Personalized Context-Aware Depression Detection via Hierarchical Temporal Contrastive Learning

Synopsis

Name: Vikas Kumar Tyagi
Mentor Name: Dr. Vivek Kumar Dwivedi

Introduction

Traditional approaches to detecting depression often rely on self-reported questionnaires and periodic clinical assessments, which are limited by their subjectivity and infrequency. These methods may miss subtle or early signs of depressive episodes.

In contrast, modern technologies such as smartphones and wearable devices offer a promising alternative by enabling the continuous, passive collection of behavioral and physiological data.

This project aims to leverage these data streams to detect depressive symptoms by identifying deviations from an individual's typical behavior patterns.

Scope and objectives

The scope of this project lies in the development of a robust, personalized mental health monitoring system capable of detecting early signs of depression through passive sensing technologies. Unlike generic models that apply uniform thresholds across populations, this approach emphasizes tailoring detection mechanisms to individual behavioral patterns.

Objective:
To design a **novel depression detection model that incorporates**:

- **Personalized Behavioral Modeling**: Captures and learns an individual's unique behavioral baseline to improve sensitivity to mood-related deviations.
- **Context-Aware Signal Interpretation**: Integrates contextual factors (e.g., time, location, events) to reduce false positives caused by routine or environmental changes.
- **Temporal Contrastive Learning**: Utilizes self-supervised learning to identify significant temporal shifts in behavior linked to depressive symptoms.

The goal is to detect mental health changes early and unobtrusively, supporting timely intervention and personalized care.

Scope and objectives continued…

What Has Been Done Till Now

- Models trained and evaluated on **generalized populations**.
- Focus on **group-level behavioral trends** for mental health prediction.
- Uses **predefined features** and **conventional ML classifiers** (e.g., Random Forest, Logistic Regression).
- Limited personalization—same model applied across all users.
- Emphasis on **population-level statistical patterns** rather than individual variability.

What We Are Going to Do

- Shift from population-level to **individual-level behavioral modeling**.
- Use **Temporal Contrastive Learning (TCL)** for capturing **subtle, personalized behavior changes**.
- Incorporate **user-specific baselines** and **context-aware personalization** using neural adapters.
- Move from conventional ML to **deep learning-based self-supervised temporal models**.
- Focus on **early detection** via modeling of intra-individual deviations over time.

Scope and objectives continued…

Ethical Considerations

- **Anonymized, Public Dataset Use :** The dataset is fully anonymized and shared with prior informed consent, ensuring ethical data sourcing and research integrity.
- **Privacy & Confidentiality Compliance:** Strict adherence to privacy regulations and participant confidentiality ensures secure handling of sensitive behavioral and physiological data.
- **Non-Invasive Passive Monitoring:** No active user input is required—behavioral data is collected passively, minimizing disruption and preserving participant comfort and autonomy.
- **Assistive, Not Replacement Tool:** The system supports clinicians and individuals by providing early insights, but does not replace expert medical judgment or therapy.

All Variables Are Measurable

All behavioral, physiological, and contextual variables in the dataset are quantifiable, enabling reliable modeling and interpretation of depressive symptom patterns.

- **Behavioral Features (Smartphone) :** Screen time, call frequency, Bluetooth proximity
- **Physiological Signals (Wearables) :** Step count, sleep duration, physical activity levels
- **Contextual Metadata :** Day type (weekday/holiday), pandemic periods, special events
- **Mental Health Labels:** PHQ-4, BDI-II scores (validated clinical instruments)

Scope and objectives continued… [Questions]

- **How can behavioral deviations be detected at a personalized level?**
  This is the central question of the study, addressing the use of real-world behavioral data for unobtrusive mental health monitoring.

- **How does context affect behavioral interpretation in depression detection?**
  This supports the context-awareness aspect and helps reduce false positives due to routine or situational changes.

- **Can temporal contrastive learning identify early signs of depression?**
  This focuses on the novel learning approach used contrastive learning and addresses the challenge of sparse mood labels.

- **What is the impact of personalization on model accuracy?**
  This focuses on the approach of tailoring the model to individual behavioral baselines rather than applying generalized thresholds across the population.

Sample Size Calculation

To ensure the results are statistically reliable and not due to random variation, sample size estimation is performed using confidence interval analysis. This method calculates the minimum number of observations needed to achieve a desired level of precision, reducing uncertainty and improving the accuracy of model predictions and generalizability.

Dataset Used: GLOBEM dataset

Total Participants: **497 individuals**

Data collected over **4 years** (2018–2021)

Confidence Interval-Based Estimation

Note: Use **Confidence Interval-Based Estimation** when you want to estimate a population parameter (like a mean or proportion) with a range that reflects the uncertainty of the sample data.

Assuming:

Margin of Error (E): **±5%**

Estimated proportion (p): **0.5** (maximum variability)

Z-score(Z) : 1.96

Note: The **Z-score of 1.96** is used in the **confidence interval formula** because it corresponds to a **95% confidence level** in a standard normal distribution (also known as the Z-distribution).

Sample Size (n) formula for estimating population proportion is given by:

$$n = \frac{Z^2 \cdot p(1-p)}{E^2} = \frac{1.96^2 \cdot 0.5 \cdot (1-0.5)}{0.05^2} = \frac{3.8416 \cdot 0.25}{0.0025} = \frac{0.9604}{0.0025} \approx 384.16$$

Conclusion: Thus, the dataset size of 497 exceeds the minimum required sample size, ensuring sufficient statistical power for model training and evaluation.

Sample Size Calculation Continued…

RQ1: How can behavioral deviations be detected at a personalized level?
Goal: Detect individual-level behavioral deviations using **temporal modeling** and **personalized neural adapters**.

Note: Use Cohen's d (two-group comparison) formula when you want to calculate the sample size or effect size for comparing the means of two independent groups.
Assumption: Medium effect size (Cohen's $d = 0.5$), $\alpha = 0.05$, Power = 80%
Using Cohen's d – Two group comparison formula

$$n = \frac{2\left(Z_{\alpha/2} + Z_{\beta}\right)^2}{d^2} = \frac{2(1.96 + 0.84)^2}{0.5^2} = \frac{2(2.8)^2}{0.25} = \frac{2 \times 7.84}{0.25} = \frac{15.68}{0.25} = 62.72$$

Description:
n: Sample size per group
$Z_{\alpha/2}$: Z-score corresponding to the desired confidence level (e.g., 1.96 for 95%)
Z_{β}: Z-score corresponding to the desired power (e.g., 0.84 for 80%)
d: Cohen's d (effect size)

Required N: ~63 per group
Available: 497 participants.
Conclusion: Hence sample size is sufficient to compare personalized vs. generic models

Sample Size Calculation Continued…

RQ2: How does context affect behavioral interpretation in depression detection?

Goal: Assess if adding contextual metadata improves model accuracy and reduces false positives.

Note: Sample size calculation formula used when planning a study with Analysis of Variance (ANOVA)

Assumption: Medium effect size (Cohen's $f = 0.25$), $\alpha = 0.05$, Power = 80%

Using Cohen's f – ANOVA design formula

$$n = \frac{\left(Z_{\alpha/2} + Z_{\beta}\right)^2}{f^2} = \frac{2(1.96 + 0.84)^2}{0.25^2} = \frac{2(2.8)^2}{0.0625} = \frac{7.84}{0.0625} = \frac{7.84}{0.0625} = 125.44$$

Description:

n: Sample size

$Z_{\alpha/2}$: Z-score corresponding to the desired confidence level (e.g., 1.96 for 95%)

Z_{β}: Z-score corresponding to the desired power (e.g., 0.84 for 80%)

f: Cohen's f (ANOVA effect size)

Required N: ~126 participants

Available: 497 participants which is **Statistically sufficient**

Conclusion: Hence data is sufficient for contextual comparisons across model settings.

Sample Size Calculation Continued…

RQ3: Can temporal contrastive learning identify early signs of depression?
Goal: Identify early signs of depression using temporal contrastive learning (**TCL**) with anchor-positive-negative windows.

Note: This approach uses time-series self-supervised learning rather than mean comparison, so sample size per group is not the primary concern; instead, temporal resolution and window length are more critical.

Requirements: High-frequency, long-term behavioral data per individual enables accurate modeling of personal patterns and detection of subtle temporal deviations.

At least several weeks of data per user must be used for meaningful contrastive learning
Dataset spans: 4 years, 497 participants
Conclusion: Dataset is highly suitable for temporal modeling with TCL.

Sample Size Calculation Continued…

RQ4: What is the impact of personalization on model accuracy?

Goal: Evaluate whether personalization layers improve accuracy over generic models.

Note: Use Cohen's d – Paired Difference when comparing the means of two related conditions or measurements taken from the same group or subjects, such as before-and-after tests or model performance with and without personalization.

Assumption: Medium effect size (Cohen's d = 0.5), α = 0.05, Power = 80%

Using Cohen's d – Pair Difference formula

$$n = \frac{\left(Z_{\alpha/2} + Z_{\beta}\right)^2}{d^2} = \frac{2(1.96 + 0.84)^2}{0.5^2} = \frac{(2.8)^2}{0.25} = \frac{7.84}{0.25} = 31.36$$

Description:

n: Sample size per group

$Z_{\alpha/2}$: Z-score corresponding to the desired confidence level (e.g., 1.96 for 95%)

Z_{β}: Z-score corresponding to the desired power (e.g., 0.84 for 80%)

d: Cohen's d (effect size)

Required N: ~32 participants

Available: 497 participants**.**

Conclusion: The dataset provides sufficient depth and diversity to support a robust evaluation of the impact of personalization on model performance.

Data Description

The GLOBEM dataset is a comprehensive, multi-modal dataset collected from 497 individuals over a period of four years (2018-2021). It includes passive sensing data from smartphones and wearable devices, capturing behavioral, physiological, and contextual information relevant to mental health monitoring. Behavioral data comprises phone usage patterns such as calls, screen time, and Bluetooth proximity. Physiological data is gathered through Fitbit devices, tracking steps, sleep, and activity levels. Contextual information includes weekday vs. weekend, holidays, and COVID-affected periods.

Source: GLOBEM Dataset (2018-2021)
Participants: 497 individuals
Data Types:
- **Behavioral**: Smartphone usage (calls, screen activity, Bluetooth usage)
- **Physiological**: Fitbit signals (steps count, sleep patterns, activity)
- **Contextual**: Time-based data (weekdays, holidays, COVID-19 periods)

Labels: PHQ-4 and BDI-II depression screening scores

Data Reference: The dataset used in this study can be accessed and downloaded from the following official source.
https://the-globem.github.io/datasets/overview

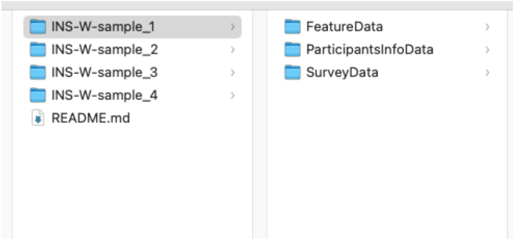
Data Description [GLOBEM Dataset Structure]

Dataset Variants

Four datasets released: INS-W_1(2018), INS-W_2(2019), INS-W_3(2020), INS-W_4(2021)

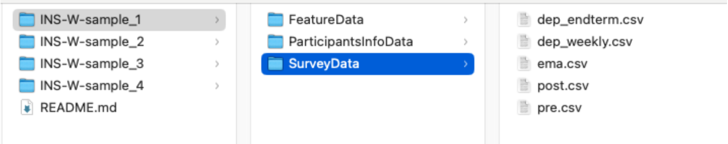
Each dataset contains three core folders:

- SurveyData
- FeatureData
- ParticipantInfoData



SurveyData: Contains participants' mental health self-assessments and questionnaire responses:

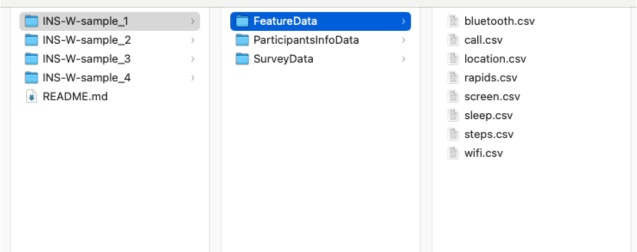
File Name	Description
dep_weekly.csv	Depression labels (from post-study & EMA surveys)
dep_endterm.csv	Depression labels (post-study only – used in end-term predictions)
pre.csv	Pre-study questionnaire responses
post.csv	Post-study questionnaire responses
ema.csv	Weekly EMA surveys (delivered midweek or weekend)



Data Description [GLOBEM Dataset Structure continued …]

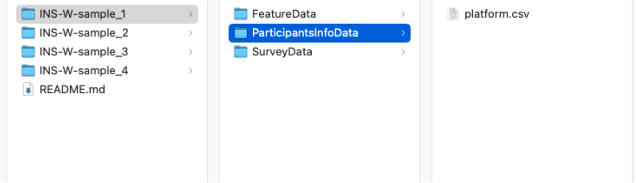
FeatureData: Includes behavior-derived features from smartphone and wearable data:

File Name	Description
rapids.csv	Features extracted using the RAPIDS tool
location.csv	Phone signals
screen.csv	Phone signals
call.csv	Phone signals
bluetooth.csv	Fitbit signals
steps.csv	Fitbit signals
sleep.csv	Fitbit signals
wifi.csv	Phone signals



ParticipantInfoData: Include participant platform information.

File Name	Description
platform.csv	Device platform info (iOS or Android)



Data Description [Data Dictionary]

The dataset includes key behavioral features such as step count, sleep duration, screen time, call frequency, and Bluetooth proximity. Contextual features include day type (weekday/holiday) and device platform. These features, extracted using the RAPIDS tool, are critical for modeling behavioral changes linked to depressive symptoms in a personalized manner.

Variable	Description	Type
step_count	Daily number of steps	Numerical
sleep_duration	Hours slept per night	Numerical
screen_time	Duration of screen activity	Numerical
call_count	Number of calls per day	Integer
bluetooth_devices	Unique Bluetooth devices detected	Categorical
day_type	Weekday, weekend, holiday, etc.	Categorical
phq4_score	Depression screening score (0–12)	Ordinal
bdi2_score	Depression inventory score	Ordinal

Analytic approach

Temporal Contrastive Learning

Temporal Contrastive Learning is a self-supervised technique that trains models to recognize subtle behavioral changes over time by comparing sequential time windows. It uses anchor, positive, and negative samples to distinguish between stable and deviating behavioral states, enabling earlyand context-aware detection of mental health variations without requiring extensive labeled data.

Personalized Behavioral Modeling

Personalized Behavioral Modeling focuses on learning each individual's unique behavioral patterns to detect deviations that may indicate mental health changes. By using model adapters, it tailors the analysis to personal baselines rather than population averages, enhancing detection accuracy and reducing false alarms caused by normal inter-individual variability in daily routines.

Context-Aware Signal Processing

Context-Aware Signal Processing incorporates contextual information—such as weekends, holidays, or special events—into the model to interpret behavioral data more accurately. By recognizing that certain deviations may be due to external circumstances rather than mental health changes, this approach helps reduce false positives and enhances the model’s sensitivity and reliability.

Multi-modal Deep Learning

Multi-modal Deep Learning integrates data from smartphones and wearables—such as activity, sleep, and phone usage—over time to capture richer behavioral patterns for prediction.

Statistical Validation

Statistical validation involves hypothesis testing, correlation analysis, and effect size estimation to assess the significance, strength, and practical relevance of the model’s predictive results.

Analytic approach [Research Question 1]

RQ1: How can behavioral deviations be detected at a personalized level?
This approach leverages **personalized behavioral modeling** to capture individual's unique baseline patterns over time. Rather than relying on population-wide thresholds, the model uses **neural adapters** that tailor predictions to individual users.

Proposed Solution:

- **Implement personalized behavioral modeling using individual-specific adapters**
  This approach uses neural adapters within the model architecture to fine-tune predictions based on individual behavior profiles, enabling context-sensitive detection through transfer learning.
- **Track deviations from baseline behavioral vectors over time**
  Temporal embeddings are generated for each user to monitor variations from their historical baseline using techniques like contrastive loss, aiding in early detection of depressive trends.

Null Hypothesis (H_0):

Personalized modeling does not improve detection accuracy compared to generic models

Expected Outcome:

- Improvement in identifying subtle mood shifts
- Reduced false positives due to individualized behavioral baselines

Analytic approach [Research Question 2]

RQ2: How does context affect behavioral interpretation in depression detection?
This approach integrates **contextual metadata**—such as weekday vs. weekend, holidays, and pandemic periods—into the depression detection model to enhance accuracy. Behavioral patterns can naturally vary due to external events or time-related factors; without accounting for this, models may generate false positives.
To address this, the model uses **attention-based context encoders** that dynamically weigh behavioral signals depending on their context. By comparing the model's performance **with and without contextual input**, researchers can evaluate the significance of context-aware interpretation in reducing misclassifications.

Proposed Solution:

- Integrate contextual metadata such as weekdays, holidays, and pandemic periods into the model input
- Use attention-based context encoders to dynamically weigh behavioral signals in varying contexts
- Compare model outputs with and without contextual features to evaluate their influence

Null Hypothesis (H_0):

Contextual information has no significant impact on the accuracy of depression detection

Expected Outcome:

- Improved accuracy and reduced false positives
- Better disambiguation of behavioral changes driven by routine/context vs. mental health shifts
- Enhanced model robustness across diverse life situations

Analytic approach [Research Question 3]

RQ3: Can temporal contrastive learning identify early signs of depression?

This approach applies Temporal Contrastive Learning (TCL), a self-supervised method that learns behavioral patterns over time by comparing different time windows. Using contrastive loss, the model learns to differentiate between stable and deviated behavioral states, helping detect subtle mood shifts before they become clinically significant. This reduces reliance on labeled data while capturing meaningful temporal dynamics.

Proposed Solution:

- Apply temporal contrastive learning using a triplet loss framework to learn temporal behavioral representations
- Construct anchor, positive, and negative time windows to distinguish between stable and deviated states
- Leverage self-supervised learning to detect behavioral shifts before clinical symptoms manifest

Null Hypothesis (H_0):
Temporal contrastive learning does not outperform traditional methods in early depression detection

Expected Outcome:

- Enhanced sensitivity to subtle, pre-clinical behavioral deviations
- Reduced dependence on labeled data
- Improved early warning capability in personalized mental health monitoring systems

Analytic approach [Research Question 4]

RQ3: What is the impact of personalization on model accuracy?
This approach compares the performance of depression detection models **with and without personalization layers** to evaluate the effect on accuracy, precision, recall, and F1-score. Personalized models use **user-specific neural adapters** that adjust predictions based on an individual's behavioral baseline, while generic models apply the same rules to all users.

Proposed Solution:
- Compare model performance with and without user-specific adapters for personalized modeling
- Evaluate metrics like precision, recall, F1-score, and AUC across personalized vs. generic models
- Perform ablation studies to isolate the contribution of personalization layers

Null Hypothesis (H_0):
Personalization has no statistically significant impact on model accuracy in detecting depression

Expected Outcome:
- Personalization significantly improves detection accuracy and reduces false positives
- Increased model generalizability across diverse individuals
- Demonstrated advantage of modeling user-specific behavioral patterns over population-level baselines

Analytic Approach [Evaluation metrics]

We are solving a **binary classification problem**—predicting whether an individual is experiencing depressive symptoms or not. The chosen evaluation metrics (Precision, Recall, F1-score, AUC) are essential due to the **class imbalance** and the **critical cost of false negatives**, ensuring both sensitivity and reliability in mental health predictions.

To assess the model's performance, standard classification metrics will be used, especially suitable for imbalanced mental health datasets:

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Indicates the proportion of true positives among predicted positives — important to reduce false alarms.

Recall (Sensitivity)

$$\text{Recall} = \frac{TP}{TP+FN}$$

Critical for identifying actual depressive cases, minimizing false negatives.

Analytic Approach [Evaluation metrics continued…]

F1 Score

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Balances precision and recall — ideal when false positives and false negatives are both costly.

Area Under the ROC Curve (AUC-ROC)

- Measures the model's ability to discriminate between classes at various thresholds.
- AUC close to 1.0 indicates excellent separability.

These metrics will be used to compare:

- Personalized vs. non-personalized models
- Context-aware vs. context-free settings
- Traditional vs. contrastive learning architectures

Recommendation and applications

Target Users

- Individuals at risk of depression
- Mental health clinicians and therapists
- Digital wellness platforms and app developers
- Researchers in behavioral health analytics

User Benefits

- **At-risk individuals** receive early, passive, and personalized detection of depressive symptoms—without the burden of self-reporting.
- **Clinicians** gain objective, continuous behavioral insights to support diagnosis and treatment planning.
- **App developers** can integrate real-time mental health monitoring features to enhance user engagement and care.
- **Researchers** benefit from a validated framework to study mental health patterns over time, enabling more targeted intervention strategies.

References and bibliography

Core References on Behavior Modeling and Datasets

- Saeb, S. et al. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *J Med Internet Res.*
- Wang, R. et al. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *UbiComp.*
- Aledavood, T. et al. (2017). Quantifying daily rhythms with behavioral data. *Scientific Reports.*
- Mohr, D. C. et al. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology.*
- De Choudhury, M. et al. (2013). Predicting depression via social media. *ICWSM.*

References Related to Temporal and Contrastive Learning

- He, K. et al. (2020). Momentum contrast for unsupervised visual representation learning. *CVPR.*
- Chen, T. et al. (2020). A simple framework for contrastive learning of visual representations. *ICML (SimCLR).*
- Yao, S. et al. (2021). Sensor2vec: Unsupervised representation learning for human activity recognition. *AAAI.*

References and bibliography Continued…

References on Personalization and Adaptation

- Abnar, S. et al. (2021). BERG: Towards temporal contrastive learning on physiological signals. *arXiv preprint*.
- Triastcyn, A. et al. (2020). Federated Learning with Bayesian Differential Privacy. *ICML*.
- Dey, A. et al. (2022). SEMBED: Self-supervised behavior representation learning. *arXiv preprint*.

Other Supporting Work

- Harari, G. M. et al. (2016). Using smartphones to collect behavioral data in psychological science. *Perspectives on Psychological Science*.
- Jacobson, N. C. et al. (2020). Flatten the curve: Digital mental health interventions to decrease depression and anxiety during the COVID-19 pandemic. *JMIR*.
- Zhan, Y. et al. (2022). Personalized mental health prediction using multi-task learning with behavioral data. *NeurIPS Workshop*.