EEVR: A Dataset of Paired Physiological Signals and Textual Descriptions for Joint Emotion Representation Learning

Pragya Singh, Ritvik Budhiraja, Ankush Gupta, Anshul Goswami,
Mohan Kumar, Pushpendra Singh

Mélange Lab

RIT

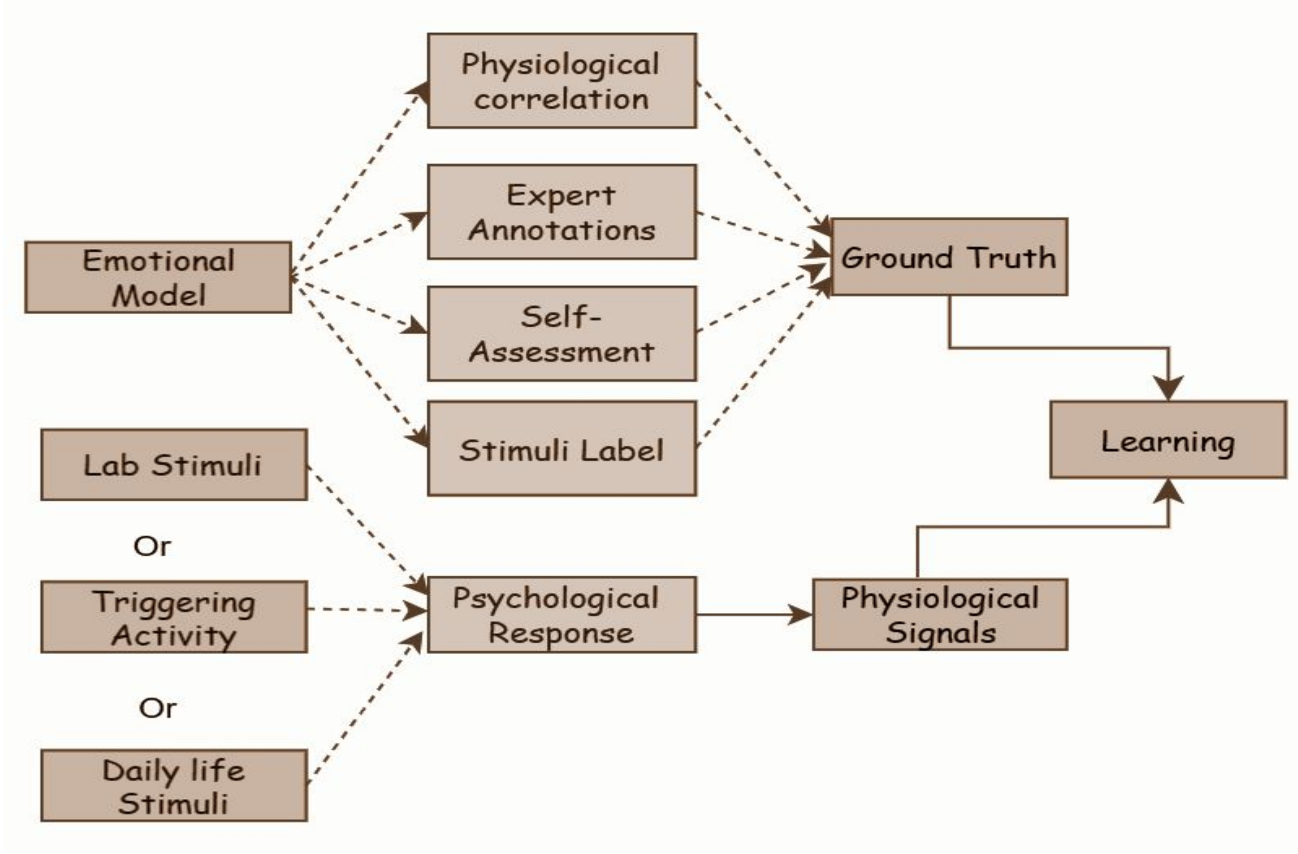
NEURAL INFORMATION PROCESSING SYSTEMS

EEVR: A Dataset of Paired Physiological Signals and Textual Descriptions for Joint Emotion Representation Learning

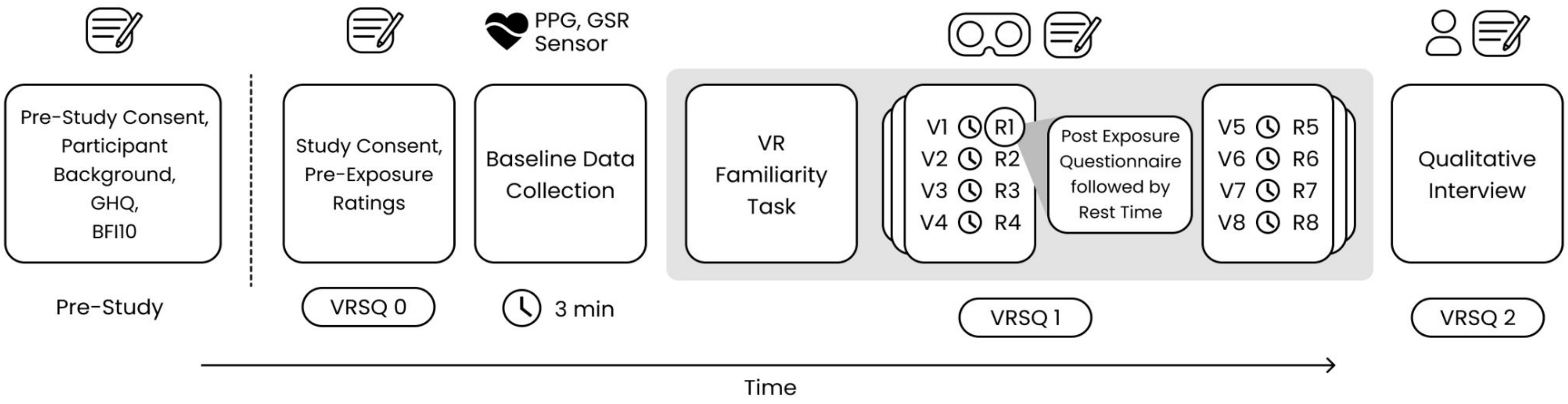
Motivation

- **Emotion Recognition** using wearables have huge potential for mental health monitoring.
- Presently, **objective scales,** stimulus-label, or self-report questionnaires are used for emotion annotation.
- These methods often **fail** to capture **mixed emotions**, **absence** of emotions, or **brief emotional** responses within the stimulus and thus lead to poor emotion representation learning.

Physiological Emotion Data Collection



Methods



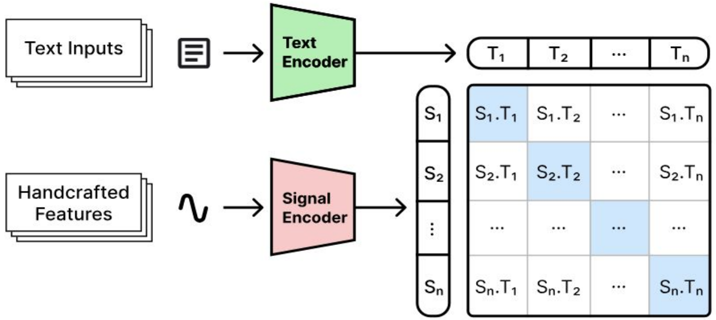
Data Curation

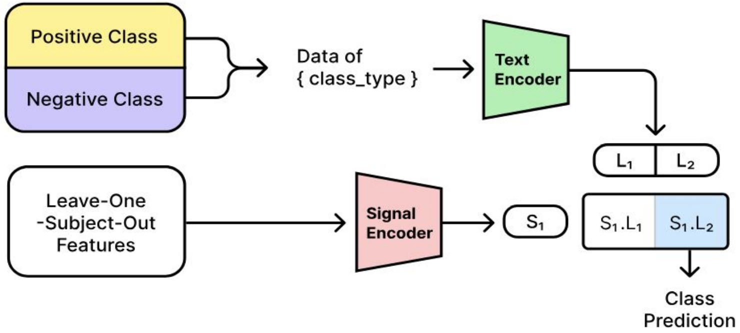
We manually extracted **textual description** data from audio recordings of semi-structured interviews for each **participant-video** pair of physiological data.

Baseline without Text-supervision

Modality	Models	Stimulus-label Accuracy	Stimulus-label F1 Score	Valence Accuracy	Valence F1 Score	Arousal Accuracy	Arousal F1 Score
EDA	**Logistic Regression**	86.78 ± 0	0.82 ± 0	61.56 ± 0	0.71 ± 0	47.41 ± 0	0.36 ± 0
	Decision Tree	85.09 ± 0.17	0.83 ± 0	58.46 ± 1.06	0.64 ± 0.01	54.05 ± 1.12	0.35 ± 0.02
	Random Forest	**90.79 ± 0.46**	**0.89 ± 0.01**	60.26 ± 1.81	0.66 ± 0.01	57.23 ± 1.19	0.28 ± 0.04
	LDA	87.69 ± 0	0.85 ± 0	**61.86 ± 0**	**0.69 ± 0**	48.97 ± 0	0.37 ± 0
	XGBoost	90.69 ± 0.52	0.89 ± 0.01	59.76 ± 0.52	0.66 ± 0.01	56.61 ± 0.34	0.37 ± 0.01
	SVM	85.29 ± 0	0.81 ± 0	59.16 ± 0	0.71 ± 0	51.66 ± 0	**0.44 ± 0**
	MLP	87.39 ± 0	0.85 ± 0	61.86 ± 0	0.68 ± 0	**57.27 ± 0**	0.39 ± 0
PPG	**Logistic Regression**	81.08 ± 0	0.77 ± 0	61.26 ± 0	0.70 ± 0	**56.29 ± 0**	**0.42 ± 0**
	Decision Tree	68.87 ± 0.35	0.65 ± 0	54.35 ± 0.30	0.59 ± 0.01	49.43 ± 0.32	0.26 ± 0.01
	Random Forest	75.88 ± 0.35	0.69 ± 0.01	**61.66 ± 1.93**	**0.70 ± 0**	49.27 ± 0.42	0.18 ± 0.01
	LDA	**81.08 ± 0**	**0.78 ± 0**	58.96 ± 1.73	0.67 ± 0.06	54.47 ± 3.72	0.40 ± 0.02
	XGBoost	49.44 ± 0	0.68 ± 0	57.26 ± 0.76	0.64 ± 0.01	47.89 ± 7.57	0.26 ± 0.13
	SVM	80.48 ± 0	0.75 ± 0	59.86 ± 1.91	0.70 ± 0.05	47.99 ± 3.78	0.32 ± 0.10
	MLP	78.68 ± 0	0.75 ± 0	56.76 ± 0	0.66 ± 0	54.16 ± 0	0.38 ± 0
PPG + EDA	**Logistic Regression**	85.89 ± 0	0.82 ± 0	60.06 ± 0	0.69 ± 0	55.23 ± 0	0.41 ± 0
	Decision Tree	83.78 ± 0.80	0.83 ± 0.01	**62.77 ± 0.30**	0.66 ± 0	**58.13 ± 0.70**	0.40 ± 0.01
	Random Forest	**90.69 ± 0**	**0.89 ± 0**	61.06 ± 1.35	0.70 ± 0.01	56.78 ± 1.56	0.26 ± 0.01
	LDA	84.89 ± 1.39	0.82 ± 0.01	57.56 ± 2.95	0.66 ± 0.06	55.48 ± 1.04	**0.42 ± 0.01**
	XGBoost	87.19 ± 2.73	0.85 ± 0.03	61.36 ± 4.79	0.67 ± 0.04	58.0 ± 1.66	0.36 ± 0.06
	SVM	87.29 ± 1.39	0.84 ± 0.02	62.16 ± 2.08	**0.72 ± 0.02**	55.97 ± 3.44	0.38 ± 0.04
	MLP	83.48 ± 0	0.81 ± 0	58.86 ± 0	0.63 ± 0	56.89 ± 1.47	0.36 ± 0.03
Text	**DistillBert**	**97.44 ± 0.69**	**0.97 ± 0.01**	**91.73 ± 1.73**	**0.88 ± 0.02**	**89.94 ± 1.17**	**0.88 ± 0.02**
	XLMBert-a Base	97.32 ± 0.34	0.97 ± 0	89.46 ± 1.60	0.70 ± 0.15	76.50 ± 9.59	0.70 ± 0.15

Baseline with Text-supervision





- We introduce the **Contrastive Language-Signal Pre-training** (CLSP) method for extracting more contextualized representations.
- The model was trained on physiological signals and text pairs to learn a joint embedding space, where both modalities are closely aligned using a **contrastive loss function**.

Results

Modality	Model	Stimulus-label		Valence		Arousal	
		Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
EDA	**HC+NN**	**87.39**	**0.85**	61.86	0.68	57.27	0.39
PPG	**HC+NN**	78.68	0.75	56.76	0.66	54.16	0.38
EDA+PPG	**HC+NN**	83.48	0.81	58.86	0.63	58.58	0.40
EDA+Text	**CLSP**	64.19	0.68	**70.38**	**0.73**	**77.25**	**0.81**
PPG+Text	**CLSP**	56.95	0.53	64.74	0.64	69.91	0.62
EDA+PPG+Text	**CLSP**	53.50	0.48	64.87	0.60	69.64	0.64

Zero-shot Transfer

Dataset (Signal Type)	Method	Arousal Accuracy	Arousal F1 Score	Valence Accuracy	Valence F1 Score
Emognition (EDA)	MLP	52.80	0.57	**61.89**	0.36
	Zero-shot CLSP	**53.23**	**0.59**	50.32	**0.49**
Emognition (PPG)	MLP	49.94	0.53	50.63	0.28
	Zero-shot CLSP	48.19	0.47	51.88	0.41
Emognition (EDA + PPG)	MLP	51.53	0.54	55.12	0.34
	Zero-shot CLSP	50.94	0.52	53.58	0.41
WESAD (EDA)	MLP	85.00	0.84	96.67	0.97
	Zero-shot CLSP	53.33	0.67	51.67	0.67
WESAD (PPG)	MLP	80.00	0.80	75.00	0.75
	Zero-shot CLSP	70.00	0.68	66.67	0.72
WESAD (EDA + PPG)	MLP	**91.67**	**0.91**	**98.33**	**0.98**
	Zero-shot CLSP	75.00	0.71	86.67	0.86
Nurse (EDA)	MLP	39.88	0.32	71.83	0.03
	Zero-shot CLSP	**55.48**	**0.58**	**84.93**	0.20
Nurse (PPG)	MLP	45.10	0.38	72.08	0.05
	Zero-shot CLSP	53.08	0.48	75.34	0.23
Nurse (EDA + PPG)	MLP	48.35	0.43	76.04	0.23
	Zero-shot CLSP	53.08	0.45	84.59	**0.42**

EEVR: A Dataset of Paired Physiological Signals and Textual Descriptions for Joint Emotion Representation Learning

Takeaways

- Objective annotations based supervised learning algorithms often **fail** to capture the **subtle complexities** of emotion data.

- **Incorporating subjective annotations**, such as textual descriptions, provides a new opportunity to enhance the quality of representations learned from physiological signals.

Thank You :)

Pragya Singh

email: pragyas@iiitd.ac.in

Github: https://github.com/alchemy18/EEVR/

See You All at Our Poster Session!