Name: Kyle Cwik

CAP 4628/5627 Affective Computing

Project 2

Due 05/01 by 11:59pm

Project Description

You are to develop a new system that can identify (i.e. classify) pain from physiological data that will be collected from wearable devices. You will have access to the type of data that will be collected to train and test your system.

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

   SVM performed the best out of the three classifiers from the last project, and so it seemed appropriate to use again for this project.

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.

   For my experiments, DIA performed the best with an accuracy of 61%. This makes sense as DIA represents blood pressure. In stress, blood pressure rises as an automatic response from the body, and pain tends to induce a stress reaction. Below is a more in detail of the experiment results:

   [56 42]
   [4 18]
   Precision: 0.6948051948051948
   Recall: 0.6166666666666667
   Accuracy: 0.6166666666666667
   Classifier Score: 0.43902439024390244

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)?

   The Fusion data ended up performing second best overall, with a precession of 65% and

accuracy of 60%. It likely performed this way because it was fed a lot of data that might not be as relevant to pain sensing than pure dia. Lots of outlier data could have influence fusion to underperform. Below is the data from the fusion experiment:

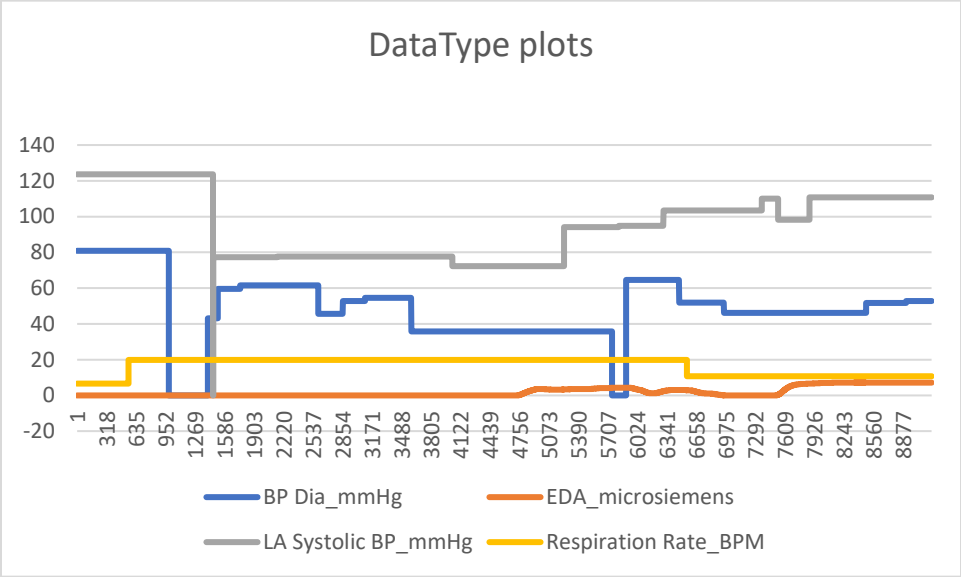
[[53 40]
[7 20]]
Precision: 0.6553166069295102
Recall: 0.6083333333333333
Accuracy: 0.6083333333333333
Classifier Score: 0.4597701149425288

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 shows the variability of each feature. (Search for box plot in python to see how to do this).

   There is indeed a lot of variation in the features when looking at the box plots (these plots can be found in the Experiments folder). Min ended up have the largest spread of data, while variance and Max excessive outlier data. This is likely due to a lot of missing data in the original data file, zero values, and simple outliers in the original data that could cause large numbers to be generated during feature creation.

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 on 1 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.

   Blood pressure appeared to have the most variability overall, followed by LA. This signal is associated with pain, and as shown previously, BP had the most accurate results for the classifier in predicting pain. Again, as already discussed, blood pressure is related to a stress reaction by the body. This is usually done to help aid in situations where a person may need to fight or flee an area. Pain is very often the cause of stress reactions in the body. BP can also change quickly which can help make it easier to see on a chart where a pain response may have occurred.


DataType plots
140
120
100
80
60
40
20
0
-20
1
318
635
952
1269
1586
1903
2220
2537
2854
3171
3488
3805
4122
4439
4756
5073
5390
5707
6024
6341
6658
6975
7292
7609
7926
8243
8560
8877
BP Dia_mmHg
EDA_microsiemens
LA Systolic BP_mmHg
Respiration Rate_BPM