Data Mining Project

a multimodal sensor dataset for continuous stress detection of nurses in a hospital

29 january 2025

Team Members

Chau Nguyen
chau.nguyen@student.oulu.fi

Hans Karawitage
Hans.Madalagama@student.oulu.fi

Ata Jodeiri Seyedian
ata.jodeiri@student.oulu.fi

Fatemeh Soufian
Fatemeh.Soufian@student.oulu.fi
Leader

Introduction

Background

Stress in the workplace:

- Can cause negative impact on employee well-being and productivity.
- Leads to long-term health risks due to high workload and long hour shifts.

Importance of monitoring stress:

- By constantly monitoring through physiological signals/surveys can identify causes.
- Analyze and derive conclusions on the collected data to minimize workload and stress.

Introduction

description of the project

- The data contains multimodal physiological signals gathered from a wearable device attached to the regular working female nurses in the nursing department in a hospital.
- A self-reported survey results completed daily by the nurses after their regular shift ends.
- The data collection is done during the second wave of the COVID-19 outbreak.

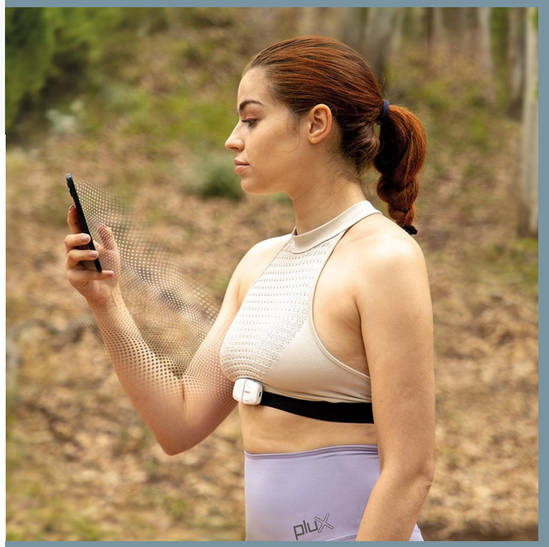
❑ **Our aim is to analyze these collected data, pre-process and extract valuable information.**

Related Work

- The **AffectiveRoad** dataset studies the effect of driving conditions on stress levels.
    - Data Collected: From 10 drivers during a 1-hour 26-minute driving test.
    - Devices Used: Empatica E4 and Zephyr Bioharness 3.0.
    - Signals Monitored: Electrodermal activity, heart rate, skin temperature, respiration, hand movement.
    - Purpose: Develop context-aware systems for monitoring driver stress and arousal.

Related Work

- The **WESAD** dataset is a multimodal dataset for wearable stress and affect detection.
    - Data Collection: Recorded from 15 participants during a lab study, including activities such as watching a movie and taking the TSST test.
        - The Trier social stress test (TSST) is a laboratory procedure used to reliably induce stress in human research participants.
    - Devices Used: RespiBAN and Empatica E4.
    - Signals Monitored: Blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration.
    - Purpose: To analyze three affective states: neutral, stress, and amusement, bridging the gap between lab studies on stress and emotions.

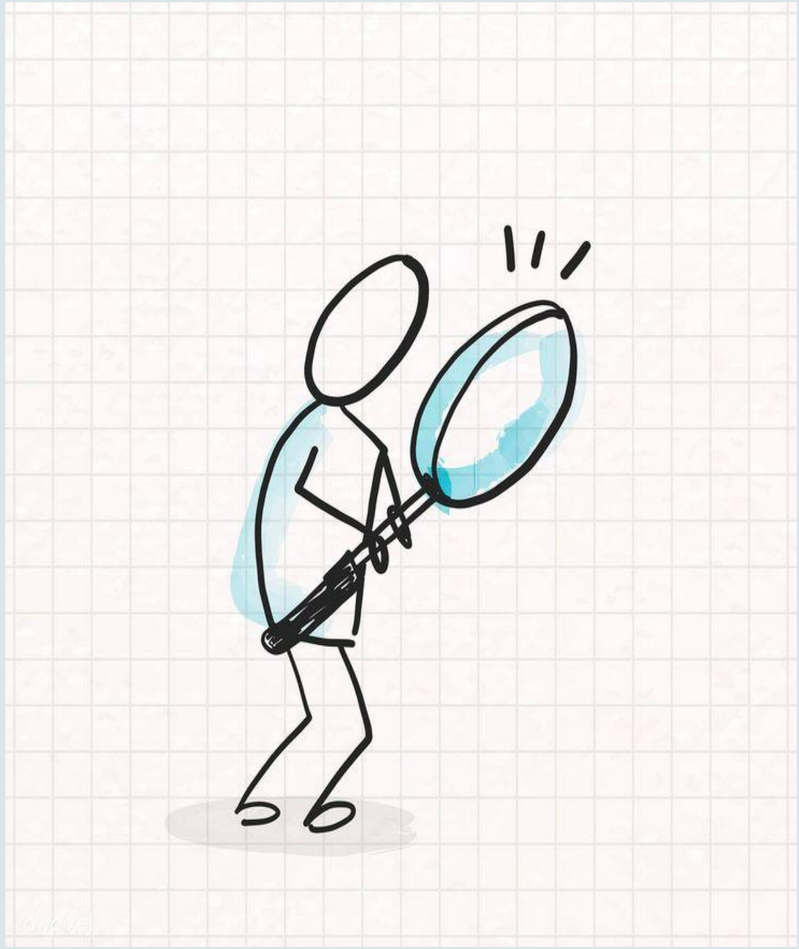


Objectives of the Research

Research Questions and Expected Results

? How do physiological signals correlate with self-reported stress levels among nurses during the COVID-19 pandemic?

- *Physiological signals, such as heart rate (HR), heart rate variability (HRV), and electrodermal activity (EDA), are expected to show significant correlations with self-reported stress levels.*
- *Increased stress may manifest as elevated HR, reduced HRV, higher EDA, and changes in skin temperature.*
- *Overall: The study is expected to validate physiological signals as reliable markers for real-time stress monitoring.*



Objectives of the Research

Research Questions and Expected Results

? What are the primary contextual factors(e.g., COVID-related challenges, workload, and patient crises) that contribute to stress among nurses, and how are these factors reflected in physiological data?

- *The stressors (e.g., COVID-related challenges, workload, and patient crises) are likely to manifest physiologically through elevated heart rate (HR), reduced heart rate variability (HRV), and increased electrodermal activity (EDA).*

- *Overall: The findings are expected to validate the impact of these stressors and offer actionable insights for developing targeted interventions to reduce stress in healthcare environments.*

Description of the used dataset

- The dataset capture biometric data from nurses (heart rate, skin temperature, electrodermal activity, accelerometer, Inter-beat interval, blood volume pulse) and associated context pertaining to the stress events (type of tasks, coworker, workload, patient families,...).
- Time range:
  - 2 study sessions 1 week long: Apr-May 2020 and Nov-Dec 2020
  - In total: 1,250 hours of data
- Participant: 15 female nurses
- Device: Empatica E4 wearable
- Delivery and storage
  - Data from the device sensors are collected and stored to 6 CSV files
  - Survey is done everyday to identify stress type and stored to 1 XLSX file

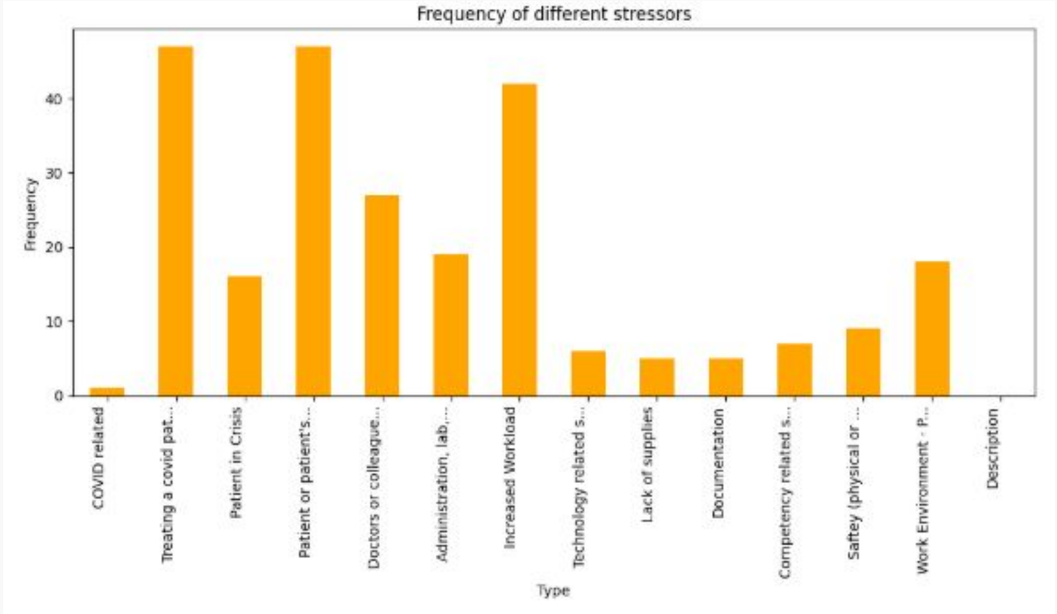
Dataset Source: https://doi.org/10.5061/dryad.5hqbzkh6f

Visualization

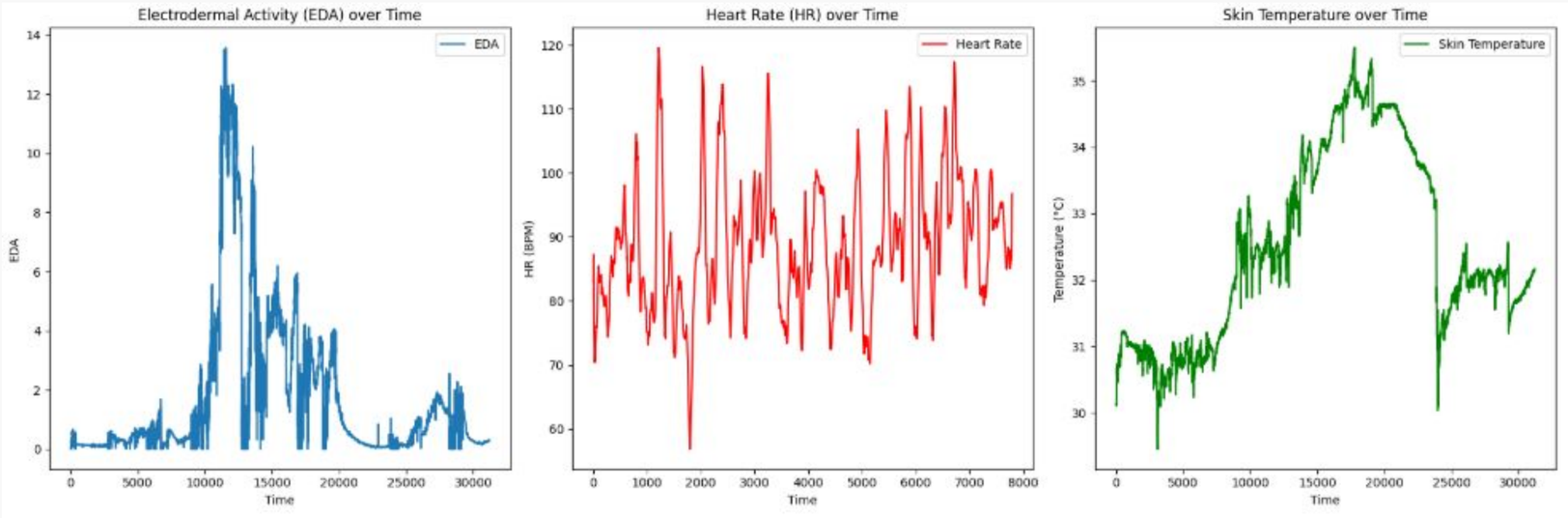
- Empatica E4 Signals and Frequency chart

Signal	Abbreviation	Frequency (Hz)
Electrodermal Activity	EDA	4.0
Heart Rate	HR	1.0
Skin Temperature	ST	1.0
Accelerometer	ACC	32.0
Inter-Beat Interval	IBI	64.0
Blood Volume Pulse	BVP	64.0

- Survey responses to contributing factors of stress

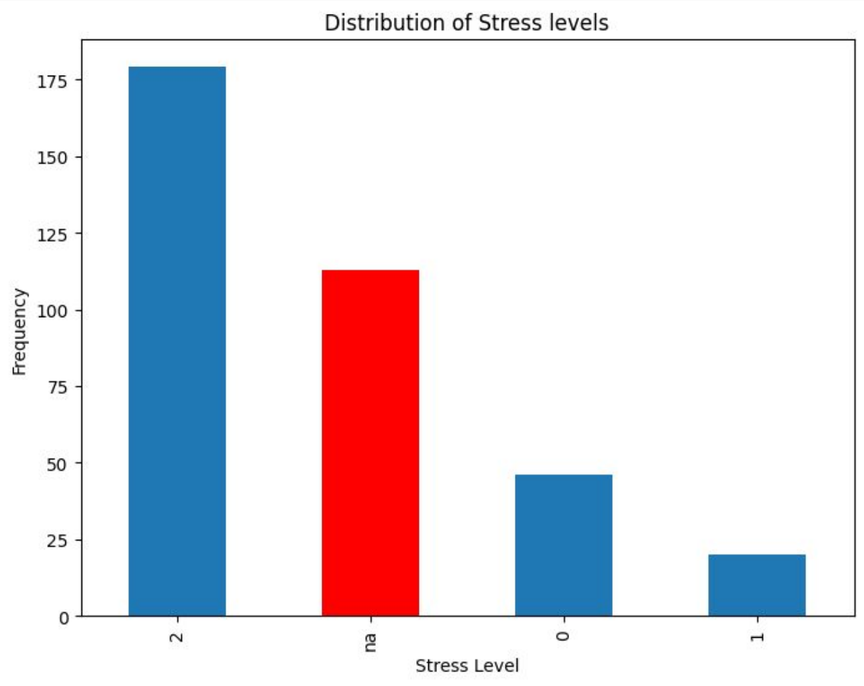


- Physiological data distribution (EDA, HR and ST) of a single subject over time

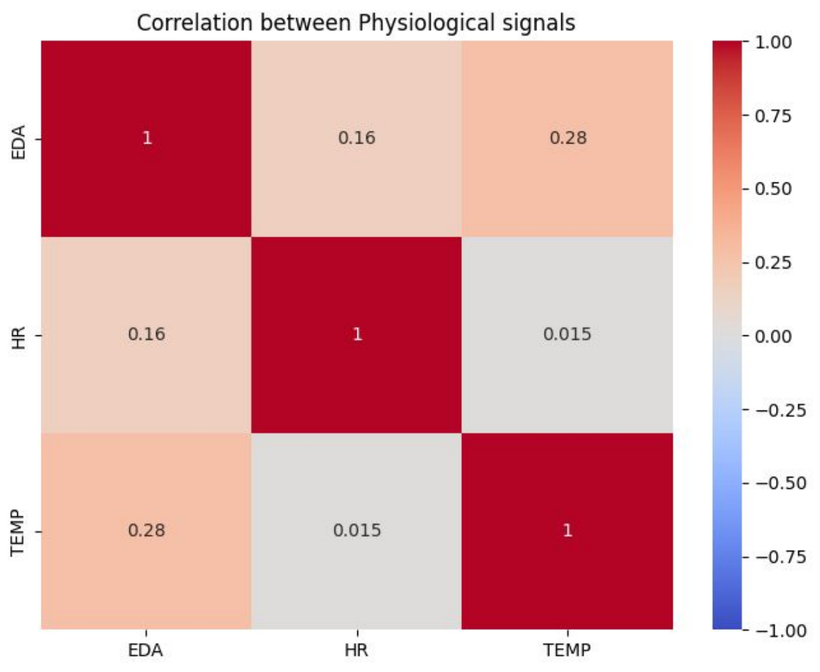


Challenges

- Good ol' missing data



- Low correlation pre-analyzing



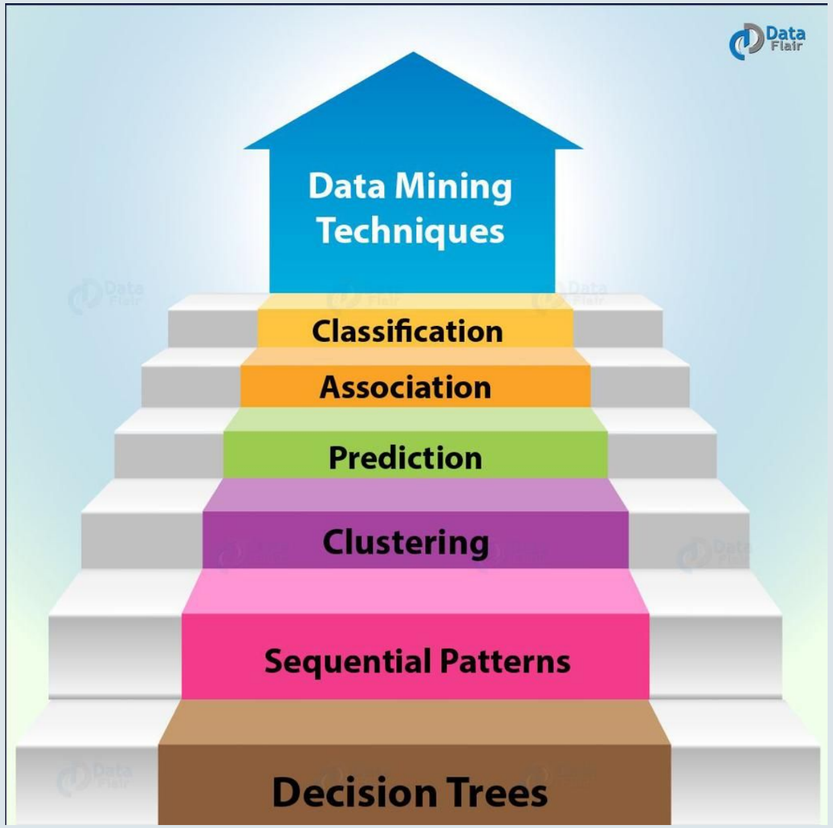
- Noise and outliers

Searching.....

Methods

Initial plan for methods to be used

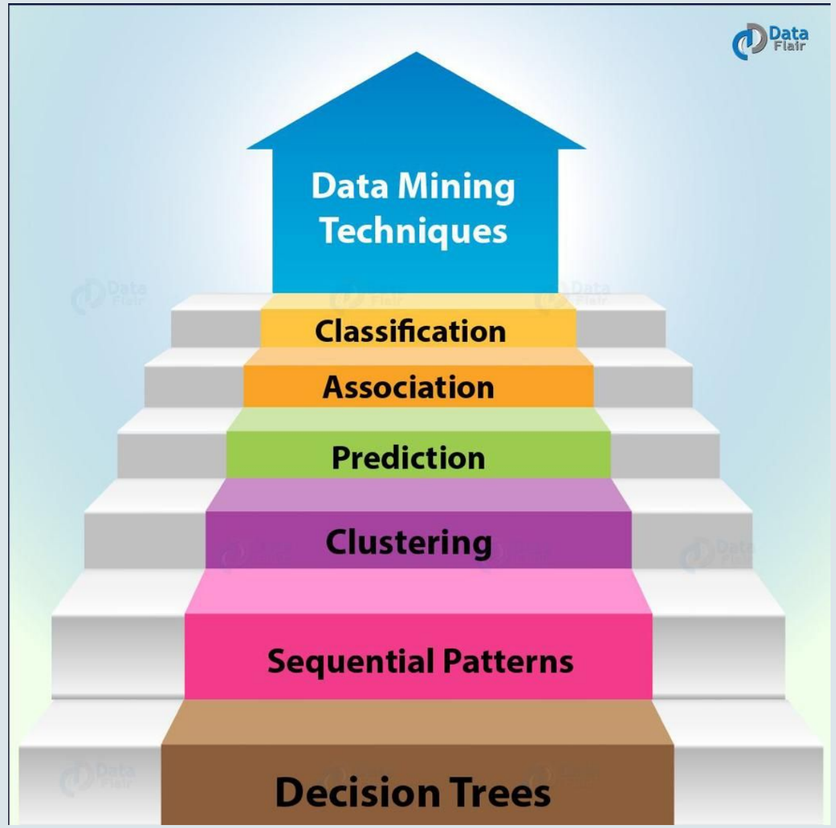
- *Preprocessing:*
  - *Data Integration:*
    - *Per Nurse: Merge all CSV files for each nurse, compute timestamps using the starting timestamp and sampling intervals, and unify data based on timestamps.*
    - *All Nurses: Combine datasets for all 15 nurses into a master dataset using a composite key (NurseID + Timestamp) for unique records.*
  - *Synchronize Units: Convert survey dates to match the timestamp format in sensor data.*
  - *Data Cleaning:Handle missing values and remove duplicate rows from merging or other processes.*



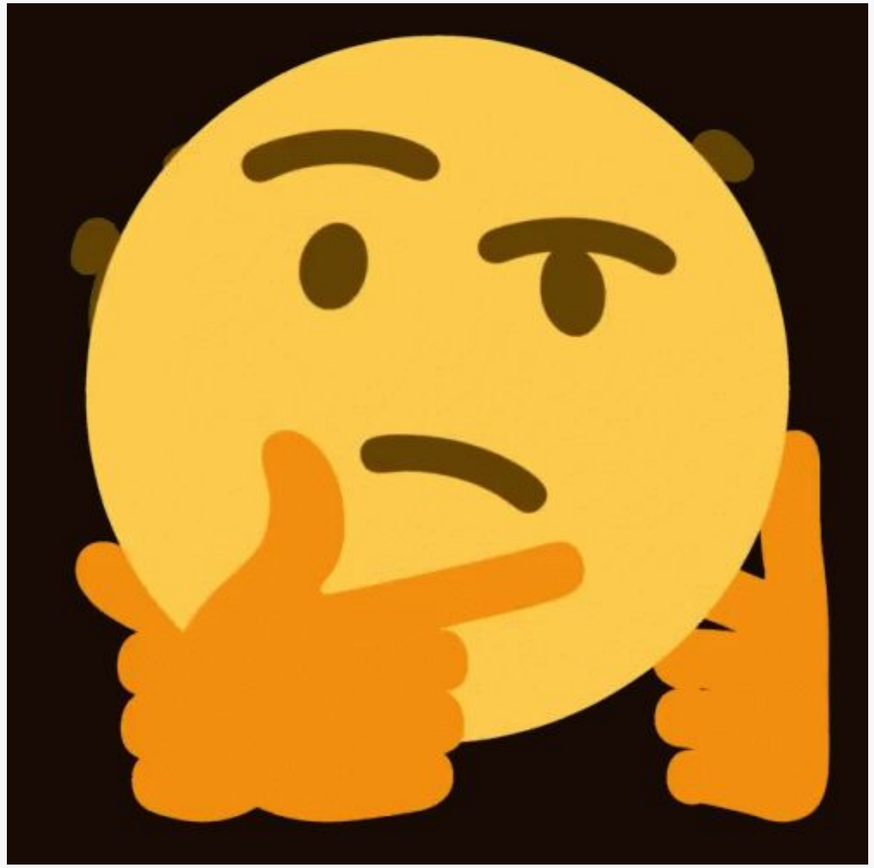
Methods

Initial plan for methods to be used

- *Statistical Analysis:*
  - *Correlation Analysis: With Pearson or Spearman correlation => relationships between physiological signals and stress.*
  - *Regression Analysis: Apply linear regression or generalized additive models GAMs => predict stress scores from physiological signals.*
- *Visualization*



Questions ?



Thank you