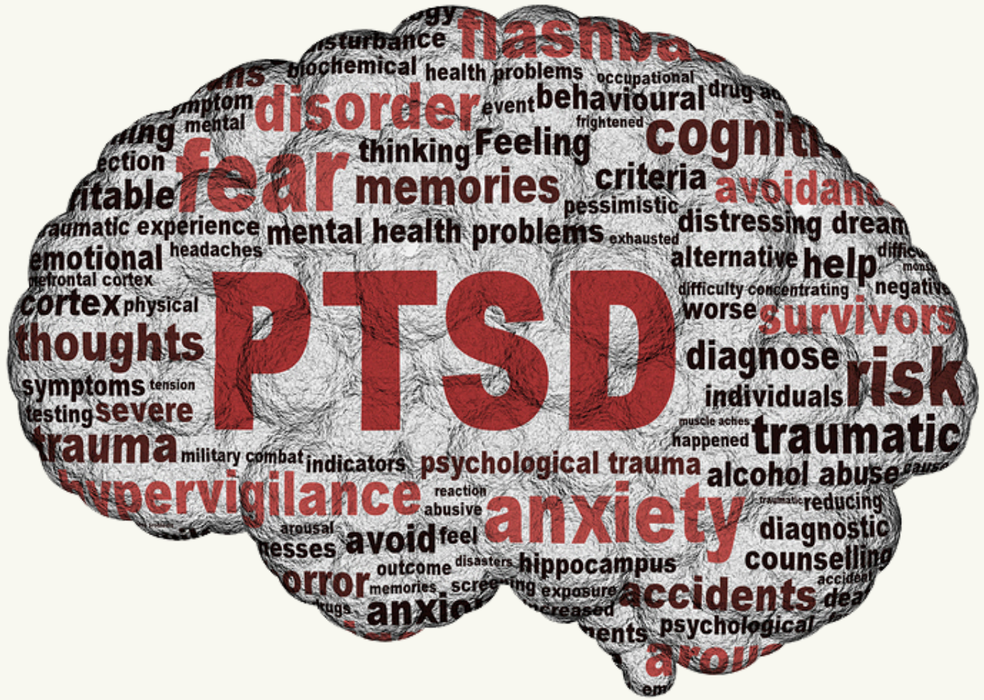
Team 40



MULTI MODAL APPROACH TO PTSD DETECTION

Presented by Aman, Arun, Sandeep

UNDERSTANDING PTSD

PTSD

(noun)

post traumatic stress disorder is a mental health condition that develops after experiencing or witnessing a traumatic, frightening, or life-threatening event

MOST COMMON SYMPTOMS ARE GROUPED INTO FOUR TYPES:

- INTRUSIVE MEMORIES
- AVOIDANCE
- NEGATIVE CHANGES IN THINKING/MOOD
- CHANGES IN PHYSICAL/EMOTIONAL REACTIONS

- https://www.who.int/news-room/fact-sheets/detail/post-traumatic-stress-disorder

PROBLEM RELEVANCE

Growing Concern

- AN ESTIMATED 3.9% OF THE WORLD POPULATION HAS HAD POST-TRAUMATIC STRESS DISORDER (PTSD) AT SOME STAGE IN THEIR LIVES. (KOENEN ET AL., 2017)
- 5% OF ADOLESCENTS AFFECTED, WITH RATES INCREASING FROM 3.7% (AGES 13-14) TO 7% (AGES 17-18). (YUAN ET AL., 2021)
- MILITARY VETERANS AND TRAUMA SURVIVORS SHOW SIGNIFICANTLY HIGHER RATES

Economic Burden

- TOTAL ECONOMIC BURDEN: $232.2 BILLION ANNUALLY (2018 DATA)
- POPULATION DISTRIBUTION:
  - CIVILIAN POPULATION: $189.5 BILLION (81.6%)
  - MILITARY POPULATION: $42.7 BILLION (18.4%)

- https://www.therecoveryvillage.com/mental-health/ptsd/ptsd-statistics/
- https://www.psychiatrist.com/jcp/economic-burden-posttraumatic-stress-disorder-united-states-societal-perspective/
- https://www.healio.com/news/psychiatry/20190424/speechbased-technologies-could-detect-ptsd-in-veterans

PROBLEM STATEMENT

Our objective is to develop an objective, multimodal approach that leverages the temporal fusion of behavioral markers to improve the accuracy of PTSD detection.

Project Aim: To create a robust machine learning framework that integrates audio, visual, and physiological signals with their temporal dynamics to objectively detect PTSD symptoms with clinical validity.

APPLICATIONS AND IMPACT

Objective screening tool for primary care settings.

Reduced diagnostic delays.

Early warning system for symptom escalation.

Potential reduction in the $232.2 billion annual economic burden (US).

- https://pmc.ncbi.nlm.nih.gov/articles/PMC11082170/
- https://www.psychiatrist.com/jcp/economic-burden-posttraumatic-stress-disorder-united-states-societal-perspective/

LITERATURE SURVEY

PAPER 1:

Speech-Based Markers for Posttraumatic Stress Disorder in U.S. Veterans

Methodology:

- Used audio recordings from clinical interviews with 52 male warzone-exposed veterans with PTSD and 77 controls.
- Extracted over 40,000 speech features.
- Built a classifier using Random Forest based on selected speech markers.

Observations:

- Found 18 key voice markers discriminating PTSD.
- Veterans with PTSD exhibited slower speech production, more monotonous speech, and flatter speech features compared to controls.

Performance Metric used:

- Area Under the ROC Curve (AUC) = 0.954.
- Overall correct classification rate = 89.1%.

Limitations:

- Only used speech data (unimodal).
- Focused on a specific veteran population (limited generalizability).
- Did not analyze temporal dynamics within speech sequences (used aggregated features for classification).

- https://pubmed.ncbi.nlm.nih.gov/31006959/

Summary statistics of 18 voice markers for the PTSD - and the PTSD + groups

	PTSD −			PTSD +			Wilcoxon Test * = p<.05
Variable	Mean	Median	Std Dev	Mean	Median	Std Dev	
Var1	−0.964	−0.973	0.030	−0.982	−0.987	0.024	*
var2	−0.937	−0.942	0.035	−0.965	−0.970	0.022	*
var3	0.936	0.945	0.053	0.967	0.972	0.021	*
var4	−0.065	−0.081	0.084	−0.039	−0.059	0.095	*
var5	409.400	240.187	557.587	1026.070	630.206	1154.760	*
var6	2.682	2.139	2.498	2.862	2.744	0.757	*
var7	−0.959	−0.967	0.038	−0.980	−0.983	0.014	*
var8	0.279	0.269	0.048	0.249	0.250	0.039	*
var9	−1.430	−1.364	0.336	−1.763	−1.657	0.632	*
var10	0.004	0.003	0.002	0.003	0.003	0.001	*
var11	−1.810	−1.883	0.463	−2.316	−2.271	1.120	*
var12	0.929	0.945	0.074	0.940	0.970	0.196	*
var13	355.616	208.173	364.526	897.390	605.526	769.039	*
var14	12.838	12.131	4.279	17.006	15.581	5.937	*
var15	0.00024	0.00018	0.00017	0.00018	0.00015	0.00019	*
var16	0.035	0.164	1.617	−0.131	−0.208	1.883	*
var17	0.0040	0.0037	0.0015	0.0032	0.0031	0.0010	*
var18	0.170	0.169	0.012	0.171	0.169	0.006	

Marmar CR, Weiss DS, Schlenger WE, Fairbank JA, Jordan BK, Kulka RA, Hough RL. Peritraumatic dissociation and posttraumatic stress in male Vietnam theater veterans. Am J Psychiatry. 1994 Jun;151(6):902-7. doi: 10.1176/ajp.151.6.902. PMID: 8185001.

Methodology:

- Explored multimodal PTSD prediction by combining:
  - Neurophysiological signals (EEG, ECG, GSR).
  - Head motion.
  - Speech data.
- Used various stimuli, including trauma-specific content, to elicit responses.

Observations:

- Multimodal data systematically improved prediction performance.
- Trauma-specific stimuli (image + audio) were most effective for discriminating PTSD from controls.

Performance Metric used:

Focused on comparative improvement with added modalities rather than a single accuracy figure.

Limitations:

- Required specialized and potentially invasive sensors (e.g., EEG, ECG), limiting scalability and ease of use.
- Use of trauma-specific stimuli may not reflect naturalistic symptom presentation and could be distressing for participants.

- https://ieeexplore.ieee.org/document/6854279

PAPER 2:

MULTI-MODAL PREDICTION OF PTSD AND STRESS INDICATORS

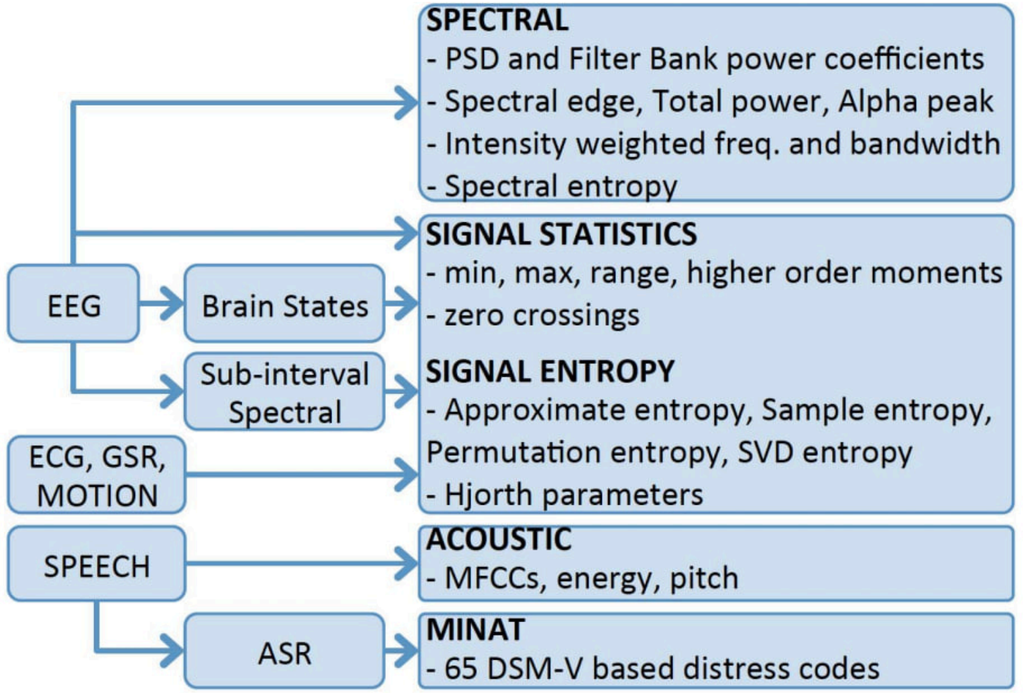


Fig. 2: *Overview of the modalities and features used*

Marmar CR, Weiss DS, Schlenger WE, Fairbank JA, Jordan BK, Kulka RA, Hough RL. Peritraumatic dissociation and posttraumatic stress in male Vietnam theater veterans. Am J Psychiatry. 1994 Jun;151(6):902-7. doi: 10.1176/ajp.151.6.902. PMID: 8185001.

PAPER 3:

Multimodal Approach to Identifying Malingered Posttraumatic Stress Disorder: A Review

Methodology:

- Focused on differentiating true PTSD from malingered PTSD. Proposed a multimodal assessment framework combining:
  - Careful clinical interview techniques (avoiding leading questions).
  - Collateral data (police reports, military records, employment files).
  - Psychometric tests (e.g., MMPI-2, SIRS, M-FAST).
  - Physiologic testing (e.g., heart rate response to sudden loud tones).

Observations:

- PTSD diagnosis is vulnerable to malingering, especially with financial incentives (e.g., VA compensation).
- High rates of symptom exaggeration reported (up to 75% in some VA PTSD claimants).
- DSM-5 criteria critiqued for poor malingering detection.

Performance Metric used:

- Framework/review-oriented; discusses sensitivity/specificity of psychometric tests.
- Eg: M-FAST reported 92% accuracy for malingering in war PTSD populations.

Limitations:

- Diagnostic methods reviewed are subjective and vulnerable to feigning.
- Proposed solution is complex, requiring extensive data gathering beyond simple clinical interaction.

- https://pmc.ncbi.nlm.nih.gov/articles/PMC4382135/

TABLE 5. Suggested interview techniques that may identify malingering

Use open-ended questions. Avoid leading questions.
Use an empathic interview style (e.g., highlight the temptation to exaggerate).
Be objective. Refrain from showing suspicion/skepticism.
Obtain detailed history of symptoms.
Elicit the patient's capacity to work vs. enjoy recreational activities.
Elicit the course of patient's illness.
Elicit the patient's premorbid functioning.
Conduct a mental status exam. Look for hypervigilance, concentration deficits, irritability, and avoidance.
Interview the patient separately from third parties.

Marmar CR, Weiss DS, Schlenger WE, Fairbank JA, Jordan BK, Kulka RA, Hough RL. Peritraumatic dissociation and posttraumatic stress in male Vietnam theater veterans. Am J Psychiatry. 1994 Jun;151(6):902-7. doi: 10.1176/ajp.151.6.902. PMID: 8185001.

HOW WE ADDRESS IT?

Our Novel Approach to PTSD Detection

- **Limitations of Current Methods:**
  - Single modality approaches miss critical behavioral markers
  - Many require specialized or invasive equipment
  - Limited temporal analysis of symptom manifestation
- **Our Solution: Temporal Fusion Architecture**
  - Combines speech and facial features from naturalistic clinical interviews
  - Non-invasive data collection suitable for telehealth applications
  - Captures subtle interplay between vocal and facial expressions
- **Multimodal Feature Integration:**
  - Speech: openSMILE eGeMAPS vocal biomarkers, DenseNet-derived deep audio representations
  - Facial: OpenFace-derived action units, pose, gaze features, VGG and ResNet-derived deep visual representations

HOW WE ADDRESS IT?

APPROACH 1

- A dual-branch LSTM network with cross-modal attention is used to model how PTSD symptoms appear over time, capturing links between vocal stress and delayed facial micro-expressions.
- This method improves diagnostic accuracy and clinical interpretability by revealing specific multimodal patterns linked to PTSD.

APPROACH 2

- Transformer models are effective at capturing global dependencies and long-range interactions within sequences.
- We applied separate Transformer models to each modality (speech and visual) before late fusion, allowing each to learn complex patterns specific to its domain without early cross-modal influence.
- Unlike BiLSTM, which processes data sequentially, Transformers process all sequence elements in parallel, enabling faster training and the ability to identify diverse temporal patterns.
- This approach can reveal key moments relevant to PTSD indication by allowing each modality's Transformer to focus on its unique contextual features.

DATASET AND FEATURES PREPROCESSING

E-DAIC WOZ DATASET OVERVIEW

Dataset Selection Rationale:

- E-DAIC WOZ dataset - gold standard for multimodal mental health assessment.
- Specifically designed for computational assessment of psychological distress, including PTSD.
- Collected in controlled laboratory settings with ethical oversight.
- Includes audio, video, and text transcriptions of clinical interviews.

Dataset Statistics:

- 275 participants (163 training, 56 test, 56 dev).
- PTSD prevalence: 30% of participants.
- 30-60 minute semi-structured interviews per participant.
- Interview conducted by virtual interviewer (Ellie) operated by Wizard-of-Oz protocol.
- Total of 200 hours of multimodal recordings.

Ethical Considerations:

- IRB approval from University of Southern California.
- Informed consent for recording and research use.
- De-identification protocols applied.
- Protected health information safeguarded.

E-DAIC WOZ DATASET FEATURES

Feature Set	Modality	Feature Type	Description
Bag-of-audio-words eGeMAPS (Schmitt et al, 2017)	Audio	Bag-of-words	eGeMAPS features processed and summarized over a block of 4-second length duration for each step of 1 second
Bag-of-audio-words MFCCs (Schmitt et al, 2017)	Audio	Bag-of-words	MFCC features processed and summarized over a block of 4-second length duration for each step of 1 second
Bag-of-visual-words Pose Gaze AUs (Schmitt et al, 2017)	Visual	Bag-of-words	Pose/Gaze/AU features processed and summarized over a block of 4-second length duration for each step of 1 second
CNN ResNet (He at al, 2016)	Visual	Deep Representations	Aligned face images are fed to the pretrained ResNet-50 model with frozen weights, and the output of the first FC layer is extracted as representation.
CNN VGG (Simonyan et al, 2014)	Visual	Deep Representations	Aligned face images are fed to the pretrained VGG-16 model with frozen weights, and the output of the global average pooling layer is extracted as representation.
Densenet (Huang et al, 2017)	Audio	Deep Representations	The speech files are first transformed into mel-spectrogram images with 128 mel-frequency bands, a window width of 4 seconds and a hop size of 1 second. The spectral-based images are fed to the densenet 201 pretrained network, and a feature vector is obtained from activations of the last average pooling layer of DenseNet.
OpenFace - Pose, Gaze, AUs (Baltrusaitis et al, 2018)	Visual	Expert Knowledge	The intensities of 17 FAUs for each video frame, along with a confidence measure are extracted using OpenFace
extended Geneva Minimalistic Acoustic Parameter Set (eGeMaPS) (Eyben et al, 2016)	Audio	Expert Knowledge	Contains 88 measures covering spectral, cepstral, prosodic, and voice quality information
MFCCs (Eyben et al, 2013)	Audio	Expert Knowledge	MFCCs 1-13, including their first and second order derivatives (deltas and double-deltas) are computed as a set of acoustic LLDs, using the OpenSMILE toolkit
VGG-16 (Simonyan et al, 2014)	Audio	Deep Representations	The speech files are first transformed into mel-spectrogram images with 128 mel-frequency bands, a window width of 4 seconds and a hop size of 1 second. The spectral-based images are fed to the densenet 201 pretrained network, and a feature vector is obtained from activations of the second fully connected layer in VGG16.

FEATURE SELECTION

Audio Features Used:

- ·OpenSMILE2.3.0_egemaps.csv: Selected key eGeMAPS vocal biomarkers (pitchrelated: F0, voicing; voice quality: jitter, shimmer, HNR; spectral: balance, slope, dynamics). List a few example feature names from your notebook.
- densenet201.csv: First 100 features from DenseNet (deep audio representation).

Visual Features Used:

- OpenFace2.1.0_Pose_gaze_AUs.csv: Selected Action Units (intensities _r for AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45), selected pose features (pose_Tx, pose_Ty, pose_Tz, pose_Rx, pose_Ry, pose_Rz), and selected gaze features (gaze_0_x, gaze_0_y, etc.). List a few example feature names.
- CNN_VGG.mat (converted to CSV/array): First 100 features from VGG (deep visual representation).

PREPROCESSING

- **Feature Loading**: Loaded selected CSV and MAT files for each participant from E-DAIC.
- **Normalization**: Z-score standardization (StandardScaler) applied to all selected numerical features.
- **Missing Data Handling**: Mean imputation (fillna(dataset_mean) for each feature column) used for any gaps.
- **Sequence Creation:**
  - Fixed-length temporal windows of 20 time steps.
  - Stride of 5 time steps used to create overlapping sequences.
  - Audio and visual feature sequences were aligned to the minimum length of the two for each participant to ensure consistent input dimensions for each sequence.
- **Data Augmentation** (to address class imbalance in training):
  - 2x augmentation of PTSD positive samples in the training set.
  - Added 14,240 augmented positive samples.
  - Final training set: 45,060 samples (target class: 47% positive).

ML METHODOLOGY & IMPLEMENTATION

BASELINE MODELS

MULTIMODAL DATASET

```
--- Training Random Forest (Combined) ---
Accuracy: 0.5455
Precision: 0.4545
Recall: 0.2083
F1 Score: 0.2857
AUC-ROC: 0.4382

Confusion Matrix:
[[25  6]
 [19  5]]

Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.81      0.67        31
           1       0.45      0.21      0.29        24

    accuracy                           0.55        55
   macro avg       0.51      0.51      0.48        55
weighted avg       0.52      0.55      0.50        55

...
weighted avg       0.32      0.56      0.41        55
```

```
--- Training SVM (Combined) ---
Accuracy: 0.5636
Precision: 0.0000
Recall: 0.0000
F1 Score: 0.0000
AUC-ROC: 0.3616

Confusion Matrix:
[[31  0]
 [24  0]]

Classification Report:
              precision    recall  f1-score   support

           0       0.56      1.00      0.72        31
           1       0.00      0.00      0.00        24

    accuracy                           0.56        55
   macro avg       0.28      0.50      0.36        55
weighted avg       0.32      0.56      0.41        55
```

```
--- Training Logistic Regression (Combined) -
Accuracy: 0.3455
Precision: 0.2500
Recall: 0.2500
F1 Score: 0.2500
AUC-ROC: 0.3199

Confusion Matrix:
[[13 18]
 [18  6]]

Classification Report:
              precision    recall  f1-score   support

           0       0.42      0.42      0.42        31
           1       0.25      0.25      0.25        24

    accuracy                           0.35        55
   macro avg       0.33      0.33      0.33        55
weighted avg       0.35      0.35      0.35        55
```

BASELINE MODELS

AUDIO ONLY DATASET

```
--- Training Random Forest (Audio Only) ---
Accuracy: 0.4727
Precision: 0.3333
Recall: 0.2083
F1 Score: 0.2564
AUC-ROC: 0.4718

Confusion Matrix:
[[21 10]
 [19  5]]

Classification Report:
              precision    recall  f1-score   support

           0       0.53      0.68      0.59        31
           1       0.33      0.21      0.26        24

    accuracy                           0.47        55
   macro avg       0.43      0.44      0.42        55
weighted avg       0.44      0.47      0.45        55
```

```
--- Training SVM (Audio Only) ---
Accuracy: 0.5818
Precision: 0.5714
Recall: 0.1667
F1 Score: 0.2581
AUC-ROC: 0.4409

Confusion Matrix:
[[28  3]
 [20  4]]

Classification Report:
              precision    recall  f1-score   support

           0       0.58      0.90      0.71        31
           1       0.57      0.17      0.26        24

    accuracy                           0.58        55
   macro avg       0.58      0.53      0.48        55
weighted avg       0.58      0.58      0.51        55
```

```
--- Training Logistic Regression (Audio Only) ---
 Accuracy: 0.4364
 Precision: 0.3793
 Recall: 0.4583
 F1 Score: 0.4151
 AUC-ROC: 0.4126

 Confusion Matrix:
 [[13 18]
  [13 11]]

 Classification Report:
               precision    recall  f1-score   support

            0       0.50      0.42      0.46        31
            1       0.38      0.46      0.42        24

     accuracy                           0.44        55
    macro avg       0.44      0.44      0.44        55
 weighted avg       0.45      0.44      0.44        55
```

BASELINE MODELS

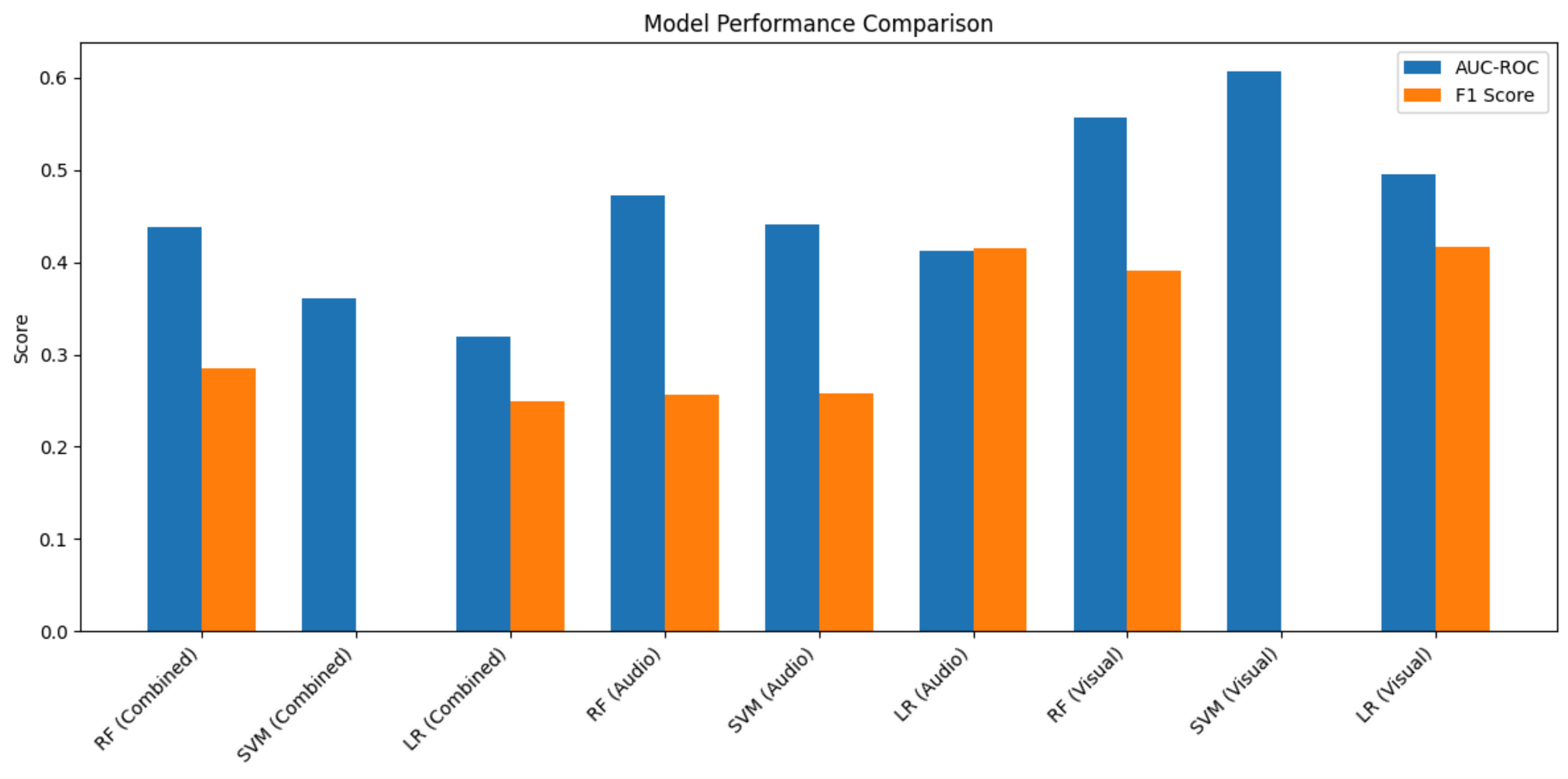
VISUAL ONLY DATASET

```
--- Training Random Forest (Visual Only) ---
Accuracy: 0.4909
Precision: 0.4091
Recall: 0.3750
F1 Score: 0.3913
AUC-ROC: 0.5565

Confusion Matrix:
[[18 13]
 [15  9]]

Classification Report:
              precision    recall  f1-score   support

           0       0.55      0.58      0.56        31
           1       0.41      0.38      0.39        24

    accuracy                           0.49        55
   macro avg       0.48      0.48      0.48        55
weighted avg       0.49      0.49      0.49        55
```

```
--- Training SVM (Visual Only) ---
Accuracy: 0.5636
Precision: 0.0000
Recall: 0.0000
F1 Score: 0.0000
AUC-ROC: 0.6075

Confusion Matrix:
[[31  0]
 [24  0]]

Classification Report:
              precision    recall  f1-score   support

           0       0.56      1.00      0.72        31
           1       0.00      0.00      0.00        24

    accuracy                           0.56        55
   macro avg       0.28      0.50      0.36        55
weighted avg       0.32      0.56      0.41        55
```

```
--- Training Logistic Regression (Visual Only) ---
Accuracy: 0.4909
Precision: 0.4167
Recall: 0.4167
F1 Score: 0.4167
AUC-ROC: 0.4960

Confusion Matrix:
[[17 14]
 [14 10]]

Classification Report:
              precision    recall  f1-score   support

           0       0.55      0.55      0.55        31
           1       0.42      0.42      0.42        24

    accuracy                           0.49        55
   macro avg       0.48      0.48      0.48        55
weighted avg       0.49      0.49      0.49        55
```

BASELINE MODEL COMAPRISON



FINAL APPROACH

1. BI LSTM
2. TRANSFORMER

BI-LSTM MODEL ARCHITECTURE

Feature Selection | Feature Preprocessing | Model Selection | Feature Fusion | Output label

Audio Features

'unknown'	0	1.414973	0	0	0
'unknown'	0.1	1.414973	0	0	0
'unknown'	0.2	1.414973	0	0	0
'unknown'	0.3	1.414973	0	0	0
'unknown'	0.4	1.414973	0	0	0
'unknown'	0.5	1.414973	0	0	0

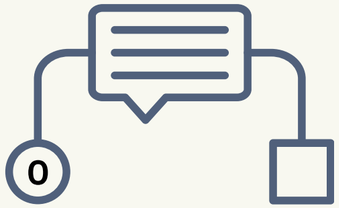
eGeMAPS

Bidirectional LSTM with Attention

Action Units, pose, gaze - OpenFace

Bidirectional LSTM with Attention

Feature Fusion & Classification



PTSD Prediction (0/1)

1

Visual Features

ne	timestamp	confidence	success	pose_Tx	pose_Ty	pose_Tz	pose_Rx	pose_Ry	pose_Rz	gaze_0_x	gaze_0_y	g
1	0	0.98	1	69.5	37.7	576.8	0.221	0.036	-0.068	0.015324	0.298824	
2	0.033	0.98	1	69.4	37.4	577.5	0.219	0.037	-0.067	0.006174	0.29452	
3	0.067	0.98	1	69.4	37.3	576.7	0.223	0.039	-0.069	0.005192	0.298043	
4	0.1	0.98	1	69.3	37.3	576.5	0.225	0.039	-0.069	0.005523	0.299662	
5	0.133	0.98	1	69.4	37.3	576.6	0.225	0.039	-0.07	0.005461	0.300506	
6	0.167	0.98	1	67.7	38	573.6	0.235	0.041	-0.07	0.003653	0.328589	
7	0.2	0.98	1	66.5	38.9	575.2	0.24	0.043	-0.068	-0.0003	0.333997	
8	0.233	0.98	1	65.3	39.4	578	0.251	0.047	-0.069	-0.00738	0.328653	
9	0.267	0.98	1	64.1	39.3	577.4	0.247	0.046	-0.07	0.018701	0.338622	

BI LSTM RESULTS

Applied on Training and Dev set

```
EPOCH 9/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9978 - AUC: 0.9997 - LOSS: 0.0071 - PRECISION: 0.9925 - RECALL: 0.9980 - VAL_ACCURACY: 0.4003 - VAL_AUC: 0.4217 - VAL_LOSS: 5.0247 - VAL_PRECISION: 0.4207 - VAL_RECALL: 0.1220 - LEARNING_RATE: 5.0000E-04
EPOCH 10/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9984 - AUC: 0.9996 - LOSS: 0.0064 - PRECISION: 0.9956 - RECALL: 0.9977 - VAL_ACCURACY: 0.4020 - VAL_AUC: 0.4250 - VAL_LOSS: 5.5418 - VAL_PRECISION: 0.4259 - VAL_RECALL: 0.1236 - LEARNING_RATE: 5.0000E-04
EPOCH 11/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9984 - AUC: 0.9999 - LOSS: 0.0050 - PRECISION: 0.9956 - RECALL: 0.9977 - VAL_ACCURACY: 0.3959 - VAL_AUC: 0.4209 - VAL_LOSS: 6.2504 - VAL_PRECISION: 0.4101 - VAL_RECALL: 0.1225 - LEARNING_RATE: 5.0000E-04
EPOCH 12/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9985 - AUC: 0.9996 - LOSS: 0.0055 - PRECISION: 0.9959 - RECALL: 0.9977 - VAL_ACCURACY: 0.3950 - VAL_AUC: 0.4250 - VAL_LOSS: 6.4345 - VAL_PRECISION: 0.4061 - VAL_RECALL: 0.1193 - LEARNING_RATE: 2.5000E-04
EPOCH 13/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9996 - AUC: 0.9998 - LOSS: 0.0023 - PRECISION: 0.9987 - RECALL: 0.9995 - VAL_ACCURACY: 0.3948 - VAL_AUC: 0.4291 - VAL_LOSS: 6.8340 - VAL_PRECISION: 0.4033 - VAL_RECALL: 0.1159 - LEARNING_RATE: 2.5000E-04
EPOCH 14/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9997 - AUC: 1.0000 - LOSS: 8.7206E-04 - PRECISION: 0.9989 - RECALL: 1.0000 - VAL_ACCURACY: 0.3933 - VAL_AUC: 0.4230 - VAL_LOSS: 6.8085 - VAL_PRECISION: 0.4036 - VAL_RECALL: 0.1218 - LEARNING_RATE: 2.5000E-04
EPOCH 15/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9999 - AUC: 1.0000 - LOSS: 5.1027E-04 - PRECISION: 0.9997 - RECALL: 0.9998 - VAL_ACCURACY: 0.3928 - VAL_AUC: 0.4252 - VAL_LOSS: 7.4722 - VAL_PRECISION: 0.3978 - VAL_RECALL: 0.1148 - LEARNING_RATE: 2.5000E-04
EPOCH 16/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9996 - AUC: 1.0000 - LOSS: 0.0014 - PRECISION: 0.9990 - RECALL: 0.9991 - VAL_ACCURACY: 0.3969 - VAL_AUC: 0.4240 - VAL_LOSS: 6.7441 - VAL_PRECISION: 0.4167 - VAL_RECALL: 0.1299 - LEARNING_RATE: 2.5000E-04
EPOCH 17/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 0.9997 - AUC: 0.9999 - LOSS: 0.0017 - PRECISION: 0.9992 - RECALL: 0.9995 - VAL_ACCURACY: 0.3986 - VAL_AUC: 0.4273 - VAL_LOSS: 6.9812 - VAL_PRECISION: 0.4164 - VAL_RECALL: 0.1218 - LEARNING_RATE: 1.2500E-04
EPOCH 18/50
964/964 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6S 7MS/STEP - ACCURACY: 1.0000 - AUC: 1.0000 - LOSS: 1.1385E-04 - PRECISION: 1.0000 - RECALL: 1.0000 - VAL_ACCURACY: 0.3994 - VAL_AUC: 0.4296 - VAL_LOSS: 7.3125 - VAL_PRECISION: 0.4183 - VAL_RECALL: 0.1218 - LEARNING_RATE: 1.2500E-04
```

INTERPRETATION OF BI LSTM OUTPUT

Problem

- The large gap between training and test performance indicates the model has memorized the training data rather than learning generalizable patterns. So there was Severe Overfitting
- There was Poor precision on the test data. Only 12.82% of predicted PTSD cases are actual positives, meaning almost 7 out of 8 predictions are false alarms.
- Class imbalance was a major issue. Our training data contained only about 23% PTSD-positive samples, and the test set was even more skewed at 13%.

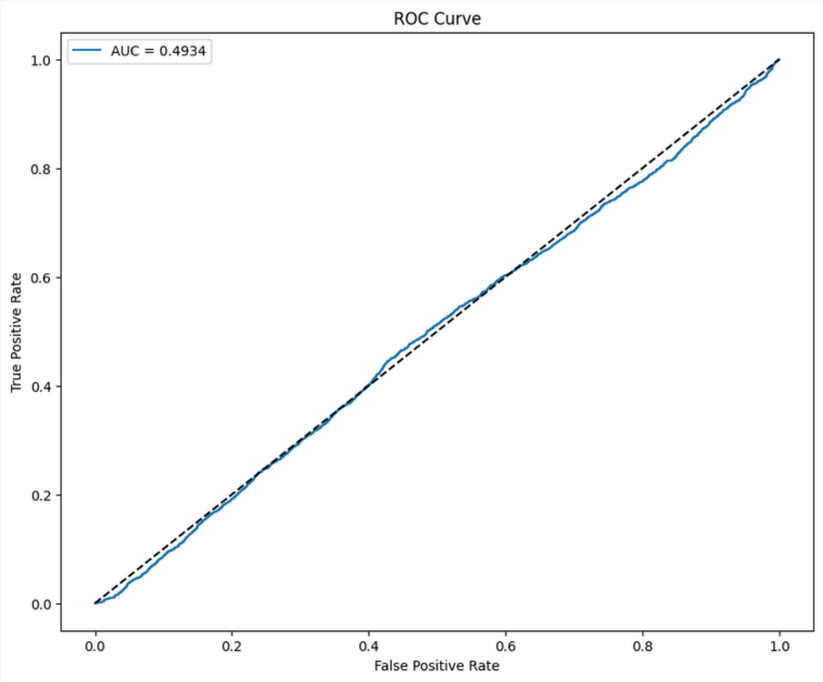
Steps to Mitigate

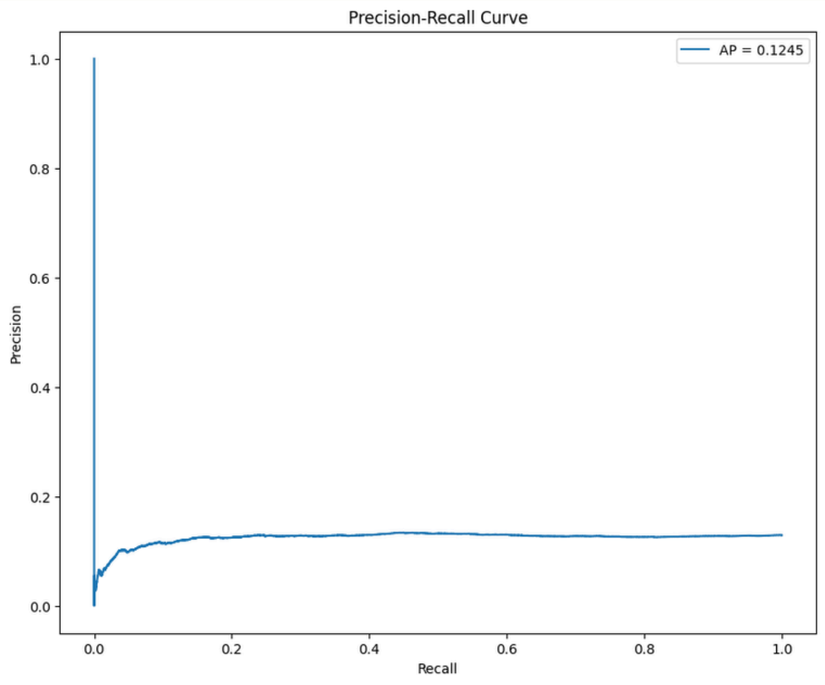
- Our first priority was tackling this overfitting. We implemented several techniques:
- Added Dropout Layers: We initially added dropout at 0.2, but quickly realized this wasn't enough and increased to 0.3 throughout the network.
- Batch Normalization: We added batch normalization after each LSTM layer to standardize the activations and improve gradient flow.
- Reduced Model Complexity: Our initial model had 128 LSTM units per layer, which we reduced to 64 to prevent the model from having too much capacity to memorize the training data.
- Early Stopping: We implemented early stopping based on validation AUC with a patience of 10 epochs.

RESULTS AFTER MITIGATION

```
Using batch size of 32 for memory efficiency
Processing data with GPU memory constraints and data augmentation...
Applying data augmentation to balance classes...
Class distribution before augmentation: 7120 positive, 23700 negative
Augmentation factor: 2x
Added 14240 augmented positive samples
Class distribution after augmentation: 21360/45060 positive samples

Number of training samples after augmentation: 45060
Number of validation samples: 17442
Using class weights: {0: np.float64(0.950632911392405), 1: np.float64(1.0547752808988764)}

Starting training with TensorFlow Dataset API (memory-efficient)...
Epoch 1/30
I0000 00:00:1746283082.833519   49741 cuda_dnn.cc:529] Loaded cuDNN version 90300
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.8788 - auc: 0.7445 - loss: 0.2502 - precision: 0.6757 - recall: 0.3479
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. W
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 13s 7ms/step - accuracy: 0.8788 - auc: 0.7446 - loss: 0.2501 - precision: 0.6759 - recall: 0.3482 - val_accuracy: 0.4416
Epoch 2/30
1408/1409 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.9450 - auc: 0.8282 - loss: 0.0400 - precision: 0.7505 - recall: 0.6711
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. W
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9450 - auc: 0.8284 - loss: 0.0399 - precision: 0.7509 - recall: 0.6715 - val_accuracy: 0.4649
Epoch 3/30
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9629 - auc: 0.8423 - loss: 0.0266 - precision: 0.7755 - recall: 0.7214 - val_accuracy: 0.4413
Epoch 4/30
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9684 - auc: 0.8460 - loss: 0.0212 - precision: 0.7818 - recall: 0.7280 - val_accuracy: 0.4479
Epoch 5/30
1402/1409 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - accuracy: 0.9735 - auc: 0.8473 - loss: 0.0203 - precision: 0.7956 - recall: 0.7561
Epoch 5: ReduceLROnPlateau reducing learning rate to 0.00010000000474974513.
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9735 - auc: 0.8482 - loss: 0.0203 - precision: 0.7967 - recall: 0.7572 - val_accuracy: 0.4412
Epoch 6/30
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9781 - auc: 0.8498 - loss: 0.0198 - precision: 0.7933 - recall: 0.7947 - val_accuracy: 0.4379
Epoch 7/30
1409/1409 ━━━━━━━━━━━━━━━━━━━━ 9s 6ms/step - accuracy: 0.9891 - auc: 0.8532 - loss: 0.0122 - precision: 0.8232 - recall: 0.8173 - val_accuracy: 0.4321
Epoch 7: early stopping
Restoring model weights from the end of the best epoch: 2.
```

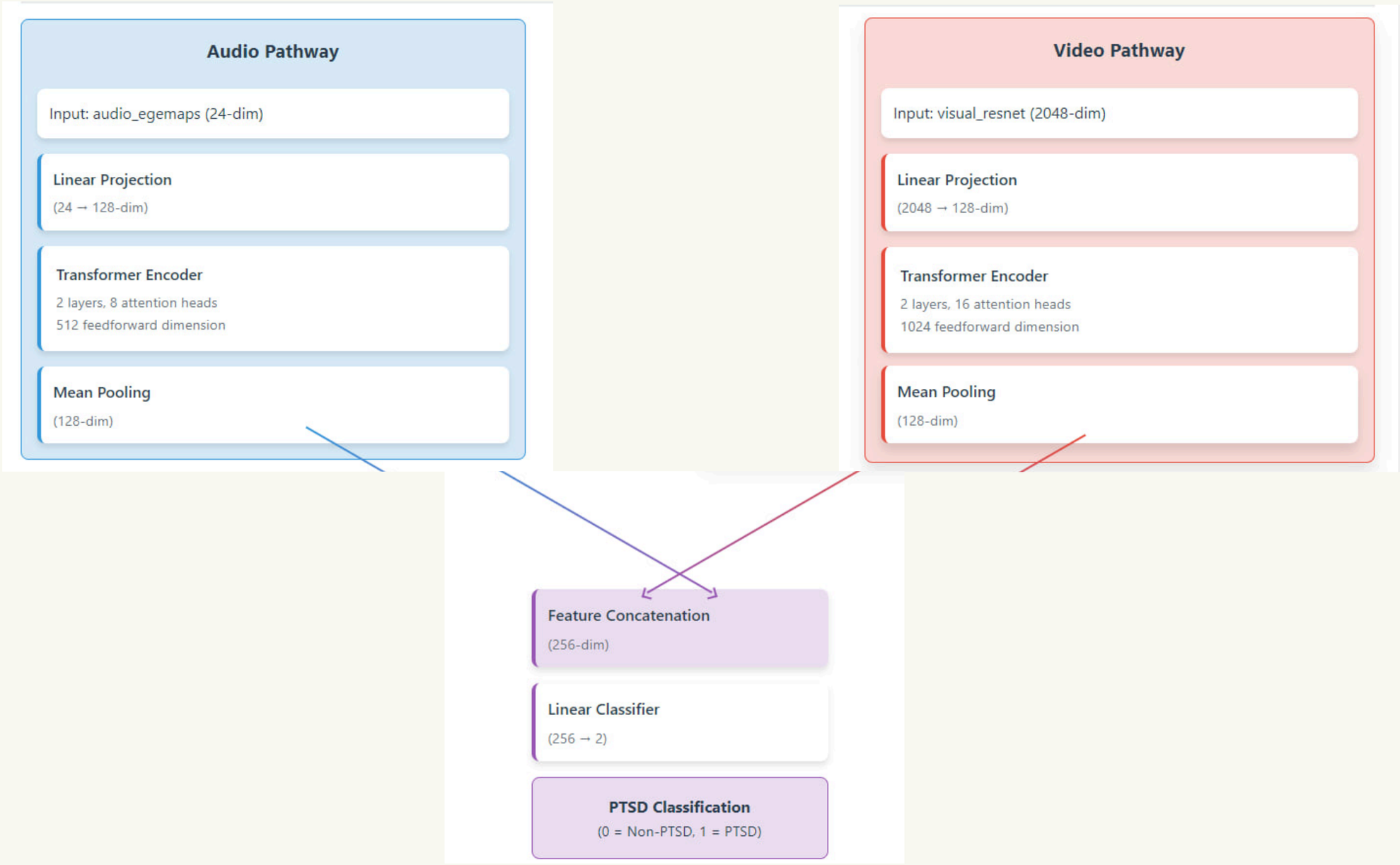
TEST SET RESULTS





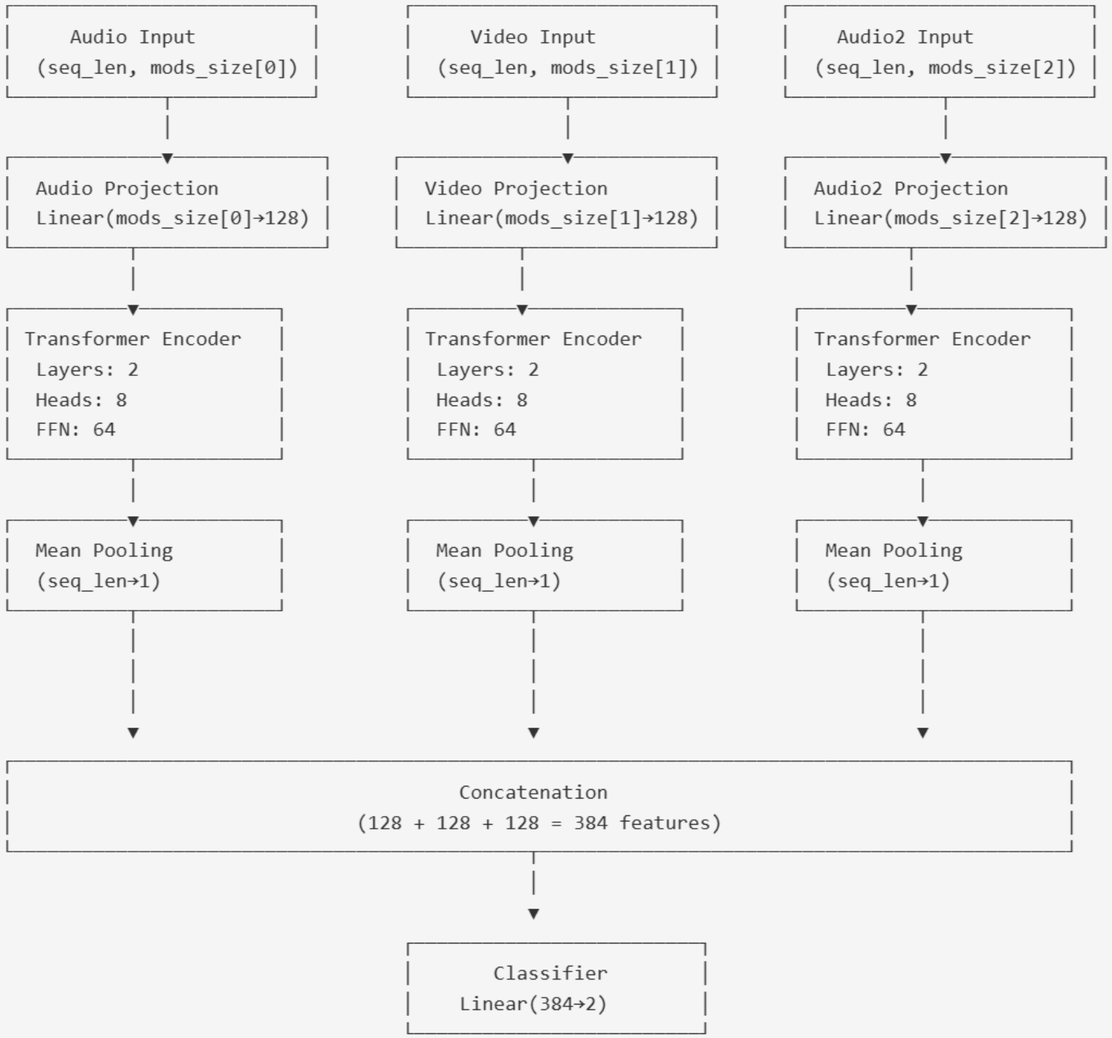

```
Evaluating on 19772 test samples
618/618 ━━━━━━━━━━━━━━━━━━━━ 1s 2ms/step

Test Set Evaluation:
Accuracy: 0.6469
Precision: 0.1282
Recall: 0.2987
F1 Score: 0.1794
Confusion Matrix:
[[12028  5190]
 [ 1791   763]]
```

TRANSFORMER ARCHITECTURE



TRANSFORMER ARCHITECTURE

Model architecture



TRANSFORMER RESULTS

Baseline:

```
Epoch 3/10
Train - Loss: 0.7218, Acc: 0.7362, F1: 0.2182, Recall: 0.1224, AUC: 0.7680
Val   - Loss: 0.8106, Acc: 0.6909, F1: 0.1905, Recall: 0.1176, AUC: 0.7260
batch idx: 0 done; train loss: 0.016976982355117798; avg train loss: 0.0169769823
```

Final:

```
Current learning rate: 0.0005

Epoch 1/10
Train - Loss: 0.7379, Acc: 0.5951, F1: 0.6887, Recall: 0.8588, AUC: 0.6605
Val   - Loss: 0.9711, Acc: 0.4727, F1: 0.5085, Recall: 0.8824, AUC: 0.6037
Epoch 2/10
Train - Loss: 1.2441, Acc: 0.5337, F1: 0.6885, Recall: 1.0000, AUC: 0.6109
Val   - Loss: 1.8210, Acc: 0.3273, F1: 0.4789, Recall: 1.0000, AUC: 0.5588
Epoch 3/10
Train - Loss: 1.1062, Acc: 0.6748, F1: 0.7415, Recall: 0.9048, AUC: 0.7577
Val   - Loss: 1.7563, Acc: 0.5273, F1: 0.5357, Recall: 0.8824, AUC: 0.6486
Epoch 4/10
Train - Loss: 0.8939, Acc: 0.6012, F1: 0.6701, Recall: 0.8250, AUC: 0.6527
Val   - Loss: 1.1795, Acc: 0.4727, F1: 0.5246, Recall: 0.9412, AUC: 0.7523
Epoch 5/10
Train - Loss: 0.6549, Acc: 0.6933, F1: 0.7126, Recall: 0.8052, AUC: 0.7535
Val   - Loss: 1.1115, Acc: 0.4909, F1: 0.5172, Recall: 0.8824, AUC: 0.7430
```

INTERPRETATION OF RESULTS

Handling Class Imbalance

- Used WeightedRandomSampler to address class imbalance in PTSD detection.
- The sampler assigns higher weights to the minority class (PTSD) during training.
- Implementation example:
  - Calculate class weights:
  - weights_per_class = 1.0 / class_counts.float()
  - Assign sample weights:
  - sample_weights = [weights_per_class[label] for label in all_labels]

Results Obtained

- Training accuracy: 74%
- Validation accuracy: 71%
- Validation F1 score: 0.20
- Validation AUC: 0.73
- Validation recall (PTSD class): 88%

FINAL CONCLUSION

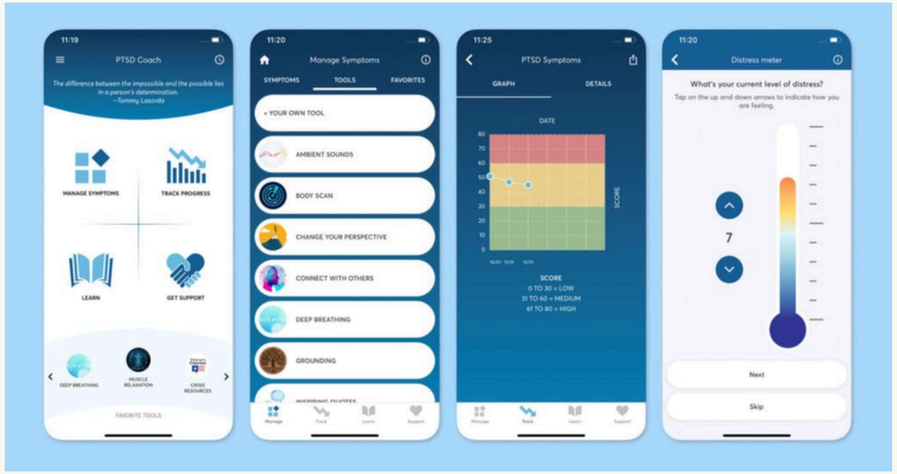
Our research demonstrates that there's no single architecture for PTSD detection that provides an optimized result:

- BiLSTM Approach: Better at identifying potential PTSD cases (higher recall) but with many false positives
- Transformer Approach: Better at distinguishing between classes (higher AUC) and generalizing to new data, but more conservative in predictions

- The Transformer demonstrates better generalization with higher AUC (0.73) and a smaller training-validation gap.
- Despite data augmentation and weighted sampling techniques, both models struggle with the severe class imbalance.
- Choosing eGEMAPS and ResNet features turned out to be effective than other features.

- The BiLSTM's higher recall might be preferred in initial screening contexts where missing cases is especially concerning, while the Transformer's better discrimination ability could be valuable in contexts where precision is prioritized.

DEPLOYMENT OPTIONS FOR MULTIMODAL PTSD DETECTION



Telehealth Integration:
Deploy as a video telehealth application integrated with existing platforms

Smartphone Applications:
Deploy as a mobile screening application accessible on personal devices

Virtual Interview Systems:
Implement AI-driven virtual interviewer through video conferencing

FUTURE WORK

- Collecting more dataset or use advanced class imbalance handling (e.g., focal loss, SMOTE for sequences).
- Advanced Fusion Strategies: Exploring more sophisticated ways to combine multimodal information
- Utterance-level chunking for more meaningful temporal units.
- Ensemble models to leverage complementary strengths of BiLSTM and Transformer.
- Robust cross-validation and external validation for generalizability.