Self-supervised ECG Representation Learning for Affective Computing

Pritam Sarkar
Master of Applied Science
Department of Electrical and Computer Engineering
Queen's University, Kingston, Canada
Supervisor: Prof. Ali Etemad

Ambient Intelligence and
Interactive Machines (Aiim) Lab

Queen's
UNIVERSITY

Outline

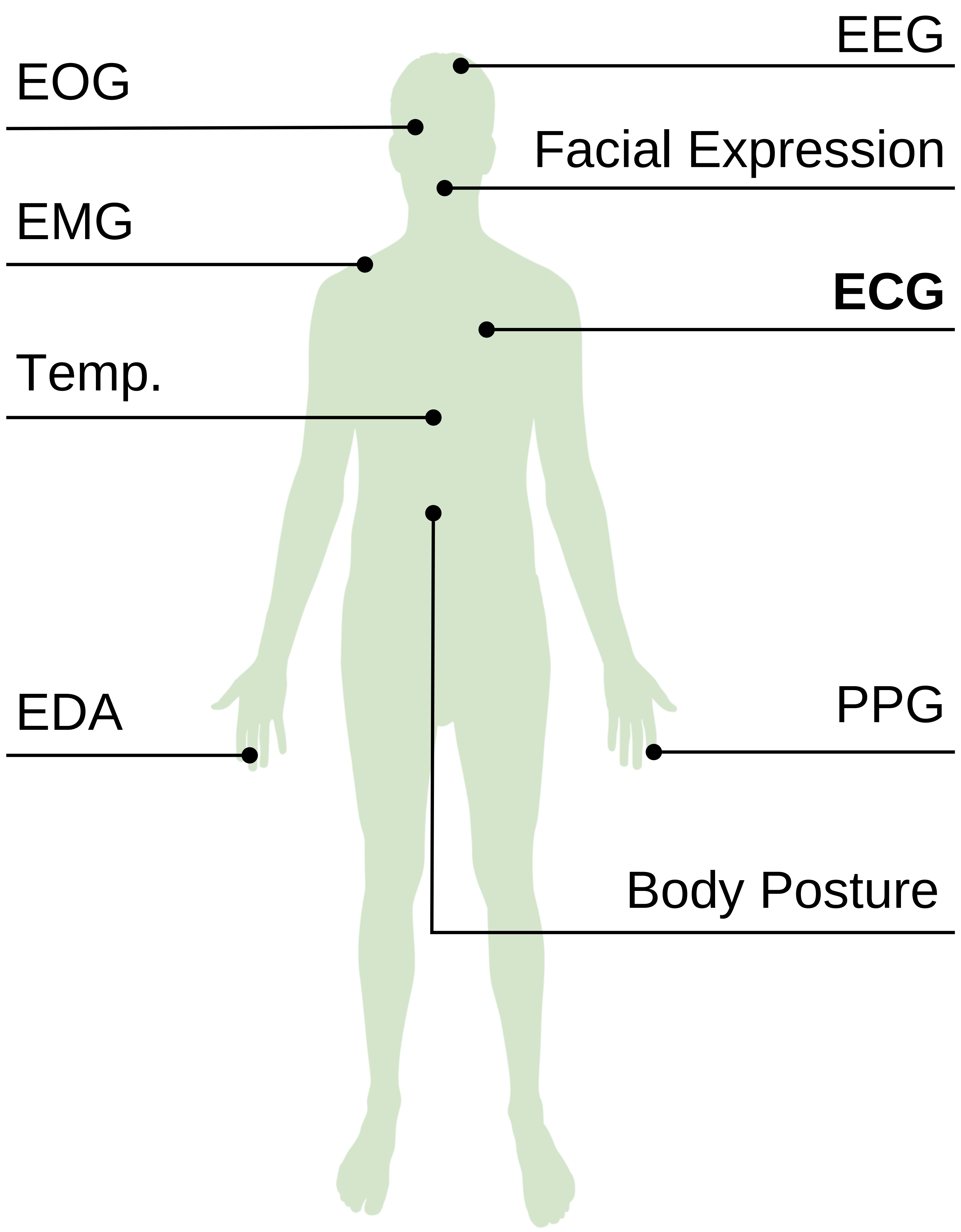
- Affective Computing (Modalities and Applications)
- Problem and Motivation
- Contribution
- Literature Review
- Proposed Framework
- Datasets
- Performance and Analysis
- A Case Study
- Summary

Affective Computing

“I call “affective computing,” computing that relates to, arises from, or influences emotions.”

R. W. Picard, Affective computing, MIT Press, 2000

Modalities and Applications



Modalities

Applications

Image Source:
https://www.clipart.email/download/7343608.html
https://www.deviantart.com/gnomegod98/art/Intense-Gaming-455698094
https://www.allacronyms.com/987276robot.png
https://www.zdnet.com/article/samsung-galaxy-watch-how-to-adjust-settings-and-configure-your-personal-preferences/

Problem and Motivation

Limitations of fully-supervised learning:

- Human annotated labels are required to learn data representations; the learned representations are often very task specific.
- Larger labelled data are required in order to train deep networks; smaller datasets often result in poor performance.

Advantages of self-supervised learning:

- Models are trained using automatically generated labels.
- Learned representations are high-level and generalized; therefore less sensitive to inter or intra instance variations (local transformations).
- Larger datasets can be acquired to train deeper and sophisticated networks.

Problem and Motivation

Limitations of fully-supervised learning:

- ❑ Human annotated labels are required to learn data representations; the learned representations are often very task specific.
- ❑ Larger labelled data are required in order to train deep networks; smaller datasets often result in poor performance.

Advantages of self-supervised learning:

- ❑ Models are trained using automatically generated labels.
- ❑ Learned representations are high-level and generalized; therefore less sensitive to inter or intra instance variations (local transformations).
- ❑ Larger datasets can be acquired to train deeper and sophisticated networks.

Contribution

- We propose a self-supervised framework for emotion recognition based on multi-task ECG representation learning for ***the first time*** and achieve ***state-of-the-art results*** in four public datasets.

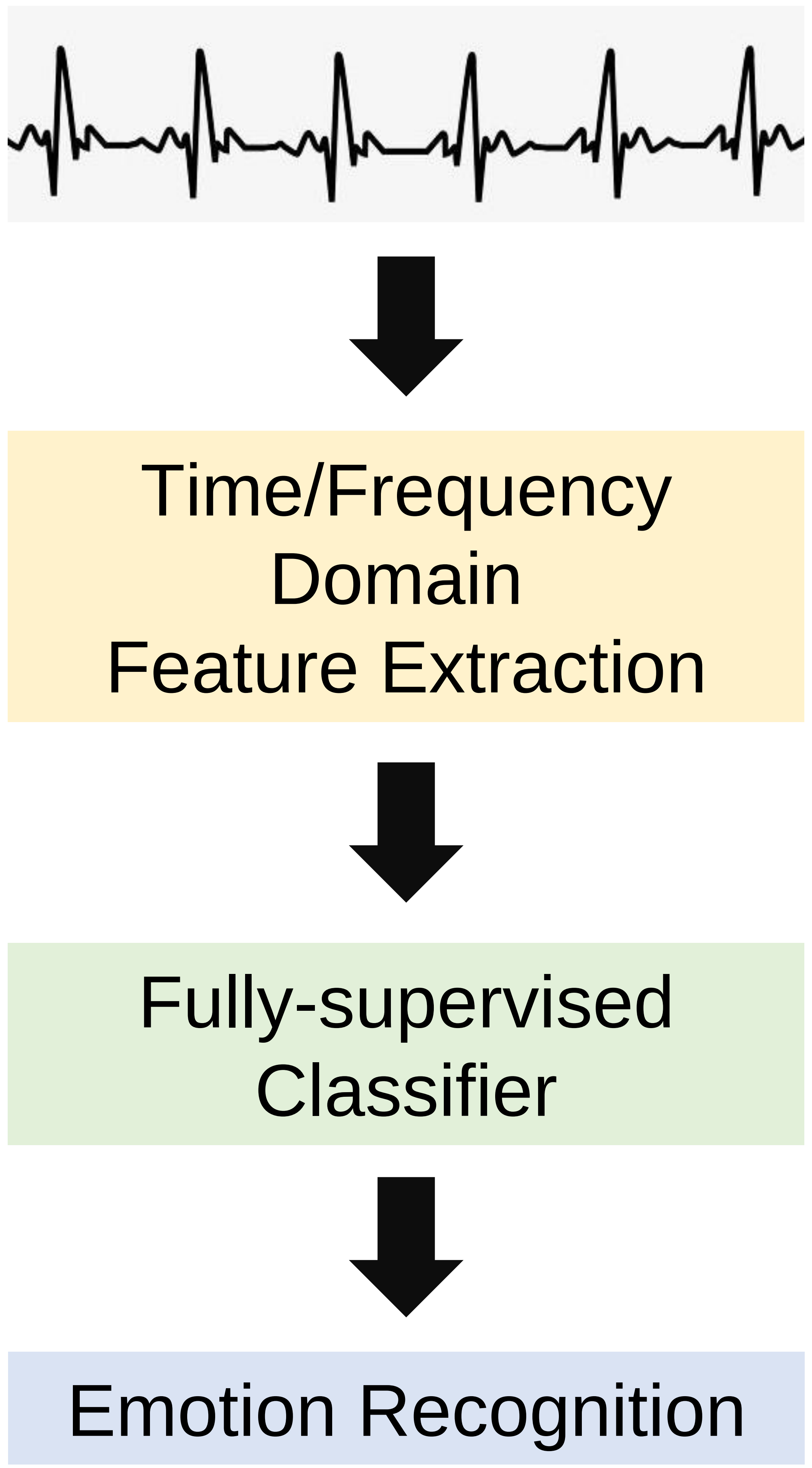
  - P. Sarkar and A. Etemad, "Self-supervised learning for ECG-based emotion recognition", *IEEE 45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2020.
  - P. Sarkar and A. Etemad, "Self-supervised ECG representation learning for emotion recognition'', under review in *IEEE Trans. Affective Computing.*

- As a case study, we propose a ***novel end-to-end framework*** for ***adaptive simulation*** for training trauma responders, capable of dynamically adapting to the cognitive load and the level of expertise of individuals.

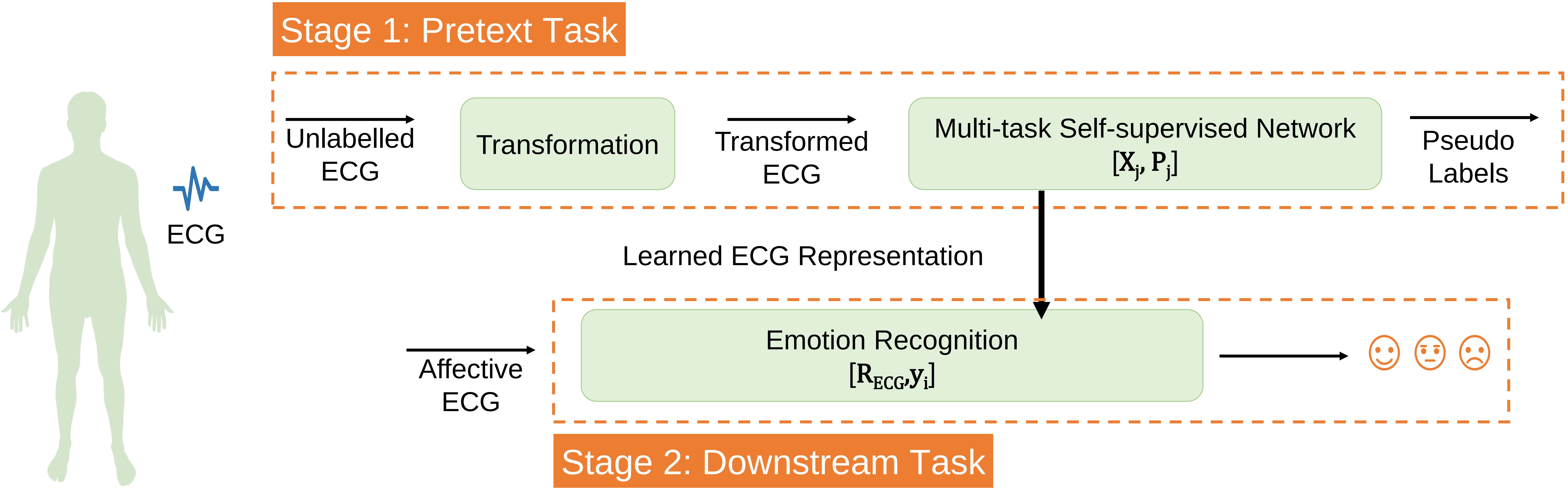
  - P. Sarkar, K. Ross, et al., "Classification of cognitive load and expertise for adaptive simulation using deep multitask learning," *IEEE 8th International Conference on Affective Computing and Intelligent Interaction (ACII),* 2019.
  - K. Ross, P. Sarkar, et al., "Toward dynamically adaptive simulation: Multimodal classification of user expertise using wearable devices", *J. Sensors,* 2019.

Literature Review

- *Healey et al., 2005*:
  - Stress detection during driving task
  - Time/frequency domain features
  - LDA classifier
- *Liu et al., 2009*:
  - Affect based gaming experience
  - Time/frequency domain features
  - RF, KNN, BN, SVM classifiers
- *Santamaria et al., 2018:*
  - Movie clips were used to elicit emotional state
  - Time/frequency domain features
  - Deep CNN classifier
- *Siddharth et al., 2019:*
  - Affect recognition
  - HRV and spectrogram features
  - Extreme learning machine classifier



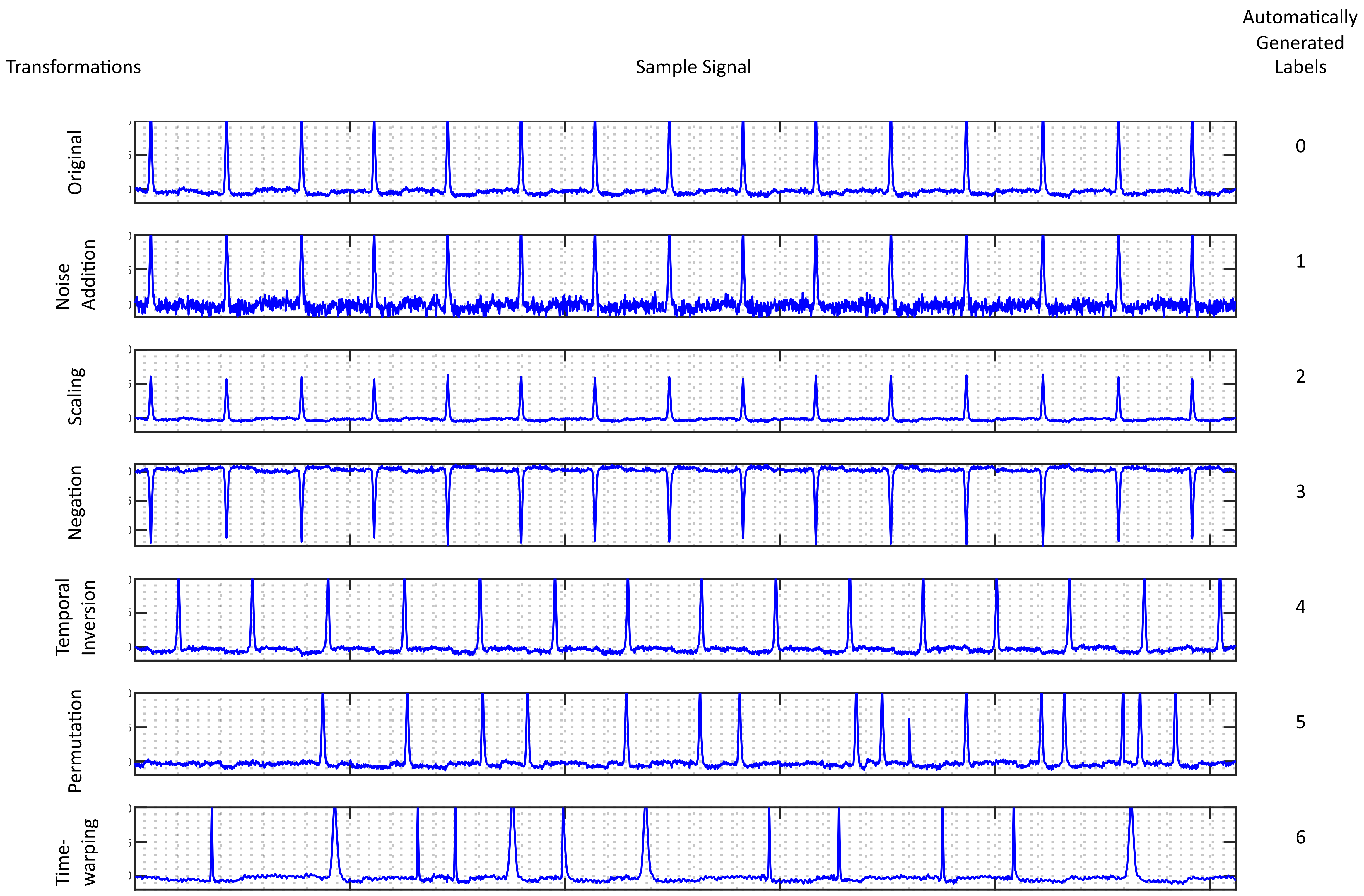
Proposed Framework



Our proposed framework.

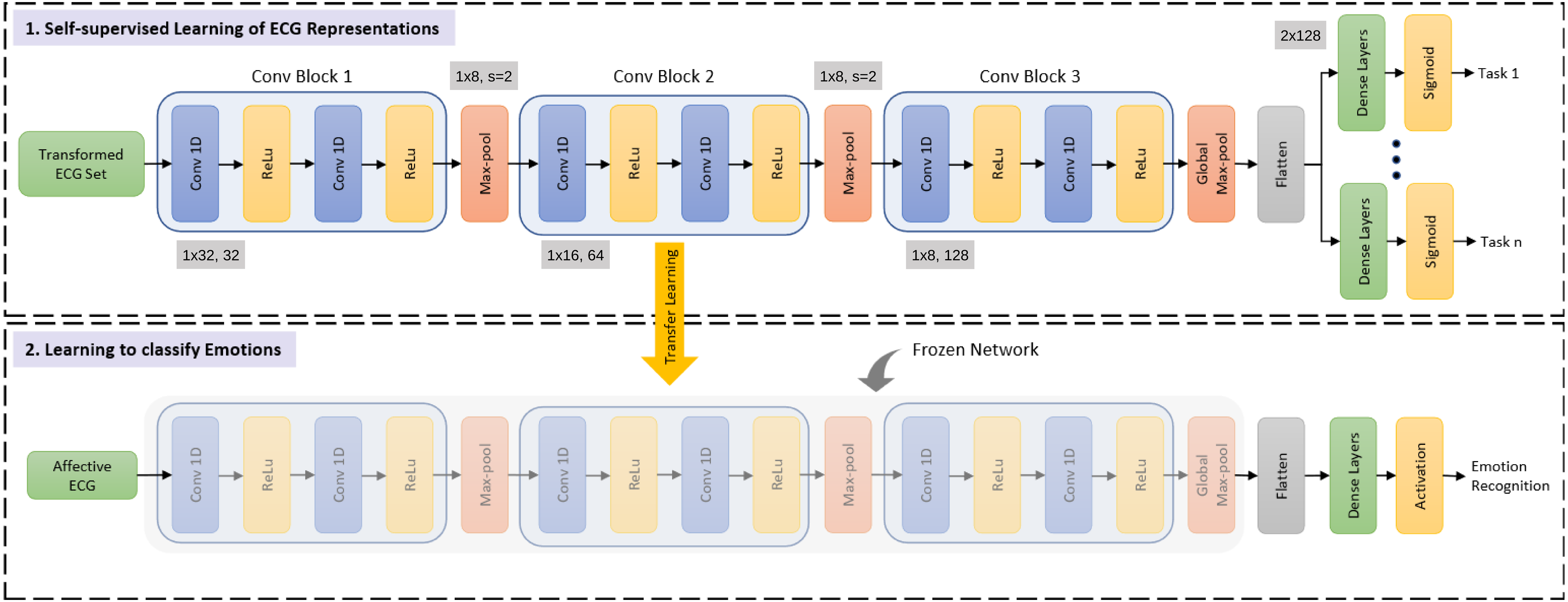
Transformations

- ❑ Noise Addition [SNR = 15]
- ❑ Scaling [scaling factor = 0.9]
- ❑ Negation
- ❑ Temporal Inversion
- ❑ Permutation [no. of segments = 20]
- ❑ Time-warping [no. of segments=9, stretching factor = 1.05]



A sample of an original ECG signal with the six transformed signals along with automatically generated labels are presented.

Proposed Architecture



Our proposed architecture.

Datasets

The summary of the four datasets used are presented.

Dataset	Participants	Attributes	Classes
AMIGOS	40	Arousal	9
		Valence	9
DREAMER	23	Arousal	5
		Valence	5
WESAD	17	Affect State	4
SWELL	25	Stress	3
		Arousal	9
		Valence	9

Transformation Recognition Results

Signal transformation recognition across the four datasets are presented.

Transformation	**All datasets combined**	
	Acc.	F1
Original	0.980 ± 0.003	0.927 ± 0.007
Noise Addition	0.995 ± 0.000	0.979 ± 0.003
Scaling	0.982 ± 0.003	0.932 ± 0.010
Temporal Inversion	0.998 ± 0.000	0.992 ± 0.004
Negation	0.998 ± 0.000	0.990 ± 0.000
Permutation	0.998 ± 0.000	0.989 ± 0.003
Time-warping	0.997 ± 0.003	0.992 ± 0.006
Average	0.992 ± 0.001	0.972 ± 0.005

Emotion Recognition Results

Multi-class emotion recognition results are presented for each of the four datasets.

Dataset	Attribute	Classes	Acc.	F1
AMIGOS	Arousal	9	0.796	0.777
	Valence	9	0.783	0.765
DREAMER	Arousal	5	0.771	0.740
	Valence	5	0.749	0.747
WESAD	Affect State	4	0.950	0.940
SWELL	Arousal	9	0.926	0.930
	Valence	9	0.938	0.943
	Stress	3	0.902	0.900

Comparison

The results of our self-supervised method on all the datasets are presented and compared with prior works including the state-of-the-art, as well as a fully-supervised CNN as a baseline.

A: AMIGOS

Ref.	Method	Arousal		Valence	
		Acc.	F1	Acc.	F1
[5]	GNB	–	0.545	–	0.551
[29]	CNN	0.81	0.76	0.71	0.68
Ours	Fully-Supervised CNN	0.844	0.835	0.811	0.809
	Self-Supervised CNN	**0.889**	**0.884**	**0.875**	**0.874**

B: DREAMER

Ref.	Method	Arousal		Valence	
		Acc.	F1	Acc.	F1
[23]	SVM	0.624	0.580	0.624	0.531
Ours	Fully-Supervised CNN	0.707	0.708	0.666	0.658
	Self-Supervised CNN	**0.859**	**0.859**	**0.850**	**0.845**

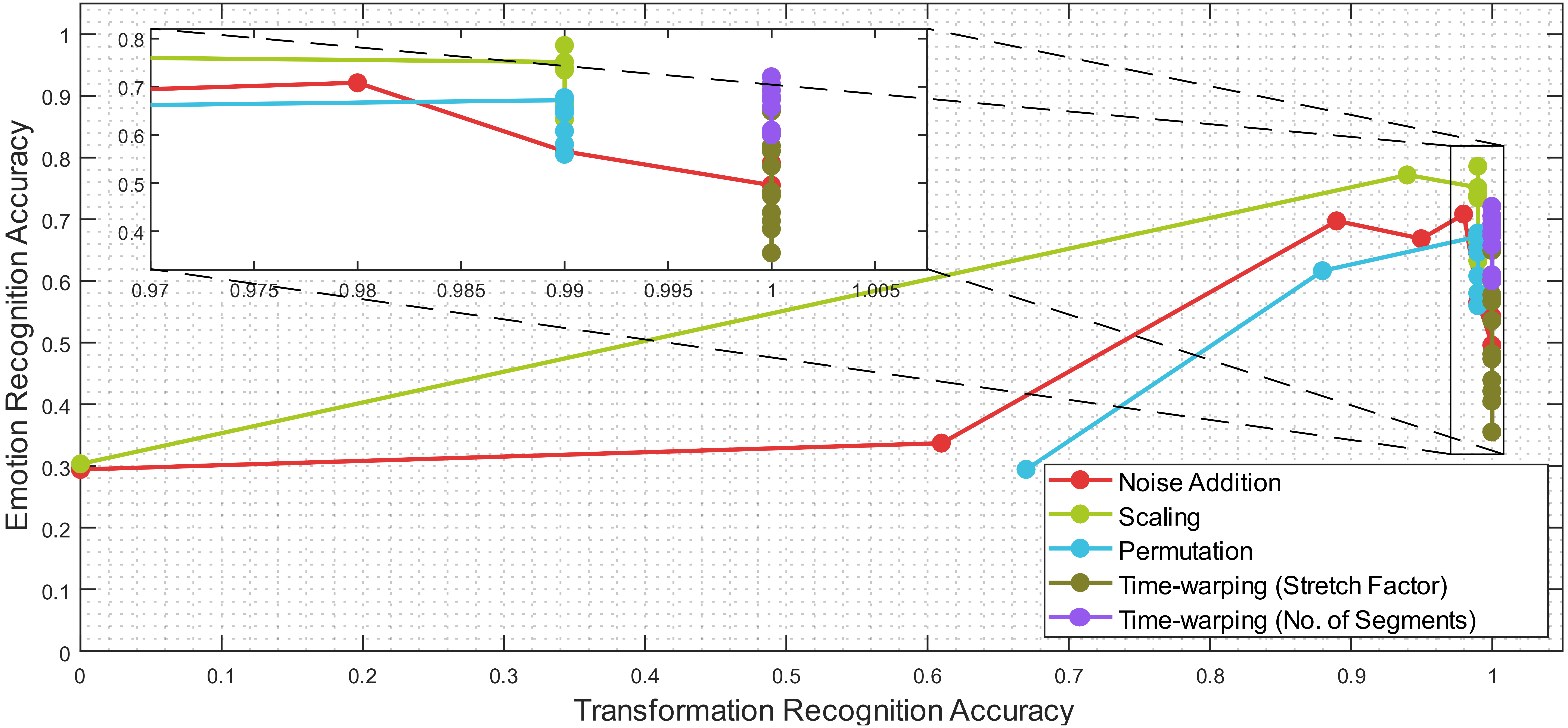
C: WESAD

Ref.	Method	Affect State	
		Acc.	F1
[24]	*k*NN	0.548	0.478
	DT	0.578	0.517
	RF	0.604	0.522
	AB	0.617	0.525
	LDA	0.663	0.560
[31]	CNN	0.83	0.81
Ours	Fully-Supervised CNN	0.932	0.912
	Self-Supervised CNN	**0.969**	**0.963**

D: SWELL

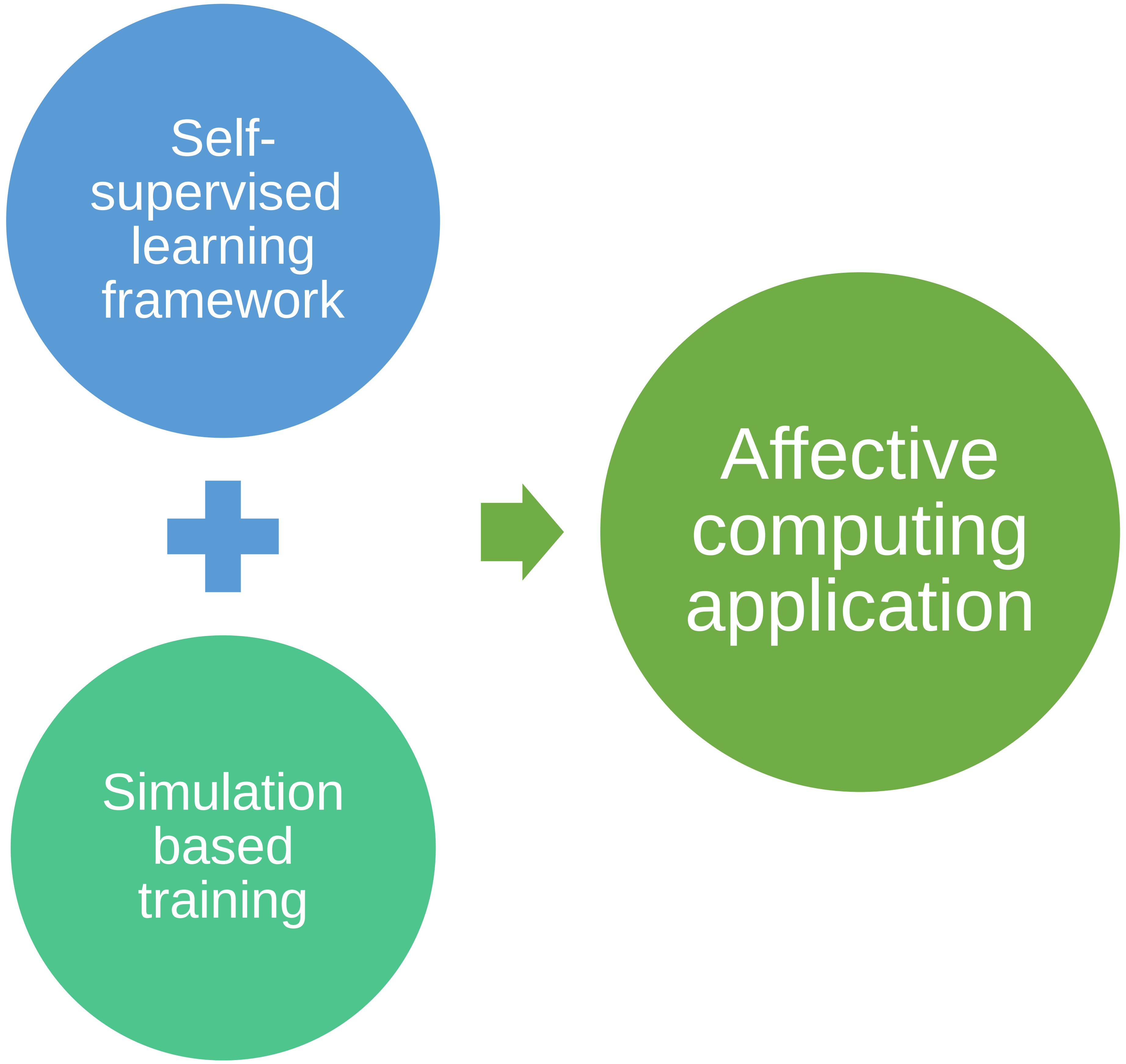
Ref.	Method	Stress		Arousal		Valence	
		Acc.	F1	Acc.	F1	Acc.	F1
[32]	*k*NN	0.769	–	–	–	–	–
	SVM	0.864	–	–	–	–	–
Our	Fully-Supervised CNN	0.894	0.874	0.956	0.962	0.961	0.956
	Self-Supervised CNN	**0.933**	**0.924**	**0.967**	**0.964**	**0.973**	**0.969**

Relationship Between Pretext Task and Downstream Task



The relationship between emotion recognition accuracy and transformation recognition is presented.

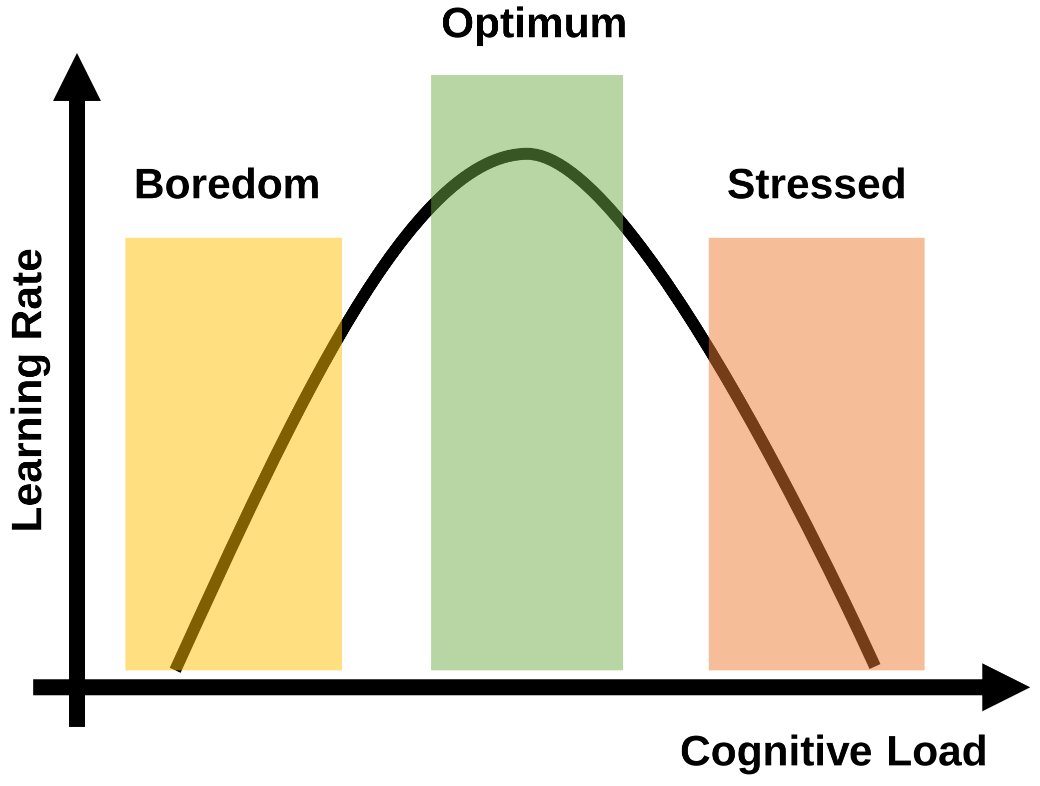
A Case Study



Self-supervised learning framework
Simulation based training
Affective computing application

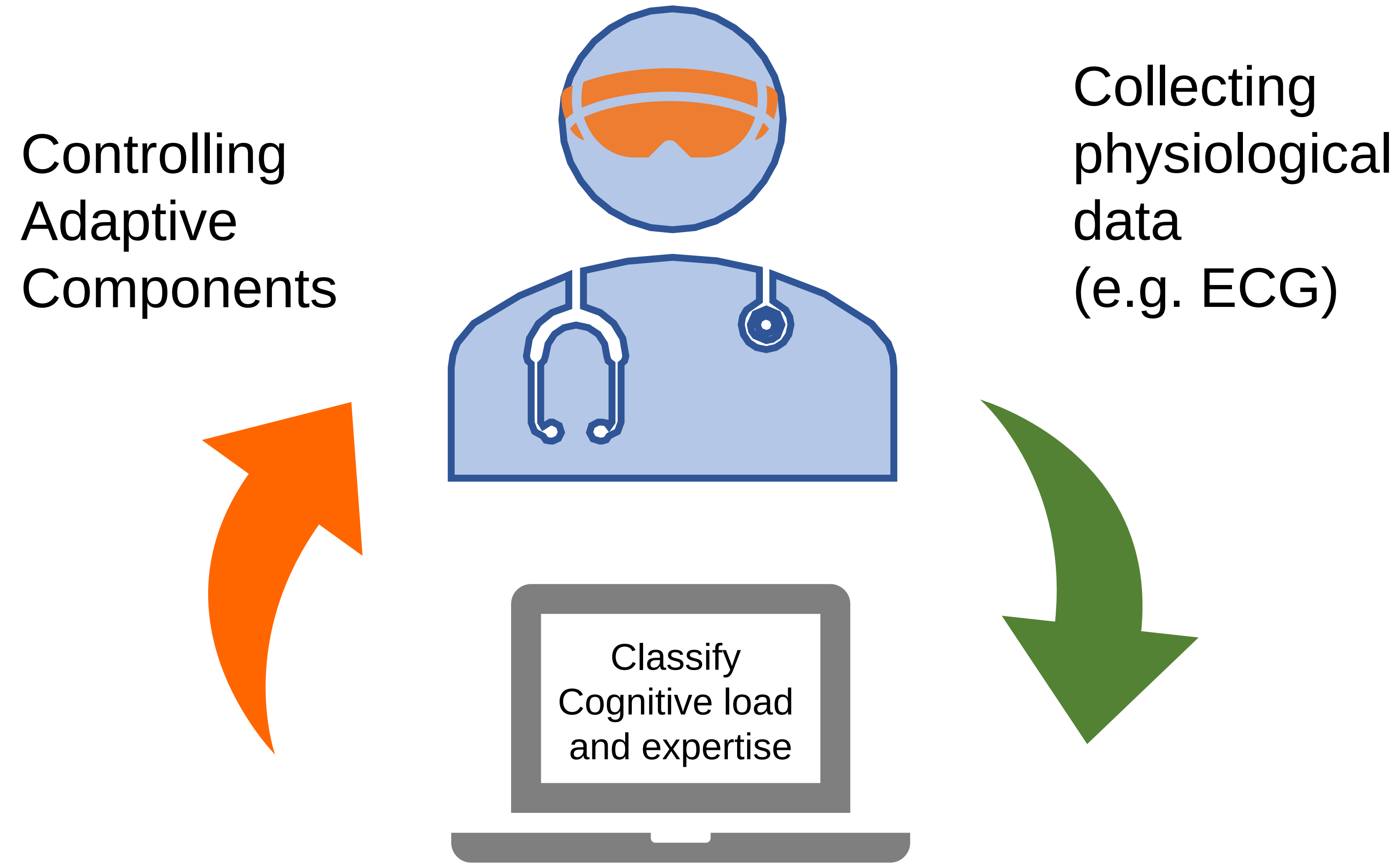
Adaptive Simulation

What we expect:



What we have:

We propose:



End-to-end framework for the development of adaptive simulation that actively classifies a participant's level of cognitive load and expertise.

Experiment Setup

Participants during simulation, this picture was taken from the control room.

Distractors were introduced to give superfluous information during simulation.

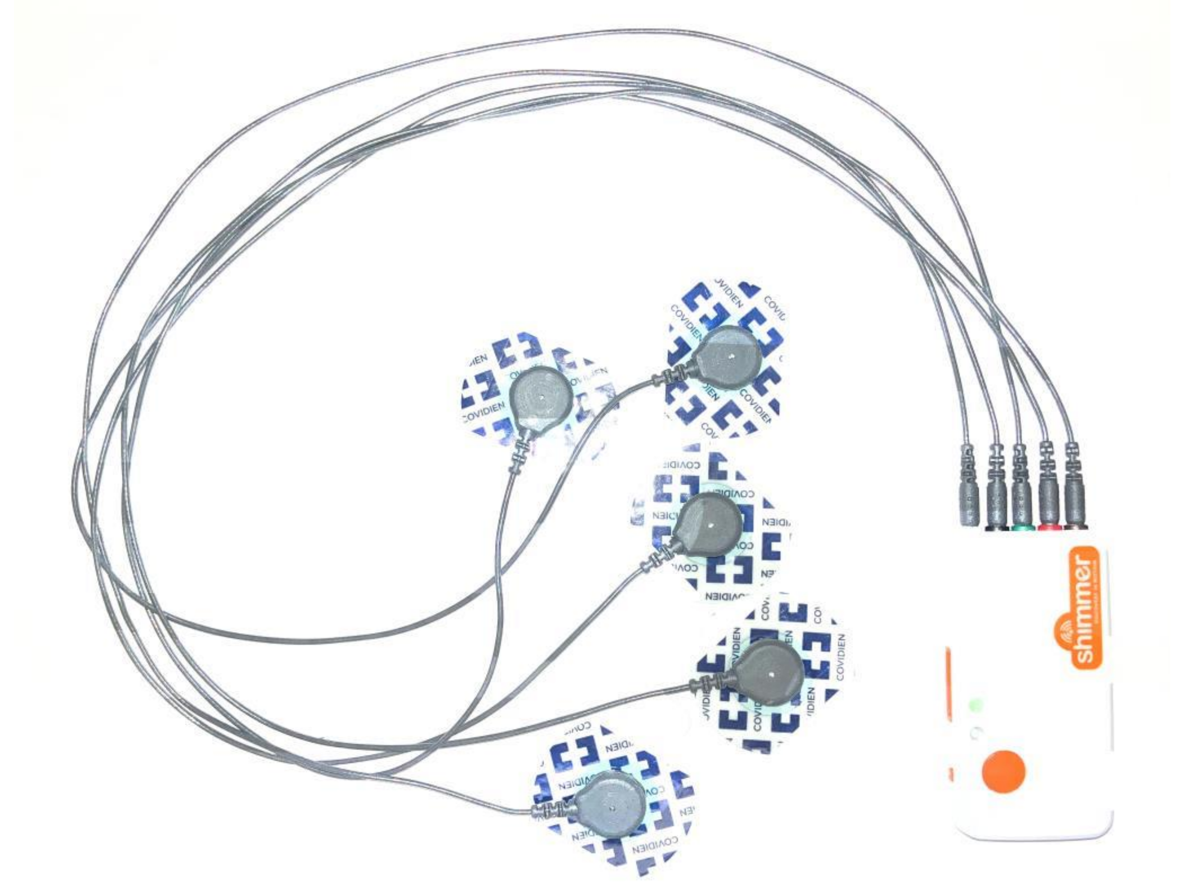
AR object to control severity of patient’s respiratory problem.

CLEAS – Data collection protocol

Cognitive Load and Expertise for Adaptive Simulation

Signing of consent form	Filling demographic information	Attaching sensor and HoloLens	Baseline data (2 mins)	Scenario explanation	Simulation 1 (10 mins)	Scenario explanation	Simulation 2 (10 mins)	Removing sensor and HoloLens	Debrief session

CLEAS Dataset	
Attributes	ECG, Cognitive load, Expertise
Total participants	9
Expert (Physicians) Novice (4th year students)	5 4

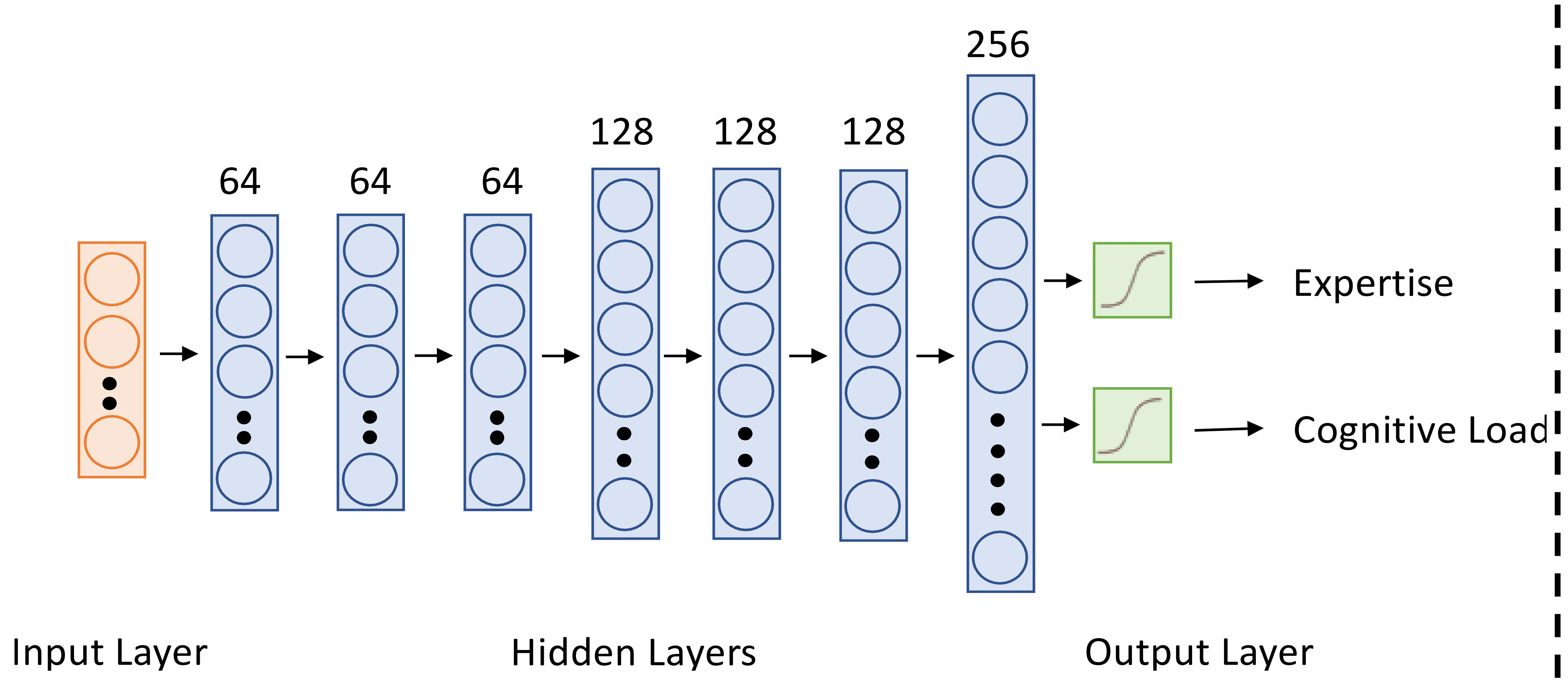


Shimmer Sensors to collect ECG Signal

Method: Fully-supervised

Steps:

- ❑ Segmented into 10 seconds window with 50% overlap.
- ❑ Used Pan Tompkins algorithm for R peaks detection.
- ❑ Time and Frequency domain features were extracted.
- ❑ Features were normalized using baseline data.
- ❑ Utilised a deep multi-task neural network for the classification of expertise and cognitive load.

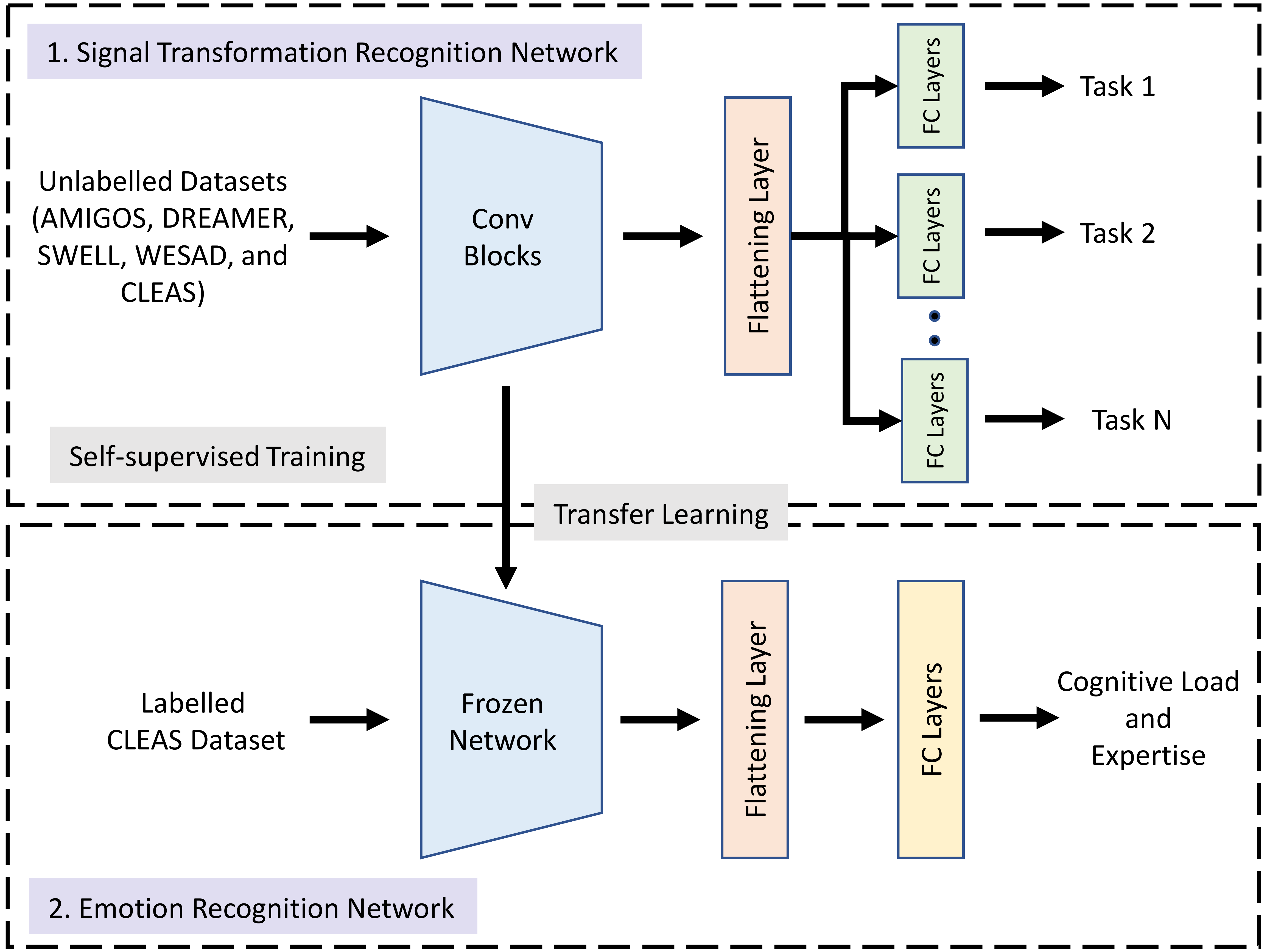


Our proposed Deep Multi-task Neural Network.

Method: Self-supervised

Steps:

- Combined CLEAS dataset with AMIGOS, DREAMER, SWELL and WESAD to perform self-supervised learning.
- Obtained the learned ECG representation from self-supervised network and utilized for classification of cognitive load and expertise.



CLEAS: Fully-supervised Learning Results

Comparison of our proposed Deep Multitask Neural Network (DMNN) with previous approaches and baseline.

Ref.	Task	Attribute	Signals	Method	Acc.
[29]	Mental Task	Cog. Load	ECG, EMG, GSR, Temp	kNN	50.4%
				NB	56.3%
				RF	57.8%
[26]	Computer Game	Anxiety	ECG, GSR, Temp	kNN	80.4%
				BN	80.6%
				RT	80.4%
				SVM	88.9%
[23]	Driving task	Stress	ECG, EMG, GSR	LDA	97.3%
[95]	Arithmetic Task	Stress	GSR	SVM	81.3%
				LDA	82.8%
Ours	Training Simulation	Expertise	ECG	SVM	89.9%
		Cog. Load			75.1%
		Expertise		DMNN	96.6%
		Cog. Load			89.4%

CLEAS: Self-supervised Learning Results

Transformation Recognition

Transformation	Acc.	F1
Original	0.962 ± 0.004	0.866 ± 0.013
Noise Addition	0.992 ± 0.001	0.971 ± 0.007
Scaling	0.963 ± 0.004	0.865 ± 0.021
Temporal Inversion	0.998 ± 0.000	0.992 ± 0.000
Negation	0.996 ± 0.000	0.987 ± 0.002
Permutation	0.995 ± 0.000	0.983 ± 0.002
Time-warping	0.995 ± 0.001	0.983 ± 0.005
Average	0.986 ± 0.002	0.950 ± 0.007

Emotion Recognition

Ref.	Method	Expertise		Cognitive Load	
		Acc.	**F1**	**Acc.**	**F1**
Ours	Fully-Supervised CNN	0.882	0.937	0.886	0.899
	Self-Supervised CNN	**0.954**	**0.954**	**0.961**	**0.961**

Summary

- We proposed a novel ECG-based self-supervised learning framework for affective computing for the first time.
- We achieved state-of-the-art results on 4 public datasets (AMIGOS, DREAMER, WESAD, SWELL).
- We presented insightful and in-depth analysis of our proposed self-supervised framework.
- We proposed a novel end-to-end framework for an adaptive simulation for training trauma responders for the first time.

Thank you!