Modeling Regularity and Predictability in Human Behavior from Multidimensional Sensing Signals and Personal Characteristics

Jingyi Gao
University of Virginia, USA
etc6bd@virginia.edu

Runze Yan
University of Virginia, USA
ry4jr@virginia.edu

Afsaneh Doryab
University of Virginia, USA
ad4ks@virginia.edu

***Abstract*—Forecasting behavioral patterns can help humans comprehend their habits and tendencies, allowing them to intervene before problems arise. Among the numerous techniques of modeling human behavior, cyclic modeling can better identify recurring patterns, providing regularity and predictability in human behavior. However, existing approaches to cyclic modeling ignore the effects of human characteristics, such as resilience and coping abilities, which substantially influence the stability of behavior. To explore the value of adding such information in behavior modeling and prediction, we introduce a transformer-based architecture with an advanced attention mechanism and parallel operation that models regularity in human behavior from multidimensional sensing signals and predicts future behavior patterns. The architecture further transforms static human characteristics metadata into dynamic time series that conform to behavioral patterns and serve as covariates to assist prediction.**

Our experiments with a wearable dataset indicate that our architecture 1) is more accurate in forecasting human behavior patterns than current time-series models and 2) enables us to investigate the impact of human characteristics on behavior patterns. Specifically, we analyze the influence of resilience and coping strategies on behavioral regularity and predictability. We show that supplementing bio-behavioral wearable data with resilience and coping scores in the forecasting model increases the convergence speed of the model and decreases prediction loss.

***Index Terms*—Resilience, Coping, Human Behavior, Cyclic Time Series**

I. Introduction

Numerous human behavior exhibit regularity, the most prominent being the circadian rhythm or sleep-wake cycle, based on a 24-hour interval. Investigating cyclic behavior is beneficial for detecting and preventing various health conditions, such as mood disorders [1] and cancer [2]. Modeling human behavior patterns can facilitate the regularization of future trajectories and provide effective health interventions [3]. With the advancement of mobile and wearable sensing technology, more bio-behavioral signals have been utilized to capture daily human behavior [4]. However, most studies have focused on detecting abnormalities [5], such as sudden changes in daily activities or in specific physical function data; and their approaches to identifying and forecasting behavior issues using mobile sensing technologies do not consider individuals' subjective characteristics, such as resilience and coping, which can greatly influence the way behaviors are performed.

In this paper, we introduce a transformer-based approach to model regularity and predictability in human behavior that considers human characteristics as covariates to enhance the model's performance. We design an architecture that uses bio-behavioral data from wearable devices to generate cyclic time series directly, i.e., data streams reflecting the regularity of people's activities. We then build a forecast model with a convolutional transformer that uses a local attention mechanism to focus on capturing changes and to implement parallelism to increase the speed of execution substantially. We run experiments with a wearable dataset and compare several different time series models to predict human future behavior patterns with circadian cycles. We find that the results obtained from the transformer forecasting model most closely approximate the real-world data.

In addition, we evaluate our behavior modeling approach by incorporating resilience and coping scores in the transformer model as covariates. A person's resilience is a capacity to withstand or bounce back from a significant threat to their stability, viability, or development [6]. Coping is inherently a behavior that serves as a conscious or unconscious strategy used to lessen unpleasant feelings or handle difficult or essential life situations [7]. Past literature [8] has confirmed that regular activity is associated with human resilience and coping strategies. They impact each other and can assist individuals in adapting to changes or adversities that occur in their lives to achieve stability in life.

To incorporate this static information into our behavior forecasting architecture, we develop a transforming module to convert those scores into dynamic time series. We show that including this information as covariates improves our transformer-based forecasting model, as evidenced by a decrease in the model's loss value and a faster training convergence. Overall, our contributions are as follows:

- We propose a cyclic time series forecasting architecture, which is the first to forecast human behavior patterns based on a Transformer. Results indicate that our architecture significantly outperforms existing time series forecasting models.
- We introduce a module that transforms static resilience and coping into dynamic series that conform to behavioral tendencies. Incorporating the transformed series

1946-0759/23/$31.00 ©2023 IEEE

DOI 10.1109/ICMLA58977.2023.00045

as covariates in the forecasting architecture allows the predictions to accurately reflect the actual behavioral patterns.

- We investigate the relationship between resilience, coping, and the regularity of human behavior. We validate that high resilience indicates more consistent behavior patterns. However, we do not observe a consistent relationship between coping strategies and behavior regularity.
- We analyze the degree to which different scores of resilience and coping affect the predictive power of human cyclical behavior. We report on the variable effect of these factors and their impact on behavior forecasting.

II. Related Work

Several research studies have improved standard machine learning-based algorithms to address the human activity recognition (HAR) challenge. Ordóñez and Roggen [9] constructed a CNN-LSTM framework that integrates homogeneous and multimodal sensor modalities. Yin et al. [10] suggested a two-phase anomaly detection approach using a single-class SVM to filter out most regular activities, followed by Kernel Nonlinear Regression (KNLR) detection of the remaining abnormal activities to balance the detection rate and false alarm rate. Duond et al. [11] proposed a switching hidden semi-Markov model (S-HSMM), a two-layered version of the hidden semi-Markov model (HSMM), to learn and recognize human activities of daily living (ADL) while eliminating the necessity for pre-segmented training data. A Hierarchical Context Hidden Markov Model (HC-HMM) [12], another variation of the HMM, has achieved a 100% true-positive alarm when describing the behavior of the elderly. These earlier work opened up good ideas, but some of the recent work is also very powerful. Although these models can measure shifts in day-to-day activity behaviors and parameters within the physical body, they are limited in their ability to express and summarize the regularity of human activity. Moreover, they disregard the potential influence of time and period on human behavior.

Research studies have also modeled cyclic behavior to find changes that cover general human trends. Huang et al. [13] suggested a Hidden Markov Model combining circadian oscillators to analyze and monitor participants' circadian rhythms in chronobiology and chronotherapeutic health research using daily physical activity data. Pierson et al. [14] introduced CyHMMs, a cyclic hidden Markov model that uses a set of multivariate time series as input and extrapolates a discrete latent cycle state from the measured time series. In addition to these, Hadj-Amar et al. [15] proposed the Bayesian observation method of oscillatory behavior based on Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithms and non-smooth time series, which can use human skin temperature data to detect sleep duration throughout the day and identify nocturnal oscillatory behavior. Later, an RJMCMC sampler developed by Taylor et al. [16] also further identified change-point events with periodicity to ascertain daily patterns; it performed well at longer cycle lengths. Even though each of the above studies deals with the periodicity and variability of the data, they are focused on calculating the cycle length. Our work is the first to concentrate on cyclic activity pattern forecasting.

Transformer modelshave also been used for behavior modeling. For example, BEHRT [17], an advanced Bidirectional Encoder Representations from Transformers (BERT)architecture was introduced to predict future diagnoses based on electronic health records (EHR). BEHRT is an interpretable personalized model trained using clinic visit data, allowing for predicting several diseases. Moreover, Zhang et al. [18] used the IMU fusion block in combination with a convolutional transformer to solve the problem of inadequate multi-sensor modal fusion and achieved better performance in human activity recognition and classification compared with other state-of-art methods. In this paper, we utilize the self-attention mechanism in the transformer model to help with time series forecasting, which is a novel attempt to use Transformer in predicting behavioral patterns.

III. Architecture of Model

The Transformer-based time series architecture consists of three sub-components, including cyclic modeling, covariate transformation, and Transformer prediction, which is visualized in Figure 1. First, we use the Cosinor method to simulate cyclic patterns with collected mobile sensing data in the cyclic model generation component. Second, we convert the static data into dynamic streams in the covariate transformation component. Finally, we apply the transformer model to the generated cyclic data with the time-dependent covariates to forecast behavior in the Transformer prediction component.

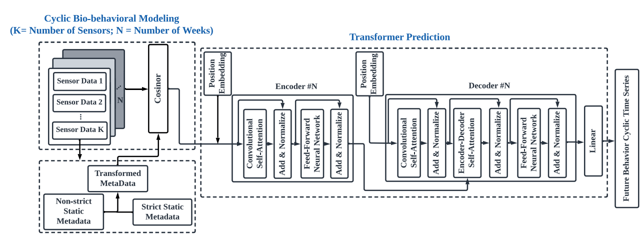

Fig. 1: Architecture of Transformer Cyclic Time Series Forecasting Model with Covariates

A. Cyclic Time Series Generation

This phase aims to fit a cosine model to the passive time series data and represent participants' overall behavior regularity. Our experiment focuses on 24-hour circadian behavior, but the model is flexible to any periodic interval. We adapt **Population-Mean Cosinor** method [19] to generate the cyclic patterns for a series of sensor streams with fixed time granularity. Population-Mean Cosinor is a statistical procedure that implements sine regression to signal streams. The basic Cosinor is modeled as [19]:

$$Y(t) = M + \mathrm{A}\cos(\frac{2\pi t}{\tau} + \phi) + e(t) \tag{1}$$

for $Y(t)$ is time series, M is MESOR, A is amplitude, ϕ is acrophase and τ is period.

The population-mean Cosinor merges the pattern of multiple data streams. Each sensor series is evaluated using a single-component Cosinor to produce estimates of a vector $\hat{u} = \{\widehat{\mathrm{M}}_{\mathrm{i}}, \widehat{\beta}_{1\mathrm{i}}, \widehat{\gamma}_{1\mathrm{i}}, \widehat{\beta}_{2\mathrm{i}}, \widehat{\gamma}_{2\mathrm{i}}, \ldots, \widehat{\beta}_{\mathrm{pi}}, \hat{\gamma}_{\mathrm{pi}}\}$ for $\mathrm{i} = 1, 2, \ldots, \mathrm{k}$ and p variates in regression.

Since the sample sizes for all sensors in this research are the same, the population estimates are unweighted averages of the individual sensor parameters [19]:

$$\widehat{u}^* = \sum_j \frac{\widehat{u}_j}{k} \tag{2}$$

for $j = 1, 2, \ldots, k$. Thus the Equation 1 could be written as:

$$Y^*(t) = \widehat{M}^* + \widehat{\mathrm{A}}^* \cos(\frac{2\pi t}{\tau} + \widehat{\phi}^*) + e(t) \tag{3}$$

where $Y^*(t)$ is the generated population cyclic time series.

We aim to investigate the rhythmic changes in behavior across weeks. Figure 2 shows our study's cyclic time series generation structure. For each series, we set T as a week (7 days), t ranging from 1 to 168 hours, and $\tau = 24h$ as the circadian period. We then fit K different passive feature values for each participant as input to model on N weeks. For each week, the Population-Mean Cosinor method generates a cyclic time series formed by all sensor data.

Thus we get a set of amplitudes $\{\widehat{\mathrm{A}}^* : \widehat{\mathrm{A}}_1^*, \widehat{\mathrm{A}}_2^*, \widehat{\mathrm{A}}_3^*, \ldots, \widehat{\mathrm{A}}_N^*\}$, a set of acrophases $\{\widehat{\phi}^* : \widehat{\phi}_1^*, \widehat{\phi}_2^*, \widehat{\phi}_3^*, \ldots, \widehat{\phi}_N^*\}$, and a set of mesors $\{\widehat{\mathrm{M}}^* : \widehat{\mathrm{M}}_1^*, \widehat{\mathrm{M}}_2^*, \widehat{\mathrm{M}}_3^*, \ldots, \widehat{\mathrm{M}}_N^*\}$ for each person.

Finally, using Equation 1, the model generates $\{\widehat{\mathrm{Y}}^* : \widehat{\mathrm{Y}}_1^*, \widehat{\mathrm{Y}}_2^*, \widehat{\mathrm{Y}}_3^*, \ldots, \widehat{\mathrm{Y}}_N^*\}$, a total of N cyclic time series (one per week) for each participant.

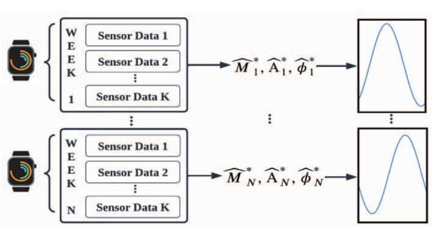

Fig. 2: Cyclic Time Series Generation

B. *Dynamic Transformation of Static Metadata*

Lim et al. [20] showed it is possible to use static covariate directly to fuse and improve the transformer model for multi-horizon forecasting. These findings motivate our efforts to develop a dynamic mapping of covariate parameters (here, static human characteristics) to the transformer model, which could potentially boost the performance of our cyclic behavior forecasting.

We observe some human characteristics are more stationary and less varying than others. For example, although resilience and coping are both internal to humans and resemble stability. However, while resilience may not change over time, coping strategies may. Therefore, we develop a strict static metadata transformation model and a non-strict static metadata transformation model. The common goal of both models is to convert static data into a time series of a certain length, thus assisting in predicting the periodic data generated in the main model.

1) Strict Static Metadata Transformation: As shown in Figure 3, when the data introduced as covariates are stationary, the transformation step only requires the expansion of the static data into a sequence of the desired time length. To maintain the balance with the primary data, normalization is required.

$$\text{Strict Static Metadata } c \xrightarrow{\text{dilation}} c, c, \ldots, c \xrightarrow{\text{normalization}} \tilde{c}, \tilde{c}, \ldots, \tilde{c}$$

(Transformed Metadata)

Fig. 3: Strict-Static Transformation

2) Non-strict Metadata Transformation: If the data is not stationary (the actual data is time-varying, but the input is static), we can transform it into a time-varying form.

In this model, we need to rely on the bio-behavioral cyclic time series to help our non-strict static metadata to find the moments which can reflect the change in its actual form. We use Penalized change-point detection [21] to calculate the exact number of alterations in cyclic data for individuals through the entire time series. Let the set $\{\widehat{\mathrm{A}}_\tau^*\}$ and $\{\widehat{\phi}_\tau^*\}$ be two signals each containing n samples. We assume that $\{\widehat{\mathrm{A}}_\tau^*\}$ and $\{\widehat{\phi}_\tau^*\}$ are not stationary and are composed of several successive regimes. Let $\mathcal{P}$ be the set of partitions of $\{1, \ldots, n\}$ that consist solely of integer intervals. For a partition $Z = \left\{\underline{z}_1, \ldots, \underline{z}_{|Z|}\right\} \in \mathcal{P}, \underline{z}_1, \ldots, \underline{z}_{|Z|}$ represent different regimes. The number of regimes in Z is $|Z|$, and the number of change points is $|Z| - 1$. Change-point detection aims to retrieve the various signal regimes and moments of signal transitions from one to the other. Since the number of change points is unknown in the data, we use a penalized empirical quadratic risk here [22]:

$$\underset{Z \in \mathcal{P}}{\arg\min} \sum_{\underline{z} \in Z} \sum_{i \in \underline{z}} \left(\widehat{\mathrm{A}}_{\tau i}^* - \bar{\widehat{\mathrm{A}}}_{\tau \underline{z}}^*\right)^2 + \mathrm{pen}(Z) \tag{4}$$

$$\underset{Z \in \mathcal{P}}{\arg\min} \sum_{\underline{z} \in Z} \sum_{i \in \underline{z}} \left(\widehat{\phi}_{\tau i}^* - \bar{\widehat{\phi}}_{\tau \underline{z}}^*\right)^2 + \mathrm{pen}(Z) \tag{5}$$

where $\bar{\widehat{\mathrm{A}}}_{\tau \underline{z}}^*$ and $\bar{\widehat{\phi}}_{\tau \underline{z}}^*$ are the mean value of $\widehat{\mathrm{A}}^*$ and $\widehat{\phi}^*$ on segment $\underline{z}$; $\mathrm{pen}(\cdot)$ denotes a suitable non-negative function defined on $\mathcal{P}$ and $\mathrm{pen}(Z) := \beta|Z|$ with $\beta > 0$. We choose Z to satisfy both conditions 4 and 5, and we set β as our penalty parameter, which controls the trade-off between model complexity and goodness of fit. Low values of $beta$ favor partition with many regimes, and high values of $beta$ discard most change points. The whole penalty term allows controlling the balance between signal approximation and model complexity.

After getting Z, we can calculate the unchanged periods; let's denote it as $\mathcal{L}$, which is a set containing all time lengths in Z. Since this model aims to find the dynamic response of individuals to change based on the static c, once we obtain the generated cyclic series $Y^*(t)$ for each period, we can compute the change of the data between consecutive periods. It can be denoted as:

$$D_n = |\int_0^{2\pi} Y^*(t_{n+\tau})dx| - |\int_0^{2\pi} Y^*(t_n)dx| \qquad (6)$$

Because $\mathcal{L}$ is the set of periods that one would keep relatively stable, the total change through the participant's stable periods could be represented as $D_n \cdot \mathcal{L}$. Therefore, we can get the individual dynamic time series by multiplying the non-strict static metadata c:

$$\tilde{c} = D_n \cdot \mathcal{L} \cdot c \qquad (7)$$

The structure of the non-strict static metadata transformation is shown in Figure 4.

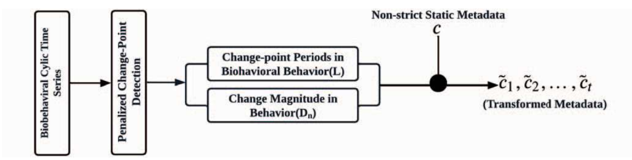

Fig. 4: Non-strict-Static Transformation

C. Transformer-Based Prediction

As shown in Figure 1, the part of the transformer in the architecture consists of a position embedding layer, followed by three stacked encoder layers and three stacked decoder layers. The positional embedding here encodes sequential information in time series data to let the subsequent encoder distinguish every time position in data.

Inside the encoder block, three identical encoder layers are stacked on top of one another. Each stacked encoder layer has a multi-head self-attention sub-layer and a fully-connected feed-forward sub-layer. In particular, inspired by [23], we use causal convolution self-attention based on the traditional self-attention to transform the cyclic time series input data into queries and keys, thereby fully leveraging local context to capture long-term dependencies. Every sub-layer has residual connections between each other, and a normalization operation follows the connections.

The decoder architecture is comparable to the encoder architecture, which contains three identical decoder layers. However, the difference is that each decoder layer has an additional encoder-decoder attention sub-layer, which takes input from the convolution self-attention and all moments output from the encoder's last layer. Finally, an output layer transfers the output of the previous decoder layer to the target time sequence. To ensure that the prediction of a time series data point will only depend on the previous data, we use look-ahead masking and a one-position offset between the decoder input and target output in the decoder. Utilizing Transformer has better parallel performance in more complex neural networks.

IV. Experimental Evaluation

We evaluate our architecture using a wearable dataset containing subjective assessments of human characteristics, including resilience and coping strategies. Specifically, we investigate the following:

- RQ1: How well does the transformer-based time series architecture perform in modeling cyclic behavior and predicting future trajectories compared to existing time-series methods? We obtain cyclic behavioral change curves from wearable data and train different time series forecasting methods. By comparing with the actual curves, we evaluate the performance of each prediction model.
- RQ2: To what extent does additional (relatively static) information about human characteristics (e.g., resilience and coping abilities) contribute to more accurate modeling of dynamic behavior? We first examine the relationship between resilience, coping scores, and behavior regularity. We then use those scores as covariates in the forecasting model to measure their impact on predicting future behavior trajectories.

A. Dataset

We use a passive biobehavioral dataset gathered from 99 student participants at an American University; Each participant received a Fitbit Sense, which they wore 24/7 (except for showering, swimming, and charging times) for at least 14 weeks to collect behavioral and physiological data. The dataset includes steps, heart rate, calories burned, and movement activity data collected in one-minute samples. Table I provides a summary of features.

Data	Description	Data Type
Heart Rate	Average heart rate each hour	Numerical
Steps	Step count each hour	Numerical
Calories	Calories burned each hour.	Numerical
Sedentary Minutes	Sedentary minutes each hour	Binary
Lightly Active Minutes	Lightly active minutes each hour	Binary
Fairly Active Minutes	Fairly active minutes each hour	Binary
Very Active Minutes	Very active minutes each hour	Binary

TABLE I: Passive Sensing Features and Description

We also use data from Brief COPE [24] and Brief Resilience Scale (BRS) [25] surveys that participants filled out once at the beginning of the study. BRS contains six questions graded on a scale of 1 to 5. A higher score indicates a greater ability to bounce back after difficult situations. Brief COPE contains 24 questions, each scored from 1 to 4, and higher scores for each question represent more frequent use of the described method. Details about the questions which can be divided into four coping strategies/factors:

- The Social support category contains eight questions and represents emotional support, instrumental support, venting, and religion.
- Problem-solving category contains four questions and represents active coping and planning.

- Avoidance category contains ten questions and represents self-distraction, denial, substance use, behavioral disengagement, and self-blame.
- Positive thinking category contains six questions and represents positive re-framing, humor, and acceptance.

B. Baseline Models

We use five time-series prediction models to predict future values of the generated cyclic time series, including ARIMA, Orbit, LSTM, NeuralProphet, and DeepAR as our baseline models. **ARIMA** [26] is the most fundamental time series forecasting model that uses lagged moving averages to smooth time series and make predictions. **Orbit** [27] is a modeling approach for the Bayesian processing of time series tasks. We use the Damped Local Trend exponential smoothing model in Orbit. **Long-short term memory (LSTM)** [28], is an improved version of recurrent neural networks (RNN) that can learn over long input series. **NeuralProphet** [29], based on Neural Networks and developed in Facebook Prophet, is a forecasting framework that supports autoregression. **DeepAR** [30] proposed by Amazon, has a structure of autoregressive recurrent networks. Instead of training a model for each input series, DeepAR fits a global model using all time-series to make accurate probabilistic forecasts of future series.

C. Experimental Setup

As previously stated, this study includes continuous data for a relatively large sample size and an extended 14 weeks for 99 participants. Since seven different sensor signals are represented in the continuous data, we denote $K = 7$. During the cyclic generating process, we normalize these multidimensional signals from -1 to 1 to manage them. In addition, to minimize the risk of overfitting, we use leave-one-participant-out cross-validation in all forecasting models to train on the data of 98 participants and test the data of the remaining participant.

D. The forecasting performance of Transformer-based time series behavior model (RQ1)

We compare baseline models with the transformer [23] one using the Cosinor-generated cyclic time series for the first 13 weeks. We choose L1 loss as the evaluation metric since it is a relatively common calculation method and provides a better understanding of the error magnitude. Table II displays the values of the five forecasting models for the 0.25 quantile, the median, the 0.75 quantile, and the mean of the participant loss data to provide different perspectives on the distribution of the loss values . According to the results, the transformer-based forecasting model has the lowest loss values in all dimensions. Specifically, the transformer-based model has an average loss value of 0.036, which is 0.02 points lower than the second best-performing model, orbit. We obtain its significant p-value of $1.74e^{-5}$ using a t-test, verifying that our transformer-based model is significantly better than other models to enhance the predictive capacity.

Models	Q1	Median	Q3	Avg
NeuralProphet	0.2133	0.2991	0.4649	0.3765
ARIMA	0.0260	0.2152	0.5431	0.3551
LSTM	0.0645	0.0792	0.1013	0.0801
DeepAR	0.0442	0.0611	0.0866	0.0675
Orbit	0.0321	0.0477	0.0685	0.0570
Transformer	**0.0191**	**0.0310**	**0.0460**	**0.0356**

TABLE II: L1 Loss Results for Models Forecasting

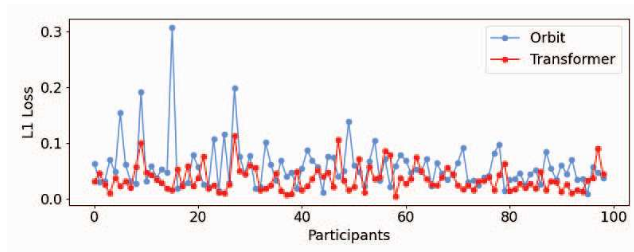

Fig. 5: L1 Loss Comparison between Orbit and Transformer Individually

To further highlight the superiority of the transformer-based model over the orbit model, we plot Figure 5. The red and blue curves depict the L1 loss for all 99 participants using the transformer and orbit model, respectively. The more stable and higher the prediction performance, the fewer variations in the curves and the closer the loss values are to zero. In most participant data, the orbit model has higher loss values than the transformer-based model and abnormally large loss values for a few people. These observations imply that our Transformer-based method can model cyclic behavioral patterns more closely than other methods.

E. The Impact of Supplementing Resilience and Coping Scores with Biobehavioral Data on Behavior Prediction Accuracy (RQ2)

To investigate the impact of adding static behavioral information (here, resilience and coping scores) on the model's performance, we first examine the strength of the relationship between resilience and coping scores with behavior regularity.

Given that regularity can have a measurable impact on prediction, a comprehensive analysis of the interplay between resilience/coping and behavioral regularity can serve as a critical tool for improving the prediction process. In essence, this exploration of these two fundamental characteristics can provide valuable insights into the predictability of a person's behavior.

1) Measuring Behavior Regularity: We measure behavior regularity by calculating frequency and intensity change in each person's time series. The frequency and intensity change with which people perform their daily activities are common indicators of the regularity of human behavior.

Frequency refers to the regular occurrence of behavioral patterns repeated over time in this study. Our architecture utilizes the Cosinor method to calculate losses, which evaluate the degree to which participants' behavioral patterns align with the 24 hours. By investigating the relationship between resilience, coping, and loss, we aim to understand how these characteristics interact and contribute to participants' ability to maintain regular behavioral routines over time. The errors generated from the Cosinor method when fitting passive sensing data streams to periodic curves based on the 24 hours can provide the frequency information. Different sensors collect signals with varying regularities, which influences the degree of fit of the Cosinor model to the original data. Cosinor measures the model's fit by computing the residual sum of squares (RSS), which is the sum of the squared discrepancies between the measured value Y_i and the estimated value of the model at time t. Therefore, the smaller the RSS, the better model fits the passive signal in the original data.

Intensity change in this study refers to the weekly behavior trend variations; it includes the value of shift of the amplitude and phase in our bio-behavioral time series. Our architecture generates cyclic behavioral time series in which we can compute participants' week-to-week trend consistency. By investigating the relationship between resilience, coping, and shift in trend, we aim to understand how these human characteristics interact and contribute to participants' ability to maintain stable behavioral intensity. To assess intensity changes, i.e., the changes in behavioral trends of participants over weeks, we use the Dynamic Time Warping (DTW) algorithm [31] and calculate the point-by-point distance between two consecutive two-week cyclic time series. A shorter distance indicates a more stable tendency in the cyclic time series throughout the weeks and, thus, more regular behavior.

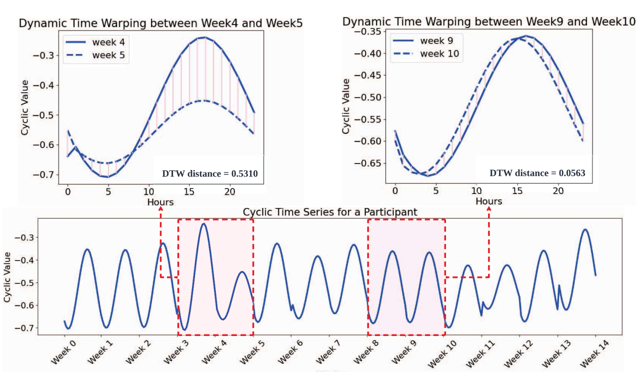

Fig. 6: Similarity of Behavior Patterns from Dynamic Time Warping

For instance, in Figure 6, the blue cyclic curve in the center displays the 14-week behavior pattern of a participant. By computing the DTW distance between two weeks for this participant, we find that the DTW distance between week 4 and week 5 is 0.53, and between week 9 and week 10 is 0.06. The lower the DTW distance, the more alike the behavior patterns are between the two weeks. Hence, we can conclude that the participant's behavior pattern fluctuated more between week 4 and week 5, indicating a more irregular behavior pattern compared to week 9 and week 10.

2) Examining Relationship between Resilience, Coping, and Behavior Regularity: Figure 7 shows the Pearson correlation coefficients regarding the scores of resilience and four coping strategies with RSS (the loss of frequency-matching in Cosinor), respectively. We observe a relatively significant($p = 0.018$) association between resilience and RSS for $r = -0.24$, suggesting more regular behavior patterns among more resilient people. However, the Pearson correlations between the four coping strategy scores and the RSS are near zero, indicating that people's coping abilities are not substantially connected to their behavior regularity.

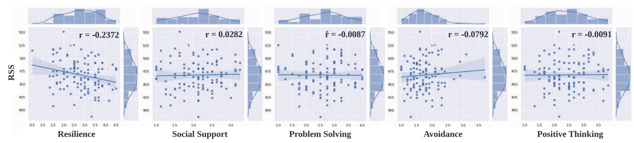

Fig. 7: Correlation Between Resilience, Coping Strategies, and RSS

We compute each participant's average inter-weekly DTW distances(the intensity change of behavioral patterns) and perform correlation tests with resilience and coping as presented in Figure 8. The results indicate that resilience is negatively correlated with DTW distance for $r = -0.22$, showing that high resilience leads to a reduction in week-to-week behavioral change and demonstrating that resilience is significantly($p = 0.030$) associated with behavioral regularity from another perspective. Our results also show a weak negative correlation between DTW distance and problem-solving ($r = -0.11$), and avoidance ($r = -0.12$), respectively. Therefore, the more frequently these two coping strategies are used, the less the week-to-week behavioral change on behalf of the participants is existed.

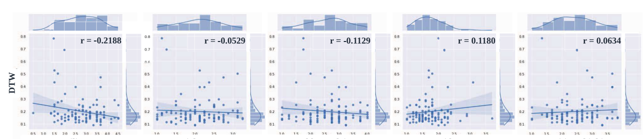

Fig. 8: Correlation Between Resilience, Coping Strategies, and DTW Distance.

3) Transforming Static Resilience and Coping scores to Dynamic Time Series:

a) Converting coping strategy scores as non-static metadata: Coping evolves as human behavior changes [32]; however, the subjective reported coping assessments are usually static scores. Therefore, we must map the static coping strategy scores into a dynamic series that varies as behavior changes over time. In our architecture, we apply coping strategy scores as the non-strict static metadata in the transformation module. Considering the variability of the time series, we use penalized change-point detection method to find the moments when participants' behavior shifts in the actual data.

In the first step, we sum and average the 24 questions in the brief COPE survey according to the categories they belonged to, thus representing the degree to which each participant scored on each coping type. In the second step, we reuse the sets of $\{\widehat{\mathrm{A}}_{\tau}^{*}\}$ and $\{\widehat{\phi}_{\tau}^{*}\}$ from Section III-A, and double check the significance of the sets to ensure that the p-value in the zero-amplitude test less than 0.05. Then, we apply the penalized change-point detection with these sets; we evaluate many β values before settling on $\beta = 0.01$ for it to reach the lowest BIC (Bayesian information criterion) among other values of $beta$. In the last step, we follow the procedure in Section III-B, substituting the average coping score as the non-strict static metadata c.

b) Converting resilience scores as static metadata: In contrast to coping, resilience is a human trait that remains stationary over long periods, so we only need to convert it to an invariant time series. In our architecture, we apply resilience scores as the strict static metadata in the transformation module. For the Brief Resilience Scale, we sum and average each individual's scores on six questions from that questionnaire and then bring the average resilience score as the strict static metadata c.

4) Forecasting Performance of Resilience and Coping Strategies as Covariates: In this section, we illustrate the process of adding transformed static resilience and coping scores as covariates to the prediction model and discuss how they improve the model's performance. Figure 1 shows the overall forecasting process, including the transformation model component. First, we transform the participants' bio-behavioral data into cyclic time series. Then, for each fold of the leave-one-participant-out cross-validation, we concatenate the participants' bio-behavioral cyclic time series with the transformed covariate series to train. In Section IV-D, we demonstrated that the transformer-based forecasting model is more accurate than other baseline models in predicting future human behavior patterns. Therefore, we use this model as a baseline to compare the performance of the model built with covariates.

The results of the transformer forecasting models with resilience and different coping strategies are shown separately in Table III. We can see that regardless of which covariate is used, it has a lower value of forecast loss than the model without covariates. It shows that the resilience and coping strategies scores can be covariates to help predict human cyclic behavioral trends. In addition, the transformer-based model with avoidance as a covariate predicted the lowest loss values among all covariates in data distributions Q2, Q3, and the average, implying that scores on avoidance coping strategies would be more helpful in predicting accuracy overall.

Figure 9 shows the box-plot of the number of epochs obtained by these models. If the number of training epochs decreases in the model after using the covariates, adding the covariates can make the prediction results converge faster.

The average number of epochs for the original transformer model without covariates is 78, the highest number of training epochs and, thus, the slowest convergence among all models

Covariates	Q1	Q2	Q3	Avg
None (Baseline)	0.0191	0.0310	0.046	0.0355
Social Support Coping	**0.0152**	0.0268	0.0419	0.0318
Problem Solving Coping	0.0178	0.0272	0.0415	0.0324
Avoidance Coping	0.0160	**0.0263**	**0.0409**	**0.0316**
Positive Thinking Coping	0.0171	0.0287	0.0420	0.0323
Resilience	0.0179	0.0277	0.0424	0.0323

TABLE III: L1 Loss for Transformer Models with Covariates

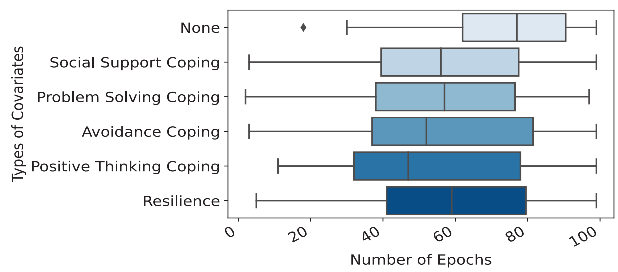

Fig. 9: Number of Epochs for Transformer Models with Covariates

with covariates. In addition, although the general distribution of the number of training epochs is quite similar across all five covariate models, the transformer-based model constructed for the positive thinking covariate yields a lower average number of training epochs, indicating that positive thinking can achieve faster convergence. It gives us some insight that faster convergence from positive thinking coping strategy will provide a competitive advantage in terms of time efficiency in the future if larger sample sizes and longer-term observations are used. It is one criterion for assessing the model's predictive performance. In summary, it further illustrates that incorporating resilience and coping strategies can improve the predictive power of the entire structure and help reproduce the cyclical behavioral trends of the real situation.

V. Discussion

Our research demonstrates that the use of the Transformer models for predicting human behavior from wearable sensing data outperforms existing state-of-the-art approaches. Moreover, the integration of human characteristics such as resilience and coping enhances the prediction of future behavior. While our model is currently configured with resilience and coping, it is expandable to include other characteristics such as emotional stability, thereby capturing the intricacy of human behavior and allowing for more accurate predictions. This adaptability enables the model to be trained on new dataset, handling diverse events and personal characteristics more effectively. Also, our methodology reduces the necessity for frequent subjective assessments, replacing daily questionnaires with only once or twice surveys and enabling more objective modeling.

Our study, first in incorporating underlying human behavior characteristics for modeling, underscores the potential for accurate future behavior predictions, thereby informing preventative technology for timely intervention. However, limitations exist; our participant pool is limited to students from a single local university, leading to regional biases and less visible behavioral variability. In the future, we will investigate more diverse demographic and behavioral samples and consider adding more types of sensors and bringing more flexible time windows.

VI. Conclusion

In our research, we utilize personal characteristics like resilience and coping strategies to analyze the regularity and predictability of human behavior patterns from sensor data, leveraging a custom transformer-based architecture for cyclic time series forecasting. Compared to conventional methods, our approach more accurately replicates behavior patterns and effectively transforms static characteristics into dynamic time series consistent with behavioral changes. Moreover, integrating resilience and coping data enhances the model's predictive capability, demonstrated by a reduction in loss and expedited convergence. Future endeavors aim to refine our models for detailed behavioral analysis, potentially facilitating personalized, timely interventions in critical situations marked by rapid behavioral changes.

Acknowledgment

This research was supported by the National Science Foundation (NSF-IIS-1816687).

References

[1] H.-J. Lee, C.-H. Cho, T. Lee, M.-G. Kim, H. In, and L. Kim, "Mood prediction of patients with mood disorder by machine learning using passive digital phenotypes based on circadian rhythm: a prospective observational cohort study (preprint)," *Journal of Medical Internet Research*, vol. 21, 05 2018.

[2] P. Innominato, C. Focan, T. Gorlia, T. Moreau, C. Garufi, J. Waterhouse, S. Giacchetti, B. Coudert, S. Iacobelli, D. Genet, M. Tampellini, P. Chollet, M.-A. Lentz, M.-C. Mormont, F. Lévi, and G. Bjarnason, "Circadian rhythm in rest and activity: A biological correlate of quality of life and a predictor of survival in patients with metastatic colorectal cancer," *Cancer research*, vol. 69, pp. 4700–7, 07 2009.

[3] W. T. Riley, D. E. Rivera, A. A. Atienza, W. Nilsen, S. M. Allison, and R. Mermelstein, "Health behavior models in the age of mobile interventions: are our theories up to the task?" *Translational Behavioral Medicine*, vol. 1, no. 1, pp. 53–71, 02 2011. [Online]. Available: https://doi.org/10.1007/s13142-011-0021-7

[4] A. Doryab, D. Villalba, P. Chikarsel, J. Dutcher, M. Tumminia, X. Liu, S. Cohen, K. Creswell, J. Mankoff, D. Creswell, and A. Dey, "Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: A three-fold analysis (preprint)," 12 2018.

[5] S. Gashi, E. Di Lascio, B. Stancu, V. D. Swain, V. Mishra, M. Gjoreski, and S. Santini, "Detection of artifacts in ambulatory electrodermal activity data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, jun 2020. [Online]. Available: https://doi.org/10.1145/3397316

[6] A. Masten, "Masten as. ordinary magic. resilience processes in development. am psychol.56(3):227-238," *The American psychologist*, vol. 56, pp. 227–38, 04 2001.

[7] C. Snyder, *Coping: The Psychology of What Works*. Oxford University Press, 1999. [Online]. Available: https://books.google.com/books?id=wM_hBwAAQBAJ

[8] M. Lancaster and P. Callaghan, "The effect of exercise on resilience, its mediators and moderators, in a general population during the uk covid-19 pandemic in 2020: a cross-sectional online study," *BMC Public Health*, vol. 22, 04 2022.

[9] F. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, p. 115, 01 2016.

[10] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[11] T. Duong, H. Bui, D. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-markov model," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 838–845 vol. 1.

[12] P.-C. Chung and C.-D. Liu, "A daily behavior enabled hidden markov model for human behavior understanding," *Pattern Recognition*, vol. 41, pp. 1572–1580, 05 2008.

[13] Q. Huang, D. Cohen, S. Komarzynski, X.-M. Li, P. Innominato, F. Lévi, and B. Finkenstädt, "Hidden markov models for monitoring circadian rhythmicity in telemetric activity data," *Journal of The Royal Society Interface*, vol. 15, p. 20170885, 02 2018.

[14] E. Pierson, T. Althoff, and J. Leskovec, "Modeling individual cyclic variation in human behavior," vol. 2018, 04 2018, pp. 107–116.

[15] B. Hadj-Amar, B. Finkenstädt, M. Fiecas, F. Lévi, and R. Huckstepp, "Bayesian model search for nonstationary periodic time series," *Journal of the American Statistical Association*, vol. 115, pp. 1320–1335, 01 2020.

[16] S. Taylor, R. Killick, J. Burr, and L. Rogerson, "Assessing daily patterns using home activity sensors and within period changepoint detection," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 70, 02 2021.

[17] Y. Li, S. Rao, J. R. Ayala Solares, A. Hassaïne, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, and G. Salimi-Khorshidi, "Behrt: Transformer for electronic health records," *Scientific Reports*, vol. 10, 04 2020.

[18] Y. Zhang, L. Wang, H. Chen, A. Tian, S. Zhou, and Y. Guo, "If-convtransformer: A framework for human activity recognition using imu fusion and convtransformer," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 2, jul 2022. [Online]. Available: https://doi.org/10.1145/3534584

[19] G. Cornélissen, "Cosinor-based rhythmometry," *Theoretical biology medical modelling*, vol. 11, p. 16, 04 2014.

[20] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, 06 2021.

[21] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Processing*, vol. 85, pp. 1501–1510, 08 2005.

[22] C. Truong, L. Oudre, and N. Vayatis, "Penalty learning for changepoint detection," 08 2017, pp. 1569–1573.

[23] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," 06 2019.

[24] K. Baumstarck, M. Alessandrini, Z. Hamidou, P. Auquier, T. Leroy, and L. Boyer, "Assessment of coping: A new french four-factor structure of the brief cope inventory," *Health and Quality of Life Outcomes*, vol. 15, 01 2017.

[25] B. Smith, J. Dalen, K. Wiggins, E. Tooley, P. Christopher, and J. Bernard, "The brief resilience scale: Assessing the ability to bounce back," *International journal of behavioral medicine*, vol. 15, pp. 194–200, 02 2008.

[26] P. Zhang, "Zhang, g.p.: Time series forecasting using a hybrid arima and neural network model. neurocomputing 50, 159-175," *Neurocomputing*, vol. 50, pp. 159–175, 01 2003.

[27] H. C. S. Y. S. S. Edwin Ng, Zhishi Wang, "Orbit: Probabilistic forecast with exponential smoothing," 2020.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997.

[29] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir, and R. Rajagopal, "Neuralprophet: Explainable forecasting at scale," 2021. [Online]. Available: https://arxiv.org/abs/2111.15397

[30] V. Flunkert, D. Salinas, and J. Gasthaus, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, 04 2017.

[31] W. Meert, K. Hendrickx, T. Van Craenendonck, and P. Robberechts, "Dtaidistance," Aug. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.7158824

[32] R. Lazarus, "Coping theory and research: Past, present, and future," *Psychosomatic medicine*, vol. 55, pp. 234–47, 01 1998.