Deep Learning-Based Classification of Social Anxiety Disorder Using Continuous Self-Reported Anxiety in Virtual Reality

1st Marco Pardini
[1]*University of Pisa*
Pisa, Italy
marco.pardini@phd.unipi.it

2nd Sergio Frumento
[1]*University of Pisa*
Pisa, Italy
sergio.frumento@med.unipi.it

3rd Matteo Martini
[2]*University of Genoa*
Genoa, Italy
matteo.martini@edu.unige.it

4th Gianluca Rho
[1]*University of Pisa*
Pisa, Italy
gianluca.rho@phd.unipi.it

5th Valerio Vatteroni
[1]*University of Pisa*
Pisa, Italy
v.vatteroni2@studenti.unipi.it

6th Krishant Tharun
[2]*University of Genoa*
Genoa, Italy
5168143@studenti.unige.it

7th Martina Alaimo
[4]*University of Pisa*
Pisa, Italy
m.alaimo4@studenti.unipi.it

8th Federico A. Galatolo
[1]*University of Pisa*
Pisa, Italy
federico.galatolo@unipi.it

9th Martina De Marinis
[1]*University of Pisa*
Pisa, Italy
martina.demarinis@phd.unipi.it

10th Enzo Pasquale Scilingo
[1,3]*University of Pisa*
Pisa, Italy
enzo.scilingo@unipi.it

11th Danilo Menicucci
[4]*University of Pisa*
Pisa, Italy
danilo.menicucci@unipi.it

12th Mario G.C.A. Cimino
[1]*University of Pisa*
Pisa, Italy
mario.cimino@unipi.it

13th Manuela Chessa
[2]*University of Genoa*
Genoa, Italy
manuela.chessa@unige.it

14th Alberto Greco
[1]*University of Pisa*
Pisa, Italy
alberto.greco@unipi.it

***Abstract*—Social Anxiety Disorder (SAD) is a mental disorder characterized by excessive fear and avoidance of social situations. Traditional assessment methods rely on retrospective self-reports, which may not fully capture moment-to-moment variations in perceived anxiety. To address this, we designed a novel Virtual Reality (VR) scenario to simulate a real-life social situation, specifically a waiting room that gradually fills with virtual characters. A continuous measure of self-reported anxiety was collected via joystick throughout the VR experience, allowing for real-time monitoring of subjective social anxiety. A one-dimensional convolutional neural network (1D-CNN) was trained to classify individuals with SAD based on their reported anxiety trajectories. The model was evaluated using a Leave-One-Subject-Out (LOSO) cross-validation strategy, achieving an F1-score of 0.82, recall of 0.89, and precision of 0.77, demonstrating strong classification performance. These findings suggest that self-reported anxiety alone is a viable signal for distinguishing individuals with SAD, paving the way for more accessible, sensor-free assessment tools in virtual environments. Future work will explore advanced feature extraction from the anxiety signal, integrate physiological markers, and investigate adaptive VR scenarios that dynamically respond to user-reported distress.**

***Index Terms*—Social Anxiety Disorder, Virtual Reality, 1D-CNN, Leave-One-Subject-Out, Anxiety Classification.**

I. Introduction

Social Anxiety Disorder (SAD) is a mental issue characterized by an exaggerated fear for social situations, in particular those in which the individual is exposed to possible scrutiny by other people [1]. This disorder gained new relevance after 2020, when pandemic restrictions 1) induced a sense of guilt and anxiety at social gatherings and 2) made it socially acceptable for adolescents to avoid exposure to social situations (e.g., school, parties) that typically soften SAD symptoms to subclinical levels: this resulted in an impressive 25.6% increase in anxiety disorders [2] that significantly affected the prevalence of SAD too, especially among women and low income earners [3].

[1] Department of Information Engineering
[2] Department of Informatics, Bioengineering, Robotics, and Systems Engineering
[3] Research Center "E. Piaggio"
[4] Department of Surgical, Medical, Molecular and Critical area pathology

The research leading to these results has received partial funding from the Italian Ministry of Education and Research (MIUR) in the framework of projects ForeLab Project (Departments of Excellence), and has been supported by projects PRIN2022 BRAVE (funded by Italian Minister of University MUR with project code n. 2022PTSX4L) and PNRR—M4C2-Investimento 1.3, Partenariato Esteso PE00000013 — 'FAIR - Future Artificial Intelligence Research' - Spoke 1 'Human-centered AI', funded by the European Commission under the NextGeneration EU programme. No conflicts of interest are reported for this study. We thank the Clinical Psychologist Sara Said and the Biomedical Engineer Arianna Galigani for their contribution in collecting part of the data analyzed in the present paper.

The key symptom of SAD – as well as other anxiety disorders [4] – consists of avoiding the feared situation, a behavior that brings immediate relief but in the long term fosters a vicious circle. On the other hand, exposure to feared situations represents the elective and effective treatment for SAD [5]. This centrality of generating controlled anxiogenic stimuli to induce social anxiety and, thus, reduce SAD symptoms motivates more and more researchers to set up virtual scenarios to immerse patients in social situations, as Virtual Reality (VR) offers a more rigorous and standardized exposure maintaining therapeutic effects comparable to those of in-vivo protocols [6].

However, most of these scenarios reproduce social situations that are irrealistic (e.g., a Freud-like therapist that splits like a cell and then morphs in other people [7]), too aversive to be acceptable for the most severe patients (e.g., reproducing particularly intense social situations such as a job interview [8], a party [9], a concert stage [7]), or focused on public speaking only (e.g., [10], [11]). Importantly, subjective anxiety was typically assessed only before and after the experimental session, thus missing all information about how this feeling possibly changed during the whole exposure.

To overcome these limitations, we recently proposed a virtual scenario representing an everyday-life social situation (i.e., a waiting room) whose online anxiety assessment [12] allows deep learning signal analysis due to the continuous nature of the signal itself. Indeed, while more objective (e.g., psychophysiological) correlates have some advantages (e.g., allowing a continuous data recording), self-reports are nevertheless a direct assessment of the most relevant symptom – i.e., the subjective feeling of anxiety [13]. On the other side, questionnaires alone cannot grasp the complexity of a mental disorder by merely asking patients to rate their behavior in a certain situation (e.g., for a fail in introspection [14]): that's why the assessment of anxiety disorders significantly benefits from an immersion of patients in the feared situation [4] and a continuous real time feedback of their psychological status.

This all premised, in the present study volunteers with more-or-less SAD symptoms were immersed in a virtual scenario aimed at inducing an acceptable but meaningful level of social anxiety: their self-reported discomfort registered using the joystick was analyzed through a deep learning approach involving one-dimensional convolutional neural network (1D-CNN). With respect to the previous literature, this study allowed the classification of each participant based on the anxiety felt along the whole virtual exposure: this online assessment conveys more information than the pre-post questions typically used (e.g., [15], [11]) to measure subjective anxiety, allowing a potentially better characterization of participants. In future applications of the same protocol, this data will add also psychophysiological correlates to obtain the most complete characterization of Social Anxiety Disorder.

II. Methods

Experimental sample

In accordance with the bioethical committee of the University of Pisa (protocol 0017548/2024 released on 02/12/2024), participants were recruited through online advertisements and preliminarily assessed for the presence of inclusion criteria – through the Liebowitz Social Anxiety Scale (L-SAS [16]) – and for the absence of other psychopathologies – through the Symptoms Check-List 90 revised (SCL-90-R [17]) – representing potential confounding factors.

After this preliminary screening, a total of 63 volunteers took part in the experimental session. Each participant was assigned to either the anxious (L-SAS score $\geq$55) or the control (L-SAS score $<$55) group.

Experimental procedure

Participants were welcomed and invited to read and sign the informed consent, reassuring them about the possibility to interrupt the experimental procedure in any moment without giving explanation: in practice, they could exit the virtual environment by pressing two controller's buttons simultaneously, in order to guarantee the tolerability of anxiogenic exposure. They were then made to wear a virtual reality headset (Meta Quest Oculus 3) connected via USB-C cable to a Windows workstation running the Unity3D engine (Figure 2). Few preliminary steps allowed each participant to 1) choose the virtual avatar (male or female) better representing the gender in which they identified, 2) get practiced with the green bar whose height was adjusted in real-time by the controller's joystick to express subjective anxiety in any moment, and 3) get used to the virtual environment by a first immersion in the empty waiting room. After that, the volunteer was immersed in the same scenario, which is initially empty (for 1 minute) and is then progressively filled by more and more virtual avatars along 9 minutes (details about the sitting order, the features of the environment and of non-player characters can be found in the scenario complete description [12]). New virtual avatars entered at jittered intervals until the room was totally filled, and then started soft verbal (e.g., complaining about waiting timings) and non-verbal (e.g., eye contact with the volunteer) interactions: after 10 minutes from the virtual immersion, the environment faded away and the participant was invited to take off the headset. The task was made passive to avoid potential confounders, as the only instruction was to use the controller's joystick to express our own anxiety level during the whole experiment.

Architecture

Deep learning has proven to have high potential in 1D signal analysis [18] [19] [20]. The present study inspects a deep learning architecture, specifically a 1D-CNN.

As shown in Figure 1, the proposed deep learning model consists of four sequential blocks, each comprising a one-dimensional convolutional layer, a Sigmoid activation function and a Dropout layer. Following these convolutional blocks, the resulting tensor is reshaped into a one-dimensional vector,

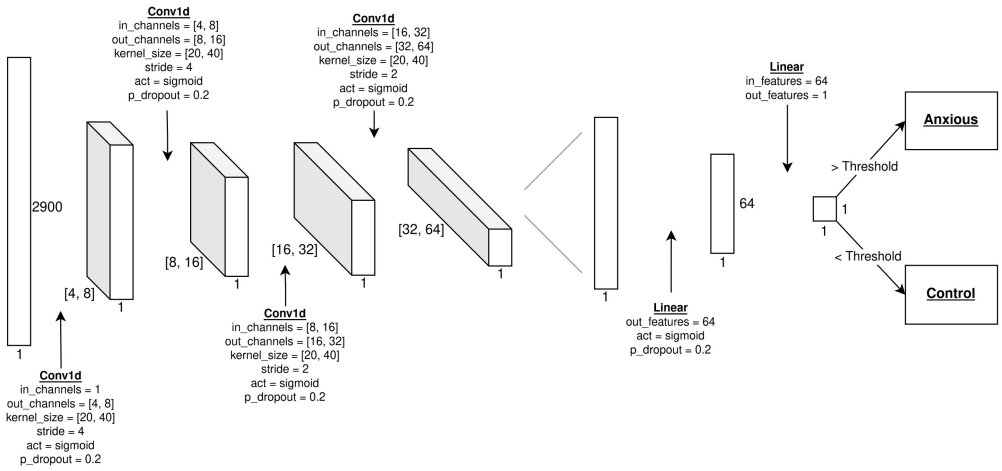

Fig. 1. The architecture consists of four 1D convolutional layers, followed by a flattening step that converts the output into a vector, which is then passed through two fully connected layers.

Fig. 2. An example of the experimental set: the participant wears a Virtual Reality headset showing a waiting room that is progressively filled with virtual avatars more and more interacting between each other. The controller's joystick allows a continuous recording of the participant's level of subjective anxiety, represented through a green bar (partially visible in the computer's screen) shown in the periphery of the visual field (the positioning can be manually adjusted by participants to fit their own visual field)

which is then processed by two fully connected (Linear) layers. The loss used was BCEWithLogitsLoss with pos_weight parameter equal to 26/37, to mitigate the class imbalance, while the optimizer employed was Adam.

Training Procedure

A Leave-One-Subject-Out (LOSO) cross-validation was employed, with the dataset of 63 subjects (each with 2900 time steps) split into training and testing sets.

Given the limited dataset size (63 subjects), we opted for a repeated stratified 80%-20% split instead of a traditional k-fold cross-validation.

While a 5-fold cross-validation would have resulted in validation sets of similar size (12-13 subjects per fold), the performance estimates would be limited to just 5 evaluations, leading to high variability across folds. Increasing the number of folds (e.g., 10-fold or more) would further reduce the validation set size (e.g., only 6-7 subjects in a 10-fold CV), making performance estimation even more unstable. Instead, by randomly splitting the training set into an 80%-20% training-validation split repeated 20 times, we obtained more robust and less variable performance estimates, as the model was validated on multiple different subject subsets.

This approach helped mitigate the dependency on specific data splits, providing a more stable validation process and a more reliable selection of hyperparameters.

The validation set served three key purposes during training:

- to determine the optimal epoch for early stopping based on the validation loss;
- to identify the optimal classification threshold based on the F1-score;
- to conduct a grid search to optimize the following model's hyperparameters:
  - Batch size: [16, 8, 4];
  - kernel size: [20, 40];
  - Learning rate: [1e-3, 5e-4, 1e-4];
  - Number of output channels in the first convolutional layer: [4, 8]. In each successive block, the number of output channels is doubled compared to the preceding convolutional layer.

  Fixed values were assigned to the remaining hyperparameters: the output dimension of the first Linear layer was set to 64, the stride to 4, and the dropout probability to 0.2. Training was performed for a maximum of 1000 epochs, with early stopping applied based on a patience of 100 epochs. Specifically, early stopping was triggered if the BCE loss on the validation set failed to show any improvement over 100 consecutive epochs, at which point training was halted.

To reduce the impact of variability across folds, the median validation loss was used to determine the stopping epoch,

and the median F1-score was used to select the classification threshold. This approach helped mitigating the risk that the selected values were not overly influenced by fluctuations in individual folds, leading to a more stable model selection process.

After identifying the best hyperparameter combination for each subject (determined by averaging the minimum BCE Loss across all 20 folds and comparing parameter combinations), a final training round was performed for each subject. In this step, the model was trained on the entire training set, excluding the left-out subject, and using the optimal number of epochs determined from the validation set. This resulted in a total of 63 subject-specific models, which were then evaluated.

III. RESULTS

Both training and validation losses across the 20 folds show a consistent trend between the Anxious and Control classes, as shown by two exemplary test subjects in Figure 3, 4. Of note, the validation loss typically reaches its minimum shortly after the training loss sharply decreases.

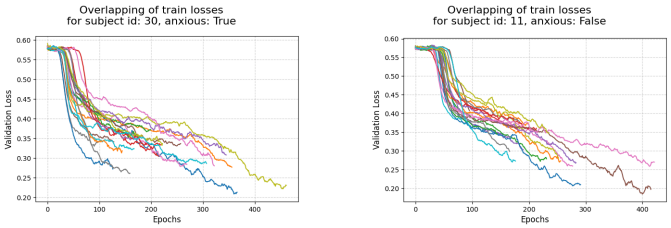

Fig. 3. Train loss of each of the twenty folds related to one exemplary test anxious and control subject

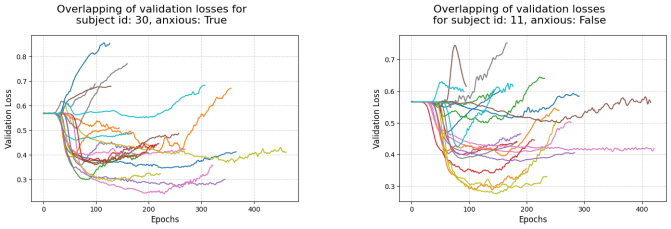

Fig. 4. Validation loss of each of the twenty folds related to one exemplary test anxious and control subject

The bar plot in Figure 5 illustrates the frequency of model selections for each subject. Notably, in 56 out of 63 cases, the selected parameter combination features a kernel size of 40 rather than 20. Additionally, a learning rate of 1e-4 was never selected, likely due to its requirement for an excessive number of epochs to begin convergence.

We obtain a total of 63 models, each trained and validated using a LOSO approach. As shown in Figure 6, the models achieve an average precision of approximately 0.77, a recall of 0.89, and an F1-score of 0.82 for the Anxious class. On the other hand, the Controls class, showed a precision of 0.80, a recall of 0.62, and an F1-score of 0.70.

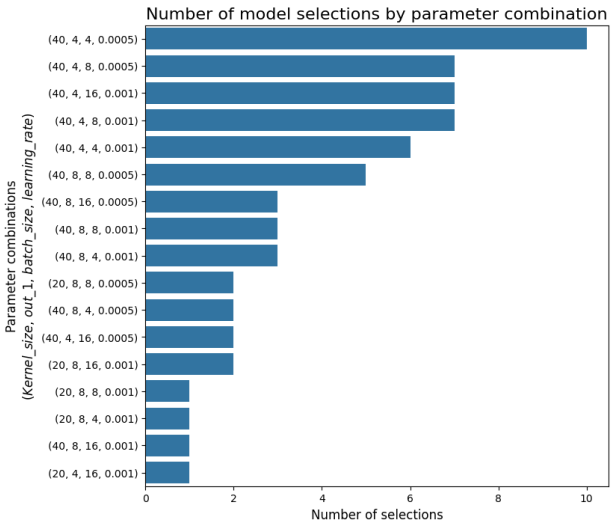

Fig. 5. Barplot of best model selected for each tessubject

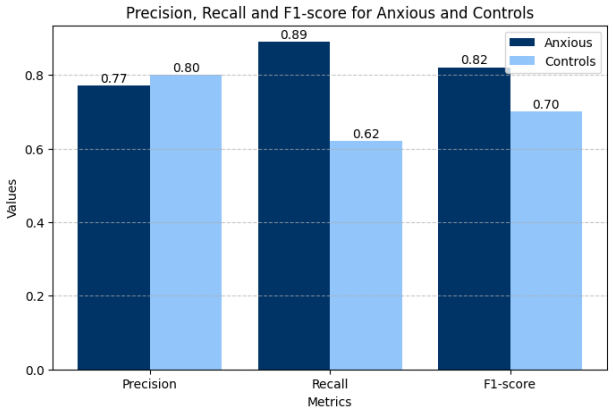

Fig. 6. F1-score, Precision and Recall for Anxious and Controls classes

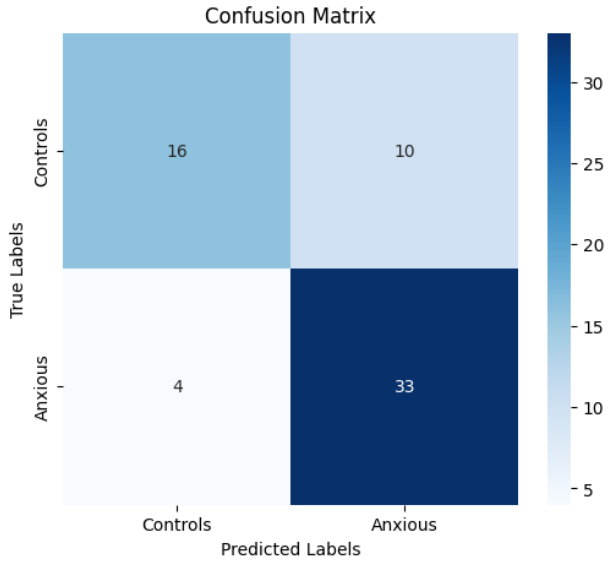

Fig. 7. Confusion Matrix

The Confusion Matrix in Figure 7 reveals that 16 control subjects are correctly predicted as controls, while 10 are misclassified as anxious. Conversely, 33 anxious subjects are correctly predicted as anxious, while 4 are misclassified as controls.

IV. Discussion and conclusions

The VR-based scenario developed for this study enables a continuous recording of subjectively perceived anxiety. This novel feature allows for a robust classification of social anxiety using only the anxiety reported through the VR joystick, making the system easily deployable in any virtual environment without the need for additional sensors or specialized equipment. This flexibility enhances its potential application in both clinical and non-clinical settings, such as therapy, screening, and real-time adaptation of virtual environments.

The choice of 1D-CNN was beneficial for their ability to capture temporal patterns in sequential data, even when the number of samples is limited. Indeed, recruiting participants for experimental studies, particularly in psychology and clinical research, is inherently challenging, leading to relatively small datasets. Although time series data provide numerous individual samples, this limitation necessitates the use of deep learning architectures capable of learning from a small number of participants while capturing meaningful temporal dependencies within the data. A key aspect that emerged was the model's preference for a kernel of size 40 in 56 out of 63 subjects, suggesting that larger time windows are effective in capturing the dynamics of perceived anxiety. This result indicates that social anxiety manifests patterns over extended time scales, and that the chosen architecture is adequate to model such characteristics.

Given such constraint on subject availability, we implemented a LOSO cross-validation approach to rigorously evaluate model generalization on unseen subjects, which is crucial in clinical applications. This approach allowed us to assess the variability in model performance across different individuals, ensuring that our method is robust to inter-subject differences. Instead of producing a single final model, this procedure provides a subject-independent evaluation of the proposed framework, demonstrating its ability to generalize to new subjects.

Once the evaluation procedure is validated, a final model could be trained on the entire dataset, leveraging the insights gained from the LOSO approach and applying it to new future data.

The model achieved for the SAD class a high recall (0.89), F1-score (0.82), and precision (0.77), demonstrating strong performance in classifying individuals with social anxiety, while for Controls class the model showed lower performance. Importantly, SAD class recall is the most critical metric in this context, as correctly identifying socially anxious individuals is a priority. False negatives (i.e., misclassifying anxious individuals as controls) could result in missed opportunities for intervention and treatment, whereas false positives (misclassifying controls as anxious) are less critical, as further clinical assessment can refine these predictions. The confusion matrix further supports this interpretation: the model correctly classifies 33 out of 37 anxious individuals, misclassifying only 4 as controls. Conversely, 10 control subjects were misclassified as anxious, which lowers precision but is a reasonable trade-off given the primary goal of detecting social anxiety. Moreover, this misclassification pattern may be influenced by the threshold selection for the L-SAS scale. Indeed, other studies suggested a lower threshold (30 instead of 55) to represent the best balance between specificity and sensitivity, i.e., to more accurately label individuals with and without SAD [21]. Future work could adjust this threshold within our model to enhance precision.

Despite the promising results, the dataset size, although larger than typical social anxiety studies, remains relatively small, and model generalization could further improve with a larger and more diverse sample. Additionally, the reliance on self-reported anxiety levels introduces a subjective component that may not fully capture the complexity of social anxiety. This approach eliminates the *need* for physiological sensors, as it records a continuous signal directly related to the subjective symptom (thus summing the advantages of physiological correlates and self-reports): nevertheless, future work could explore hybrid models that integrate both self-reported and physiological data for more comprehensive assessment of SAD behavioral symptoms (and of their therapy-induced decrease, known to be underestimated by self-reports [22]). Finally, alternative architectures such as Long short-term memory-based models or Transformers could be investigated to improve the temporal modeling of perceived anxiety fluctuations.

Beyond classification, a key next step is to embed this framework directly within the VR scenario, enabling real-time adaptation based on the user's perceived anxiety. This would allow for dynamic adjustments in the virtual environment, such as modulating social stimuli in response to the user's anxiety state, paving the way for personalized VR-based interventions for social anxiety disorder.

References

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: Autor, 5th ed. ed., 2013.

[2] D. F. Santomauro, A. M. M. Herrera, J. Shadid, P. Zheng, C. Ashbaugh, D. M. Pigott, C. Abbafati, C. Adolph, J. O. Amlag, A. Y. Aravkin, *et al.*, "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the covid-19 pandemic," *The Lancet*, vol. 398, no. 10312, pp. 1700–1712, 2021.

[3] R. Kindred and G. W. Bates, "The influence of the covid-19 pandemic on social anxiety: a systematic review," *International journal of environmental research and public health*, vol. 20, no. 3, p. 2362, 2023.

[4] S. Frumento, A. Iannizzotto, A. Greco, E. P. Scilingo, A. Gemignani, and D. Menicucci, "Development of a behavioral avoidance test in virtual reality (vr-bat)," in *2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pp. 949–953, 2023.

[5] N. Chowdhury and A. H. Khandoker, "The gold-standard treatment for social anxiety disorder: A roadmap for the future," *Frontiers in Psychology*, vol. 13, p. 1070975, 2023.

[6] K. P. Wong, C. Y. Y. Lai, and J. Qin, "Systematic review and meta-analysis of randomised controlled trials for evaluating the effectiveness of virtual reality therapy for social anxiety disorder," *Journal of Affective disorders*, vol. 333, pp. 353–364, 2023.

[7] D. Banakou, T. Johnston, A. Beacco, G. Senel, and M. Slater, "Desensitizing anxiety through imperceptible change: feasibility study on a paradigm for single-session exposure therapy for fear of public speaking," *JMIR formative research*, vol. 8, p. e52212, 2024.

[8] C. Lacey, C. Frampton, and B. Beaglehole, "A self-guided virtual reality solution for social anxiety: Results from a randomized controlled study," *Journal of psychiatric research*, vol. 180, pp. 333–339, 2024.

[9] H. S. Jeong, J. H. Lee, H. E. Kim, and J.-J. Kim, "Appropriate number of treatment sessions in virtual reality-based individual cognitive behavioral therapy for social anxiety disorder," *Journal of clinical medicine*, vol. 10, no. 5, p. 915, 2021.

[10] P. Lindner, A. Miloff, S. Fagernäs, J. Andersen, M. Sigeman, G. Andersson, T. Furmark, and P. Carlbring, "Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial," *Journal of anxiety disorders*, vol. 61, pp. 45–54, 2019.

[11] A. L. McGlade, M. Treanor, R. Kim, and M. G. Craske, "Does fear reduction predict treatment response to exposure for social anxiety disorder?," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 79, p. 101833, 2023.

[12] M. Martini, E. Viola, F. Bossi, S. Frumento, A. Iannizzotto, S. Said, A. L. Callara, F. Solari, E. P. Scilingo, A. Greco, *et al.*, "Designing an immersive virtual reality scenario for social anxiety elicitation and modeling: a preliminary evaluation," in *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pp. 435–440, IEEE, 2024.

[13] O. Corneille and B. Gawronski, "Self-reports are better measurement instruments than implicit measures," *Nature Reviews Psychology*, pp. 1–12, 2024.

[14] S. L. Koole and K. Rothermund, *The psychology of implicit emotion regulation*. Psychology Press, 2011.

[15] M. B. Powers, N. F. Briceno, R. Gresham, E. N. Jouriles, P. M. Emmelkamp, and J. A. Smits, "Do conversations with virtual avatars increase feelings of social anxiety?," *Journal of anxiety disorders*, vol. 27, no. 4, pp. 398–403, 2013.

[16] R. G. Heimberg, K. Horner, H. Juster, S. Safren, E. Brown, F. Schneier, and M. Liebowitz, "Psychometric properties of the liebowitz social anxiety scale," *Psychological medicine*, vol. 29, no. 1, pp. 199–212, 1999.

[17] A. Prunas, I. Sarno, E. Preti, F. Madeddu, and M. Perugini, "Psychometric properties of the italian version of the scl-90-r: a study on a large community sample," *European psychiatry*, vol. 27, no. 8, pp. 591–597, 2012.

[18] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sensing of Environment*, vol. 221, pp. 430–443, 2019.

[19] Y. Xi, C. Ren, Q. Tian, Y. Ren, X. Dong, and Z. Zhang, "Exploitation of time series sentinel-2 data and different machine learning algorithms for detailed tree species classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7589–7603, 2021.

[20] S. Abbas, S. Ojo, A. A. Hejaili, G. A. Sampedro, A. Almadhor, M. M. Zaidi, and N. Kryvinska, "Artificial intelligence framework for heart disease classification from audio signals," *Scientific Reports*, vol. 14, no. 1, p. 3123, 2024.

[21] D. S. Mennin, D. M. Fresco, R. G. Heimberg, F. R. Schneier, S. O. Davies, and M. R. Liebowitz, "Screening for social anxiety disorder in the clinical setting: using the liebowitz social anxiety scale," *Journal of anxiety disorders*, vol. 16, no. 6, pp. 661–673, 2002.

[22] S. Frumento, A. Gemignani, and D. Menicucci, "Perceptually visible but emotionally subliminal stimuli to improve exposure therapies," *Brain Sciences*, vol. 12, no. 7, p. 867, 2022.