Classification of Alcoholics and Non-Alcoholics using Heart Rate Variability Analysis on ECG Signals

Akarsh N Kolekar, **Apoorv Vatsal** and **Rakshith Vishwanatha**
Guided by – Dr. B. Niranjana Krupa

Abstract: Several medical studies reveal alcohol consumption has pronounced effects on the heart rate variability (HRV) of the consumer. In this article, electrocardiogram (ECG) samples of chronically alcoholic subjects and normative subjects are collected for HRV analysis and feature extraction. The features extracted are fed to machine learning algorithms to enable the algorithms to classify new subjects into alcoholic or normative classes. For this classification problem, Support Vector Machines and Extreme Learning Machines have been trained, and their performance has been compared. While time domain, frequency domain and non-linear features are generally extracted from ECG signals for HRV analysis, in this study a new set of features obtained from Autoregressive Modelling (using Exogenous Inputs) have also been used to improve the accuracy of the algorithms being trained. It has been ensured that the subjects whose ECG has been taken do not have any heart disorder, to increase the likeliness that the HRV occurring in the subject is due to the effect of alcohol and no other cause.

Keywords: Electrocardiogram, Heart Rate Variability, Classification, Support Vector Machine, Extreme Learning Machine, Autoregressive Modelling with Exogenous Input

1 INTRODUCTION

The effects of consumption of alcohol in large amounts in a short period of time or gradually for a prolonged time in humans have been studied and documented extensively in [1], [2], and [3]. The Parasympathetic Nervous System (PNS) and Sympathetic Nervous System (SNS), which are parts of the Autonomic Nervous System (ANS), are inputs to the Sino-Atrial (SA) Node in the heart which controls the variation in inter-beat-interval. PNS is known to lower heart rate among other physiological changes, while SNS is known to increase it [4].

Alcohol consumption acts as a depressant on the brain and nervous tissue, which results in increased SNS activity, decreased PNS activity, which causes increased heart rate, and decreased Heart Rate Variability (HRV), which is the variation or change in the inter-beat interval of the heart. Studies [5] and [6] have detailed the correlation between the amount of alcohol consumed and the extent of change in the HRV, while accounting for factors like the subject's gender, weight, BMI, etc. From the results of these studies, it is possible to draw conclusions about the differences in alcoholic and normative subjects based purely on HRV.

In this article, HRV analysis has been performed on the ECG signals to obtain three different types of features. They include the time domain, frequency domain and non-linear features, which were used to train the Support Vector Machine (SVM) and Extreme Learning Machine (ELM). This feature set yielded accuracies higher than that obtained in [7] for SVM and similarly high accuracies for ELM. A new set of features were extracted from the ECG signals using autoregressive modelling with exogenous inputs (ARX). Addition of ARX features generalized well for both classifiers and improvements in their accuracies was observed. A comparative study has been made between both the algorithms in the two cases where the usual time domain, frequency domain and non linear features were used to train the classifiers, to the case when the autoregressive model coefficients were also included with the feature set.

2 METHODOLOGY

2.1 Dataset Description

In order to classify test subjects as alcoholics or normative with a reasonable accuracy, the classifier was trained with ECG data recorded at the Autonomic Lab, Department of Neurophysiology, NIMHANS, Bengaluru. The dataset consists of 67 ECG samples, of which 38 were samples of alcoholic patients and 29 samples were that of normative people. The ECG samples had a sampling frequency of 1kHz and each sample was recorded for approximately 5 minutes (300 seconds). From the dataset of 67 samples, an equal number of alcoholic and normative samples (28 each) were used to train the classifiers to prevent biasing of the algorithms.

2.2 Pre-processing

The ECG dataset contains disturbances like baseline wandering and power-line noise which need to be removed to obtain clean ECG signals. Wavelet transform, a technique to break up a signal into shifted and scaled versions of the original signal, is used to remove such sections of the signal.

Here, discrete wavelet transform of wavelet analysis is used in which the scale and position of the wavelets are varied in powers of two. The Daubechies wavelet of the wavelet family is selected because the shape of the ECG signal and that of db5 is the same. Eight level wavelet decomposition is performed on the signal, and the eight component when removed from the original signal filters the signal of baseline wandering. The power-line noise has a much smaller amplitude than the total swing in the ECG signal, and is thus left unfiltered.

2.3 Feature Extraction

Features for the classifiers are extracted from the filtered signal using some of the techniques mentioned in [8]. Four types of features have been obtained. They are obtained as mentioned in the sections 2.3.1 to 2.3.4.

2.3.1 Time Domain Features

In this type of feature extraction, the time instants of occurrences of R peaks is stored in an array, and the RR intervals (the time passed between two R peaks) is calculated. The series of RR intervals is used to calculate the time domain features. The following features are extracted:

a. <u>Mean of RR intervals (RR_mean)</u>

This feature is the arithmetic mean calculated on the RR interval series and was calculated as follows

$$RR_{mean} = \frac{1}{n}\left(\sum_{i=0}^{n} RR_i\right) \qquad (1)$$

b. <u>Standard Deviation of RR intervals (RR_std)</u>

This feature is the standard deviation of the RR interval series and was calculated using

$$RR_{std} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n}(RR_i - RR_{mean})^2} \quad (2)$$

c. <u>Mean Heart Rate (HR_mean)</u>

From the RR interval series, the average frequency of occurrence of the RR intervals per minute was calculated using (3). The inverse was then taken which resulted in the mean heart rate.

$$HR_{mean} = \frac{60 \times 1000}{RR_{mean}} \qquad (3)$$

d. <u>Standard Deviation of Heart Rate (HR_std)</u>

The standard deviation of the heart rate was calculated in a similar manner as the mean heart rate. First the standard deviation of the of the frequency of occurrence of the RR intervals per minute was calculated. Then, the inverse was taken to obtain the standard deviation of the heart rate.

$$HR_{std} = \frac{60 \times 1000}{RR_{std}} \qquad (4)$$

e. <u>Root Mean Square of RR intervals (RR_rms)</u>

The square root of the mean of the sum of the squares of all the entries in the RR interval series results in the RMS of the RR interval series. The formula for the same is as follows:

$$RR_{rms} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n-1}(RR_{i+1} - RR_i)^2} \quad (5)$$

f. Number of Intervals Varying by Larger than a Threshold (RR_50)

This feature is slightly different from the rest, in the sense that it involves an additional step of taking differences. While obtaining the RR interval series required taking successive differences of the time instants at which R peaks occurred, here, successive differences are taken for the values in the RR interval series itself. On this new series of difference, the number of time differences that are larger than 50ms are counted to yield the RR_50 feature.

g. Relative Number of Intervals Varying Larger than a Threshold (RR_r50)

The previous feature obtained divided by the total length of the RR interval series gives rise to the final time domain feature. This can be represented by the following equation:

$$RR_{r50} = \frac{RR_{50}}{n-1} \times 100\% \qquad (6)$$

2.3.2 Non-Linear Features

Three non-linear features have been extracted. There were two types of analysis done, one of which yielded two features while the other gave rise to a single feature. These two methods are as given below:

a. Poincare Plot

This feature is extracted 'graphically'. A graph of Poincare points is plotted, where the horizontal axis value of all the points is some i-the value in the RR interval series, while the vertical axis value of the point is the (i+1)-th value of the RR interval series. Then, the standard deviations of the points along two different axes are calculated to yield two of three non-linear features denoted by SD1 and SD2 respectively. The axes along which the standard deviations are calculated are the x = y line and the x = -y line. A sample of the Poincare plot obtained for the first alcoholic sample in the dataset has been shown below.

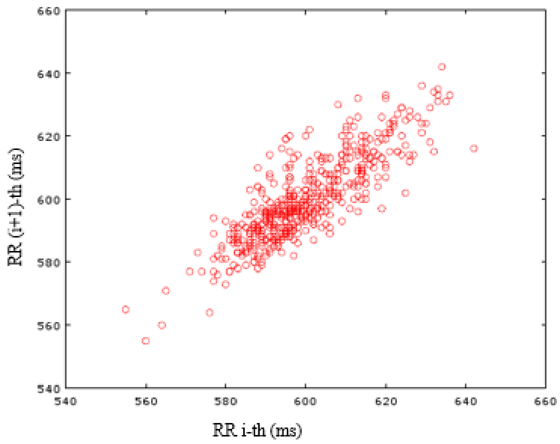

Fig. 1. Poincare Plot

b. Approximate Entropy

This feature is essentially a measure of how much irregularity exists within the RR interval series. The following steps outline very briefly how the feature is calculated. The exact equations to implement the steps mentioned below are provided in [7]

a. From the RR interval sequence, a set of $N + m - 1$ vectors of length m are formed. N is the total length of the RR interval series.
b. Then the maximum of the element-wise differences of all the pairs of vectors is found and stored.
c. Then a quantity C is calculated that counts the number of distance metrics that were lesser than a threshold r.
d. The natural logarithm of the quantity C is calculated
e. The sum of all such logarithmic values corresponding to all the vectors is obtained.
f. Similar steps are performed for vectors of length $m + 1$
g. The difference between the results for vectors of length m and $m + 1$ gives the approximate entropy.

2.3.3 Frequency Domain

The frequency domain features calculated are obtained from the power spectral density (PSD) of the RR interval series. Some features are calculated for the full frequency range of the

PSD, while the rest are calculated only within some frequency bands. The three main frequency bands where features are calculated are the very low frequency (VLF) band, the low frequency (LF) band and the high frequency band (HF). VLF has a range from 0Hz to 0.04Hz, LF has a range from 0.04Hz to 0.15Hz, and HF is the range of all frequencies above 0.15Hz up to 0.4Hz.

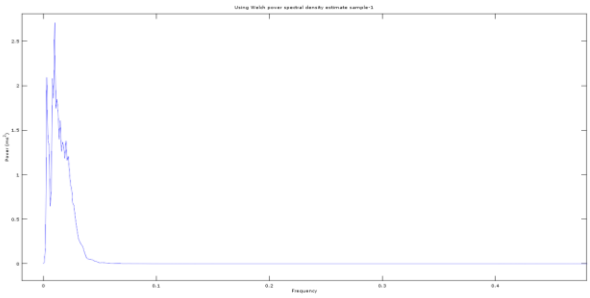

Fig. 2. Power Spectral Density of ECG signal

The frequency domain features that have been extracted are:

a. <u>Peak Frequency in VLF, LF, HF (pk_freq_vlf, pk_freq_lf, pk_freq_hf):</u>

The amplitude of the peak frequency values in very low, low, and high frequency ranges of the PSD of the signal form the first three frequency domain features.

b. <u>Absolute Power in VLF, LF, HF (ab_pow_vlf, ab_pow_lf, ab_pow_hf):</u>

The total power contained in the very low, low and high frequency ranges of the PSD of the signal form the next three frequency domain features.

c. <u>Total Power of the Signal (pw_ttl):</u>

The seventh frequency feature used is the total power contained in the PSD of the signal.

d. <u>Relative Power in VLF, LF, HF (rp_vlf, rp_lf, rp_hf):</u>

The relative power of a certain band is calculated as the power in that band divided by the total power present in the signal. The folmula for VLF is given below (7). The formulae for LF and HF are similar.

$$rp_{vlf} = \frac{ab\_pow_{vlf}}{pw_{ttl}} \qquad (7)$$

e. <u>Normalized Power in LF, HF (norm_lf, norm_hf):</u>

The normalized power of the LF and HF range is the absolute power present in that range divided by the difference in the total power of the signal and the power contained in the VLF band. The equation for the normalized power in the LF range is given by (8) and that for the HF band is also similar.

$$norm_{lf} = \frac{ab\_pow_{lf}}{pw_{ttl} - ab\_pow_{vlf}} \qquad (8)$$

f. <u>Ratio of Absolute Power of LF and absolute power of HF (ratio):</u>

This last feature is obtained by taking the ration pf the power present in the LF range and that present in the HF range.

$$ratio = \frac{ab\_pow_{lf}}{ab\_pow_{hf}} \qquad (9)$$

2.3.4 Coefficients of Autoregressive Model with Exogenous Input

The autoregressive (AR) model establishes a connection between a particular output variable and its previous values. Using AR model, a signal sequence y(n) can be represented as:

$$y(n) = a\_y(n-1) + a\_y(n-2) + a\_y(n-p) + \varepsilon(n)$$

where $a\_(k = 1,2,\dots,p)$ are the model coefficients and $\varepsilon(n)$ is a white noise series. The AR model is further expanded to include ECG signal as the input signals and adapted into an Auto-regressive exogenous (ARX) input model. The ECG signal obtained from the subject is divided into two halves. One half of the signal serves as input to the system and the other half serves as the output. The model that relates the input sequence to the output sequence now forms the ARX model.

After the ECG signal is free from base line wandering (through the pre-processing stage), the signal is loaded into the System Identification Toolbox in MATLAB to calculate the ARX coefficient. These coefficients are later used as additional features to train the classifier.

What ARX seems to be capturing in its model is the long term variations in the ECG that occur over the entire length of the signal. By splitting the signal into two halves, one being the input and the other the output, an ARX model is built to link the first half of the signal to the second half of the signal. This we believe, captures the long term gradual change that occurs through the span of the first half of the ECG signal to the second half. In essence, this seems to be another form of HRV analysis being performed over longer periods of time rather than the shorter beat to beat analysis performed in time, frequency and non-linear methods.

2.4 Classifiers

Two classifier algorithms, each having a different ideology behind it has been trained on the dataset. One of the classifiers is the Support Vector Machine (SVM) which uses the idea of hyperplanes and decision boundaries, while the other is based off neural network concepts and is the Extreme Learning Machine (ELM).

2.4.1 Support Vector Machine

A support vector machine with regularization was implemented on the dataset of 28 alcoholic and 28 normative samples. An RBF kernel was used [9] in order to better separate the data points in a higher dimension. The SVM used the Simplified SMO algorithm [10] to solve the Lagrangian problem and obtain the weights for the hyperplane.

The dataset was divided randomly into k-folds. One fold was used as the validation set, while the rest were used to train the system. The averaged result of k-such validation accuracies resulted in the k-fold validation accuracy. Such a process of obtaining the k-fold validation accuracy was conducted a large number of times (two hundred times). In each of the 200 iterations, the dataset was relabelled randomly into k new folds, and the optimal C and σ for such a labelled dataset was obtained by looking at which (C, σ) pair yielded the highest k fold cross validation accuracy. After all the 200 loops had been iterated, the pair of C and σ that was selected the most (ie. The pair that yielded the highest k-fold accuracy most consistently was used to train the final SVM classifier model.

2.4.2 Extreme Learning Machine

The Extreme Learning Machine (ELM) was also trained on the same data (28 alcoholic and 28 normative sample) that was used for the SVM. Here again, the data was passed through an RBF Kernel to obtain better training and validation results.

In the ELM algorithm [11], the input weights are set randomly and the values to which they are set can affect the accuracy of the classifiers significantly. Variations up-to and sometimes beyond ten percent can be seen due to changes in the random assignment of the input weights. To obtain the most accurate classifier for the given dataset, the ELM algorithm was trained several times and the input weights yielding the best accuracy were used.

Two parameters other than the input weights also affect the accuracy of the ELM system. They are the number of hidden neurons and the variance parameter used for the RBF kernel. The best combination of input weights, hidden number of neurons, and variance parameter were arrived at by 'brute force'. The variance parameter was ~~varied~~ changed from 0.01 to 30 in steps of three, the number of hidden neurons was varied from 1 to 100 in steps of five, and the ELM was retrained repeatedly for all combinations of variance and hidden neurons. This process (of retraining the ELM for different number of hidden neurons and variance parameter was repeated a large number of (two hundred) times to increase the probability of obtaining a model with optimum input weights. For all the models that were trained, k-fold (7-fold) cross validation was performed and the model that resulted in the highest k-fold cross validation accuracy was saved.

2.4.3 Validation

While the training the classifier systems k-fold cross validation was used to verify/validate the accuracy of the model that was trained. For both classifiers $k = 7$ was used while performing k-fold cross validation, allowing each fold to contain eight samples with four samples from each of the two classes.

Along with the accuracy obtained through leave-one-out and k-fold cross validation, the sensitivity and specificity (10) and (11) were obtained from the confusion matrix.

$$Sensitivity = \frac{True\ Positive}{True\ Positive+False\ Negative} \quad (10)$$

$$Specificity = \frac{True\ Negative}{True\ Negative+False\ Positive} \quad (11)$$

Calculation of sensitivity and specificity ensured that classification accuracy remained high for both the positive and negative classes, and that the system was not biased toward a particular class.

3 RESULTS

The results obtained in the steps undertaken while training classifiers to distinguish ECG samples of alcoholic and normative subjects have been provided in this section.

3.1 Results of Pre-Processing

Applying eight step wavelet decomposition on the ECG signal proved fruitful in separating the baseline wandering component from the signal. The eight level of decomposition when subtracted from the original signal resulted in a flat base ECG signal whose peaks could be detected accurately and PSD could be calculated.

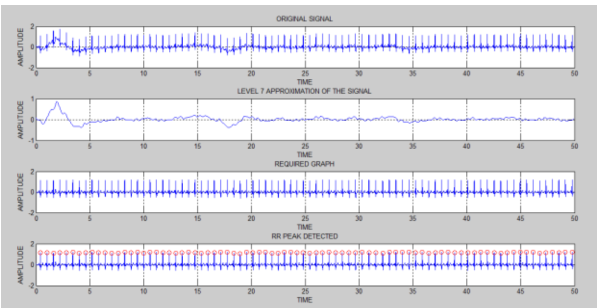

Fig. 3 Wavelet decomposition for baseline wandering removal

The figure below shows the complete steps involved in removing the base line wandering of the signal using Wavelet transform and detecting the R-peaks. The first subplot of the Fig. 3 shows the original ECG signal. The second subplot shows the level8 approximation obtained after Wavelet analysis on the ECG signal. The de-trended signal is shown in the third subplot. The RR-peaks is detected after de-trending the signal as shown by fourth subplot.

3.2 Results of Feature Extraction

Seven time domain, three non-linear, thirteen frequency domain and four ARX features were extracted from each of the fifty six ECG signals used to train the system. The array below gives a consolidated list of all the features that were extracted:

[RR_mean, RR_std, HR_mean, HR_std, RR_rms, RR_50, RR_r50, pk_freq_vlf, pk_freq_lf, pk_freq_hf, ab_pow_vlf, ab_pow_lf, ab_pow_hf, pw_ttl, rp_vlf, rp_lf, rp_hf, norm_lf, norm_hf, ratio, sd1, sd2, ApEn, ARX_coeff1, ARX_coeff2, ARX_coeff3, ARX_coeff4, ARX_coeff5, ARX_coeff6]

A point to note is that ARX modelling of order n results in $n + 3$ coefficients. In the array given above ARX coefficients for a fifth order system has been provided. The fifth order system provided eight coefficients, however only six were used as two coefficients were the same (either one or zero for all samples) for all the samples.

3.3 Results of SVM

The process of selecting the C and σ pair, training the SVM, obtaining the k-fold validation accuracy have been mentioned in Section 2. The output of the two hundred iterations (per C value considered on the feature set with ARX order 5 features), to obtain a count of which (C, σ) was chosen most and yielded the highest k-fold cross validation accuracy for the algorithm is given in Fig. 4.

C/Sigma	0.01	0.03	0.1	0.3	1	3	10	30
0.01	0	2	1	12	170	6	4	5
0.03	0	2	1	11	171	8	4	3
0.1	0	1	1	12	177	2	3	4
0.3	0	2	1	6	186	3	1	1
1	0	5	0	7	131	47	8	2
3	0	7	2	13	26	126	16	10
10	0	11	5	30	12	0	114	28
30	0	9	3	21	20	0	123	24

Fig. 4 Distribution of choice of regularization parameter and RBF variance for SVM

From the distribution above it is clear that $(0.3, 1)$ pair yielded the highest accuracy the most number of times and thus worked best for feature set that included ARX order 5 features.

In the tables that follow, the results obtained for different groups of feature sets. The accuracies and optimal C and σ pair obtained for the different types of feature sets are as given the table below Table 1.

Features Used	Optimal ($\boldsymbol{C}$, $\boldsymbol{\sigma}$) Pair	7-Fold Accuracy
Time, non-linear, frequency	(0.1, 0.3)	80%
Time, non-linear, frequency, ARX order 3	(0.3, 1)	82%
Time, non-linear, frequency, ARX order 5	(0.3, 1)	86%

Table 1. 8-fold Accuracy of SVM for different feature sets

From the table above (Table 1), it is seen clearly that the accuracy of the system improves when ARX coefficients are used along with the time domain, non-linear and frequency domain features.

As mentioned earlier, it is important to understand how many samples of each class were classified correctly. For this, The confusion matrix was obtained and the sensitivity and specificity (10) and (11) were calculated for each of the cases given above in Table 2.

Prediction /Actual	Actual +ve	Actual -ve	Sensitivity/ Specificity
Predicted +ve	25	3	89%
Predicted -ve	5	23	82%

Table. 2 Confusion matrix, sensitivity and specificity of SVM with ARX order 5 featrues

The table Table 2 shows that the system is very slightly biased toward the positive class since a few more samples in the negative class are being misclassified. However, this is not of much concern as both sensitivity and specificity are well above 80% accuracy.

3.3 Results of ELM

The process of obtaining good input weights, selecting the optimal number of neurons in the hidden layer and choosing the RBF kernel's variance parameter σ to train and validate the ELM have been provided in Section 2. A graph representing the accuracy versus hidden neurons plot Fig. 5 for a specific σ showed that accuracy of the algorithm peaked when the number of hidden neurons were maintained between one to twenty. This behaviour of the accuracy peaking for one to twenty neurons could be generalized for any σ, as similar plots were obtained irrespective of σ. The analysis drawn from such a plot allowed finer tuning of the system by iterating only through one to twenty neurons in steps of one unlike the previous iteration of one to 100 in steps of five.

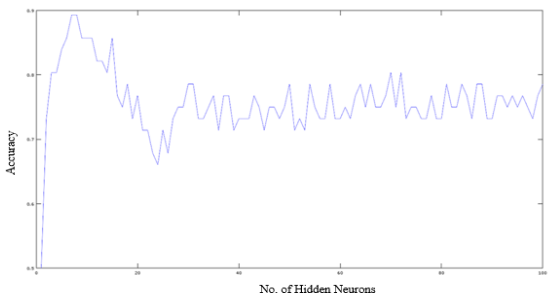

Fig. 5 Accuracy v/s hidden number of neurons

It must be noted that the effect of σ was not seen in the shape of the accuracy v/s number of hidden neurons graph, since the accuracy value always peaked in the one to twenty neurons range. However, variation in σ caused an overall increase or decrease in the graph's accuracy value. Thus, a smaller but finer set of iterations could be performed over the number of hidden neurons, but the full length of σ still needed to be iterated to obtain high accuracies.

The tables given in this section tabulate the results obtained for specific groups of feature sets. The accuracies and optimal number of hidden layer neurons and σ obtained for the different types of feature sets are as given the table below Table 3.

Features Used	Optimal (No. of Hidden Layer Neurons, $\boldsymbol{\sigma}$)	7-Fold Accuracy
Time, non-linear, frequency	(6, 1.4)	89.29%
Time, non-linear, frequency, ARX order 3	(7, 2)	92.86%
Time, non-linear, frequency, ARX order 5	(15, 1.8)	94.64%

Table 3. 7-fold Accuracy of ELM for different feature sets

Similar to the SVM classifier, the table Table 3 shows that the accuracy of the system improves when ARX coefficients are used along with the time domain, non-linear and frequency domain features.

Again, the confusion matrix was obtained and the sensitivity and specificity were calculated for each of the cases given above in Table 4.

Prediction/ Actual	Actual +ve	Actual -ve	Sensitivity/Specificity
Predicted +ve	26	2	92.86%
Predicted -ve	4	24	85.71%

Table. 4 Confusion matrix, sensitivity and specificity of ELM with ARX order 5 features

The confusion matrix above shows how the system is classifying samples from both classes well and thus, is not biased.

3.4 Comparative Results and Points of Discussion

With the results that have been obtained, it becomes clear that HRV analysis is a very viable method to extract features from ECG signals for the application of classifying alcoholics and non-alcoholic subjects. A comparison made between the SVM and ELM classifiers in Table 6 and Table 7 show that the ELM outperformed the SVM when trained without ARX features and when trained with the ARX features. Even the sensitivity and specificity of the ELM classifier was consistently superior to that of the SVM.

Feature Used	SVM	ELM
Without ARX Coefficients	80%	89%
With ARX Coefficients of order 5	86%	94%

Table 6. Comparative results of SVM and ELM with and without ARX (order 5) features

Parameter	SVM	ELM
Sensitivity	89%	92.86%
Specificity	82%	85.71%

Table 7. Comparative results of sensitivity and specificity of SVM and ELM

The most significant observation is the effect that the ARX features had on the classifiers. For both classifiers, ARX features yielded greater accuracies.

REFERENCES

[1] Ping Shi, Ying Chen, Ming-Ming Guo and Hong-Liu Yu, “Acute Effects Of Alcohol On Heart Rate Variability: Time-Related Changes And Gender Difference”, Biomedical Engineering: Applications, Basis and Communications,Vol. 26, No. 3 (2014) 1450048 (10 pages)

[2] Kusuma Ramanna1, Fazal M Gahlot2, Nagaraja Puranik1, “Electrocardiogram changes and heart rate variability during moderate exercise in chronic alcoholics”, International Journal of Medical Science and Public Health Vol 4, Issue 4 (2015) pp. 492-495

[3] Phyllis K. Stein, et. al., “Heart Rate Variability and Measure of Autonomic Tone”, American Heart Journal, vol. 127 no. 5 (Sept. 1993) pp. 1376-1381

[4] Johnson RH, Eisenhofer G, Lambie DG. “The effects of acute and chronic ingestion of ethanol on the autonomic nervous system.”, National Center for Biotechnology, (Dec. 1986), pp. 319-28.

[5] Jon T. Ingjaldsson, Jon C. Laberg, and Julian F. Thayer, “Reduced Heart Rate Variability in Chronic Alcohol Abuse: Relationship with Negative Mood, Chronic Thought Suppression, and Compulsive Drinking”, Society of Biological Psychiatry, (2002), pp. 1427-1436

[6] Katsuyuki Murata, Philip J. Landrigan, and Shunichi Araki, “Effects of age, heart rate, gender, tobacco and alcohol ingestion on R-R interval variability in human ECG”, Journal of the Autonomic Nervous System, 37 (1992) pp.199-206

[7] Chung Kit Wu, et. al. “A Precise Drunk Driving Detection Using Weighted Kernel based on Electrocardiogram”, *Sensors*. [Online]. 16(5), pp659. Available: http://www.mdpi.com/1424-8220/16/5/659/htm

[8] Mika P. Tarvainen and Juha-Pekka Niskanen, “Kubios HRV Analysis version 2.0 beta USER’S GUIDE”, Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Finland

[9] Andrew Ng, “Support Vector Machines”, 2011. [Online] Available: http://cs229.stanford.edu/notes/cs229-notes3.pdf Accessed: 10-Feb-2016

[10] Andrew Ng, “The Simplified SMO Algorithm”, 2012. [Online] Available: http://cs229.stanford.edu/materials/smo.pdf

[11] G.-B. Huang, “What are Extreme Learning Machines? Filling the Gap between Frank Rosenblatt's Dream and John von Neumann's Puzzle,” Cognitive Computation, vol. 7, pp. 263-278, 2015.