Affective Computing Project 2

Name: Rikshith Tirumanpuri

1. **Why did you choose the classifier that you did?**
   - I played around with a few models, such as SVM and Decision Trees. There are benefits and drawbacks to every model.
   - I chose to use logistic regression because the problem we are trying to solve falls into the binary classification category. Logistic regression is used to solve problems in this domain. For example, figuring out whether a message is spam or not.
   - Our objective is to discover a method that enables us to use factors such as respiration, diastolic and diastolic blood pressure, fusion of all data types, and EDA to classify a patient as either in pain or not. Validity and reliability have been shown by these metrics in the categorization of pain.
   - By evaluating these signals and applying logistic regression, it may be able to distinguish between states of pain and states without pain. It also has benefits like interpretability, efficiency with little data, and probabilistic predictions. These make it a valuable tool for classifying pain.

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.
   - With an accuracy of 66% for the sys data type, I was able to attain the highest accuracy in my case. Systolic blood pressure and pain can have different relationships depending on the circumstances and individual factors. Blood pressure can occasionally rise as a result of pain because of the body's stress reaction. Adrenaline and other hormones are released during this reaction, which may cause brief elevations in blood pressure and heart rate. Overall, even though there may be a connection between pain and systolic blood pressure, this accounts for the highest accuracy found in all datatypes.

	Predicted Pain	**Predicted No Pain**
Actual Pain	**47**	**27**
Actual No Pain	**13**	**33**

Precision: 0.6762632197414806
Recall: 0.6666666666666667
Accuracy: 0.6666666666666666

- With an accuracy of 65%, the Dia data type ranks second in my ranking.The body's stress response during acute pain may cause the diastolic blood pressure (DBP), the lower number in blood pressure readings, to momentarily rise. If chronic pain is not adequately treated, it can also have a long-term impact on DBP. On the other hand, hypertension may increase the likelihood of developing some forms of pain.

	Predicted Pain	**Predicted No Pain**
Actual Pain	**46**	**28**
Actual No Pain	**14**	**32**

Precision: 0.6586368977673325
Recall: 0.65
Accuracy: 0.65

- My third highest is **Eda** data type with an **accuracy** of 65%. Electrodermal activity (EDA) reflects the body's stress response to pain, with increased skin conductance indicating heightened arousal. EDA monitoring provides an objective measure of pain intensity, aiding in both assessment and the development of biofeedback techniques for pain management.

	Predicted Pain	**Predicted No Pain**
Actual Pain	**45**	**28**
Actual No Pain	**15**	**32**

Precision: 0.6486447099970853
Recall: 0.6416666666666666

Accuracy: 0.6416666666666667

Eda data type, with an accuracy of 65%, ranks third for me. The body's stress response to pain is reflected in electrodermal activity (EDA), where higher skin conductance denotes a higher state of arousal. An objective measure of pain intensity is provided by EDA monitoring, which helps with assessment and the creation of biofeedback methods for pain management.

3. **Fusing data is a common approach in machine learning. How did your fusion features (e.g.**
   all from the command line) perform? If it had the highest accuracy (from question 1) why did
   this happens (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)?

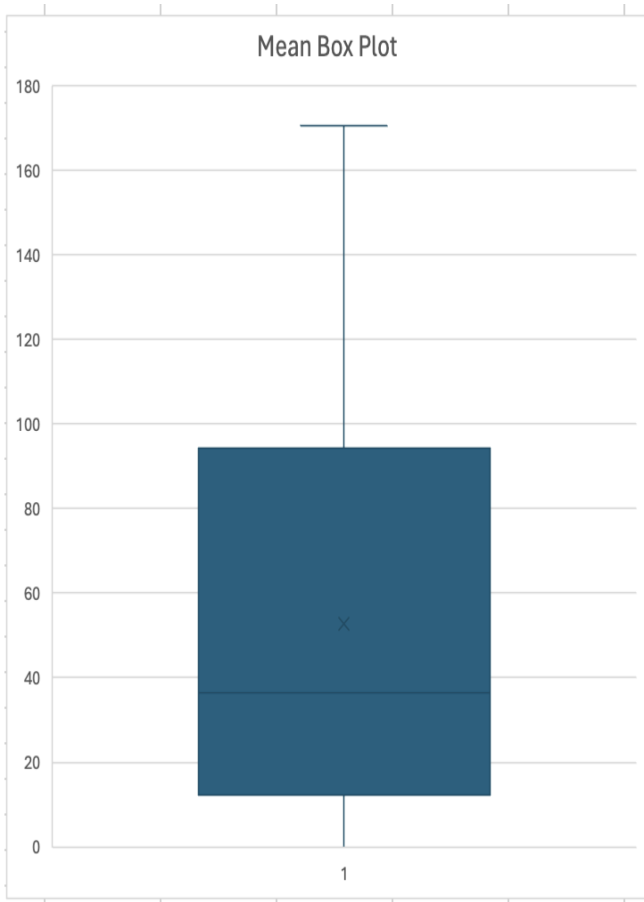
   - An astounding performance was obtained by fusing data into 16 feature vectors, yielding an accuracy of 73%.
   - Complementary Information:
     - To fill in any gaps in the patterns, one piece of the data completes another.
     - Complementary information is combined to achieve enhanced understanding.
   - Decrease in Uncertainty:
     - The data's noise and uncertainty are reduced by fusion.
     - One way to reduce inconsistencies and outliers is to combine data from multiple sources.
   - Risk mitigation:
     - Steers clear of relying too much on one feature.
     - Diversification reduces the possibility of erroneous findings and corrupted data.
     - Robustness and reliability are increased when reliance is spread across several features.

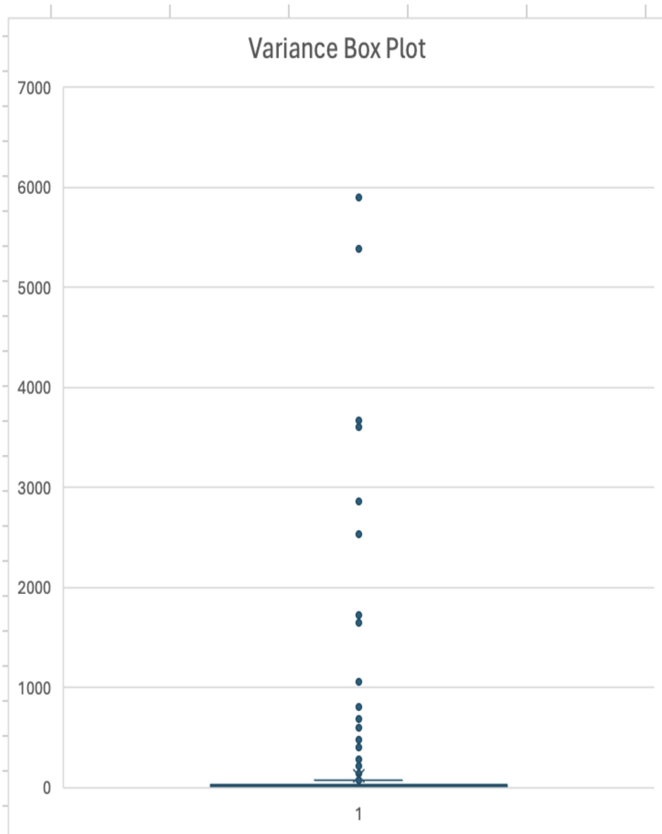
	Predicted Pain	**Predicted No Pain**
Actual Pain	**10**	**3**

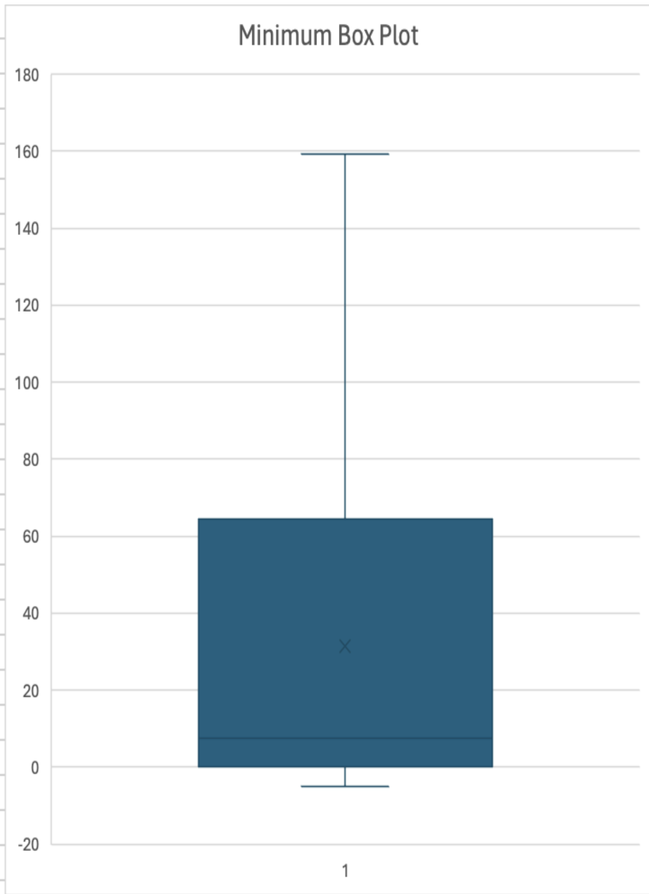
Actual No Pain	**5**	**12**

Precision: 0.7375565610859729
Recall: 0.7333333333333334
Accuracy: 0.7333333333333333

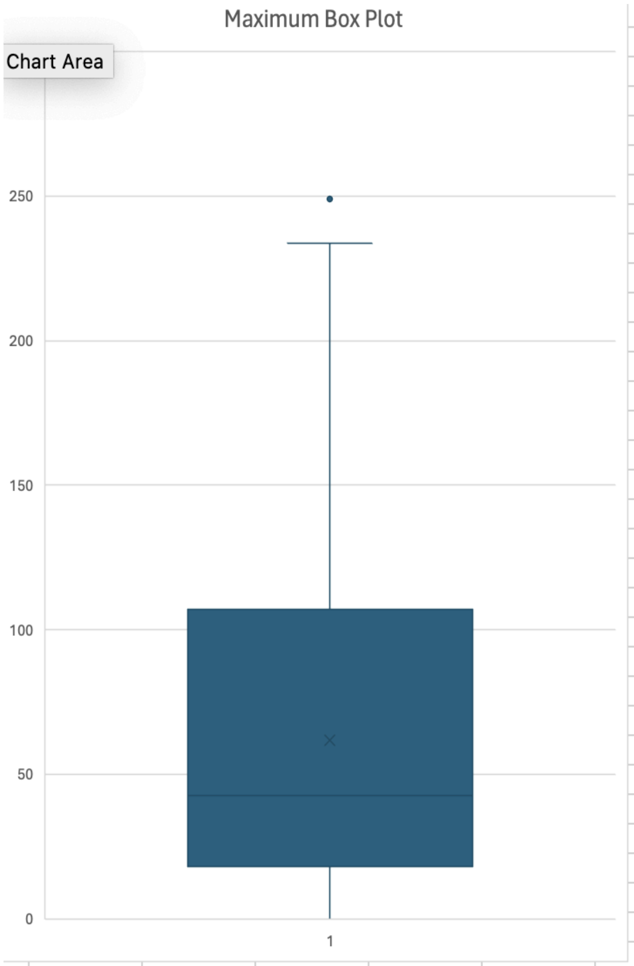
4. **Is there a lot of variability in the features that you created? Why do you think this is? To answer this, create a box plot that contains all the features. In other words, the plot will have 1 box for each feature type which will include lines coming from them that show the variability of each feature.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

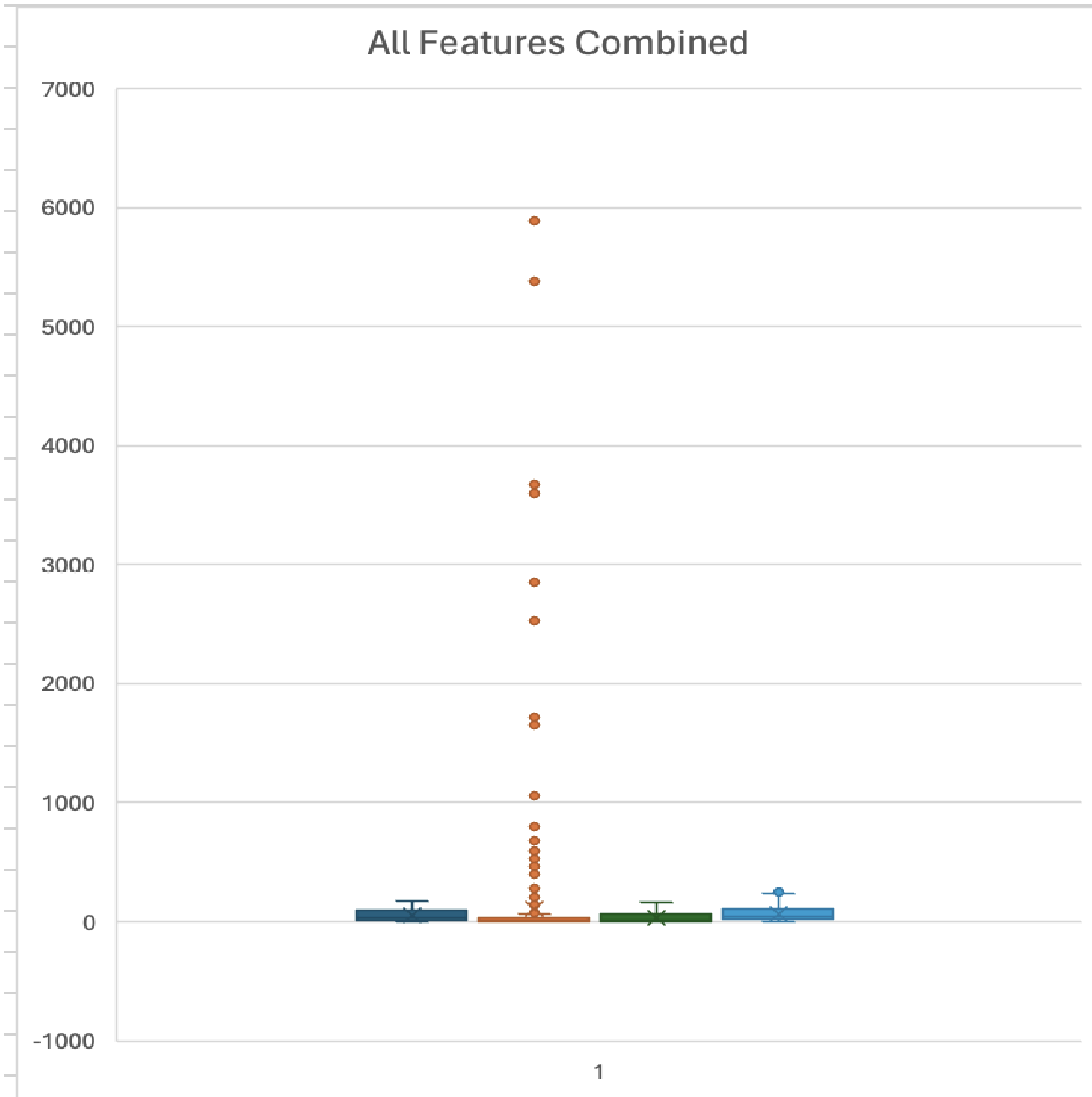

Mean Box Plot
180
160
140
120
100
80
60
40
20
0
1


Variance Box Plot
7000
6000
5000
4000
3000
2000
1000
0
1



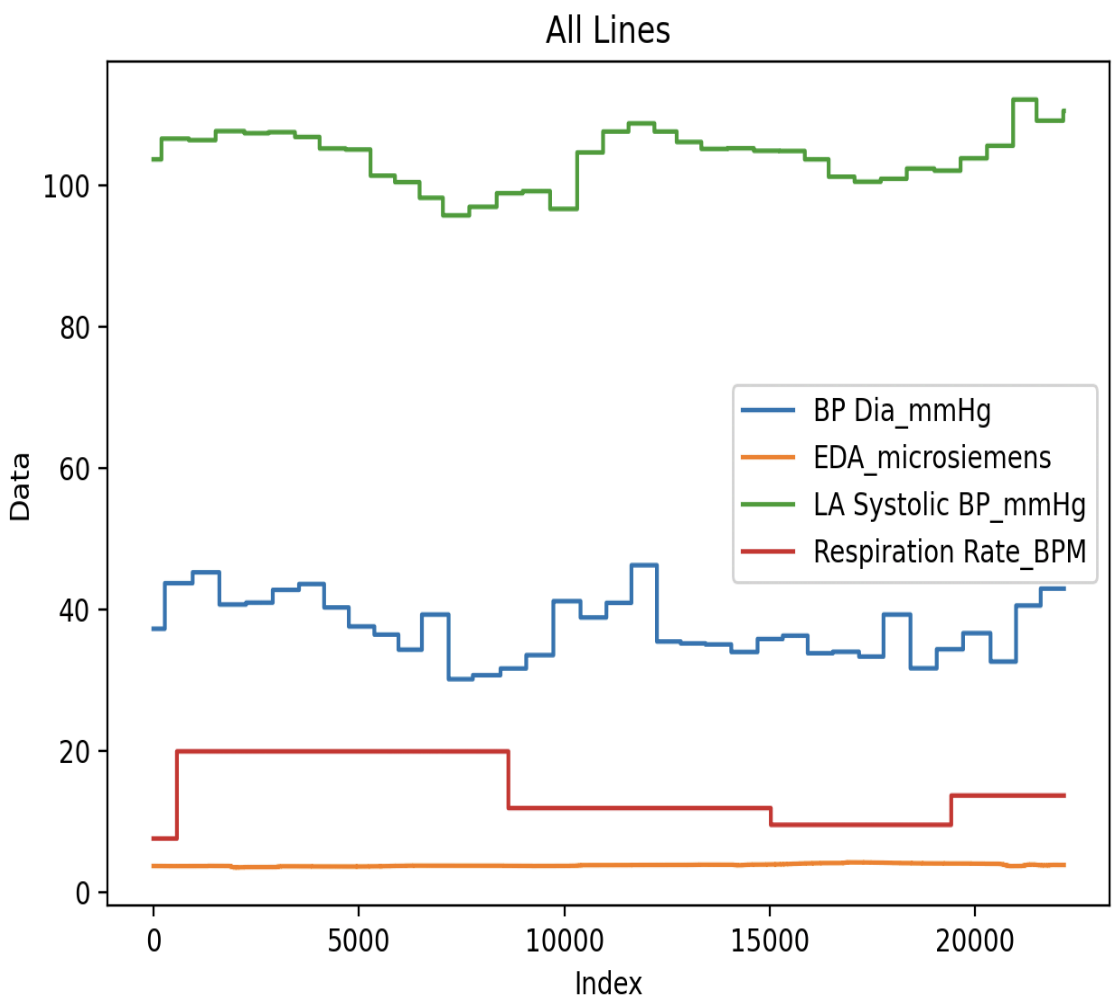
Minimum Box Plot
180
160
140
120
100
80
60
40
20
0
-20
1

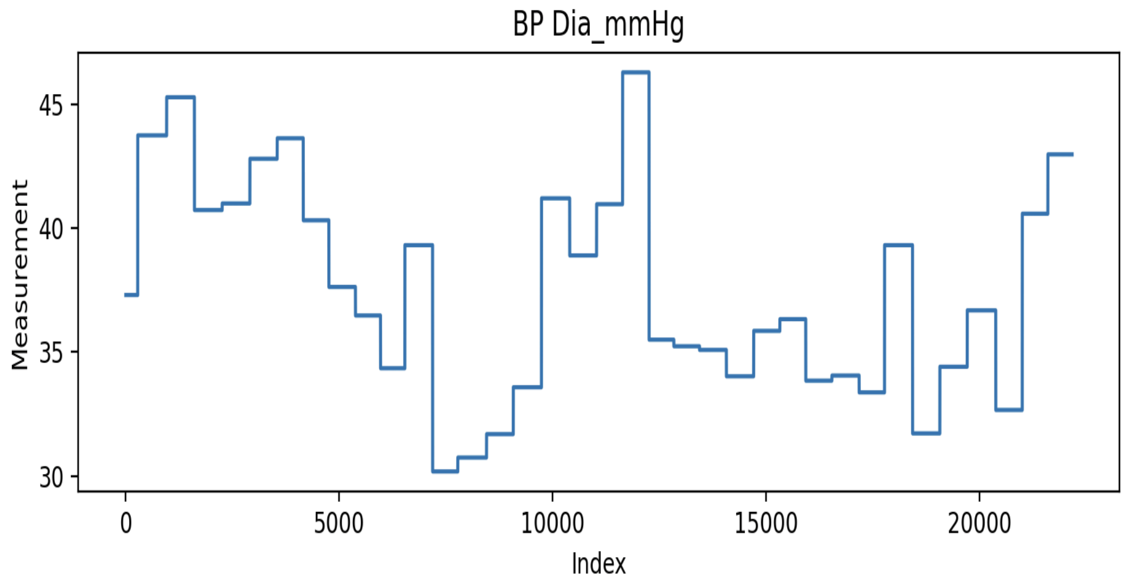

Maximum Box Plot
Chart Area
250
200
150
100
50
0
1

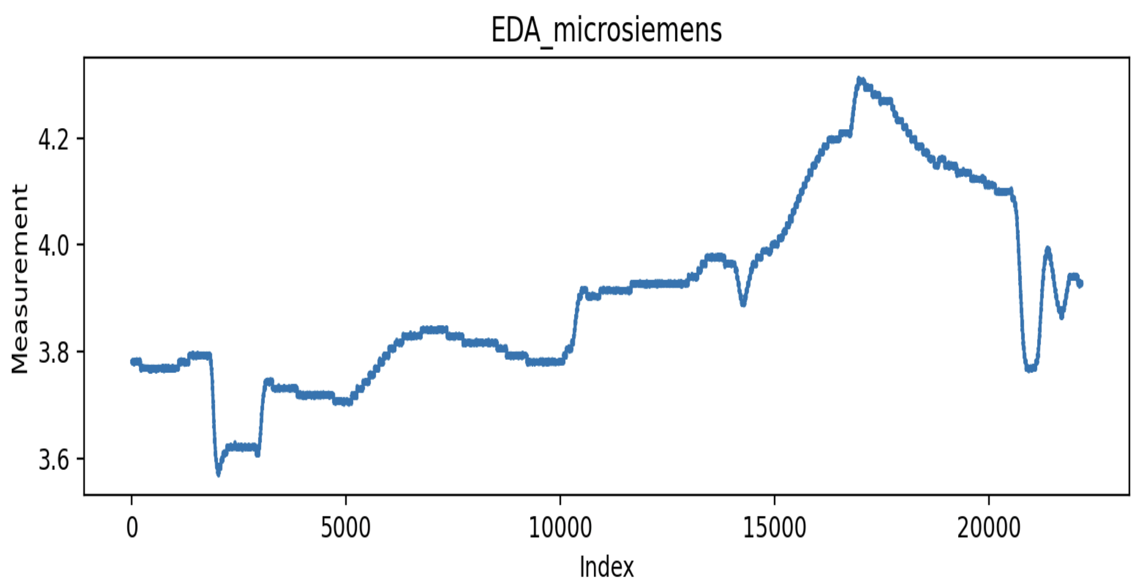


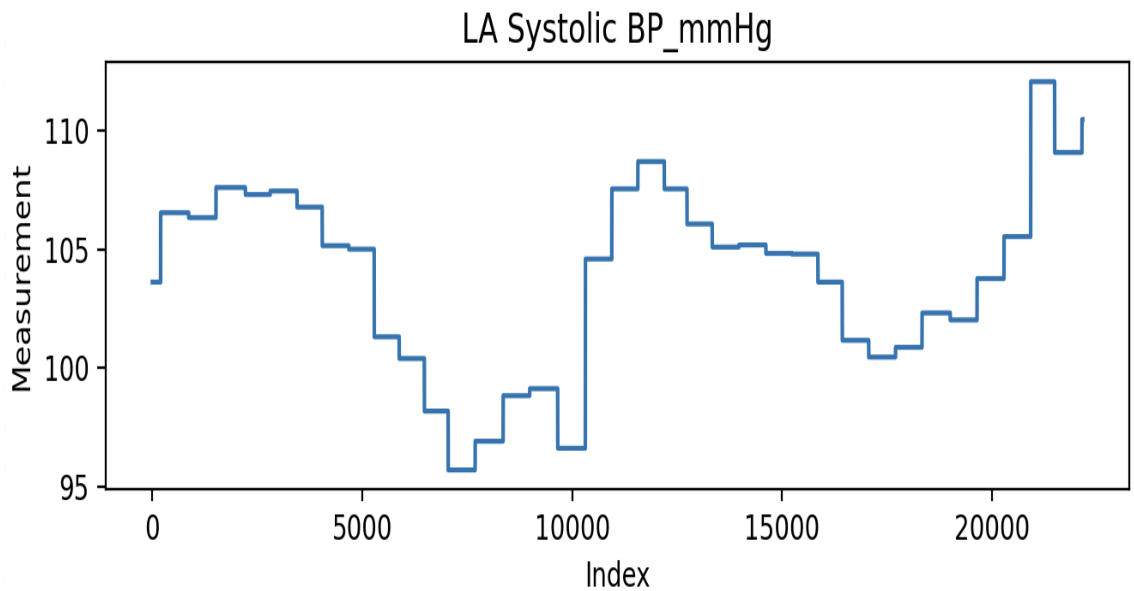
- High variance indicates that physiological signals change dramatically over time.
- Unlike fixed statistics such as mean, max, and min, this variance is regarded as normal.
- This variability is influenced by elements like the environment, physical activity, emotions, and health.
- Because physiological signals are non-stationary, they alter over time in a constant manner.
- By indicating the degree to which data points deviate from the average, variance illustrates these shifts.
- Naturally, there is a lot of variance in signals because physiological processes are dynamic and complex.

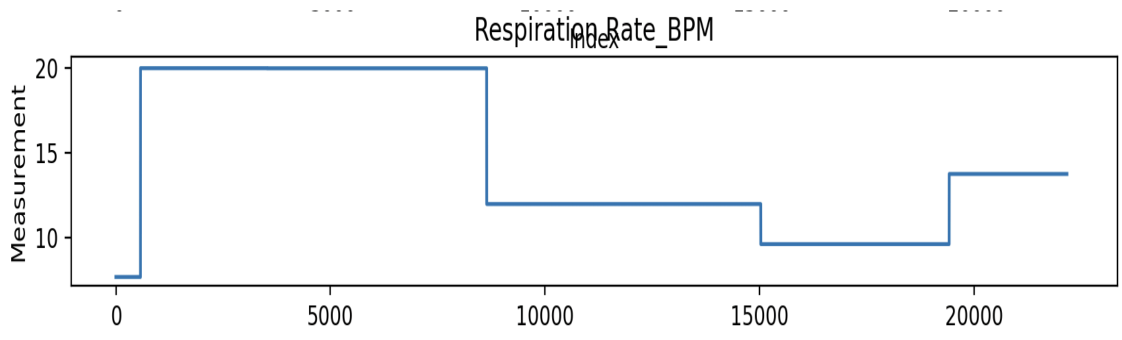
5. **Which physiological signal can visually be seen to have the most variability? To answer**
this, take a random instance of the original physiological signals and plot them in one line
graph. Include a key to show which signal is which (can use different colors for each). Is
the signal that looks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




BP Dia_mmHg
Measurement
45
40
35
30
0
5000
10000
15000
20000
Index


EDA_microsiemens
Measurement
4.2
4.0
3.8
3.6
0
5000
10000
15000
20000
Index





- The most variable signal is respiration rate, which is influenced by variables other than pain, such as exercise and anxiety.
- Although pain can alter breathing patterns, there are other factors that can cause changes in respiration rate.
- In this data, there is no discernible variation in blood pressure readings—which are frequently used to measure pain—when compared to respiration rate.
- Therefore, it is impossible to conclusively link pain to the high variability in respiration rate alone.

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, backup your answer with at least one citation from a published paper.**

- The paper explores using multivariate correlation analysis of physiological signals for emotion recognition.
- It suggests that pain elicits a complex interplay of physiological responses beyond facial expressions.
- Multivariate correlation helps capture the intricate relationship between different physiological signals.
- This approach aims to enhance understanding and recognition of emotional states like pain.

- **Reference:** W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan and W. Huang, "Emotion Recognition Based on Multivariate Correlation of Physiological Signals," in IEEE Transactions on Affective Computing, vol. 5, no. 2, pp. 126-140, 1 April-June 2014, doi: 10.1109/TAFFC.2014.2327617. keywords: {Physiology;Videos;Films;Data acquisition;Heart rate;Feature extraction;Affective pattern recognition;local scaling dimension;multivariate correlation;physiological signal},