Proceedings of the 89th Annual International Meeting of the Psychometric Society, Prague, Czech Republic, 2024 (2025).

PSYCHOMETRIC SOCIETY

Wavelet-Based Deep Learning for Multi-Time Scale Affect Forecasting

Sy-Miin Chow,*† Young Won Cho,† Xiaoyue Xiong,† Yanling Li,† Yuqi Shen,† Jyotirmoy Das,† Linying Ji,‡ and Soundar Kumara†

†Pennsylvania State University
‡Montana State University
*Corresponding author. Email: symiin@psu.edu

Abstract

We present results using the scattering transform, a machine learning approach that integrates wavelet analysis with deep learning models in a single step, enabling efficient forecasting and classification. Because coefficients in the deep neural network are fixed to known coefficients in the wavelet analysis, computational burden and expenses are greatly reduced, with useful results found even with sample sizes that are comparably small for standard machine learning applications. Using illustrative and empirical examples designed to mirror multi-temporal and non-stationary changes in individuals' physiological and perceived (self-report) affect arousal, we propose a multi-subject extension of a feature activation heatmap proposed previously for convolutional network models, and illustrate its utility in displaying the time-varying importance of multiple physiological signals' frequency components in forecasting individuals' self-report affect arousal during a laboratory emotion induction task.

Keywords: affect forecasting, wavelet, machine learning, multi-time scale, scattering transform, deep learning

1. Introduction

Affective processes have been reported to show distinct changes across multiple temporal scales. The phrase "moods nag at us, emotions scream at us" (Larsen, 2000) was used to clarify the distinctions between emotions, which are short–lived, relatively intense, and are triggered by specific events or targets; and moods, which reflect longer–term feelings that may not have a specific cause. Indeed, past research has indicated that human affect exhibits changes in multiple temporal resolutions, including relatively gradual variations shaped by personality, life experiences, as well as ebbs and flows as triggered by personal, environmental, and situational contexts (e.g., daily stress, weather), as well as their interactions (Knapova et al., 2024; Kuppens et al., 2012; Ram et al., 2014).

From the frequency components of electrocardiogram (ECG) recordings (Bigger et al., 1995) to diurnal (Lu et al., 2015) and weekly cycles in emotions (Chow et al., 2005), cyclic regularity has been observed in the affective changes of individuals in controlled and natural environments. Unfortunately, these cycles may also show nonstationarity over time. For example, college students' weekly fluctuations in positive emotion might be disrupted during final exam weeks (Chow et al., 2009).

Dynamic features such as maximum level shift, maximum variance shift, and standard deviation of the first derivative of the time series have been shown to improve the predictive power of machine learners in classifying device non-wear using time series of individuals' actigraphy data (Das et al., 2025). To extract dynamic features that specifically target periodicity in time series data over time, we consider *wavelet scattering*, a class of machine learning methods that incorporates wavelet analysis features into deep learning models (Andreux et al., 2020; L. Liu et al., 2018; Z. Liu et al., 2020; Oyallon et al., 2013; Sepúlveda et al., 2021), to reveal whether and in what ways physiological signals such as ECG predict individuals' self-report affective arousal over time. Because most of the coefficients in wavelet analysis are fixed to known values, deep learning models consisting of such wavelet components are characterized by fewer coefficients to be estimated, and have been shown to produce useful results even with sample sizes that are comparably small for machine learning methods (Andreux et al., 2020).

2. Scattering Transform with Deep Learning for Capturing Non-stationarity in Multiple Temporal Resolutions Over Time

We propose and evaluate a deep learning model architecture that integrates wavelet-based feature extraction with a deep learning model to predict a continuous dependent variable over time across multiple individuals. We use the scattering transform functions from the Python package, Kymatio, which provides an efficient implementation of wavelet transformations within a machine learning framework (Andreux et al., 2020; Bruna & Mallat, 2013; Mallat, 2012), and is readily integrated with other deep neural network modeling functions in PyTorch (Imambi et al., 2021) and TensorFlow (Abadi et al., 2015), to implement the proposed modeling architecture.

Wavelet analysis is a popular approach for capturing time-varying or other sources of heterogeneity in the frequency components of a time series (Mallat, 1999; Suh et al., 1999). Wavelet analysis utilizes wavelets (denoted as ψ), which are oscillatory mathematical functions associated with distinct temporal (or frequency) resolutions, to approximate a time series. By systematically applying wavelet extraction operations at a targeted range of frequency bands as dictated by user-specified hyperparameters, the scattering transform implemented in Kymatio provides a stable representation to capture temporal changes across multiple frequency resolutions.

2.1 Scattering Transform for Feature Extraction

Our proposed modeling architecture first applies the scattering transform to each feature independently. Kymatio's scattering transform applies the discrete wavelet transform (DWT) to extract stable, multiresolution features from a time series in three orders. These features are known as scattering coefficients, and they capture aspects of the signal at different levels of granularity.

The *Zeroth-Order Coefficients*, denoted as $S_J[0]x$, is computed as:

$$S_J[0]x[t] = x \star \phi_J, \tag{1}$$

where $\star$ denotes convolution, and ϕ_J is a low-pass filter that allows low frequencies to pass through, as determined by a downsampling parameter, J. The convolution operation in (1) can be thought of as applying global averaging of the signal to produce a baseline feature. The parameter J determines the largest scale of the scattering transform, such that the maximum temporal (or spatial) scale captured is 2^J samples (e.g., time steps). Thus for a time series of length T, the scattering transform extracts summary coefficients downsampled by a factor of 2^J, resulting in summary outputs over roughly $\frac{T}{2^J}$ time windows. A larger J corresponds to a coarser time resolution.

The 1st-order scattering coefficients, $S_J[1]x$, is computed by convolving the signal with a bandpass wavelet, ψ_{λ_1}, followed by taking the modulus (denoted as |.|), and then smoothing with ϕ_J:

$$S_J[1]x[t, \lambda_1] = |x \star \psi_{\lambda_1}| \star \phi_J, \tag{2}$$

where $x \star \psi_{\lambda_1}$ represents convolution of the signal with a bandpass filter, ψ_{λ_1}, centered at frequency

λ_1 as:

$$(x \star \psi_{\lambda_1})[t] = \sum_{\tau} x[\tau]\psi_{\lambda_1}[t-\tau] \tag{3}$$

where τ sums over the time region supported by ψ_{λ_1}. Thus, the convolution "slides" the bandpass filter across time, computing a weighted sum at each position. In doing so, the filter extracts components of $x[t]$ that match the shape of ψ_{λ_1} and fall into the frequency range targeted by the filter. The modulus ("amplitude") of the convolutions is then taken, followed by the application of a low-pass filter to enhance the stability of the scattering coefficients. Another hyperparameter that governs the granularity of the frequencies extracted is Q, which controls the number of wavelets used per octave (a broad frequency range where the upper limit is twice the lower limit; e.g., 1-2 Hertz or Hz). A larger Q increases the number of wavelets, producing more frequency bands and associated scattering coefficients that capture more granular differences in frequencies. In total, approximately JQ first-order scattering coefficients are extracted for each of the $\frac{T}{2^J}$ time windows to collectively capture the dominant frequencies in the signal.

A set of second-order scattering coefficients is computed by applying a second wavelet transform ψ_{λ_2} to the modulus transformed first-order output, followed by smoothing with ϕ_J:

$$S_2x[t, \lambda_1, \lambda_2] = \left| |x \star \psi_{\lambda_1}| \star \psi_{\lambda_2} \right| \star \phi_J. \tag{4}$$

The approximately $\frac{J(J-1)}{2}Q^2$ second-order scattering coefficients capture interactions between different frequency bands for each of the $T' = \frac{T}{2^J}$ time windows. Once computed, the scattering coefficients are flattened and used as features in a deep learning model. The low-dimensional, stable features extracted through this process improve robustness to noise and small deformations in time-series tasks such as classification and regression. For further details, see Andreux et al. (2020).

To summarize, DWT in Kymatio applies multi-scale wavelet convolutions, modulus computation, and low-pass filtering to generate scattering coefficients. J controls time resolution and downsampling, while Q controls frequency resolution and the number of wavelets used within each frequency band. The 0th-order coefficients represent the global average, the 1st-order captures frequency energy, and the 2nd-order encodes frequency interactions. These features are then passed into dense neural networks for downstream learning tasks.

2.2 Deep Neural Network

Deep Neural Networks (DNNs) are a class of machine learning models characterized by multiple layers of interconnected neurons, which enable the modeling of complex, non-linear relationships within the data. The term "deep" refers to the presence of multiple hidden layers between the input and output layers. The closely related Multilayer Perceptrons (MLPs) are a class of feedforward DNNs consisting of an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to every neuron in the subsequent layer (Ivakhnenko, 1971; Rosenblatt, 1958). In this article, we use the terms DNN and MLP interchangeably,

In our proposed model architecture, following scattering transform, the scattering coefficients for all features across all time windows are subjected to an activation function (defined below), and subsequently flattened into a vector, denoted herein as x_{MLP}, and passed as input through a sequence of dense layers. The input data, x_{MLP}, consist of a collection of $x_{\text{MLP},i,k,s,t'}$, which denotes the activated scattering coefficient of frequency band s ($s = 1, \ldots, S$) for feature (or independent variable) k ($k = 1, \ldots, K$) at time t' ($t' = 1, \ldots, T'$ time windows) for individual i ($i = 1, \ldots, N$).

The values and strengths of the feature-specific scattering coefficients that pass through layers of a DNN are controlled by an *activation function*, expressed as $\sigma(\cdot)$. As part of the hyperparameter tuning process, we considered two plausible activation functions: Rectified Linear Unit (ReLU) and

Exponential Linear Unit (ELU). The ReLU (Nair & Hinton, 2010) activation function is defined as:

$$f(x) = max(0, x), \tag{5}$$

In ReLU, neurons with negative inputs always output zero, meaning some neurons stop learning (i.e., "dead neurons"). ELU introduces smooth non-linearity and reduces the severity of the "dead neurons" problem in ReLU by allowing small negative outputs. ELU activation function (Clevert et al., 2016) is defined as:

$$f(x) = \begin{cases} x, & \text{if } x > 0, \\ \alpha(\exp(x) - 1), & \text{if } x \leq 0. \end{cases} \tag{6}$$

Here, α controls the saturation value for negative inputs (default $\alpha = 1.0$).

The values in x_{MLP} are first passed through a dropout layer to allow for some initial feature selection, followed by the first dense layer of a deep neural network, and subjected to the activation function of choice. The activated output from this first dense layer, denoted as $a^{[1]}_{i,h}$, for person i and "neuron" h, can be obtained as:

$$a^{[1]}_{i,h} = \sigma(\sum_{k=1}^{F} \sum_{s=1}^{S} \sum_{t'=1}^{T'} W^{[1]}_{MLP,h,k,s,t'} x_{\text{MLP},i,k,s,t'} + b^{[1]}_{h}) \tag{7}$$

where h indexes a specific neuron ("hidden" or "latent" variable) in layer 1 ($h = 1, \ldots, H^{[1]}$), where $H^{[1]}$ is the hidden dimension of this layer, and we set $H^{[1]}$ to D to provide an initial layer of consolidation of scattering coefficients by the number of output variables. $W^{[1]}_{MLP,h,k,s,t'}$ is the weight (held invariant across individuals) for independent variable k from frequency band s at time window t' on neuron h; and $b^{[1]}_{h}$ is the intercept (also termed "bias") for the layer.

After the first dense layer, each subsequent hidden layer l outputs its corresponding activated output as:

$$a^{[l]}_{i,h} = \sigma(\sum_{h'=1}^{H^{[l-1]}} W^{[l]}_{h,h'} a^{[l-1]}_{i,h'} + b^{[l]}_{h}) \tag{8}$$

The number of layers, L, and $H^{[l]}$, the size of the hidden dimensions in layer l, are the hyperparameters to be tuned. For regularization purposes, a dropout layer is specified after each dense layer, in which a fraction of the activations is randomly set to zero during training. The dropout rate controls this fraction and is among the hyperparameters we tune.

Finally, a fully connected output layer maps the output of the last hidden dense layer to each of the $d = 1, \ldots, D$ dependent variable at each time point as:

$$\hat{y}_{i,d,t} = \sigma(\sum_{h=1}^{H^{[L]}} W^{[L]}_{d,t,h} a^{[L]}_{i,h} + b^{[L]}_{d,t}) \tag{9}$$

where $\hat{y}_{i,d,t}$ contains the prediction for the dth dependent variable for individual i at time t.

We considered and evaluated several strategies for tuning the number of layers and hidden dimension in each layer. One direct strategy considered was to remove all hidden layers and simply retain the output layer in (9), with $x_{\text{MLP},i,k,s,t'}$ as the input. A close alternative was to allow for only a single hidden layer (i.e., Equation (7) with as many "neurons" as the size of $x_{\text{MLP},i,k,s,t'}$ to benefit from the use of a dropout layer to reduce model complexity. These options entail minimal decisions to be made on hyperparameters, but did not yield good performance in our evaluations. Two options that yielded better performance were: (1) a partially confirmatory approach in which we removed all but the first hidden layer, in which $H^{[1]}$ was set to D, the number of dependent variables to be predicted; and (2) a "doubling-halving" structure. Using this doubling-halving procedure, we retained the first layer and only tuned the hidden dimension of the second layer, $H^{[2]}$ (through a hyperparameter optimization process to be described next). The number of hidden neurons in each subsequent hidden layer then followed a doubling and halving pattern. That is, in the first half of the layers, every subsequent layer was specified to have twice the number of hidden dimensions as the previous layer. For the second half of the layers, every subsequent layer was specified to have half of the number of hidden dimensions as the previous layer.

2.3 Hyperparameter Tuning

Hyperparameter tuning is a critical component of machine learning methods. One possible way to tune hyperparameters is to find a set of hyperparameters that minimizes a loss function of choice. We used the mean squared error averaged across K-folds (in which we set K to 5) resampling of the training data as a loss function. Hyperopt, a Tree-structured Parzen Estimators (Bergstra & Bengio, 2012; Snoek et al., 2012), is used to optimize the following hyperparameters over a specified number of trials and epochs per trial. The search space determining possible ranges of values of the hyperparameters to be optimized was specified as: hidden dimension ($H^{[1]}$): 1 to 24; number of layers (L): 3 to 6; dropout rate (dropout_rate): 0.0 to 0.5; activation function (activation): ReLU or ELU; learning rate (learning_rate): log(−3) to log(−1); and L2 regularization rate (l2_reg): log(−3) to log(−1). For the partially confirmatory approach, the number of layers and size of the hidden dimensions were not tuned but determined a priori.

2.4 Interpretations of Feature Importance

DNNs and other machine learning models are generally highly underidentified models. Although these models can be arbitrarily made more complex, it is critical to select models with good performance in predicting new independent data sets. Viton et al. (2020) proposed using a feature activation heatmap to facilitate interpretations of feature importance in using convolution neural networks (CNNs) to perform cross-sectional classification. We extended the graphical tool proposed by these researchers that pools information across time to perform cross-sectional binary classification (mortality outcome), to allow longitudinal, person- and time-specific predictions of a continuous outcome by pooling data, weights, and hyperparameter settings across multiple participants. We also integrated hyperparameter tuning using hyperopt to explore the "optimal" hyperparameter settings (e.g., dropout rate) to be used in the scattering transform and DNNs.

We extracted the weighted activation, namely, the inputs to layer 1 in (7), as:

$$\text{Weighted Activation}_{i,k,s,t'} = W^{[1]}_{k,s,t'} x_{\text{MLP},i,k,s,t'}, \tag{10}$$

by setting $H^{[1]} = D = 1$ in our example. Multiplication of $x_{\text{MLP},i,k,s,t'}$ with $W^{[1]}_{k,s,t'}$ conveys some information concerning the directionality of the influence of each scattering coefficient in $x_{\text{MLP},i,k,s,t'}$ on subsequent, and eventually, the final output layer. The absence of individual index i in $W^{[1]}_{k,s,t'}$ serves to highlight our constraints for person-invariant weights.[1]

Summing across frequency bands provides the feature importance value for feature k and individual i across all frequency bands over the T' time windows as:

$$\text{Person-Specific Feature Importance}_{i,k,t'} = \sum_{s=1}^{S} \text{Weighted Activation}_{i,k,s,t'} \tag{11}$$

In some scenarios, it may be beneficial to sum over a subset of frequency bands, such as the top three bands with the largest scattering coefficients, as measured either by their maximum value or by their norm (e.g., l_2-norm) over time windows.

In a similar vein, averaging across the weighted activated values across individuals provides some insights on the average importance of each feature at each time window across all individuals as:

$$\text{Sample Feature Importance}_{k,t'} = \frac{1}{N}\sum_{i=1}^{N} \text{Person-Specific Feature Importance}_{i,k,t'} \tag{12}$$

3. Illustrations with Simulated Data

3.1 Constant Frequency

As a simple simulation, we simulated a cosine time series with $2^8 = 256$ time points at a constant frequency of .1 Hz (i.e., a period of 10 seconds to complete one cycle), with a sampling rate of 1 sample per second (see time series plot in Figure 1(A)). We specified $J = 3$, $Q = 2$.

The central frequencies of the scattering transform's bandpass filters can be computed explicitly using J and Q. In Kymatio, the central frequencies of the wavelets are given by: (L. Cohen, 2020; Destouet et al., 2021; Lostanlen et al., 2021; Mallat, 1999)

$$f_c^{(j,q)} = \frac{f_s}{2^{J-j+q/Q+1}} \tag{13}$$

where: f_s is the sampling frequency (typically set to 1 in scattering transform, normalized); j is the scaling index ($j = 1, \ldots, J$); q is the wavelet index within a time window ($q = 0, 1, Q-1$), with low and upper limits of the frequency band given by: $f_c^{(j,q)} \cdot 2^{\pm 1/(2Q)}$ (A. Cohen & Daubechies, 1993; Selesnick, 2011)

The weighted activation heatmap portraying only the zeroth and first-order scattering coefficients is shown in Figure 1(B). The heatmap highlighted the sustained high scattering coefficient magnitude of the dominant frequency (of approximately 0.1 Hz) that persisted across all time windows, reflecting the constancy of this frequency in this illustration.

3.2 Change Point in Frequency

The second illustration serves to demonstrate a scenario in which a low-frequency sine wave (where $T = 1000$) is interrupted by a high-frequency transient at $t = 500$ (see Figure 2(A)) . The first half of the signal consists of a low-frequency cosine wave with a frequency of 0.05 Hz. The second half contains a high-frequency component with a frequency of 0.2 Hz.

The weighted activation heatmap in Figure 2(B) shows the scattering coefficients' strengths over time. The sudden transition to a faster frequency $t = 500$ is reflected in the weighted activation heatmap as a sudden change in the dominant frequency band at approximately $t' = 64$.

3.3 Feature Importance Using Weighted Activation Map

In this illustration, we generated time series data for 15 hypothetical participants contaminated with Gaussian noise, as dependent on three (features 4-6) out of 6 possible features that comprised structured sinusoidal signals during specific time spans (see Figure 3). We tested the proposed procedures of splitting of the 15 participants into a training set and a test set, and optimization of the hyperparameters through Hyperopt over 15 trials with 30 epochs each.

Plots of the scattering activations by frequency band, the maximum scattering coefficients for each feature within each time window; and the sample feature importance map based on Equation (12) are shown in plots (A)–(C), respectively, in Figure 4. These plots indicated that the proposed graphical

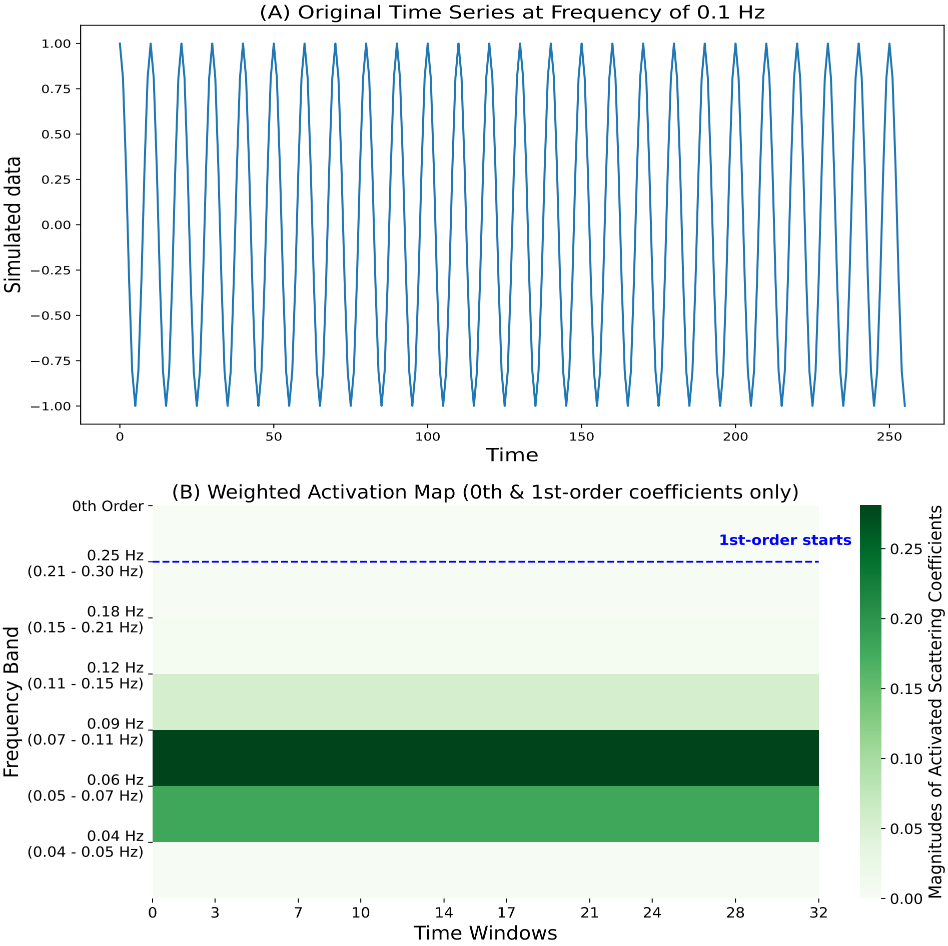


Figure 1. Simulated data with a constant frequency throughout the entire time span.

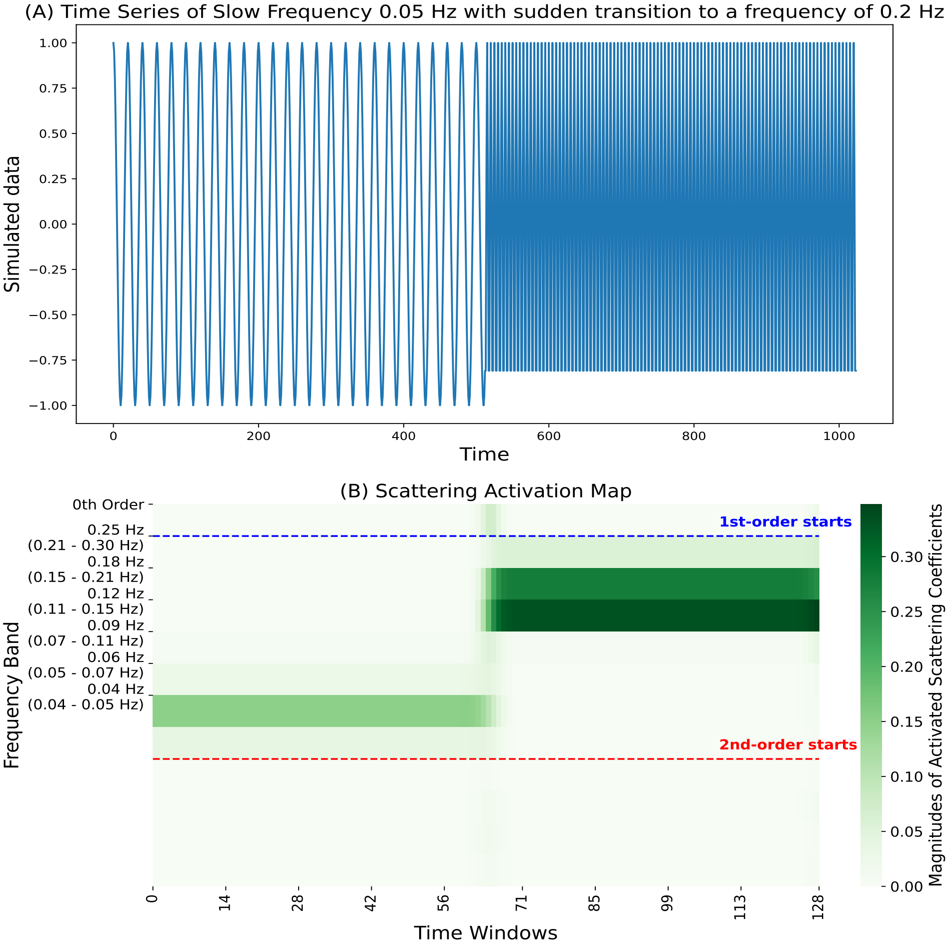


Figure 2. Simulated data in which a time series of a constant frequency shows sudden transition to a faster frequency at the mid-length of the time series.

tool (in plot C) could capture the localized, time-varying influence of each of the features, even though some of the influence might be attentuated (e.g., from feature 6). The *partially confirmatory* and *doubling-halving* structures both yielded similar R^2 values. The partially confirmatory structure was thus preferred for reasons of parsimony. The R^2 values from using the estimated model to predict self-reports for participants in the training and test set (note that the model was *not* re-estimated after the estimation with training data) were .91 and .89, respectively, suggesting reasonable generation of training results from the partially confirmatory model to independent participants in the test set.

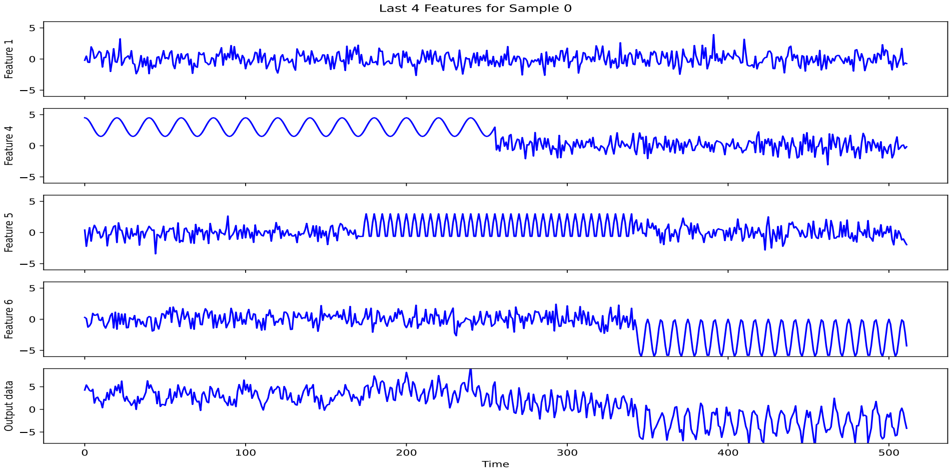


Figure 3. Simulated data in which a time series with Gaussian noise was influenced directly by three selected features with distinct frequencies during targeted time windows. Only one of the three remaining spurious features was plotted.

4. Illustrative Empirical Example with Affect Forecasting: Multiple Features with Time-Varying Influence

In this study, we forecast self-report data from a group of $n = 160$ participants from part of the Affective Dynamics and Individual Differences (ADID). Participants were asked to provide continuous self-reports of their perceived affect intensity levels while watching slide shows consisting of negative stimuli from the International Affective Picture System (IAPS; Lang et al., 2005) following (1) a neutral movie, (2) a low positive affect (PA) movie and (3) a high PA movie. Their physiological data were collected concurrently. Only data from the negative slide show following the low PA (LPA) induction procedure were used. All data were aggregated over every 50 milliseconds (msec) in all subsequent analysis, and followed the data pre-processing procedures adopted in a previously published pilot study (Yang & Chow, 2010). The following *within-person standardized* physiological signals collected concurrently as the self-reports were used as potential features: electrodermal activity (EDA), facial EMG activities in two major muscle groups, corrugator supercilii (CS, associated with frowning) and zygomaticus major (ZM, associated with smiling; Cacioppo et al., 1986), ECG RR-intervals, heart rate, skin temperature, and normative slide valence and arousal ratings (stimuli-specific ratings provided by the IAPS developers; Lang et al., 2005) .

We performed pairwise exploratory wavelet coherence analysis using the R package, *WaveletComp* (Rösch & Schmidbauer, 2016). For each participant, we examined the pairwise coherence (i.e., cross-correlation in the frequency domain) between the participant's self-reports and each physiological signal in turn to reveal potential frequency scales that show substantial associations in the frequency domain. A plot of the time series of slide valence ratings, skin temperature, and self-reports for one selected participant is shown in Figure 5(A). As shown in the plot of wavelet coherence (see Figure

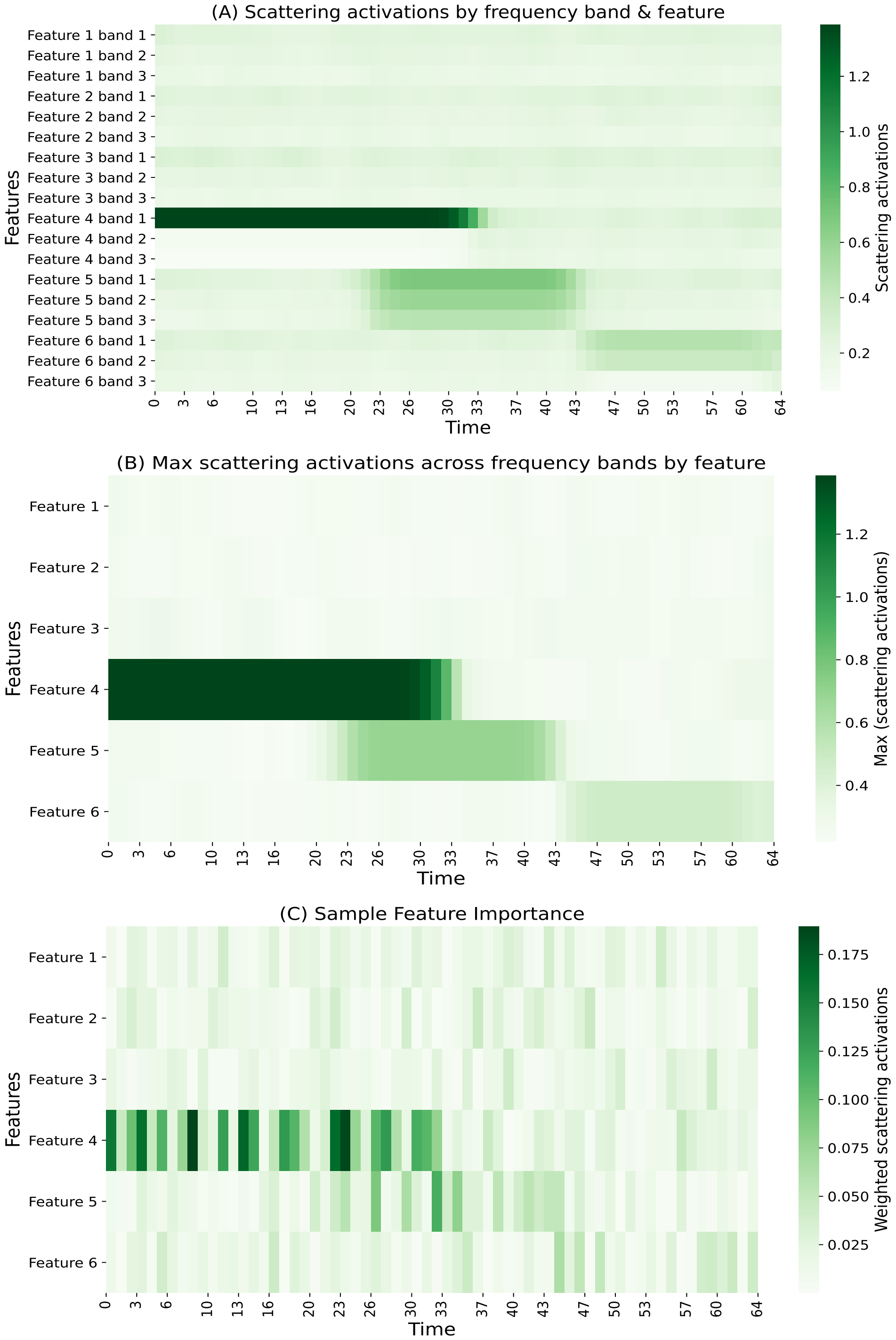
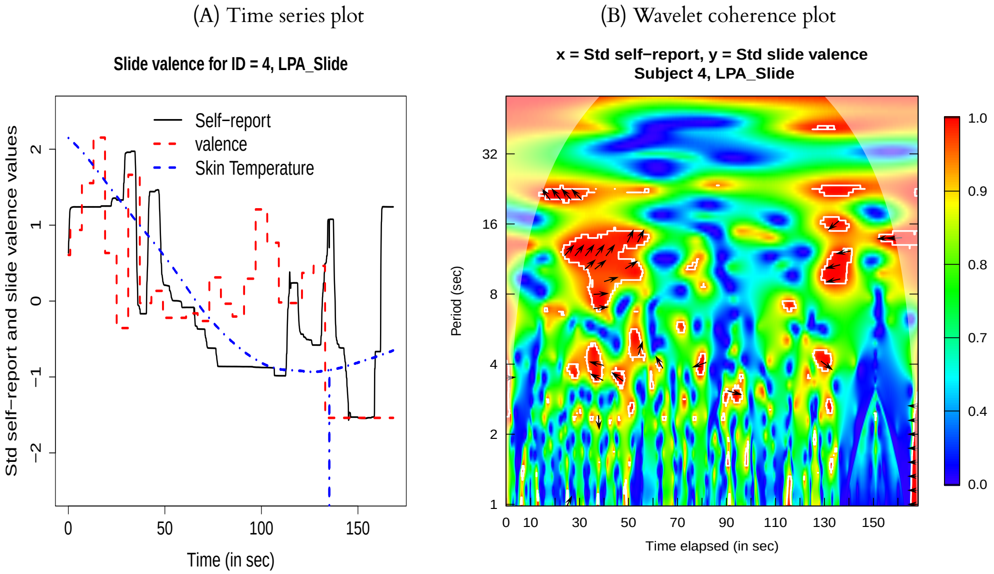


Figure 4. Plots of: (A) the scattering activations by frequency band; (B) the maximum scattering coefficients for each feature within each time window; and (C) the sample feature importance map.

5(B)), statistically significant in-phase (i.e., synchronous, in which peaks align with peaks) coherence was found between individuals' self-report levels and the normed slide valence ratings around t = 20 and 60 in the 8 to 16-second frequency bands (shown in Figure 5(B) as arrows pointing from left to right, coinciding with the alignment between the peaks and valleys of the two series during this time span in Figure 5(A). However, the association became attenuated at later time points, with ongoing changes in the slide valence as part of the experimental design of the study, but little corresponding changes in the participant's self-reports. Thus, the two processes fluctuated between in-phase (i.e., synchronously, with arrows pointing from left to right) and anti-phase (asynchronously, with arrows pointing from right to left) at different points of the experiment, but neither patterns persisted throughout the study span^2.

We used the proposed deep scattering transform model to predict the participants' self reports over 3345 time points, with every time step corresponding to 50 milliseconds. We split data from the participants equally into a training set and a test set with 80 participants each. Mean squared error as aggregated across 5 validation folds was used as the objective functioning for optimizing hyperparameters with Hyperopt. Based on our exploratory wavelet analysis and our experimental design, we expected some dominant frequencies to emerge in the range of 5 seconds (0.2Hz). Kymatio normalizes frequencies by setting the sampling rate to 1 Hz (dimensionless frequency). Thus, with a sampling rate of 1/.05 second = 20 Hz, rescaling to Kymatio's default sampling rate of 1, we expect to see some relative frequencies in the range of 0.2/20 = 0.01 in Kymatio's frequency representation. This motivated our choice to set J and Q to 4 and 3, respectively, to capture frequencies in this approximate range. We computed scattering coefficients separately for the following physiological signals and used them as features to predict the participants' self-reports.


(A) Time series plot

(B) Wavelet coherence plot

Figure 5. Plots of (A) experimentally-induces changes in slide valence experienced by one participant during the negative emotion induction procedure and the participant's corresponding fluctuations in self-reports and skin temperature; and (B) wavelet coherence between that participant's self-reports and slide valence.

The scattering activation map depicting the features' importance across all participants is shown in Figure 6(A). The results indicated that the normed slide valence and arousal ratings of the slides were among the key features in predicting fluctuations in the participants' self-reports. ECG RR intervals showed some initial importance, but their importance was transient and was observed primarily in the earlier time windows. Using scattering coefficients across the physiological signals helped explain approximately 10% of the variability in the self-reports in the training (R^2 = .11) and test (R^2 = .13) data sets relative to using the mean of each participant's time series of self-reports alone. This demonstrated the considerable disconnect between individuals' self-reports and their underlying physiological changes, and the highly time-localized characteristic of the associations between individuals' subjective perceived affect intensity and their physiological responses. Nevertheless, the improvement in R^2 from the training to test data underscored robustness of prediction results using scattering transforms when applied to new independent samples.

The low to moderate R^2 values obtained were related in part to the substantial differences between individuals in the associations between self-reports and physiological data. Considerable heterogeneity was observed in the person-specific feature importance maps (see Figures 6 (B)-(C) for examples). The person-specific feature importance map of participant 4 (see Figure 6(B)) underscores the concordance between the slow-varying downward declines in this participant's skin temperature and self-reports over time, as reflected also in Figure 5(A). As another example, the person-specific feature importance map of participant 21 (see Figure 6(C)) highlighted the surprising importance of activities in the participant's Zygomaticus region, which typically serves as a marker of smile, joy, or in some scenarios, expression of smile with mixed emotions (e.g. smiles with disdain, or under bittersweet memories). Such differences underscored the need to balance the modeling of group and individual dynamics despite the challenges of limited sample sizes.

5. Discussion

In this paper, we presented a deep neutral network architecture that integrates scattering transforms and hyperparameter tuning via K-fold cross-validation, as well as graphical display to elucidate the time-varying importance of different frequency components of experimental stimuli and multiple physiological signals in influencing individuals' perceptions of their affective arousal levels.

One limitation of this study stems from the mismatch between the frequencies of self-reports and the physiological predictors used. Most of the physiological signals considered in this study were characterized by very fast frequencies relative to those associated with the self-reports. Fluctuations in human self-reports are naturally limited in temporal granularity by factors such as the reaction time of the participants, and the participants' emotional expressivity (Feldman Barrett et al., 2001; Gross & John, 1997). Consistent with findings in the affect literature highlighting the discrepancies between subjective reports and the physiology of affect, the physiological characteristics considered in this study account for approximately 10% of the variability in self-reports compared to the use of static, subject-specific means. The high non-stationarity of the data poses challenges even for the wavelet-based methods used in this study: the frequent and ongoing shifts in associations among the predictors and self-reports over time provide insufficient data for identification of meaningful predictors with consistently strong effects across participants. The stochastic, noisy nature of the data further complicates interpretation and extraction of meaningful patterns. Other pre-processing techniques such as smoothing may need to be used to improve the robustness of the feature identification process.

Future research should explore ways to consolidate and account for heterogeneity in dominant frequency patterns across different features and participants. Individual variability may lead to inconsistencies in extracted frequency components, suggesting the need for methods that adaptively align or cluster frequency patterns across subjects. For feature importance, the SHAP (SHapley Additive exPlanations)(Lundberg & Lee, 2017) and LIME (Local Interpretable Model-agnostic Explanations)(Ribeiro et al., 2016) methods have been used to interpret deep learning models. In this study, we employed activation maps instead of feature-wise methods, such as SHAP or LIME, due to computational costs. SHAP computes the marginal contribution score of each feature by considering all possible coalitions of features. In our case, the features are the wavelet coefficients from the different frequency bands across the time horizon, resulting in more than 1,000 features. Hence,

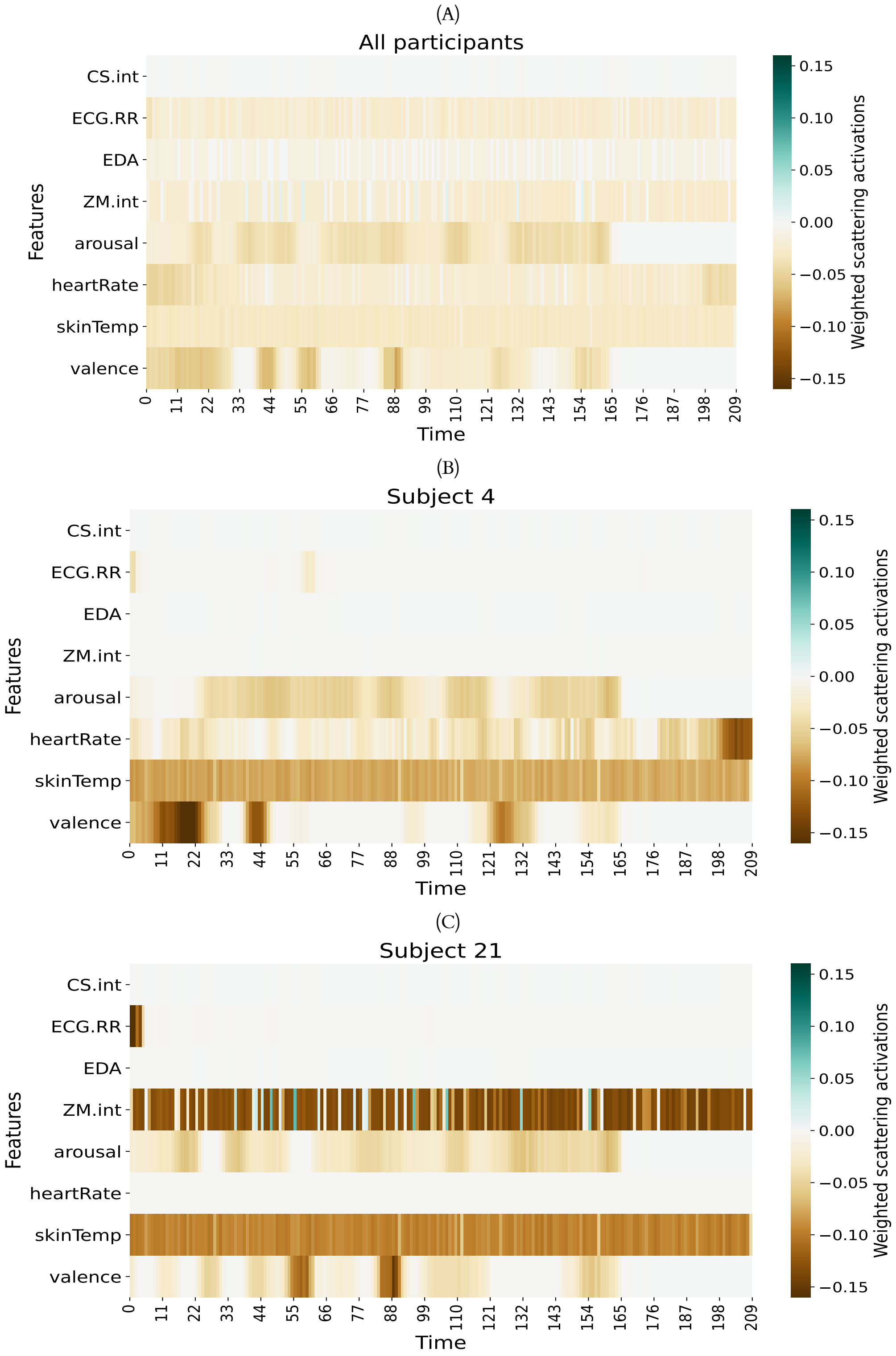


Figure 6. (A) Sample feature importance map across all participants in the training set from the ADID study. (B)-(C): Person-specific feature importance maps from two selected participants.

it becomes computationally expensive and time-consuming to compute all the possible coalitions of 1000 features. In comparison, LIME is faster than SHAP, but we still need to retrain a new explainable model for each data instance. In contrast, activation maps can be calculated directly from the weights in the trained neural network. For future research, instead of using explanability tools, the neural network model can be replaced with neural additive models, which are based on constraining a neural network model onto a generalized additive model (Agarwal et al., 2020). These additive models are inherently interpretable and can help visualize the decision making of the neural net. Another approach can be learning interpretable embeddings from the signals using an encoder (Alvarez-Melis & Jaakkola, 2018). Extensions to accommodate multiple outcomes and missing data are also warranted. These advances would improve the applicability, robustness, and interpretability of frequency-based machine learning methods in social and behavioral sciences.

Funding Statement Funding for this study was provided by National Institutes of Health grant U24AA027684, National Science Foundation grants DUE-2417294, the National Center for Advancing Translational Sciences under UL1TR002014-06, and the National Institute of Diabetes, Digestive & Kidney Diseases under U01DK135126 and R01DK134863.

Competing Interests The authors declare that there are no conflicts of interest.

Notes

[1] We also considered replacing $W^{[1]}$ with propagated weights from the final fully connected layer to the first layer to obtain a feature important weight, $IN^{[L]}_{i,k,s,t'}$, through repeated matrix multiplication of the weight matrices from the last (L) layer to the first as:

$$\text{Backpropagated } \mathrm{W}^{[L]}_{k,s,t'} = (W^{[L]} \cdot W^{[L-1]} \cdots W^{[1]})_{k,s,t'} \tag{14}$$

where Backpropagated $\mathrm{W}^{[L]}_{k,s,t'}$ denotes the backpropagated weight for the k feature, s scattering coefficient at time window t', extracted from the (k, s, t')th element after the series of backpropagated weight matrix multiplication. However, due to the multilayer and fully connected nature of DNNs, the backpropagated weights were found to yield overly diffuse portrayal of the importance across multiple features in our preliminary simulations, and at times, spurious features that contained "spilled-over" influence from truly important features.

[2] Arrows pointing northeast from left to right, and southwest from right to left both suggest that series 1 (self-reports in this case) is "leading" series 2 (valence in this example); with flat arrows suggesting no clear lead-lag order. However, in this case, the lead-lag directionality suggested by the exploratory wavelet coherence analysis might reflect arbitrary rises and declines in the participant's ratings as they transitioned into the beginning and end of a slide show.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org].

Agarwal, R., Frosst, N., Zhang, X., Caruana, R., & Hinton, G. E. (2020). Neural additive models: Interpretable machine learning with neural nets. *CoRR*, *abs/2004.13912*. https://arxiv.org/abs/2004.13912

Alvarez-Melis, D., & Jaakkola, T. S. (2018). Towards robust interpretability with self-explaining neural networks. *CoRR*, *abs/1806.07538*. http://arxiv.org/abs/1806.07538

Andreux, M., Angles, T., Exarchakis, G., Leonarduzzi, R., Rochette, G., Thiry, L., Zarka, J., Mallat, S., andèn, J., Belilovsky, E., Bruna, J., Lostanlen, V., Chaudhary, M., Hirn, M. J., Oyallon, E., Zhang, S., Cella, C., & Eickenberg, M. (2020). Kymatio: Scattering transforms in python. *Journal of Machine Learning Research*, *21*(60), 1–6. http://jmlr.org/papers/v21/19-047.html

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2), 281–305.

Bigger, J. T. J., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Schneider, W. J., & Stein, P. K. (1995). Rr variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction. *Circulation*, *91*(7), 1936–1943.

Bruna, J., & Mallat, S. (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1872–1886.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*(2), 260–268.

Chow, S.-M., Hamaker, E. J., Fujita, F., & Boker, S. M. (2009). Representing time-varying cyclic dynamics using multiple-subject state-space models. *British Journal of Mathematical and Statistical Psychology*, *62*, 683–716.

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, *5*(2), 208–225. https://doi.org/10.1037/1528-3542.5.2.208

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). *Fast and accurate deep network learning by exponential linear units (elus)* [preprint at: https://arxiv.org/abs/1511.07289].

Cohen, A., & Daubechies, I. (1993). Orthonormal bases of compactly supported wavelets iii. better frequency resolution. *SIAM Journal on Mathematical Analysis*, *24*(2), 520–527.

Cohen, L. (2020). Time-frequency analysis: What we. *Landscapes of Time-Frequency Analysis: ATFA 2019*, 75.

Das, J. N., Ji, L., Shen, Y., Kumara, S., Buxton, O. M., & Chow, S. (2025). Performance evaluation of a machine learning-based methodology using dynamical features to detect nonwear intervals in actigraphy data in a free-living setting. *Sleep health*.

Destouet, G., Dumas, C., Frassati, A., & Perrier, V. (2021). Wavelet scattering transform and ensemble methods for side-channel analysis. *Constructive Side-Channel Analysis and Secure Design: 11th International Workshop, COSADE 2020, Lugano, Switzerland, April 1–3, 2020, Revised Selected Papers 11*, 71–89.

Feldman Barrett, L., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, *15*(6), 713–724.

Gross, J. J., & John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings and behavior. *Journal of Personality and Social Psychology*, *72*(2), 435–448.

Imambi, S., Prakash, K. B., & Kanagachidambaresan, G. (2021). Pytorch. *Programming with TensorFlow: solution for edge computing applications*, 87–104.

Ivakhnenko, A. G. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-1*(4), 364–378. https://doi.org/10.1109/TSMC.1971.4308320

Knapova, L., Cho, Y. W., Chow, S.-M., Kuhnova, J., & Elavsky, S. (2024). From intention to behavior: Within-and between-person moderators of the relationship between intention and physical activity. *Psychology of Sport and Exercise*, *71*, 102566.

Kuppens, P., Sheeber, L. B., Yap, M. B. H., Whittle, S., Simmons, J., & Allen, N. B. (2012). Emotional inertia prospectively predicts the onset of depression in adolescence. *Emotion*, *12*, 283–289.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Instruction manual and affective ratings*. University of Florida, Technical Report A-6, The Center for Research in Psychophysiology.

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, *11*, 129–141.

Liu, L., Wu, J., Li, D., Senhadji, L., & Shu, H. (2018). Fractional wavelet scattering network and applications. *IEEE Transactions on Biomedical Engineering*, *66*(2), 553–563.

Liu, Z., Yao, G., Zhang, Q., Zhang, J., Zeng, X., et al. (2020). Wavelet scattering transform for ecg beat classification. *Computational and mathematical methods in medicine*, *2020*.

Lostanlen, V., Cohen-Hadria, A., & Bello, J. P. (2021). One or two frequencies? the scattering transform answers. *2020 28th European Signal Processing Conference (EUSIPCO)*, 2205–2209.

Lu, Z.-H., Chow, S.-M., Sherwood, A., & Zhu, H. (2015). Bayesian analysis of ambulatory cardiovascular dynamics with application to irregularly spaced sparse data. *Annals of Applied Statistics*, *9*, 1601–1620. https://doi.org/10.1214/15-AOAS846

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, *65*(10), 1331–1398.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 807–814. https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf

Oyallon, E., Mallat, S., & Sifre, L. (2013). Generic deep networks with wavelet scattering. *arXiv preprint arXiv:1312.5940*.

Ram, N., Conroy, D. E., Pincus, A. L., Lorek, A., Rebar, A., Roche, M. J., Coccia, M., Morack, J., Feldman, J., & Gerstorf, D. (2014). Examining the interplay of processes across multiple time-scales: Illustration with the intraindividual study of affect, health, and interpersonal behavior (isahib). *Research in human development*, *11*(2), 142–160.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Rösch, A., & Schmidbauer, H. (2016). Waveletcomp 1.1: A guided tour through the r package. *URL: http://www. hsstat. com/projects/WaveletComp/WaveletComp_guided_tour. pdf*.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Selesnick, I. W. (2011). Wavelet transform with tunable q-factor. *IEEE transactions on signal processing*, *59*(8), 3560–3575.

Sepúlveda, A., Castillo, F., Palma, C., & Rodriguez-Fernandez, M. (2021). Emotion recognition from ecg signals using wavelet scattering and machine learning. *Applied Sciences*, *11*(11), 4945.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 2951–2959.

Suh, J. H., Kumara, S. R., & Mysore, S. P. (1999). Machinery fault diagnosis and prognosis: Application of advanced signal processing techniques. *CIRP Annals*, *48*(1), 317–320. https://doi.org/https://doi.org/10.1016/S0007-8506(07)63192-8

Viton, F., Elbattah, M., Guérin, J.-L., & Dequen, G. (2020). Heatmaps for visual explainability of cnn-based predictions for multivariate time series with application to healthcare. *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, 1–8.

Yang, M., & Chow, S.-M. (2010). Using state-space model with regime switching to represent the dynamics of facial electromyography (EMG) data. *Psychometrika: Application and Case Studies*, *74*(4), 744–771.