Pilot Stress Detection Through Physiological Signals Using a Transformer-Based Deep Learning Model

- IEEE Sensors Journal, 2023
- Cited by: 22

- Presenter: Nooshin Taheri
- 8/14/2025

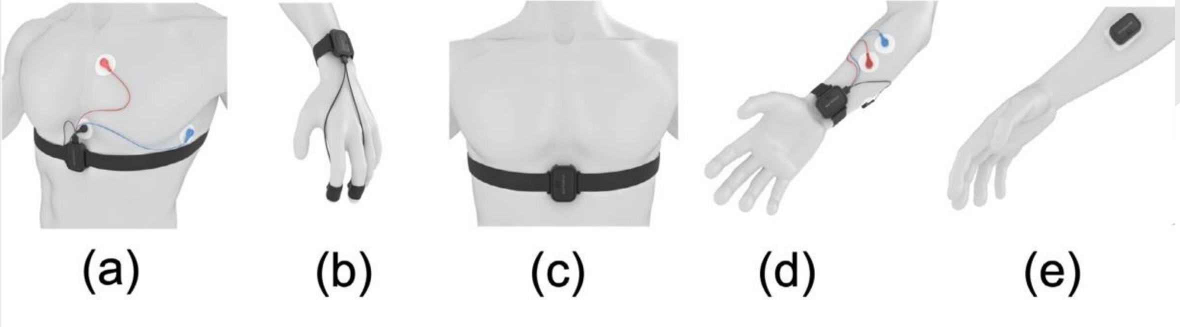
Why Detecting Pilot Stress Matters

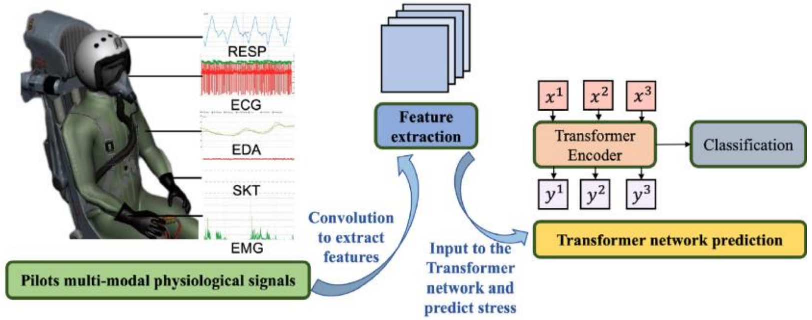
• Pilot stress detection is **crucial** for flight safety and performance.
• **High automation + fewer crew members** → pilots must handle multiple simultaneous tasks, leading to increasing stress.

- Pilot **high stress** leads to:
  - Poor decision-making
  - Reduced situational awareness
  - Decreased performance

• Early detection enables flight systems to adjust workload and prevent performance decline during acute stress events.
• With **wearable physiological sensors** and **AI**, stress can now be detected in real time.

Aim of This Study





- Detect pilot stress caused by mission difficulty or abnormal flight conditions using **multimodal physiological signals.**
- **Signals chosen** for portability & effectiveness:
    - Electrocardiography (ECG)
    - Electromyography (EMG)
    - Electrodermal activity (EDA)
    - Respiration (RESP)
    - Skin temperature (SKT)
- **Contribution of the study:**
    - **Novel CNN + Transformer architecture**
        - CNN module → extracts **local features** and patterns within each signal.
        - Transformer module → captures **global dependencies** and relationships between multiple signals.

Experiment Protocol

Maneuver	Overall Rating		Comparative Rating	
	μ	σ	μ	σ
Level flight	1.93	0.83	1.29	0.47
Roll	1.93	0.73	2.07	0.83
Hovering	3.14	0.66	3.43	0.85
Takeoff	3.36	0.63	3.29	0.83
Somersault	4.50	0.52	4.93	0.27

- Participants: 14 professionally trained flight cadets
- Equipped with wearable physiological measurement devices.
- Completed State-Trait Anxiety Inventory (STAI-Y1) to confirm stress-free baseline.
- Five maneuvers performed in the simulator (**Takeoff, Level flight**, **Roll, Hovering**, **Somersault)**
- **Stress Assessment**
  - **Questionnaire after each maneuver**
    - Overall Rating (OR): 1 = No stress, 5 = High stress
    - Comparative Rating (CR): Rank maneuvers by stress level
  - Scores are standardized using **Z-score**
  - The mean and standard deviation of the scores were calculated

Labeling Method

Stress score computed **per second** based on:

- Aircraft attitude change
- G-force variation
  - G-force is the acceleration of an object relative to the Earth's gravitational force, which is output by the flight simulation software.
- Skin conductance variation

Verification:

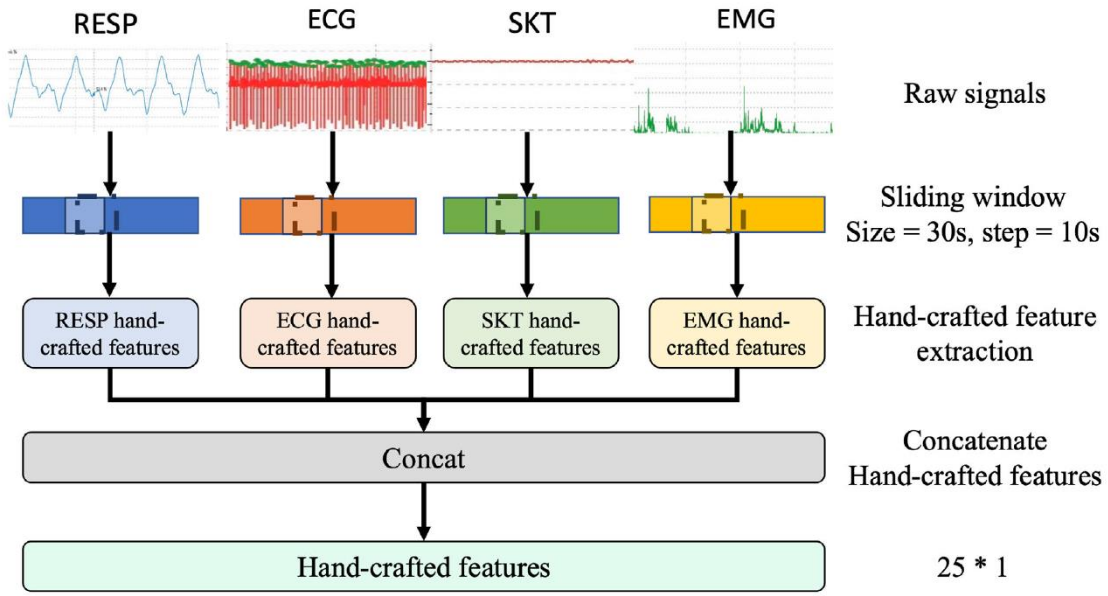
- Average stress scores for each maneuver were calculated.
  - Results exceeded the acceptable threshold.
  - This shows the scoring method is **feasible** and produced **consistent results** agreed upon by all scorers.

Maneuver	Level flight	Roll	Hovering	Takeoff	Somersault
Average Stress Score	1.2	2.4	3.9	3.5	5.2

Classes defined:

- **2-class**: Low vs High stress
- **3-class**: Low, Medium, High stress
- **4-class**: None, Low, Medium, High stress

Data Processing & Features



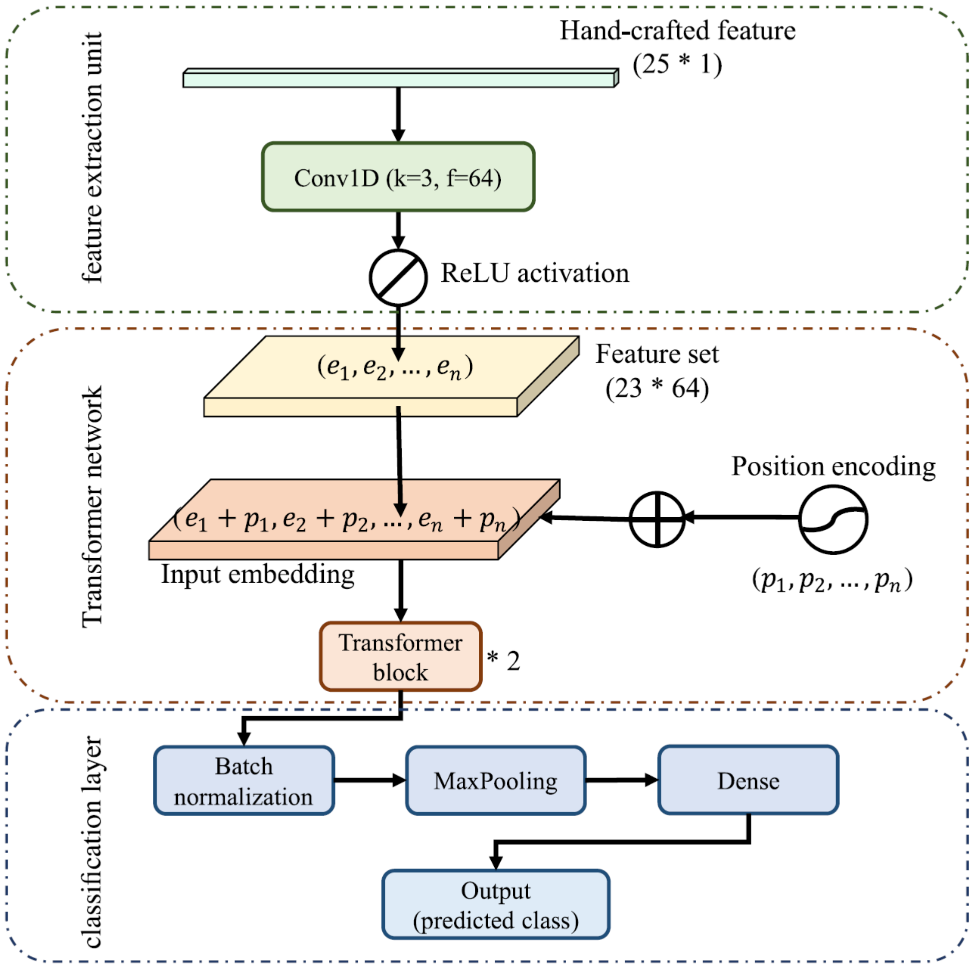
- **Preprocessing:**
  - Handle missing data (interpolation filling)
  - denoising (wavelet, filtering)

- **Features:**
  - **Handcrafted (**25 features per time window**):**
    - **ECG/HRV:** Heart rate, NN intervals, SDNN, RMSSD, pNN50, pNN20, VLF, ULF, LF, HF, LF/HF ratio
    - **EMG:** Voltage mean, std, RMS, integral EMG, median frequency, mean power frequency
    - **RESP:** Respiratory rate (mean, std), respiration power
    - **SKT:** Skin temperature

  - **Deep features:** Extracted via CNN layers.

Model Architecture

- **Pipeline:**
  - **Feature Extraction Unit** – 1D CNN → captures local patterns. (extract hidden features)
  - **Transformer Block** – self-attention → models dependencies between modalities.
  - **Classification Layer** – predicts stress level.
- Due to the relatively few features and the individual differences between subjects, dropout is replaced by layer normalization (LN) and regularization.
- more attention would be paid to positions with higher activation.



Results

- **Baseline Models Compared**
  - Modified **AlexNet** (1D conv layers for signals)
  - Modified **ResNet18**
  - Light-ResNet (simplified ResNet for fairness in computation)
- All models evaluated with **10-fold cross-validation**.
- A higher number of classes increases difficulty, but the transformer advantage grows.
- Comparable performance to ResNet18 but with **faster training**.
- Position embedding improves accuracy

CLASSIFICATION PERFORMANCE OF THREE MODELS

Model	Task	Accuracy	FLOPs (MFLOPs)	Prams
Transformer	2-class	93.28%	12.159	287,298
	3-class	88.75%	12.162	288,771
	4-class	84.85%	12.165	290,244
Modified AlexNet	2-class	82.86%	3.262	912,642
	3-class	77.62%	3.263	913,155
	4-class	70.30%	3.264	913,668
Modified ResNet18	2-class	92.17%	64.191	3,853,442
	3-class	86.32%	64.203	3,856,003
	4-class	80.45%	64.214	3,858,564
Light-ResNet	2-class	91.18%	12.266	711,938
	3-class	85.45%	12.269	713,219
	4-class	78.20%	12.271	714,500

RESULTS OF WITH/WITHOUT POSITION EMBEDDING

Position Embedding	2-class	3-class	4-class
With	93.28%	88.75%	84.85%
Without	91.01%	86.62%	82.22%

Discussion

- Transformer module captures dependencies between multimodal physiological signals better than CNN alone.

- **Feature extraction unit** (handcrafted + CNN deep features) improves accuracy by:
  - Representing both **time-domain** and **frequency-domain** info.
  - Capturing local correlations before transformer processing.
- **Position embedding**:
  - Encodes location info of features.
- **Attention mechanism**:
  - Focuses on global relationships.
  - The transformer’s attention mechanism can directly link any two points in a signal, no matter how far apart they are, making it easier to spot relationships between events separated in time.

Conclusion

- Developed a **CNN + Transformer** model for detecting pilot stress from **multimodal physiological signals** (ECG, EMG, EDA, RESP, SKT).
- **CNN** extracts local patterns; **Transformer** captures global relationships between different signals.
- Achieved **high accuracy**: 93.28% (2-class), 88.75% (3-class), 84.85% (4-class), outperforming AlexNet and matching or exceeding ResNet18 with faster training.
- Combining **handcrafted features** with deep-learned features improves robustness.
- Position embedding and attention mechanism are key for performance.
- **Potential use:** real-time pilot stress monitoring to enhance safety and decision-making.