EMPATICA E4 STRESS DETECTION

BY GROUP 7

WHAT ARE WE TRYING TO DO?

- Objective: Detect stress using only wristwatch sensor data
- Dataset Used: WESAD (Wearable Stress and Affect Detection)which includes data from both chest and wrist sensors (but we use only wrist).
- Why: Eliminate need for bulky, impractical devices like chest belts

Why it matters?

- Stress has become a part of daily life, with burnout, anxiety, and fatigue being common.
- Wristwatches provide a comfortable and wearable solution.

1. Improve workplace well-being
2. Support mental heath care
3. Be used in real time interventions

INSPIRATION & CREATIVITY

- Mental health is often silent.
- We believed a wristwatch could do more than tell time—it could listen.
- Not just for data, but for signs of stress.
- We chose empathy over complexity,
- Simplicity over chest belts,
- Found clarity in the chaos of wrist signals.

WHO CARES? WHO BENEFITS?

1. 👩‍💼 **Corporate Wellness Programs**

Companies can provide wristbands to monitor employee stress and trigger personalized wellness activities (e.g., break reminders, de-stress alerts).

2. 📱 **Personal Stress Tracking**

Individuals gain real-time insights into stress patterns, enabling self-awareness, lifestyle changes, and data-driven therapy support.

3. 🧠 **Mental Health Companion Apps**

Integrates with mental health platforms to detect prolonged stress and alert caregivers or recommend breathing/meditation routines.

4. 🏥 **Recovery Monitoring for Patients**

Ideal for patients recovering from anxiety, cardiac events, or surgery—non-intrusive, wearable tracking of physiological stress indicators.

USUAL PROCESS

1. Chest-worn belts (RespiBAN): reliable but impractical
2. high training accuracy, poor real-world use
3. Problems: Uncomfortable hardware, messy signals

OUR PROCESS

Extract

- Sensor data (EDA, TEMP, BVP,ACC
- Metadata(Subject ID, Timsestamps
- Emotion labels(Stress, Baseline, Neutal, etc.)

Transform

- Resampling
- Baseline Identification
- Feature Engineering(Label mapping, Rolling stats, and cleanup)

Load

- Train Model – Random Forest, XGboost, LSTM
- Evaluate Metrics (Accuracy, Recall ,Precision
- Visualize (Graphs, Confusion matrix

```
Processing: S4...
Target Frequency : 40Hz
ACC_x : 256920
ACC_y : 256920
ACC_z : 256920
BVP : 256920
EDA : 256920
TEMP : 256920
labels : 256920
S4 Resampled
```

```
Processing: S2...
Target Frequency : 40Hz
ACC_x : 243160
ACC_y : 243160
ACC_z : 243160
BVP : 243160
EDA : 243160
TEMP : 243160
labels : 243160
S2 Resampled
```

```
Processing: S7...
Target Frequency : 40Hz
ACC_x : 209520
ACC_y : 209520
ACC_z : 209520
BVP : 209520
EDA : 209520
TEMP : 209520
labels : 209520
S7 Resampled
```

Sensor Chaos

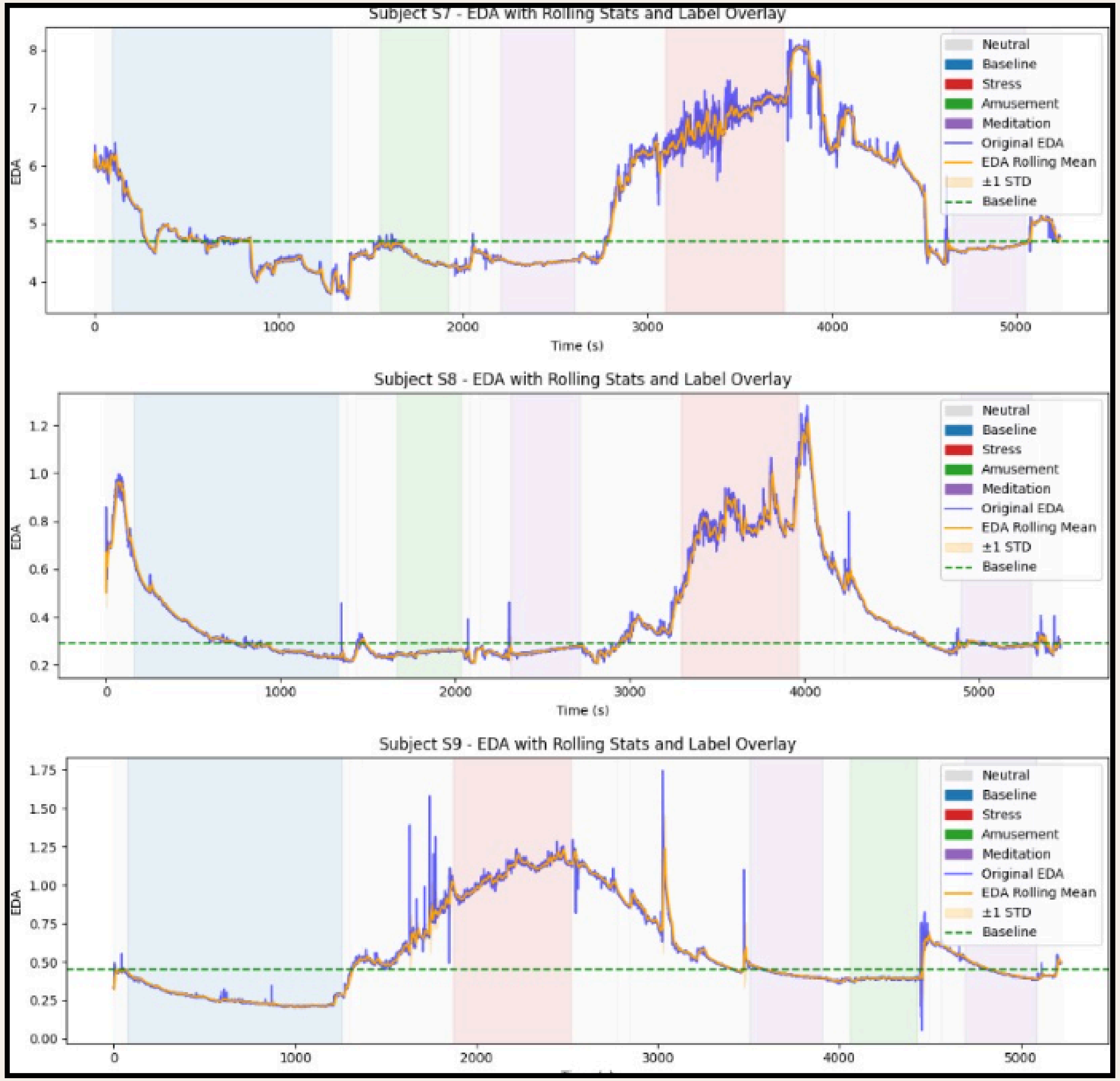
- Problem: Each sensor had different sampling rates (e.g., BVP at 64Hz, EDA at 4Hz)
- Struggle: Signals couldn’t be combined meaningfully
- Solution: Resample all to a uniform 40Hz

Why feature engineering ?

- One-Hot Encoding: Converted emotion labels for model compatibility
- Rolling Stats: Smoothed noisy signals (EDA, Temp) to reduce random spikes
- StandardScaler: Standardized sensor data to a common scale

Raw	**Transformed**	**Output**
Emotion	One-hot encoding	Neutral-0 Baseline-1 Stress-2 Amusement-3
EDA	Rolling Stats(Mean, median, min, max, diff)	EDA_mean, EDA_max,EDA_diff
Temp	Rolling Stats(Mean, median, min, max, diff)	Temp_mean, Temp_max, Temp_diff
Acc_x/y/z	Rolling Stats(Mean, median, min, max, diff)	Acc_mean, Acc_max(x/y/z)
Subject Metadata	Binary Encoding	Gender, Smoker, Drank Coffee

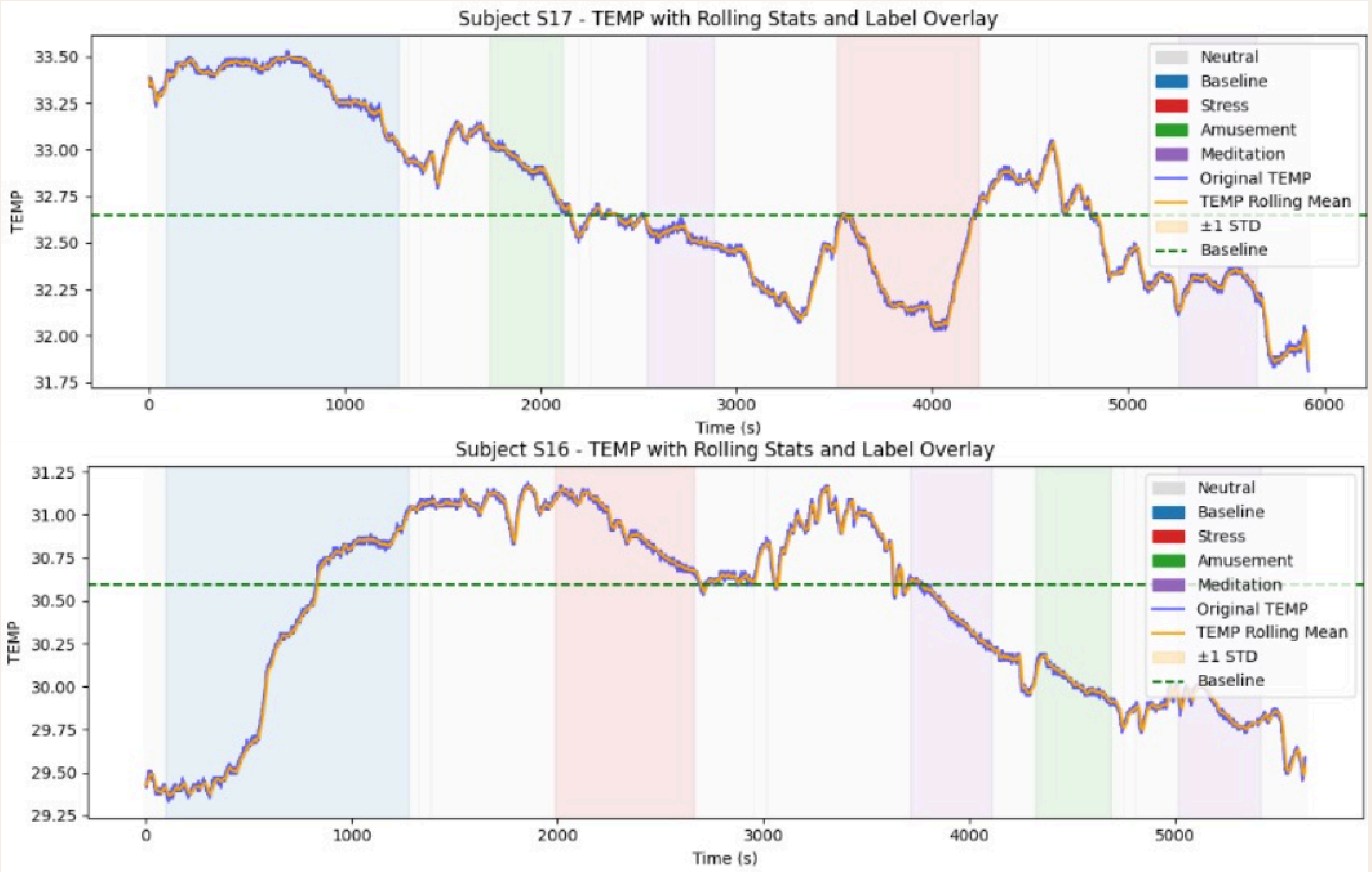
EDA
RAW VS SMOOTHED SIGNALS



Chaotic, Spiky Signals

- Problem: Raw signals (especially EDA, Temp)were noisy and unstable
- Struggle: Poor model performance due to irrelevant spikes
- Solution: Feature smoothing-Applied rolling features to smooth erratic spikes and reduce noise.

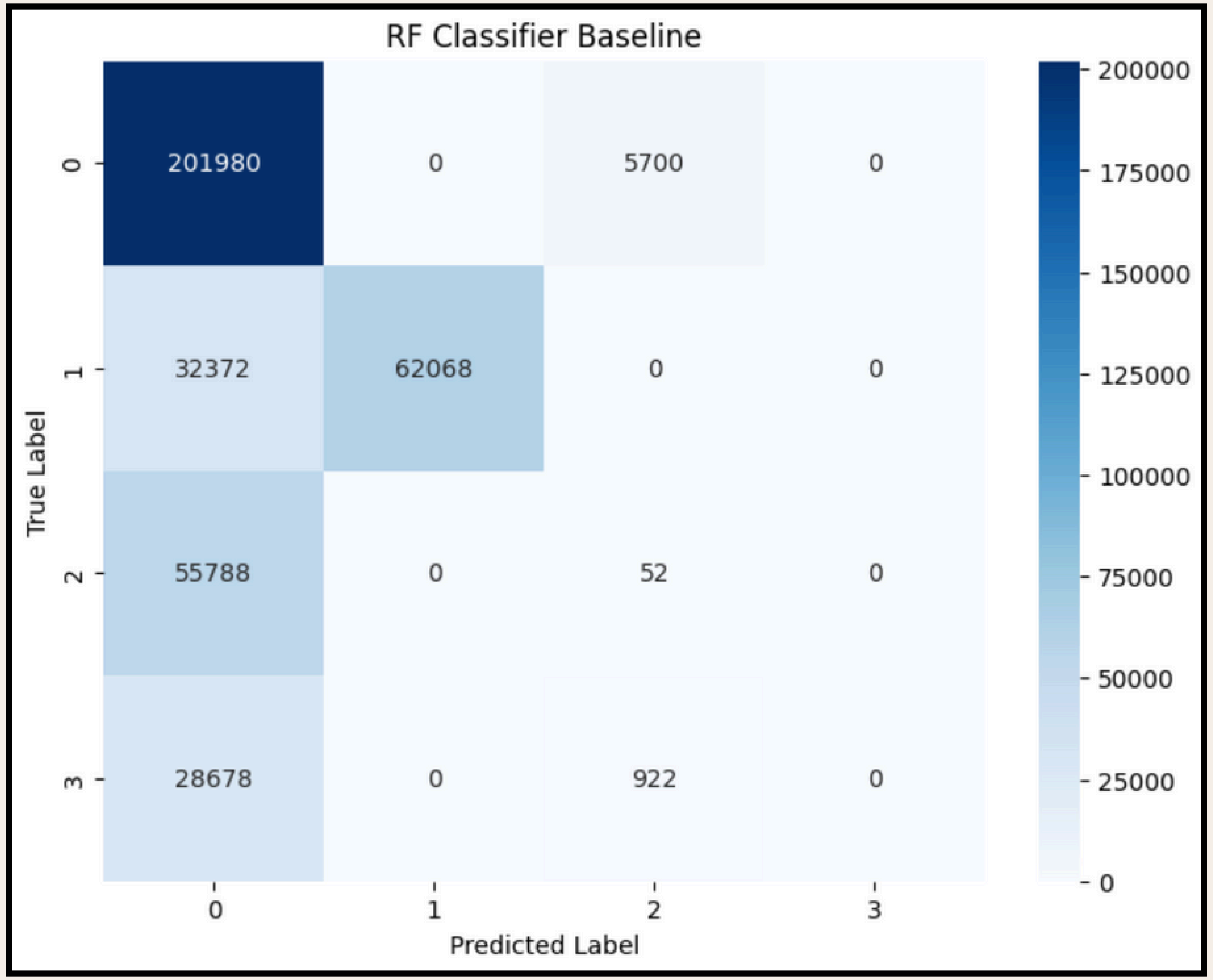
Temp
RAW VS SMOOTHED SIGNALS



BASELINE MODELS FOR RANDOM FOREST

accuracy	precision	recall	f1
0.999962	0.999962	0.999962	0.999962
0.999962	0.999962	0.999962	0.999962
0.999962	0.999962	0.999962	0.999962
0.999962	0.999962	0.999962	0.999962

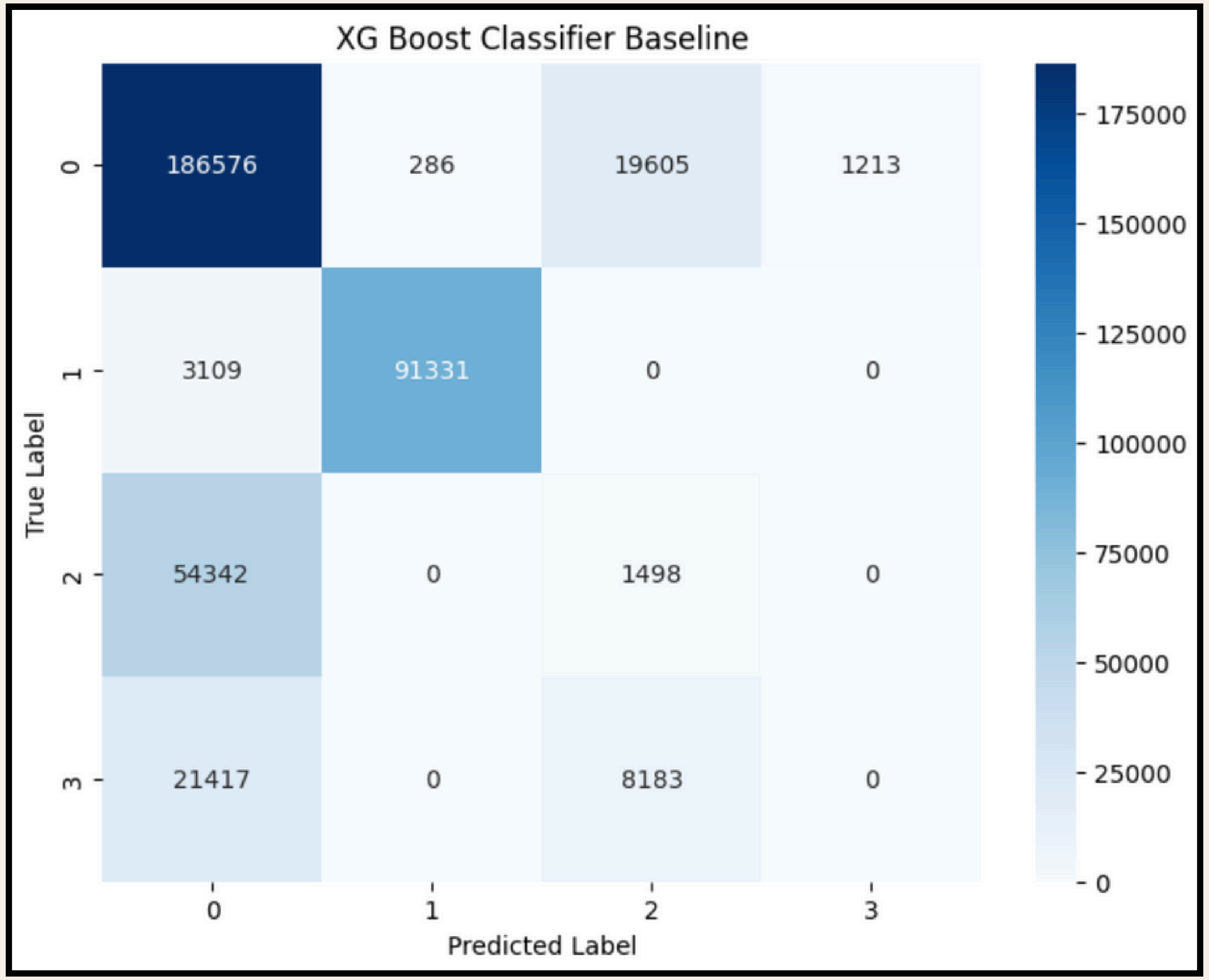
TEST RESULTS



BASELINE MODELS FOR XGBOOST

accuracy	precision	recall	f1
0.998331	0.998331	0.998331	0.998331
0.998331	0.998331	0.998331	0.998331
0.998331	0.998331	0.998331	0.998331
0.998331	0.998331	0.998331	0.998331

TEST OUTPUT



BASELINE MODELS FOR LSTM

TEST OUTPUT

	precision	recall	f1-score	support
0.0	0.98	0.97	0.97	13726
1.0	0.98	0.98	0.98	6100
2.0	0.95	0.99	0.97	3428
3.0	0.91	0.96	0.93	1934
accuracy			0.97	25188
macro avg	0.96	0.97	0.96	25188
weighted avg	0.97	0.97	0.97	25188

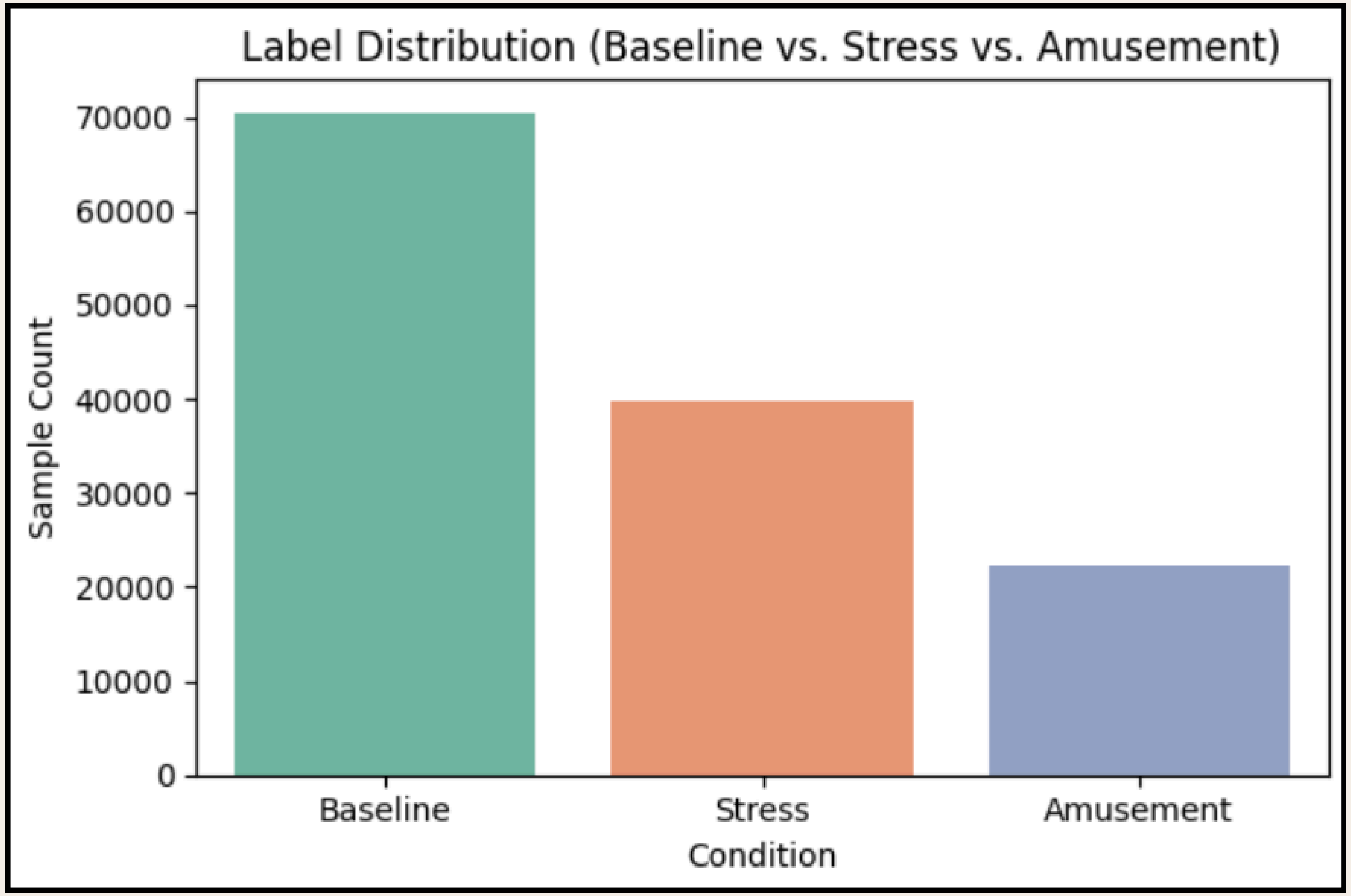
LSTM on Eval:

	precision	recall	f1-score	support
0.0	0.64	0.98	0.77	10382
1.0	0.94	0.50	0.65	4722
2.0	0.90	0.25	0.40	2792
3.0	0.41	0.01	0.02	1480
accuracy			0.69	19376
macro avg	0.72	0.44	0.46	19376
weighted avg	0.73	0.69	0.63	19376

False Accuracy

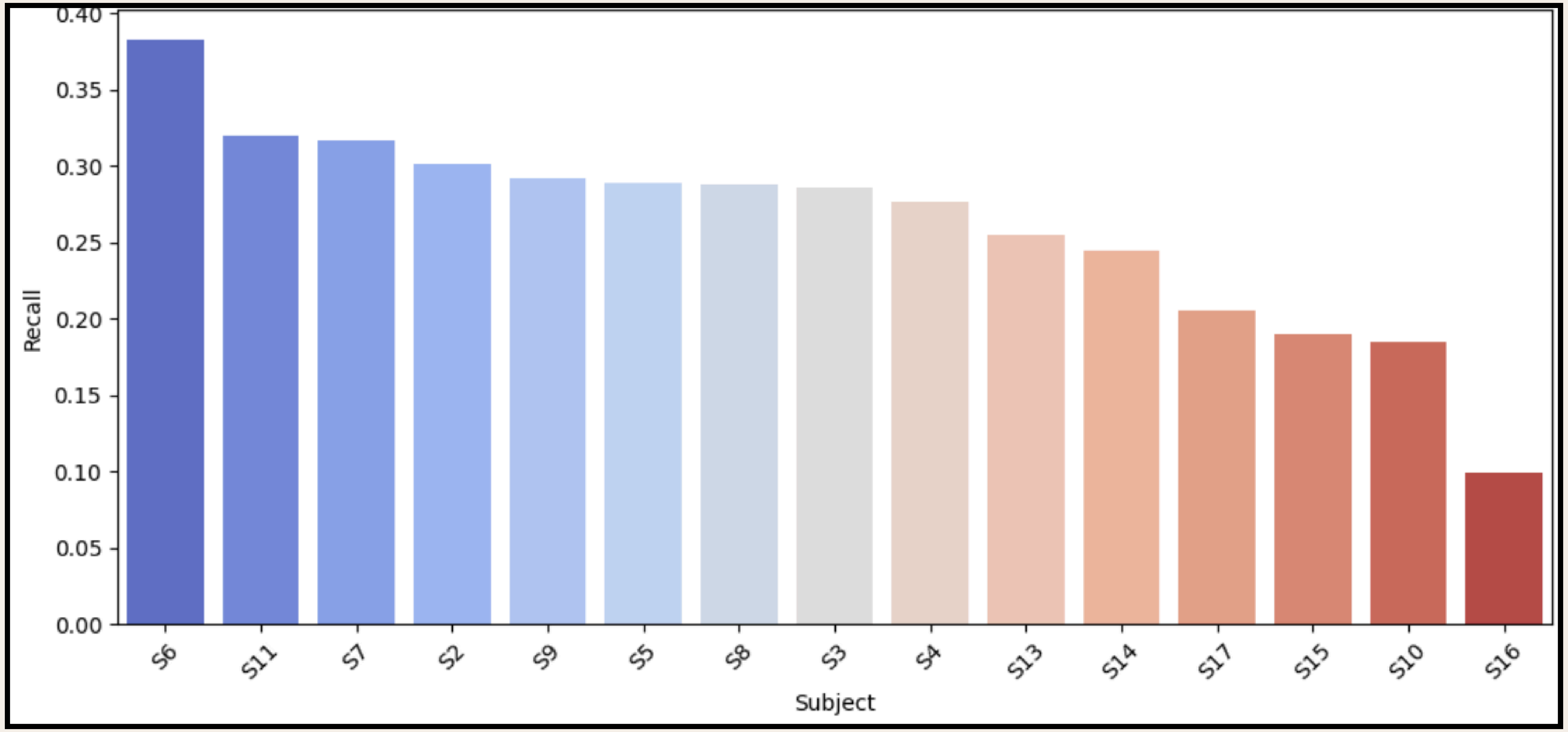
- Problem: Our model looked great until we tried it on someone new. What went wrong? model is overfit
- Struggle: Models failed on new people
- Solution:Leave-One-Subject-Out (LOSO) cross validation, SMOTE, Rolling Features — test model on unseen individuals

Why SMOTE?



- We used SMOTE (Synthetic Minority Over-sampling Technique) to balance the class distribution (as shown in the graph).
- Since stress and amusement had far fewer samples than baseline, SMOTE created synthetic examples for these underrepresented classes—
- Preventing model bias toward the dominant class (baseline)
- Improving generalization and fairness in classification

Why LOSO?

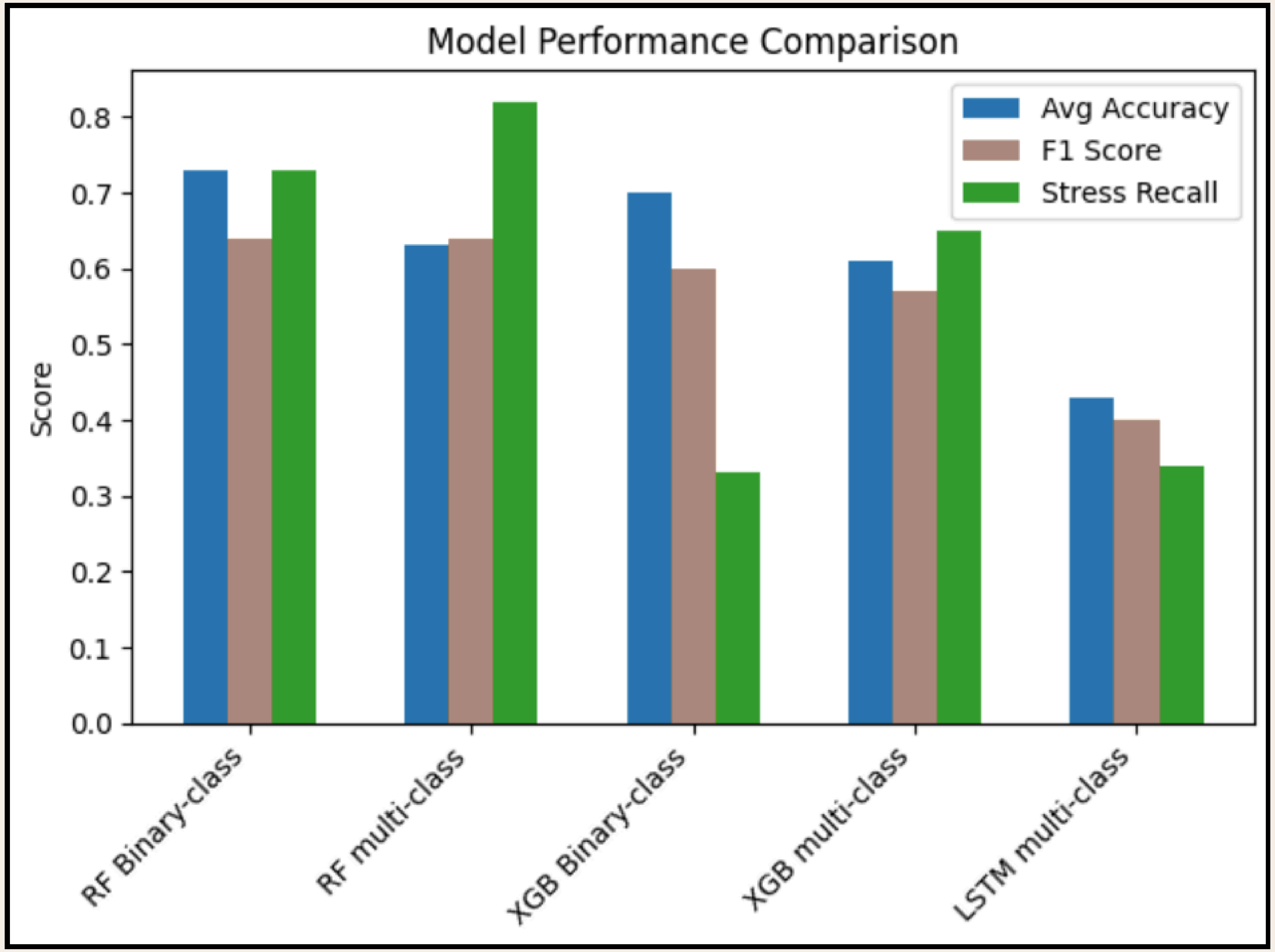


- In this chart, each bar shows recall performance for individual subjects.
- The variation proves that some subjects (e.g., S6, S11) are recognized well, while others (e.g., S16, S10) perform poorly—
  highlighting subject variability and the importance of evaluating generalization.
- LOSO helps avoid overfitting to specific people and ensures the model works across a diverse population.

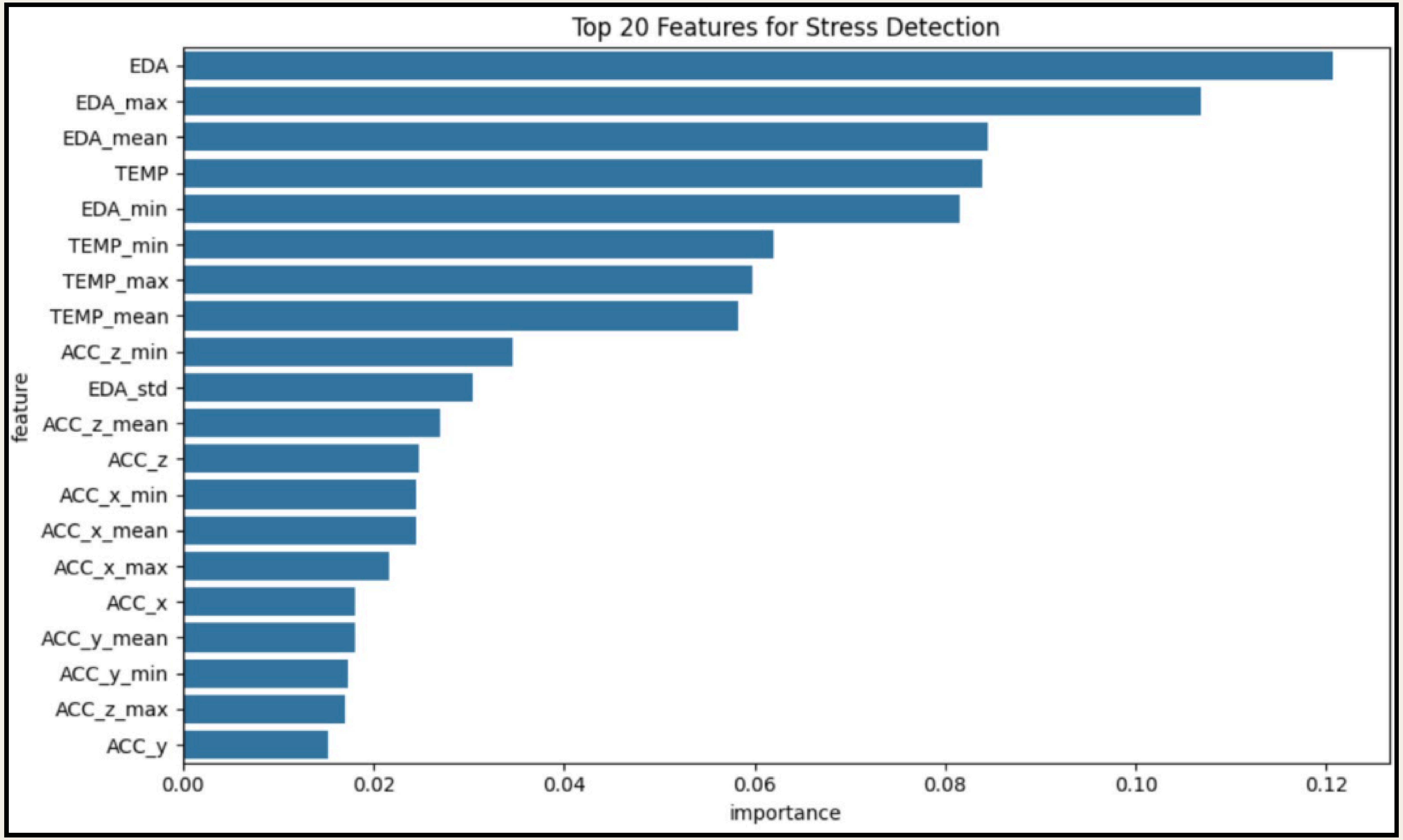
SUCCESS CASES RANDOM FOREST BINARY CLASSIFICATION

- For a production system where you want the **best trade-off** between overall accuracy, F1, and decent stress recall, go with **RF Binary-class**.
- If you **absolutely cannot miss stress events** and can tolerate more false-positives, try **RF Multi-class**.

Models	Avg Acc	F1	Stress Recall	Precision
Binary RF	**0.73**	**0.64**	**0.77**	0.54
Multi RF	0.63	0.64	**0.82**	0.53
Binary XGB	0.70	0.60	0.33	**0.70**
Multi XGB	0.61	0.57	0.65	0.50
Multi LSTM	0.43	0.40	0.34	0.49



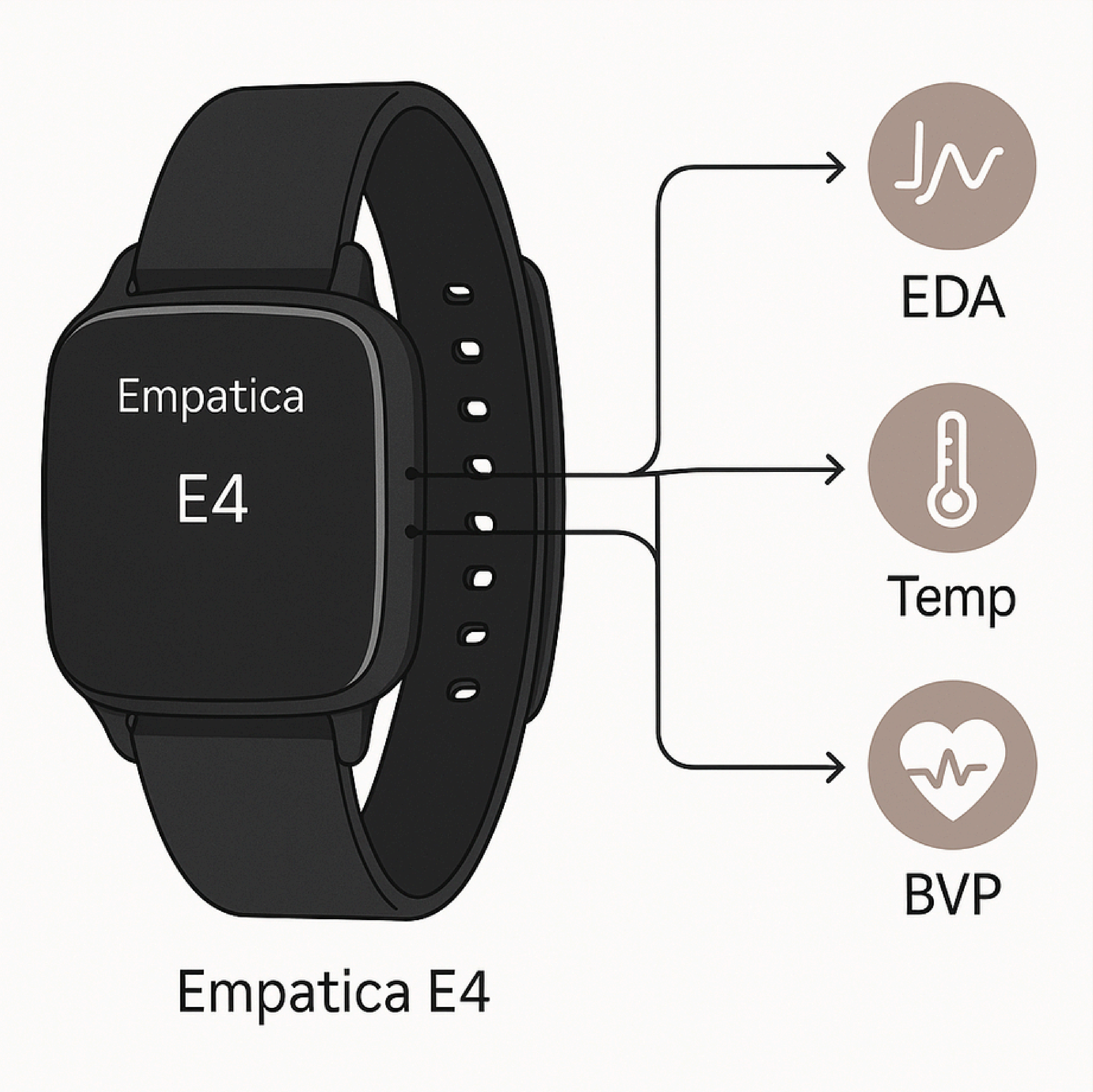
Feature Importance



EDA, TEMP and ACC along with their engineered features, impact the model highly, while categorical feature add almost no impact.

CONCLUSION

- Thoughtful design enables wrist-based stress detection.
- The combination of engineering and biology presents a genuine challenge.
- Models should honor human variability rather than solely pursue accuracy.
- The future of mental health monitoring will be wearable, personalized, and continuously evolving.



THE TEAM

KARTHIK

AVANI

SOWMYA

REFERENCES

- Github Link
- https://scholar.google.com/scholar?as_q=Scalable+deep+learning+for+stress+and+affect+detection+on+resource-constrained+devices&as_occt=title&hl=en&as_sdt=0%2C31
- https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10097874
- https://dl.acm.org/doi/abs/10.1145/3242969.3242985?casa_token=kPyQ-ocOAoQAAAAA:JtCCP1GcRMjf_U4RC8JHWVWbTI_-Ps633f6Bzf_ezIV-fZJMwk6tzsgTRqMaW3M-ZgF3tPmudpk8
- https://ieeexplore.ieee.org/abstract/document/5871728?casa_token=rhi7NqrK83IAAAAA:rHCdUsTNz78-fMVFq2TMSClnPLoFt-TyXz3yjBj_pUcfH6i_Gxx_OHZuukiha5yjR3FBZiJL
- https://dl.acm.org/doi/abs/10.1145/2968219.2968306?casa_token=RpwjAhAQ7MUAAAAA:Hv7KzdHf6Ff-4ZEbOgDBLzsBj85KXCYhgGP0VB9M7OciW1iSglBEK4m51cKiZHinoOwX7BNGF-YC
- https://www.sciencedirect.com/science/article/pii/S2215016124000359
- https://arxiv.org/pdf/2205.03974
- https://asharahmedd.medium.com/from-stressed-to-serene-how-i-built-a-real-time-stress-detection-system-using-machine-learning-c6e05220fbc6
- http://www.fit4work-aal.eu/
- https://www.apa.org/topics/stress/body
- WESAD Dataset
- Medium post

THANK YOU.