Privacy Preservation in Emotion Recognition Under Federated Learning Settings

REU fellow: Jason Nuñez[1]
Faculty Mentors: Dr. Houwei Cao[2]
Affiliation: [1.] Amherst College, [2.] School of Engineering and Computing Science, NYIT
Emails: jasonnunez105@gmail.com, hcao02@nyit.edu

ABSTRACT

As more data is produced, internet companies and machine learning (ML) systems are increasingly accessing user data to train and improve their prediction models. Communicating user data to a centralized cloud-based or physical server runs the risk of exposing demographic or cultural identifiers. People do not want to disclose sensitive information but the utility of many of these models relies on having direct access to this data to make more accurate predictions. To satisfy users' privacy interests and continue the efficacy of prediction algorithms, such as emotion recognition (ER) as studied in this paper, model updates should be implemented under federated learning (FL) settings.

BACKGROUND

SPEECH EMOTION RECOGNITION aims to identify the high-level affective status of biometric input from the low-level features.
Feature extraction often includes

- pitch-related features
- energy-related features
- Mel-frequency cepstrum coefficients [1]

FEDERATED LEARNING is a distributed learning frameworks where local data never leaves the owners device and model updates use only an aggregate of participating party's weights.

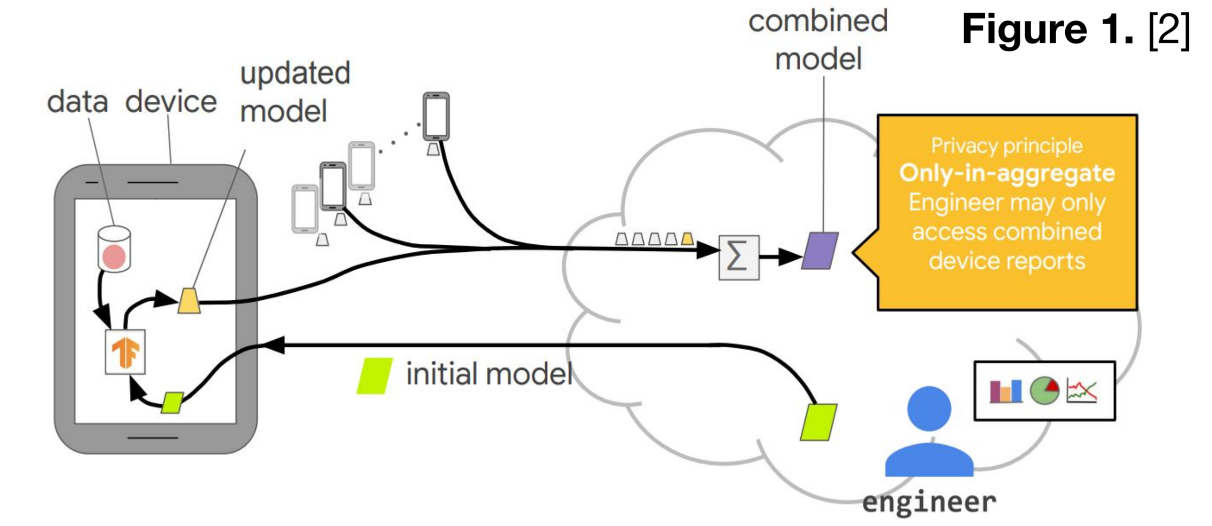


Figure 1. [2]

EXPERIMENTAL SET-UP

DATA The Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) is a large multimedia stimulus set produced by 91 professional actors of various ages, races, and ethnicities. We will solely use the audio files for speech ER. Each actor and their respective recordings will represent a single party.
FEATURES To extract low-level features, we used openSMILE and the provided ComParE-16 configuration.
BASELINE In the dependent experiment, the CREMA-D will be split into training and test sets with a 70:30 ratio. Such an experiment will best simulate a speech ER model that does not have direct access to user data. In the subject-independent trial we will use the leave-one-out cross validation process. Because the model is trained and tested on each party, this trial best simulates when user data is exposed to the centralized ML model. These experiments represent traditional methods to ER.
FEDERATED LEARNING To implement the speech ER model in a FL setting, we will be using stochastic gradient descent for model optimization in order to simulate a FedAvg algorithmic approach [3].

METHODOLOGY

Feature Extraction

openSMILE:) by audEERING™

Or the Munich Open-Sourced Large-Scaled Multimedia Feature Extractor

Configuration provided by the INTERSPEECH 2016 Computational Paralinguistics Challenge (ComParE-16)

Dependent experiment

Baseline Dependent Experiment

- Using SKLearn's linear SVM
- Data is split into training and test sets 70:30

Baseline Independent Experiment

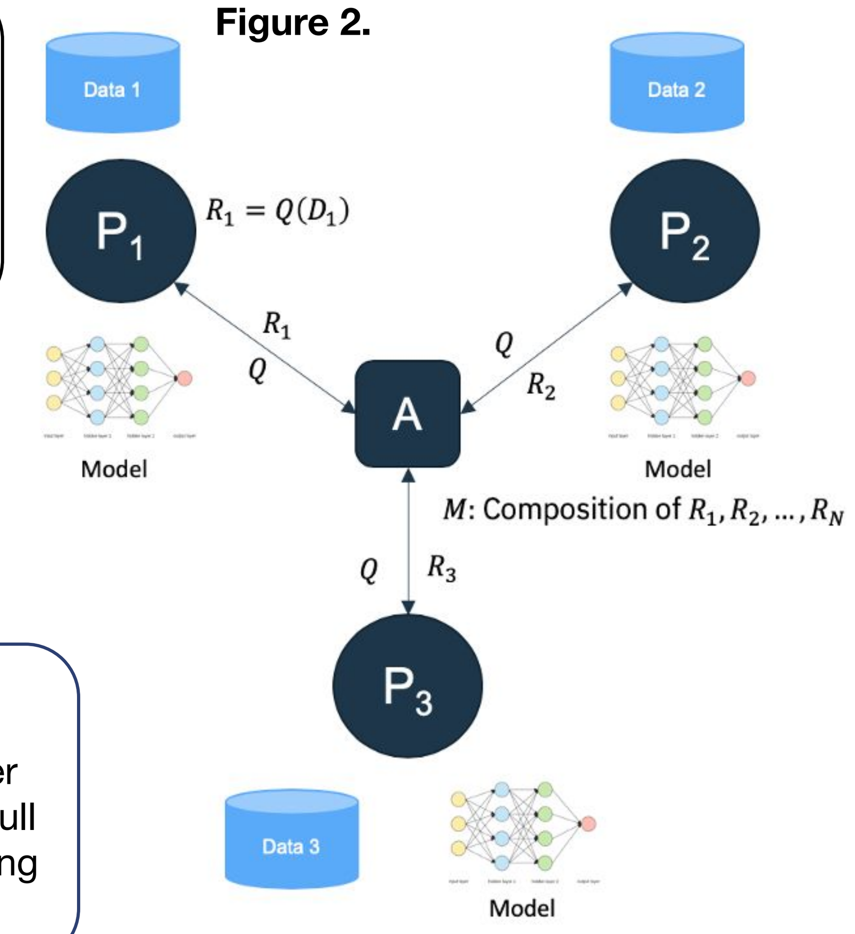
- Using SKLearn's linear SVM
- Leave-one-out cross validation

Federated Learning Experiment

- Using SKLearn's SGDclassifier
- IBM's package provides full implementation of an FL setting

Figure 2. [4]

Figure 2.



RESULTS

Dependent experiment

```
UAR: 0.4999693350390774
              precision    recall  f1-score   support

         ANG       0.58      0.80      0.67       364
         DIS       0.40      0.37      0.39       364
         FEA       0.45      0.38      0.41       364
         HAP       0.50      0.35      0.41       452
         NEU       0.53      0.43      0.48       361
         SAD       0.47      0.66      0.55       364

    accuracy                           0.49      2269
   macro avg       0.49      0.50      0.48      2269
weighted avg       0.49      0.49      0.48      2269


Confusion Matrix:

[[293  19   9  32  10   1]
 [ 45 135  40  28  40  76]
 [ 28  49 138  27  24  98]
 [124  51  59 156  36  26]
 [ 11  30  26  60 157  77]
 [  4  50  34   6  28 242]]
```

Independent experiment

```
UAR: 0.6100588916385649
              precision    recall  f1-score   support

         ANG       0.66      0.81      0.73      1271
         DIS       0.65      0.48      0.56      1271
         FEA       0.56      0.47      0.51      1271
         HAP       0.59      0.59      0.59      1271
         NEU       0.60      0.70      0.65      1087
         SAD       0.58      0.60      0.59      1271

    accuracy                           0.61      7442
   macro avg       0.61      0.61      0.60      7442
weighted avg       0.61      0.61      0.60      7442


Confusion Matrix:

[[1030   49   31  105   50    6]
 [ 139  615   89  116  134  178]
 [ 123   62  603  171   75  237]
 [ 210   46  136  754   93   32]
 [  36   60   51   80  758  102]
 [  22  113  175   46  151  764]]
```

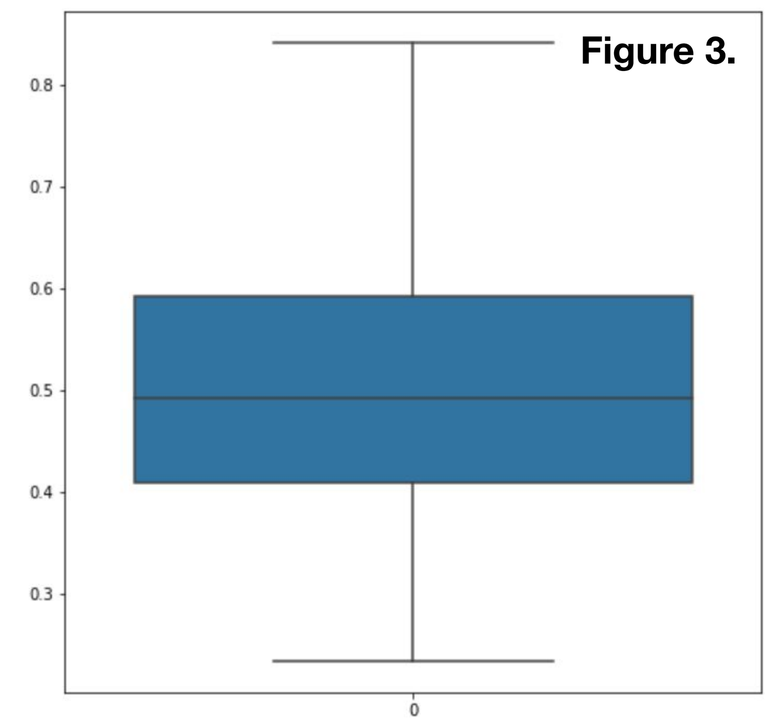


Figure 3.

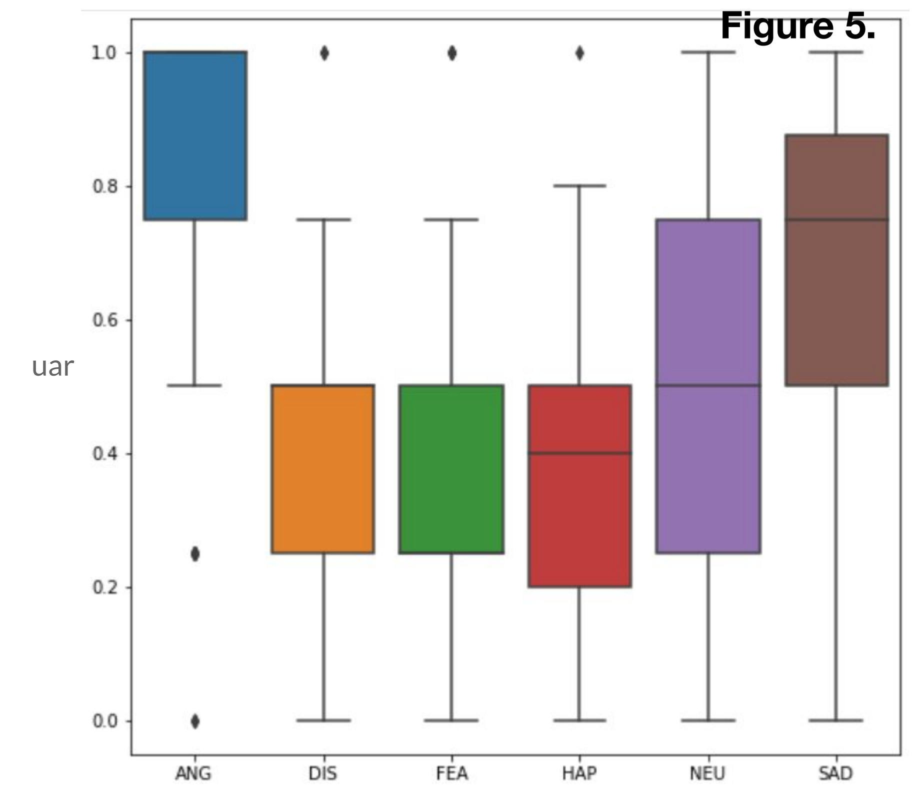


Figure 5.

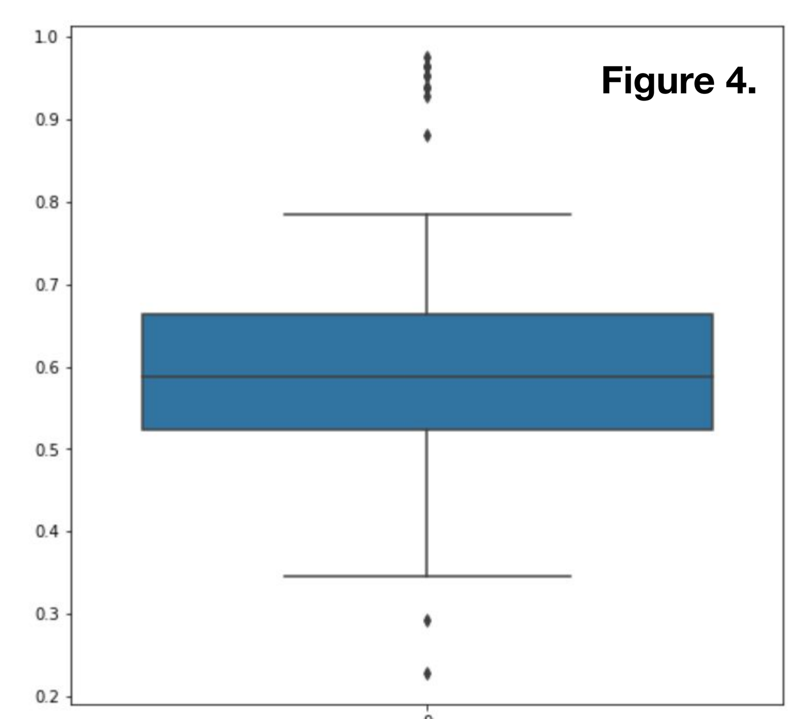


Figure 4.

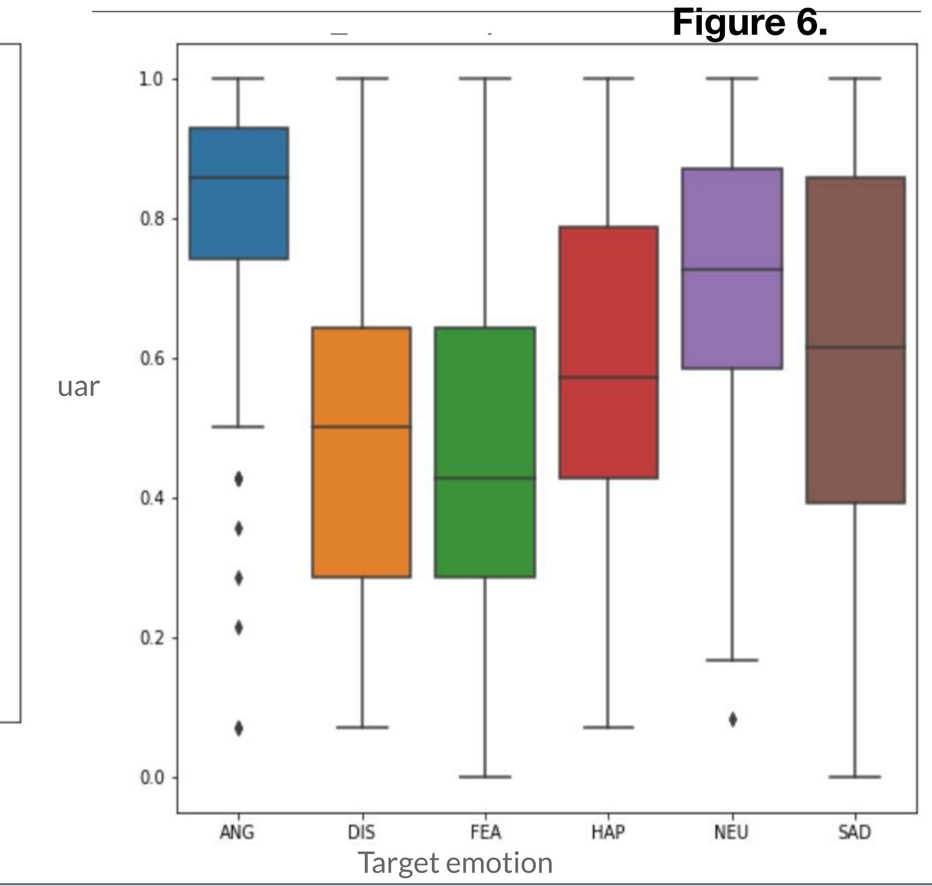


Figure 6.

DISCUSSIONS

- The baseline results show good UAR scores considering this only consider features from audio input, not from the visual input, the actors' facial expressions
- The independent trial, as expected, achieves better accuracy but more variability between subjects
- The results of the FL experiment should be no worse than the dependent trial, ideally better than the independent

FUTURE WORK

- Use various low-level feature configurations, such as
- Improve overall utility by extracting visual features from CREMA-D
- Perform speech ER under FL settings using other packages, such as TensorFlow Federated or Flower

CONCLUSIONS

In this work, our aim is to reveal that there does not exist significant costs to machine learning utility in federated learning settings. Because of time and resource constraints of the project we were not able to fully perform FL experiments. Future work in this project encourages the use of such privacy methods.

REFERENCES

[1] Han, Kun, Dong Yu, and Ivan Tashev. "Speech emotion recognition using deep neural network and extreme learning machine." In Interspeech 2014. 2014.
[2] "Federated Learning with Tensorflow Federated (TF World '19)." YouTube, YouTube, 10 Dec. 2019, https://www.youtube.com/watch?v=m17IgaHaoTI.
[3] Aledhari M, Razzak R, Parizi RM, Saeed F. Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications. IEEE Access. 2020;8:140699-140725. doi: 10.1109/access.2020.3013541. Epub 2020 Jul 31. PMID: 32999795; PMCID: PMC7523633.
[4] "IBM Federated Learning – Machine Learning Where the Data Is." IBM Research Blog, 21 Aug. 2020, https://www.ibm.com/blogs/research/2020/08/ibm-federated-learning-machine-learning-where-the-data-is/.

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation (NSF) Research Experiences for Undergraduates (REU) program. We would like to thank all of the mentors and research fellows at the New York Institute of Technology who have provided their helpful insight and expertise that greatly assisted with our research. We want to extend our thanks to the helpful post graduate students, Ziqing Yang and Michael Trzasko, for their constant help over the program's duration.

This project is funded by National Science Foundation Grant No. CNS-1852316 and New York Institute of Technology.