





# MoodPupilar: Predicting Mood Through Smartphone Detected Pupillary Responses in Naturalistic Settings

Rahul Islam , Tongze Zhang , Priyanshu Singh Bisen , Sang Won Bae   
Charles V. Schaefer, Jr. School of Engineering and Science  
Stevens Institute of Technology  
Hoboken, USA

**Abstract**—MoodPupilar introduces a novel method for mood evaluation using pupillary response captured by a smartphone’s front-facing camera during daily use. Over a four-week period, data was gathered from 25 participants to develop models capable of predicting daily mood averages. Utilizing the GLOBEM behavior modeling platform, we benchmarked the utility of pupillary response as a predictor for mood. Our proposed model demonstrated a Matthew’s Correlation Coefficient (MCC) score of 0.15 for Valence and 0.12 for Arousal, which is on par with or exceeds those achieved by existing behavioral modeling algorithms supported by GLOBEM. This capability to accurately predict mood trends underscores the effectiveness of pupillary response data in providing crucial insights for timely mental health interventions and resource allocation. The outcomes are encouraging, demonstrating the potential of real-time and predictive mood analysis to support mental health interventions.

## I. INTRODUCTION

Smartphones are central to modern life, enabling everything from communication to managing finances and accessing healthcare services. This ubiquity provides a unique opportunity to monitor mental health by analyzing usage patterns and capturing candid facial expressions and pupillary responses through the front-facing camera. Unlike posed selfies, these candid captures offer a genuine snapshot of emotional states, free from social masking. Advances in machine learning and artificial intelligence enhance affective computing, improving the accuracy of mood detection [1]–[4]. This accuracy facilitates interventions like mood-based music recommendations or prompts for social support, expanding the role of smartphones in human-centered computing, ambient intelligence, and interactive design.

The current state of mood tracking relies on self-reported questions, such as those available on Apple Fitness or Google Fit, which depend on subjective biased recall of one’s mood at any given moment. Typically, these reported moods are used as a mood diary for user reference. Most recently, research has been focusing on detecting mood states using sensors that collect personal social and behavioral data. Studies [1]–[4] often use data from smartphone app usage, wearable devices, and GPS to predict mood, achieving significant accuracy and F1 scores in classifying various mood states like valence and arousal. This reflects a move towards less intrusive, sensor-

based mood detection technologies that offer accurate and immediate evaluations with minimal user effort.

However, mood characterization extends beyond social and behavioral constructs to include physiological signals. Studies indicate that physiological signals from pupillary response correlate with mood states, demonstrating nuanced interactions between pupil dilation and emotional, cognitive processing. For instance, sustained pupil dilation linked to depression shows heightened sensitivity to negative stimuli, decreases after induced negative moods, suggesting a mechanism for maintaining depressive states [5]. These observations in controlled settings highlight potential insights into natural environments.

In our work, we aim to bridge the gap between lab-based research and real-world applications by integrating physiological signals to detect mood using mobile sensing. We employ an affective mobile sensing system [6], [7] that operates in the background of a user’s smartphone, opportunistically collecting images of the pupil during the user’s daily routine. We then use the pupillary response measures computed from these images to create a model for detecting the individual’s average daily mood. Our proposed method supports the use of pupillary response as a physiological signal for mood detection, arguing for its incremental utility in mobile sensing when combined with other sensors to enhance mood detection.

## II. RELATED WORK

### A. Understanding Mood Using Pupillary Response

In the exploration of mood detection through pupillary response, several recent studies have demonstrated the nuanced relationship between pupil dilation and emotional and cognitive processing. For instance, a study has shown that individuals with a history of depression exhibit more sustained pupil dilation to negative emotional stimuli, a response that tends to decrease following a negative mood induction, indicating a heightened sensitivity and subsequent cognitive blunting in response to prolonged negative stimuli [5]. In social contexts, pupillary contagion—where observers’ pupils dilate in response to others’ dilated pupils—has been observed regardless of the emotional expression presented, with socioeconomic status influencing the degree of contagion in response to emotionally neutral stimuli [8]. Moreover, physiological arousal linked to very light exercise has been cor-

related with pupil dilation, identifying the ascending arousal system’s activation at surprisingly low levels of physical exertion [9]. Furthermore, sustained pupil dilation has been linked to self-reported rumination in depressed individuals, suggesting a potential mechanism for the maintenance of depressive states through prolonged emotional processing [10]. The most recent advances include employing sophisticated eye-tracking technology to measure pupillary responses during exposure to emotional video scenarios, achieving significant classification accuracies in detecting emotions such as fear, anger, and surprise. These findings underscore the potential of pupillometry in enhancing our understanding of mood and emotional reactivity, promising applications in affective computing and mental health diagnostics. However, this research into pupillary response and mood has primarily taken place in controlled laboratory environments. Nonetheless, observing human behavior and physiological reactions in natural settings can offer distinct insights.

### B. Mood Detection Using Mobile Sensing

Traditionally, Mood detection via mobile sensing has relied on behavioral and social data gathered from smartphone sensors, with several studies achieving noteworthy accuracies in mood prediction [1]–[4]. For instance, LiKamwa et al. [1] reached up to 67% accuracy in mood inference using smartphone app usage data to model mood. Kang et al. [2] achieved an F1 score of 0.543 for valence and 0.534 for arousal utilizing data from both smartphones and wearable devices. More recently, Meegahapola et al. [3] conducted an extensive study to develop a model that classifies valence into high and low categories, achieving an AUC of 0.51. Additionally, Jacob et al. [4] leveraged IMU data from smartphones to predict mood states, obtaining an F1 score of 0.897 for both valence and arousal. These developments highlight a trend towards more unobtrusive, sensor-based mood tracking technologies that reduce the effort required from users while delivering accurate and immediate mood evaluations.

Behavioral data captures long-term trends but often misses rapid emotional changes that physiological signals like pupillary responses can detect. Integrating both data types could improve mood detection systems’ robustness. Studies in controlled settings, such as those by Lee et al. [11], show that eye-tracking metrics can effectively classify emotions with high accuracy. However, the limited real-world applicability of these controlled studies underscores the need for more ecologically valid research. This calls for developing methods that analyze behavioral and physiological data in naturalistic settings to better reflect actual user experiences and behaviors.

## III. METHOD

### A. Data Collection

Our research utilized the affective mobile sensing system, FacePsy, which operates in the background on Android smartphones. This system opportunistically collects images of the pupil when the user unlocks their phone or opens any of 35 predefined trigger apps categorized into communication,

social media, productivity, entertainment, and health. After activation, FacePsy captures eye images for 10 seconds, processes them to detect and crop the images of the left and right eyes, and transmits these for Pupil-Iris Ratio (PIR) estimation as a measure of pupillary response. This estimation employs the deep learning-based methods detailed in our PupilSense [7] research. For further details, please see FacePsy [6] and PupilSense [7].

### B. Research Design and Participant Demographics

1) *Protocol*: The study was conducted remotely during the COVID-19 Public Health Emergency, approved by the Institutional Review Board (IRB). It involved participants aged 18 or older, owning an Android phone. Participants completed baseline activities and engaged with daily mood surveys using the Circumplex Model of Affect (CMA) three times daily for four weeks. Compliance was incentivized with up to \$135 in compensation. Further procedural details are available in the [6].

2) *Participants*: We recruited 38 participants, with 25 completing the study. Reasons for non-completion included excessive battery usage, personal reasons, and failure to meet survey requirements. The participant demographics consisted of 11 males, 8 females, and 6 unspecified, with an average age of 27.88. Most were Asian (15) or Caucasian (4), with varying education levels from high school diplomas to master’s degrees. Details on mental health ratings and substance use are reported in the study results.

### C. Dataset

Our system triggered 15,995 data collection instances, capturing eye images during phone interactions. Out of these, 8,299 instances successfully underwent Pupil-Iris Ratio (PIR) estimation, averaging 11.85 daily PIR estimations per participant. We filtered these instances based on the acceptable PIR range of 0.2 (highly constricted) to 0.7 (highly dilated) established by prior research [12], resulting in 6,657 usable instances.

Data were segmented into four daily periods—midnight, morning, afternoon, and evening—analyzing statistical features such as sum, minimum, maximum, average, median, and standard deviation of PIR estimations. Data gaps from participant inactivity or errors in estimation were addressed by imputing missing values with daily means. Each participant’s daily data included 48 features and was labeled with their average mood scores. The total potential data coverage spanned 616 participant days. Of these, 55 days lacked eye images due to participant-reported unavailability (e.g., vacations), and 33 days were excluded due to poor image quality or non-responsiveness to mood surveys. This left 528 analyzable days in our dataset. Participants completed mood assessments via the CMA survey three times daily over four weeks. Despite collecting 577 days of mood data, due to incomplete daily data, we matched mood reports with PIR data for only 470 days.

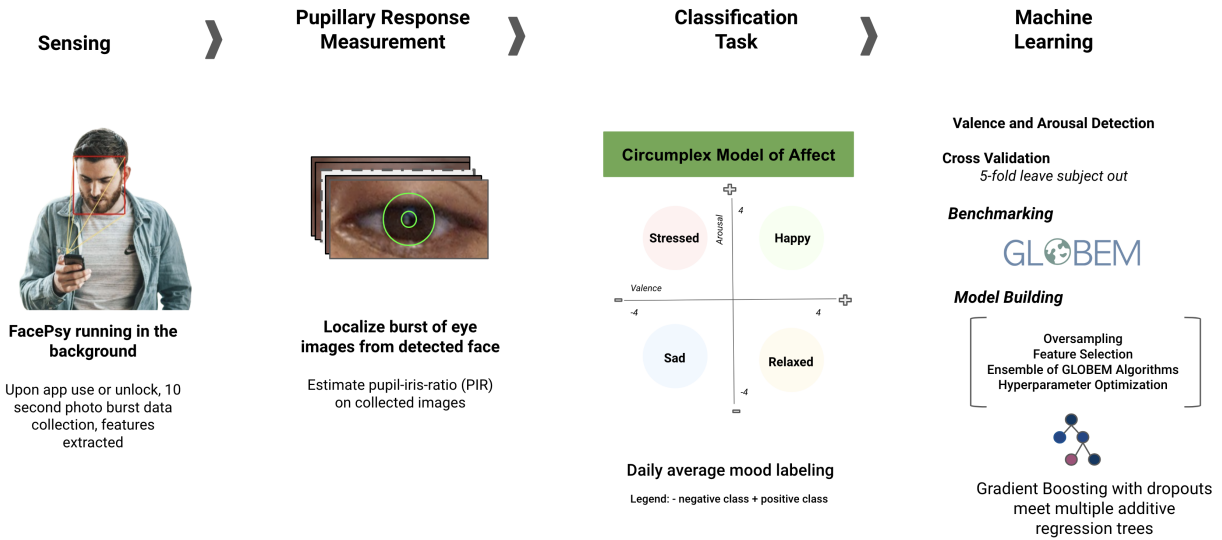


Fig. 1: Our mood sensing pipeline through pupillary response.

#### D. Classification Framework

Our study employs the Circumplex Model of Affect (CMA) to evaluate mood across two primary dimensions: valence and arousal. Each dimension is scored on a scale from -4 to 4. Consistent with methodologies established in prior research such as Likamwa et al. [1] and Meegahapola et al. [3], we adapt these scales into binary classifications for analytical purposes. Specifically, a score below 0 is categorized as low valence or arousal, whereas a score of 0 or above is categorized as high. This binary classification facilitates the development of mood prediction models that aim to capture the daily averages of valence and arousal. Reliable predictions of daily mood averages could serve multiple purposes. They reflect a user’s current emotional state and help identify patterns or shifts in mood, indicating changes from their typical emotional behavior.

1) *Benchmarking*: With the daily average mood model described above, we benchmark the valence and arousal detection on the GLOBEM Platform [13], which supports the flexible and rapid evaluation of existing behavior modeling methods. We specifically benchmark our dataset using existing machine learning algorithms i.e Canzian et al. [14], Farhan et al. [15], Xu et al. [16], Lu et al. [17], Saeb et al. [18], Wahle et al. [19], Wang et al. [20] supported by GLOBEM.

2) *Our Model*: We further propose our model for valence and arousal detection, which is an ensemble model with feature selection using LightGBM, a gradient boosting algorithm. We employed algorithms supported in GLOBEM [13] (i.e Canzian et al. [14], Farhan et al. [15], Xu et al. [16], Lu et al. [17], Saeb et al. [18], Wahle et al. [19], Wang et al. [20]) as base learners, using the best hyperparameters. LightGBM serves as the meta learner.

Informed by prior research [13] in behavior modeling, we use a *5-fold leave subject out cross validation* to evaluate each model in benchmarking and our proposed model. This ensures

that all data from a single participant are exclusively allocated to either the training, validation, or testing phase, but never shared across these subsets. Additionally, we implement hyperparameter optimization in our training dataset for tuning hyperparameters. This subject-independent partitioning approach, combined with cross-validation, enhances the robustness our results compared to methods using a simple train-test split. To evaluate the performance of our model, we use balanced accuracy and Matthew’s Correlation Coefficient (MCC) as our metrics. These metrics were selected because they offer a thorough evaluation of model performance. Specifically, MCC incorporates all elements of the confusion matrix, making it a strong alternative to the F1 score for binary classification tasks, while balanced accuracy equally considers the rates of true positives and true negatives.

#### IV. RESULTS AND DISCUSSION

Benchmarking our smartphone-collected pupillary response data for mood detection on the GLOBEM Platform against existing behavioral modeling algorithms indicates generally low performance across the board. Notably, the algorithms by Canzian et al. [14], Saeb et al. [18], Wahle et al. [19], and Wang et al. [20] received negative MCC scores, suggesting that these models might be performing no better than random guessing in some cases, particularly in their ability to reliably predict mood based on the data. In contrast, our model, MoodPupilar, shows improved performance with the highest scores in both Balanced Accuracy and Matthews Correlation Coefficient among the compared models. It achieved a BA of 0.63 and an MCC of 0.15 for valence, and a BA of 0.56 and an MCC of 0.12 for arousal. This indicates a more reliable and effective prediction capability in mood detection using pupillary response compared to the existing approaches on the GLOBEM Platform, which is specifically designed for depression. While individual models on the GLOBEM

platform may not perform well, an ensemble model built with the GLOBEM algorithm could offer superior performance.

TABLE I: Result on the GLOBEM Platform and our model for 5-fold leave-subject-out cross-validation for valence and arousal (BA: Balanced Accuracy, MCC: Matthews correlation coefficients).

Model	Valence		Arousal	
	BA	MCC	BA	MCC
GLOBEM Benchmark				
Canzian et al. [14]	0.60	-0.02	0.50	-0.01
Farhan et al. [15]	0.60	0.03	0.53	0.06
Xu et al. [16]	0.51	0.11	0.55	0.09
Lu et al. [17]	0.63	0.08	0.51	-0.03
Saeb et al. [18]	0.60	-0.01	0.50	-0.03
Wahle et al. [19]	0.60	-0.01	0.50	-0.01
Wang et al. [20]	0.60	-0.01	0.50	-0.03
Our Model				
MoodPupilar	<b>0.63</b>	<b>0.15</b>	<b>0.56</b>	<b>0.12</b>

Current mood-tracking methodologies range from self-reported questionnaires, like those in Apple Fitness or Google Fit, to sensor-based approaches using smartphone and wearable data. Traditional methods rely on users recalling and reporting their mood, often used as a personal mood diary. In contrast, recent research [1]–[4] employs sensors such as smartphone app usage, wearable device metrics, and IMU data to objectively assess mood states, achieving notable accuracies in detecting emotional states such as valence and arousal. This shift towards less intrusive, mobile sensing-based technologies allows for more accurate and immediate mood assessments with reduced user involvement [1]–[4]. However, the current mobile sensing-based solution only explores social and behavioral constructs of mood characterization. Research [5], [8] shows that physiological signals, such as pupillary response, are linked to mood states, revealing complex relationships between pupil dilation and mood. The potential use cases for our findings are extensive. Implementing a model like MoodPupilar could automate mood tracking, providing more accurate, objective, and real-time assessments of mood that reflect a user’s immediate emotional state without the need for manual input. This could revolutionize the way mood variations are monitored, offering new insights into mental health and well-being.

## REFERENCES

- [1] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, “Moodscope: Building a mood sensor from smartphone usage patterns,” in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, 2013, pp. 389–402.
- [2] S. Kang, W. Choi, C. Y. Park, N. Cha, A. Kim, A. H. Khandoker, L. Hadjileontiadis, H. Kim, Y. Jeong, and U. Lee, “K-emophone: A mobile and wearable dataset with in-situ emotion, stress, and attention labels,” *Scientific data*, vol. 10, no. 1, p. 351, 2023.
- [3] L. Meegahapola, W. Droz, P. Kun, A. De Götzen, C. Nutakki, S. Diwakar, S. R. Correa, D. Song, H. Xu, M. Bidoglia et al., “Generalization and personalization of mobile sensing-based mood inference models: an analysis of college students in eight countries,” *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 6, no. 4, pp. 1–32, 2023.
- [4] S. Jacob, P. Vinod, A. Subramanian, and V. G. Menon, “Affect sensing from smartphones through touch and motion contexts,” *Multimedia Systems*, vol. 29, no. 5, pp. 2495–2509, 2023.
- [5] D. Steidtmann, R. E. Ingram, and G. J. Siegle, “Pupil response to negative emotional information in individuals at risk for depression,” *Cognition and Emotion*, vol. 24, no. 3, pp. 480–496, 2010.
- [6] R. Islam and S. W. Bae, “Facepsy: An open-source affective mobile sensing system – analyzing facial behavior and head gesture for depression detection in naturalistic settings,” in *Proceedings of the ACM International Conference on Mobile Human-Computer Interaction (MobileHCI)*, 2024.
- [7] M. R. Islam and S. W. Bae, “Pupilsense: Detection of depressive episodes through pupillary response in the wild,” in *International Conference on Activity and Behavior Computing*, 2024.
- [8] C. Fawcett, E. Nordenswan, S. Yrttiaho, T. Häikiö, R. Korja, L. Karlsson, H. Karlsson, and E.-L. Kataja, “Individual differences in pupil dilation to others’ emotional and neutral eyes with varying pupil sizes,” *Cognition and Emotion*, vol. 36, no. 5, pp. 928–942, 2022.
- [9] R. Kuwamizu, Y. Yamazaki, N. Aoike, G. Ochi, K. Suwabe, and H. Soya, “Pupil-linked arousal with very light exercise: pattern of pupil dilation during graded exercise,” *The Journal of Physiological Sciences*, vol. 72, no. 1, p. 23, 2022.
- [10] G. J. Siegle, S. R. Steinhauer, C. S. Carter, W. Ramel, and M. E. Thase, “Do the seconds turn into hours? relationships between sustained pupil dilation in response to emotional information and self-reported rumination,” *Cognitive Therapy and Research*, vol. 27, pp. 365–382, 2003.
- [11] C.-L. Lee, W. Pei, Y.-C. Lin, A. Granmo, and K.-H. Liu, “Emotion detection based on pupil variation,” *Healthcare*, vol. 11, no. 3, 2023. [Online]. Available: <https://www.mdpi.com/2227-9032/11/3/322>
- [12] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, “Pupil dilation degrades iris biometric performance,” *Computer vision and image understanding*, vol. 113, no. 1, pp. 150–157, 2009.
- [13] X. Xu, X. Liu, H. Zhang, W. Wang, S. Nepal, Y. Sefidgar, W. Seo, K. S. Kuehn, J. F. Huckins, M. E. Morris et al., “Globem: Cross-dataset generalization of longitudinal human behavior modeling,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–34, 2023.
- [14] L. Canzian and M. Musolesi, “Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis,” in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 1293–1304.
- [15] A. A. Farhan, C. Yue, R. Morillo, S. Ware, J. Lu, J. Bi, J. Kamath, A. Russell, A. Bamis, and B. Wang, “Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data,” in *2016 IEEE wireless health (WH)*. IEEE, 2016, pp. 1–8.
- [16] X. Xu, P. Chikersal, A. Doryab, D. K. Villalba, J. M. Dutcher, M. J. Tumminia, T. Althoff, S. Cohen, K. G. Creswell, J. D. Creswell et al., “Leveraging routine behavior and contextually-filtered features for depression detection among college students,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–33, 2019.
- [17] J. Lu, C. Shang, C. Yue, R. Morillo, S. Ware, J. Kamath, A. Bamis, A. Russell, B. Wang, and J. Bi, “Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.
- [18] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, D. C. Mohr et al., “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study,” *Journal of medical Internet research*, vol. 17, no. 7, p. e4273, 2015.
- [19] F. Wahle, T. Kowatsch, E. Fleisch, M. Rufer, S. Weidt et al., “Mobile sensing and support for people with depression: a pilot trial in the wild,” *JMIR mHealth and uHealth*, vol. 4, no. 3, p. e5960, 2016.
- [20] R. Wang, W. Wang, A. DaSilva, J. F. Huckins, W. M. Kelley, T. F. Heatherton, and A. T. Campbell, “Tracking depression dynamics in college students using mobile phone and wearable sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–26, 2018.