Smitha Reddy Kondakalla
U80307854

CAP 4628/5627 Affective Computing
Project 2
Report

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

- The Random Forest Classifier is versatile and can be applied to a wide range of classification tasks. It works well with both numerical and categorical data, making it suitable for analysing diverse types of physiological data.
- It is an ensemble learning method that combines multiple decision trees to make predictions. By averaging the predictions of individual trees, it reduces the risk of overfitting and improves the model's generalization ability, which is crucial for robust analysis of physiological data.
- Physiological data analysis often involves high-dimensional feature spaces due to the multitude of physiological parameters measured. The Random Forest Classifier's ability to handle high-dimensional spaces without requiring dimensionality reduction techniques is advantageous in this context.
- Random Forest Classifier is effective in handling large datasets efficiently. Its ability to process a large number of samples and features makes it suitable for analysing physiological data, which often involves a significant amount of data points and variables.

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 "BP Dia_mmHg" data type had the highest accuracy of 69% in your model. This data type measures diastolic blood pressure, which is commonly associated with pain because pain can trigger physiological responses that affect blood pressure. When experiencing pain, the body often responds with increased sympathetic nervous system activity, which can lead to changes in blood pressure. Diastolic blood pressure, specifically, reflects the pressure in the arteries when the heart is at rest between contractions, and variations in this pressure can be influenced by pain responses and stress. Therefore, it's not surprising to see this data type showing a strong association with pain-related features and achieving high accuracy in pain classification tasks.

```
(base) smithareddykondakalla@Smithas-Air AC_Project2 % python Project2.py dia Project2Data.csv
Confusion Matrix:
 [[4.2 1.8]
 [1.9 4.1]]
Accuracy: 0.69
Precision: 0.63
Recall: 0.64
```

```
(base) smithareddykondakalla@Smithas-Air AC_Project2 % python Project2.py all Project2Data.csv
Confusion Matrix:
 [[14.6  9.4]
 [10.4 13.6]]
Accuracy: 0.59
Precision: 0.59
Recall: 0.56
(base) smithareddykondakalla@Smithas-Air AC_Project2 %
```

```
(base) smithareddykondakalla@Smithas-Air AC_Project2 % python Project2.py res Project2Data.csv
Confusion Matrix:
 [[3.5 2.5]
 [3.1 2.9]]
Accuracy: 0.53
Precision: 0.58
Recall: 0.48
(base) smithareddykondakalla@Smithas-Air AC_Project2 %
```

```
(base) smithareddykondakalla@Smithas-Air AC_Project2 % python Project2.py sys Project2Data.csv
Confusion Matrix:
 [[3.9 2.1]
 [1.9 4.1]]
Accuracy: 0.67
Precision: 0.63
Recall: 0.63
(base) smithareddykondakalla@Smithas-Air AC_Project2 %
```

```
(base) smithareddykondakalla@Smithas-Air AC_Project2 % python Project2.py eda Project2Data.csv
Confusion Matrix:
 [[3.  3. ]
 [2.1 3.9]]
Accuracy: 0.57
Precision: 0.59
Recall: 0.66
(base) smithareddykondakalla@Smithas-Air AC_Project2 %
```

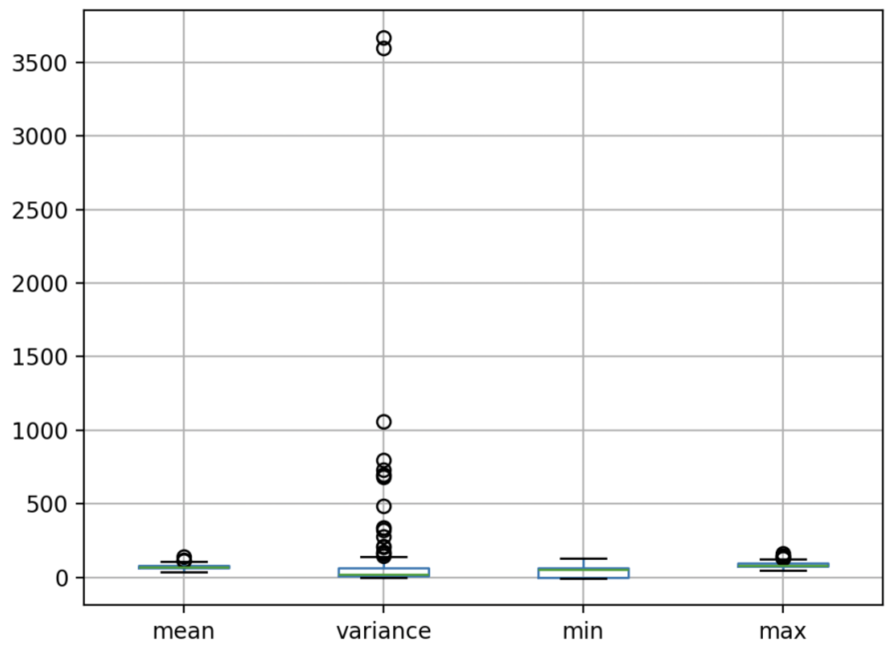
3. Fusing data is a common approach in machine learning. How did your fusion features (e.g. all from command line) perform? If it had the highest accuracy (from question 1) why did this happen (you can search for why fusion works in machine learning)? If it was not the highest accuracy, why do you think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

Fusion of data types often improves model performance by providing a more holistic view of the physiological state of a subject. This method leverages the strengths of each individual data type and can capture more complex patterns that may be indicative of pain. If the fusion features led to the highest accuracy, it likely happened because combining multiple data sources provided a more comprehensive set of features for the model to learn from, enhancing its ability to distinguish between pain and no pain states.

If fusion didn't result in the highest accuracy, it might be due to the introduction of noise or irrelevant information from less predictive data types, which can dilute the predictive power of more informative ones. The effectiveness of data fusion can depend on how correlated the data types are and whether they contribute unique information about the target variable.

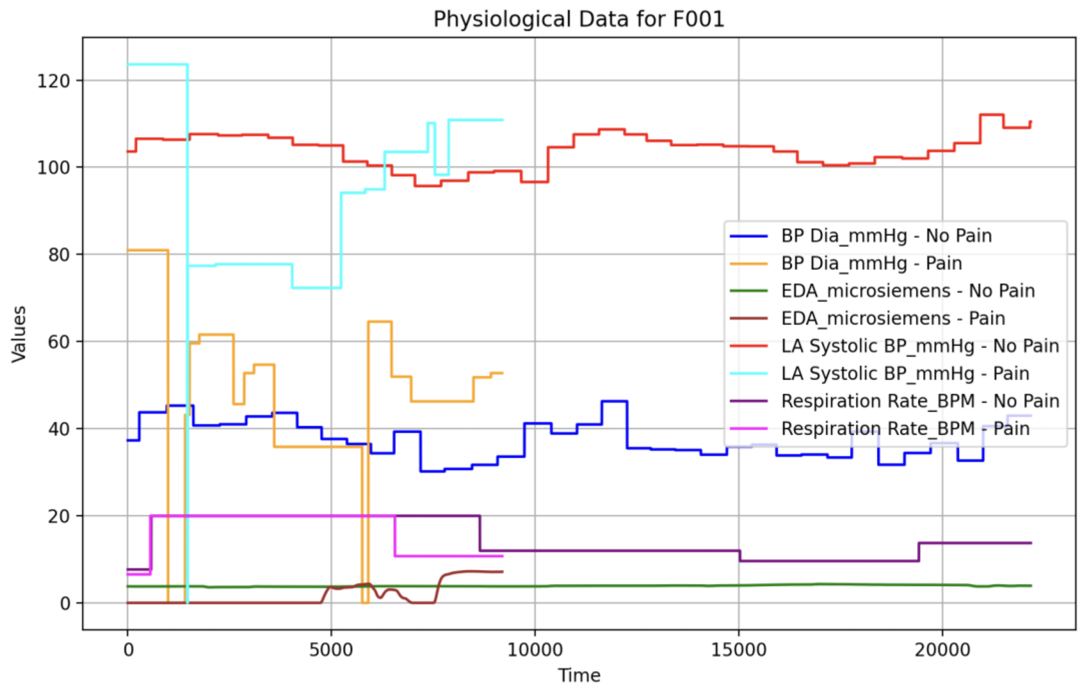
4. Is there a lot of variability in the features that you created? Why do you think this is? To answer this, create a box plot that contains all the features. In other words, the plot will have 1 box for each feature type which will include lines coming from them that show the variability of each feature.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

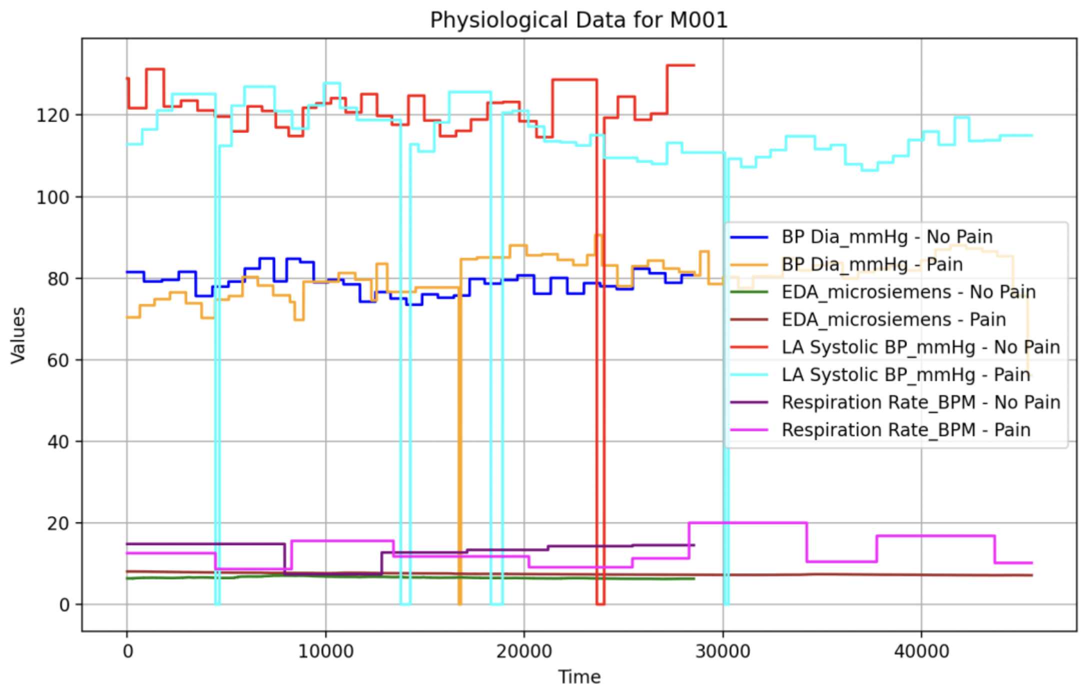
The box plot provides a visual representation of the variability in the features created from the data. Each box in the plot represents a feature type (mean, variance, min, max), and the lines within each box show the spread or variability of values for that feature type across the dataset. Variability can also stem from noise in the data or errors in measurement devices, especially in physiological data where small changes can lead to significant differences in measured values. Variance shows more variability because it measures how spread out data points are from the mean, emphasizing larger deviations and being sensitive to outliers. This sensitivity to the spread of values and the statistical calculation method leads to higher variance values and a perception of greater variability compared to other features like mean, min, and max.



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.

LA Systolic BP_mmHg exhibited the most variability among the plotted physiological signals for a random instance. This variability is often associated with changes in blood pressure, which can occur due to various factors such as physical activity, stress, or health conditions. Pain, while it can affect blood pressure, may not always be the primary factor driving such variability, especially in a random instance where specific pain-related patterns may not be evident.





6. (CAP 5627 only) There is some evidence that some physiological signals are correlated with facial movement (e.g., expressions) during levels of intense emotion (e.g. pain in this case). Give your thoughts and critiques on this. For this question, back up your answer with at least one citation from a published paper.

Research has shown that certain physiological signals, such as heart rate or EDA, can correlate with facial expressions during intense emotions, including pain. For example, a study might find that facial grimacing correlates with spikes in EDA, suggesting a link between the autonomic response and facial expressions of pain. Here, it is crucial to consider both the physiological and behavioral aspects to understand the pain experience fully. A relevant citation for this could be:

- Craig, K. D., & Prkachin, K. M. (2004). Nonverbal measures of pain. In P. McGreevy (Ed.), *Pain: Psychological Perspectives* (pp. 173-202). Lawrence Erlbaum Associates Publishers.
- This work discusses the correlation between nonverbal expressions, like facial movements, and the experience of pain, highlighting the interplay between physiological and expressive indicators of pain.