Privacy-Preserving Federated Learning for Depression Assessment

Murat Şahin - 22301345
Melih Coşğun - 22301344

Presentation Structure

- Introduction
- Dataset
- Centralized Learning
  - Preprocessing
  - Methodology
- Federated Learning
  - Background
  - Methodology
- Evaluation
- Future work
- Conclusion

Introduction - DL Motivation

Depression is a serious condition

- Depression affects millions of people worldwide, harms overall life
- Curable after the correct diagnosis
- Assessment is done by popular questionnaires - they require collaboration

- Deep learning - is good - can use multimodal features

Introduction - FL Motivation

Addressing data sparsity + privacy

- Federated learning allows training ML models without seeing the whole data
- It can help clinics or hospitals to collaborate without compromising privacy
- In real world scenarios, this can lead usage of more data
  and allows for more generalizable models

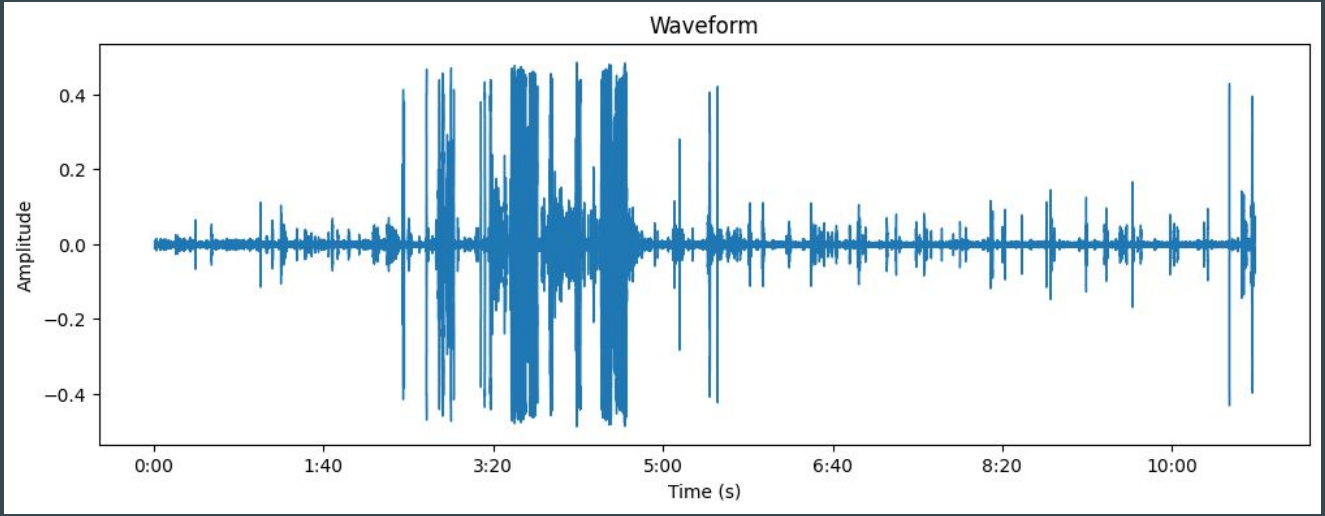
Dataset

- **E-DAIC** dataset, which is used in **AVEC 2019** challenge
- Clinical interviews done by a virtual agent, ~300 samples
- Contains video features, audio recordings and transcripts
- Binary classification (depressed or non depressed)

Centralized Learning

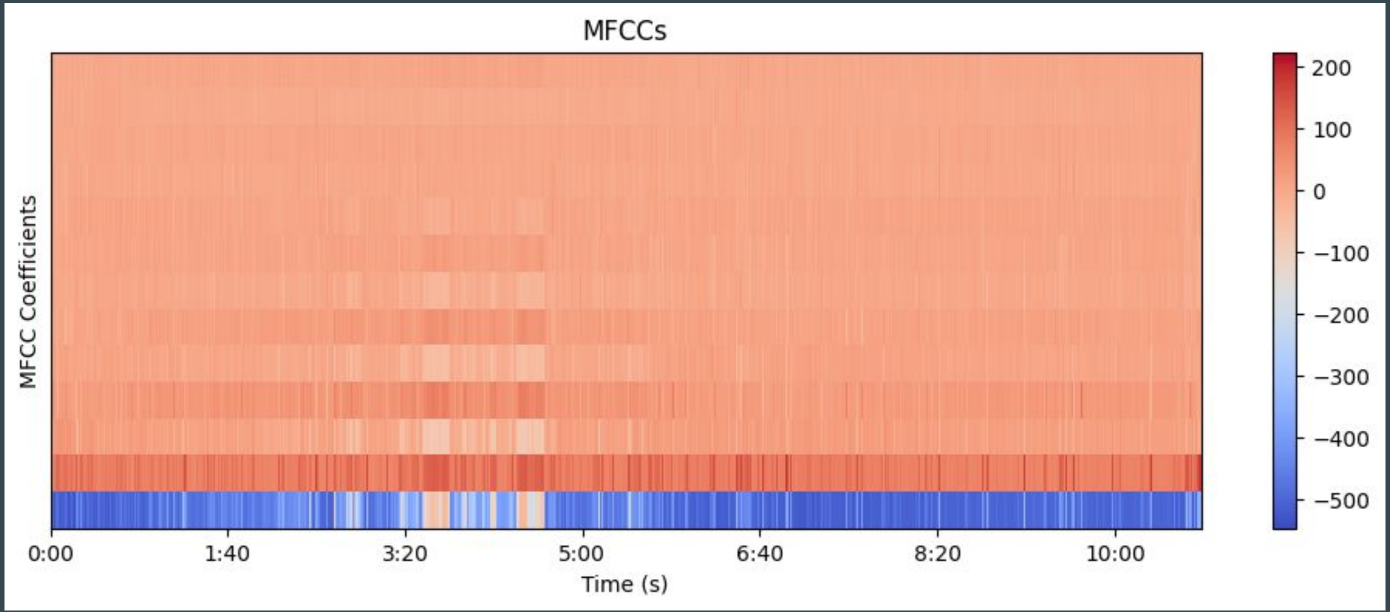
Centralized Learning - Preprocessing

- How to use audio data? This is what you have:



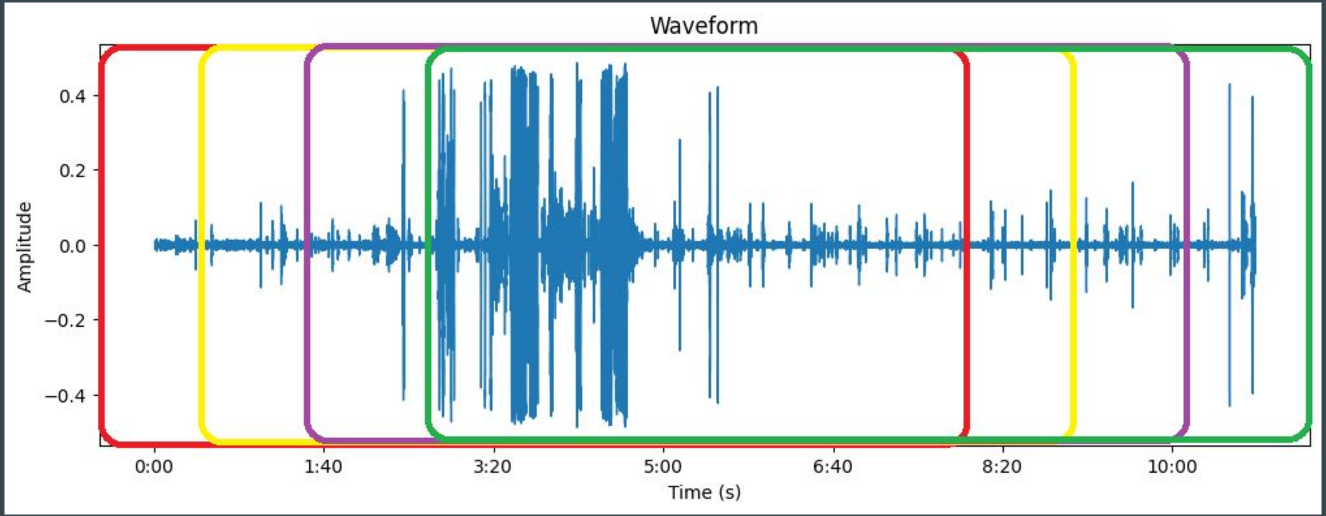
Centralized Learning - Preprocessing

- MFCC coefficients can be extracted from the audio signal for further usage:



Centralized Learning - Preprocessing

- Our dataset has **~15 minute** samples
- Samples with **8 minutes** of speech extracted using **Weighted Random Sampling**



Centralized Learning - Preprocessing

- What is **Weighted Random Sampling**?

E-DAIC dataset is not balanced, nearly **75%** of samples are from **non-depressive** individuals.

We extract 8 minutes of random speech segments from each sample in following quantities:

10 sample if non-depressive, **30 sample** if depressive

This helps model to generalize well.

Centralized Learning - Preprocessing

- What is **Weighted Random Sampling**?

	Non-Depressive	Depressive
Training Samples	126	37
Validation Samples	44	12
Test Samples	39	17

TABLE I
DISTRIBUTION BEFORE SAMPLING

	Non-Depressive	Depressive
Training Samples	1240	1110
Validation Samples	440	360
Test Samples	390	170

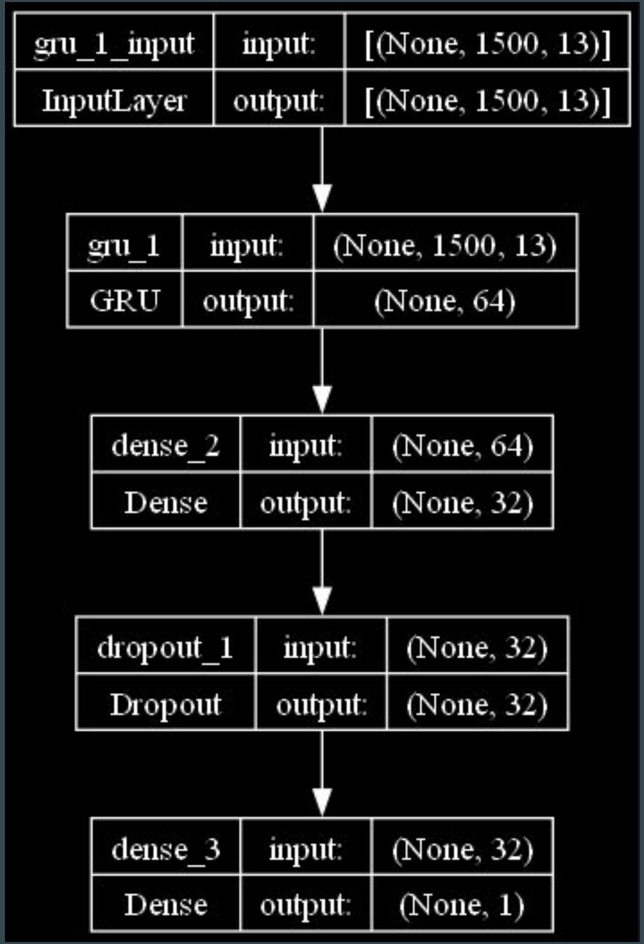
TABLE II
DISTRIBUTION AFTER SAMPLING

Centralized Learning - Methodology

- We tried different convolutional and recurrent architectures
- **Single layer GRU** and a following fully-connected layer worked the best
- We tuned parameters over validation loss and results are comparable to earlier studies in the domain

Optimizer:

Adam, lr = 0.0003, bs=32, label smoothing (0.1), early-stopping

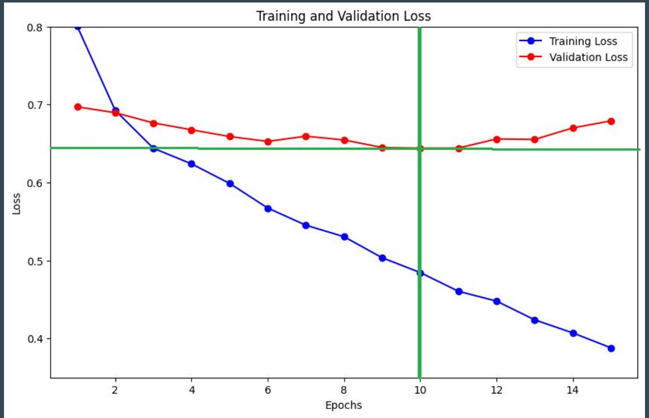


Centralized Learning - Failed Trials

- **Excluding segments** that subject is not speaking did not worked well: model overfits to frequency jumps
- Using **convolutional networks** over audio features did not perform well during our testings
- Re-implementing the **work from Lin et. al.** (1D CNN) including preprocessing and training did not yield meaningful results

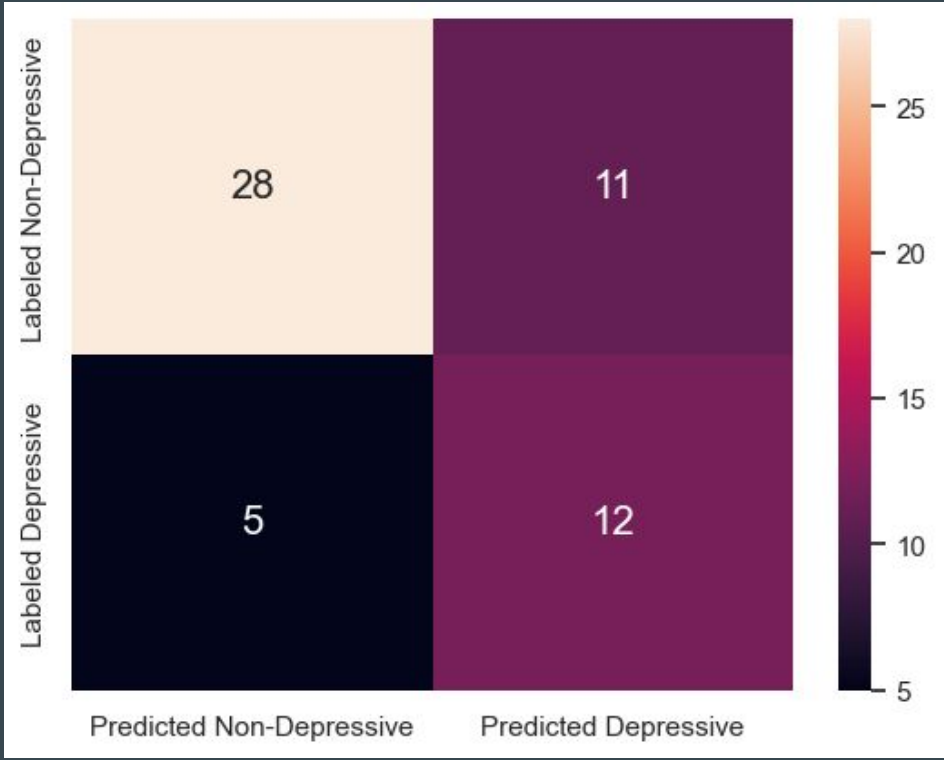
Lin, L., Chen, X., Shen, Y., & Zhang, L. (2020). Towards Automatic Depression Detection: A BiLSTM/1D CNN-Based Model. Applied Sciences.

Centralized Learning - Test Results



Class	F1	Prec.	Rec.	Acc.
Non-Depressive	0.78	0.85	0.72	0.71
Depressive	0.60	0.52	0.71	0.71

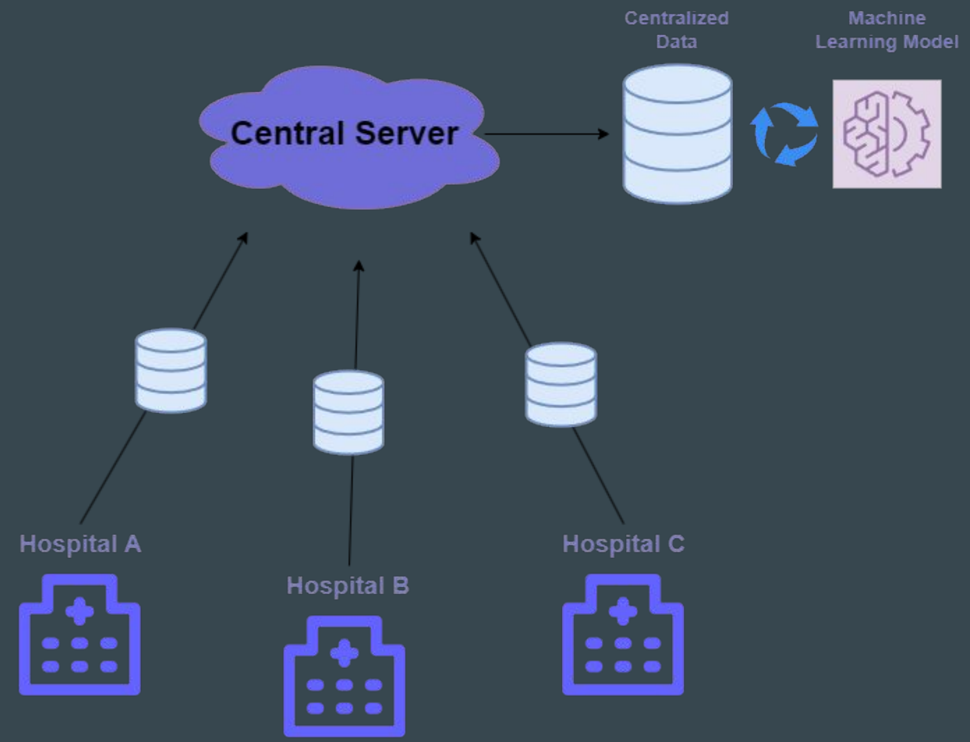
TABLE IV
RESULTS ON CENTRALIZED



Federated Learning

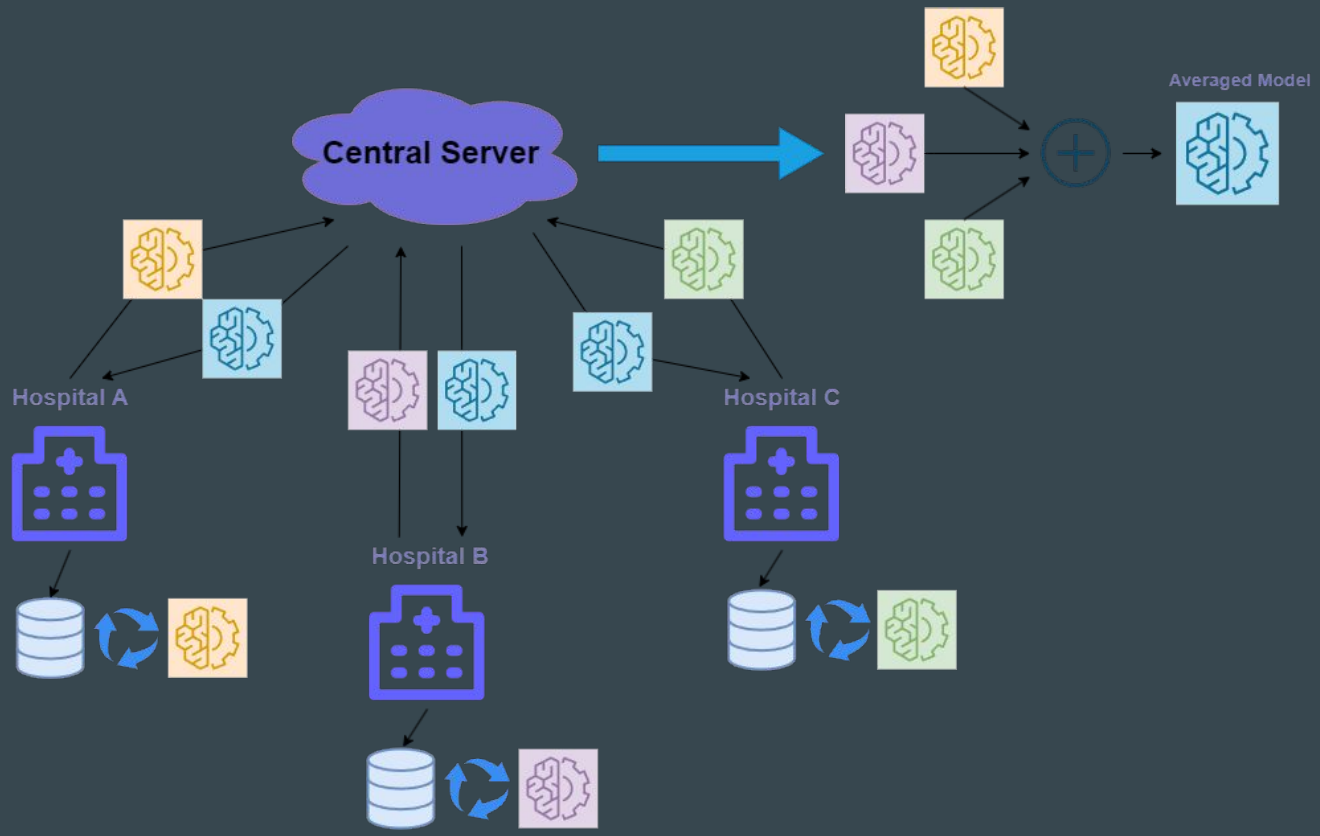
FL Background

Traditional ML Training on Depression Data



FL Background

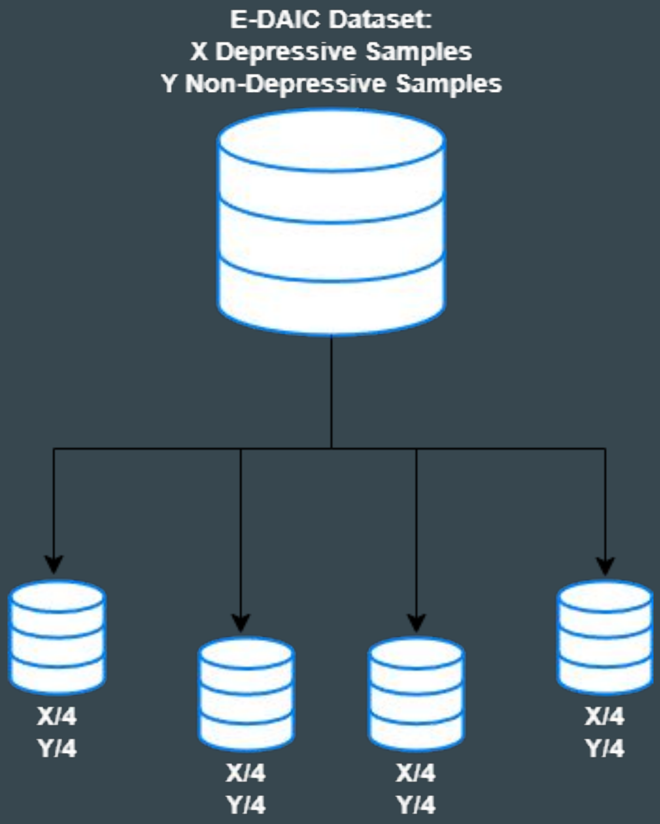
Federated Approach



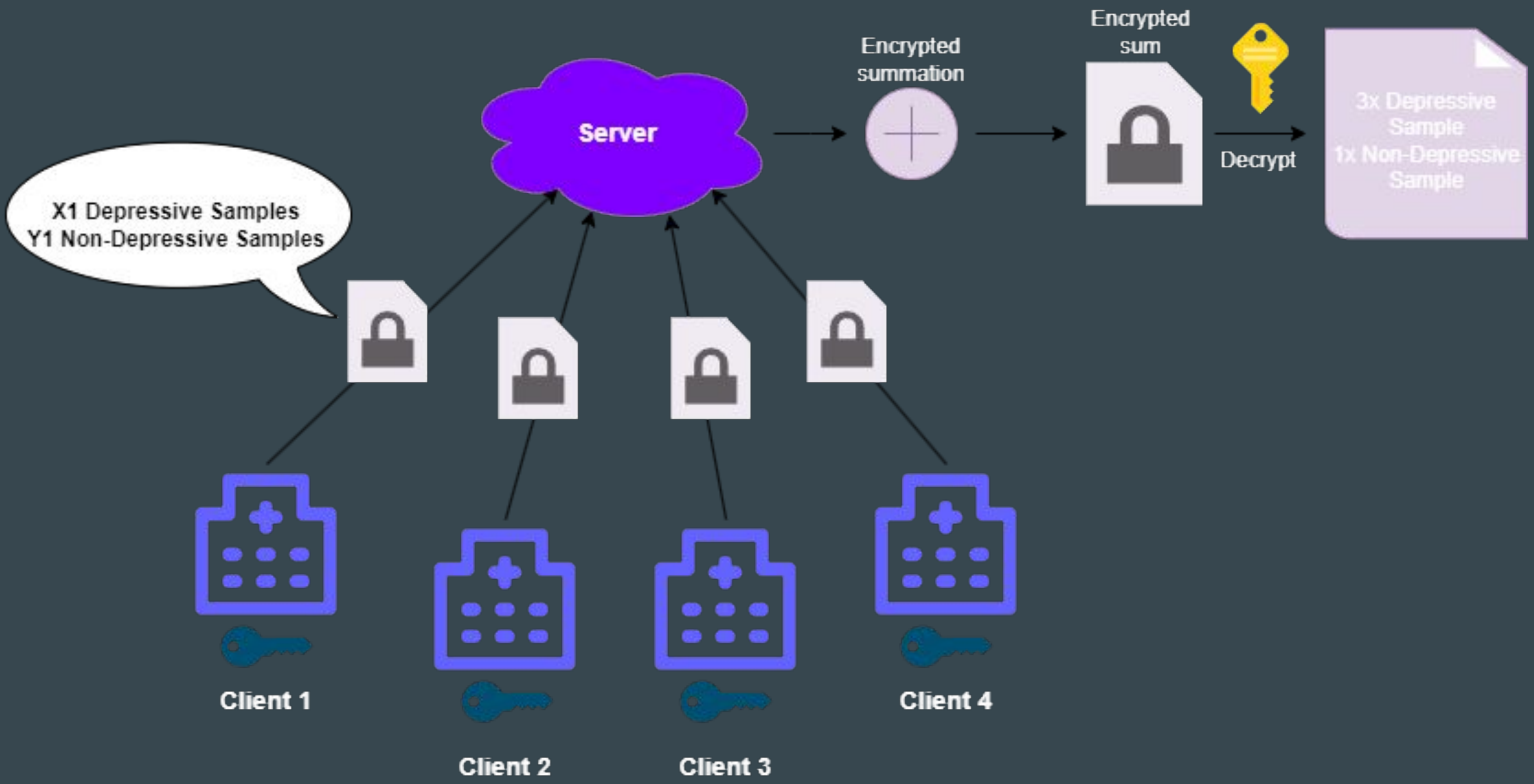
Federated Learning Methodology

- Data partitioning
- Secure Random Sampling
- FL environment setup

FL Methodology: Data partitioning



FL Methodology: Secure Random Sampling



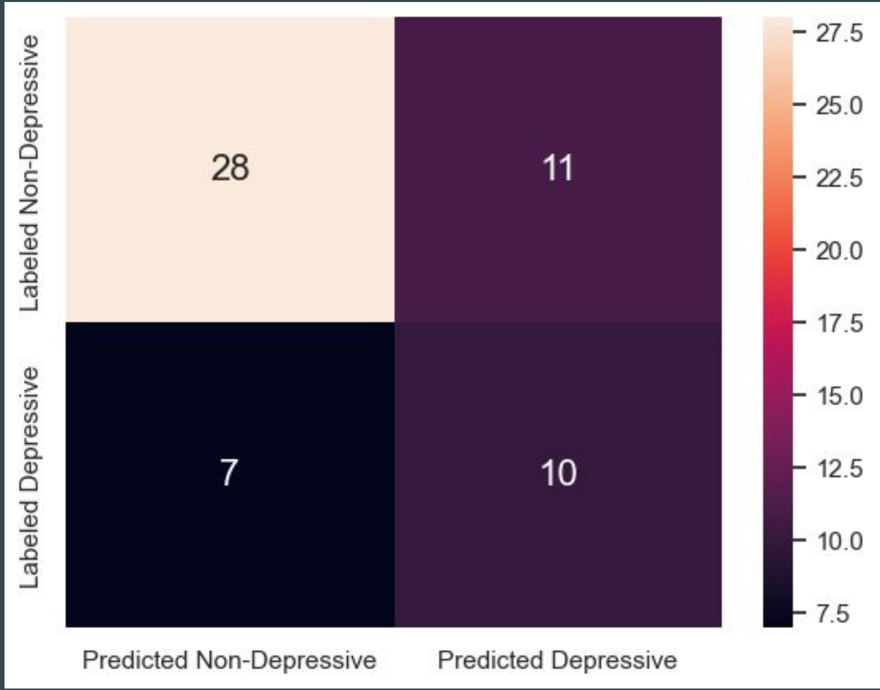
FL Methodology: Environment setup

- Used **Flower** framework with TensorFlow
- Centralized evaluation (server holds the validation data)
- Run X epochs, get best weights based on centralized validation loss
- Tuned learning rate and batch size based on number of clients

Evaluation

Evaluation: 4-client case

- Learning rate = 0.0005, Batch Size = 32
- All other parameters are same with centralized model

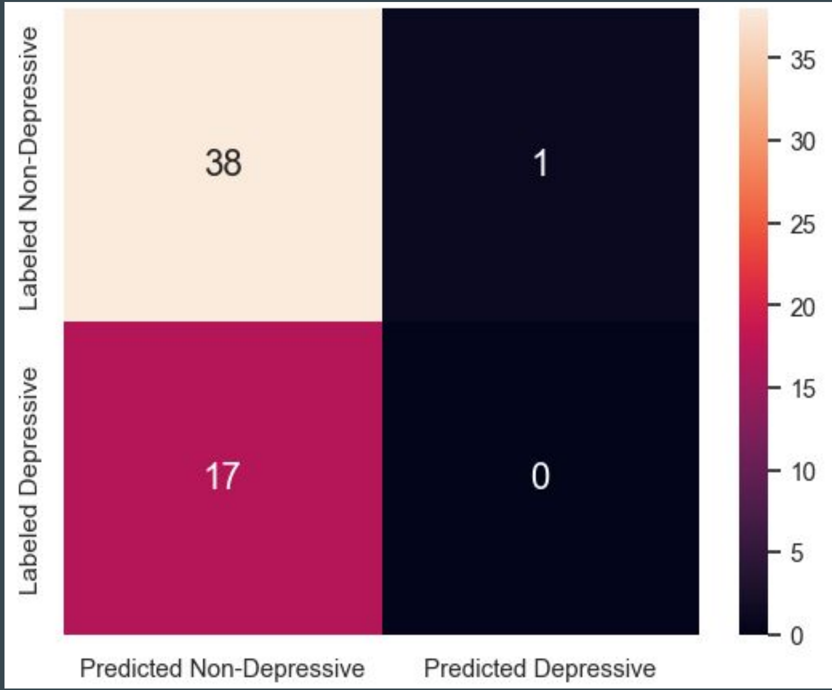


- F1 score: 0.53 (0.76)
- Precision: 0.48 (0.80)
- Recall: 0.59 (0.72)
- Accuracy: 0.68

Values in brackets represents the non-depressive class

Evaluation: 4-client with differential privacy

- Learning rate = 0.0005, Batch Size = 32
- VectorizedDPKerasSGDOptimizer is used from TensorFlow-privacy library
- Default parameters are used for DP



- F1 score: 0.0 (0.81)
- Precision: 0.0 (0.69)
- Recall: 0.0 (0.97)
- Accuracy: 0.68

Values in brackets represents the non-depressive class

Evaluation: Results comparison

Case	F1	Prec.	Rec.	Acc.
Centralized	.60 (.78)	.52 (.85)	.71 (.72)	.71
4-Client	.53 (.76)	.48 (.80)	.59 (.72)	.68
8-Client	.33 (.68)	.32 (.70)	.35 (.67)	.57
4-Client DP	.0 (.81)	.0 (.69)	.0 (.97)	.68

Values in brackets represents the non-depressive class

- 4-Client case is very promising
- Only 3% accuracy loss compared to centralized training

Future Work

- Implementing more defence techniques to FL environment
- Improving centralized model performance
- Using different data modalities (video, text)
- Evaluate different metrics (communication cost, privacy cost)

Conclusion

- Depression Assessment is an important topic
- Datasets are not enough for Deep Learning
- Using Federated Learning for depression assessment is applicable!

Thanks for listening.