Stress Monitoring in Free-Living Environments

Ramesh Kumar Sah, Michael J. Cleveland, Hassan Ghasemzadeh

IEEE Journal of Biomedical and Health Informatics, Dec 2023

Research Focus: Data-driven stress detection for individuals with Alcohol Use Disorder (AUD) in real-world settings using CNN-based models

Presenter: Nooshin Taheri

12/17/2025

Motivation & Background

- Stress is a **major risk factor** for both physical and mental health.
- It is especially critical for individuals with **Alcohol Use Disorder (AUD)**, where stress can trigger relapse.
- **Lab-based stress studies** don't reflect how people actually live their daily lives.
- **Wearable sensors** make continuous, real-world stress monitoring possible.

But in real life:

- **Many sensors** → higher cost, energy use, and complexity.
- Stress is **subjective and self-reported**, so labels are noisy.
- The **exact timing of stress** relative to the signal is uncertain.

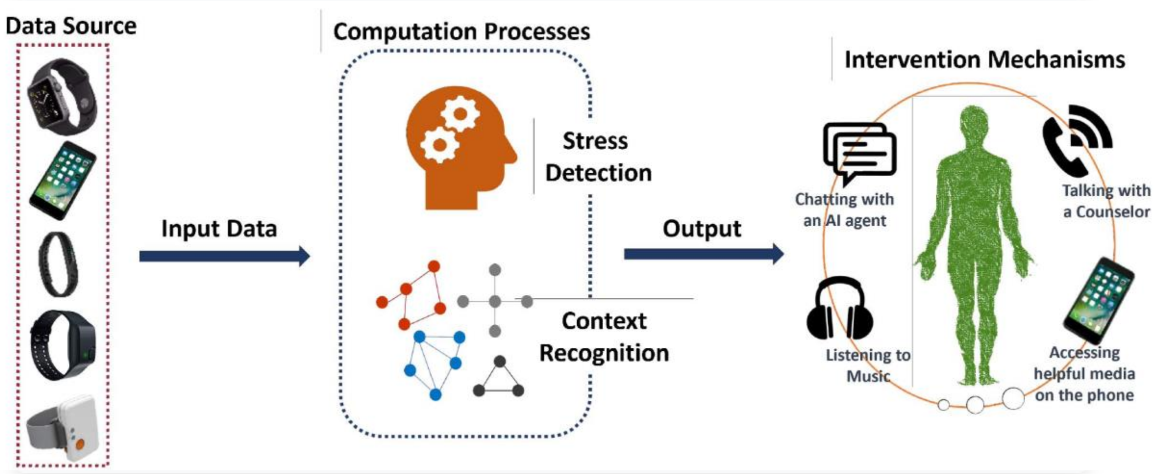


Figure 1: Mobile health (mHealth) system for stress detection and intervention. Wearable sensors collect data, computational processes detect stress, and personalized interventions help individuals remain abstinent from alcohol.

Gaps in Current Research

❌ Problem 1: Lab-Only Studies

Prior research conducted in controlled environments (counting backward, speech preparation)

Issue: Fails to generalize to real-world settings

❌ Problem 2: Healthy Participants

Lack of research with vulnerable populations

Issue: Doesn't address needs of high-risk groups like AUD patients

❌ Problem 3: Multiple Sensors

Reliance on multiple sensor modalities

Issue: High computational cost and energy requirements

❌ Problem 4: Known Duration

Assumes stress event duration is known

Issue: Stress response is subjective and varies by individual

Research Aims

Problem Statement

Develop a practical method to detect stress episodes in people with AUD in their daily lives, while:

- Identifying which sensor channel(s) are actually needed.
- Finding how much data around a self-reported stress event should be used.

Key Questions

1. Which sensor channel is most informative for stress detection?

2. What is the optimal segment length around stress events?

Main Contributions

1. Polynomial-Time Sensor Channel Selection Algorithm

Developed **KCS and PCS algorithms** to identify the best sensor channel(s) for stress detection, reducing complexity from $O(2^n)$ to $O(n^2)$.

2. Method to Find Optimal Stress Segment Length

Created an iterative search algorithm to determine the **optimal time window** around self-reported stress events.

3. Evaluation on Real-World Data

First study using data from **individuals with AUD** in free-living conditions (1,698 hours, 409 stress events).

4. Public Release of Dataset + Code

Made dataset and analysis code **publicly available** to encourage further research in this area.

Machine Learning Approach

Supervised Learning Framework: Train a model *f* to learn the mapping **f : X → Y**

```
X (sensor data) → f (model) → Y (stress/not-stress)
```

Dataset **D(X, Y)** from ADARP study: real-world data from individuals with AUD

Why CNN?

- **Feature extraction block:** Automatically learns representations from raw sensor data
- **Classification block:** Maps features to stress/not-stress classes
- **No manual feature engineering:** Bypasses cumbersome domain-expert-driven feature selection
- **End-to-end learning:** From raw signals to predictions

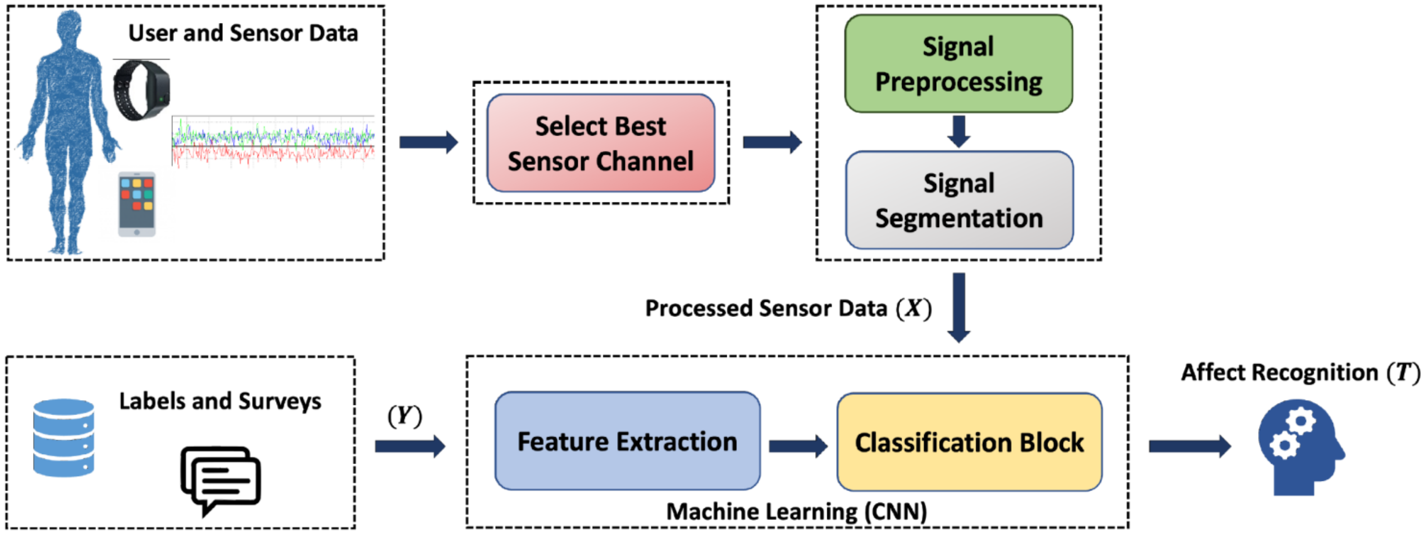


Figure 2: CNN architecture with feature extraction and classification blocks

Key Challenges in Real-World Settings

Challenge 1: Which Sensors to Use?

Problem: Wearable devices provide multiple sensor modalities, but which are actually needed?

- Continuous monitoring requires efficient resource consumption
- Unnecessary sensors increase cost, energy use, and complexity
- Need to find the **optimal sensor channel combination**

Challenge 2: How Much Data Around Stress Events?

Problem: Stress timing is uncertain due to subjective self-reports

- Stress ground truth from EMA and E4 event marker button
- Users may report stress **before, during, or after** the physiological peak
- Duration of acute stress varies person-to-person
- Need segment large enough to capture stress, but not baseline data

Figure 4: Event marker timing vs. actual physiological stress peak - shows delay and uncertainty in self-reported stress

Optimal Sensor Channel Selection

The Challenge

Multi-modal systems perform better BUT come with significant costs

❌ Drawbacks of Multiple Sensors

- Higher power consumption
- Poor battery life
- Complex processing routines
- Increased system cost
- Lower user adoption

✓ Goal of Optimization

- Minimize number of sensors
- Maintain performance requirements
- Reduce computational load
- Extend battery life
- Encourage adoption

⚡ Critical for wearable systems where computation and battery are constrained

Channel Selection Problem Definition

Problem 1: Top-K Channel Selection (KCS)

Find the best k channels without performance constraints

$$C^*_k = \arg\max P(C_k)$$

where $k < n$ (total channels)

Goal: Select k channels that maximize model performance

Problem 2: Performance-Guaranteed Selection (PCS)

Find minimum channels that meet performance threshold $\bar{P}$

$$\text{minimize } |C_p|$$
$$\text{subject to: } P(f(I_p)) \geq \bar{P}$$

Goal: Use fewest channels while maintaining required performance

Performance Metrics: Accuracy, F1-score, Mean Absolute Error

Available Sensor Channels in Empatica E4 Wristband

7 Sensor Channels Available

🫀 Cardiovascular

BVP: Blood Volume Pulse (includes HRV)

HR: Heart Rate

⚡ Electrodermal

EDA: Electrodermal Activity (Skin Conductance)

🌡️ Temperature

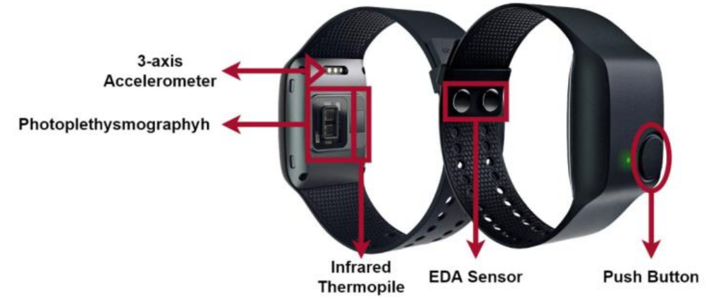
TEMP: Skin Temperature

🏃 Motion

ACC-X:X-axis Acceleration

ACC-Y:Y-axis Acceleration

ACC-Z:Z-axis Acceleration

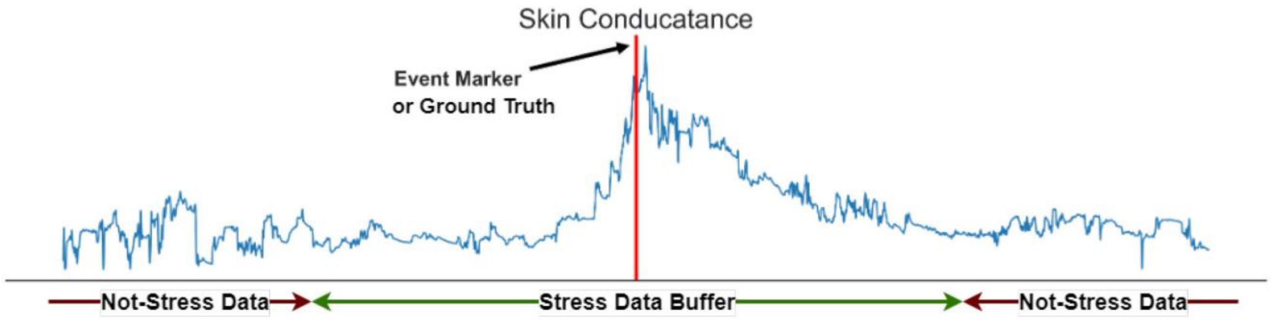


Question: Which channel(s) are most effective for stress detection?

Optimal Stress Segment Length

The Problem

In real-world settings, the duration of stress events is **not known in advance**



⚠️ Challenges

- **Subjective nature of stress:** Same stimuli → different effects
- **Temporal variations:** Event markers may be before, during, or after actual stress peak
- **Unknown duration:** How long does a stress episode last?
- **Individual differences:** Stress response varies person to person

🔬 Approach

- Test segment lengths: **60s to 3600s** (60s increments)
- Train CNN model for each length
- Select length with best performance
- Use EDA sensor data (best channel)

🎯 Goal

Find segment length **L** that:

- Captures all stress event variations
- Minimizes generalization error
- Maximizes model performance

Optimization: Find length L such that generalization error $\varepsilon_g(f(x)) \leq \delta$

User Study and Dataset

📋 Study Design

Participants: 11 individuals receiving treatment for mental health and Alcohol Use Disorder (AUD)

Duration: 14 days of continuous monitoring

Device: Empatica E4 wristband

📊 Data Collection Protocol

- **Continuous physiological monitoring** via E4 wristband
- **Event markers:** Participants pressed E4 button when feeling stressed
- **EMA surveys:** 4 times daily for self-reported outcomes

📈 Dataset Statistics

1698
Hours of physiological data

409
Stress moments identified

🔧 Filtering and Noise Removal

Preprocessing applied to EDA signals

🎛 Low-Pass Butterworth Filter

- **Filter type:** Second-order Butterworth
- **Cutoff frequency:** 1.25 Hz
- **Purpose:** Remove high-frequency noise from EDA signals

✓ Data Quality Assessment

- **Quality estimation tools:** EDA Explorer and LedaLab
- **Clean signal proportion:** 87.86% of EDA signals
- **Analysis methods:** Trough-to-Peak (TTP) and Continuous Decomposition Analysis (CDA)
- **Additional processing:** Only low-pass filtering applied (no other routines needed)

Note: High data quality (87.86% clean signals) enabled minimal preprocessing pipeline

Dataset Construction

🔨 Dataset Construction

Sensor data partitioned based on event markers (stress ground truth)

📍 Data Partitioning

- **Stress class:** Data within variable segment length around event markers
- **Not-stress class:** Data outside 60-minute buffer zone (starts 60 min before and 60 min after event marker)
- **No overlap:** Between stress and not-stress classes

✓ Label Verification Challenge

- **Cross-validation:** E4 event markers vs. EMA survey responses
- **Verification rate:** Only 26% of E4 stress events corroborated by surveys
- **Decision:** Used only E4 event markers (more temporally accurate)
- **Rationale:** E4 button pressed during actual stress moment vs. survey completed later

⚠️ Limitation

Participants may not always realize stress in free-living environments or forget to press the button

Class Imbalance

⚖️ Class Imbalance Problem

Large imbalance between stress and not-stress classes due to low number of stress events

📊 Imbalance Characteristics

• Number of stress samples depends on segment length

• Smaller segment lengths → fewer stress samples → higher imbalance

• **Consequence of real-world data collection**

🔧 Balancing Strategies

1. Majority Class Undersampling

• Randomly drop samples from not-stress class

• **Advantage:** Preserves variance in both classes

• **Expected:** Better generalization

2. Minority Class Oversampling (SMOTE)

• Generate synthetic stress samples

• **Disadvantage:** Creates similar samples, reduces variance

• **Risk:** Model overfits to noise, poor generalization

Key Insight: Class imbalance is a natural consequence of real-world stress monitoring where stress events are relatively rare

Results: Optimal Sensor Channel

🎯 Experimental Setup

- **Algorithm:** KCS (k=1) with CNN model
- **Channels tested:** 7 (EDA, BVP, HR, TEMP, ACC-X, ACC-Y, ACC-Z)
- **Stress segment:** 40 minutes around event markers
- **Window size:** 60 seconds with 50% overlap
- **Data split:** 70% training, 30% testing
- **Balancing:** Random undersampling

💡 Key Finding

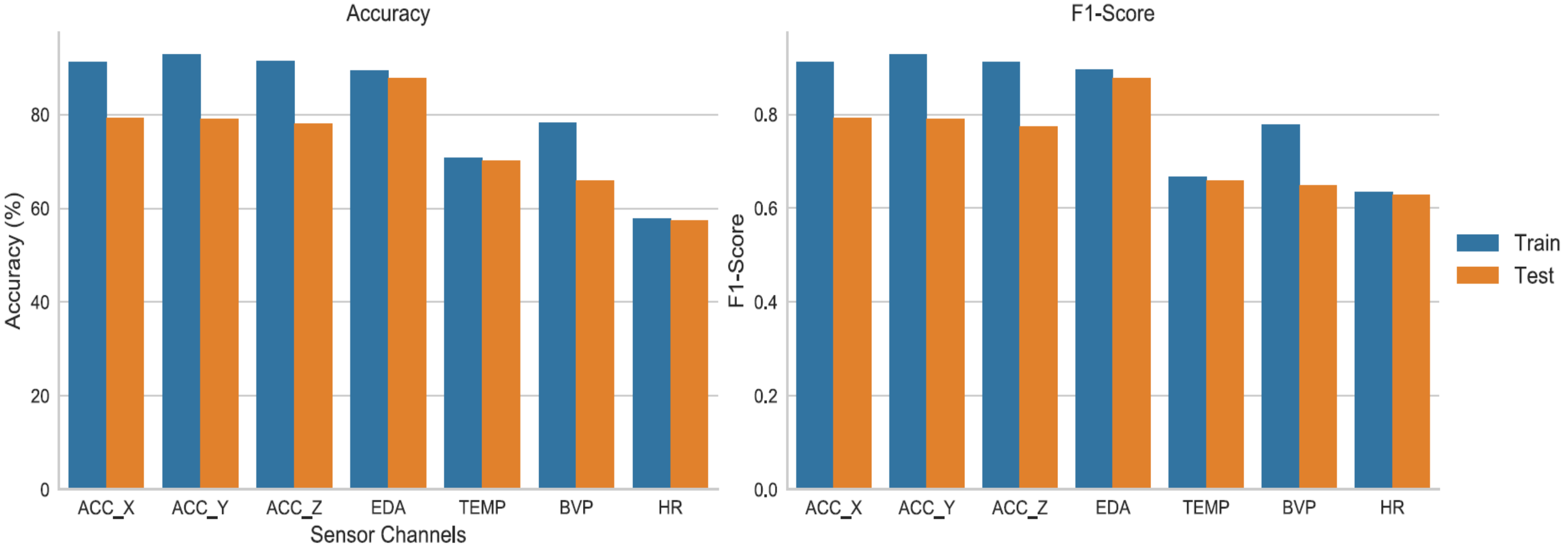
EDA (skin conductance) identified as the single best sensor modality for stress classification in real-world settings

🏆 Winner: EDA (Electrodermal Activity)

✓ **Best performance on test set**
✓ Similar training performance to ACC-Y
✓ Superior generalization capability
✓ Most indicative of stress responses

📊 Why EDA Outperforms ACC

- ACC signals have more noise
- ACC models prone to overfitting
- EDA better captures stress physiology
- EDA provides cleaner signal for stress detection



Conclusion: Single-channel EDA provides optimal balance between performance and system efficiency

Results: Optimal Stress Segment Length

🔬 Experimental Setup

- **Segment lengths tested:** 60s to 3600s (60s increments)
- **Sensor channel:** EDA (best channel identified)
- **Test set:** Fixed across all experiments (39 stress samples, 39 not-stress samples)
- **Balancing method:** Random undersampling

🏆 Winner: 60 Seconds

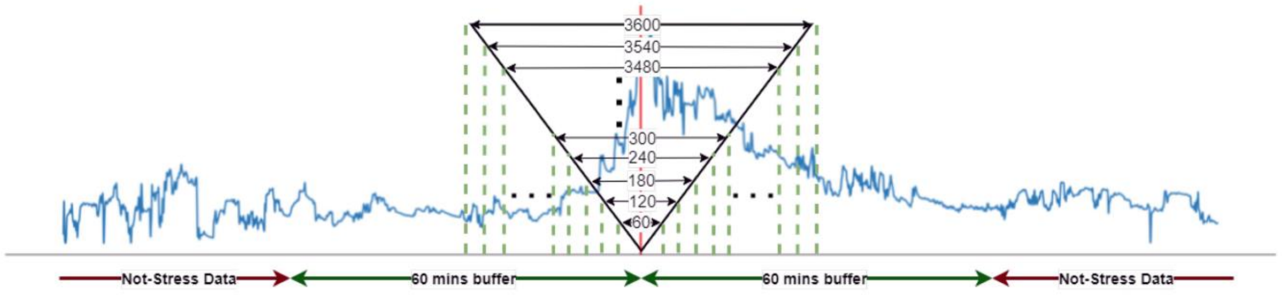
✓ **Highest accuracy and F1-score**
✓ Best performance on both training and test sets
✓ Optimal for capturing stress events
✓ Sufficient to accommodate temporal variations

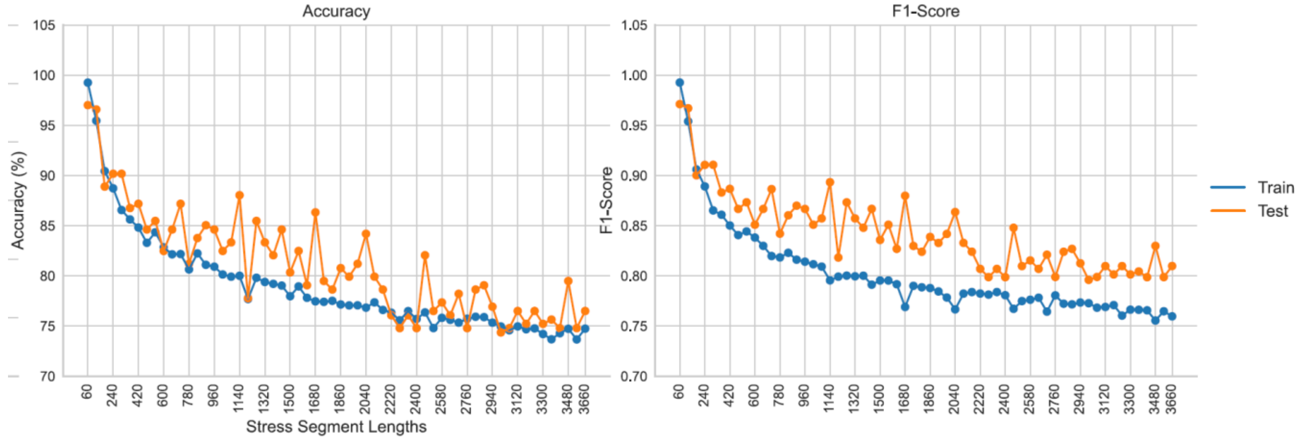
📉 Performance Trend

- **As segment length increases:** Performance decreases
- Longer segments include more baseline/not-stress data
- Reduces discriminative power between classes
- Optimal window captures stress peak without noise

💡 Key Insight

The **60-second segment** provides the best balance between capturing subjective stress variations and maintaining classification accuracy





Conclusion: 60-second segment length around event markers yields optimal stress detection performance

Results: Stress Classification

🎯 Final Model Configuration

- **Sensor channel:** EDA (Electrodermal Activity)
- **Segment length:** 60 seconds around event markers
- **Model:** CNN with 5-fold cross-validation
- **Training samples:** 203
- **Test samples:** 50
- **Balancing:** Random undersampling

✨ Key Achievements

✓ Excellent generalization (no overfitting)
✓ Balanced precision and recall
✓ Single-channel solution (efficient)
✓ Real-world data validation

💡 Significance

First successful stress detection model using **real-world data** from individuals with **Alcohol Use Disorder**, achieving near-perfect accuracy with minimal sensor requirements

TABLE I
STRESS CLASSIFICATION WITH EDA STRESS SEGMENT LENGTH OF 60 SECONDS AND RANDOM UNDERSAMPLING TO BALANCE THE CLASSES

Dataset	Loss	Accuracy (%)	Precision	Recall	f1-Score
Training	0.009	99.707	0.994	1.0	0.997
Testing	0.075	99.215	0.985	1.0	0.992

Conclusion: EDA-based CNN model with 60-second segments achieves 99% accuracy for stress detection in free-living environments

Results: Oversampling Comparison

🔄 SMOTE Oversampling Setup

- **Method:** Synthetic Minority Over-sampling Technique (SMOTE)
- **Best sensor:** EDA (same as undersampling)
- **Optimal segment:** 180 seconds (accuracy), 660 seconds (F1-score)
- **Mixed results** compared to undersampling approach

⚠️ Why SMOTE Underperformed

- **Overfitting:** Model learns specific details/noise of synthetic stress samples
- **Low variance:** SMOTE creates highly similar stress samples
- **Poor generalization:** Higher generalization error on unseen data
- **Misclassification:** Low recall (0.52) indicates not-stress samples classified as stress

💡 Key Insight

Undersampling preserves real data variance and leads to better generalization, while SMOTE creates artificial patterns that don't generalize well in real-world settings

TABLE II
Stress Classification With EDA Stress Segment Length of 180 Seconds and SMOTE Oversampling to Balance the Classes

Dataset	Loss	Accuracy (%)	Precision	Recall	f1-Score
Training	0.018	99.33	0.992	0.994	0.993
Testing	3.974	76.25	0.99	0.529	0.689

Conclusion: Majority undersampling outperforms minority oversampling (SMOTE), highlighting challenges of working with real-world datasets

Conclusion

🎯 Key Findings

✓ **EDA (Electrodermal Activity)** identified as the most indicative sensor modality for stress detection

✓ **60-second segment length** around stress events provides optimal performance

✓ **99% accuracy and 0.99 F1-score** achieved with majority undersampling

✓ **Single-channel solution** enables efficient, low-power wearable systems

💡 Impact

First successful data-driven stress detection approach for individuals with **Alcohol Use Disorder** in free-living environments, demonstrating the feasibility of efficient mobile health interventions using minimal sensor requirements.

⚠️ Challenges Highlighted

Performance with minority oversampling (SMOTE) dropped to **76.25% accuracy**, highlighting the challenges of working with real-world datasets and the importance of preserving data variance.

This work provides a viable pathway for efficient stress monitoring systems that balance performance with practical deployment considerations in real-world settings