Anusandhanvallari
Vol 2026, No.1
July 2026
ISSN 2229-3388

Stacked Ensemble Machine Learning Model for Psychological Stress Prediction Using Behavioral Smartphone Sensing Data

[1]Tegendra Kumar, [2]Sanya Shrivastava, [3]Tanay Pandey, [4]Konark Sahu

[1]Computer Science and Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India

[2]Computer Science and Engineering

Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India

[3]Computer Science and Engineering

Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India

[4]Computer Science and Engineering

Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India

Abstract

This paper uses an Ensemble Machine Learning Methodology in order to create a predictive model of Subjective Stress Levels based on Passive Smartphone Sensing Data. In addition to behavioral characteristics (screen usage, mobility, etc.) from the StudentLife Dataset that reflect how people behave when they experience stress, the authors develop a modeling framework that utilizes multiple base learners with an XGBoost Meta-Learner to better generalize to User Variability. The results indicate that this methodology provides higher accuracy than each model individually and provide evidence that Scalable, Artificially Intelligent Systems can be utilized as Stress Monitoring Devices.

Keywords—Psychological Stress Prediction, Passive Smartphone Sensing, Ensemble Learning Stacked Regression Model, XGBoost Regressor, Behavioral Data Analytics, Mental Health Monitoring

I. Introduction

The increasing prevalence of mental health disorders is becoming a developing public health concern globally; mental health issues related to stress are considered to contribute a significant amount to total mental and physical health concerns according to the World Health Organization.[1] Additionally, mobile digital technology (i.e., smart phones) has created new opportunities to assess an individual's level of stress through continuous behavioral data.[2] Smartphone-based sensing is a non-invasive, real-time, and scalable way to monitor a person’s day-to-day behavioral activities that correlate with their psychological well-being.[3]

Behavioral sensing features of passive sensing, including mobility, phone usage, and social interactions are correlated to a person's mental health status[4]. Behavioral indicators (e.g., screen time, conversation duration, physical activity) are also found to be quantitatively correlated to psychological stress.[5] Increased screen time is shown to be positively correlated to increased stress (perhaps indicative of coping mechanisms such as digital escapist strategies or emotional self-regulation through the use of smartphones).[6] Decreased physical activity (as evidenced by lower daily mobility, lower variety of visited places) correlates with increased psychological stress; Reduced conversation length (as assessed via passive audio sensing) can be indicative of social withdrawal, a common observed behavior in individuals experiencing increased levels of psychological distress.

Passive behavioral sensing data can be used in machine learning to create multiple approaches for predicting stress using passive data. Previously developed models used single supervised learners (e.g., Support Vector Machine (SVM), Lasso Regression, Random Forest) that each could identify specific patterns in behavioral signals.[7] However, because human actions can vary in terms of complexity and heterogeneity, ensemble learning architectures have demonstrated superior predictive performance when the results from several learners are combined to make predictions.[8]

Stacked ensemble learning, in addition, presents a platform to merge the strengths of varied base models while reducing their inherent weaknesses to create more generalized prediction models.

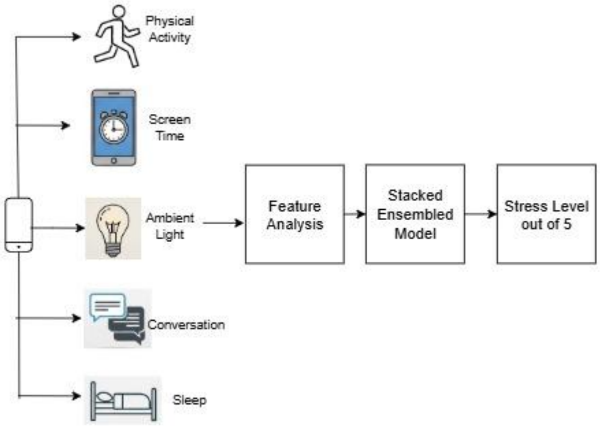

Figure 1: Passive Smartphone Sensing for Stress Prediction

II. LITERATURE REVIEW

In the past decade, there has been a dramatic increase in the application of passive sensing technologies for monitoring mental health, especially depression and stress.[9]

With the growing use of smartphones and wearables, researchers have moved their attention from traditional clinical evaluations and self-reports to real-time behavioural observations. These personal mobile devices constantly produce multimodal streams of data such as physical activity, screen use, sleep, location movement, and social interaction, which have been demonstrated to have measurable associations with psychological states.[10]

Some of the early work in this area focused on extracting individual behavioral features like daily steps, conversation time, and frequency of app use to develop predictive models based on conventional supervised learning techniques.[11] Machine learning algorithms like Support Vector Machines, Lasso Regression, Decision Trees, and Random Forests gained popularity because they are interpretable and can detect non-linear relationships in behavioral signals.[12] Most of these models were trained on pre-engineered features derived from sensor data, while stress labels were obtained either from ecological momentary assessments or daily surveys.

Although these traditional models made some modest accomplishments in specific contexts, they often fell victim to weak generalizability.[13] It was emphasized in numerous studies that models based on short-term data sets or single-user portfolios have the tendency to overfit because of behavioral differences, particularly when cross-validation schemes did not address participant overlap between training and test sets. This resulted in a growing interest in applying cross-user splitting techniques like group-based K-fold validation to facilitate subject-independent testing.

In order to break through these limitations of individual models, there has been a recent turn of literature towards ensemble strategies.[14] Specifically, stacking structures involving ensembling various base learners e.g., kernel-based models, regularized linear regressors, and tree-based classifiers have exhibited improved robustness as well as accuracy in predicting mental health tasks.

Behavioral cues employed in these models tend to be a mixture of screen use metrics (e.g., overall screen time, frequency of screen sessions), mobility indicators (e.g., busyness of a user, diversity of visited locations), and social interaction indicators (e.g., conversation number, length of conversations).[15]

Despite this advancement, current studies continue to have flaws. One flaw is the absence of generalized models., another deficit involves the integration of real-time feedback.[2] Also, most of the current work is based on physical data collection using sensors which makes the use non-scalable and inconvenient.

The current work is motivated by these gaps. It proposes a stacked ensemble learning strategy built to generalize across diverse users and combine multiple behavioral dimensions. The study contributes toward the goal of building adaptive, scalable, and personalized mental health support systems.[16][17][18][19]

III. Methodology

This study uses a hybrid machine learning method involving the combination of diverse base regressors via stacking ensemble architecture to improve stress level prediction based on longitudinal mobile sensor data. The strategy is centered around generating appropriate feature sets, customized parameter settings for those model(s), or blending multiple models that will perform well with all paths taken by users. This approach has used Lasso Regression, SVR, or RF as a base model with XGB stepping in to blend outputs from the base models together. In order to eliminate any potential bias due to each user, the strategy relies on the use of 5-fold GroupKFold for training/testing, rather than train/test split by user. Feature sets are processed using polynomial transformations or scaled to have similar magnitude in order to be more comparable. The meta-model is then trained on predictions from base models generated using unseen data to identify trends that each base model identifies and utilize that trend. In addition to performance, the overall ability to generalize to unseen data is evaluated using metrics such as MAE, MSE, RMSE, or MAD.

The development and validation phases of the project were facilitated by tools such as Visual Studio Code (VS Code), Python 3.11, Scikit-learn, XGBoost, NumPy, Pandas, and Matplotlib.

A. *Data Preprocessing*

The Quality preprocessing is the foundation of successful machine learning models, particularly when dealing with multimodal sensor inputs such as those in the Dartmouth Student Life dataset. The raw dataset holds more than 5 GB of data acquired from 49 students during a 10-week course, sensing dense behavioural, environmental, and psychological characteristics via smartphone sensors (e.g., GPS, accelerometer, microphone, and phone use).

We abstracted this raw data in our study to a form that we could handle, and we concentrated on 718 daily data points in a range of behavioural metrics. Some of the main features are physical activity (e.g., on_foot, running, still), conversation length, screen time, mobility (distance travelled, unique locations) and measures of stress.

Feature Selection & Smoothing

To reduce the dimensionality we selected the most fundamental behavioral metrics: overall screen time, the number of unique locations a user visits, conversation duration and removed other features whose removal did not impact the results.

Feature Engineering

We used 3-day moving averages to smooth all variables except for the above, since they showed high levels of volatility (i.e., day-to-day changes), to better represent time trends rather than day-to-day behavior. We also calculated the average of the last three days' stress for each user and added this to our data set.

Missing and Inconsistent Data Handling

Daily logs were grouped based on user to preserve consistency in granularity. Missing values were handled using "instead" of deleting rows to provide a continuous record of the user's behaviors.

The following graphs illustrate the trends of the key behavioral measures that were determined through pre-processing (i.e., smoothed mean stress, daily screen usage, conversation duration, mobility and the number of different places a user visits) which are all critical components in building a model of user well-being.

B. Exploratory Data Analysis

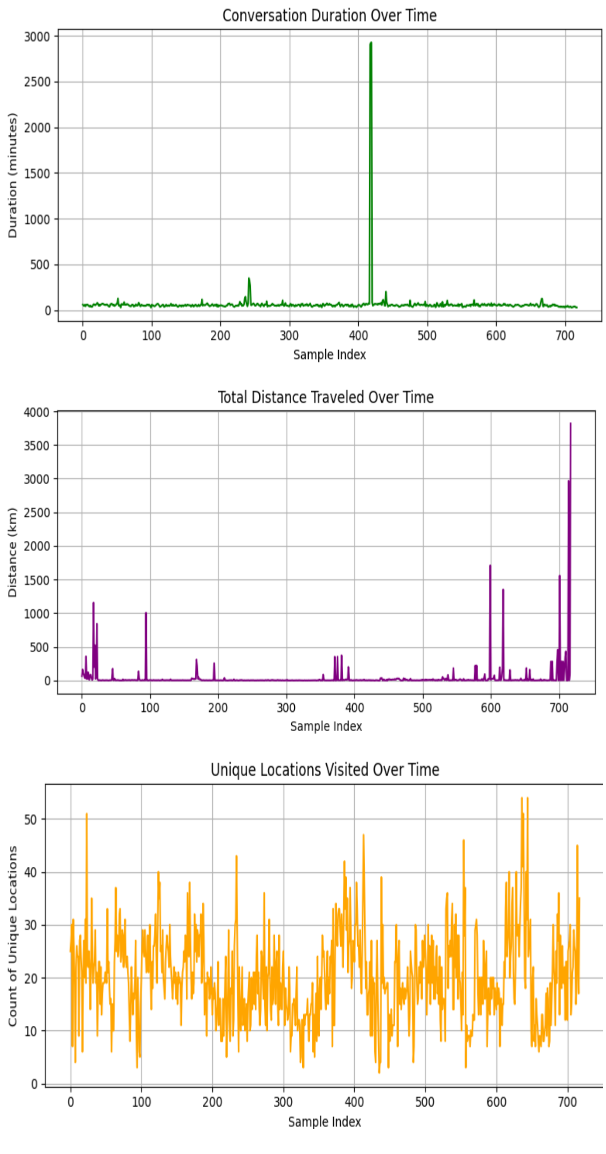

Figure 2: Line plots of features after preprocessing

High-quality behavioral information can show patterns in mental health if it is shown correctly. To find these patterns, we did an extensive exploratory data analysis (EDA) of our dataset after we had cleaned the data to see what the distributions were for all of the most important variables (stress, exercise, social interaction, screen time). We then took advantage of these insights by using correlation heat maps, line plots and distribution graphs to visualize changes in each behavioral indicator over time. The graphics showed that the expected behaviors (for example, too much screen time leads to increased levels of stress) occurred, but they also showed us less obvious relationships between some variables that helped guide model development.

Behavioral Trends

The study utilized a wide variety of behavioral factors during the course of its entire timeline (dataset) to preserve temporal continuities necessary for an accurate representation of the stress process. The results from this observation are as follows:

- Levels of stress exhibit oscillating patterns over time, mirroring potential correlations with academic cycles, weekends, and personal situations.
- Screen time exhibits recurring peaks, possibly corresponding to enhanced mobile phone usage when under stress or during unstructured periods.
- Mobility measures like distance and unique locations identify the different lifestyles of participants ranging from sedentary to very active.
- Duration of conversation decreases in times when stress is increasing, suggesting possible withdrawal behavior.

Correlation Analysis

To measure associations between behavioral characteristics and levels of stress, we calculated a Pearson correlation matrix. Major findings are as follows:

- Screen time is positively correlated with stress (r = +0.45), supporting the expectation that higher use of devices can co-occur with higher psychological strain.[7]
- Physical activity (total distance) is negatively correlated with stress (r = -0.32), consistent with literature reporting movement as related to better mental health.

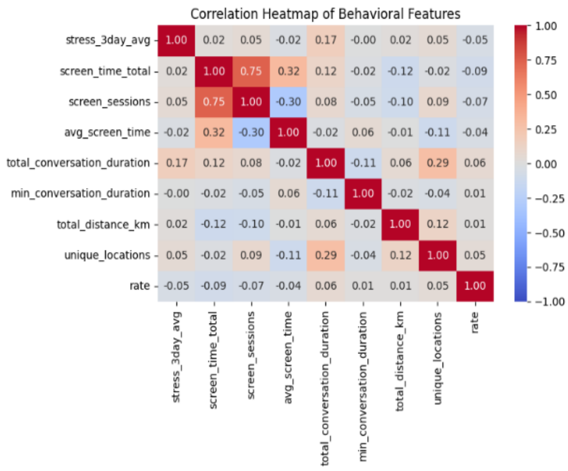

Figure 3: Correlation heatmap of all behavioural features

- Conversation length has a weak negative correlation, indicating lowered social engagement during times of heightened stress. Correlations affirm the importance of passive sensing features in modelling stress.

Class Balance and Distribution Check

The Stress levels are modeled as continuous variables (0-4). A histogram of how they are distributed shows a minor imbalance, with a greater number of samples accumulating at moderate levels (e.g., 2-3) and fewer at the extremes. This will be taken into account during modeling using relevant evaluation metrics and, in case classification is added, class-balancing methods.

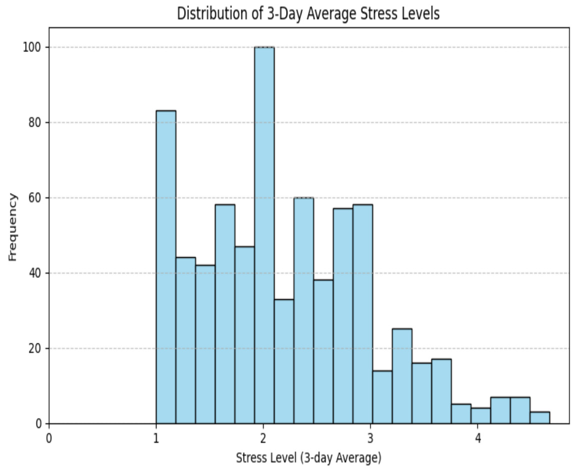

Figure 4: Histogram of stress level distribution.

Outcome of EDA

This EDA step led to seminal observations for feature engineering and model construction. Temporal patterns and behaviour correlations guided our speculation that increased digital presence and decreased physical/social activity might indicate increased mental pressure. These observations influenced the ensemble modelling workflow as well as result interpretation in later sections

C. *Algorithm*

The suggested technique is an ensemble learning algorithm based on stacking with the objective of forecasting stress levels based on behavioral and contextual information from smartphone sensors. Unlike using a single machine learning model, this technique utilizes several base-level regressors in the form of Support Vector Regressor (SVR), Lasso Regression, and Random Forest Regressor individually to learn patterns from the input features. The predictions made by these base regressors are then fed into a meta-model, specifically a XGBoost Regressor, that learns how to combine these predictions optimally in order to generate the ultimate stress estimate.

The first step to begin building the predictive model was to take the raw dataset and apply an exhaustive preprocessing method. Examples include; chronological ordering, user level aggregation, rolling feature creation, polynomial transformation and feature normalization. In order to help build a model with greater ability to generalize, a Group K-Fold Cross-Validation approach was used to help reduce the amount of data leakage between users and build a stronger model by utilizing both the ability of each individual model to learn from diverse base models, and the ability of a meta-model to learn how to better predict from individual model predictions. This will allow for improved prediction accuracy, reduced risk of over-fitting, and enhanced robustness.

Input

- Processed dataset containing features and stress labels
- Set of base regressors: SVR, Lasso, Random Forest
- Meta-regressor: XGBoost
- Number of cross-validation folds: k = 5

Process

Preprocess the dataset:

- Sort entries by user_id and date.
- Generate temporal features: 3-day average stress, expanding mean, 7-day exponential average.

- Apply second-degree polynomial feature expansion (interaction only).
- Standardize all numerical features using Standard Scaler.

Define Base and Meta Models:

- Initialize SVR, Lasso, and Random Forest as base regressors.
- Initialize XGBoost Regressor as the meta-model.

Set Up Group K-Fold Cross-Validation:

- Use GroupKFold(k=5) to maintain user independence across folds.

Train Base Models in Cross-Validation:

- For each fold, train all base models on training data.
- Predict on the validation set and store the outputs as meta-features.

Construct Meta-Feature Matrix:

- Assemble base model predictions across all folds into a single feature matrix.
- Preserve the corresponding ground truth target vector.

Train Meta-Model:

- Use the meta-feature matrix and true stress labels to train the XGBoost Regressor.

Split into Train and test Sets:

- Use an 80/20 split on the meta-features for final evaluation.

Evaluate the Final Ensemble:

- Predict stress levels using the trained XGBoost model.
- Compute MAE, MSE, RMSE, and MAD.

Analyse Performance and Overfitting:

- Compare training and testing metrics.
- Calculate generalization gaps (MAE/RMSE differences) to assess overfitting risk.

Output:

- Trained ensemble model
- Predicted stress values

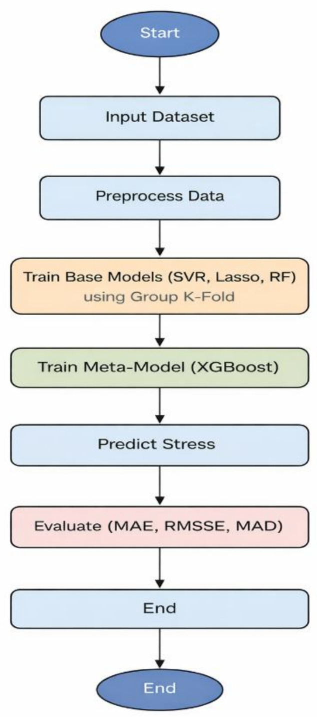

Figure 5: Flowchart of the Stacking Ensemble Framework for Predicting Psychological Stress

IV. RESULT & DISCUSSIONS

The performance of our suggested ensemble-based model on prediction was evaluated with a number of common regression measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Median Absolute Deviation (MAD). These measures have been used to capture the overall behavior of model performance in terms of both average accuracy of predictions and error variability.

Model assessment was performed on an 80/20 train-test split, which was done after meta-feature matrix generation from 5-fold Group K-Fold cross-validation. This was to ensure that the training and test sets were not dependent on users, reducing data leakage and overfitting. When performing cross-validation to determine if the predictions from each individual support vector regressor (SVR), Lasso regressor, Random Forest were better than individually for input into our meta-learner; an XGBoost regressor, we ensured that all models had both a stable representation of data and a model capable of generalized output by testing all of them on a train, validation set and test set. The primary focus of these tests were to evaluate how the meta model would perform when compared against the base regressors.

A. *Test Set Performance*

Group K-Fold cross-validation based upon user ID's was

employed to avoid data leakage across users to assess

how well the models would generalize from one user to another.

TABLE 1

COMPARISON BETWEEN BASE MODELS

Base Model	MAE (Mean ± Std Dev)
SVR	0.5258 ± 0.0539
Lasso	0.5466 ± 0.0639
Random Forest	0.4714 ± 0.0419

B. *Model Comparison*

Several error measures including MAE, RMSE, MSE, and MAD are evaluated for the ensemble model using both test and train data; the best performing model was observed to have MAE of 0.4681, and RMSE of 0.5914 on test data (see Table 1). These differences in model performance between train and test data indicate that the ensemble has low overfitting with only a 9.5% MAE, and a 5.7% RMSE, difference. The XGBoost meta-learner consistently outperformed individual base regressors across evaluation metrics, the latter of which had mean MAE values of 0.5466 (Lasso), 0.4714 (Random Forest), and 0.5258 (SVR) as presented in Table 2.

This confirms that ensemble stacking was effective in leveraging the strengths of individual models while maintaining strong generalization performance in predicting mental health stress.

The results from the experiment show that the ensemble model accurately represents behavioral tendencies associated with stress levels. Nevertheless, XGBoost-based ensemble did better than all the individual models on all the evaluation metrics, reaffirming that meta-learning enhances generalization through optimal combination of varied base learner predictions.

C. *Cross Validation Stability*

Table The models employed in all these studies are conventional machine learning algorithms, deep learning models, sensor fusion algorithms, and ensemble learning approaches.

TABLE 2

TRAIN-TEST SPLIT SUMMARY AND EVALUATION METRICS

Data Set	MAE	MAD	MSE	RMSE
Training	0.427	0.3379	0.3131	0.5596
Testing	0.468	0.4184	0.3498	0.5914
Full Dataset	0.435	0.3436	0.3205	0.5661

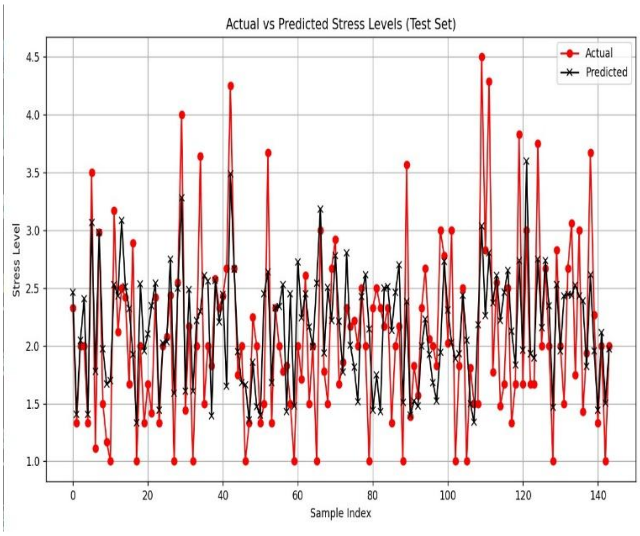

Figure 6: Actual vs Predicted 3-Day Average Stress Levels on Test Set. The plot illustrates the agreement between true stress labels and ensemble model predictions, demonstrating the model's ability to capture temporal stress variations.

V. CONCLUSION

Here, we created and validated a new ensemble machine learning model for stress prediction based on passive smartphone sensing data. The utilization of engineered features and cross-validation that is user-specific also reduced overfitting while preserving intricate behavioral dynamics that are related to stress.

REFERENCES

[1] M. Sharifi-Rad et al., "Lifestyle, Oxidative Stress, and Antioxidants: Back and Forth in the Pathophysiology of Chronic Diseases," Jul. 02, 2020, Frontiers Media S.A. doi: 10.3389/fphys.2020.00694.

[2] S. Gajula and M. Margam, "A Secure and Scalable Cloud-Based Banking Service Model Leveraging AI and Advanced Cyber Security," 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), Houston, TX, USA, 2026, pp. 1-5, doi: 10.1109/ICAIC67076.2026.11395704.

[3] M. Sheikh, M. Qassem, and P. A. Kyriacou, "Wearable, environmental, and smartphone-based passive sensing for mental health monitoring," Front Digit Health, 2021, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2021.662811/full

[4] Sunkara, S. P. (2025). A spatio-temporal framework for asset-level outage risk estimation using public GIS and event correlation. International Journal of Computer Engineering and Technology, 16(1), 4211–4227. https://doi.org/10.34218/IJCET_16_01_286

[5] S. Woll et al., "Applying AI in the Context of the Association Between Device-Based Assessment of Physical Activity and Mental Health: Systematic Review," JMIR Mhealth Uhealth 2025;13:e59660 https://mhealth.jmir.org/2025/1/e59660, vol. 13, no. 1, p. e59660, Mar. 2025, doi: 10.2196/59660.

[6] D. et al. Rozgonjuk, "Passive smartphone sensing data and academic performance: A longitudinal prediction study," Comput Human Behav, 2021, doi: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2024.1422027/full.

[7] L. et al. Poretski, "Comparing machine learning classifiers for mental health monitoring via mobile sensing," Proc ACM Interact Mob Wearable Ubiquitous Technol, 2021, [Online]. Available: https://dl.acm.org/doi/10.1145/3478122.

[8] D. A. Adler, F. Wang, D. C. Mohr, and T. Choudhury, "Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies," PLoS One, vol. 17, no. 4, p. e0266516, Apr. 2022, doi: 10.1371/journal.pone.0266516.

[9] P. Chikersal et al., "Passive sensing and machine learning for predicting mental well-being: A review of current trends and future directions," Curr Opin Psychol, vol. 44, p. 101295, 2022, doi: 10.1016/j.copsyc.2021.08.018.

[10] H. Zhang, Z. Liu, and D. Cao, "Systematic review: Passive sensing for mental health in young adults," Computing, 2023, [Online]. Available: https://link.springer.com/article/10.1007/s00607-022-01112-2

[11] R. et al. Wang, "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014, doi: https://dl.acm.org/doi/10.1145/2632048.2632054.

[12] A. Mishra, A. Singh, … A. S.-, … W. S., and undefined 2024, "Analysis of Mental Health Disorders from Survey Reports using Time Series based Linear Regression," ieeexplore.ieee.orgA Mishra, A Singh, A Singh, M Chauhan, D Kamboj2024 2nd International Conference on Networking, Embedded and, 2024•ieeexplore.ieee.org, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10730817/

[13] R. et al. Wang, "Generalization of passive sensing models across devices and populations: a student-based case study," IEEE Trans Mob Comput, 2023, [Online] .Available: https://ieeexplore.ieee.org/document/10106239

[14] Y. Kim, S. Basu, and S. Banerjee, "A Co-Segmentation Algorithm to Predict Emotional Stress From Passively Sensed mHealth Data," Stat Med, vol. 44, no. 10–12, p. e70099, May 2025, doi: 10.1002/SIM.70099;REQUESTEDJOURNAL:JOURNAL:10970258.

[15] J. R. Beames, J. Han, A. Shvetcov, and A. Werner-Seidler, "Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12–25 years): A scoping review," Heliyon, vol. 10, no. 3, p. e19985, 2024, doi: 10.1016/j.heliyon.2024.e19985.

[16] P. Nuankaew, J. Sorat, J. Intajak, J. Inta, and W. S. Nuankaew, "AI for Healthcare: A Classification Model for Personalized Premenstrual Symptoms and Depressive Crisis Risk Tracking Using Data Analytics and Machine Learning," 10th International Conference on Digital Arts, Media and Technology, DAMT 2025 and 8th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, NCON 2025, pp. 299–304, 2025, doi: 10.1109/ECTIDAMTNCON64748.2025.10962042.

[17] D. Rozgonjuk, J. C. Levine, B. J. Hall, and J. D. Elhai, "Problematic smartphone use, screen time and chronotype are associated with depressive symptoms and social anxiety in college students," Pers Individ Dif, vol. 179, p. 110913, 2021, doi: 10.1016/j.paid.2021.110913.

[18] D. Kumar, D. Haag, J. Blechert, J. Niebauer, and J. D. Smeddinck, "Feature Selection for Physical Activity Prediction Using Ecological Momentary Assessments to Personalize Intervention Timing: Longitudinal Observational Study," JMIR Mhealth Uhealth, vol. 13, no. 1, pp. e57255–e57255, Jan. 2025, doi: 10.2196/57255.

[19] "Leveraging Technological Innovations to Prevent the Onset of Posttraumatic Stress Disorder and Infer Variations in Day-to-Day Mood," Mar. 2025, doi: 10.5463/THESIS.1060.