**Observational Learning for Personalized Generative Anxiety Therapy**

Samuel Solomon1, Yadong Xu1, Ruixiao Liu1, Jadelynn Dao1, José Lasalde Ramírez1, Jihong Min1, Dickson Yao1, Sarah Solomon2, Canran Wang1, Wei Gao1\*

**Affiliations.**  
1Andrew and Peggy Cherng Department of Medical Engineering, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA.

2Dartmouth Hitchcock Medical Center and Clinics, Adult Psychiatry Residency, Lebanon, NH, USA.

\*Corresponding author. Email: [weigao@caltech.edu](mailto:weigao@caltech.edu).

**Flow:**

1. **Why do we need to have a universal way of handling physiological data for emotion modeling to be successful. Why must this work between physiology and emotions. The consequences of success or failure. Emphasize the issue of small datasets.**
2. **From a psychological perspective (foundational theory), from a machine learning perspective (datasets compared to other successful models), what is the current state and consensus around physiological emotion modeling.**
3. **Our approach is to use observational learning to overcome these challenges**
4. **Introduction to the datasets and therapy.**
5. **Summation**

**Abstract.**

**Summary paragraph.**

Advances in wearable technology have enabled the collection of various anxiety-based physiological datasets across multiple sensors and emotional questions; however, reliably associating these signals with emotional states remains an unsolved problem. This study presents a new neural architecture called Observational Learning for merging physiological datasets that enable researchers to work on fragmented out-of-domain datasets that currently cannot be meta-trained together. This work demonstrates the potential to share wearable data across platforms not only within mental health, but also other continuous health monitoring systems, enhancing the capabilities of wearable healthcare and offering insights into designing more robust, adaptive systems that respond to complex physiological patterns in real time.

**Main.**

Physiological responses are foundational to any emotional state, yet associating time-dependent biometric patterns with affective experiences remains challenging due to the fragmented, biomarker-specific, emotion-specific, and sparsely sampled nature of existing datasets, which impedes clinical efforts toward data-driven, autonomous mental health interventions. However, without any mathematical framework or high-precision wearables in 1872, Darwin visually identified universal physiological reactions and facial expressions across the animal kingdom that are strongly linked to discrete emotional experiences, suggesting that some affective and biological states may have inherent correlations1. With the recent improvements in text-based sentiment analysis that have enabled models to capture affective cues within longer sentences, researchers have begun to revisit the problem of decoding physiological-based emotions across broader time windows with diverse biometric patterns. Unfortunately, the transformer-based architectures that dominate contemporary sentiment-focused large language models are permutation-invariant and emphasize semantic correlations over temporal trends in time-series signals, limiting the direct application of these frameworks for physiological-based affective computing2. Consequently, 150 years after Darwin’s first observations, the scientific community still lacks consensus on which subset of biomarkers reliably indicate emotions, thereby hindering collaboration on any open-source physio-emotion model, which has been vital to the success of similar widely used sentiment text analysis frameworks such as Google’s Bidirectional Encoder Representations from Transformers (BERT) model.

Despite the lack of transparency within physio-emotion modeling, psychologists as early as Sigmund Freud have contended that momentary state anxiety (s-anxiety) is a compilation of emotional states: feelings that originate within physiological perturbations (Schachter and Singer). Paul Ekman found that one could model these emotions as universal facial expressions, which Matsumoto and Willingham found are not socially learned, as even blind and deaf athletes exhibit predictable biological reactions to similar emotional stimuli. This suggests that some emotional expressions are linked from an evolutionary adaptation of the body reacting to potential threats—such as squinting out of disgust to block harmful stimuli from entering. Unfortunately, despite extensive theoretical and empirical evidence, the continuous monitoring of s-anxiety through wearable devices has not been widely accepted due to poor generalization across out-of-domain examples. Meanwhile, other well-known health monitoring frameworks, such as the Framingham Risk Score, utilize larger datasets with fixed biomarkers and health outcomes that enable actionable medical insights within poorly correlated health trends. Meanwhile, mental health modeling utilizes subjective emotional observations, recorded via different wearable platforms. This lack of any standardized biomarkers and emotion questions for s-anxiety promotes the use of dataset-specific architectures that prevents data compilation across researchers for an open-source generalizable physio-emotion model.

We introduce Observational Learning (OL) as a foundational framework designed to address the challenges of integrating out-of-domain time-series datasets from wearables. OL achieves this by decoupling biomarker weights and emotion weights, enabling the model to accommodate individual input signals and output emotions. Within OL models, each physiological feature (observable event) is treated as the outcome of a hidden biological perturbation, termed the physiological profile (p-profile). This approach mimics the scientific method when solving inverse problems by treating the p-profile as the underlying causal state. The system is "probed" by observing how perturbations in the p-profile manifest as changes in time-series signals (input features). Missing signal data, whether partial or full, do not invalidate other observations, since the p-profile is down-sampled to the observations at the end of training. This reduces the number of discarded samples from motion artifacts and lost data packets, which is inherently common in real-time wearable recordings.

An OL framework learns the p-profile by assessing how well the hidden state justifies the collectively observed signals. This is achieved through a series of reversible, learnable transformations that connect input and output states. By mapping all the biometric responses to the p-profile, OL models effectively learn a unified representation of a patient’s physiological state while maintaining the ability to new biomarkers and timepoints as needed. Each observable event could therefore be considered a weak classifier of the p-profile, where more observations increase the prediction confidence. When data is missing, which is a problem in any real-time wearable platform, the weak classifier cannot be used to reconstruct the physiological profile. Fortunately, once the model weights are trained and frozen, few-shot learning can be employed during inference to approximate the p-profile and provide a causal explanation (e.g., Granger causality) for each observation.

We present observational learning (OL) as a new foundational architecture for overcoming the current challenges with merging out-of-domain time-series datasets by decoupling biomarker and emotion weights, providing the model with a trainable path for different input signals and output emotions as seen within separate physio-emotion datasets currently in the literature. In OL models, each physiological feature (observable event) is treated as a response from a hidden biological perturbation, which we refer to as the physiological profile (p-profile). Observational learning therefore mimics the scientific method when solving an inverse problem (hidden state dynamics) by probing a system (p-profile, output state) and observing how the resulting perturbations (time series signals, input features) respond. Missing full or partial features do not invalidate other observations, allowing the model to backpropagate on incomplete datasets, minimizing the number of discarded samples. OL models can therefore learn the p-profile based on how well the hidden state can justify each observed vital sign collectively, using a core set of reversible learnable transformations between the output and input states. This allows us to condense all relevant physiological information within a common structure while maintaining the flexibility to add or remove new biomarkers and timepoints in the future. Once the p-profile has established a path to all observations, the network can be inverted to reconstruct the p-profile from any of these weak classifiers. Unfortunately, if samples are missing, as is common in real-time recordings, the model cannot affectively project the samples into the wavelet domain. Therefore, after the weights of the model are fixed, during inference, few-shot learning of the physiological profile can still explain (granger-cause) each observation.

The forward method cant be done.

By training the inverse model using a reversible map from the output latent manifold to a given input biomarker, OL architectures easily adapt to a new set of biomarkers as well as missing biomarker data, allowing researchers to meta-train on different real-world physiological datasets with a common set of shared weights.

Unfortunately, within physiological datasets, there inherently exist motion artifacts and noise that can bias data interpretations, which is only exacerbated by the relatively small size of physio-emotion datasets compared to traditional large language models. To learn how to deal with these events, we employ meta-learning techniques to share weights and integrate multiple datasets together. Each path to the physiological profile can be broken up into 3 main sections: a signal-specific layer, shared meta-learning layer, and another signal-specific layer. Most of the model weights lie within the meta layer, removing the burden of small datasets, while the signal specific layers provide dataset-specific expressibility. Each single-specific layer address subproblems within the broader task of physiological-emotion mapping, enabling few-shot learning on previously unseen datasets. Our core model has demonstrated strong performance after training on five physiological meta-datasets from the literature (WESAD, EMOGNITION, DAPPER, AMIGOS, and CASE) and has shown few-shot learning capabilities on two out-of-domain datasets that we compiled (EMPATCH). We aim to further enhance our model's generalizability and utility for the broader research community by establishing the first domain-agnostic physio-emotion model.

By understanding the complex relationship between physiology and emotions, we gain deeper insights into the intricacies of human emotional responses, creating opportunities for more personalized mental health interventions. Our approach was validated through generative virtual reality, music, heat, and voice therapy sessions, all of which were personalized to the subject’s real-time emotion and biological profile. In early tests, we demonstrated the ability to modulate and reduce negative affective emotions and anxiety. We further validated our model’s predictions against the gold standard State-Trait Anxiety Inventory (STAI) Y2 questionnaire as well as the Positive and Negative Affect Schedule (PANAS), showing strong alignment between our automated therapy outcomes and patient-reported results. This validation not only supports Ekman’s theoretical framework linking physiological and psychological domains but also establishes a new benchmark for closed-loop personalized therapeutic interventions, paving the way for automated, in-home, tailored mental health therapies.

Spielberger used this as a foundation for quantifying s-anxiety through the State-Trait Anxiety Inventory (STAI) questionnaire, which has undergone multiple independent assessments for validity and reliability across different cultures, age groups, occupations, race, gender, and languages. Paul Ekman mapped these emotions to facial expressions within the Facial Action Coding System (FACS), proving that certain physiological responses are universally associated with emotions across cultures. Further analysis by

The brain can therefore process an increased heart rate or rapid breathing as anxiety or excitement, depending on the situation.

In 1895, Sigmund Freud realized that anxiety can be fully explained by discrete emotions, which according to the 1894 James-Lange theory meant that anxiety can be fully explained by physiology.

researchers have been unable to isolate single physiological signals that yield deterministic emotional states within a subject, partly due to the subjective interpretation of emotions and the complexity of physiological responses within the broader context of relatively small and noisy datasets.

In contrast, mental health lacks comparable rigor, partly due to the subjective nature of emotions.

While physiological theories like Schachter-Singer’s two-factor model and Ekman’s Facial Action Coding System suggest measurable links between physical responses—such as heart rate or facial expressions—and emotions, attempts to identify deterministic biomarkers have failed due to the fragmented, noisy, and sparsely sampled nature of datasets.

In parallel, Schachter and Singer found that emotions are causal to physiological reactions, with the mind interpreting the body’s reaction as a ‘feeling,’ fundamentally link anxiety and physiological responses.

The continuous monitoring of mental health through wearable devices poses a unique and pressing challenge as compared to tracking physical well-being. While there is extensive literature documenting biomarker implications for physical illnesses, such as the Framingham Risk Score3, Susceptible-Infectious-Recovered model4, and the progression of diseases like Covid-19, mental health lacks clear, unbiased, and reproducible models. This knowledge gap becomes critical when addressing long-term anxiety, as acting on poorly understood models may exacerbate mental deterioration and lead to more severe conditions such as depression.

Without any standardized approach, physiological-based emotion modeling has been widely accepted within the psychological community

Anxiety modeling has been a longstanding endeavor in psychology. In 1895, Sigmund Freud defined anxiety as arising from emotional states, linking sensations such as tension, nervousness, and apprehension5. In 1966, Cattell distinguished anxiety that arise from emotional states and personality traits6. Building upon these definitions, Spielberger introduced the State-Trait Anxiety Inventory (STAI) in 1983 (Fig. 1), a well-investigated questionnaire that has undergone multiple independent assessments for validity and reliability across different cultures7–9, age groups9, occupations, races8, gender9, and languages. The STAI exam numerically quantifies both short-term, emotionally based state anxiety (S-anxiety) and long-term, personality-based trait anxiety (T-anxiety)10. Despite its population-based correlation to S-anxiety, the STAI survey relies on subjective user ratings of emotions, introducing biases to individual scores.

We validated the generalizability of the to new datasets via commercial and novel electronic skin wearable devices.

to decode well-observed physiological patterns within humans

The p-profile represents a signal-agnostic manifold: a latent construct capable of explaining each biological reaction independently.

Here, we present the first bidirectional wavelet neural operator (BWNO) trained within a newly proposed observational learning (OL) framework designed to overcome the current limitations of wearable datasets to develop the first open-sourced physio-emotion model as well as real-time personalized anxiety therapy.

, suggesting that some affective and physiological states are intrinsically linked. Spiel,

The problem of developing a reliable universal affective computing pipeline is hindered by sparse, biomarker-specific, and unevenly sampled wearable datasets, limiting the applicability of machine learning models that require large datasets and rely on domain-specific attributes.

By training the inverse problem, observational learning solves a key issue within wearable data analysis as the architecture does not require evenly sampled nor specific input features. This is achieved by projecting the PP manifold into the Fourier domain, solving the inverse problem, and mapping the oversampled solution back into the spatial domain at unevenly sampled time points. The key to convergence for OL models relies on the set of information being strongly correlated across different signal combinations, so that one can isolate a common domain with minimal signal-specific weights. For physiological data, prior literature has demonstrated this correlation through state-space models and multivariate regression, where heart rate data can be hallucinated through respiratory information. We use a 1:16 ratio for dataset-specific to shared weights, allowing for small datasets to be easily interfaced within our framework with minimal training. Within this framework, each biomarker serves as evidence – a weak classifier – for the underlying perturbation rather than as a mandatory input, where the model is trained from the hidden variable to the observed signals. We have demonstrated this through partially and fully masking out a given signal, while converging to a similar PP manifold to reconstruct the missing points.

The problems in compiling enough trainable data to associate physical and mental states prevent simple classification systems such as decision trees and fully connected networks from converging and generalizing across datasets. The main obstruction for using more complex predictors, is the lack of standardization across the literature for which physiological stimuli (sweat-metabolites, facial expressions, vital signs) as well as emotional responses (positive and negative affective states) should be recorded for a complete and universal physio-emotional dataset. Currently, there is no universal physiological model for emotion classification as many datasets compiled in the literature – WESAD, EMOGNITION, DAPPER, AMIGOS, and CASE – are signal-specific.

Compiling large datasets is hard and time consuming.

, each emotion model is only trained on a selective and highly specialized dataset, where it is impossible to generalize across publications. This problem is exacerbated by black-box machine learning techniques, which hinders how relevant any emotion study is to one another. Further issues include resampling features to match a neural architecture, random information loss during data transmission, sampling bias from emotion questionnaires, as well as data sparsity compared to the number of learnable parameters. Modern advancements in machine learning techniques including text-based sentiment analysis as well as neural operators offer a new way for overcoming these issues within the broader context of universal affective computing.

In this work, we present the first bidirectional wavelet neural operator (BWNO) trained through observational learning (OL).

Unlike fixed models, each new observation acts as a weak classifier, increasing the confidence that the manifold represents the body’s current state. We claim that missing or unaligned data points, which typically degrade model performance, have a limited impact on this flexible architecture.

One key innovation that allows observational learning the flexibility to generalize across input domains is bidirectional training. Through diffeomorphic transformations, OL models independently map the common ancestor to each observable output using a set of invertible weights, allowing for both inference and interpretation depending on the direction. During training, the model learns how to remove information from the manifold to validate each observation. While projecting different spaces into a common latent space is extremely difficult, by inverting the weights in observational learning during inference, the model can transform each individual signal into the manifold’s space. Each projection inherently contains noise that is smoothed out when considering all observations together. This approach mirrors scientific experiments, where multiple correlated observations are used to reconstruct the underlying properties of a system. Consequently, this architecture provides an interpretable framework for machine learning researchers, emphasizing not only the final predictions but also the mappings across each space, offering deeper insights into physiological events and their associated emotional responses.

**Discussion.**

Here we present our analysis from capturing real-time physiological signals in response to four unique experimental stressors, including music, cold pressure, exercise (biking), and virtual reality, to investigate the relationship between physiology and emotional responses. The custom-designed patch integrated four key sensors—electrooculography (EOG), electroencephalography (EEG), electrodermal activity (EDA), and temperature—enabling the continuous recording of both electrical and thermal signals (Fig 3a). These sensors provided rich, multidimensional data that form the foundation of the analysis.

**Device Design and Signal Processing**

Our device is composed of EOG, EEG, EDA, and temperature sensors that conformally adhere to the surface of the skin. Signals are collected and processed through a series of filters, amplifiers, voltage regulators, and a microcontroller unit (Fig 3bc). This robust design ensured that signals were processed in real-time and transmitted for further analysis.

Signal preprocessing, illustrated on the top right, involved filtering the raw physiological data to remove noise and artifacts while preserving essential features related to emotional states. The filtering and amplification stages were particularly important in handling high-frequency noise in EEG and EOG recordings, while EDA and temperature data required low-pass filtering to stabilize readings. These preprocessing steps ensured the data was clean and interpretable for the subsequent analyses of emotional responses.

**Real-Time Data Collection from Stressors**

The figure presents representative datasets from four different stressors—cold pressure, exercise, music, and virtual reality—demonstrating the robustness of the device in capturing physiological signals across various conditions. The datasets are displayed in the middle portion of the figure, showing the raw signal patterns for each stressor.

In the cold pressure and exercise experiments, we observed clear variations in physiological responses. For example, alpha power from the EEG readings, which are known to decrease during times of anxiety, distinctly dropped during the cold pressure test, indicating increased sympathetic nervous system activity. Similarly, temperature fluctuations were evident during the cold pressure experiment, correlating with participant discomfort. In contrast, music and virtual reality experiments yielded more moderate physiological changes, suggesting these stressors were less intense.

**Positive and Negative Affectivity Scores**

We also analyzed affective responses through positive and negative affectivity scores, focusing on how these scores varied across the cold pressure and exercise experiments. As shown in figure 3e, affectivity scores derived from physiological data showed distinct patterns. Cold pressure led to an overall increase in both positive and negative affectivity, while exercise was able to boost positive affectivity and decrease negative affectivity. The graphs show overlapping yet distinguishable distributions of affective states, with cold pressure eliciting a strong negative response and exercise producing mixed emotional states.

**Emotional Distributions and Correlations**

The violin plots (3g) represent the distribution of emotional states as measured by the State-Trait Anxiety Inventory (STAI). These distributions capture the range of emotional intensities experienced by participants, highlighting the variability across different stressors. Cold pressure induced a broader distribution of negative emotions compared to exercise, which showed a wider spread across both positive and negative affective states.

Lastly, the correlation analysis from the Positive and Negative Affect Schedule (PANAS) revealed significant relationships between emotional states (3h). The correlations confirm the model’s ability meta-learn a core set of transformations, while leaving only a small set of weights to differentiate each emotion.

**Implications for Personalized Interventions**

The data gathered from this study not only validate our device’s capacity to measure physiological responses but also demonstrate its potential in developing automated, personalized mental health interventions. By accurately mapping physiological signals to emotional profiles in real time, our approach could be used to tailor interventions such as music or virtual reality therapy to a patient's emotional state, offering a novel method for treating anxiety and mood disorders.

The results also set a strong foundation for future work, where further validation across larger and more diverse datasets will be critical. The combination of wearable sensors and machine learning can provide increasingly precise mappings between physiology and emotional states, ultimately leading to more robust, domain-agnostic models for therapeutic applications.

The intrinsic softness, thickness, and gas permeability of skin-interfaced electronics hold promise for long-term and continuous high-fidelity monitoring of physiological states. Recently, there has been many exciting progresses in the fabrication of ultrathin and soft wearables, as evidenced by elastomeric nanofiber mats1 and elastic conductor with microcracked structures2 and semiconductors3. While these devices exhibited seamless contact with skin, the fabrication typically involves electrospinning, iterative spin coating, and thermal evaporation, which are laborious and time-consuming. Moreover, the manufacture of large patches to cover the head or face presents challenges due to limitations in dimensions imposed by instruments such as spin-coaters and oxygen plasma cleaners. While the float assembly method shows promise4,5, achieving in-situ generation of microscale porosity remains fundamentally challenging.

Here, we present a phase-separated float assembly (PSFA) technique that achieves in situ phase separation of silver nanowires (AgNWs) from thermoplastic polyurethane (TPU) matrices at water-air interface. This process enables rapid (in minutes), facile and large-area (>200 cm2) fabrication of ultrathin elastic patches (~1 µm) and phase-separated porous elastic sensors (~6 µm). The resulting ultrathin porous electronic skin exhibits substantial improvement in unobtrusiveness, comfort, and intimate contact with skin. As illustrated in Fig. 2a, the fabrication begins with the preparation of a precursor solution containing polymer solution (TPU in tetrahydrofuran (THF)) and conductive filler solution (AgNWs in ethanol). Upon the injection of the nanocomposite solution into water bath, phase separation initiates due to the evaporation and dissolution of the volatile solvent (THF) and non-solvent (ethanol) into water bath, whereas water-immiscible TPU stays on water surface. This results in macroscopic Marangoni flow and microscale liquid-liquid demixing simultaneously (Fig. 2b). On the one hand, dissolution of THF and ethanol lowers the local surface tension, which results in a circular surface tension gradient near the droplet (Fig. 2b, (i)). This gradient induces Marangoni flow from the center to the boundary, which pushes the boundary expansion and mass flow of the precursor solution. This leads to a partially dried ultrathin nanocomposite film floating at the water-air interface. On the other hand, concurrent liquid-liquid demixing results in phase separation of the TPU-rich and TPU-poor phases (Fig. 2b, (ii)). Ag NWs with amphiphilic ligands (that is, polyvinyl pyrrolidone) reside in the PU-poor phase due to its immiscibility with TPU solution. This process generates co-continuous phases in three dimensions, completed by drying that creates continuous porous structures within the PU matrices.

The resulting porous nanocomposite consists of randomly distributed AgNWs confined within porous polymer matrices (Fig. 2c), forming conductive percolation networks that bridge interconnected pores across multiple scales. Unlike conventional spin-coating processes, the PSFA method enables the fabrication of large, thin, yet adhesive electronic skins suitable for multimodal stress therapy. These devices are stretchable, ultrasoft, and provide conformal contact with the skin for enhanced comfort and unobtrusiveness (Fig. 2d–f). Additionally, the porous conductor demonstrates exceptional durability and reliability, with only a slight resistance change (R/R₀ = 4.8) after 3,000 cycles of 25% stretching, compared to the non-porous control (R/R₀ > 200; Fig. 2h). The multiscale interconnected cellular structure further enhances porosity and breathability, facilitating skin perspiration (Fig. 2i) and improving long-term biocompatibility (Fig. 2j, k).

To evaluate the performance of our porous conductor for electrophysiological signal recording, we first examined the electrode-skin contact impedance. The ultrathin porous electrode showed notably reduced impedance compared to its thicker counterpart, primarily due to its intimate skin contact (Fig. 2l). Furthermore, electrooculogram (EOG) signals recorded before and after sweating using the porous on-skin sensors demonstrated quantitatively comparable patterns to those obtained with Ag/AgCl gel electrodes (Fig. 2m, n). Lastly, we present a comprehensive comparison of our Empatch with state-of-the-art materials fabricated by other methods, such as spin-coating and float assembly4, in terms of processing efficiency, large-area fabrication, breathability, wearability, sweat resistance, and electrical conductivity (Fig. 2o).

To accommodate dynamic skin deformations, we utilize strain-resilient interconnects composed of a porous liquid metal composite6, which enables reliable and robust sensing and therapeutic capabilities. This performance is demonstrated by the proper operation of a light-emitting diode array (Fig. 2p), which can be attributed to the stable electrical resistance of the interconnect under strain (Fig. 2q) and its resilience to various damage scenarios, including scalpel punctures, hammer impacts, and tweezer pricks (Fig. 2r). Furthermore, electrocardiogram signals recorded using the liquid metal composite wiring showed negligible variation before and after stretching (Fig. 2s).

**Methods.**

While rapid processing and reaction to physiological cues are a core function of the human body, the intricacies of how this occurs are not fully understood.

Modern approaches using machine learning and wearable devices are now exploring how everyday biomarkers contribute to one’s inner psychological state.

**Reference**

1. Darwin, C. *The Expression of the Emotions in Man and Animals*. (J. Murray, 1904).

2. Zeng, A., Chen, M., Zhang, L. & Xu, Q. Are Transformers Effective for Time Series Forecasting? Preprint at https://doi.org/10.48550/arXiv.2205.13504 (2022).

3. Wilson, P. W. F. *et al.* Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* **97**, 1837–1847 (1998).

4. Vallée, A., Faranda, D. & Arutkin, M. COVID-19 epidemic peaks distribution in the United-States of America, from epidemiological modeling to public health policies. *Sci. Rep.* **13**, 4996 (2023).