NEVIN ADA CAKMAK

SEPTEMBER 2024

On Using Physiological Sensors and AI to Monitor Emotions in a Bug-Hunting Game

A scientific paper on understanding emotions, motivating students and software testing.

Abstract, Goal, Motivation

— PART 1

Abstract & Goal

Although software testing is crucial, many students find it boring and stressful—a better understanding of what testers “feel” when learning the skill can be revolutionary.

An innovative approach involving motivational feedback, physiological sensors, and of course, AI can do the trick.

2 experiments were conducted to understand the problem better and develop a realistic solution.

Problem

WHAT THEY WANT TO ACHIEVE

Ways to motivate students to enjoy testing and not give up learning soon to contribute to a more sentiment-aware education.

"DBugIT"

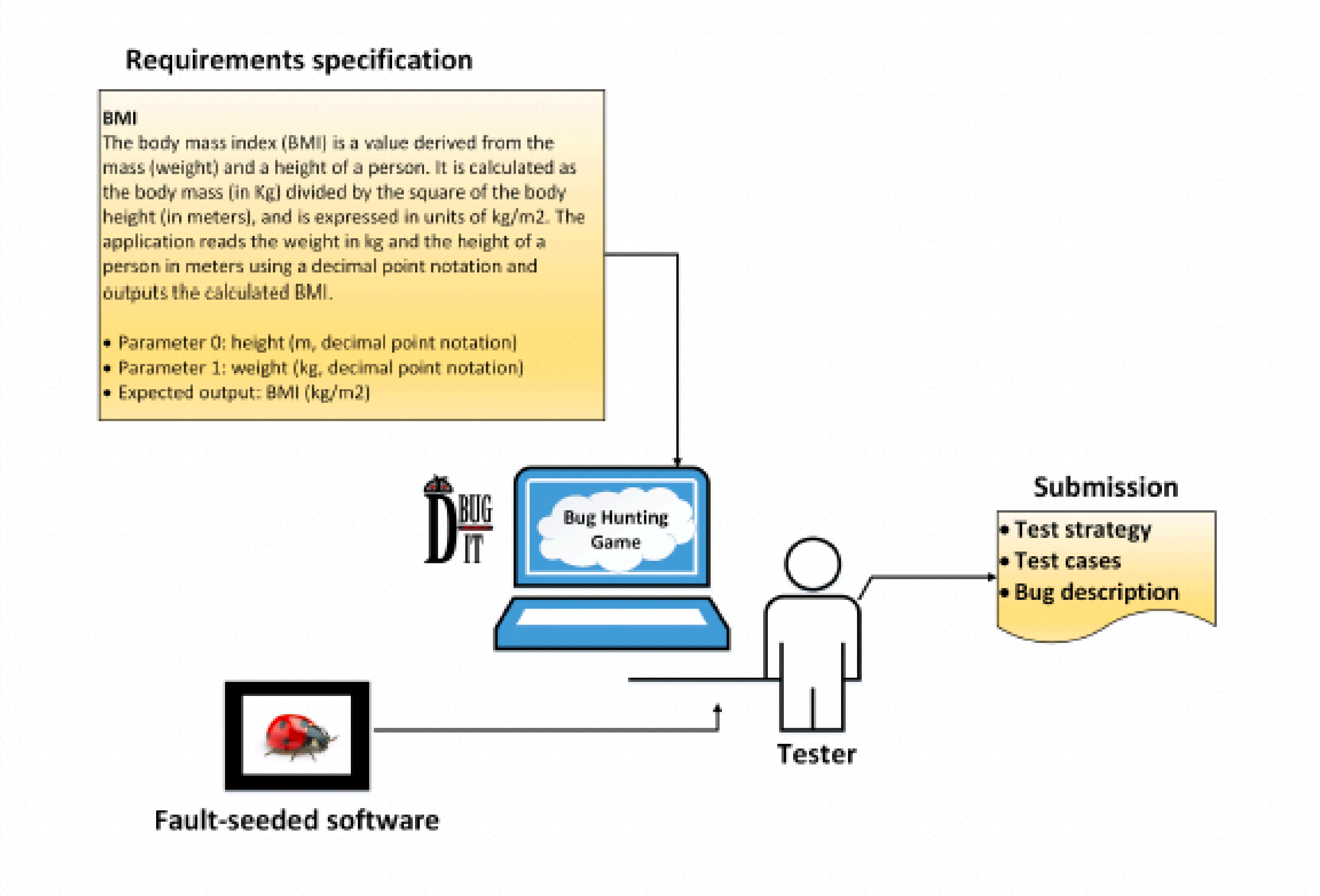
Hypothesis

WHAT THEY WANT TO PROVE OR DISPROVE

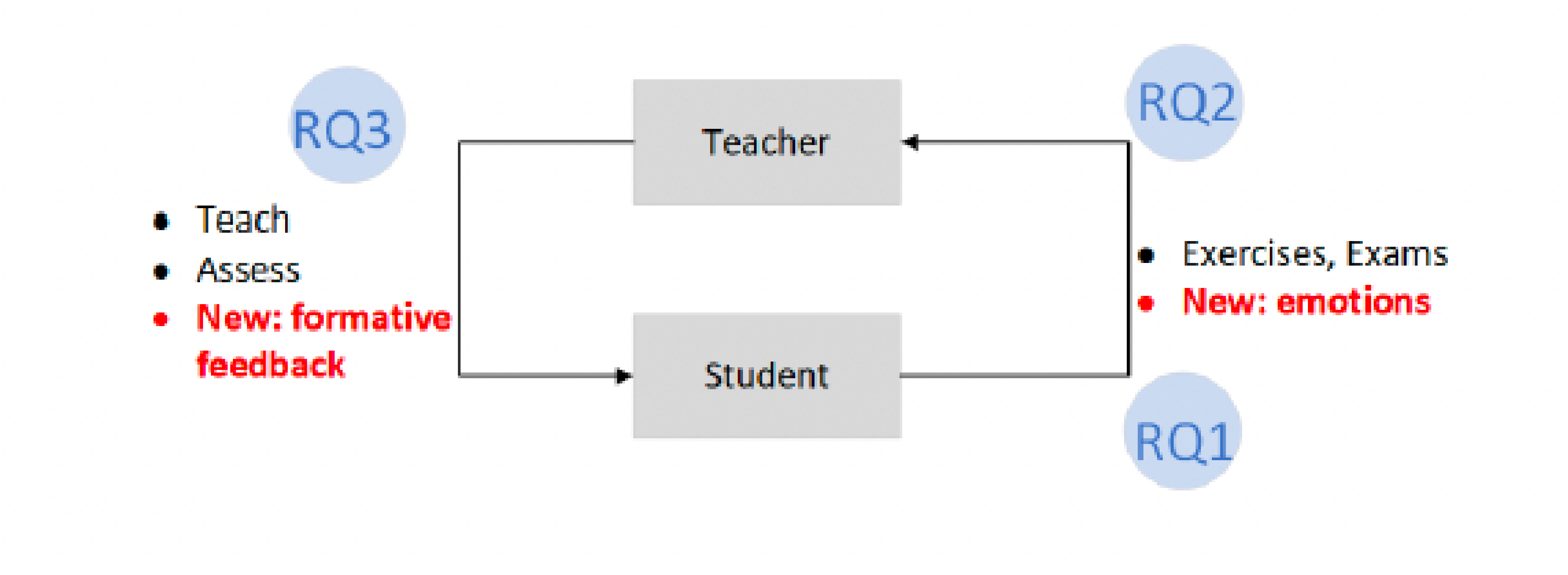
The integration of a game-based learning tool, combined with real-time affective feedback from wearable sensors, will improve students' motivation and performance.

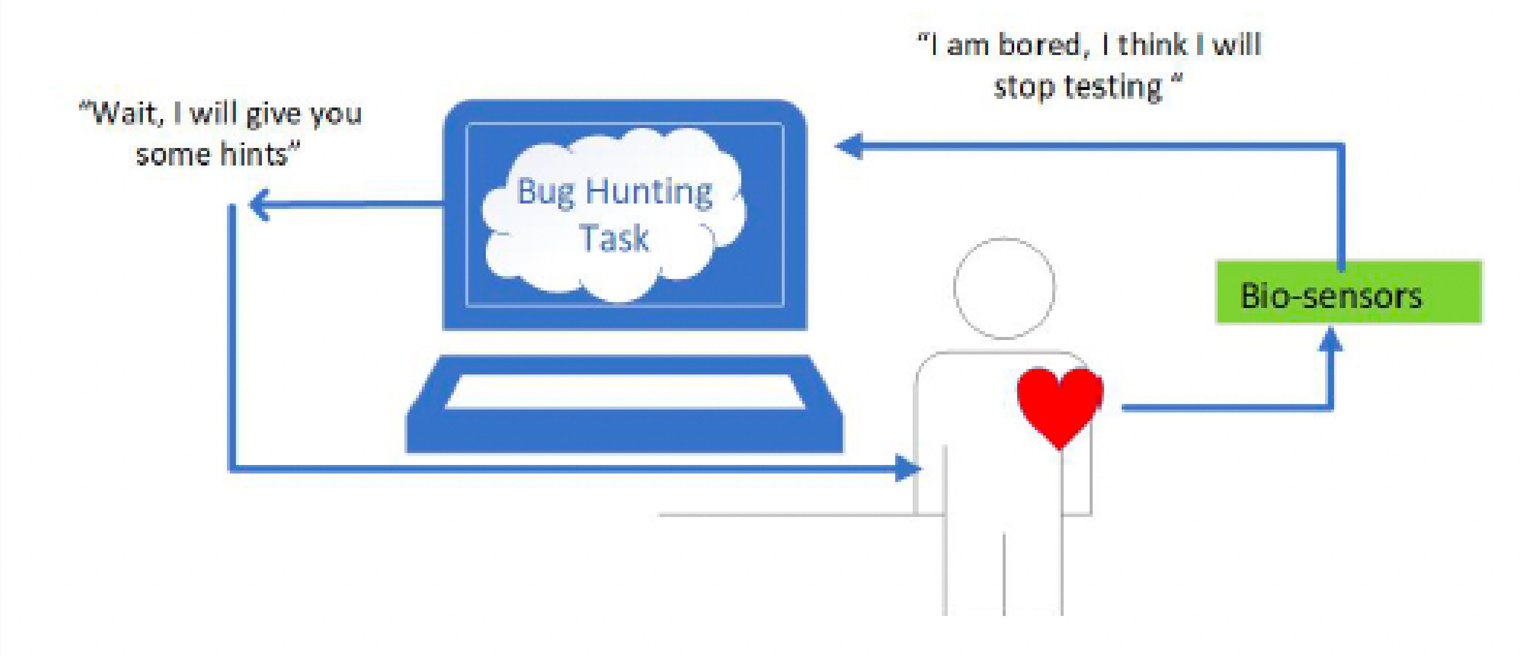
What is a Bug-Hunting Game?

An interactive tool where students find bugs in faulty software. They design tests, run them, and analyze the results to locate and describe the bug. It's a hands-on way to learn software testing, with grades based on how well they plan, test, and report the bug.



Models for the Envisioned Teaching Plan





How to Design Such Models?

RQ1. How to sense the affective state of a tester engaged in a bughunting game?
RQ2. How to classify a tester's emotions based on the sensed biometric data ?
RQ3. How to generate motivating feedback based on a tester's emotions?

A emotion-carrying channel from students towards the teacher, based on physiological sensors, self-reports and machine learning.

Answering RQ3 is the future work, the aim of this study was to answer RQ1 and RQ2.

How to Implement an Automated Emotion Recognition System?

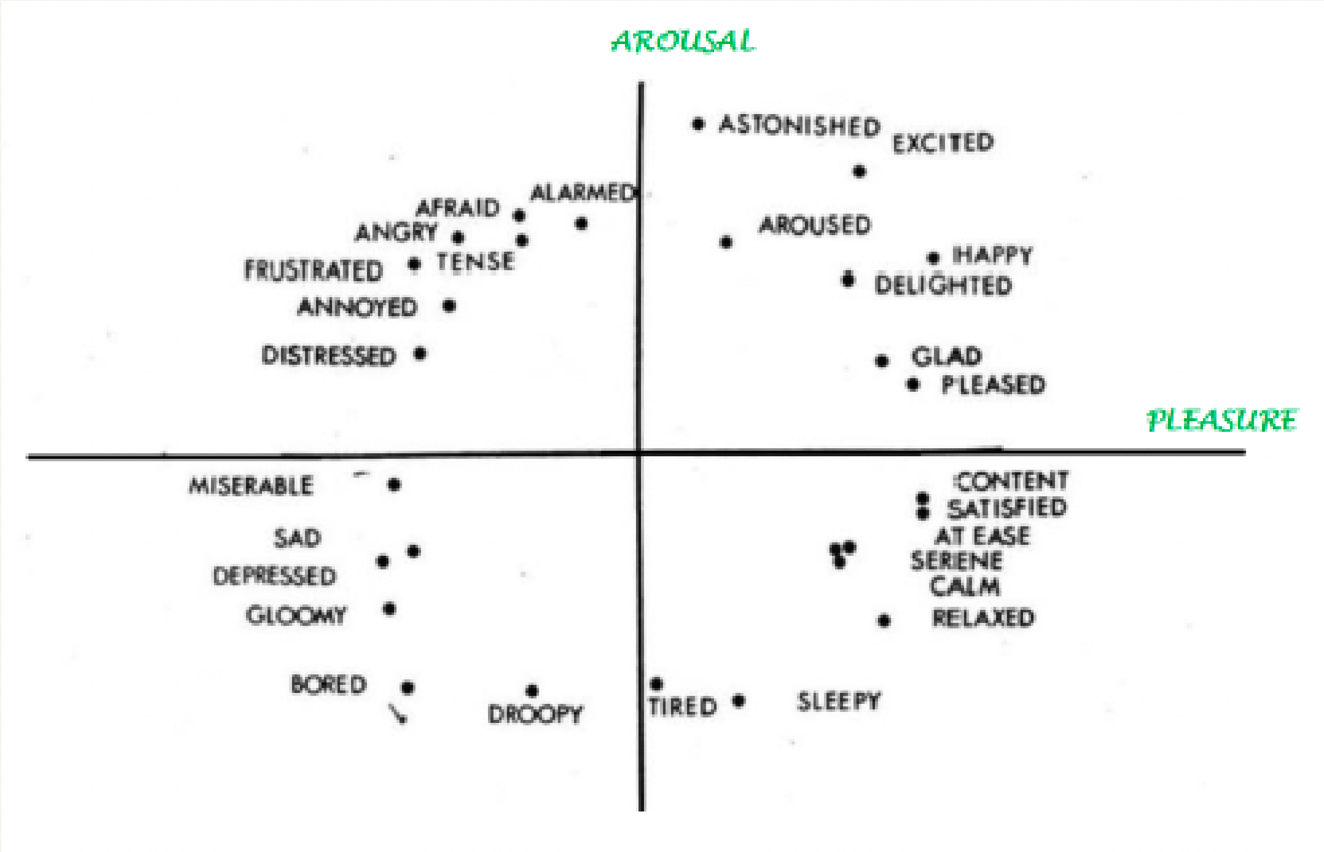
PROBLEM POINTS, AND POTENTIAL SOLUTIONS

SENSORS → **EMOTIONS**

Emotions can be detected through physical signals like facial expressions or body posture using a camera, but this method can be unreliable.

More reliable approaches involve analyzing speech, text, or physiological signals like brain activity (EEG), heart activity (ECG), or skin conductivity (GSR/EDA).

Common emotion models include classifying emotions as positive or negative, or using three factors: pleasure (valence), arousal (intensity), and dominance (control).



How to Implement an Automated Emotion Recognition System?

PROBLEM POINTS, AND POTENTIAL SOLUTIONS

CLASSIFIERS

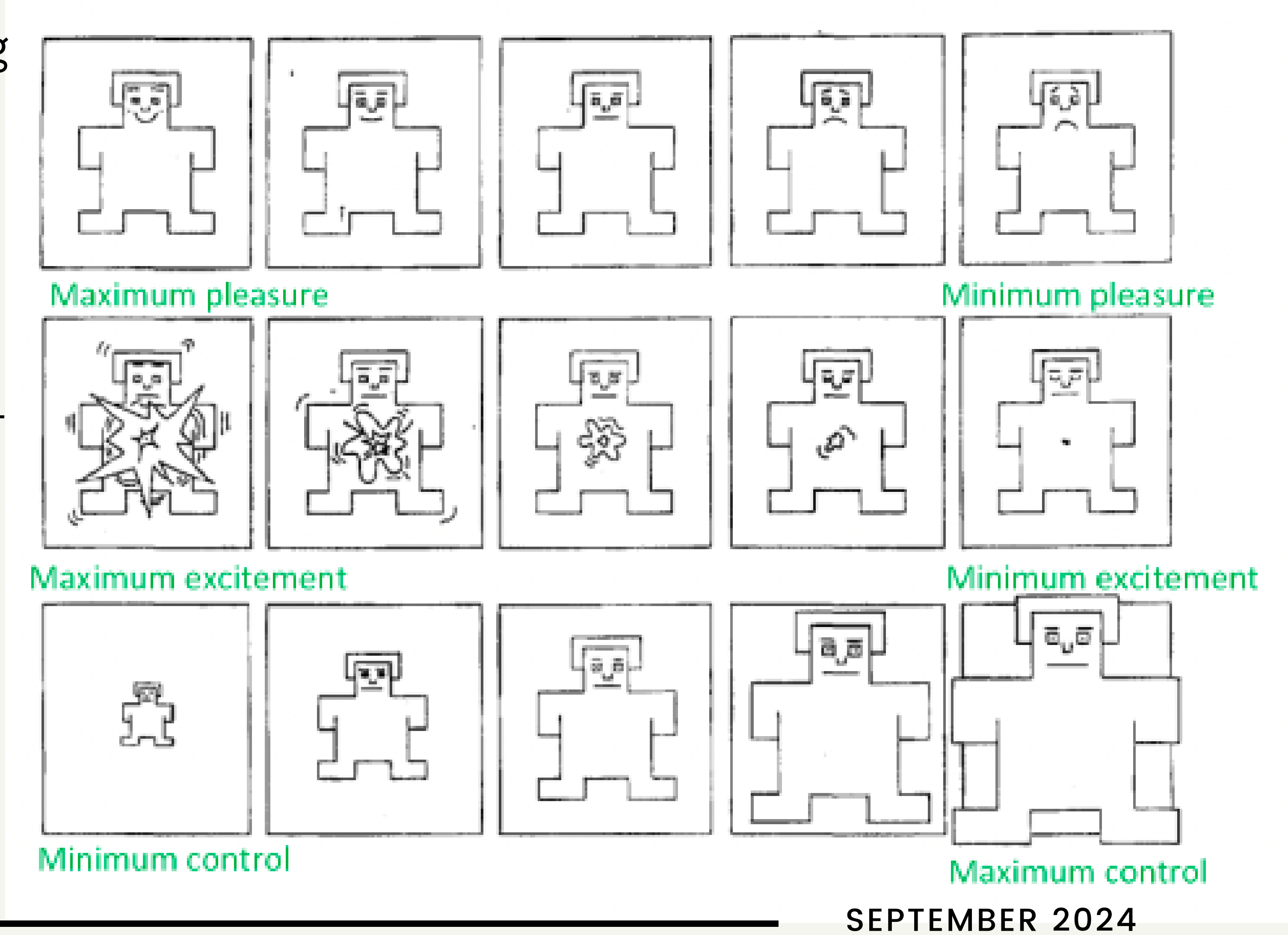
A classifier is an algorithm that categorizes new data into known classes. Simple sentiment analysis can use rule-based classifiers like VADER for social media.

More advanced methods include template matching, probabilistic models like Bayes Networks, and machine learning approaches such as neural networks or deep learning algorithms like CNNs and RNNs.

EVALUATION

The biggest challenge in evaluating sentiment analysis is determining the ground truth of a subject's feelings.

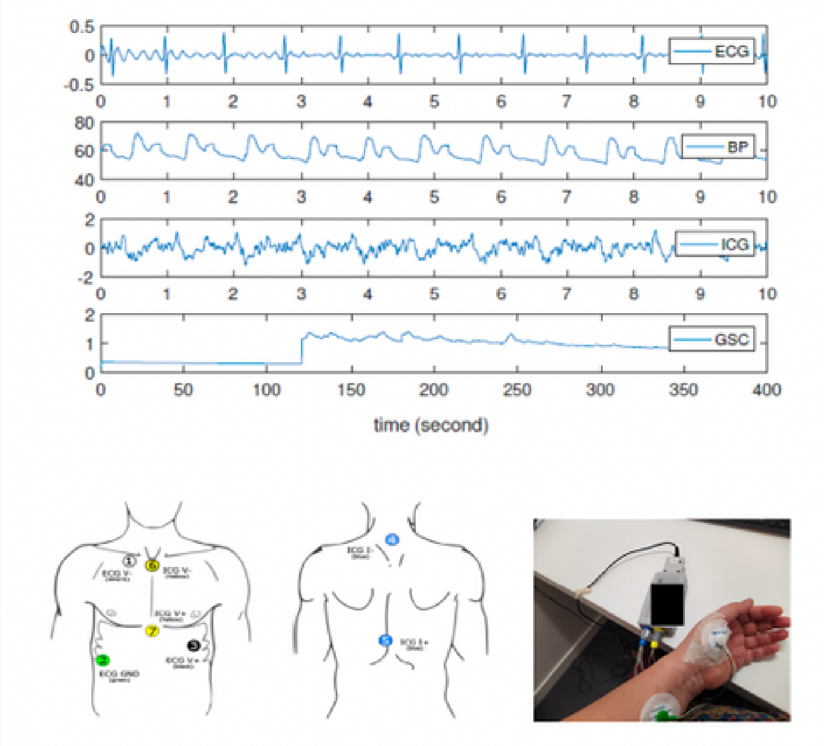
An alternative is the Self-Assessment Manikin (SAM), a non-verbal, picture-based tool that measures pleasure, arousal, and dominance.

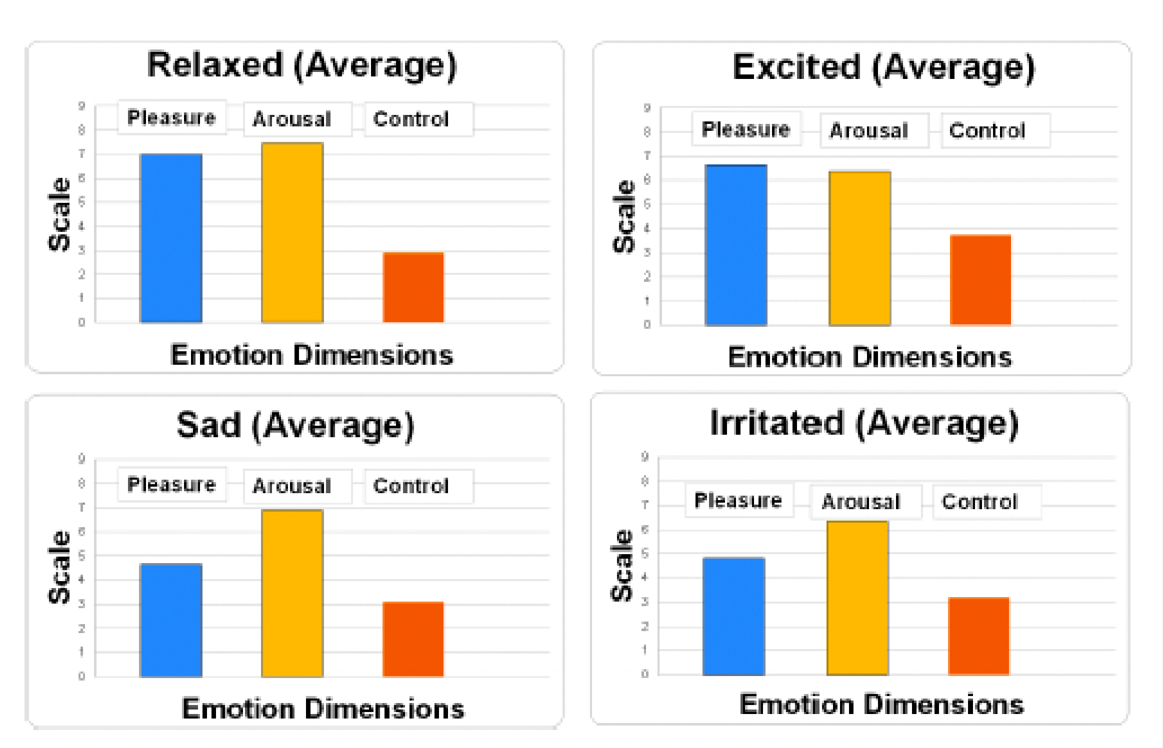


Experiments & Results & ML Aspect

— PART 2

Experiment#1. Monitoring emotions in real-time





01 Eleven CS students volunteered as testers. They wore a VU-AMS device while following a guided script on a computer. Continuous recording of physiological signals (ECG, EDA, ICG).

02 Induced four emotions (irritation, sadness, excitement, relaxation) using IAPS images. Participants completed a 50-minute task in DBugIT. Participants rated their emotional state every five minutes using a SAM pop-up. for the “ground truth”.

03 Results matched intended emotions, but distinguishing emotions based on arousal and pleasure was challenging. Variability in pleasure and arousal during the task was noted, indicating emotional fluctuations.

Experiment#2 Training a deep-learning emotions classifier

01 The experiment aimed to answer RQ2 by applying machine learning algorithms to classify emotions from cardiac activity signals. They used a validated dataset of ECG-EDA-ICG recordings.

02 Participants completed tasks like screen flares, stair climbing, and vacuum cleaning while their physiological responses were recorded. Emotions were labeled, covering four positive emotions and five negative emotions.

03 They trained a deep-learning classifier based on a type of Recurrent Neural Network (RNN) called Long Short-Term Memory (LSTM) Network. Two models were used for emotion categorization: a binary model (positive-negative) and a four-emotion model (anxious, down, enthusiastic, relaxed).

Conclusion, Future Work, Related Work

— PART 3

Relationship to Other Studies

CONNECTION WITH SIMILAR EXPERIMENTS

In particular, Muller et al. and Girardi used EEG signals to get insights into the behaviour of programmers.

Grassi et al. used EDA signals to assist Agile teams in their retrospective meetings.

Fritz used a combination of eye trackers, EEG, EDA and ECG sensors to assess task difficulty in coding.

Vrzakova used a multimodal sensing combination (eye trackers, GSR and pressure sensors) and machine learning to monitor the emotions of participants in a code review session in a large company.

All authors reported promising results, with an accuracy of around 85%.

Results and Conclusion

OUTCOMES OF THE STUDY

Two experiments confirmed that it is possible to monitor the emotional state of testers working on a bug-hunting task using a multimodal (ECG, EDA and ICG) combination of physiological sensors, and that deep-learning algorithms can be trained to make sense of the raw biometric data.

A multidisciplinary approach involving computer science, biological psychology, and education expertise will be needed to turn this idea into success.

Areas of Improvement

SUGGESTIONS FOR FUTURE RESEARCH

Optimizing the emotion recognition process, involving a larger number of participants, increasing the validity of self-reported ground-truth emotion labels, adding less-intrusive sensors to increase scalability, and exploring ways to generate emotion-aware motivating feedback.

On the long term, the approach can become interesting for other educators who experiment with sentiment analysis or capture-the-flag (CTF) style learning initiatives and gamification.

Thank you!

QUESTIONS? COMMENTS? FEEDBACK?