Smart Health 32 (2024) 100464

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Integrating wearable sensor data and self-reported diaries for personalized affect forecasting

Zhongqi Yang [a,*], Yuning Wang [d], Ken S. Yamashita [b], Elahe Khatibi [a], Iman Azimi [a], Nikil Dutt [a], Jessica L. Borelli [b], Amir M. Rahmani [a,c]

[a] Department of Computer Science, University of California, Irvine, United States
[b] Department of Psychological Science, University of California, Irvine, United States
[c] School of Nursing, University of California, Irvine, United States
[d] Department of Computing, University of Turku, Finland

ARTICLE INFO

Keywords:
Affective computing
Multimodal machine learning
NLP
Wearable sensor data
Digital mental health

ABSTRACT

Emotional states, as indicators of affect, are pivotal to overall health, making their accurate prediction before onset crucial. Current studies are primarily centered on immediate short-term affect detection using data from wearable and mobile devices. These studies typically focus on objective sensory measures, often neglecting other forms of self-reported information like diaries and notes. In this paper, we propose a multimodal deep learning model for affect status forecasting. This model combines a transformer encoder with a pre-trained language model, facilitating the integrated analysis of objective metrics and self-reported diaries. To validate our model, we conduct a longitudinal study, enrolling college students and monitoring them over a year, to collect an extensive dataset including physiological, environmental, sleep, metabolic, and physical activity parameters, alongside open-ended textual diaries provided by the participants. Our results demonstrate that the proposed model achieves predictive accuracy of 82.50% for positive affect and 82.76% for negative affect, a full week in advance. The effectiveness of our model is further elevated by its explainability.

1. Introduction

Affect is a broad term referring to one's subjective feeling states. Affective states can influence moment-to-moment emotional experience as well as long-term mood and have impacts on overall well-being (Ekici et al., 2014). Prolonged negative affect has been demonstrated to yield adverse consequences for both physical and mental health outcomes, including increased susceptibility to mental health conditions such as depression (Frijda, 1987; Remes et al., 2021; Shankar et al., 2016; Smidt et al., 2015), which ranks among the leading causes of disability (Karrouri et al., 2021). Accurate identification, detection, and prediction of affect are crucial for effective intervention and prevention strategies.

In recent years, wearable and mobile devices, along with machine learning-based approaches, have been employed to track individuals' affective states (Wang et al., 2022). Existing works focus on gathering and analyzing objective physiological and behavioral data through the use of laboratory-based or wearable/portable devices. Such research utilizing multimodal data, including electroencephalogram (EEG), photoplethysmogram (PPG), has primarily been conducted in laboratory settings (Deb et al., 2017; Filippini et al., 2022; Yang et al., 2020). Expanding beyond the confines of laboratory environments, wearable and mobile devices such as smartwatches, smart rings, wristbands, and smartphones have been increasingly used for real-time, remote monitoring of

* Corresponding author.
E-mail address: zhongqy4@uci.edu (Z. Yang).

https://doi.org/10.1016/j.smhl.2024.100464
Received 16 March 2024; Accepted 18 March 2024
Available online 23 March 2024

individuals' affective states in free-living settings (Jafarlou et al., 2023; Kanjo et al., 2019; Taylor et al., 2017; Wang et al., 2014, 2024). Notably, Jafarlou et al. (2023) demonstrates the feasibility of using fully objective measurements from commercial wearable devices to predict affect status in everyday settings. This study employs smart rings, smartwatches, and smartphones to continuously monitor physiological and environmental information, as well as sleep, metabolic, and physical activity patterns. Machine learning models are leveraged for data analysis to predict positive and negative affect statuses for the next day. This research highlights that utilizing data collected from wearable devices enables a purely objective method to predict one's emotional state the following day. Although the accuracy of this approach is not exceptionally high, it demonstrates the potential of wearable technology in forecasting affect status.

However, there are two limitations of the existing studies. Firstly, while current remote affect monitoring systems primarily rely on objective data gathered from wearable devices in everyday settings, they often ignore the valuable insights that can be obtained from human-generated textual data in daily life, which are crucial for mental health analysis. Such textual data play a pivotal role in providing information, such as life events or emotional narratives, which offer essential information for a comprehensive understanding of an individual's mood trajectory (Baboo et al., 2022; Suhasini et al., 2020). Incorporating such textual data (including diaries, notes, etc.) complementing the objective measurements still requires further exploration. Secondly, existing research concentrates on the instantaneous detection or short-term prediction of affective states. The reliance on fully collected data from the next day limits their predictive capabilities to no earlier than the following day. This timeframe may not be adequate for timely prevention strategies. There is a pressing need for research focused on forecasting mental health conditions significantly in advance of their onset to enhance preventive measures and overall mental well-being.

In this paper, we present a multimodal deep learning model which is based on a transformer encoder and a pre-trained language model, DistilBERT, to forecast affect status one week in advance. We accomplish this by fusing objective features from wearable and mobile devices and self-reported diaries. The objective features encompass a range of data, including sleep patterns, physiological metrics, physical activity, metabolic rates, and environmental factors. For the diary data, we extract features from two primary aspects: the diary submission frequency and the diary content. To evaluate the effectiveness of our proposed affect forecasting model, we collect a dataset featuring multimodal information from wearable devices and self-reported diaries detailing daily affective highs and lows. This system also includes 20 unique discrete affect ratings provided by participants daily and weekly. The affect ratings are aggregated to establish labels for positive and negative affect statuses. Furthermore, we investigate model explainability by analyzing the contribution of features as indicated by Shapley values, and the attention scores assigned to keywords in the diaries.

2. Data collection

We collected data for assessing mental health and affect status of college students including 25 undergraduate students between the ages of 18 and 22 who could speak and write English fluently. Participants could not be parents or married or returning to school after a 3 year period or longer to maintain a homogeneous sample that reflects the college student population. Participants must not have participated in any of our previous studies on this topic.

Over the course of the study, we tracked participants' emotional states, physiological patterns, and behavioral habits through smart devices and ecological momentary assessments (EMAs). Participants were fitted with the Oura ring and Samsung Gear Sport smartwatch and downloaded the corresponding Oura and Samsung Android mobile apps in an effort to capture an accurate depiction of their daily physical habits, sleep, and health. These devices were integrated into the ZotCare mHealth platform (Labbaf et al., 2023). The Oura ring collects data on sleep and cardiovascular activity that are used in this study to assess sleep quality by measuring sleep duration, average heart rate during sleep, and heart rate variability during sleep.

The Samsung watch complements the physical activity feature set by adding walk and run steps and captures the atmospheric pressure, which enriches the environmental features available for analysis. The detailed list of the objective features is the same as those described in Jafarlou et al. (2023).

Participants utilized the ZotCare application to submit self-reports functioning as daily diaries through EMAs. These entries vary from expressions of personal sentiments and moods to narratives of daily experiences. Participants document key moments of their day, detailing the high and low points of their day's affective experiences, overall emotional experiences, and notable daily events such as physical activities, academic exams, or social interactions.

Participants were also asked to engage with the ZotCare app to submit their daily and weekly affect status through EMAs that were used as labels with a one-week delay. The affect status is evaluated using a scale ranging from 0 ("Very Slightly") to 100 ("Extremely"), , where 0 meant "Very Slightly" and 100 "Extremely", including 20 affect words like "inspired" and "nervous", based on the Positive and Negative Affect Schedule (PANAS) (Sano, 2016; Watson et al., 1988). Each affect word is scored individually and as part of a composite, with Positive Affect (PA) and Negative Affect (NA) calculated as the average of 10 positive and negative items, respectively. Binary classifications for PA and NA are assigned based on whether an emotion's score surpassed or did not meet the median for all participants, thereby maintaining an equitable binary label distribution. In line with standard practice, we exclude the central 20% of values from the overall distribution prior to assigning labels.

3. Affect forecasting model

This section introduces the affect forecasting model designed for the analysis of our multimodal data. The workflow of the model is illustrated in Fig. 1.

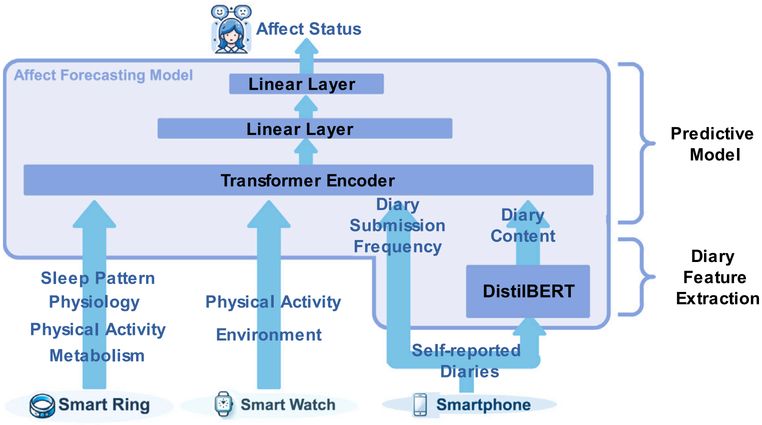

Fig. 1. Affect forecasting model workflow.

3.1. Diary feature extraction

Participants submitted daily self-reported diaries to describe their daily feelings through the ZotCare application on smartphones. We introduce the self-reported diary feature extraction of the affect forecasting model for two key features from the diaries: Diary Content and Diary Submission Frequency.

Diary content. The analysis of self-reported diaries poses a unique challenge in affective computing as they usually do not explicitly reveal emotional status. Moreover, these diaries rarely directly indicate physiological or behavioral factors that could predict future affect status. Considering the model efficiency, we employ DistilBERT (Sanh et al., 2019) complemented by a linear layer to extract meaningful insights from the diaries. DistilBERT functions to parse and interpret the underlying sentiments of the diaries and offers a high-dimensional vector representation of the input. Then the following linear layer acts on the high-dimensional representation produced by DistilBERT, transforming them into a streamlined, singular value vector as the Diary Content feature.

Diary submission frequency. Another challenge in our study is the variability in the frequency of diary submissions by participants, serving as a factor that, as highlighted in recent studies, reflects the mental health status of individuals (Vishnubhotla et al., 2022). We calculate the diary submission frequency based on how many days a participant submits in a one-week window.

3.2. Predictive model

The architecture of the predictive model is built on a Transformer Encoder (Vaswani et al., 2017) to aggregate the features from a wide range of modalities efficiently through its self-attention mechanism. Following the transformer encoder, we integrate a 2-layer multi-layer Perceptron to project the complex embeddings generated by the transformer into a singular, interpretable value that directly corresponds to the affect status.

The input for the predictive model amalgamates diary features derived from Self-reported Diary Processing with objective features, as detailed in Section 2. To achieve this, we concatenate both objective features and diary features to create an extensive feature vector.

3.3. Model training strategy

The integration of diary features in our predictive model is challenged by the initial ineffectiveness of DistilBERT. To address this, we propose a sequential training process that is executed in two steps. Initially, DistilBERT is fine-tuned independently to adapt its pre-trained weights to our specific affective context. We then train the entire model in a joint manner. This sequential approach ensures a balanced learning process to prevent any type of feature from disproportionately influencing the model's development. This process is also illustrated in Fig. 2.

Step 1: Fine-tuning DistilBERT. We observe that it is challenging when training DistilBERT and the predictive model jointly. Given that DistilBERT is a general-purpose model trained on a diverse corpus, we must align DistilBERT to the affect forecasting task. However, jointly training DistilBERT with the predictive model presents challenges. Initially, DistilBERT's performance in identifying affect-related information from diary entries is suboptimal, leading the predictive model to assign attention lower scores to the diary features. This leads to weak gradient signals being sent back to DistilBERT during training that may result in omission of the diary information.

To address this challenge, it is essential to customize DistilBERT specifically for the affect forecasting task. Our approach involves initially fine-tuning DistilBERT to better suit affect forecasting. We accomplish this by integrating a temporary linear layer on top of

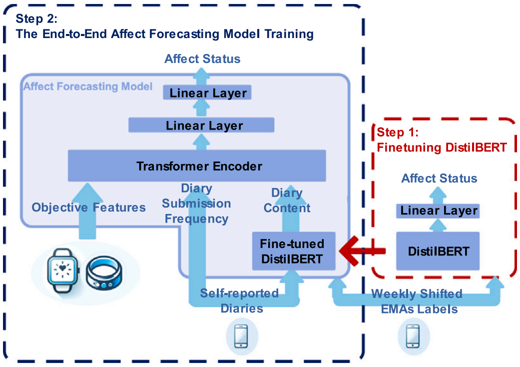

Fig. 2. Training process of the affect forecasting model.

Table 1
Forecasting accuracy results.

	Personalized (%)	Non-personalized (%)
Forecasting PA w/ Diary	**82.50**	81.20
Forecasting NA w/ Diary	**87.76**	80.22
Forecasting PA w/o Diary	81.08	79.03
Forecasting NA w/o Diary	82.11	75.31

DistilBERT, which serves to translate embeddings into predicted affect status and facilitates gradient propagation. Both DistilBERT and this linear layer are then fine-tuned together on the training dataset.

Step 2: The end-to-end affect forecasting model training. Following the initial fine-tuning of DistilBERT, Step 2 jointly trains the whole affect forecasting model encompassing the fine-tuned DistilBERT, the transformer encoder, and the linear layers. Diaries are processed through DistilBERT to extract diary content feature, which are subsequently concatenated with objective features and diary submission frequency. The combined inputs are then fed through the transformer encoder, followed by linear layers that map these inputs to the corresponding affect status.

3.4. Model explainability

We demonstrate the explainability of the proposed model in two aspects. We first highlight how individual features impact the forecasted affect status. We accomplish this by employing path-dependent feature perturbation algorithms provided by the SHAP (SHapley Additive exPlanations) library (Lundberg et al., 2017) to calculate the Shapley values for each feature. Additionally, to gain a deeper understanding of which keywords the model identifies as crucial for predicting affect status, we delve into the significance of specific words in the diary entries. This is achieved by analyzing the attention scores allocated by the fine-tuned DistilBERT model to these keywords.

4. Results

We evaluate the proposed model using leave-one-subject-out cross-validation and a personalized cross-validation strategy. In the leave-one-subject-out cross-validation approach, each iteration involves training the model on the data from all but one subject and testing on the data from the excluded subject. For the personalized cross-validation, we utilize the latter half of a participant's data for testing, while the initial half, combined with data from other participants, is employed for training. This method ensures that each iteration of the model incorporates historical data from an individual participant for training, fostering a personalized model. We refer to the models derived from personalized cross-validation as 'Personalized,' while those obtained from leave-one-subject-out cross-validation are designated as 'Non-personalized.' The accuracy of our proposed model for forecasting affect status are shown in Table 1.

We observe an enhancement in accuracy in the personalized cross-validation process that the personalized models, distinct for the training on the participant's unique data set, outperform generic, non-personalized models. This result suggests that models tailored to individual behavioral and emotional patterns yield more precise and reliable predictions, emphasizing the value of personalization in predictive analytics.

Additionally, to illustrate the impact of the diary features, we compare the proposed model against a purely objective model, which excluded the diaries and relies solely on objective features, designated as 'Forecasting w/o Diary'. Our findings indicate that the inclusion of diary features leads to a boost in the model's forecasting accuracy.

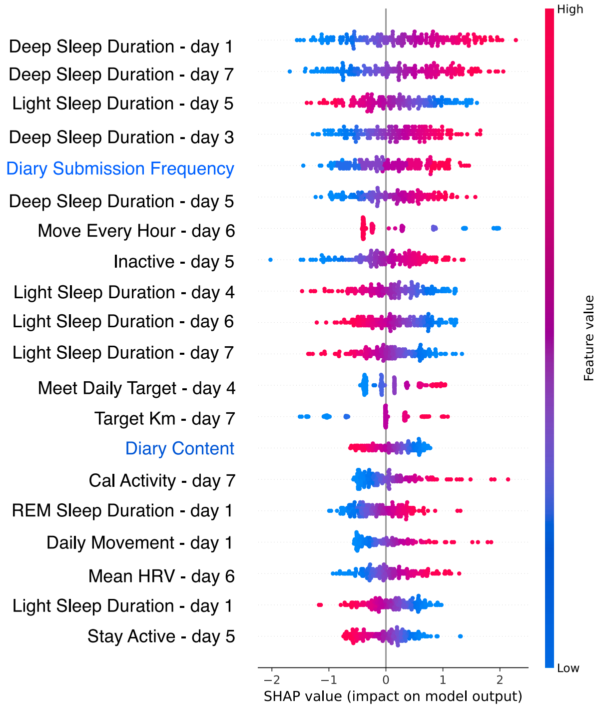

Fig. 3. Feature shapley values for positive affect forecasting.

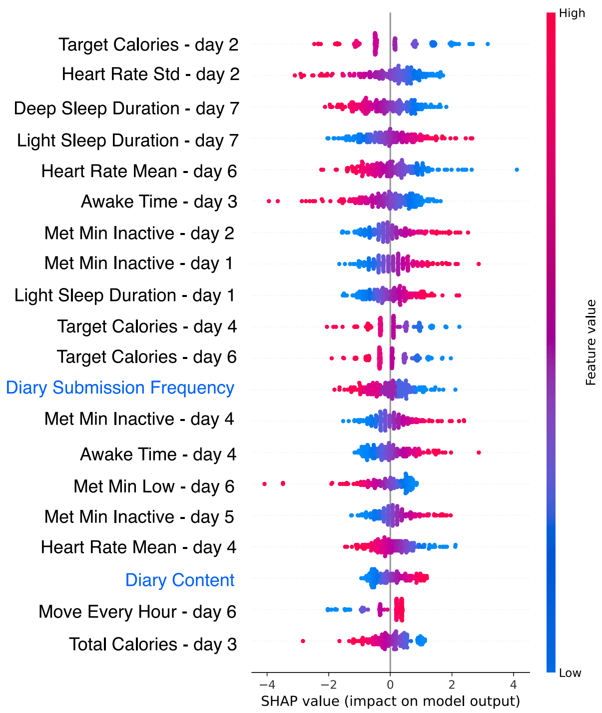

Fig. 4. Feature shapley values for negative affect forecasting.

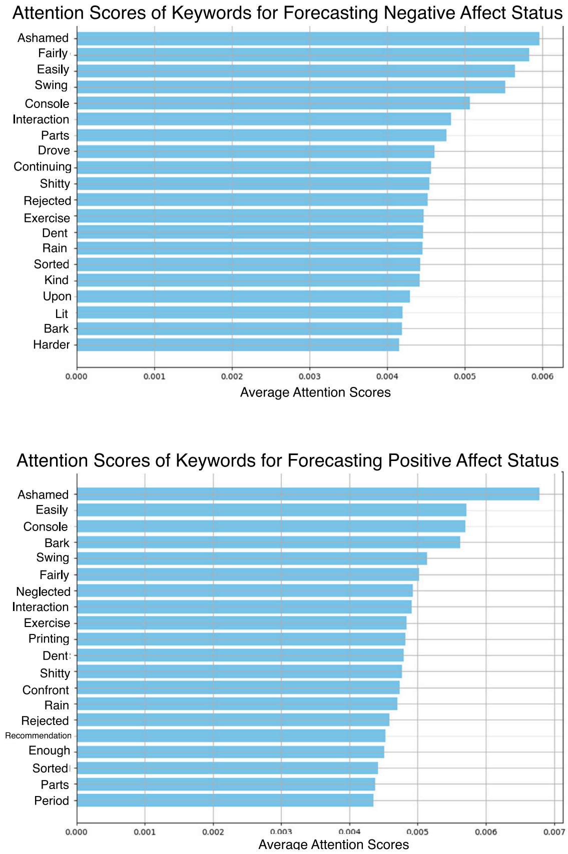

Fig. 5. Attention scores of keywords.

4.1. Features' shapley value

The Shapley values distribution of features are shown in Figs. 3 and 4. We observe that both light and deep sleep duration emerge as top-ranked predictors, with higher duration of deep sleep and lower light sleep duration contributing strongly to the prediction of a positive affect status, and lower duration is correlated to negative affect status. Diary Submission Frequency also holds considerable weight as the higher frequency of submitted diaries indicates an association with positive affect status while a lower frequency correlates with an increase in negative affect status. The physically active time and the total movement contribute to our model's positive affect predictions. Moreover, we observe that dietary metrics, such as calorie intake, along with activity metrics like MET and daily target calorie expansion, are influential in determining negative affect status, underscoring the multifaceted nature of affect status predictions.

4.2. Diary keywords attention scores

Recognizing the critical role of diary content in forecasting, we analyze the most influential keywords identified by their attention scores computed by DistillBERT. Figs. 5 demonstrate the attention scores derived from the DistilBERT model for forecasting both positive and negative affect status. The most influential keywords are largely similar across both positive and negative affect forecasting. It can be seen that the words demonstrating emotional status, such as "ashamed", "easily", and "shitty" obtain higher attention scores. Additionally, keywords pertaining to specific life events or activities, like "drove", "console", "exercise", and "rain", are also identified as significant in affect forecasting.

5. Discussion

The results of this study not only validate the feasibility of one-week forecasting but also align with the accuracy of next-day predictions reported in our previous study (Jafarlou et al., 2023), highlighting the effectiveness of our multimodal approach, especially incorporating self-reported diaries. Although forecasting a single loneliness level was previously shown to be feasible using objective features (Yang et al., 2023), affect status forecasting, as a multi-facet case, is still challenging. We show that the integration of diary data with objective modalities, such as physiological, environmental, and physical activity information, gathered from wearable and mobile devices represents a novel and efficient approach to affective computing.

In terms of explainability, we find that levels of sleep had the greatest impact on the model output. This is consistent with previous literature relating the quality of sleep to mood, where sleep deprivation is found to have a strong influence on the positive affective system (Finan et al., 2017).

Our findings also highlight certain keywords with higher attention scores in affect forecasting. The significant terms are similar across both positive and negative affect. Words demonstrating emotional status (i.e., ashamed) have a higher attention score in general, which is consistent with previous research suggesting that emotionally intense words can influence behavior and cognition (Carretié et al., 2008). Additionally, studies have found that negative words are generally stronger in extremity in comparison to positive words (Yang et al., 2013). This aligns with our findings, as the highest score of emotion-tied terms has a negative valence. We find keywords relating to day-to-day activities (i.e., rain, console) are significant in affect forecasting. Additionally, this study indicates that a higher frequency of diary submissions is associated with an increased likelihood of positive affect status and a decreased likelihood of negative affect status. Some research has found links between writing and mood fluctuations; for instance, a study found that completing daily diary entries reduced self-reported depression levels as well as influenced emotion regulation strategies (Suhr et al., 2017). Although neutral terminology may not seem significant, simply writing and sharing daily activities may have implications on mood (Yau et al., 2021). Future work could expand on specific keyword frequency, studying if the increased presence of a specific valence (positive or negative) can be associated with positive affect or negative affect forecasting, respectively.

6. Conclusion

This study introduced a multimodal deep learning model to predict affect status using a combination of data from wearable devices and self-reported diaries. The model exhibited satisfactory accuracy of 82.50% and 87.76% for positive and negative affect, respectively, affirming the practicality of forecasting affect status. A significant outcome of our study is the improvement in accuracy upon incorporating diary data, underscoring the contribution of self-reported diaries to this process. Moreover, our results stressed the importance of personalized methods in monitoring mental health, suggesting that personalized approaches yield more accurate outcomes. Future research could expand upon this study by applying the model to diverse populations and environments. This would aim to enhance the integration of both objective and self-reported diary data for more accurate predictions.

CRediT authorship contribution statement

Zhongqi Yang: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Yuning Wang:** Conceptualization, Methodology, Writing – review & editing. **Ken S. Yamashita:** Data curation, Writing – original draft, Writing – review & editing. **Elahe Khatibi:** Writing – original draft, Writing – review & editing. **Iman Azimi:** Supervision, Writing – review & editing. **Nikil Dutt:** Supervision, Writing – review & editing. **Jessica L. Borelli:** Supervision, Writing – review & editing. **Amir M. Rahmani:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Baboo, S. S., et al. (2022). Sentiment analysis and automatic emotion detection analysis of Twitter using machine learning classifiers. *International Journal of Mechanical Engineering*, *7*(2).

Carretié, L., et al. (2008). Modulation of ongoing cognitive processes by emotionally intense words. *Psychophysiology*, *45*(2), 188–196.

Deb, S., et al. (2017). Emotion classification using segmentation of vowel-like and non-vowel-like regions. *IEEE Transactions on Affective Computing*, *10*(3), 360–373.

Ekici, B., et al. (2014). The effect of emotional status and health-related quality of life on the severity of coronary artery disease. *Kardiologia Polska*, *72*(7), 617–623.

Filippini, C., et al. (2022). Automated affective computing based on bio-signals analysis and deep learning approach. *Sensors*, *22*(5), 1789.

Finan, P. H., et al. (2017). Partial sleep deprivation attenuates the positive affective system: effects across multiple measurement modalities. *Sleep*, *40*(1), zsw017.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*(2), 115–143.

Jafarlou, S., et al. (2023). Objective prediction of next-day's affect using multimodal physiological and behavioral data: Algorithm development and validation study. *JMIR Formative Research*, *7*(1), Article e39425.

Kanjo, E., et al. (2019). Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, *49*, 46–56.

Karrouri, R., et al. (2021). Major depressive disorder: Validated treatments and future challenges. *World Journal of Clinical Cases*, *9*(31), 9350.

Labbaf, S., et al. (2023). ZotCare: a flexible, personalizable, and affordable mhealth service provider. *Frontiers in Digital Health*, *5*, http://dx.doi.org/10.3389/fdgth.2023.1253087.

Lundberg, S. M., et al. (2017). A unified approach to interpreting model predictions. In *NIPS'17, vol. 30*.
Remes, O., et al. (2021). Biological, psychological, and social determinants of depression: a review of recent literature. *Brain Sciences*, *11*(12), 1633.
Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.
Sano, A. (2016). *Measuring college students' sleep, stress, mental health and wellbeing with wearable sensors and mobile phones* (Ph.D. thesis), Massachusetts Institute of Technology.
Shankar, N. L., et al. (2016). Effects of stress on students' physical and mental health and academic success. *International Journal of School & Educational Psychology*, *4*(1), 5–9.
Smidt, K. E., et al. (2015). A brief, but nuanced, review of emotional granularity and emotion differentiation research. *Current Opinion in Psychology*, *3*, 48–51.
Suhasini, M., et al. (2020). Emotion detection framework for twitter data using supervised classifiers. In *Proceedings of 3rd ICDECT-2K19* (pp. 565–576). Springer.
Suhr, M., et al. (2017). Maintaining mental health through positive writing: Effects of a resource diary on depression and emotion regulation. *Journal of Clinical Psychology*, *73*(12), 1586–1598.
Taylor, S., et al. (2017). Personalized multitask learning for predicting tomorrow's mood, stress, and health. *IEEE Transactions on Affective Computing*, *11*(2), 200–213.
Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.
Vishnubhotla, K., et al. (2022). Tweet emotion dynamics: Emotion word usage in tweets from US and Canada. arXiv preprint arXiv:2204.04862.
Wang, R., et al. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *UBICOMp 2014* (pp. 3–14).
Wang, Y., et al. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, *83*, 19–52.
Wang, Z., et al. (2024). Differential private federated transfer learning for mental health monitoring in everyday settings: A case study on stress detection. arXiv preprint arXiv:2402.10862.
Watson, D., et al. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.
Yang, J., et al. (2013). Positive words or negative words: Whose valence strength are we more sensitive to? *Brain Research*, *1533*, 91–104.
Yang, C.-J., et al. (2020). An ai-edge platform with multimodal wearable physiological signals monitoring sensors for affective computing applications. In *2020 ISCAS* (pp. 1–5). IEEE.
Yang, Z., et al. (2023). Loneliness forecasting using multi-modal wearable and mobile sensing in everyday settings. In *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)* (pp. 1–4). IEEE.
Yau, J. C., et al. (2021). Coping with stress through texting: an experimental study. *Journal of Adolescent Health*, *68*(3), 565–571.