EXTRACTING MULTIMODAL EMBEDDINGS VIA SUPERVISED CONTRASTIVE LEARNING FOR PSYCHOLOGICAL SCREENING

Manasa Kalanadhabhatta, Adrelys Mateo Santana, Deepak Ganesan, Tauhidur Rahman, Adam Grabell

University of Massachusetts Amherst

ACII 2022

Emotion regulation disorders are prevalent in early childhood.

Screening is challenging and time-consuming.

Potential to collect diagnostic information from wearables and video-based systems.

Combining multimodal information for accurate prediction of disorders.

SETUP: PREDICT CLINICAL VS NON-CLINICAL RISK

Dataset: 76 preschool-aged children playing a frustration-inducing game (up to 30 positive or negative feedback trials).

Modality	# Trials with Handcrafted Features	AUROC
GSR: Galvanic Skin Response	764	0.78
Video: Facial Action Units (AUs)	1483	0.64
functional Near Infrared Spectroscopy (fNIRS): Prefrontal cortex activation	1498	0.74
GSR + Video + fNIRS	696	0.62

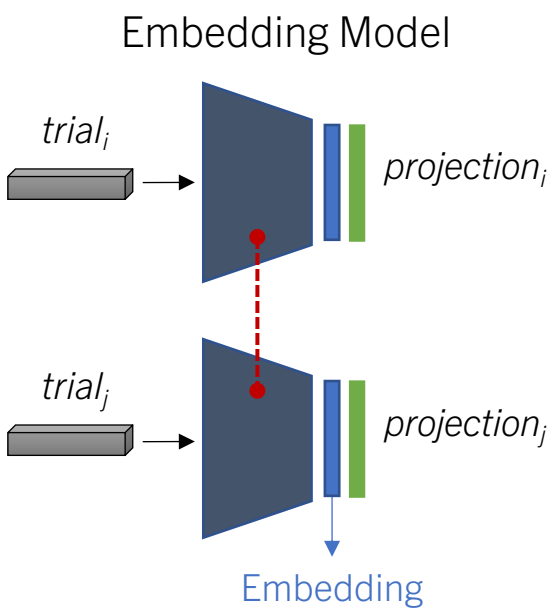
CHALLENGES IN USING HANDCRAFTED FEATURES

- Multimodal model performs worse – ineffective combination of information.
- Fewer trials with valid features – inefficient use of data.

MULTITASK SUPERVISED CONTRASTIVE LEARNING

Extract informative features from raw multimodal signals in a data-efficient manner.

STAGE I: Train Embedding Model by minimizing Multitask SupCon Loss



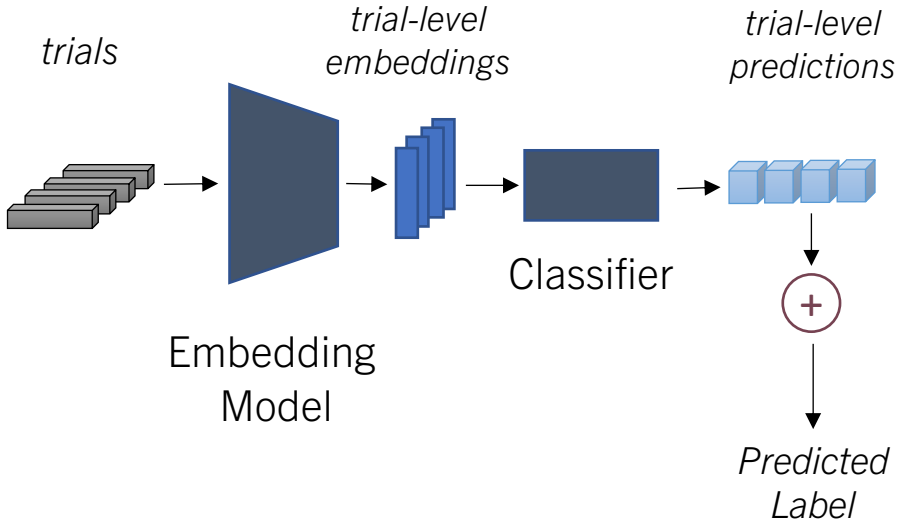
Minimize

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \lambda\, \mathcal{L}_{\text{aux}}$$

$\mathcal{L}_{\text{main}}$: Clinical vs non-clinical contrastive loss.

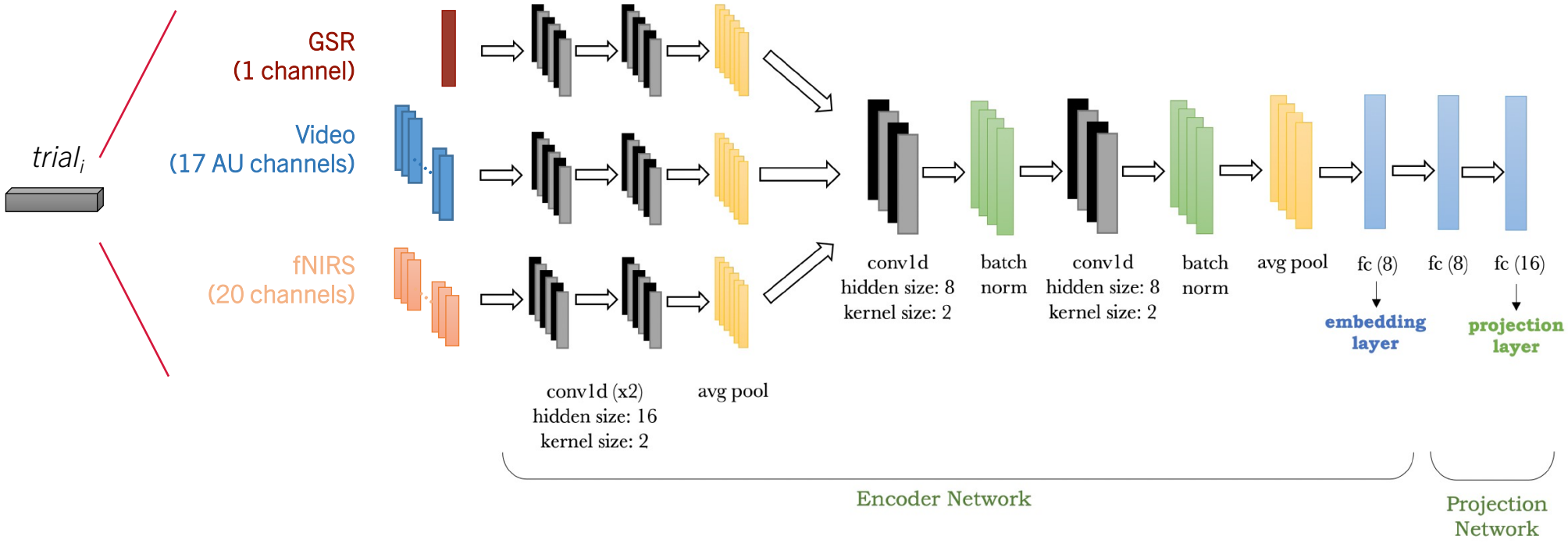
$\mathcal{L}_{\text{aux}}$: Positive vs negative feedback contrastive loss.

STAGE II: Freeze Embedding Model and Train Classifier



EMBEDDING MODEL ARCHITECTURE

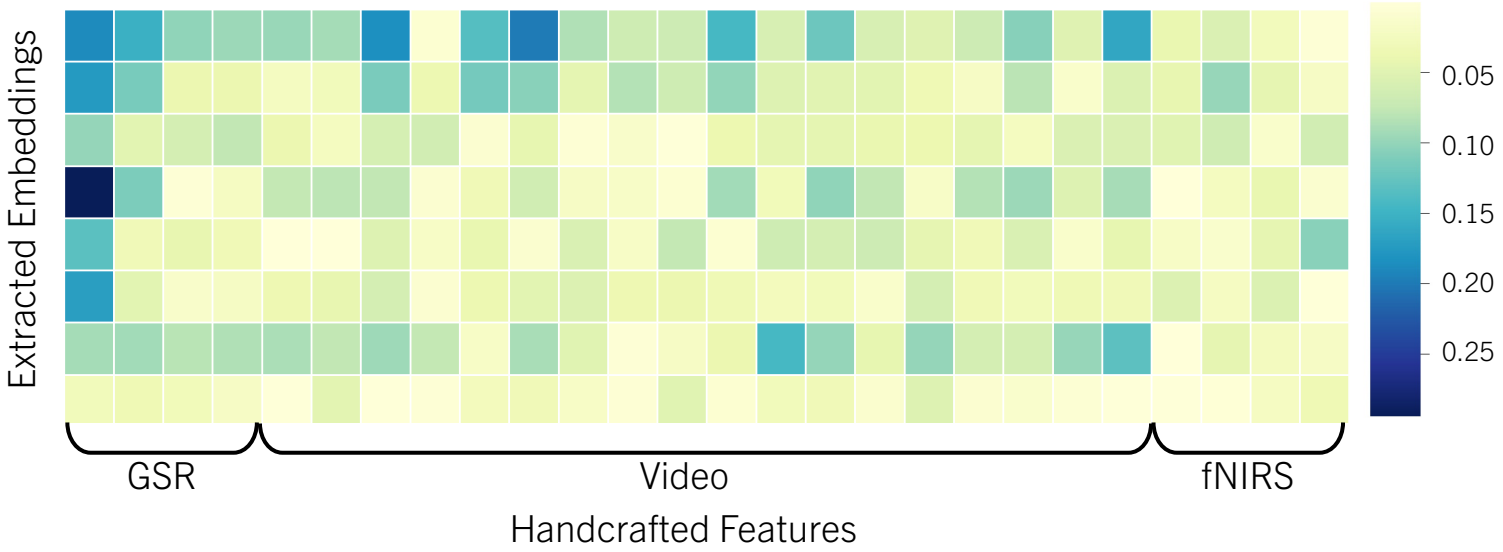
Modality-specific heads and late fusion to extract an embedding of length 8.



EVALUATING MULTIMODAL EMBEDDINGS

Modality	# Trials With Valid Embeddings	AUROC	F1 Score
GSR + Video + fNIRS	1413	0.90	0.91
compared to baseline	2.03x ↑	28% ↑	34% ↑

Embeddings capture cross-modal features.



For more details, see our paper

"Extracting Multimodal Embeddings via Supervised Contrastive Learning for Psychological Screening"

in Proc. of the 10th International Conference on Affective Computing and Intelligent Interaction (ACII)

Manasa Kalanadhabhatta

Adrelys Mateo Santana

Deepak Ganesan

Tauhidur Rahman

Adam Grabell

UMassAmherst | Manning College of Information & Computer Sciences

College of Natural Sciences
Department of Psychological and Brain Sciences