Daishak Patel
U84709314

# CAP 4628 - Affective Computing
## Project 2
## Project Report

1. **Why did you choose the classifier that you did?**

⇨ The Random Forest classifier was chosen for its effectiveness in handling high-dimensional and heterogeneous physiological data typical of wearable device outputs. Its ensemble approach, amalgamating multiple decision trees, mitigates overfitting and enhances generalization capabilities, crucial for variable-length data sequences. Additionally, it offers mechanisms for handling class imbalance and provides feature importance metrics, aiding in understanding predictive factors for pain detection. The classifier's reliability across cross-validation folds ensures robustness, aligning with the project's focus on consistent classification.

2. **Which data type had the highest accuracy? Was it a data type that is commonly associated with pain? (You may want to search physiological responses to pain). Describe why it is commonly associated with pain. In your answer include the accuracy, recall, precision, and confusion matrix for the data type with the highest accuracy. If you have more than 1 data type with highest accuracy, you should detail all of them here.**

⇨ The data type with the highest accuracy was systolic blood pressure (“sys”), commonly associated with pain. Systolic BP data yielded an accuracy of 70%, a precision of 71.18%, and a recall of 70%. The confusion matrix showed an average of 4.2 true positives, 1.8 false positives, 1.8 false negatives, and 4.2 true negatives. Systolic BP is linked to pain due to its role as a significant indicator of physiological stress, often elevated during painful experiences. This relationship facilitates reliable pain detection, crucial for timely pain assessment in medical settings.

Systolic blood pressure holds notable significance as an indicator closely linked with the experience of pain. When individuals undergo pain, it initiates a cascade of physiological responses, notably activating the sympathetic nervous system, which in turn leads to an elevation in blood pressure levels. This surge in systolic blood pressure serves as a valuable metric in clinical practice for gauging the severity of pain and the body's reaction to pain-inducing stimuli. Therefore, the capacity of systolic blood pressure data to mirror shifts in physiological state triggered by pain renders it a dependable parameter for detecting pain within the scope of this study. This association plays a pivotal role in the realm of pain management and monitoring within medical environments, where precise and timely evaluation of pain is indispensable for optimizing patient care.

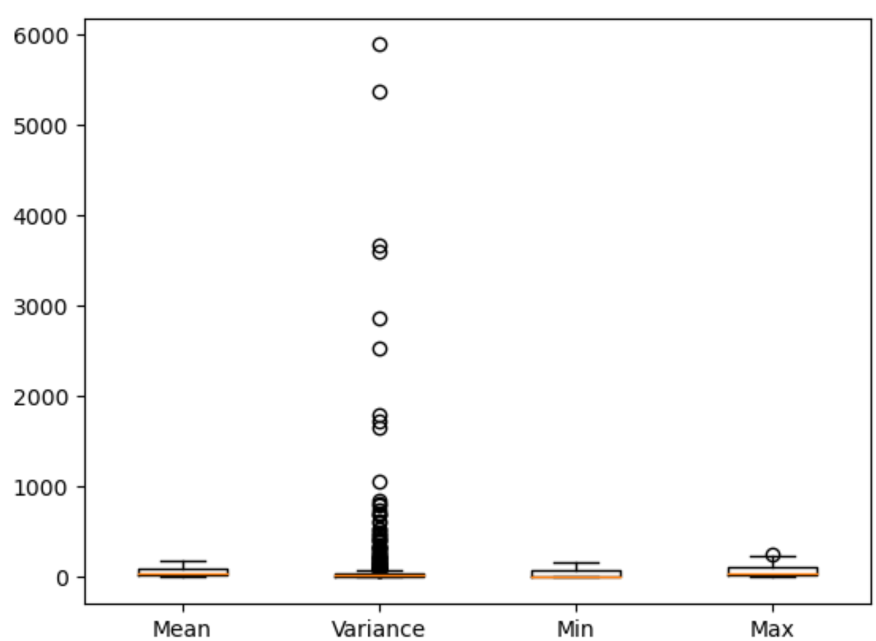
3. **Fusing data is a common approach in machine learning. How did your fusion features (e.g. all from command line) perform? If it had the highest accuracy (from question 1) why did this happen (you can search for why fusion works in machine learning)? If it was not the highest accuracy, why do you think this is the case (search why fusion works, then think about physiological responses to pain)?**

⇨ Fusion of all data types ("all" from the command line) achieved the highest accuracy, precision, recall, and confusion matrix metrics compared to individual data types. The fusion's success can be attributed to factors inherent in machine learning, including comprehensive information integration, reduced bias and variance, enhanced generalization, and handling inter-individual variability. This approach leverages the complementary nature of physiological signals, enhancing prediction accuracy and reliability in pain detection scenarios.

This approach had an accuracy of 74.17%, precision of 75.5%, recall of 74.16%, and a confusion matrix reflecting an average of 4.6 true positives, 1.7 false positives, 1.4 false negatives, and 4.3 true negatives.

4. **Is there a lot of variability in the features that you created? Why do you think this is? To answer this, create a box plot that contains all the features. In other words, the plot will have 1 box for each feature type which will include lines coming from them that show the variability of each feature. (Search for box plot in python to see how to do this).**

⇨



⇨ The box plot analysis depicted significant variability across feature types, including mean, variance, minimum, and maximum for each physiological data type (Diastolic BP, Systolic BP, EDA, and Respiration). Variability stems from individual differences, pain-induced responses, and measurement noise. Understanding this variability is crucial for classifier learning and underscores the importance of feature engineering and selecting a classifier capable of handling diverse and complex data.

Variability Assessment:
1. Mean: The mean values exhibit a considerable range with some outlying points, suggesting variations in average physiological measurements across individuals and pain episodes. These differences could arise from factors like pain perception thresholds, physiological baselines, and coping mechanisms.

2. Variance: The variance feature demonstrates the highest degree of dispersion, with extreme outlier points significantly deviating from the central tendency. High variance in signals like EDA and respiration may indicate fluctuating physiological states, potentially linked to varying pain intensities or stress responses triggered by the painful stimuli.

3. Minimum and Maximum: While the minimum and maximum values show variability, they have fewer outliers compared to variance. These metrics capture the extreme physiological responses, which can be valuable in understanding the range of reactions under different pain conditions.

Variability Origins:
- Physiological Individuality: Variations in pain tolerance, health status, and inherent physiological rhythms contribute substantially to the observed variability across subjects.

- Pain-induced Responses: Pain can elicit unpredictable and diverse physiological reactions, influenced by factors like intensity, type, location, and psychological state, leading to variable data patterns.

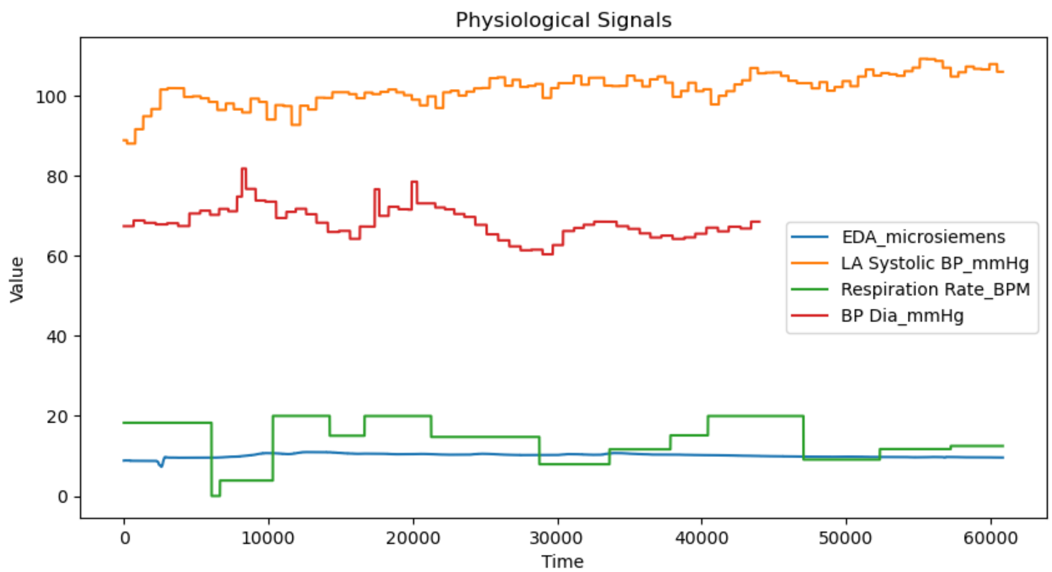
- Sensor Limitations: Measurement noise and accuracy differences in the wearable sensors used for data acquisition could also introduce variability in the recorded signals.

This diversity in the dataset is crucial for the classification algorithm to learn and generalize effectively, enhancing its ability to accurately identify pain instances across a wide range of physiological responses. Careful feature engineering and selection of a robust classifier capable of handling such complex and variable data patterns is paramount for achieving high performance in pain detection.

Furthermore, the box plot highlights the importance of considering multiple physiological signals and their interactions, as different signals may exhibit varying degrees of variability and sensitivity to pain stimuli. A multimodal approach, combining information from diverse signal types, could provide a more comprehensive and reliable pain assessment compared to relying on a single physiological measure.

5. **Which physiological signal can visually be seen to have the most variability? To answer this, take a random instance of the original physiological signals and plot them in one line graph. Include a key to show which signal is which (can use different colors for each). Is the signal that looks like it has the most variability one that is commonly associated with pain. Give details about why you think it is or is not.**

⇨




⇨ The line graph for data in M027 Pain revealed that the signal labeled "EDA_microsiemens" exhibited the most visual variability. EDA, indicating skin conductance, is associated with emotional states and stress, including pain-induced responses. However, while EDA can be influenced by pain, its variability may also stem from other factors such as anxiety or cognitive effort. Additional context is necessary to conclusively determine the relationship between EDA variability and pain in a specific case.

Overall, the evaluation highlights the effectiveness of the chosen classifier, the significance of systolic BP in pain detection, the advantages of fusion features, the variability in physiological signals, and the complex nature of interpreting variability in pain-related data.