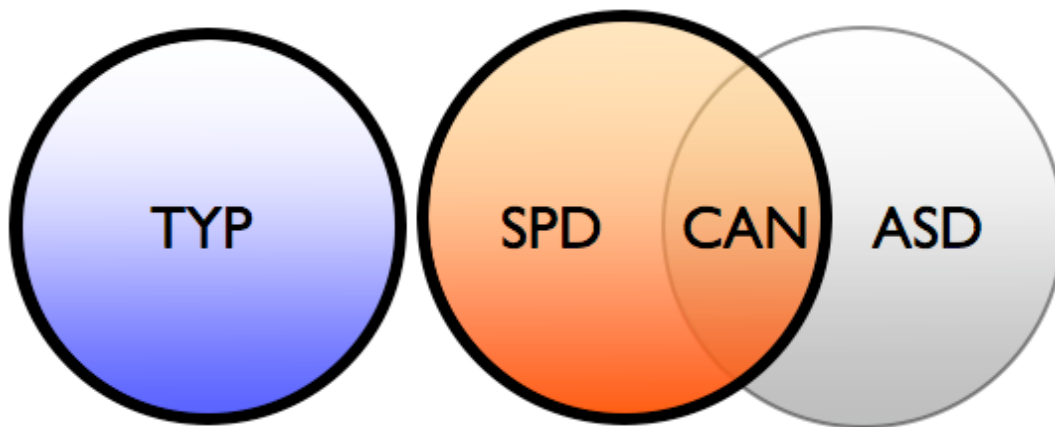# MAS 622J Final Project Report

Classifying Sensory Processing Disorders (SPD) Using Machine Learning and Feature Selection

Ming-Zher Poh, Elliot Hedman, Micah Eckhardt

Motivation

Sensory dysfunction is common among individuals with autism spectrum disorders (ASD), but there is increasing evidence to suggest that this affliction it is not limited to autistic individuals. Nonetheless, sensory processing disorder (SPD) has yet to be recognized by many clinicians as an independent medical condition and thus, children with dysfunctional sensory processing do not receive appropriate treatment. In other cases, these children are often misdiagnosed with attention-deficit hyperactivity disorder (ADHD) and are inappropriately medicated. The objective of this project was to attempt to build a classifier that can distinguish between neurotypical individuals (TYP) and individuals with SPD. Additionally, we attempted to identify features that are particularly robust in separating the two classes.

Description of Data

Dr. Lucy Jane Miller, director of the SPD Foundation, provided the data used in this project. The data consisted of continuous recordings of electrocardiogram (EKG) and electrodermal activity (EDA) obtained from participants during a sensory challenge protocol. The sensory challenge protocol starts with 3 minutes of resting to obtain baseline measurements and proceeds through 6 different sensory events (Figure 1). Each sensory event consists of 8 trials during which the subjects are presented with a stimulus. After the last sensory event, there is another 3-minute period of relaxation. A total of 10 neurotypical samples and 34 SPD samples were obtained.
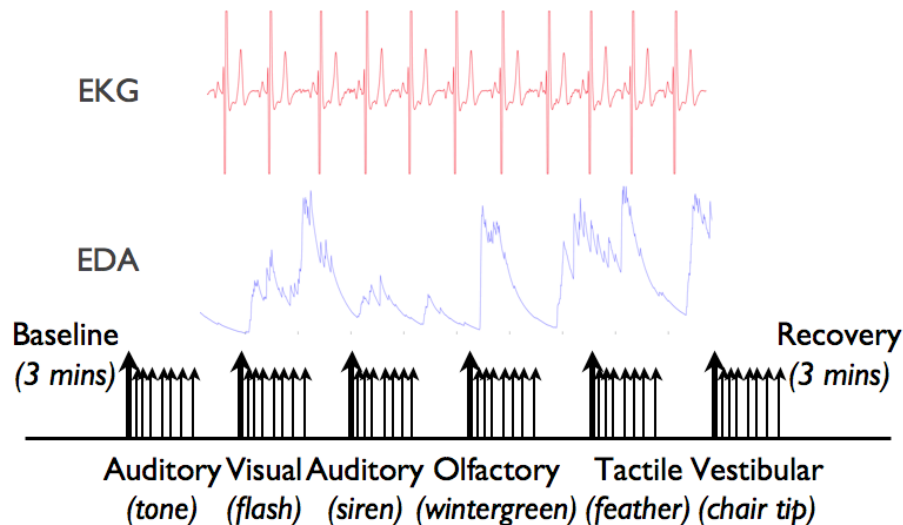

Figure 1. Sensory challenge protocol

Classifiers

Four Machine Learning Classifiers were used to classify SPD from TYP. Each of these algorithms are available through Matlab. Due to the limited number of samples we had to work

with, we used leave-one-out cross validation to determine the specificity and sensitivity of each classifier. Below is a brief summary of each classifier:

- *K Nearest Neighbor*: Each new subject is classified the same as the classification of the majority of its K nearest neighbors. The Euclidian distance between samples is used as a measure of nearness.
- *Decision Trees*: Creates categories based on separability: choosing features that separate the most number of SPD from TYP, and repeating the procedure for each branch. The full tree and the optimal sequence of pruned subtrees was computed. Impure nodes must have 10 or more observations to be split. Splitting criterion was based on Gini's diversity index.
- *Linear Discriminate*: Takes a linear combination of features for each sample and attempts to find a multi-dimensional line that separates the data
- *Support Vector Machines (SVM)*: Constructs a separating hyperplane by considering those data points who's neighbor is in the opposite class. A linear kernel function was employed.

Feature Selection

In order to create robust classifiers features needed to be selected that best predicted future data. Two feature selection algorithms were implemented: Multiple Class Principle Component Analysis and Sequential Backward Floating Selection. Both methods where used with all subsequent classification methods.

Multiple Class Principle Component Analysis (MPCA):
MPCA selects features based on covariance—those features that spread the data are emphasized more. With multiple classes, features are selected based off of what dimensions work for each class alone, and separates them the most. This method can significantly reduce the total number of features, which has significant computational implications. In our case the feature vector for each subject was typically reduced from 580 features to as few as 4.
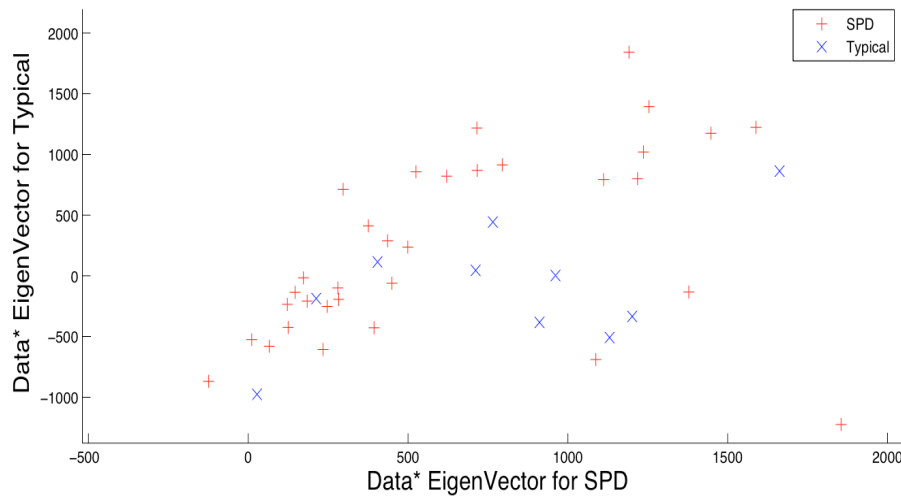
Figure 2. Projected data with selected eigenvectors

The results of MPCA are eigenvectors which can be used to project the feature data into a lower dimensional space, notice how the data begins to appear separable (Figure 2) after projection. In this case the subjects feature vector is reduced from a 580 dimensional space to two dimensions using the corresponding eigenvectors obtained from MPCA. We can use the weights of the eigenvector to help estimate robust features.

Sequential Backward Floating Selection (SBFS):
SBFS attempts to find an ideal combination of features through a stepwise exclusion and inclusion of features into a subset of features, which are then used in the proceeding step. This process continues until the "best" single feature that describes the data is found. It should be noted that though the algorithm attempts to find the "best" single feature in actuality it is often a combination of features which produces the overall lowest error rate.
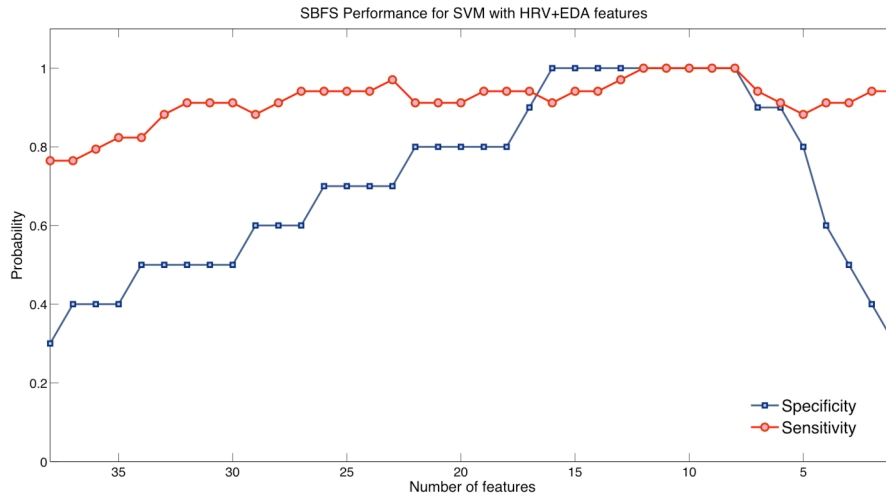
Figure 3. Performance using sequential backward floating selection

Initial classifiers, trained without feature selection, produced particularly poor performance. Figure 3 shows what happens when *SBFS* is applied. By removing misleading features accuracy can significantly improve. Also, by having two criteria functions (sensitivity and specificity) and by focusing on the weaker function, we can achieve good performances for both. In this example, 100% classification is obtained by looking at only 8 key features. Those these results are encouraging it should be considered that results are for validation testing and may not hold on generalized test sets.

Procedure

1) From the raw data provided, we classified 580 features based on events and trials within an event. Of the total features, 56 were heart rate variability (HRV) features, 32 were EDA features compiled over an entire event segment, the remaining features were specific to each stimulus presented in an event segment.
2) From the above features three subgroups were constructed: 1. HRV features 2. EDA features, 3. A combination of HRV features and event specific EDA features
3) Feature selection was performed for each group across all implemented classifiers. The features that gave the classifier the best specificity and sensitivity were reported.

Results and Discussion:

Using HRV features alone

The results obtained from various classifiers using only HRV features are shown in Table 1. The linear discriminant proved to be strongest classifying all TYP correctly and 91% of SPD subjects. SVM had the highest sensitivity (100%) but a lower specificity of 70%. The kNN

algorithm performed the poorest among all classifiers; this could be due to the unbalanced data set (10 TYPs versus 34 SPDs). Note that performance is given for validation test only, due to the small sample size there was no holdout test set.

Table 1. Accuracy of classifiers when looking at only HRV features

Classifier	Sensitivity	Specificity
kNN (k=1)	0.76	0.70
Linear Discriminant	0.91	1.00
Decision Trees	0.91	0.70
SVM	1.00	0.70

Table 2 shows the HRV features that yielded the best performances for each classifier. As can be seen low frequency (LF) and high frequency (HF) data were the most frequently selected by all four different classifiers. pNN50, which is a time domain measure of high frequency variability, was also frequently selected. This is not surprising, given the strong correlation between pNN50 and HF. HF and LF are well known measures of autonomic balance and could provide insight into the physiology of SPD.

Table 2. Most robust HRV features.

Features	kNN	DT	LD	SVM	total
SDNN					0
rMSSD			1		1
pNN50		1	3	2	6
RSA			4	1	5
LF	1	1	2	1	5
HF	2	1	5	1	9
LF/HF			3		3

Using EDA features alone

The results obtained from various classifiers using only EDA features are shown in Table 3. Once again, the linear discriminant provided the best results with 100% accuracy. Nonetheless, this must be taken with a grain of salt since we have such a small data set and the larger number of features used by LD could result in overfitting the data.

Table 3. Accuracy of classifiers when looking at only EDA features

Classifier	Senstivity	Specificity
kNN (k=1)	0.88	0.90
Linear Discriminant	1.00	1.00
Decision Trees	0.94	0.80
SVM	0.94	0.60

The EDA features that provided these results are shown in Table 4. Overall events tended to be much stronger correlates then looking at events on a trial-by-trial basis. The one exception for this was Rise Time, which was repeatedly a strong feature. Here we speculate that the Event Max Amp could be a good indicator of sensory over-responsivity. Conversely, the Event Min Amp could potentially be a good indicator for sensory under-responsivity.

Table 4. Most robust EDA features

Features	kNN	DT	LD	SVM	total
Peak Amp					0
Latency			3		3
Rise T.		2	3	1	6
1/2 Rec. T.	4				4
Mean Amp.					0
Std Amp.			4		4
Event Max Amp		1	3	1	5
Event Mean Amp			4	1	5
Event Min Amp		1	3	2	6
Habituation	1		5	2	8

Using both HRV and EDA features

Table 5 compares the success rates for all 3 groups of features. When combining HRV and EDA data, the results are very encouraging as the results for nearly all classifiers improve. This suggests that a) the data is separable when looking at HRV and EDA combined, and b) the separability is increased when looking at the combination of features. It is important to note that

while 100% accuracy was given in many cases these results may not generalize. Independent test data is need to more accurately interpret classifier performance,

Table 5. Accuracy of classifiers when looking at both HRV and EDA features

Classifier	Sensitivity			Specificity		
	HRV	EDA	Both	HRV	EDA	Both
kNN (k=1)	0.76	0.88	**0.97**	0.70	0.90	**0.90**
Linear Discriminant	0.91	1.00	**0.97**	1.00	1.00	**1.00**
Decision Trees	0.91	0.94	**0.94**	0.70	0.80	**0.80**
SVM	1.00	0.94	**1.00**	0.70	0.60	**1.00**

When looking at heart rate and EDA data combined, features were selected from both in all cases with the max peak amplitude, habituation, and frequency appearing most robust across classifier types (Table 6). It was also nice to see that the same features that were selected from HRV and EDA independently were once again chosen when the feature set was combined, with the exception of the Event Mean Amp.

Table 6. Most robust overall features

Features	kNN	DT	LD	SVM	total
SDNN					0
rMSSD			1		1
pNN50	1	1	6	5	13
RSA			4	1	5
LF	1	1	2	1	5
HF	2	1	5	1	9
LF/HF	1		6	1	8
Event Max Amp	3	3	9	1	16
Event Mean Amp	1	2	9	1	13
Event Min Amp	3		6	3	12

% Habituation	1	1	9	4	15

The algorithms consistently picked the HRV low frequency (0.04-0.15 Hz) and high frequency (0.15-0.4 Hz) features. Figure 4 shows the subjects HRV based on the average LF and HF across all event segments. Even with just these two dimensions, we see that the subjects are separable. Also, we start to notice clusters forming with the TYP in the middle and SPD's in separate groups of high LF and HF and low LF and HF. This high autonomic activation could correspond to sensory over-responsivity while low autonomic activation could indicate sensory under-responsivity.
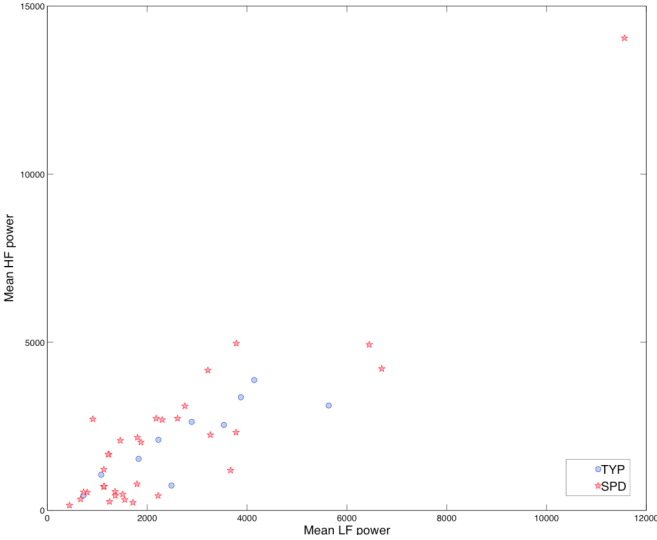

Figure 4. Distribution of subjects according to HRV frequencies

With respect to EDA features computed on an event-stimulus basis we were encouraged that the feature selection methods never selected them, naively we felt many of these features where not representative of the underlying data. The one exception was the half peak time, which was used extensively by the Nearest Neighbor algorithm. Conversely the feature selection methods indicated that the max peak amplitude and the min peak amplitude were also representative indicators. EDA habituation was also preferred by our feature selection methods.

Classification improved significantly when analyzing HRV and EDA together. This does not mean just looking at them separately and taking a vote, but looking at their correlations between them and having them in the same dimensional space. Figure 6 shows how the data becomes more separable with combined features. This data has a 70% specificity and 97% sensitivity. The accuracy increases with higher dimensions (but can't be graphed). Figure 5 shows a clear distinction between SPD and TYP.
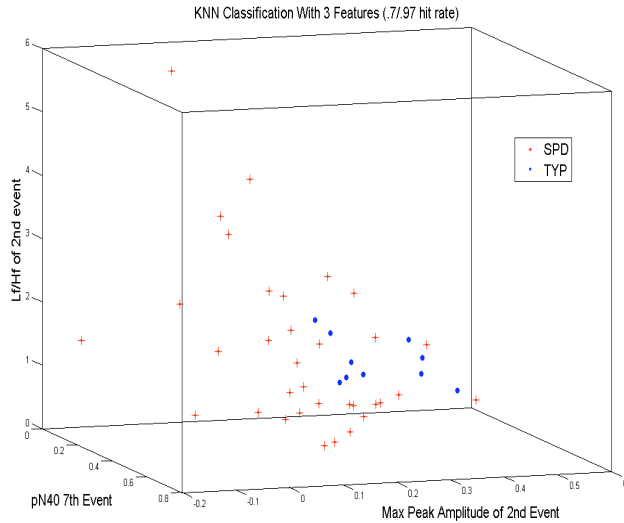
Figure 5. Distribution of subjects in three dimensions

Summary

Experimental results produced robust classifiers with validation accuracy typically above 90%. Overall, results are encouraging and imply that it is computationally possible to separate TYP from SPD. Additionally, and perhaps of more significance, the identification of representative features may be of significant importance. To our knowledge, the features that we have identified are not typically used in the field for SPD classification.

Suggestions for Future Research:

1) Add more subjects to strengthen classifiers, and test for generalization.
2) Investigate event-based features. Feature selection methods indicate that features that encompass an entire event tend to be more representative of the underlying data.
3) Try to find how data correlates. We only identified features, not how they correlate. Does high frequency positively or negatively correlate with SPD, or is it only when looked at with other features?