NON-INVASIVE STRESS MONITORING FROM VIDEO

Akshata Tiwari[1], Brian Matejek[2], Daniel Haehn[3]

[1]Department of Electrical Engineering and Computer Science, MIT,
[2]Computer Science Laboratory, SRI International, [3]Department of Computer Science, University of Massachusetts Boston

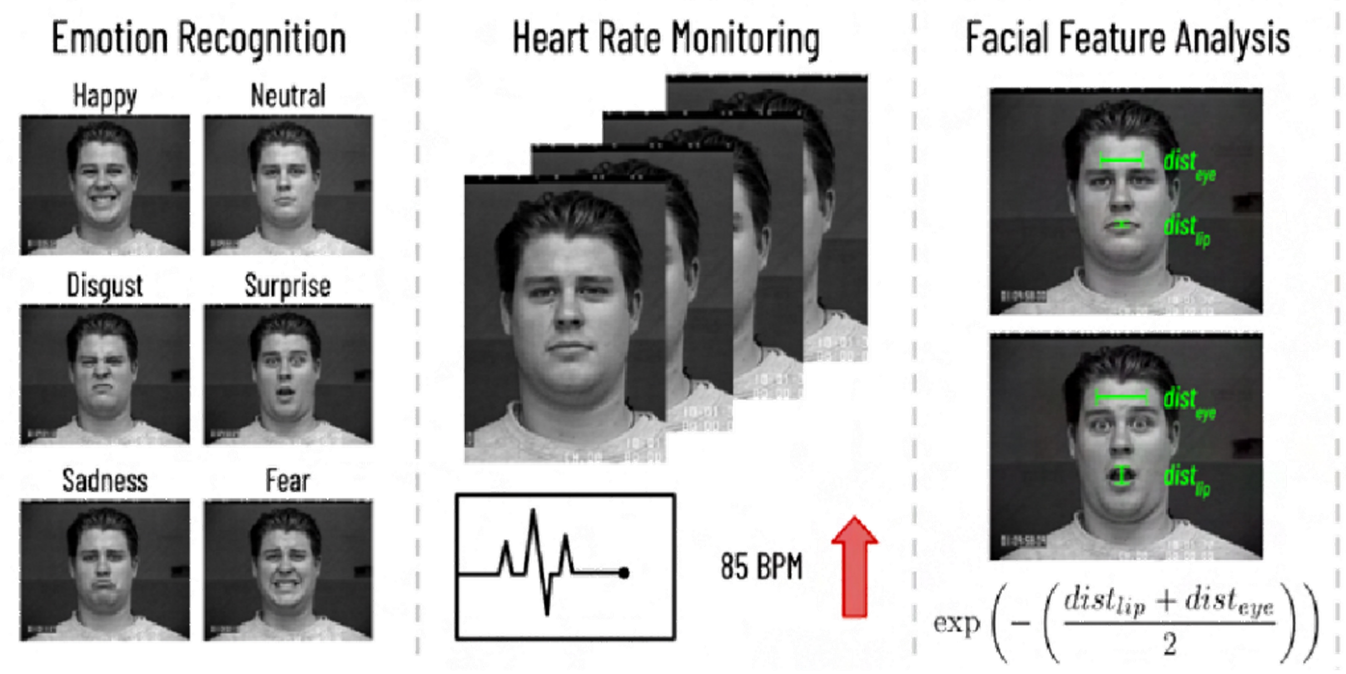
ISBI 2024 ATHENS, GREECE

Motivation

- 4 of 5 adult Americans experience medium to high-stress levels.
- High-stress levels can cause mental and behavioral changes.
- Stress measurements are still conducted through through medical devices or user questionnaires.
- An efficient, automated method to detect stress could help curb such consequences of stress.

Contributions

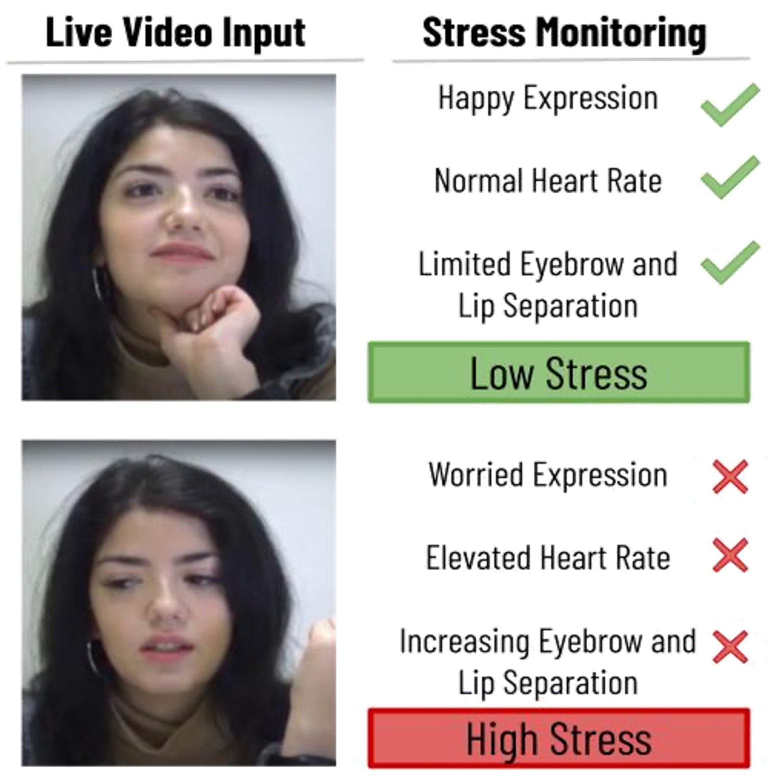
In this study, we propose a framework that achieves rapid, non-invasive stress detection through video.



1. Classify the subject's emotional state by extracting frames from video and labeling the expressions.

2. Remotely measure the user's heart rate with an algorithm that amplifies the slight changes in skin hue.

3. We calculate the distances of specific facial landmarks, such as the eyebrows and lips, in successive frames.

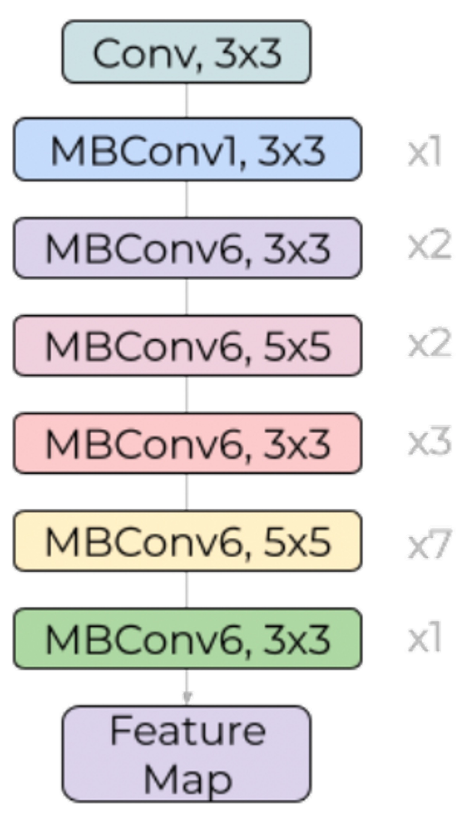


Paper

Github

Methods

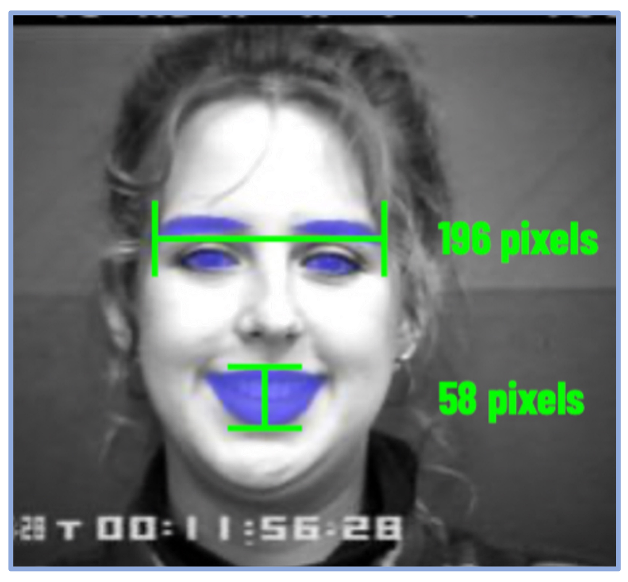
Emotion Recognition



1. Convolutional Neural Network (CNN)
2. VGG-19 with Batch Normalization
3. Support Vector Machine (SVM)
4. Decision Tree Model, and EfficientNet

Heart Rate Detection (EVM)

Facial Feature Detection



Measure the variation of facial features from each image frame.
- Eyebrows
- Lips
- Mouth

$$\mathcal{S}_{FF} = \max\left(\exp\left(-\frac{dist_{lip} + dist_{eye}}{2}\right), 0.85\right)$$

Creating convex hulls for each item, we calculate Euclidean distance between the left/right eyebrows and top/bottom lip.

Integration

$$\mathcal{S} = \begin{cases} \text{low} & \mathcal{S}_{ER} + \mathcal{S}_{HR} + \mathcal{S}_{FF} < 1.2 \\ \text{moderate} & 1.2 \leq \mathcal{S}_{ER} + \mathcal{S}_{HR} + \mathcal{S}_{FF} < 1.8 \\ \text{high} & \mathcal{S}_{ER} + \mathcal{S}_{HR} + \mathcal{S}_{FF} \geq 1.8 \end{cases}$$

$$\mathcal{S}_{ER} = \begin{cases} 0 & \text{if ER} \in \text{happiness} \\ 0.6 & \text{if ER} \in \text{neutral} \\ 0.7 & \text{if ER} \in \text{disgust, sadness} \\ 0.8 & \text{if ER} \in \text{fear} \\ 1.0 & \text{if ER} \in \text{anger, surprise} \end{cases}$$

We use a weighted correspondence between components and their stress levels.

$$\mathcal{S}_{HR} = \begin{cases} 0.4 & \text{if HRV} < 80 \\ 0.7 & \text{if } 80 \leq \text{HRV} < 100 \\ 0.8 & \text{if } 100 \leq \text{HRV} < 120 \\ 1.0 & \text{if HRV} \geq 120 \end{cases}$$

Results

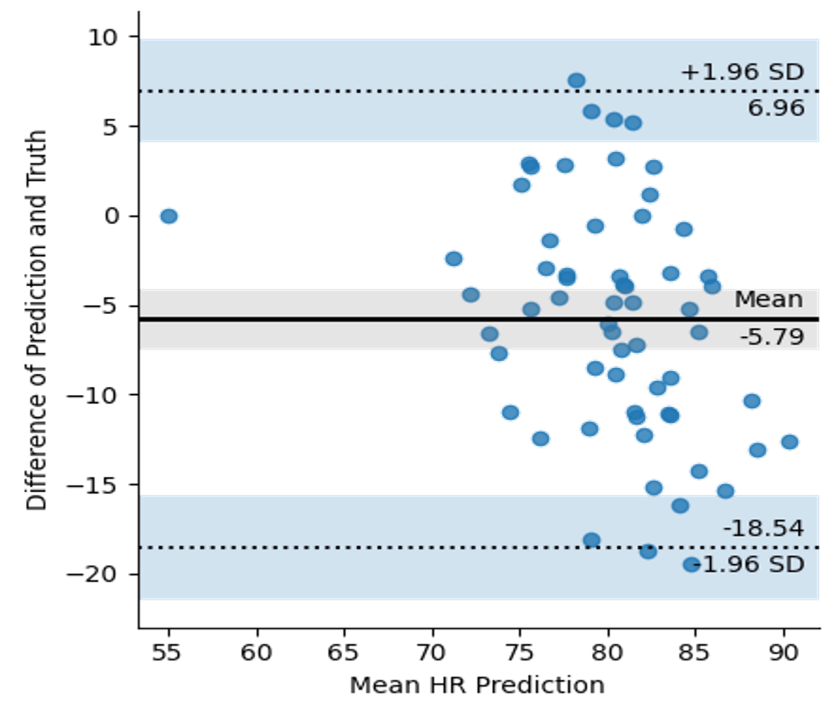
Emotion Recognition

Our framework uses the VGG-19 model, the second highest performance.

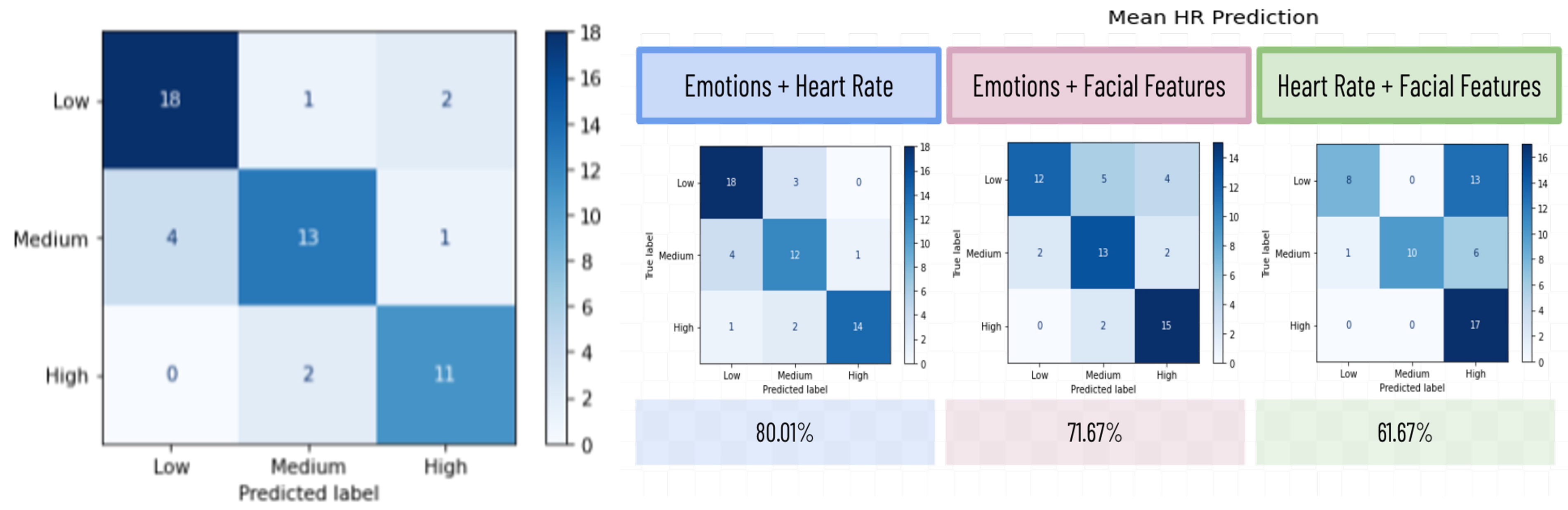
Machine Learning Model	Accuracy (↑)
Linear SVM	84.29
2D CNN	85.50
Decision Tree	80.05
EfficientNet	**98.48**
VGG-19	96.46

Heart Rate Detection

Past Works' Models	MAE (↓)
2SR	12.81
CHROM	11.36
IBIS-CNN	9.39
HR-CNN	8.72
This Study	**5.79**

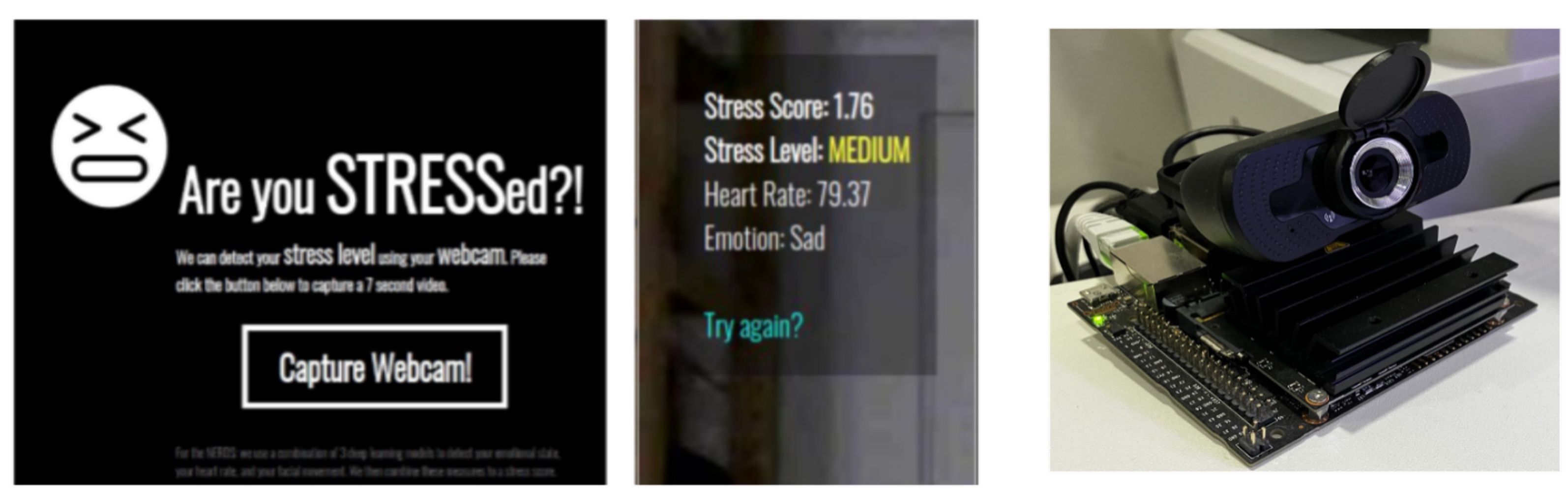


Model and Ablation Studies



The model identifies stress levels with 84% accuracy, reaching up to 90% accuracy when combining moderate/high-stress.

Discussion + Future Work



- Our package is available in a ready-to-use web application.
- Testing an embedded, low-cost device on an NVIDIA Jetson Nano connected to a mini camera.
- Integration of this software with semi-autonomous vehicles could play a role in driver settings.