Towards Reducing Continuous Emotion Annotation Effort During Video Consumption: A Physiological Response Profiling Approach

Author(s): **Swarnali Banik**, Sougata Sen, Snehanshu Saha, Surjya Ghosh

Computer Science and Information Systems, APPCAIR

BITS Pilani Goa, INDIA

UBICOMP '24: The 2024 ACM International Joint Conference on Pervasive and Ubiquitous computing

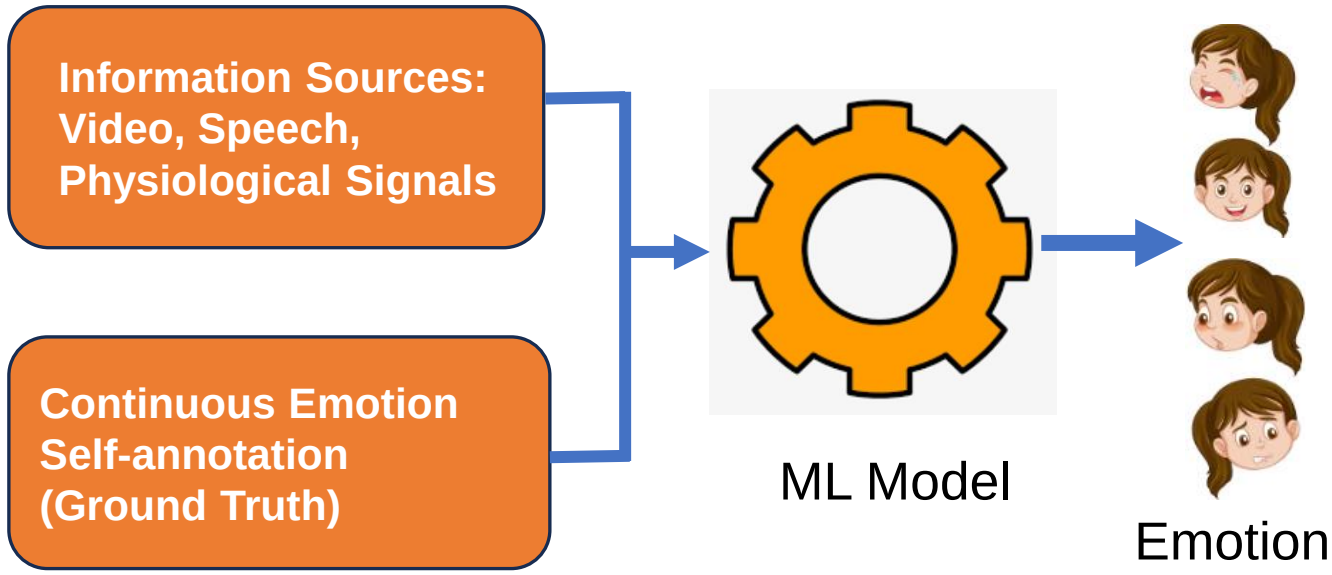
UBICOMP ISWC 2024 MELBOURNE, AUSTRALIA

APPCAIR

BITS Pilani
K K Birla Goa Campus

Motivation and Background

- Video-based applications
- Emotion Detection Using ML Models : A Multi-Modal Approach
  - Integrating Multiple Modalities
  - Emotion ground truth



Typical ML-based emotion detection model

Motivation and Background

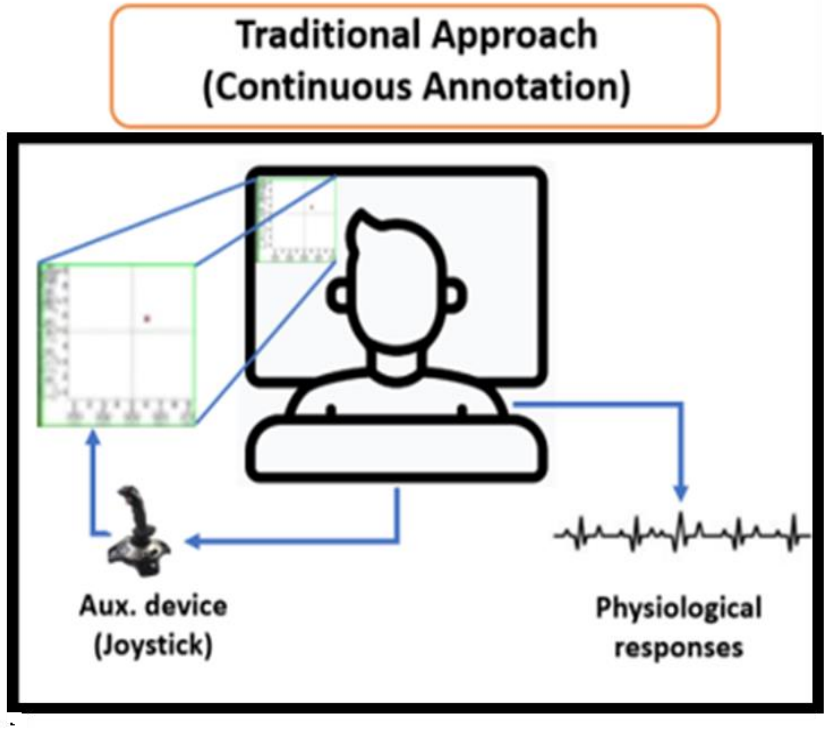
Emotion Self-Annotation

- **Post-interaction/ post stimuli**
  - Fails to capturing subtle intra-video nuances.

Motivation and Background

Emotion Self-Annotation

- **Continuous annotation**
  - Degrades viewing experience.
  - Increases cognitive load, leading to less accurate emotion labels.



Participant watching videos and annotate continuously

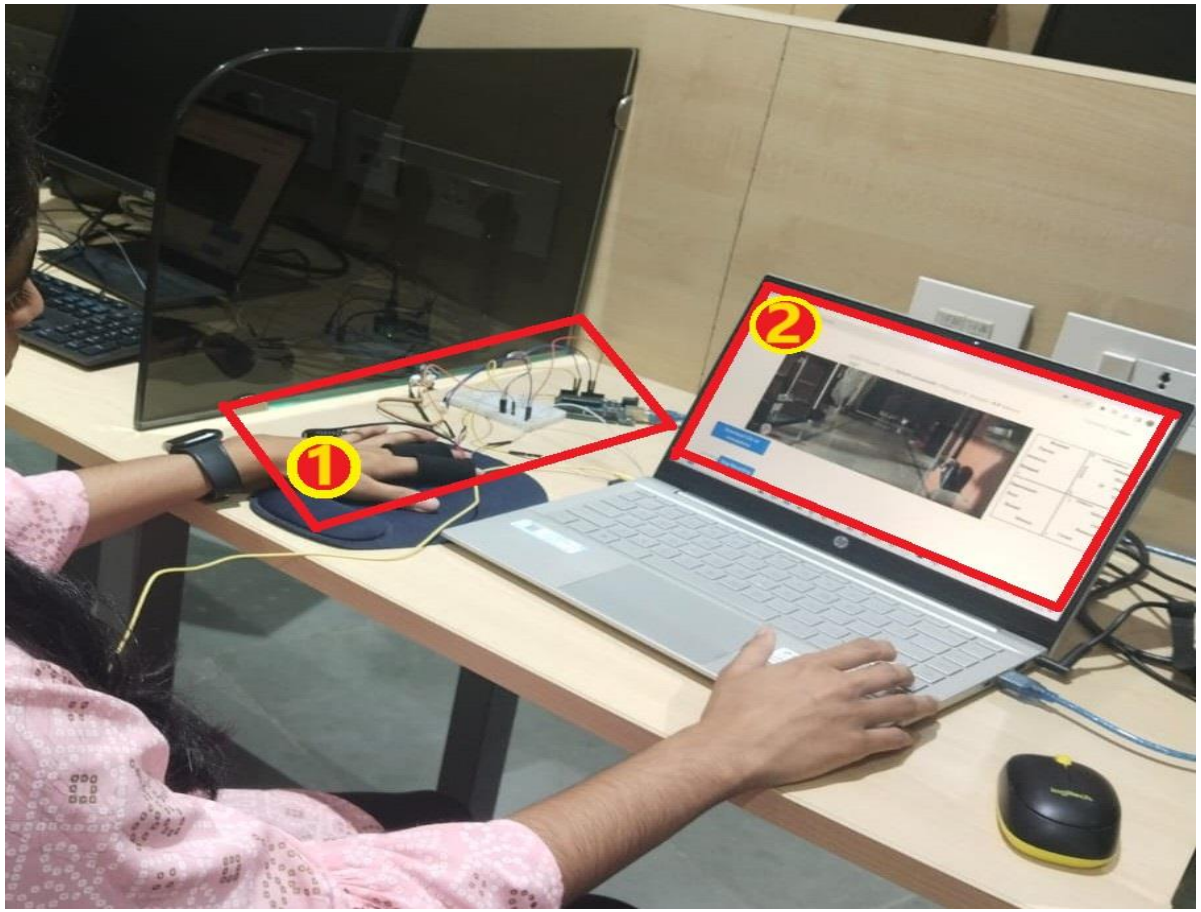
Research Question

Can we developed an intelligent annotation framework ?

- Capture ***opportune moments*** for emotion self-report collection
- Reducing the number of self-annotations (probes)
  - Decreases the cognitive workload

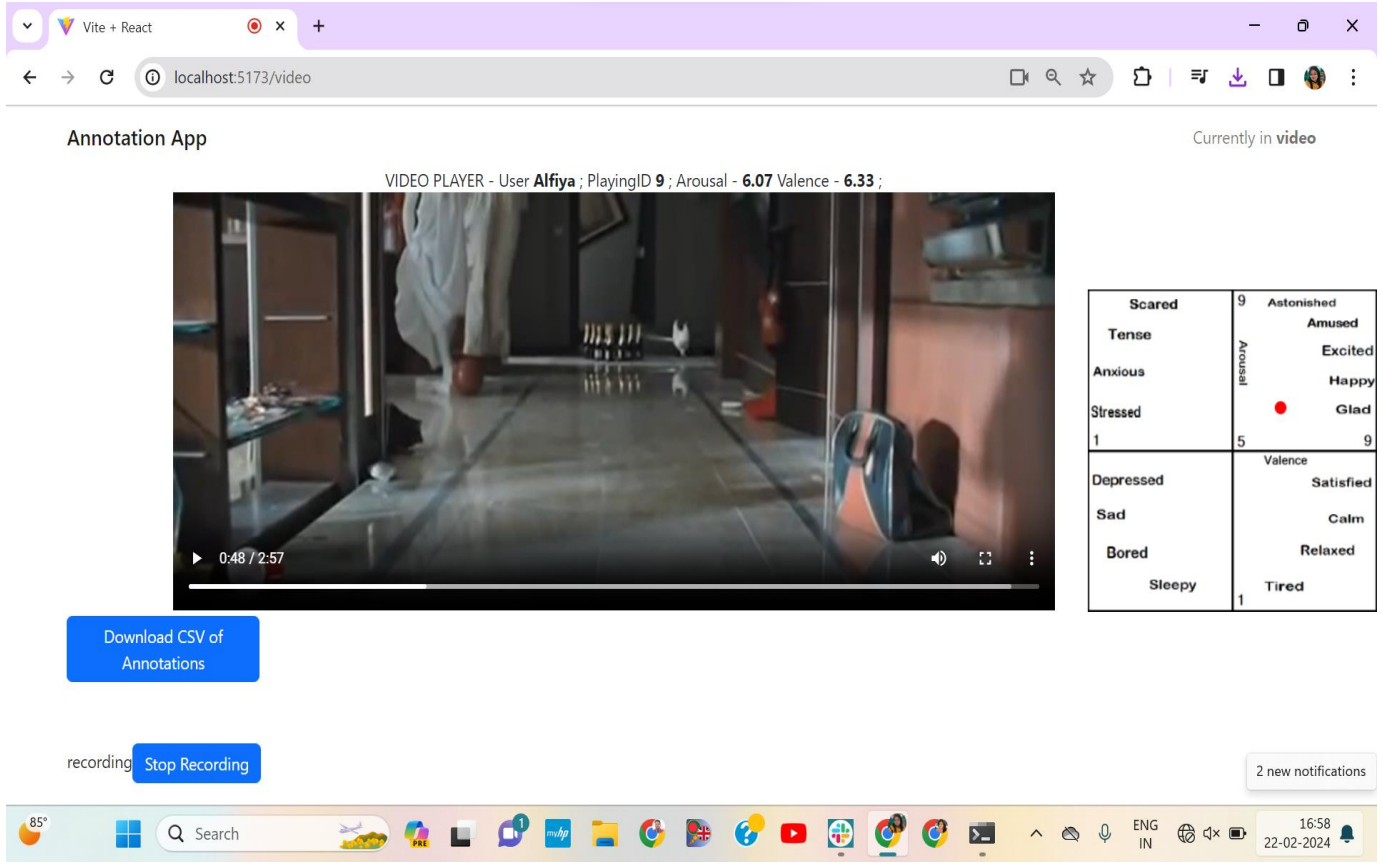
User Study I

- **Study procedure**
  - 36 participants (18M, 18F).



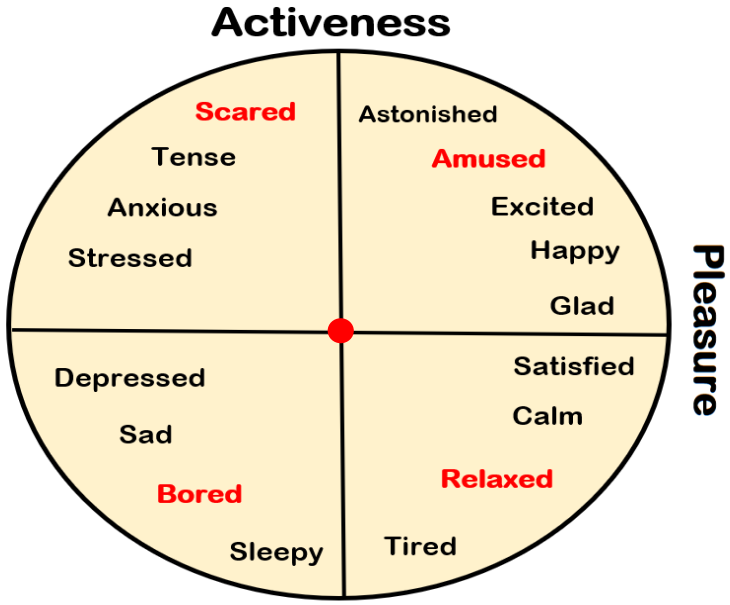
User Study I

- **Experiment apparatus**
  - Annotate app



User Study I

- **Study procedure**
  - Used Circumplex model.
  - Record valence and arousal rating on a 9-point scale.



Circumplex Model

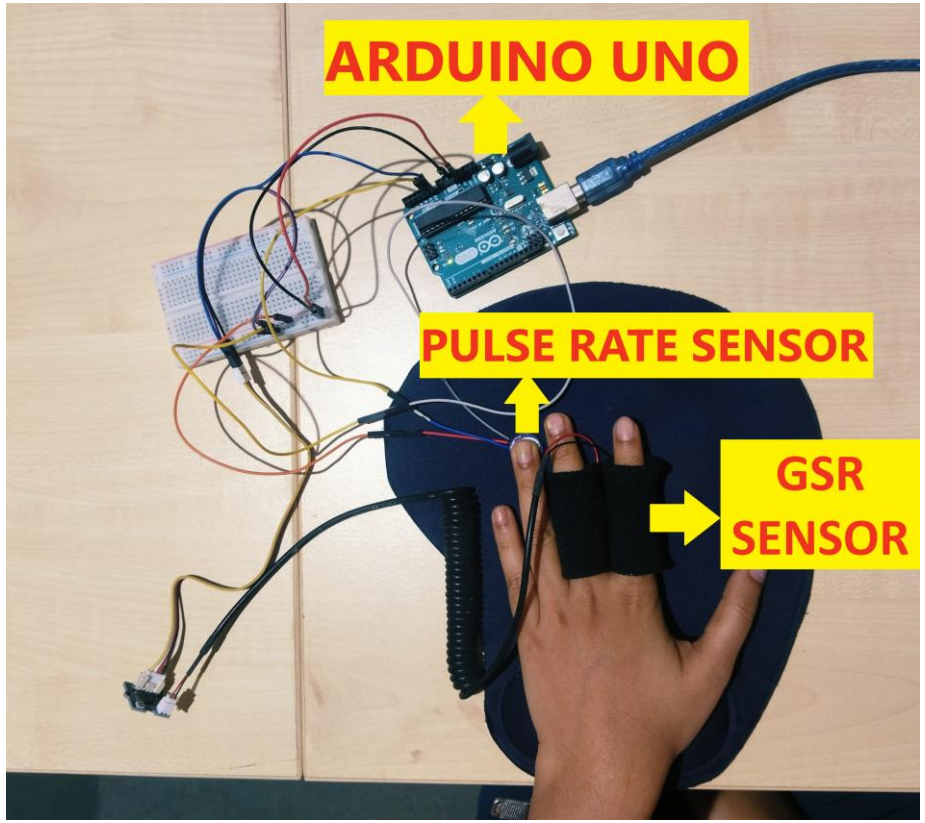
User Study I

- **Study procedure**
  - 8 stimuli videos in random order

Video id	Emotion	Valence	Arousal	Duration (in sec.)
1	amusing	med/high	med/high	185
2	amusing	med/high	med/high	173
3	boring	low	low	119
4	boring	low	low	160
5	relaxing	med/high	low	145
6	relaxing	med/high	low	147
7	scary	low	high	197
8	scary	low	high	144

User Study I

- **Experiment apparatus**
  - Galvanic skin response sensor (GSR V1.2)
  - Pulse rate sensor (HW-827)

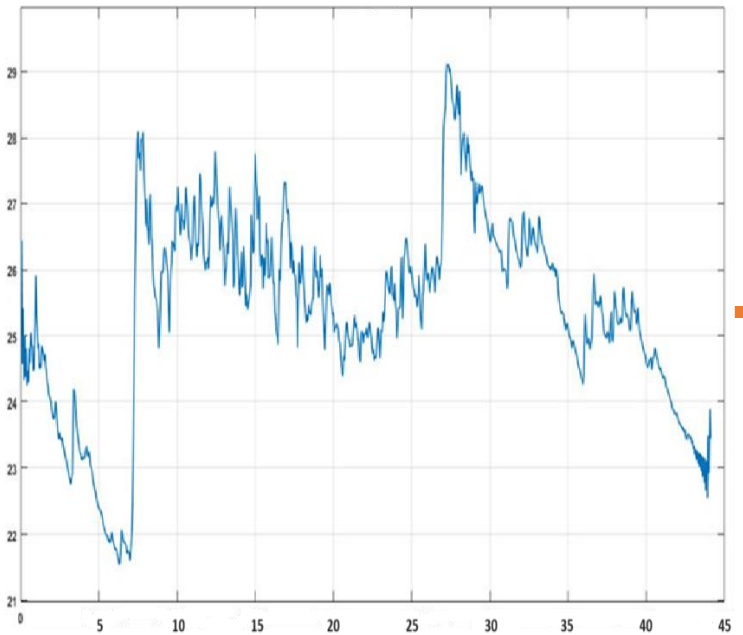


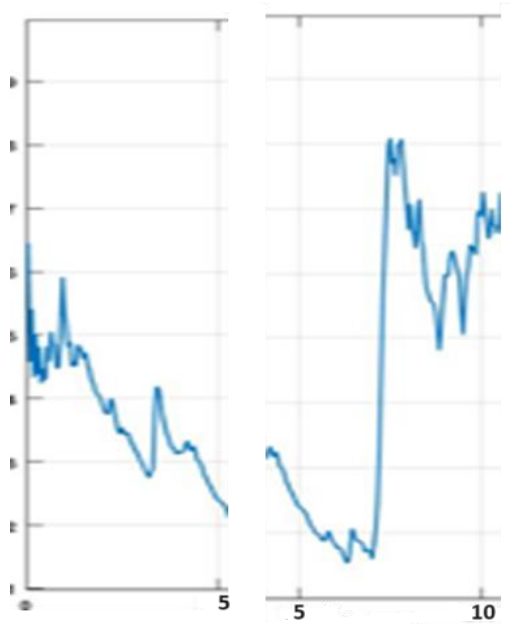
Dataset: Data Pre-processing

Physiological Responses

Segment signals into fixed size windows (5 second)

Compute the Change Point Score for two consecutive windows

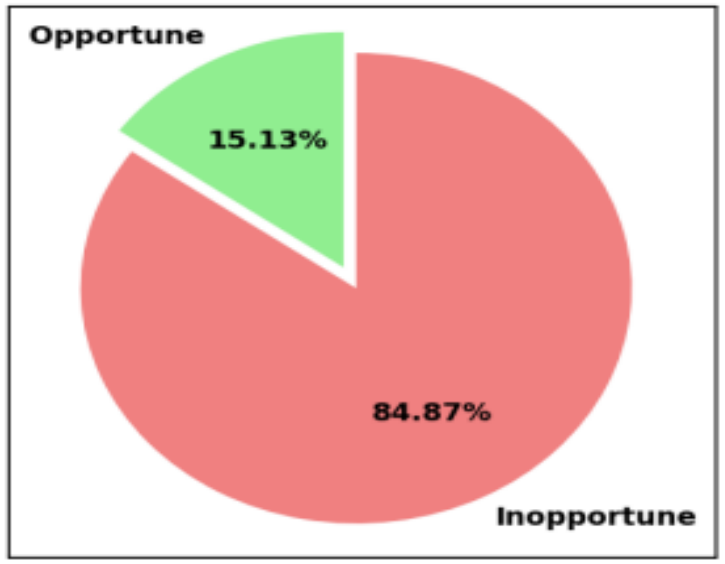




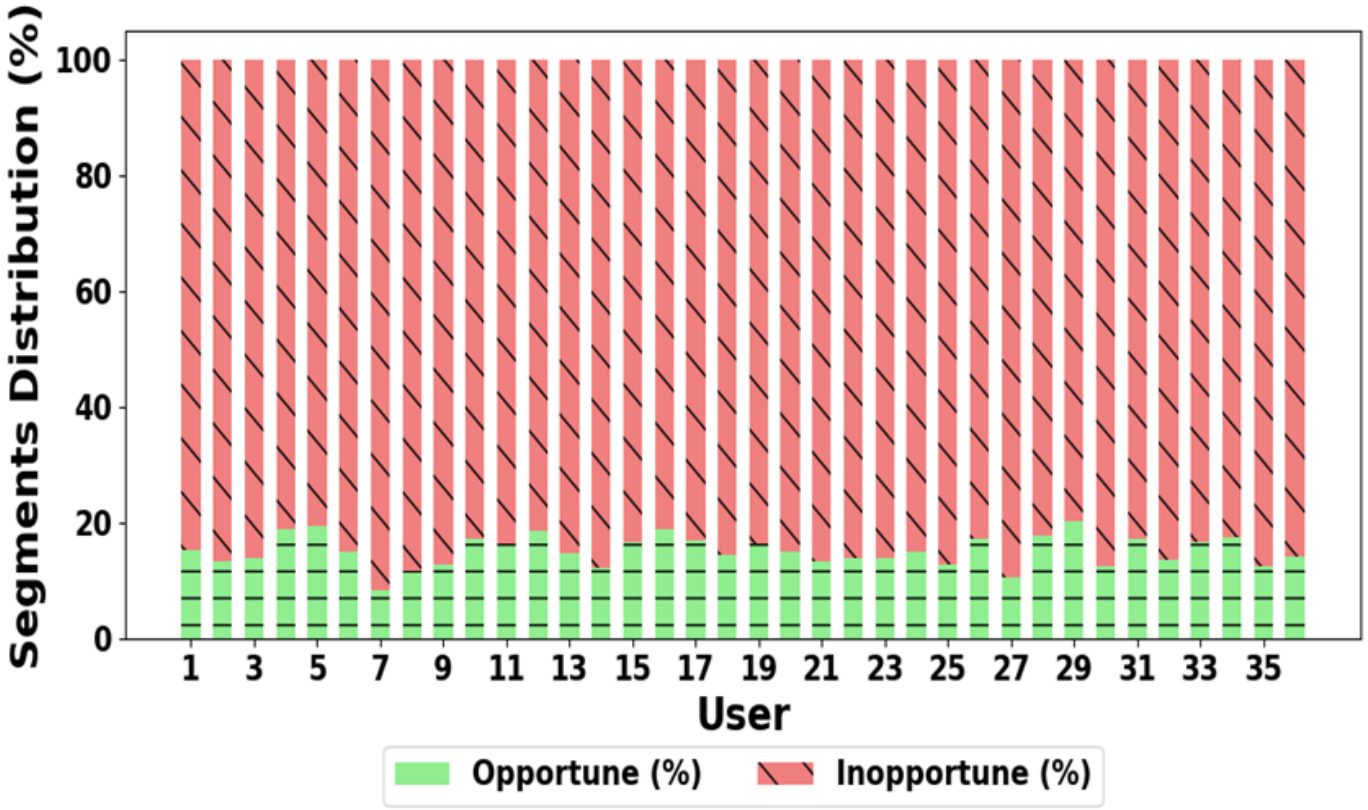
Change Point Score

Dataset: Data Description

- **Dataset description:**
  - Total 8608 segments.



Segment distribution in total



User-wise distribution of opportune and inopportune segments

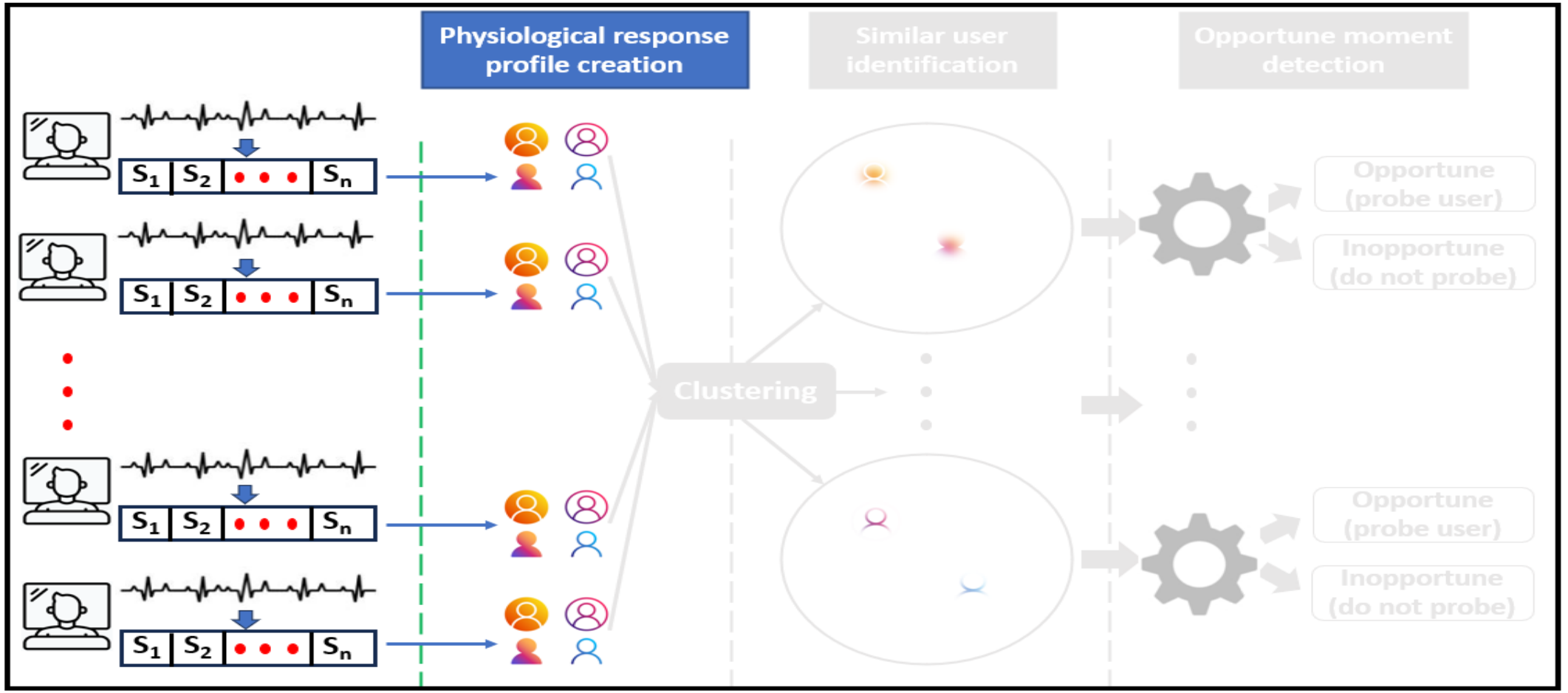
PResUP Framework

PResUP: Physiological Response based User Profiling

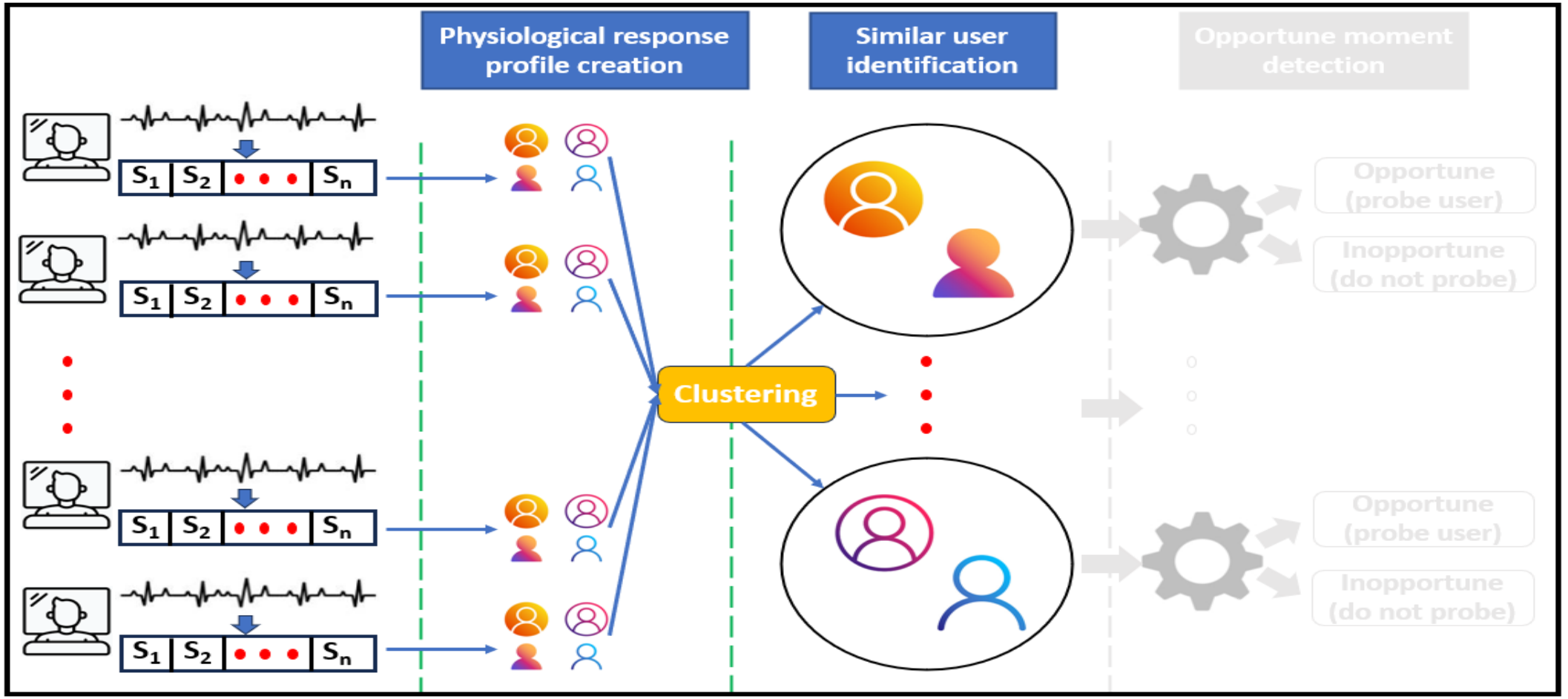
- Reduces continuous emotion self-report annotation effort
- Detects opportune moments

Overview of the **PResUP** framework

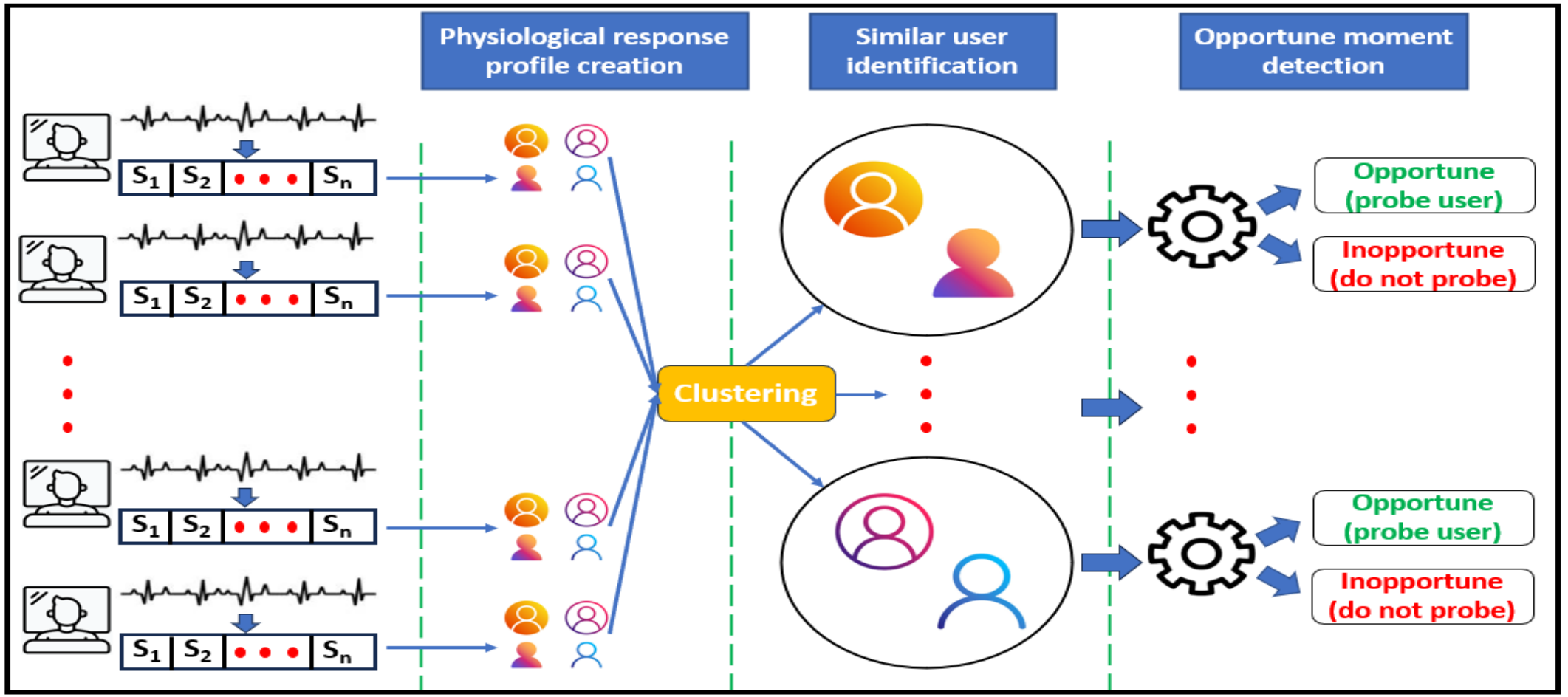
PResUP Framework



PResUP Framework

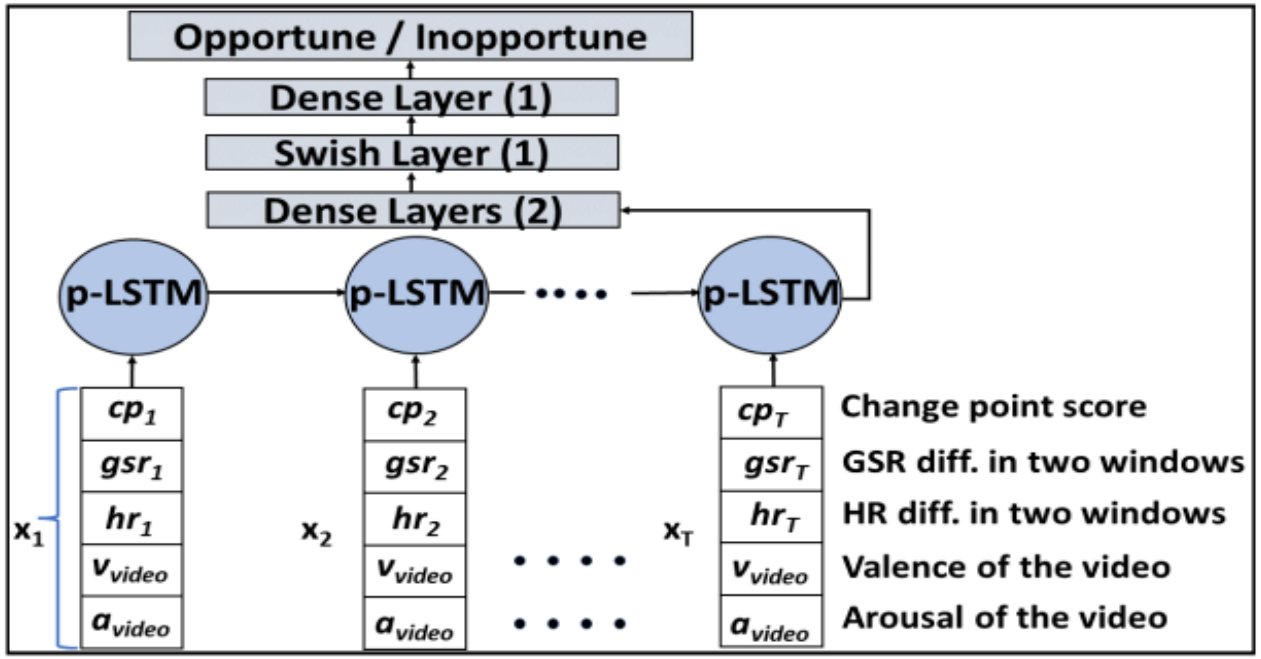


PResUP Framework



PResUP Framework

The parameterized Elliott activation functions: $f(x) = \frac{px}{1 + |x|}$

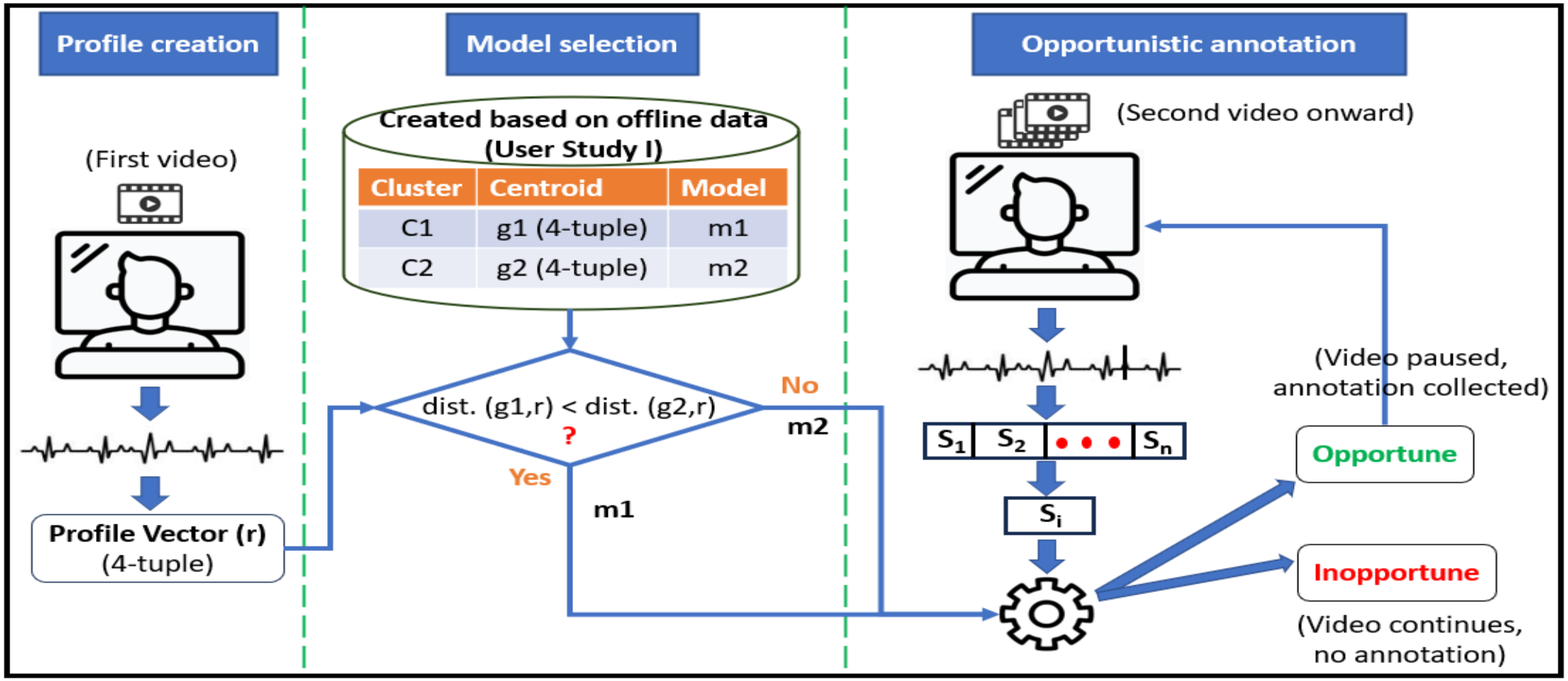


p-LSTM-based architecture used in the PResUP framework

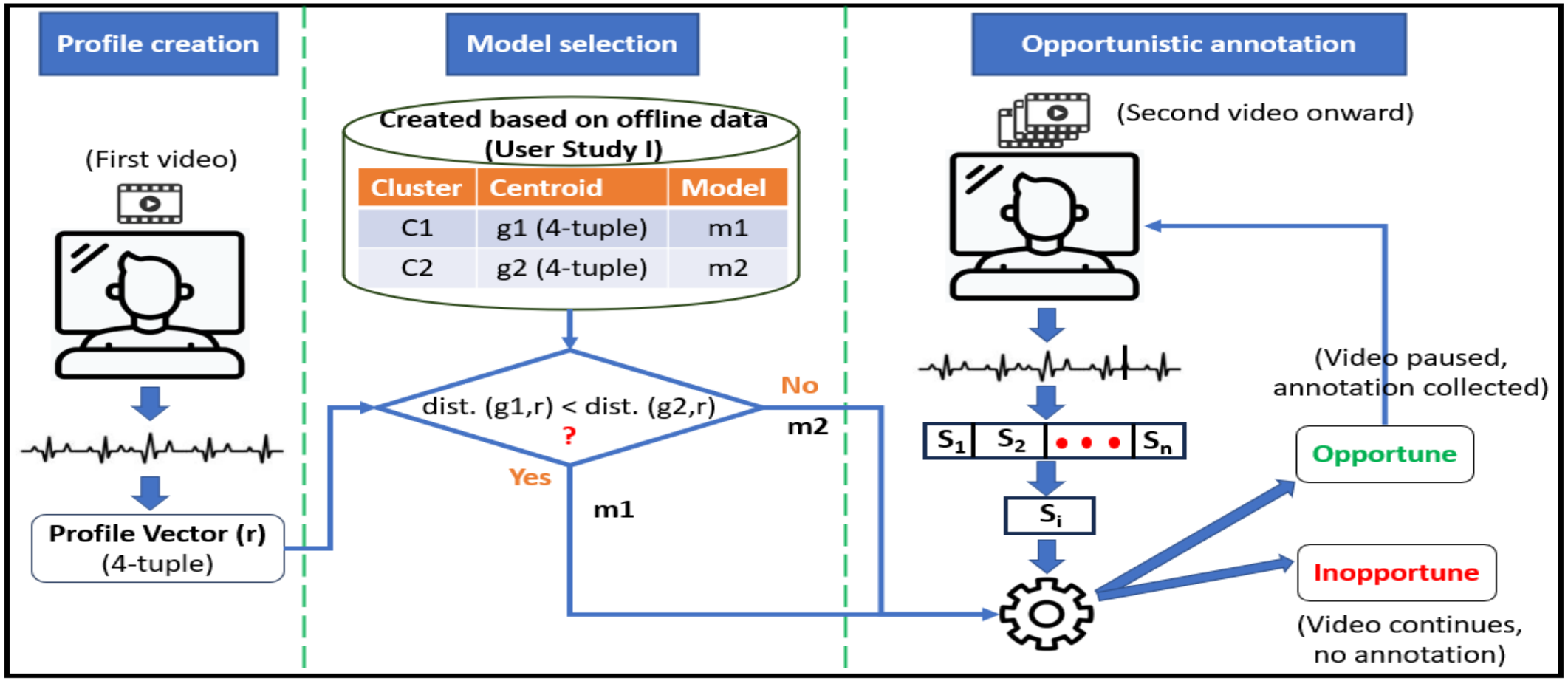
Evaluation

	Probing rate↓	TPR (%) ↑	FPR (%) ↓	LR+ ↑
TBS	29.89 (0.00)	100.00 (0.00)	100.00 (0.00)	1.00
RPS	17.93 (0.01)	60.28 (0.01)	59.98 (0.01)	1.00
RePS	28.79 (0.01)	99.39 (0.02)	95.77 (0.01)	1.04
PPS	21.59 (3.46)	71.53 (0.23)	9.96 (0.04)	7.18
FBS	5.60 (4.58)	53.23 (0.03)	9.53 (0.17)	5.59
APS	7.18 (5.32)	65.27 (0.24)	15.88 (0.16)	4.11
GBPS	6.68 (4.96)	66.84 (0.24)	14.32 (0.14)	4.67
GPS	7.79 (3.11)	76.21 (0.18)	19.72 (0.11)	3.86
RNNPS	6.08 (3.25)	61.16 (0.38)	14.08 (0.12)	4.34
GRUPS	7.65 (3.60)	67.50 (0.36)	18.83 (0.09)	3.58
CNNPS	6.13 (1.75)	68.80 (0.19)	13.13 (0.08)	5.24
PResUP	**5.48 (1.73)**	**80.07 (0.16)**	**9.30 (0.05)**	**8.61**

Deployment of PResUP Framework

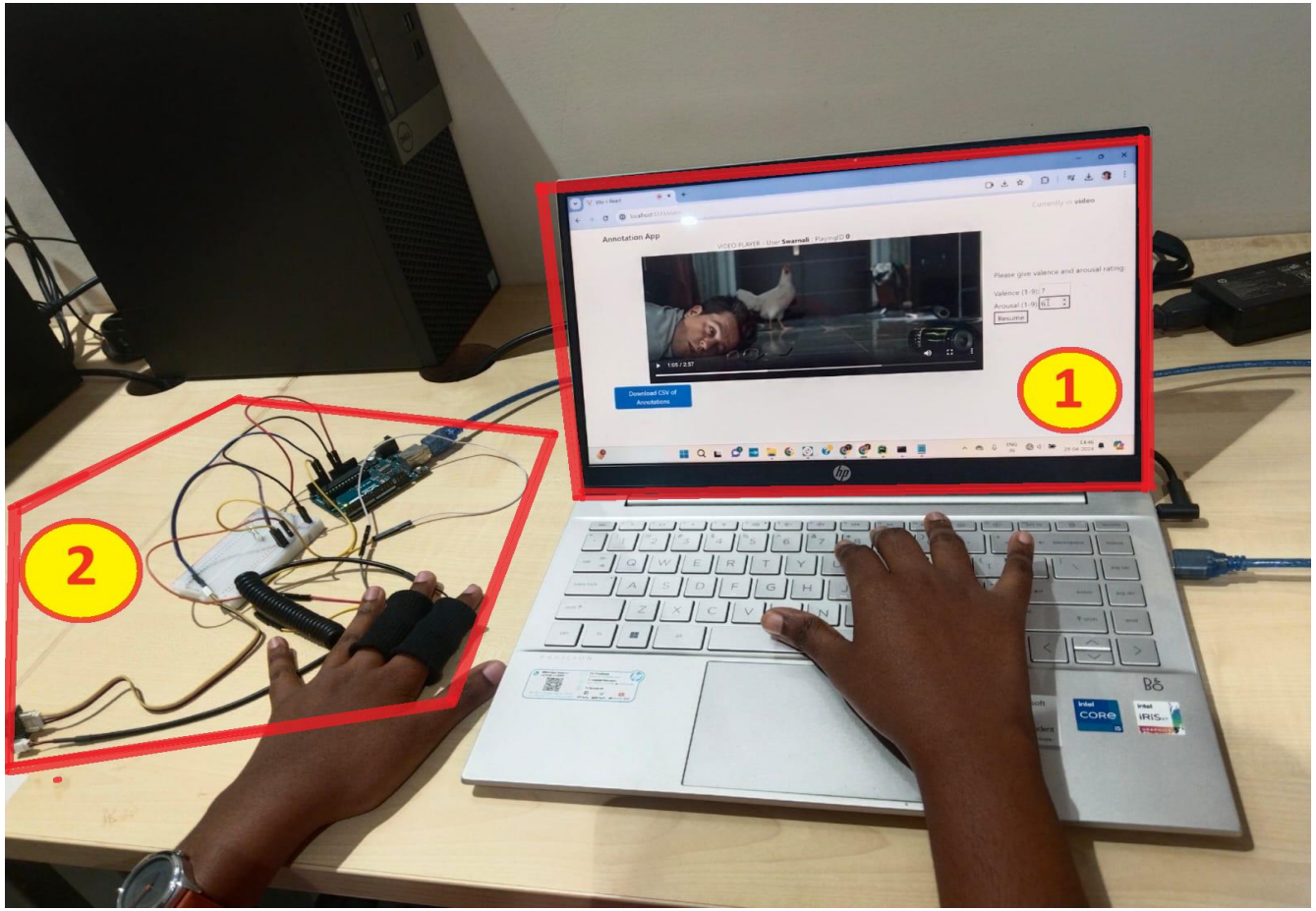


Deployment of PResUP Framework



User Study II

- **Study procedure**
  - 18 participants (13M, 5F).

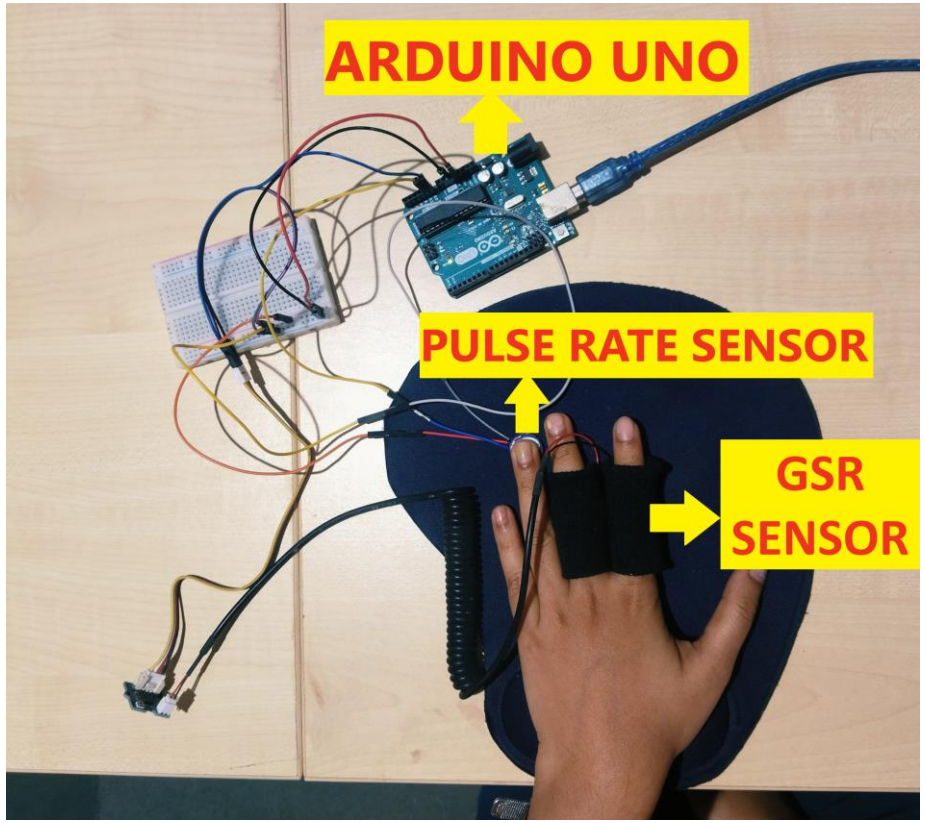


User Study II

- **Experiment apparatus**
  - Annotate app

User Study II

- **Experiment apparatus**
  - Galvanic skin response sensor (GSR V1.2)
  - Pulse rate sensor (HW-827)



Evaluation (after deployment)

	Probing rate↓	TPR (%) ↑	FPR (%) ↓	LR+ ↑
TBS	34.42 (0.00)	100 (0.00)	100 (0.00)	1.00
RPS	23.99 (0.01)	69.79 (0.01)	69.6 (0.01)	1.00
RePS	30.38 (0.10)	98.82 (0.01)	96.25 (0.00)	1.03
PPS	12.16 (1.62)	78.8 (0.18)	9.49 (0.03)	8.30
FBS	5.81 (2.57)	61.62 (0.15)	9.21 (0.02)	6.69
APS	7.77 (5.21)	71.23 (0.20)	12.11 (0.12)	5.88
GBPS	7.65 (4.94)	73.82 (0.20)	11.3 (0.11)	6.53
GPS	8.22 (4.02)	64.77 (0.41)	17.54 (0.18)	3.69
RNNPS	8.09 (3.94)	54.9 (0.36)	21.29 (0.10)	2.58
GRUPS	7.36 (3.77)	51.16 (0.39)	19.7 (0.09)	2.60
CNNPS	9.59 (1.98)	34.82 (0.18)	31.13 (0.06)	1.12
PResUP	**5.79 (2.54)**	**82.26 (0.09)**	**9.01 (0.06)**	**9.13**

Future scope

- Expand physiological signals
- Enhance user experience
- Mobile Platform Applicability

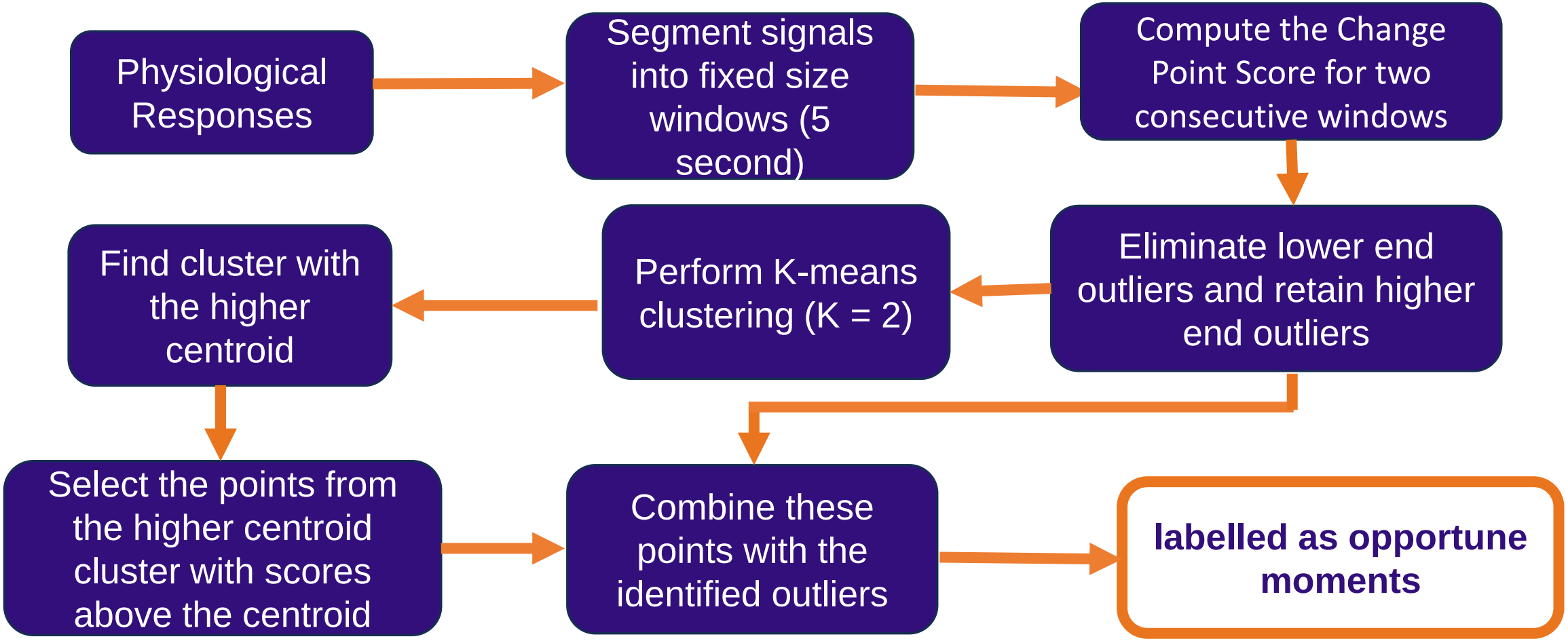
Thank You!

Email: p20210016@goa.bits-pilani.ac.in

Website: https://sites.google.com/goa.bits-pilani.ac.in/swarnalibanik/

BITS Pilani
K K Birla Goa Campus

Dataset: Data Pre-processing

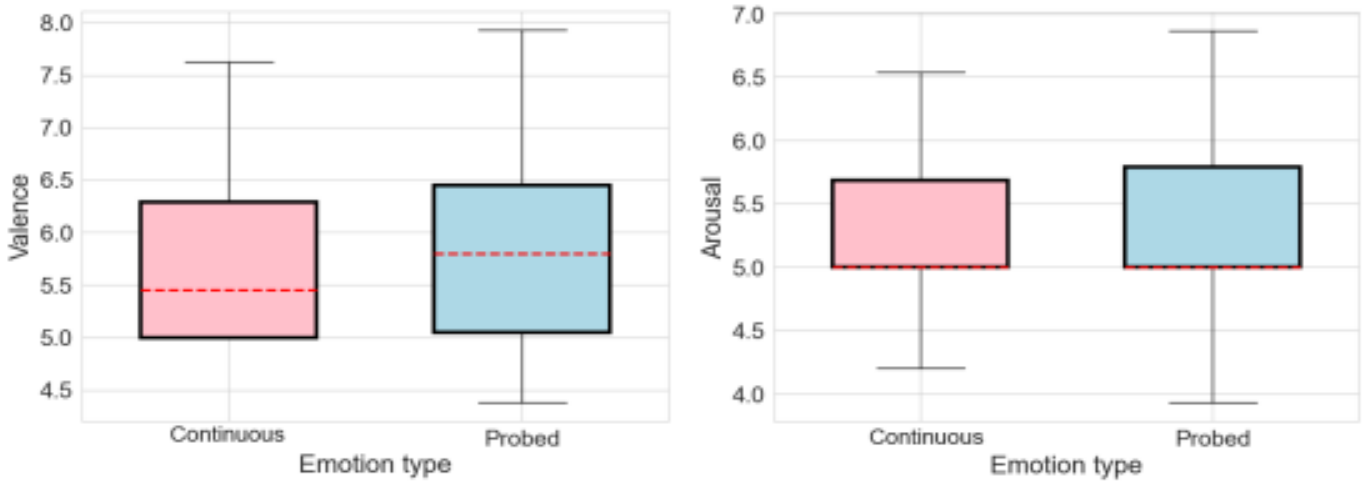


Evaluation Metrices

- Probing Rate
- True Positives Rate (TPR)
- False Positives Rate (FPR)
- Likelihood Ratio (LR+) : $\frac{\text{TPR (Sensitivity)}}{\text{FPR}(1-\text{Specif icity})}$

Evaluation

Actual : Original valence (and arousal) present in continuous annotation
Probed: Sampled valence (and arousal) using PResUP



(a) Comparison of valence scores (b) Comparison of arousal scores

No significant difference between ground truth continuous annotations and sampled values using PResUP framework

Post-study user Survey

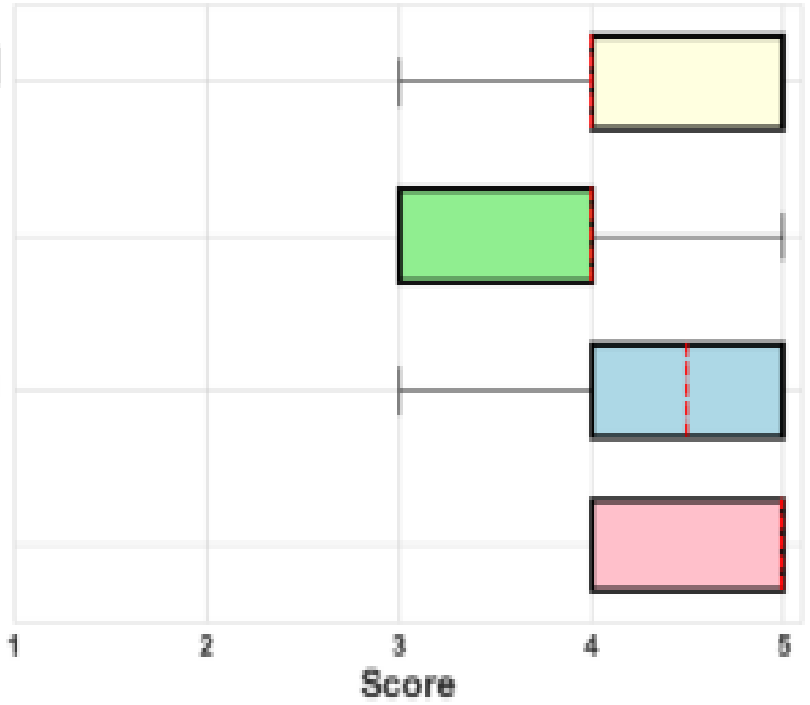
Questions

Q1. The self-reports were issued when the emotional variations occurred
[1: Strongly disagree, 5: Strongly agree]

Q2. Rate the interruption caused by the proposed opportunistic annotation method
[1: Very high, 5: Very low]

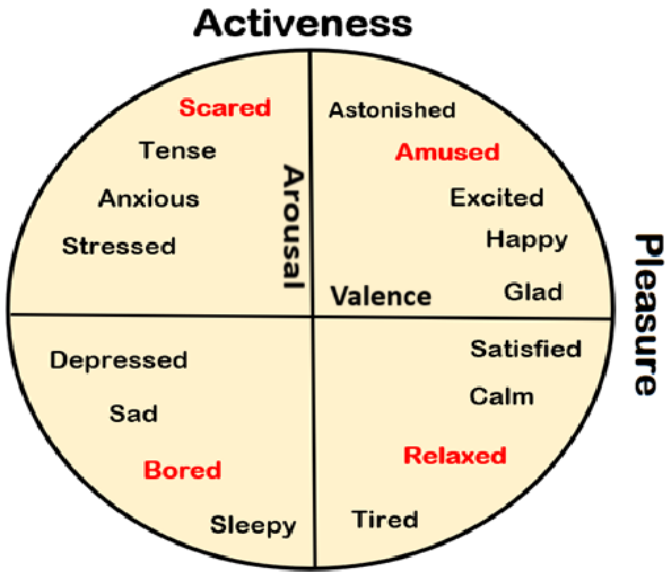
Q3. Rate the ease of use of the proposed opportunistic annotation interface
[1: Very difficult, 5: Very easy]

Q4. Rate your user experiences of the annotation interface
[1: Very dissatisfying, 5: Very satisfying]



Dataset : CASE [1]

- 30 users (15 M, 15F)
- 2D plane - Joystick Input
- Collected emotion self-report annotation



Circumplex Model

Participant watching videos and annotate continuously

[1] - Sharma, K., Castellini, C., van den Broek, E.L. et al. A dataset of continuous affect annotations and physiological signals for emotion analysis. Sci Data 6, 196 (2019). https://doi.org/10.1038/s41597-019-0209-0

Dataset : CASE [1]

- 8 videos in randomized order
- Physiological sensors data collected
  - Electrocardiograph (ECG)
  - Blood Volume Pulse (BVP)
  - **Galvanic Skin Response (GSR)**
  - Respiration (RSP)
  - Skin Temperature (SKT)
  - Electromyography (EMG)

Video id	Emotion	Valence	Arousal	Duration (in sec.)
1	amusing	med/high	med/high	185
2	amusing	med/high	med/high	173
3	boring	low	low	119
4	boring	low	low	160
5	relaxing	med/high	low	145
6	relaxing	med/high	low	147
7	scary	low	high	197
8	scary	low	high	144

The 8 videos present in the CASE dataset

[1] - Sharma, K., Castellini, C., van den Broek, E.L. et al. A dataset of continuous affect annotations and physiological signals for emotion analysis. Sci Data 6, 196 (2019). https://doi.org/10.1038/s41597-019-0209-0

Data Description

- Total 7,290 segments.
- 11.6% are opportune and 88.4% are inopportune.

Evaluation

	Probing rate↓	TPR (%) ↑	FPR (%) ↓	LR+ ↑
TBS	30.38 (0.00)	100.00 (0.00)	100.00 (0.00)	1.00
PPS	20.38 (1.33)	69.21 (32.34)	16.16 (18.72)	4.28
FBS	6.50 (3.02)	70.49 (22.49)	15.02 (8.31)	4.69
GPS	6.37 (4.10)	81.30 (21.24)	14.93 (14.15)	5.45
APS	6.10 (4.04)	71.18 (20.86)	14.35 (15.98)	4.96
GBPS	6.07 (4.07)	82.12 (16.18)	13.68 (13.19)	6.00
RNNPS	6.01 (8.72)	82.66 (11.52)	12.88(5.24)	6.41
GRUPS	6.03 (8.98)	76.14 (10.75)	13.68 (5.37)	5.56
CNNPS	6.52 (6.76)	82.55(10.39)	15.47(6.96)	5.33
PResUP	**5.80 (2.89)**	**86.07 (11.64)**	**12.27 (9.67)**	**7.01**

Performance comparison of PResUP and baselines.

Dataset : K-EmoCon [1]

- 32 Participants (12F, 20M).
- 16 paired debates on accepting Yemeni refugees in Jeju, South Korea.
- Audio-visual recordings captured the debates.
- Continuous emotion annotations were made from three perspectives: subjects, partners, and external observers.

The participants sitting at a table for a debate

[1] C. Y. Park, N. Cha, S. Kang, A. Kim, A. H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, “K-emocon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations,” Scientific Data, vol. 7, no. 1, pp. 1–16, 2020.

Dataset: Data Description

- Total 6644 segments.
- 13.6% are opportune and 86.3% are inopportune.

Evaluation

	Probing rate↓	TPR (%) ↑	FPR (%) ↓	LR+ ↑
TBS	237.28 (0.00)	100 (0.00)	100 (0.00)	1.00
PPS	74.46 (14.87)	81.26 (15.41)	13.01 (1.02)	6.25
FBS	32.32 (16.61)	80.22 (16.19)	4.04 (5.86)	19.86
GPS	33.89 (21.96)	75.65 (36.17)	5.16 (6.58)	14.66
APS	33.82 (16.41)	78.28 (17.51)	3.72 (5.74)	21.04
GBPS	33.78 (17.22)	80.01 (17.16)	3.74 (5.98)	21.39
RNNPS	34.35 (23.93)	68.39 (44.94)	6.56 (6.92)	10.42
GRUPS	40.96 (18.35)	50.51 (43.57)	11.60 (2.94)	4.35
CNNPS	44.5 (16.38)	73.53 (27.01)	11.41 (5.95)	6.44
PResUP	**31.36 (11.23)**	**82.26 (13.38)**	**3.41 (4.41)**	**24.12**

Performance comparison of PResUP and baselines.

Acknowledgement

We gratefully acknowledge the support provided by the following:

- Chanakya Ph.D. Fellowship at AI4ICPS Innovation Hub, IIT Kharagpur.
- SURE grant (SUR/2022/001965) from the Science and Engineering Research Board (SERB), Department of Science & Technology (DST), Government of India.
- CDRF grant (C1/23/152) from BITS Pilani Goa.
- CDRF grant (C1/23/114) from BITS Pilani K K Birla Goa Campus
- SERB CRG-DST, Government of India (CRG/2023/003210)