**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.

**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], [3-2], and [5-3]. 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 [add ref].

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 [4-4] and [6-5] 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 [7-6] 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. Mention that: ASSUMPTION THAT HRV IS NOT DUE TO OTHER FACTORS like heart problems?

What else needs to be elaborated in introduction? Is the transitioning correct?

1. **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 [8-7]. Four types of features have been obtained. They are obtained as mentioned in Tables add no, add no, add no. Time domain features utilize the inter-beat interval (RR interval series) to obtain a measure of the variability in the subject’s heart rate. Measures like the mean and standard deviation of the RR interval series, the mean and standard deviation of the mean hear rate and a count of the number of intervals varying by more than 50ms are some of the time domain features. The complete list of time domain features used is given in Table add no. Frequency domain features take the power spectral density of the RR interval sequence and utilize power contained in different frequency bands as features for the classifiers. A list of the frequency domain features used is given in Table add no. Non-linear features include the Poincare plot [add ref kubios manual] and approximate entropy [add ref kubios manual]. The Poincare plot is a method used to related the RRi-th interval to the RRi+1-th interval. It calculates the standard deviation along two different axes (x = y and x = -y) to obtain a measure of changes occurring in successive RR intervals.

While the three types of features mentioned above provided good k-fold cross validation accuracy, another type of feature was extracted using auto-regressive modelling. This has been described in the next section, Section 2.3.1.

**2.3.1 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 can be represented as:  
where are the model coefficients and 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. TALK ABOUT MISFIT DECREASING TILL ORDER 5 AND THEN INCREASING FOR HIGHER ORDERS ……. THUS ORDER 5 WAS USED …. IT ALSO CORRESPODED TO THE HIGHEST ACCURACY THAT THE CLASSIFERS PROVIDED (K-FOLD ACCURACY OF ORDER 3 COEFFICIENTS WASN’T AS GOOD AS ORDER 5).

|  |  |  |  |
| --- | --- | --- | --- |
| Sl. No. | Feature | Formula | Description |
| 1 | RR\_mean |  | The mean of the RR interval sequence |
| 2 | RR\_std |  | The standard deviation of the RR interval sequence |
| 3 | HR\_mean |  | The mean heart rate |
| 4 | HR\_std |  | The standard deviation of the heart rate |
| 5 | RR\_rms |  | The RMS of the RR interval series |
| 6 | RR\_50 | – | Number of RR intervals that are larger than 50ms |
| 7 | RR\_r50 |  | Normalized number of RR intervals that are larger than 50ms |

Table add no. Time Domain Features

|  |  |  |
| --- | --- | --- |
| Sl. No. | Feature | Description |
| 1 | SD1 | Standard deviation of Poincare plot along x = y line |
| 2 | SD2 | Standard deviation of Poincare plot along x = -y line |
| 3 | ApEn | Approximate entropy of RR interval sequence [Add Ref] |

Table add no. Frequency Domain Features

|  |  |  |  |
| --- | --- | --- | --- |
| Sl. No. | Feature | Formula | Description |
| 1 | pk\_freq\_vlf | – | Frequency at which the peak of PSD is seen for the VLF, LF and HF frequency bands |
| 2 | pk\_freq\_lf |
| 3 | pk\_freq\_hf |
| 4 | ab\_pow\_vlf | – | Absolute power of the VLF, LF and HF frequency bands |
| 5 | ab\_pow\_lf |
| 6 | ab\_pow\_hf |
| 7 | pw\_ttl | – | Total power contained in the signal |
| 8 | rp\_vlf |  | Ratio of the power in a particular band to the total power in the signal |
| 9 | rp\_lf |
| 10 | rp\_hf |
| 11 | norm\_lf |  | Ratio of power in a particular band to the power of the signal without considering contribution of power due to VLF band |
| 12 | norm\_hf |
| 13 | ratio |  | Ratio of the absolute power in the LF band to the absolute power of the HF band |

Table add no. Frequency Domain Features

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 [9-8] in order to better separate the data points in a higher dimension. The SVM used the Simplified SMO algorithm [10-9] 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 and for such a labelled dataset was obtained by looking at which pair yielded the highest k fold cross validation accuracy. After all the 200 loops had been iterated, the pair of and 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. This was how the most optimal was selected for any feature set used to train the SVM.

**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 [11-10], 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 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 were obtained from the confusion matrix.

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**

Results pertaining to the Pre-processing of signals using wavelet decomposition, all four types of feature extraction, and cross validation accuracies obtained for the SVM and ELM 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 without ambiguity.

The figure (add fig no) shows the 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 level-8 approximation obtained after Wavelet analysis on the ECG signal. The de-trended signal (ie. signal obtained on removing the level-8 components from the original signal) is shown in the third subplot. The RR-peaks 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 results in coefficients. In the array given above ARX coefficients for a fifth order system has been provided. The fifth order system provides eight coefficients, however only six are used as two coefficients are the same (either always one or always zero) for all the samples.

**~~3.3 Results of SVM~~**

The process of selecting the and pair, training the SVM, obtaining the k-fold validation accuracy have been mentioned in Section 2.4.1. As an example, the output of the two hundred iterations (per value considered) on the feature set with ARX order 5 features ~~to obtain a count of which was chosen most and yielded the highest k-fold cross validation accuracy for the algorithm~~ is given in Fig. 4.

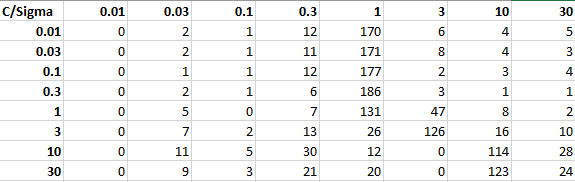
****

Fig. 4 Distribution of choice of regularization parameter and RBF variance for SVM when features (with ARX 5 were used) ADD a column to the right with the sums of all row? ALSO replace image with a regular table

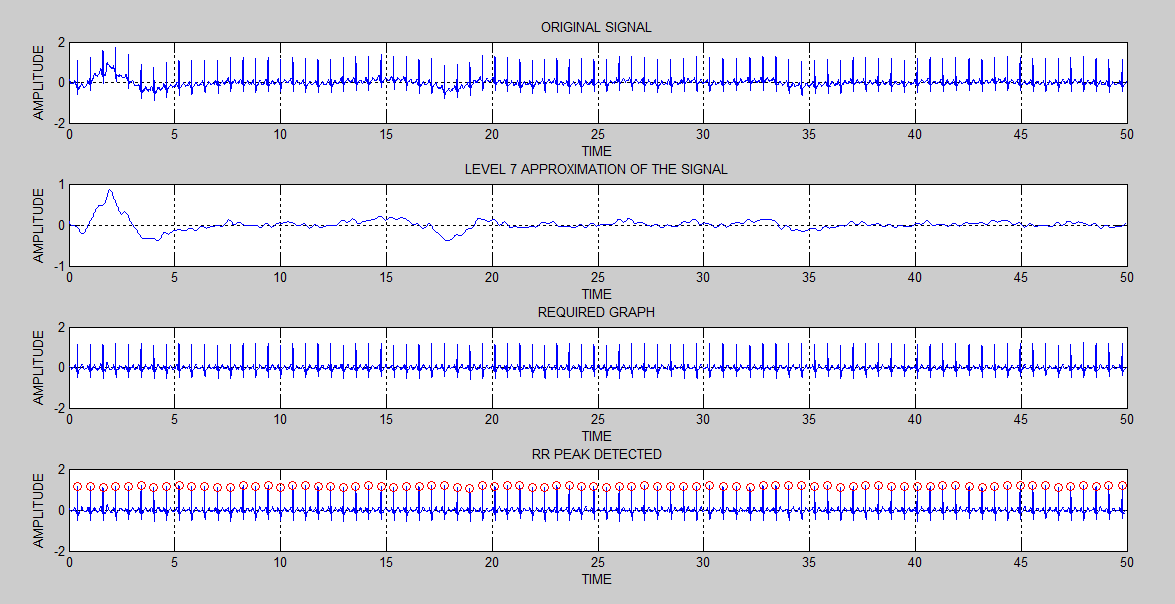


Fig. 3 Wavelet decomposition for baseline wandering removal [the title of subplot 2 says it is 7th level of approximation]

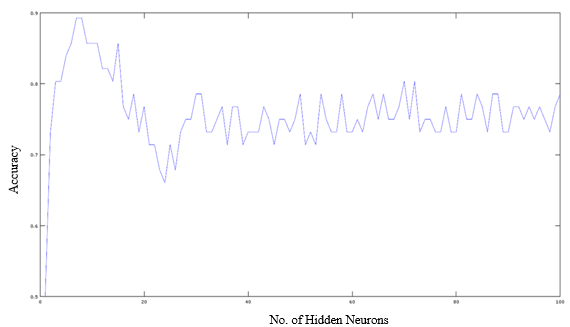
****

Fig. 5 Accuracy v/s hidden number of neurons

From the distribution of Fig. 4 it is clear that pair yielded the highest accuracy the most number of times and thus worked best for feature set that included ARX order 5 features. ~~In a similar manner, C and were picked for any feature set used to train the SVM.~~

~~In the tables that follow, the results obtained for different groups of feature sets. The accuracies and optimal and pair obtained for the different types of feature sets are as given the table below Table 1.~~ Keep in appendix or something? Useful to make the paragraph more meaningful/easy to understand.

|  |  |  |
| --- | --- | --- |
| ~~Features Used~~ | ~~Optimal (, ) Pair~~ | ~~8-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.~~

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 calculated have been provided in Table add table number of the spec and sense SVM/ELM Comparison Table? ~~(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. [remove this paragraph also?]

**~~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 to train and validate the ELM have been provided in Section 2.4.2. A graph representing the accuracy versus hidden neurons plot Fig. 5 for a specific shows the accuracy of the algorithm peaks when the number of hidden neurons are maintained between one and twenty. This behaviour of the accuracy peaking for the range one to twenty neurons is seen to generalize for any value considered, since similar plots (with significantly higher peaks in the one to twenty neuron range) were obtained irrespective of . The analysis drawn from such a plot allows finer tuning of the system by iterating only through one to twenty neurons in steps of one unlike the previous iterations of one to 100 in steps of five.

It must be noted that the effect of 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 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 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 obtained for the different types of feature sets are as given the table below Table 3.~~ Keep in appendix or something? Useful to make the paragraph more meaningful/easy to understand.

|  |  |  |
| --- | --- | --- |
| ~~Features Used~~ | ~~Optimal (No. of Hidden Layer~~  ~~Neurons, )~~ | ~~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 add table number of the spec and sense SVM/ELM Comparison Table?

|  |  |  |  |
| --- | --- | --- | --- |
| ~~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. [remove this paragraph also?]

**~~3.4 Comparative Results and Points of Discussion~~**

With the results that have been obtained, it is 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  (7-fold) | ELM  (7-fold) |
| Without ARX Coefficients | 80% | 89% |
| With ARX Coefficients of order 3 | 82% | 92% |
| 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. 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 provided useful in yielding 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. 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
5. 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
6. 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
7. 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
8. Andrew Ng, “Support Vector Machines”, 2011. [Online] Available: http://cs229.stanford.edu/notes/cs229-notes3.pdf Accessed: 10-Feb-2016
9. Andrew Ng, “The Simplified SMO Algorithm”, 2012. [Online] Available: http://cs229.stanford.edu/materials/smo.pdf Accessed: 10-Feb-2016
10. 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.

**APPENDIX**

**Change all tense to present tense**

**Table and Figure nos**

**Margins**