Phyical Activity Classification

CS470 MACHINE LEARNING

Uzair Akbar
BEE - 5A
04584
13beeuakbar@seecs.nust.edu.pk

M. Sarmad Saeed
BEE - 5A
05673
13beemsaeed@seecs.nust.edu.pk

***Abstract*—Electrical impedance plethsmographs (EIP) are the most feasible way of measuring breathing rate. But movements induce noise in the data acquired through EIP. People disregard the noise and dont take it into consideration. This paper focuses on the noise generated when measuring breathing rate, this noise is distinct to the type of physical activity. With the help of this noise we can determine different activity patterns. Using the noise data, we are able to distinguish 5 distinct physical activities (coughing, reaching, walking, eating and rolling-on bed). This helps us to simultaneously measure the physical activity along with the breathing rates of human-subject. This technique unlike other techniques does not rely on different sensors for both breathing rate and physical activity monitoring and hence saving on costs. We took data from 19 different subjects and used the data to train our system.**

Keywords*—*SVM, Machine Learning, Classification.

I. INTRODUCTION

With every passing day, we face new challenges and providing the world with adequate health resources and facilities is one of them. Many people die because there was no one near them to monitor their adverse condition. Work is being done on systems where doctors can remotely monitor their patients. Moreover, in under developed countries where there are poor health facilities and people are less aware we need devices which can help them improve their health conditions.

To solve this problem, we are trying to move towards remote health monitoring systems (RHMS) for better health care. Research is being carried out in this field to improve the quality of human life. Using RHMS lives of many patients are saved by providing their health data on time to the doctors. Asthma, heart attack, chronic obstructive pulmonary disease (COPD) and other similar life-threatening conditions which deteriorate the patients health can be prevented using this system.

An effective RHMS system which can monitor breathing patterns and physical activities simultaneously will be very useful. In non-clinical environments EIP is the best non-invasive way of monitoring breathing rate. Accelerometer are the most common way of monitoring human activity.

Our paper focuses on how different physical movements can be measured using EIP sensors. EIP has unique distortion signal for different movements. This can enable us to acquire activity data using only the EIP sensor and no need for motion sensing sensors.

The existing techniques heavily rely on accelerometer for motion sensing data. We present a novel approach where we use the data from EIP sensors to get the activity information. We used data from 19 different test subjects multiple time. We the used the data to calibrate our system and to monitor its performance.

We used the data acquired using different humans and trained our system. Our system was able to achieve 85.4 percent accuracy. We used 4-Fold validation. To classify different physical activities we used SVM classifiers.

II. METHODOLOGY

We use support vector machine (SVM) classification for this task. However, the data needs to be pre-processed before proceeding on to the training and classification phase. The main steps involved in classification are given in Fig. 1. The details of these steps are given below.

A. Pre-processing

We need to convert the scores into valid features or a format that acceptable to a classifier. This can be accomplished by the following steps:

- Concatenate the vectors HiCurve and LoCurve to form one feature vector.
- Make all feature vector equal length by discarding a few samples at the start and end so that the features in each sample are the same.

We end up with a 129×11156 matrix, where each row represents a data point or feature vector (activity session) containing 11156 features.

B. Kernel Selection

We assume a polynomial kernel. To find out the best kernel to use in this case, we vary the hyper parameters (in this case the order of the kernel) and calculate the training and validation error each time. After plotting these values, we get a graph as in Fig. 2. We select the order for which we have the first minimum in the validation error. In this case, it is the 2^{nd} order or quadratic SVM kernel.

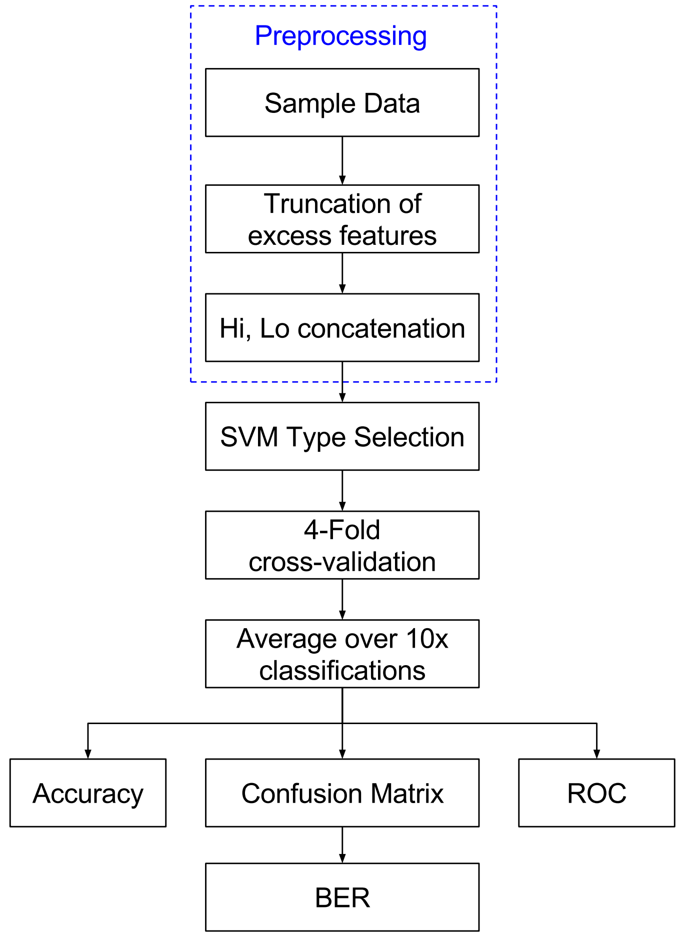

Fig. 1. Method block diagram

C. *Final Training & Evaluation*

We used 4-fold cross validation in our system. We divided our data randomly into 4 non-overlapping folds. We trained 3 of the folds and tested on the 4th fold. We repeated this process until all the folds had been in the test sent one by one. We averaged the accuracy over 4-folds and stored it. We repeated the 4 fold validation a minimum of 4 times and then presented the overall accuracy. The activity sessions were unbalance so we evaluated the Balanced error rate (BER) in addition to accuracy which in our case is more suitable. We use the following equation to calculate BER:

$$BER = 1/M \sum_{i=1}^{N} \frac{(\sum_{j=1}^{N} A_{ij}) - A_{ii}}{\sum_{j=1}^{M} A_{ij}} \tag{1}$$

III. Results

The final accuracy, BER and confusion matrix as a result of the 4-fold cross validation of the 2^{nd} order SVM training is given as follows:

TABLE I. Results

Validation Accuracy	80.85
BER	0.0372

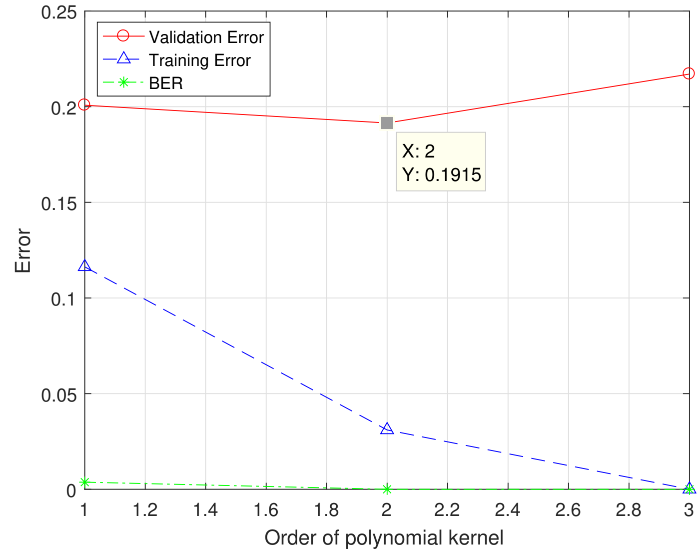

Fig. 2. Validation error and training error for various polynomial kernels

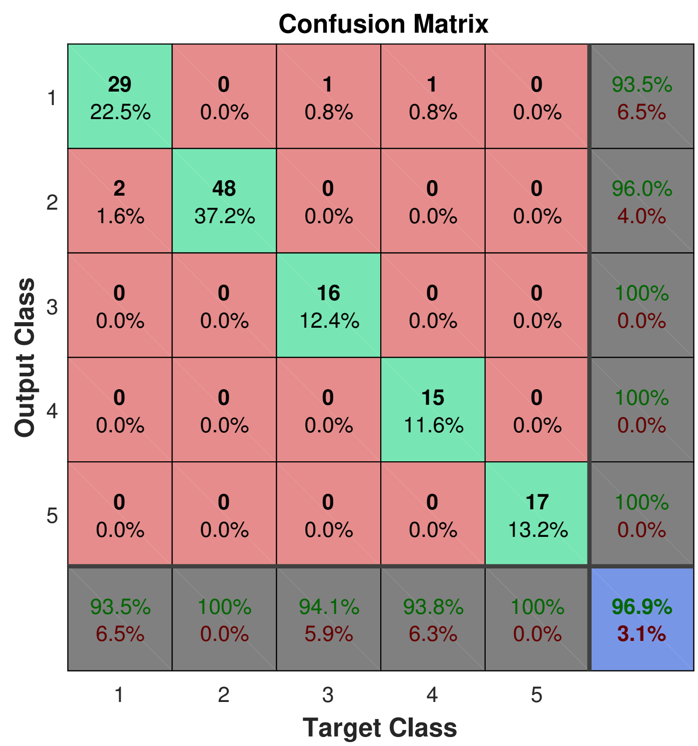

Fig. 3. The average confusion matrix of 10× classification.

We also plot the Return Operating Characteristics for each class, by successively assuming each class as the true class and all other classes as false classes. An ROC curve that is closer to the (0, 1) point is an indication of a good classifier. As shown in Fig. 3, our 2^{nd} order SVM is indeed a good classifier, with good ROC for all classes.

IV. Conclusion

Our system was able to differentiate between different activities based on the noise signal that we obtained when measuring breathing rates of different subjects. Costs can

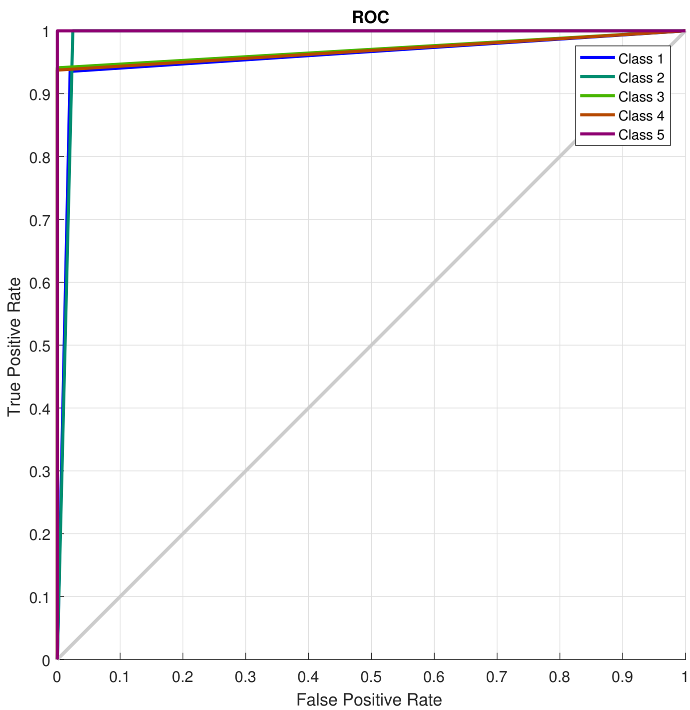

Fig. 4. Return Operating Characteristics (ROC) for every class

be saved on sensors and the prices of RHMS devices can be reduced using such techniques. We achieved an accuracy of 85.4 percent with our system, when measuring different physical activities. More work can be done in this directions and researchers can look for other similar information in the noise signals.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

V. Appendix

The code can be found at the following link:

https://github.com/uzairakbar/machine-learning-cs470/tree/master/course-project