



Multimodal Fusion with Modality-Specific Factors for IEMOCAP dataset

Abishay Reddy

College Park Scholars – Science & Global Change Program
Computer Science
areddy17@umd.edu
CPSP395G



College Park Scholars Academic Showcase, May 5, 2022

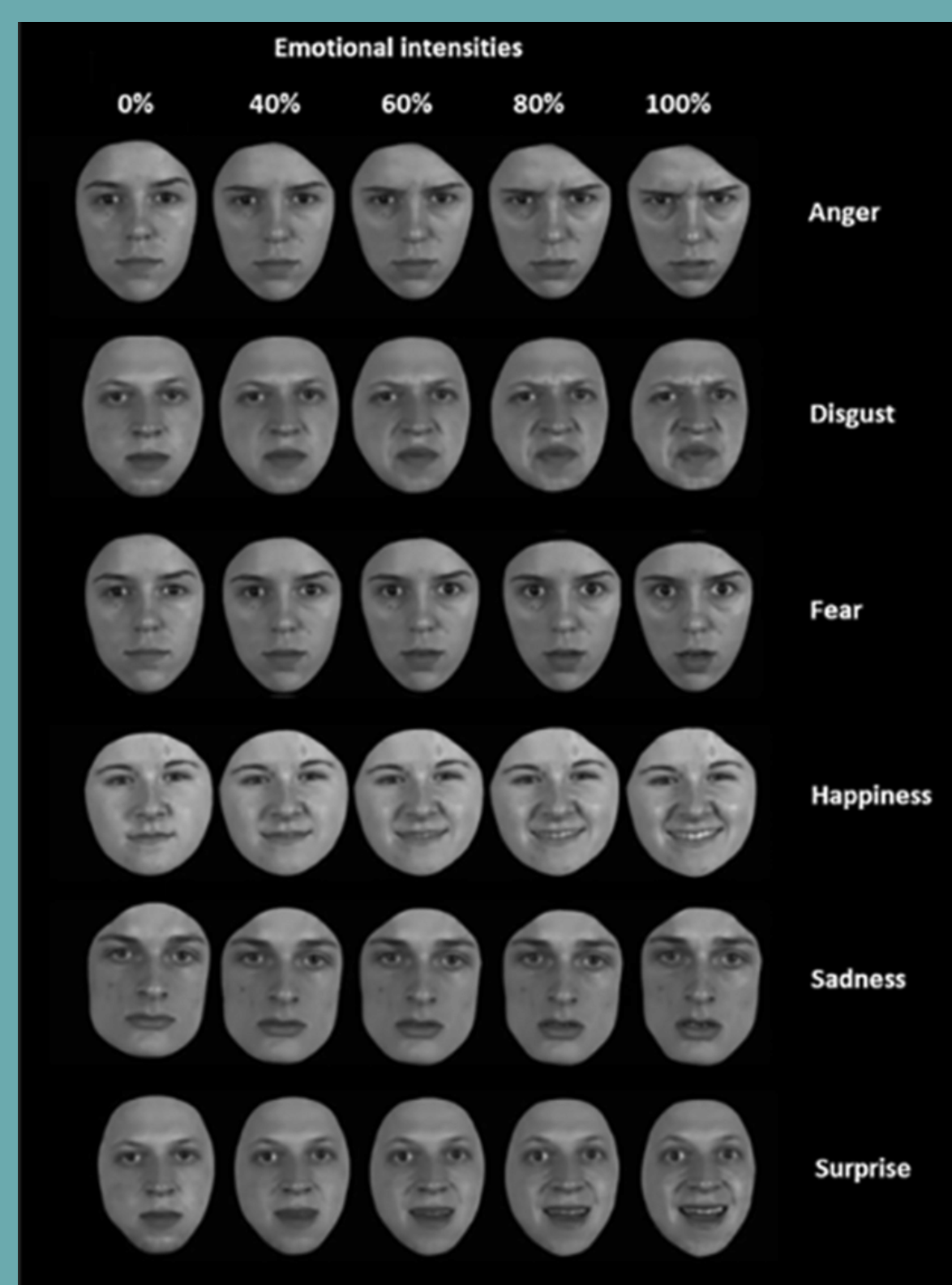
Introduction

My research project focused on developing a method for accurately recognizing human emotions from different communication modalities including video, audio, and text.

The project's main objective is to showcase accuracy of:

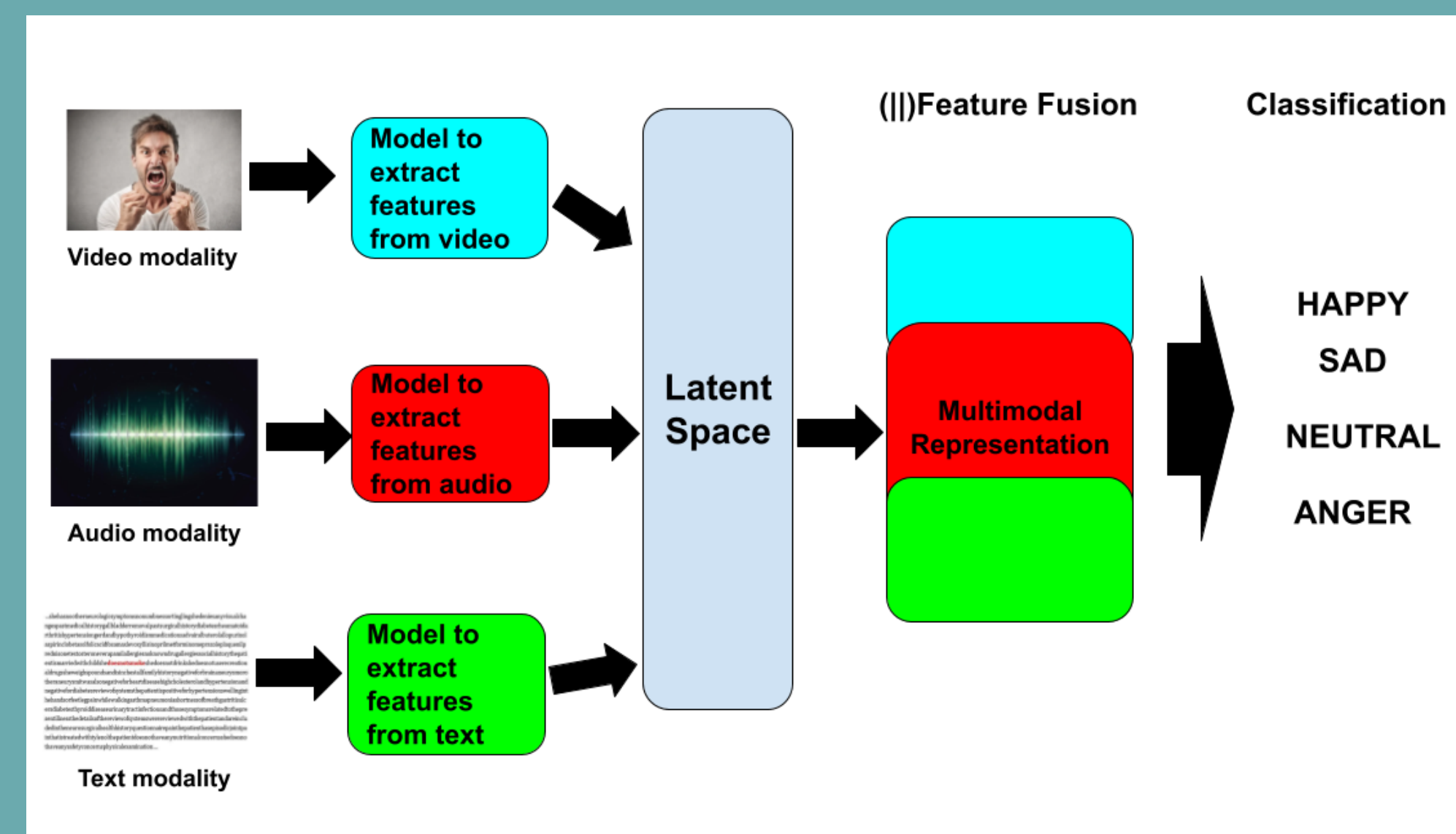
- Feature extraction,
- Selection processes
- Algorithm emotion prediction accuracy

To achieve this goal, we utilized the IEMOCAP dataset, which is a widely used benchmark dataset for emotion recognition research.



Materials:

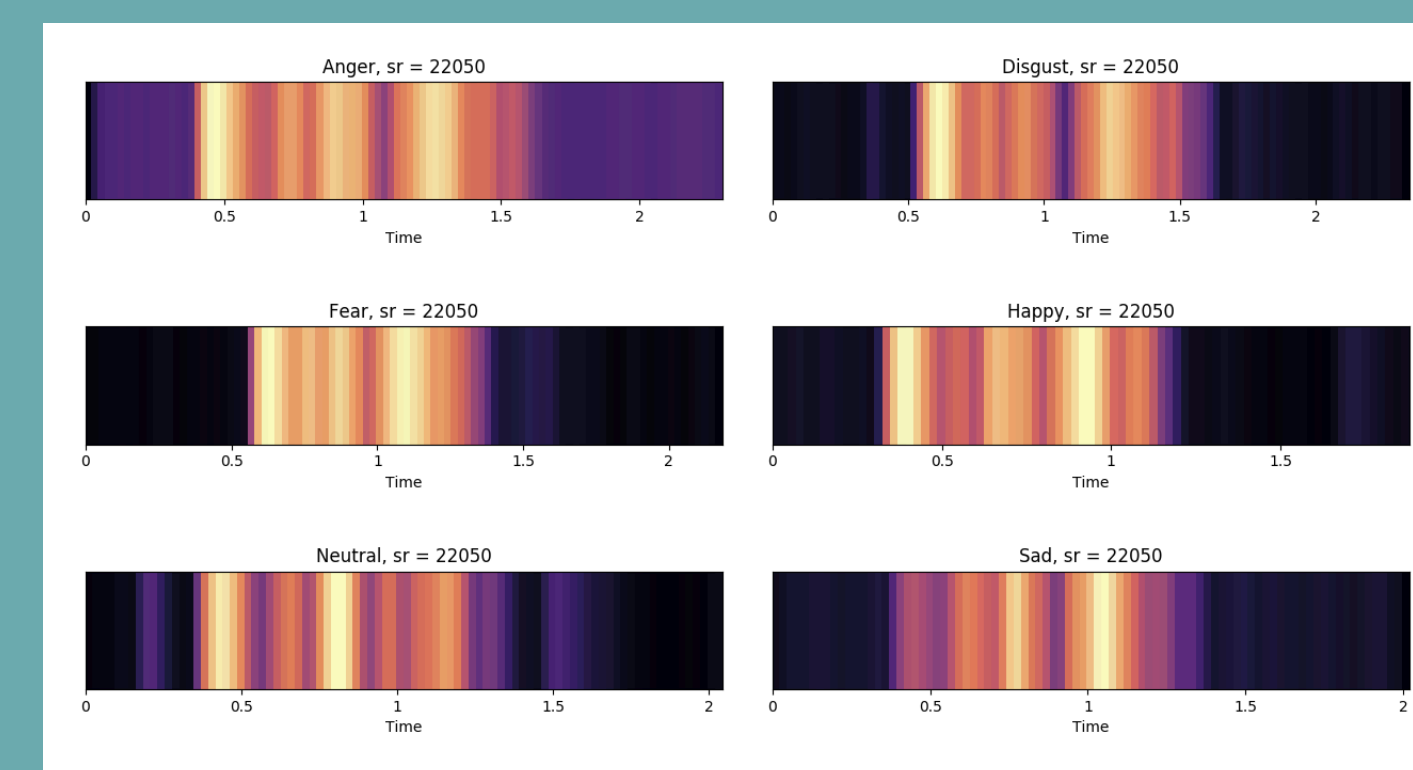
- Audiovisual data collected from IEMOCAP dataset
- Python- Librosa, NumPy
- I3d and c3d models



Methodology/Activities:

- Carry out feature extraction, selection, and fusion on audiovisual data in the IEMOCAP dataset to be used in training of emotion recognition algorithms.
- Process unimodal data samples into multimodal data representations which offer better classification results.
- Worked with a number of different tools and was exposed to different NLP(Natural Language Processing) development methods.

Results:



Discussion:

- Discussed and implemented multiple multimodal fusion methods and a variation of steps in development.
- Worked through feature extraction, selection, and fusion for audio/video modalities on the IEMOCAP dataset.
- Started work on the preliminary iteration of the classification model and visualized processed data

Future Work:

- For this project: Continue project with more dataset application and algorithm development.
- For lab: Continue work in the space of emotion recognition.

Acknowledgments:

Thank you Dr.Sarala Padi for facilitating this opportunity for me, and being with me every step of the way during the project development stage. Lastly thank you to Dr.Thomas Holtz and Dr.John Merck for overseeing my time in CPSG.

Site Information:

NIST Information Technology Laboratory
100 Bureau Dr, Gaithersburg, MD 20899
Dr.Sarala Padi

Expand upon current emotion recognition technology, by utilizing different approaches to algorithm development.

Analyze effectiveness of feature selection and fusion with the IEMOCAP dataset in focus

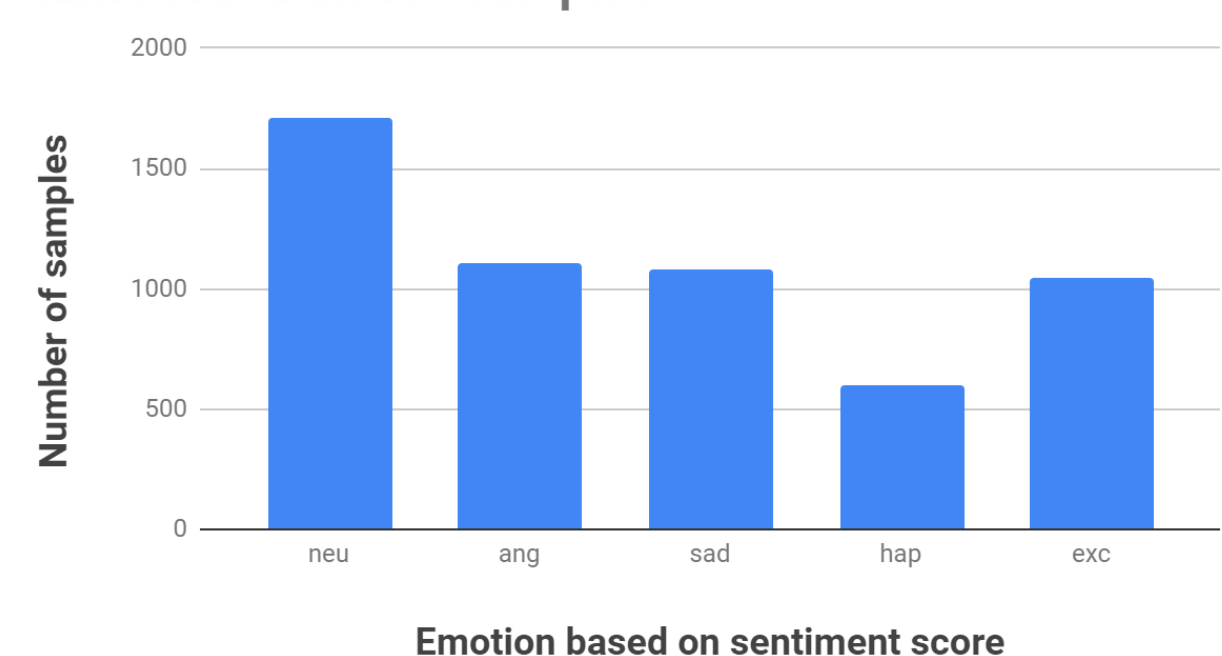
IEMOCAP

10,038 corpus samples

Database consists of twelve hours of audiovisual data

Includes video, speech, motion capture of face, and text transcriptions

IEMOCAP Dataset Samples



SCIENCE AND GLOBAL CHANGE

