A Novel Approach for the Identification of Chronic Alcoholics from ECG Signals

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

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 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. An accuracy >80% was achieved by SVM and ELM without the use of ARX coefficients, while an accuracy >85% was achieved by both classifiers when ARX coefficients were included in the feature set.

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

I.   INTRODUCTION

The effects of consumption of alcohol in large amounts in a short period of time or gradually for a prolonged time in humans has been studied and documented extensively in [1], [3-2], and [5-3]. The Parasympathetic Nervous System (PNS) and Sympathetic Nervous System (SNS), are inputs to the Sino-Atrial (SA) Node in the heart which initiates a heart beat and in turn controls the inter-beat-interval. The PNS is known to lower heart rate, while the 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 and decreased PNS activity [add ref]. This causes an increase in the heart rate and decrease in the 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 that alcoholic and normative subjects can be differentiated based purely on HRV.

Research work for the detection of drunk driving [add ref: svm and drunk driving] utilizes HRV analysis to extract certain time domain features to solve the classification problem using a support vector machine (SVM). [abrupt end]. Certain other studies [add ref: link1], [add ref: link2 and link3] also use HRV analysis to extract another type of features, non-linear features, for their classification algorithms. Recent work done to detect the effect of cannabis on the ANS of paddy workers [add ref: indicon cannabis] also utilizes HRV analysis and obtains time domain and frequency domain features for the classifier.

ARX applications…of the past

ARX has been used for so many other reasons but not for alc identification… here we have used.

ARX coeffs have never been fed as features to classifiers have they?

As done in previous studies, in this study also HRV analysis has been used to extract time domain, non-linear and frequency domain features. However, these features have been used for the purpose of training the extreme learning machine (ELM) and comparing its results with that of the support vector machine (SVM). It is seen that the ELM performs better

than the SVM by a margin of 5% to 10% [keep or remove sentence?].

Along with the above set of features, the use of a new set of features is proposed in this study to improve the accuracy of the classifiers. An Autoregressive Model with Exogenous Inputs (ARX) is developed using the ECG signal of alcoholic and normative subjects, and the model coefficients obtained are used as additional features to the classifiers to improve the classifiers' accuracy. Use of ARX model coefficients as features for the SVM and ELM classifiers improved their accuracies by >5% each[keep or remove sentence?].

This study ensures that the test subjects whose ECG signals have been used to train the classifiers are free from heart conditions such as cardiomyopathy, atrial fibrillation or pre-mature ventricular contraction, etc causing arrhythmia. This step is crucial to ensure that heart rate variations that are measured and used for the purpose of classification arise primarily due to the effect of alcohol on the subjects heart beat and not due to other pre-existing conditions.

*1*We have used ARX

*2*We have applied to both SVM and ELM to show it generalizes over all classifiers

*3*Note that ELM outperforms

II. METHODOLOGY

This section covers the steps that were followed to filter ECG signals, extract features, train the classifiers and validate their behaviour (add fig no: Fig. 1).
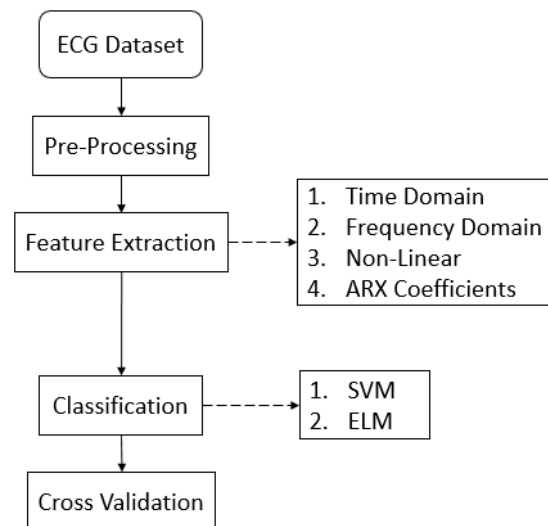

Fig. 1 Methodology

2.1 Dataset Description

To classify test subjects as alcoholics or normative, the classifier was trained with ECG data recorded at the Autonomic Lab, Department of Neurophysiology, NIMHANS, Bengaluru on ensuring informed consent adhering to the Declaration of Helsinki. The dataset consists of 56 ECG samples, of which 28 samples are of chronically alcoholic patients and the other 28 samples are that of normative people. The ECG samples had a sampling frequency of 1kHz and each sample was recorded for approximately 5 minutes (300 seconds), to capture the activity of the autonomic nervous system (ANS) [add ref: hrv explanation manual] which is primarily responsible in initiating/controlling the heart beat.

2.2 Pre-processing [shorten?]

The ECG dataset contains disturbances like baseline wandering and power-line noise which need to be removed to obtain clean ECG signals. Wavelet decomposition, a technique to break up a signal into shifted and scaled versions of itself [add ref: wavelet reference], is used to remove such sections of the signal.

Here, discrete wavelet transform with Daubechies wavelet is used. The Daubechies wavelet works well for the ECG dataset because the shape of the QRS complex in the ECG signal and the Daubechies wavelet match/resemble one another closely [add ref: ref [5] of cannabis paper? Or find another one… atleast verify and use]. Eight level wavelet decomposition is performed on the signal, and

removal of the eighth components from the original signal rids the ECG signal of baseline wandering. The power-line noise has a much smaller amplitude than the total swing in the ECG signal, and is left unfiltered.

2.3 Feature Extraction [The main reference for the medical significance doesn't have other references cited in it]

Features are extracted from the filtered signal using some of the techniques mentioned in [8-7]. Four types of features have been obtained.

Time Domain: Time domain features (add table no) utilize the inter-beat interval (RR interval series) to obtain a measure of the variability in the subject's heart rate. Measurements like the standard deviation of the RR interval series shows how quickly the heat is able to adapt to minor changes in the body and how much HRV is there in the subject. It is seen that HRV decreases in chronic alcoholics [add ref nos] which is detected primarily by measuring the standard deviation on the RR interval sequence. Other measurements that are used as features like the root mean square of the RR interval sequence give information about the subject's parasympathetic nervous system [add ref: clinical guide].

Table. 1 Time Domain Features

	Time Domain Features
1	$RR_{mean} = \frac{1}{n}\left(\sum_{i=0}^{n} RR_i\right)$ The mean of the RR interval sequence.
2	$RR_{std} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n}(RR_i - RR_{mean})^2}$ The standard deviation of the RR interval sequence.
3	$HR_{mean} = \frac{60 \times 1000}{RR_{mean}}$ The mean heart rate.
4	$HR_{std} = \frac{60 \times 1000}{RR_{std}}$ The standard deviation of the heart rate.
5	$RR_{rms} = \sqrt{\frac{1}{n-1}\sum_{i=0}^{n-1}(RR_{i+1} - RR_i)^2}$ The RMS of the RR interval series.
6	RR_{50} Number of RR intervals that are larger than 50ms.
7	$RR_{r50} = \frac{RR_{50}}{n-1}$ Normalized number of RR intervals that are larger than 50ms.

Non-Linear: Non-linear features (add table no) include the Poincare plot [add ref kubios manual] and approximate entropy [add ref kubios manual]. The Poincare plot is a graphical method to visualize the HRV of the subject [add ref: Poincare ref], and the approximate entropy gives a measure of the inter-beat irregularity [add ref: ApEn for ECG signals … Kubios HRV manual?]. ~~(Little more on Poincare?...more clumping and less spread for alcoholics)~~

Table. 2 Non-Linear Features

	Non-Linear Features
1	SD_1 – standard deviation of Poincare plot along x = y line.
2	SD_2 – standard deviation of Poincare plot along x = – y line.
3	$ApEn$ – Approximate entropy of RR interval sequence [Add Ref].

Add caption?

Frequency Domain: Frequency domain features (add table no) take the power spectral density of the RR interval sequence and utilize power contained in different frequency bands as features for the classifiers [add ref: hrv kubios guide]. These features reflect the activity of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS) [add ref: clinical guide] which are the two constituents of the ANS, which is the portion of the nervous system that controls the heartbeat [add ref: which ?].

Table. 3 Frequency Domain Features

	Frequency Domain Features
1	$pk_{freq_{vlf}}, pk_{freq_{lf}}, pk_{freq_{hf}}$
2	Frequency at which the peak of PSD
3	occurs for the VLF, LF and HF frequency bands.
4	$ab_{pow_{vlf}}, ab_{pow_{lf}}, ab_{pow_{hf}}$
5	Absolute power of the VLF, LF and HF
6	frequency bands.
7	pw_{ttl} Total power contained in the signal.
8	$rp_{vlf} = \frac{ab_{pow_{vlf}}}{pw_{ttl}}$
9	
10	Ratio of the power in a particular band to the total power in the signal.
11	$norm_{lf} = \frac{ab_{pow_{lf}}}{pw_{ttl} - ab_{pow_{vlf}}}$
12	Ratio of power in a particular band to the power of the signal without considering contribution of power due to VLF band.
13	$ratio = \frac{ab_{pow_{lf}}}{ab_{pow_{hf}}}$

	Ratio of the absolute power in the LF band to the absolute power of the HF band.

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:

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \cdots + a_p y(n-p) + \varepsilon(n)$$

where $a_{(k=1,2,...,p)}$ are the model coefficients and $\varepsilon(n)$ is a white noise series [add ref no].

In this study, 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 first and second halves of a signal are loaded into the System Identification Toolbox in MATLAB to calculate the ARX coefficients. These coefficients are ~~later~~ used as additional features to train the classifier.

ADD ARX block diagram…x, y, y-cap, polynomial p

ALSO check the following paragraph to ensure it matches with the figure

For a given input-output signal pair that is fed to the ARX system, a polynomial is generated that tries to use the input signal provided along with the polynomial generated, to estimate an output signal (y-cap) that best matches the output signal (y) fed to the system[add ref: MATLAB page?]. The polynomial generated may be of different orders, and some orders may provide a better fit (to the expected putput signal (y)) than others [add ref: MATLAB page?]. The order of the ARX polynomial that provides the least misfit while also providing the highest cross validation accuracy when used as features for the classifiers needs to be used.

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 gradual change that occurs through the span of the first half of the ECG signal to the second half. In essence, such a model seems to be another form of HRV analysis. While time domain, frequency domain and non-linear features performed HRV analysis over short spans of the ECG signal, the ARX model captures variations over the entire length of the signal.

An overview of the array of features including features for ARX model of order five is as follows:

RR_{mean}, RR_{std}, HR_{mean}, HR_{std}, RR_{rms}, RR_{50}, RR_{r50}, SD_1, SD_2, $pk_{freq_{vlf}}$, $pk_{freq_{lf}}$, $pk_{freq_{hf}}$, $ab_{pow_{vlf}}$, $ab_{pow_{vlf}}$, $ab_{pow_{vlf}}$, pw_{ttl}, rp_{vlf}, rp_{lf}, rp_{hf}, $norm_{lf}$, $norm_{hf}$, $ratio$, $ApEn$, 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 provides eight coefficients, however only six are used as two coefficients are the same (either always one or always zero) for all the samples.

2.4 Classifiers [Give one to two equations for SVM/ELM]

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 is implemented on the dataset of 28 alcoholic and 28 normative samples. An RBF kernel is used [9-8] in order to better separate the data points in a higher dimension. The SVM uses the Simplified SMO algorithm [10-9] to solve the Lagrangian problem and obtain the weights for the hyperplane. Finally, the SVM's performance is validated using k-fold ($k = 7$) cross validation (Section 2.5).

2.4.2 Extreme Learning Machine

The Extreme Learning Machine (ELM) is also trained on the same dataset as the SVM, and again, an RBF kernel is applied to the dataset before feeding it to the ELM algorithm.

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 classifier 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 (not most/best for sure… just trying to increase the probability of getting the best i/p wts) accurate classifier for the given dataset, the ELM algorithm was trained several times for different randomly generated input weights and the input weights yielding the best/highest 7-fold cross validation accuracy were used.

The accuracy of the ELM varied based on the number of hidden neurons used. A graphical plot of the accuracy of the algorithm versus the number of hidden neurons showed that the accuracy peaked when the number of hidden neurons being used was five to twenty. (Add: Fig no.).
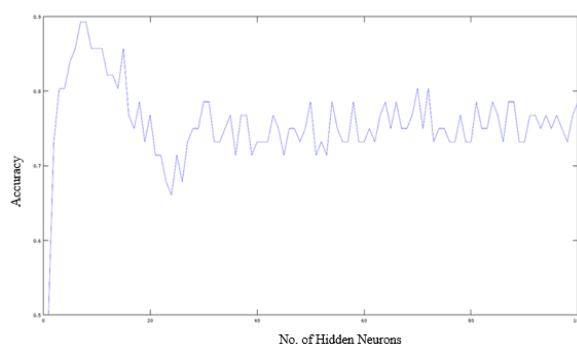

Fig. 5 Accuracy v/s hidden number of neurons [Make it small…fit in single col] [Make the X and Y axis clearer … Paint … tiff format]

2.4.3 Validation

While the training the classifiers, 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.

The dataset is divided randomly into 7-folds. One fold is used as the validation set, while the remaining folds are used to train the system. Next, the first fold is returned to the training set and a different fold (2nd fold) is used to test the algorithm's accuracy. This process is carried out cyclically to yield 7 accuracies, the average of which gives the 7-fold cross validation accuracy.

The sensitivity and specificity of the classifiers was also calculated to ensure the system was not biased and the classification accuracy remained high for the positive and negative classes.

III.   RESULTS AND DISCUSSION

Results pertaining to the Pre-processing of ECG signals using wavelet decomposition, selection of the most optimal order of ARX coefficients to include in the feature set, and cross validation accuracies obtained for the SVM and ELM have been provided in this section.

The output of applying wavelet decomposition in the pre-processing stage can be seen in (add: Fig. 3). The first subplot shows the original ECG signal with baseline wandering, the second subplot shows the level-8 approximation of the baseline wandering, and the third subplot shows the filtered signal without baseline wandering. The final subplot shows the detected R-peaks of the ECG signal.
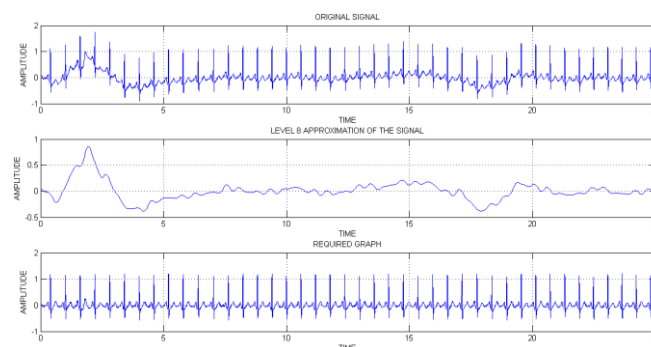

Fig. 3 Wavelet decomposition for baseline wandering removal

As mentioned earlier (Section 2.3 Feature Extraction), ARX coefficients of different orders provide different signal approximation accuracies/misfit. For the ECG dataset used, it is seen that the ARX model's misfit percentage decreases with increase in the order of the ARX model's polynomial. The least misfit is observed for order five, after which any further increase in the order results in increased misfit (add fig no.).

It was seen that using ARX coefficients of the order with the least misfit percentage also yielded the highest cross validation accuracy for the classifiers. Thus, ARX coefficients of order five have been used as features.

The SVM and ELM algorithms provided accuracies of 80% or greater when only time domain, frequency domain and non-linear features were used to train them. However, the inclusion of ARX coefficients in the feature set increased the accuracy of both algorithms by 5% or greater (add: table no Table. 6).

Feature Used	SVM (7-fold)	ELM (7-fold)
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

The sensitivity and specificity of both algorithms is seen to be well above 80% percent and this verifies that neither algorithms is biased toward either of the two classes (add table no: Table. 7).

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

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

ELM results does not mention of neurons used …. INCLUDE this

Does not highlight the use of ELM to alcoholic ECG dataset

IV.   CONCLUSION

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 shows that the ELM outperforms the SVM when trained with and without ARX features. Even the sensitivity and specificity of the ELM classifier is consistently superior to that of the SVM.

The most significant observation is the effect that the ARX features had on the classifiers. The inclusion of ARX coefficients in the feature set generalized well/worked well with both classifiers and yielded an improvement of 5% or greater in their accuracy.

FUTURE WORK

In the current study, the input weights, number of hidden neurons and hidden neuron weights for the ELM algorithm is optimized by 'brute force' through multiple trial-error steps. The plan ahead, is to prevent this randomness and utilize a technique called Meta-Cognitive Learning on ELM to make it learn more intelligently and accurately the first time around itself. Another point that can be worked on to further improve the accuracy of the classifiers, is to use non-linear ARX model coefficients. The final goal of the study is to be able to convert this binary classification paradigm into one where the algorithm is able to classify subjects into multiple classes based on the level of alcohol intake.

ACKNOWLEDGEMENT

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.

[11] Link1: https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-10-96 --- also uses Kubios!!

[12] Link2: http://ieeexplore.ieee.org/document/6968443/

[13] Link3: http://ieeexplore.ieee.org/document/7019200/

[14]

APPENDIX

Change all tense to present tense

Table and Figure nos

Margins

ARX coefficients section:

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).

Don't rely only on ALC (1) to set the misfit orders. Try to use the others as well

The HRV explanation manual given by apoorv has lots of info but no reference about where it was obtained from

Title is : Novel approach … so talk about ARX only and have just the list of time, freq, non-lin features … maybe with one or two lines to describe their purpose combined? … No need to give even this much detail?

Tables must have headings and not caption

ADD reference for SVM, ELM and ARX current use…in intro

Save all diags in TIFF format

READ NOTES in INDICON paper

How much to reduce/merge in SVM and ELM sections?

Keep it as chronic alc and normative OR chronic alc and non-alc?

**Results Discussion too brief?? But it seems to cover everything

** Change RR_mean to NN_mean??