Table II.

Table II. The performance of the pre-trained ResNet152 and Xception models with GRU and LSTM on the new dataset obtained during the virtual blood donation (N frames = 25).

Model	VVR group	Precision	Recall	F1	MCC
ResNet152 with GRU	**Low (N =29)**	**0.93**	**0.45**	**0.60**	**0.42**
	High (N=18)	**0.52**	**0.94**	**0.67**	
ResNet152 with LSTM	Low (N =29)	1	0.03	0.07	0.12
	High (N =18)	0.39	1	0.56	
Xception with GRU	Low (N =29)	0.61	0.38	0.47	0.009
	High (N =18)	0.38	0.61	0.47	
Xception with LSTM	Low (N =29)	0.77	0.34	0.48	0.19
	High (N =18)	0.44	0.83	0.58	

Table II shows that all models performed slightly better in classifying high than low VVR groups. In addition, models with GRU performed better than LSTM in classifying both low and high VVR groups with the most balanced and highest performance achieved using ResNet152 with GRU with precision of 0.52, recall of 0.94, F1 score of 0.67, and MCC score of 0.42. Thus, we further completed the error analysis using pre-trained ResNet152 with GRU (fig. 5) to identify where the model makes the mistakes in predicting vasovagal reactions.

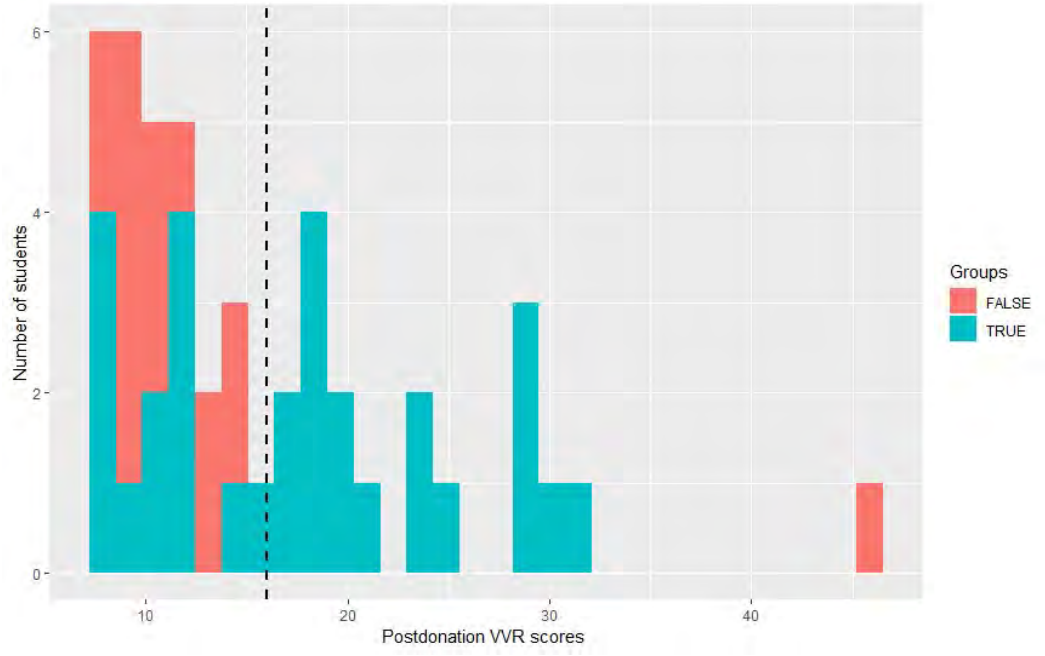

Fig. 5. Figure shows correctly (blue color) and incorrectly (red color) classified samples on the test set using 2D CNN pre-trained model on ResNet152 with GRU on 25 frames video sequence. The dashed black line separates low (on the left side) and high (on the right side) VVR groups.

IV. Discussion

In this study we aimed to assess whether it is possible to predict adverse emotional and physical reactions that may occur during blood donation from facial information prior to the donation, what type of facial features are important, and whether this data is more informative than subjective self-reports or personality factors.

Related work has shown that self-reported anxiety and fear are risk factors for experiencing VVR [2, 9-11, 60]. Similarly, our study showed that using the self-reported pre-donation VVR scores already provide a relatively good performance, with moderate agreement between predicted and actual values (MCC = 0.44, F1 score = 0.56). Classification performance using the personality questionnaire scores also resulted in a similar performance (F1 score = 0.59, MCC = 0.40). However, the overall best performing model in our study used the pre-trained ResNet152 features to classify low and high classes with GRU on 25 frames of a digital video recording, resulting in an MCC score of 0.56, and a balanced performance with a recall score of 0.58 and a precision of 0.86. Additionally, the results indicate that the eyes, lips, chin, and forehead areas are important for classification, which is in line with previous findings that the eye region and movements of the brow play an important role in classifying low and high VVR groups [15]. Movements of the brow, chin, lips, and even jaw have previously been shown to be indicative of stress, [20], and cues such as closing the eyes, lowering the brow, tightening the eyelids, or parting lips play a major role in the detection of pain [21].

We expected that ITI data would also reflect the underlying emotional and physiological reactions, especially as previous studies show that thermal imaging is less influenced by the pose variation or lightening [61] and that stronger distress can be associated with a decrease in nasal temperature [62-64]. Also, ITI can reflect breathing patterns and pick up signs of hyperventilation, which is associated with anxiety and an increased risk of vasovagal reactions [58; 59]. However, the performance using thermal imaging was not better than the baseline model using self-reported experience of VVR. Using ITI for this purpose is novel, and the role of individual differences is yet unclear [30]. In addition, people often show increased head movements when in distress [21], which can also affect the results since facial action units and thermal imaging are affected by rigid and non-rigid facial motions [65]. For example, in our study some blood donors tend to look away during the needle insertion. However, for clinical applications it is a positive finding that 'regular' facial video data using an affordable camera is a better target for predicting vasovagal reactions, achieving the overall best performance using a continuous video stream with GRU. This is corroborated by similar studies focusing on pain detection among others, indicating that measuring pain on a frame-by-frame basis and taking into account spatial-temporal features yield much better results than using static images, overcoming many challenges related to capturing spontaneous emotions or head movements (e.g. [21]). In addition, facial videos have shown to be a good target for measuring emotional states and physiological reactions such as heart-rate [24-26] or respiratory signals [66-67]. This might be also a promising avenue to explore in the future.

For clinical versus research applications it is also important to consider which model best fits the purpose. For the AI driven biofeedback solution for example, that aims to detect early signs of VVR in real-time, a model that is faster, uses less frames, is less computationally demanding and which has a high recall is preferred. In other words, it may be more important to be on the side of caution and to correctly classify all the individuals at risk for experiencing VVR at the expense of incorrectly capturing a few individuals who may actually be in the low VVR group. For other applications a higher precision or balance may be preferred, for example when the intervention is expensive or invasive, and for research purposes speed or computational considerations may not be an issue. We further validated whether these algorithms may work in a clinical setting by using a new dataset consisting of lower quality mobile phone video recordings. These were collected with the mobile app in mind, as the app uses input from the front facing camera, reflecting a real-life scenario where participants look to their smartphone screen. Taking these differences into account, the classification performance was relatively high with 0.52 precision, 0.94 recall, and 0.67 F1 score. This indicates that the model performs relatively well on a new, previously unseen dataset especially in capturing the group of people who experience higher levels of VVR. Prediction errors mostly occurred around the cut-off point. This indicates that the cut-off point may not be the most optimal choice for separating low and high VVR groups because the same mistakes were observed in training the model, and in the future it may beneficial to move away from dichotomous classification and instead use a continuous prediction. Additionally, we aim to find the best 'mix' of trait (e.g., a personality scale, self-identification as suffering from needle fear), state (e.g., a self-report rating), and video features.

One of the important limitations of our study is that our dataset consisted of a majority of blood donors who reported low VVR scores. In this study, only 8 blood donors scored above 70 points on the VVR scale. Although this is a reflection of the overall prevalence rates of VVR in blood banking, varying from 1-2% [1], other similar studies on anxiety and stress report an increase in model performance after including more subjects and more training samples. Therefore, acquiring more high VVR examples is preferred over relying on data augmentation and SMOTE techniques [49].

In conclusion, this is the first study showing that it is possible to classify vasovagal reactions from facial video data. Future work includes improving model performance and incorporating the solution into the biofeedback solution.

Acknowledgments

We thank Hasti Memarzadeh for collecting the data, Natalie de Wit and Laura Heij for her support in preprocessing the data, all blood donors who voluntarily participated in our study, and the staff at the participating blood collection centers for their hospitality in hosting this study.

References

[1] Malave, B., & Vrooman, B. (2022). Vasovagal Reactions during Interventional Pain Management Procedures—A Review of

Pathophysiology, Incidence, Risk Factors, Prevention, and Management. Medical Sciences, 10(3), 39.
[2] Thijsen, A., & Masser, B. (2019). Vasovagal reactions in blood donors: risks, prevention and management. Transfusion Medicine, 29, 13-22.
[3] Padilla-Garrido, N., Fernández-Herrera, M. D., Aguado-Correa, F., & Rabadán-Martín, I. (2021). Motivators, barriers and communication channels for blood donation in relation to students at a university in Spain. Transfusion and apheresis science, 60(6), 103270.
[4] Wevers, A., Wigboldus, D. H., De Kort, W. L., Van Baaren, R., & Veldhuizen, I. J. (2014). Characteristics of donors who do or do not return to give blood and barriers to their return. Blood transfusion, 12(Suppl 1), s37.
[5] Thayer, J. F., Åhs, F., Fredrikson, M., Sollers III, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neuroscience & Biobehavioral Reviews, 36(2), 747-756.
[6] Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1718), 20160206.
[7] Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems, 6(2), 156-166.
[8] McDuff, D. J., Hernandez, J., Gontarek, S., & Picard, R. W. (2016, May). Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4000-4004).
[9] France, C. R., France, J. L., Conatser, R., Lux, P., McCullough, J., & Erickson, Y. (2019). Predonation fears identify young donors at risk for vasovagal reactions. Transfusion, 59(9), 2870-2875.
[10] Labus, J. S., France, C. R., & Taylor, B. K. (2000). Vasovagal reactions in volunteer blood donors: Analyzing the predictive power of the medical fears survey. International Journal of Behavioral Medicine, 7, 62-72.
[11] Fisher, S. A., Allen, D., Doree, C., Naylor, J., Di Angelantonio, E., & Roberts, D. J. (2016). Interventions to reduce vasovagal reactions in blood donors: a systematic review and meta-analysis. Transfusion Medicine, 26(1), 15-33.
[12] Morand, C., N. Coudurier, C. Rolland, S. Thoret, D. Legrand, P. Tiberghien & J. Bosson (2016). Prevention of syncopal-type reactions after whole blood donation: a cluster-randomized trial assessing hydration and muscle tension exercise. Tranfusion 56, 2412-2421.
[13] Hoogerwerf, M.D.; Veldhuizen, I.J.T.; Merz, E.M.; De Kort, W.L.; Frings-Dresen, M.H.; Sluiter, J.K. Psychological and hormonal stress response patterns during a blood donation. Vox Sang. 2017, 112, 733–743.
[14] Hoogerwerf, M. D., Veldhuizen, I. J. T., Tarvainen, M. P., Merz, E. M., Huis in ‘t Veld, E. M. J., de Kort, W. L. A. M., ... & Frings-Dresen, M. H. W. (2018). Physiological stress response patterns during a blood donation. Vox sanguinis, 113(4), 357-367.
[15] Rudokaite, J., Ertugrul, I. O., Ong, S., Janssen, M. P., & Huis in ‘t Veld, E. (2023). Predicting Vasovagal Reactions to Needles from Facial Action Units. Journal of Clinical Medicine, 12(4), 1644.
[16] Sutarto, A. P., Wahab, M. N. A., & Zin, N. M. (2010). Heart Rate Variability (HRV) biofeedback: A new training approach for operator’s performance enhancement. Journal of industrial engineering and management, 3(1), 176-198.
[17] Marzbani, H., Marateb, H. R., & Mansourian, M. (2016). Neurofeedback: a comprehensive review on system design, methodology and clinical applications. Basic and clinical neuroscience, 7(2), 143.
[18] Blum, J., Rockstroh, C., & Göritz, A. S. (2020). Development and pilot test of a virtual reality respiratory biofeedback approach. Applied Psychophysiology and Biofeedback, 45, 153-163.
[19] Gavrilescu, M., & Vizireanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system. Sensors, 19(17), 3693.
[20] Giannakakis, G., Koujan, M. R., Roussos, A., & Marias, K. (2022). Automatic stress analysis from facial videos based on deep facial action units recognition. Pattern Analysis and Applications, 1-15.
[21] Lucey, P., Cohn, J. F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., & Prkachin, K. M. (2010). Automatically detecting pain in video through facial action units. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 41(3), 664-674.
[22] Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018, May). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 59-66). IEEE.
[23] Jaiswal, S., & Valstar, M. (2016, March). Deep learning the dynamic appearance and shape of facial action units. In 2016 IEEE winter conference on applications of computer vision (WACV) (pp. 1-8). IEEE.
[24] Hassan, M. A., Malik, A. S., Fofi, D., Saad, N., Karasfi, B., Ali, Y. S., & Meriaudeau, F. (2017). Heart rate estimation using facial video: A review. Biomedical Signal Processing and Control, 38, 346-360.
[25] Bousefsaf, F., Pruski, A., & Maaoui, C. (2019). 3D convolutional neural networks for remote pulse rate measurement and mapping from facial video. Applied Sciences, 9(20), 4364.
[26] Rouast, P. V., Adam, M. T., Chiong, R., Cornforth, D., & Lux, E. (2018). Remote heart rate measurement using low-cost RGB face video: a technical literature review. Frontiers of Computer Science, 12, 858-872.
[27] Li, X., Chen, J., Zhao, G., & Pietikainen, M. (2014). Remote heart rate measurement from face videos under realistic situations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4264-4271).
[28] Ioannou, S., Gallese, V., & Merla, A. (2014). Thermal infrared imaging in psychophysiology: potentialities and limits. Psychophysiology, 51(10), 951-963.
[29] Garbey, M., Sun, N., Merla, A., & Pavlidis, I. (2007). Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. IEEE transactions on Biomedical Engineering, 54(8), 1418-1426.
[30] Engert, V., Merla, A., Grant, J. A., Cardone, D., Tusche, A., & Singer, T. (2014). Exploring the use of thermal infrared imaging in human stress research. PloS one, 9(3), e90782.
[31] Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., ... & Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. Biomedical Signal Processing and Control, 31, 89-101.
[32] Cruz-Albarran, I. A., Rodriguez-Medina, D. A., Leija-Alva, G., Dominguez-Trejo, B., Osornio-Rios, R. A., & Morales-Hernandez, L. A. (2020). Physiological stressor impact on peripheral facial temperature, Il-6 and mean arterial pressure, in young people. Journal of Thermal Biology, 91, 102616.
[33] Werner, P., Lopez-Martinez, D., Walter, S., Al-Hamadi, A., Gruss, S., & Picard, R. W. (2019). Automatic recognition methods supporting pain assessment: A survey. IEEE Transactions on Affective Computing, 13(1), 530-552.
[34] Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked, 20, 100372.
[35] Ge X, Jose J M, Wang P, et al. ALGRNet: Multi-Relational Adaptive Facial Action Unit Modelling for Face Representation and Relevant Recognitions. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2023.
[36] Alneyadi, M., Drissi, N., Almeqbaali, M., & Ouhbi, S. (2021). Biofeedback-based connected mental health interventions for anxiety: Systematic literature review. JMIR mHealth and uHealth, 9(4), e26038.
[37] Almeqbaali, M., Ouhbi, S., Serhani, M. A., Amiri, L., Jan, R. K., Zaki, N., ... & Almheiri, E. (2022). A Biofeedback-Based Mobile App With Serious Games for Young Adults With Anxiety in the United Arab Emirates: Development and Usability Study. JMIR Serious Games, 10(3), e36936.
[38] Hunter, J. F., Olah, M. S., Williams, A. L., Parks, A. C., & Pressman, S. D. (2019). Effect of brief biofeedback via a smartphone app on stress recovery: randomized experimental study. JMIR Serious Games, 7(4), e15974.
[39] France, C. R., Ditto, B., France, J. L., & Himawan, L. K. (2008). Psychometric properties of the Blood Donation Reactions Inventory: a subjective measure of presyncopal reactions to blood donation. Transfusion, 48(9), 1820-1826.
[40] Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In Proceedings of the IEEE international conference on computer vision (pp. 1021-1030).
[41] Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). Neurocomputing, 307, 72-77.
[42] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[43] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[44] Reiss, S., Peterson, R. A., Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. Behaviour research and therapy, 24(1), 1-8.
[45] Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. Journal of personality and social psychology, 85(2), 348.
[46] Barsky, A. J., Wyshak, G., & Klerman, G. L. (1990). The somatosensory amplification scale and its relationship to hypochondriasis. Journal of psychiatric research, 24(4), 323-334.
[47] Mehling, W. E., Acree, M., Stewart, A., Silas, J., & Jones, A. (2018). The multidimensional assessment of interoceptive awareness, version 2 (MAIA-2). PloS one, 13(12), e0208034.
[48] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.
[49] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.
[50] Misra, P., & Yadav, A. S. (2020). Improving the classification accuracy using recursive feature elimination with cross-validation. Int. J. Emerg. Technol, 11(3), 659-665.
[51] LaValle, S. M., Branicky, M. S., & Lindemann, S. R. (2004). On the relationship between classical grid search and probabilistic roadmaps. The International Journal of Robotics Research, 23(7–8), 673–692.
[52] Cauli, N., & Reforgiato Recupero, D. (2022). Survey on videos data augmentation for deep learning models. Future Internet, 14(3), 93.
[53] Kirori, Z., & Ireri, E. (2020). Towards Optimization of the Gated Recurrent Unit (GRU) for Regression Modeling. Int. J. Soc. Sci. Inf. Technol, 157-167.
[54] Kingma, D. and Ba, J. Adam: A method for stochastic optimization. In ICLR, 2015.
[55] Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. Int. J. Adv. Trends Comput. Sci. Eng, 9(10).
[56] Ertugrul, I. O., Cohn, J. F., Jeni, L. A., Zhang, Z., Yin, L., & Ji, Q. (2020). Crossing domains for au coding: Perspectives, approaches, and measures. IEEE transactions on biometrics, behavior, and identity science, 2(2), 158-171.
[57] Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel'touch that eyes see. Nature, 391(6669), 756-756.
[58] Trost, Z., Jones, A., Guck, A., Vervoort, T., Kowalsky, J. M., & France, C. R. (2017). Initial validation of a virtual blood draw exposure paradigm for fear of blood and needles. Journal of Anxiety Disorders, 51, 65-71.
[59] Rudokaite, J., Ong, L. L. S., Janssen, M. P., Postma, E., & Huis In't Veld, E. (2022). Predicting vasovagal reactions to a virtual blood donation using facial image analysis. Transfusion, 62(4), 838-847.
[60] Ditto, B., Gilchrist, P. T., & Holly, C. D. (2012). Fear-related predictors of vasovagal symptoms during blood donation: it's in the blood. Journal of Behavioral Medicine, 35, 393-399.
[61] Cardone, D., & Merla, A. (2017). New frontiers for applications of thermal infrared imaging devices: Computational psychopshysiology in the neurosciences. Sensors, 17(5), 1042.
[62] Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. H., Olivier, B., & Kalkman, C. J. (2013). The effect of stress on core and peripheral body temperature in humans. Stress, 16(5), 520-530.
[63] Shastri, D., Papadakis, M., Tsiamyrtzis, P., Bass, B., & Pavlidis, I. (2012). Perinasal imaging of physiological stress and its affective potential. IEEE Transactions on Affective Computing, 3(3), 366-378.
[64] Zhi, R., Liu, M., & Zhang, D. (2020). A comprehensive survey on automatic facial action unit analysis. The Visual Computer, 36, 1067-1093.
[65] Yu, Z., Li, X., & Zhao, G. (2021). Facial-video-based physiological signal measurement: Recent advances and affective applications. IEEE Signal Processing Magazine, 38(6), 50-58.
[66] Liu, H., Allen, J., Zheng, D., & Chen, F. (2019). Recent development of respiratory rate measurement technologies. Physiological measurement, 40(7), 07.