The Role of Physiological Signals in Multimodal Emotion Recognition Solutions in the Era of Autonomous Driving

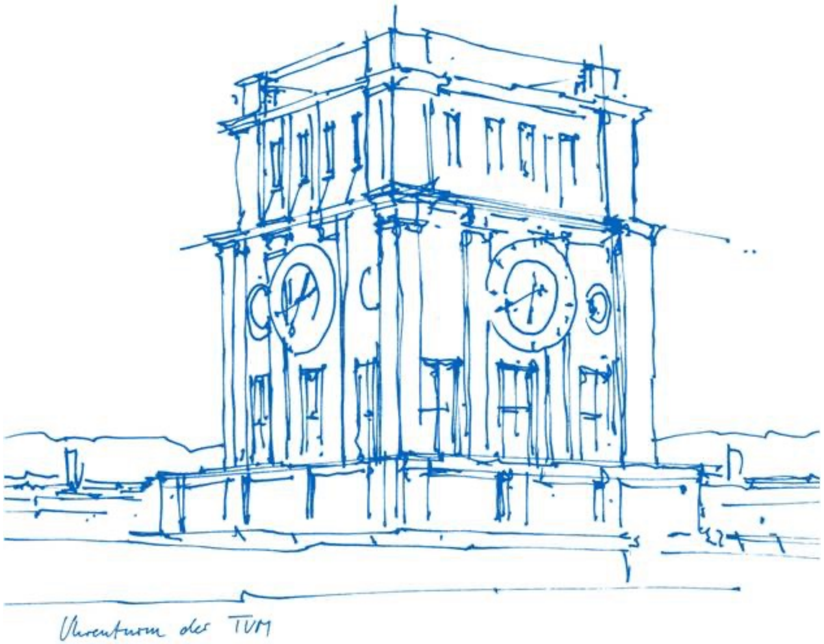
Simge Özcan

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

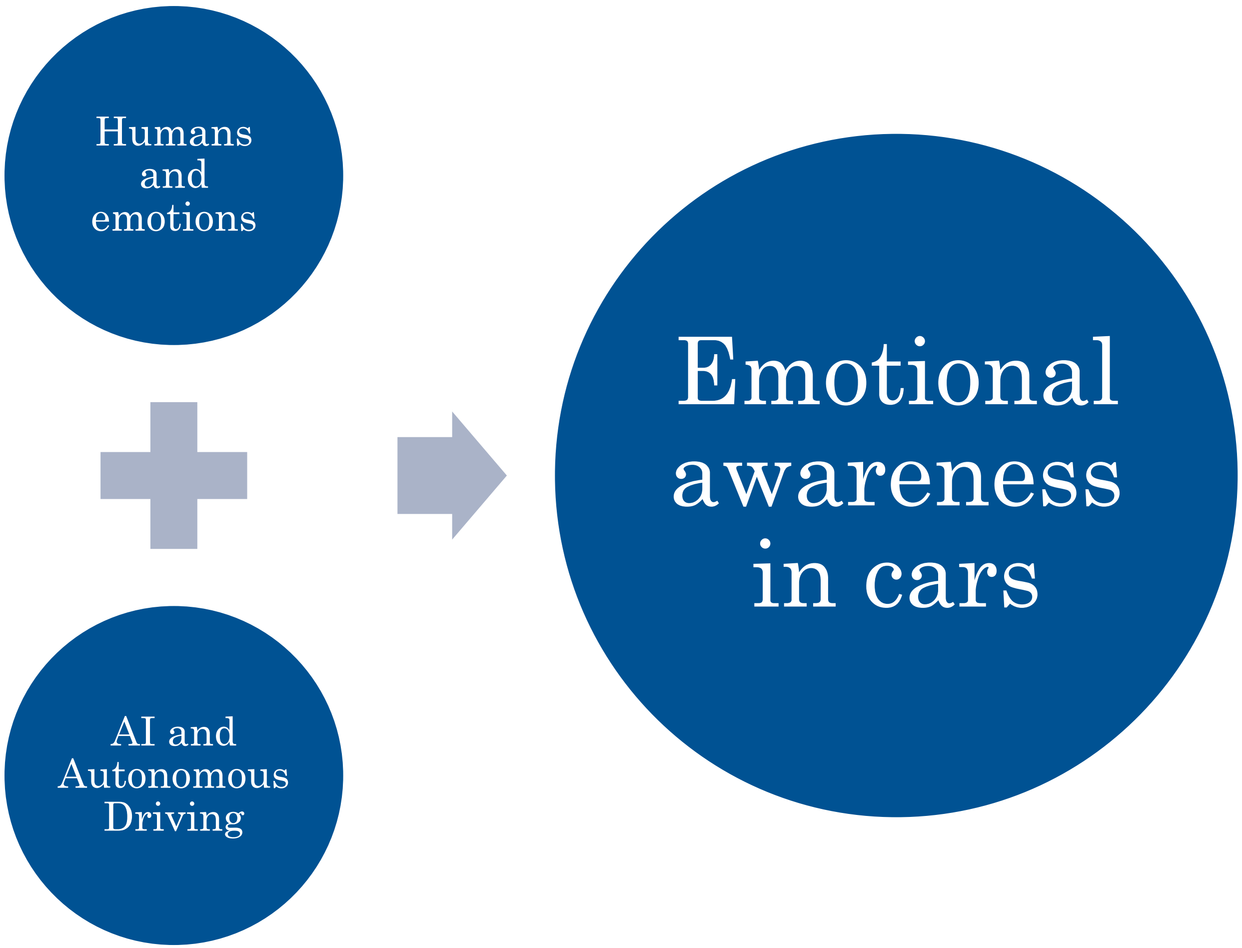
23.07.2021



Outline

- Motivation
- Problem Statement and Research Questions
- Proposed Approach and Methodology
- Experiments and Results
- Conclusion and Future Work

Motivation


Humans and emotions
AI and Autonomous Driving
Emotional awareness in cars

Motivation

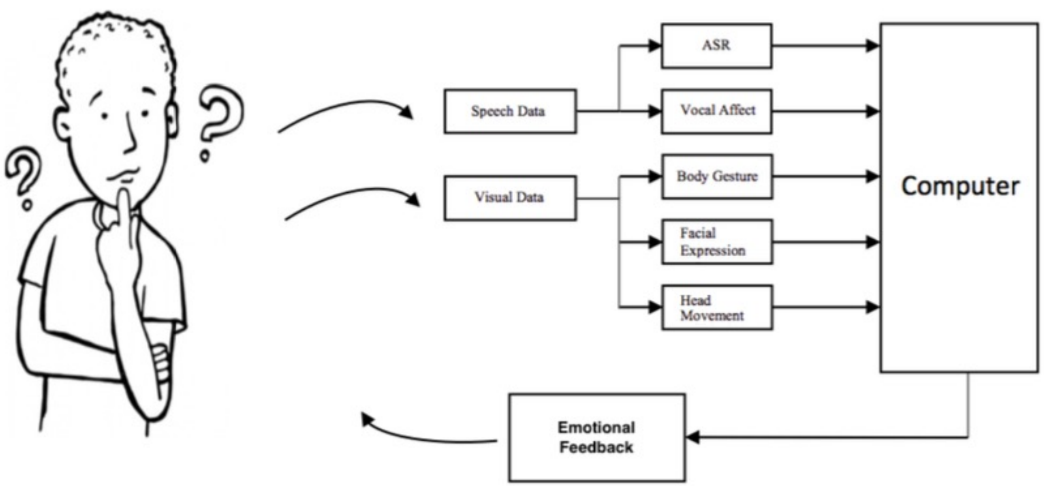
- Applications:
  - Improving the safety
  - Improving the comfort
  - Improving the mental health of the drivers
  - Providing personalized interface

- Physiological signals play a crucial role when it comes to emotions
  - Induce physiological changes
  - Linked to a particular pattern of physiological activity.
  - Heart rate, breathing rate, body temperature, and sweating level…

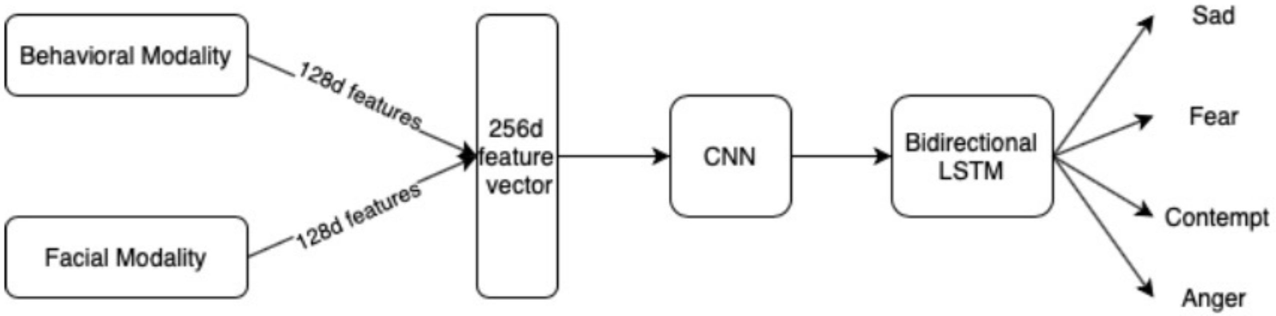
Problem Statement

- Emotions induce not only physiological but also behavioral, and cognitive changes
  - Facial expression
  - Behavioral cues
  - Physiological changes

- It is important to process information from different sensors
  - Research in Physiological signals is relatively new
  - Change of pattern in physiological signals is inevitable and detectable.
  - Non-obstructive sensors to detect the physiological signals are needed



Problem Statement: Baseline model



- Happy drivers are more likely to achieve an error-less driving experience
  - Focusing on abnormal emotions in cabin
- Emotion recognition problem is a multimodal learning task
  - Early fusion approach as baseline

Research Questions

- Which physiological signals could be used for in-cabin environment in order to recognize emotions in a non-invasive way and how can these input signals be grouped to achieve better performance?

- What common characteristics do DL architectures that perform well on physiological data have?

- How is the accuracy of DL approaches for emotion recognition from physiological signals, especially when combined with the signal preprocessing techniques before model training?

- How does fusing camera-based and behavioral approaches with physiological signals enhance the performance?

Proposed Approach

Identify and ***preprocess*** the relevant signals

Develop several ***end-to-end DL models*** utilizing ***physiological signals***

Fuse the best performing model into the baseline model

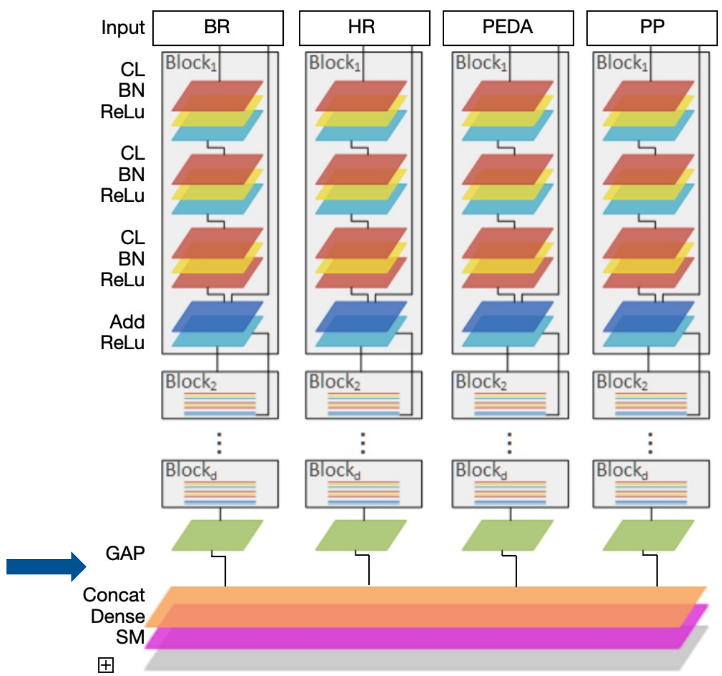
Methodology

- Preprocessing pipeline:
  - 3–97% winsorization, which removes extreme values form the signal data
  - Butterworth low-pass filter with a 10 Hz cut-off which removes components above the threshold
  - Min-max normalization
  - Sliding windows of 30 seconds with 1 second slides

- End-to-end Deep Learning methods for time-series analysis for physiological signals:
  - Fully Convolutional Network (FCN)
  - Residual Network (Resnet)
  - Spectro-Temporal Resnet (Stresnet)

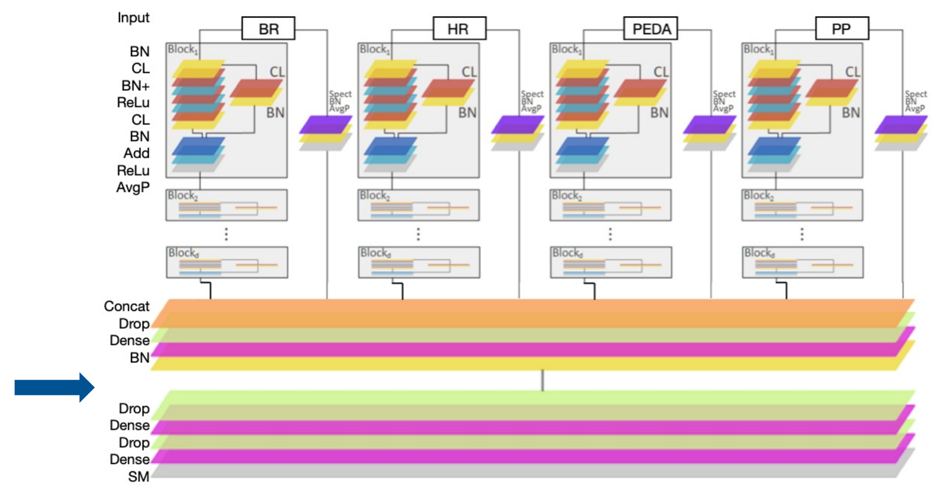
Methodology

Resnet



FCN

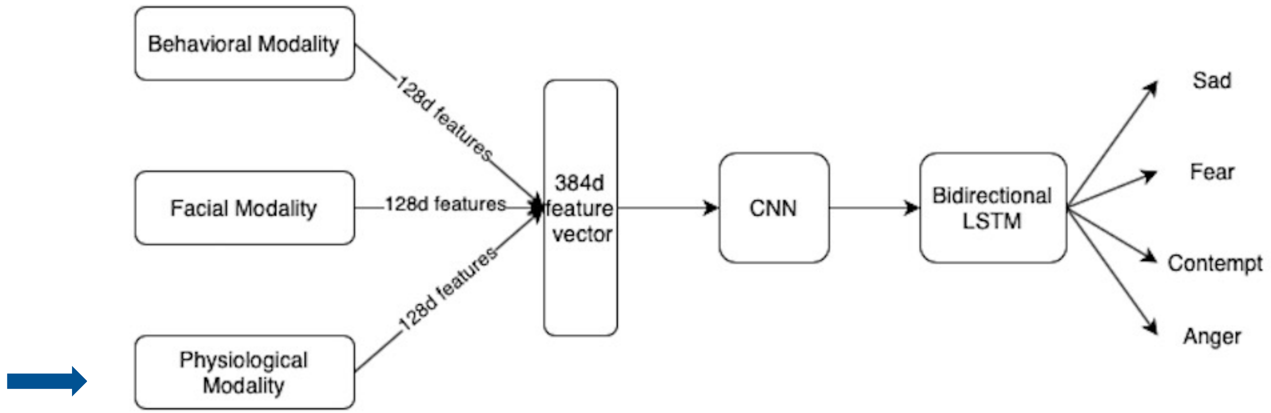
Stresnet



Methodology

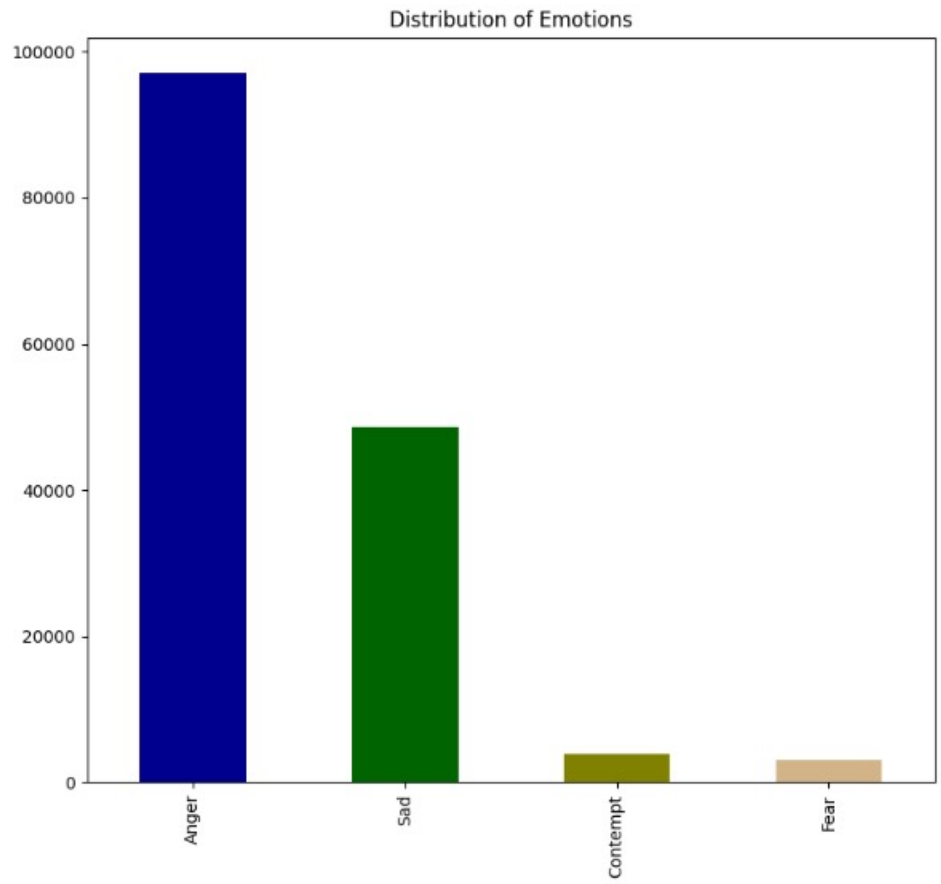
- Convolutional Neural Network (CNN) based models for
  - Facial modality
  - Behavioral modality

- Feature level fusion model for baseline and final fused models
  - Baseline model: Facial and Behavioral Modalities
  - Final fused model: Facial, Behavioral, and Physiological Modalities



Experiments and Results: Setup

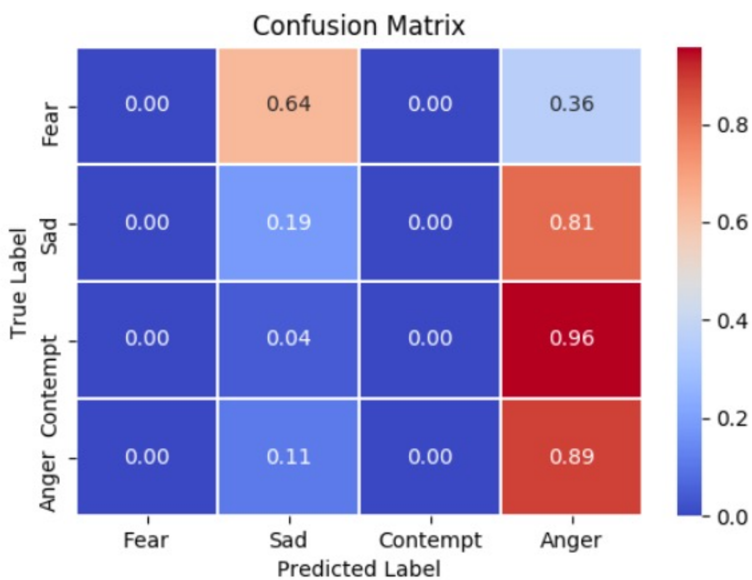
- 60% train, %20 validation, %20 test
- 50 epochs, with patience 15
- Adam optimizer
- Bayesion optimization for hyperparameters
- Class weights

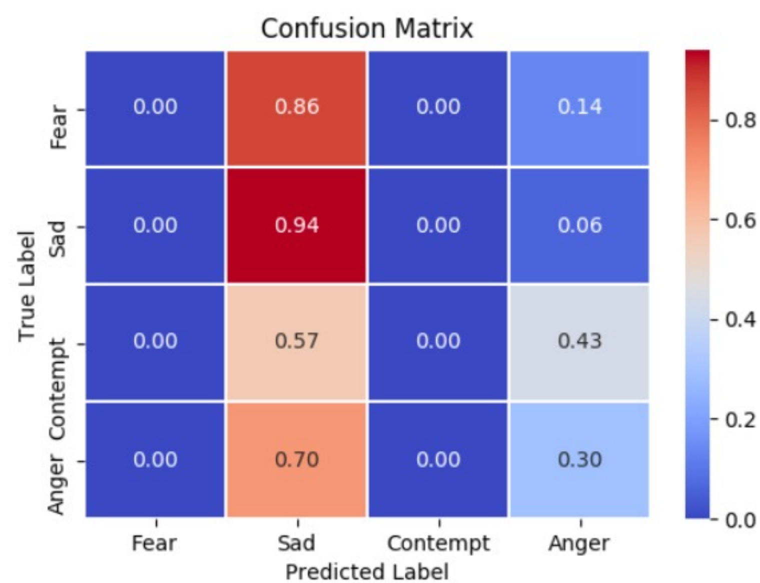


Experiments and Results: Resnet and FCN

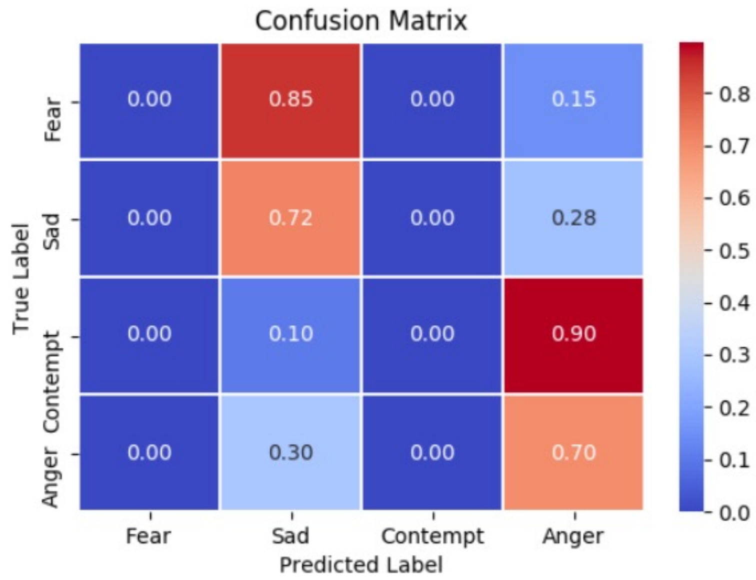
	Precision	Recall	F1-score	#Data points
fear	0.0	0.0	0.0	66
sadness	0.40	0.19	0.26	6461
contempt	0.0	0.0	0.0	1328
anger	0.67	0.89	0.76	15078
accuracy			0.64	22933
macro avg	0.27	0.27	0.26	22933
weighted avg	0.56	0.64	0.58	22933

	Precision	Recall	F1-score	#Data points
fear	0.0	0.0	0.0	66
sadness	0.35	0.94	0.51	6461
contempt	0.0	0.0	0.0	1328
anger	0.82	0.30	0.43	15078
accuracy			0.46	22933
macro avg	0.29	0.31	0.24	22933
weighted avg	0.64	0.46	0.43	22933



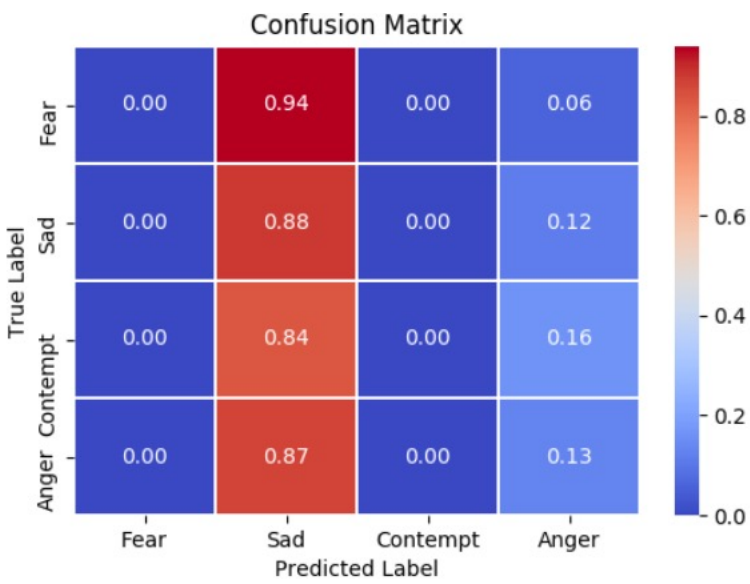


Experiments and Results: Stresnet



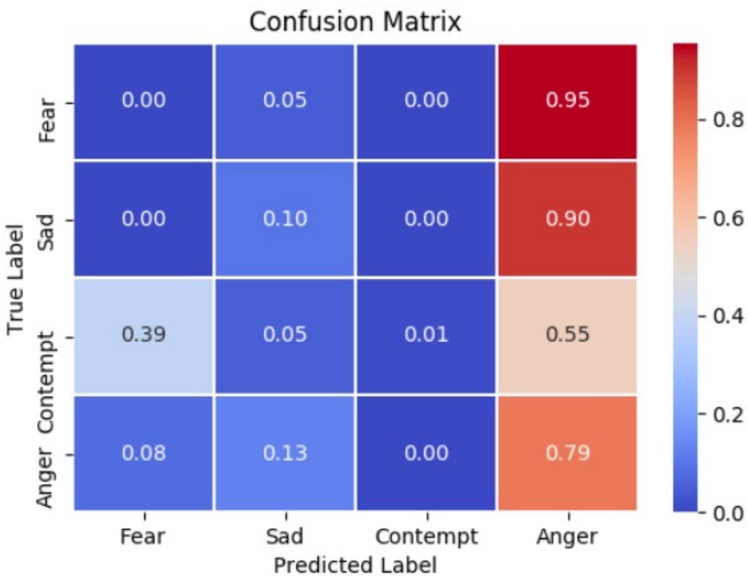
	Precision	Recall	F1-score	#Data points
fear	0.0	0.0	0.0	66
sadness	0.49	0.72	0.59	6461
contempt	0.0	0.0	0.0	1328
anger	0.78	0.70	0.74	15078
accuracy			0.66	22933
macro avg	0.32	0.35	0.33	22933
weighted avg	0.65	0.66	0.65	22933

Experiments and Results: Baseline model



	Precision	Recall	F1-score	#Data points
fear	0.0	0.0	0.0	66
sadness	0.29	0.88	0.43	6461
contempt	0.0	0.0	0.0	1328
anger	0.67	0.13	0.22	15078
accuracy			0.33	22933
macro avg	0.24	0.25	0.16	22933
weighted avg	0.52	0.33	0.27	22933

Experiments and Results: Final fused model



	Precision	Recall	F1-score	#Data points
fear	0.0	0.00	0.0	66
sadness	0.24	0.10	0.14	6461
contempt	0.60	0.01	0.02	1328
anger	0.64	0.79	0.71	15078
accuracy			0.55	22933
macro avg	0.37	0.22	0.22	22933
weighted avg	0.57	0.46	0.49	22933

Conclusion and Future Work

Conclusions

- Heart rate, breathing rate, palm EDA and perinasal EDA are relevant features and can be grouped together to classify negative emotions
- End-to-end DL models that analyzes the input data in both spectral and temporal domains achieves a better performance.
- Physiological signals can be used to enhance the emotion recognition performance in a multimodal recognition system

Scope for improvement

- A more balanced dataset
- A more sophisticated ground truth labelling scheme

Possible future works

- Experimenting with different preprocessing techniques or more NN architectures
- Experimenting with other fusion approaches, such as decision level
- Adding more modalities, such as audio modality

References

- Dziezyc, Maciej & Gjoreski, Martin & Kazienko, Przemysław & Saganowski, Stanisław & Gams, Matjaz. (2020). Can We Ditch Feature Engineering? End-to-End Deep Learning for Affect Recognition from Physiological Sensor Data. Sensors. 20. 6535. 10.3390/s20226535.

- Tzirakis, Panagiotis & Trigeorgis, George & Nicolaou, Mihalis & Schuller, Björn & Zafeiriou, Stefanos. (2017). End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. IEEE Journal of Selected Topics in Signal Processing. PP. 10.1109/JSTSP.2017.2764438.

- S. Taamneh, P. Tsiamyrtzis, M. Dcosta, P. Buddharaju, A. Khatri, M. Manser, T.Ferris, R. Wunderlich, and I. Pavlidis. “A multimodal dataset for various formsof distracted driving.” In:Scientific Data4.1 (2017), p. 170110.

- F. Eyben, M. Wöllmer, T. Poitschke, B. Schuller, C. Blaschke, B. Färber, and N. Nguyen-Thien. “Emotion on the road—necessity, acceptance, and feasibility of affective computing in the car.” In: *Advances in human-computer interaction* (2010).

- S. Poria, E. Cambria, R. Bajpai, and A. Hussain. “A review of affective computing: From unimodal analysis to multimodal fusion.” In: *Information Fusion* 37 (2017), pp. 98–125.

Thank you!

Simge Özcan

Master Thesis in Informatics

23.07.2021