Supraventricular Tachycardia Detection via Machine Learning Algorithms

Zhi Li[1], Harm Derksen[2], Jonathan Gryak[1], Mohsen Hooshmand[1], Alexander Wood[1], Hamid Ghanbari[4], Pujitha Gunaratne[6], and Kayvan Najarian[1,3,5]

***Abstract*— Methods for the classification and detection of supraventricular tachycardia were developed based on ECG signals collected from publicly available databases. Two different approaches were pursued, either by applying machine learning algorithms to heart rate variability features extracted from the ECG, or by applying a Markov chain model directly to the ECG signal.**

Keywords - Supraventricular Tachycardia, Machine Learning, Markov model, ECG, signal processing

I. INTRODUCTION

Arrhythmia is a group of conditions where the heart beats too fast, too slowly, or erratically. Supraventricular tachycardia (SVT) is a general term describing a subgroup of arrhythmias whose mechanism involves or originates above the atrioventricular node. The incidence of SVT is approximately 35 cases per 100,000 patients with a prevalence of 2.25 cases per 1,000 in the general population [1]. SVT has different arrhythmia related symptoms ranging from nonexistent to severe. Symptoms include palpitation, fatigue, light-headedness, chest discomfort and dyspnea, while on the other extreme certain paroxysmal SVT could also be asymptomatic.

Numerous arrhythmia classification algorithms have been developed during these years, include: SVM [2], [3], [4], [5], auto-regressive modeling [6], hidden Markov model (HMM) [7], [8], set of rules set by cardiologists [9], [10], optimal path forest [11], artificial neural networks [12], [13]. SVT has several subtypes, ranging from the generally benign sinus tachycardia and paroxysmal supraventricular tachycardia (PSVT), through unifocal and multifocal atrial tachycardia (MAT), to the more serious conditions of atrial flutter and atrial fibrillation (AFib) [14].

While most of these algorithms for SVT rhythm detection have been focused on AFib or atrial flutter since they are the most common and severe types of SVT, in this study we are focusing on the other more general types of SVT. We propose to classify the SVT vs. non-SVT dichotomy through two different approaches. In the first approach we will use heart rate variability features with machine learning algorithms for SVT detections; in the second approach we will use a Markov model for detection.

[1] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[2] Department of Mathematics, University of Michigan, Ann Arbor, MI, USA
[3] Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, USA
[4] Department of Internal Medicine - Cardiology, University of Michigan, Ann Arbor, MI, USA
[5] Michigan Center for Integrative Research in Critical Care University of Michigan, Ann Arbor, MI, USA
[6] Toyota Motor North America, Ann Arbor, MI, USA

II. DATA

Three publicly available datasets with SVT rhythm annotations were used in this study. The first dataset was the MIT-BIH Arrhythmia Database (mitdb) [15], which contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects enrolled in the study. The recordings were sampled at 360 per second. Two or more cardiologists have independently annotated each recording. The second database was the long-term AFib database (ltafdb) from physionet [15], [16]. This database includes 84 long-term ECG recordings of subjects with paroxysmal or sustained atrial fibrillation (AF). Each recording contains two simultaneously recorded ECG signals sampled at 128 Hz with duration around 24 to 25 hours. The third database was the MIT-BIH Malignant Ventricular Arrhythmia Database (vfdb) [15] sampled at 250 Hz. All three databases contain some annotated SVT episodes other than AFib or atrial flutter.

III. METHODS

We have used 2 different approaches for detecting SVT episodes. The first method is based on heart rate variability (HRV) and morphological features extracted from the signals, and uses machine learning algorithms based on such features for classification. The second method uses a Markov chain model which does not require feature extraction from the signal, hence the raw signal, after some standard signal processing steps, is directly fed into the algorithm.

The methods section will be divided into two main parts. In the first part we will describe the details about the HRV related features, and in the second part we will describe the Markov model for SVT detection.

A. Heart Rate Variability for SVT detection

1) Data pre-processing: During the signal pre-processing step, a fourth-order Butterworth bandpass filter with cutoff frequencies at 0.5 and 40 Hz was first applied to the raw ECG signal to remove noise, after which a double median filter with orders equal to 0.2 and 0.6 times the sampling frequency was applied to remove baseline wandering.

The signals from mitdb or vfdb were then down-sampled to 128 Hz to be consistent with ltafdb.

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

2) Data and Feature extraction: Episodes of SVT were extracted from the three databases based on the rhythm annotations. These SVT episodes were then separated into 10-second-long segments. We aim to detect these SVT events based only on the data contained within 10-second-long intervals.

There was a total of 667 10-second intervals. Among these episodes 611 were included in the training set and the remaining 56 episodes from a different patient cohort were reserved for the testing set. Additionally, 800 of the non-SVT episodes were randomly selected for the training set and 100 of non-SVT episodes were selected for the testing set.

The R-peaks and QRS complex were detected by the Pan–Tompkins algorithm [17] [18]. After identifying the R-peaks and QRS complex, we extracted a total of 16 features from the 10-second intervals (Table I).

HRV features	Description
ApEn	Approximate entropy, which measures the regularity and complexity of a time series
HR	Mean heart rate
SDHR	Standard Deviation of heart rate
RR	Mean RR interval
SDRR	Standard deviation of RR intervals
CV_RR	SDRR/meanRR intervals unitless scaled with factor 100
QS	Mean Q-S peak interval
SDQS	Standard deviation of Q-S peak interval
NN50	pairs of adjacent RR intervals differing that differ by more than 50 ms
pNN50	Percentage of successive RR intervals that differ by more than 50 ms
RMSSD	Root mean square of successive RR interval differences
SDSD	Standard deviation of differences between adjacent RR intervals
Poincaré SD1	Poincaré plot standard deviation perpendicular the line of identity
Poincaré SD2	Poincaré plot standard deviation along the line of identity
Lorentz OriginCount	Number of points (dRR(i-1),dRR(i)) within the radius normal sinus rhythm mask
Lorentz OriginCountRatio	Ratio between number of points (dRR(i-1),dRR(i)) within the radius normal sinus rhythm mask

TABLE I: HRV features

3) Machine Learning Algorithm: We applied several machine learning algorithms including random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN) on the training dataset with 5-fold cross validation. The cross validation procedure was performed on the training dataset only, enabling parameter optimization with respect to area under ROC curve (AUC) performance.

Among the three methods RF seemed to have achieved the best result on the training dataset (Fig1). Among the 16 HRV features, mean QS interval length, Poincaré SD2, mean RR interval and mean HR had the highest importance based on the trained random forest algorithm on the training dataset. Importance was calculated as the weighted average of the differences between the prediction accuracies on the out-of-bag portion or the entirety of the data. Regression was performed based on the weighted average of the differences of the respective MSEs instead.

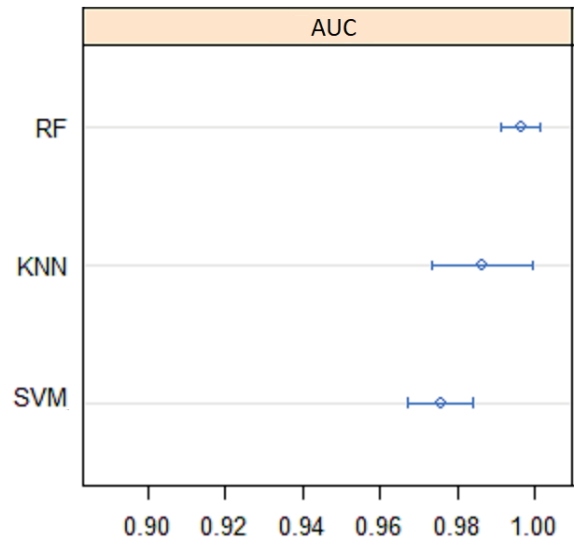

Fig. 1: Compare machine learning algorithms on training data

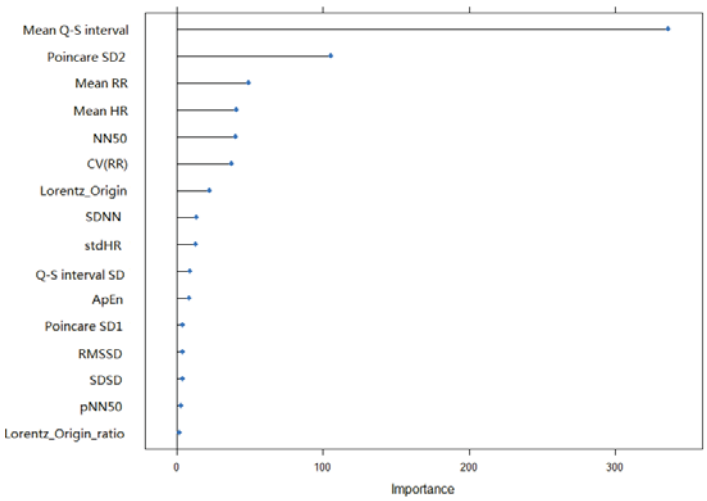

Fig. 2: Variable importance on training data

The RF model with the best AUC based on the training data was then applied on the testing data for evaluation.

B. Signal Encoding and Markov Model

A Markov chain algorithm for the classification and detection of SVT intervals has been developed. Unlike other more traditional Markov chain based models, in this model both the transition probabilities and the underlying structure, which in particular includes the state space, of the Markov chains involved are computed from the data via frequency analysis. Consequently this model is much more adaptive and more faithfully reflects the patterns contained within the data.

The input of the Markov Chain algorithm is a sequence of probability vectors

$$\begin{bmatrix} p_1 \\ q_1 \end{bmatrix}, \begin{bmatrix} p_2 \\ q_2 \end{bmatrix}, \begin{bmatrix} p_3 \\ q_3 \end{bmatrix}, \begin{bmatrix} p_4 \\ q_4 \end{bmatrix}, \ \cdots$$

where p_i denotes the peak probability and q_i denotes the non-peak probability at the ith sample point. To obtain the probability vectors, an additional round of signal processing techniques need to be applied. The various techniques involve:

1) *Baseline removal* The baseline removal step removes baseline drift from the ECG signal and enhance the R peaks in comparison to the T and P waves.
2) *Normalization* The ECG signal with baseline removed is normalized to have maximal absolute magnitude equal to 1.
3) *Non-linear filtering* Apply a non-linear filter to enhance R-peaks and suppress other peaks.
4) *Re-sampling of data* The signal is discretized and then re-sampled due to dimensionality considerations.
5) *Thresholding* A thresholding procedure is applied to systematically convert the signal into a sequence probability vectors.

Previous studies have shown that Markov chain models are capable of detecting AFib events [8]. Here we propose to apply a Markov model based algorithm on the problem of SVT detection.

The Markov model was applied to the training dataset to get the transition matrices and states. These transition matrices and states were then applied to the testing set to make predictions.

IV. Result

The result section begins by showing the algorithms' ability to detect annotated SVT episodes using HRV features and RF. Then we will also show the performance of the new Markov model for SVT detection.

A. HRV features with RF

There was a total of 667 10-second intervals. 611 of these episodes were included in the training set and the remaining 56 episodes from a different patient cohort were reserved for the testing set. The algorithm correctly classified 51 out of 56 SVT episodes with sensitivity of 91.1% and 100 out of 100 non-SVT episodes with specificity of 100% in the testing set. The F1 score was 0.95 and the AUC was 0.995.

		Annotated Label		
		SVT	Non-SVT	Total
Prediction	SVT	51	0	51
	Non-SVT	5	100	105
	Total	56	100	156

TABLE II: Confusion Matrix for SVT Detection with HRV and RF

B. SVT detection with Markov Model

We also performed the Markov model algorithm for SVT detection with publicly available databases. The process is summarized in Figure 4. We have correctly identified 41/56 SVT events (0.73 sensitivity) and 97/100 (0.97 specificity). The AUC was 0.951 and F1 score was 0.914.

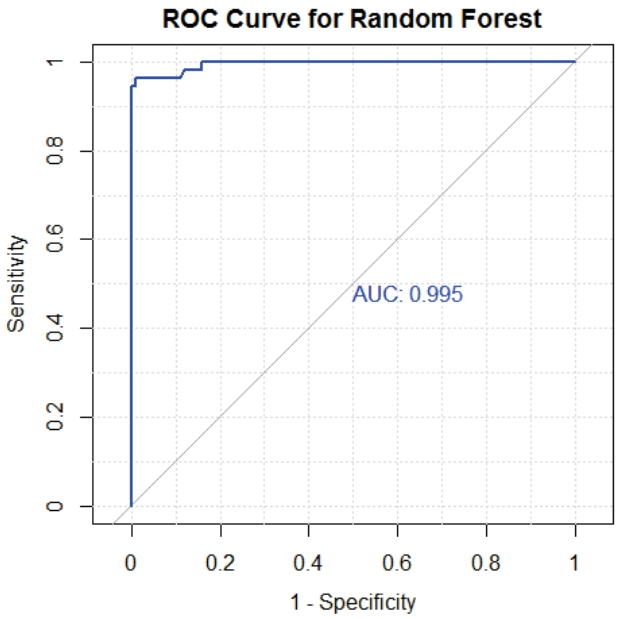

Fig. 3: SVT Detection with HRV features and RF

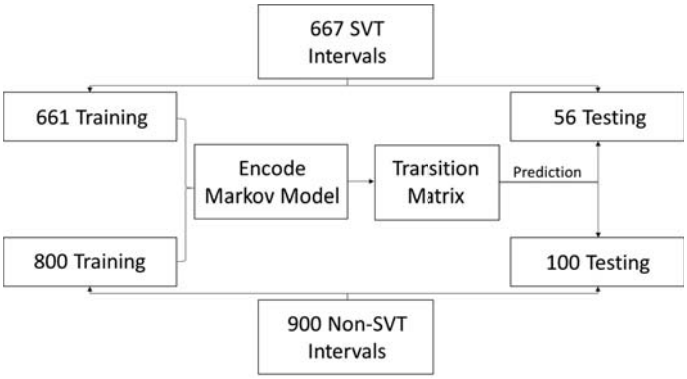

Fig. 4: Training and Testing Datasets

		Annotated Label		
		SVT	Non-SVT	Total
Prediction	SVT	41	3	44
	Non-SVT	15	97	112
	Total	56	100	156

TABLE III: Confusion Matrix for SVT detection Markov Model

V. Discussion

In this study, we have shown that HRV features with RF successfully detected the SVT episodes with 91.1% sensitivity and 100% specificity. The Markov model was able to detect the SVT episodes with high AUC of 0.95 and F1 of 0.91. The sensitivity using the Markov model is lower than using the HRV features with RF. However, the Markov model does not require calculation of hand-crafted features and can be applied easily on real-time data collected from portable devices.

As a next step, it would be very interesting to predict SVT events based on the ECG signals which are recorded up until several minutes before the onset of the SVT event

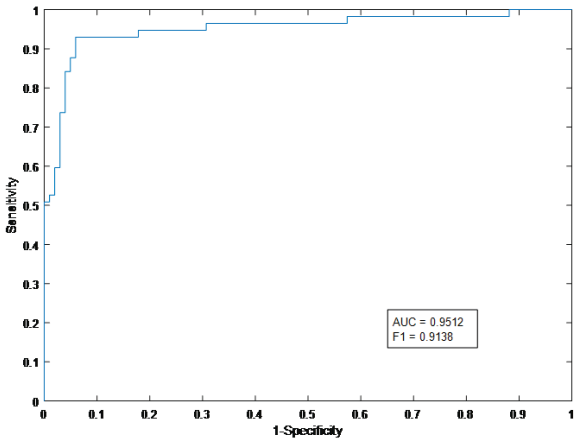

Fig. 5: SVT Detection with Markov Model

with real-time data from portable devices.

One limitation of the study is the lack of annotated samples with good quality labels. There are more databases with beat to beat annotations but not many of them has rhythm labels. Without these rhythm labels it is hard to classify different types of arrhythmia other than the beat types. We used the three publicly available databases with well annotated rhythms, however the number of SVT episodes we obtained from these records were still rather low. With limited amount of samples, the trained Markov model and RF model are more prone to problems such as generalizability and whether they can be directly applied to other datasets. Although in the testing set, we tested the algorithms on a different patient population, more ECG data with quality SVT labels are required to further validate the programs.

VI. Conclusion

SVT is a group of arrhythmia which includes AFib and atrial flutter. While most of the current studies have been focusing on AFib and atrial flutter, we proposed 2 approaches for detection of other types of SVT. This study aims to develop algorithms to classify and detect SVT using ECG signals collected from 3 publically available databases. Among the two methods considered, the HRV based features yield an overall accuracy of 91.1% for prediction with 0.995 AUC, while the Markov model has an accuracy of 73% for prediction with 0.951 AUC. It would be interesting to extend the models to predict SVT events based on ECG recordings several minutes before onset.

Acknowledgement

The research work in this paper was funded by Toyota Motor North America.

References

[1] Daniel Sohinki and Owen A Obel. Current trends in supraventricular tachycardia management. *The Ochsner Journal*, 14(4):586–595, 2014.

[2] Can Ye, Miguel Tavares Coimbra, and BVK Vijaya Kumar. Arrhythmia detection and classification using morphological and dynamic features of ecg signals. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 1918–1921. IEEE, 2010.

[3] Jalal A Nasiri, Mahmoud Naghibzadeh, H Sadoghi Yazdi, and Bahram Naghibzadeh. Ecg arrhythmia classification with support vector machines and genetic algorithm. In *Computer Modeling and Simulation, 2009. EMS'09. Third UKSim European Symposium on*, pages 187–192. IEEE, 2009.

[4] Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, and Sun Kook Yoo. Support vector machine based arrhythmia classification using reduced features. *International Journal of Control, Automation, and Systems*, 3(4):571–579, 2005.

[5] N Ozlem Ozcan and Fikret Gurgen. Fuzzy support vector machines for ecg arrhythmia detection. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2973–2976. IEEE, 2010.

[6] Dingfei Ge, Narayanan Srinivasan, and Shankar M Krishnan. Cardiac arrhythmia classification using autoregressive modeling. *Biomedical engineering online*, 1(1):5, 2002.

[7] Douglas A Coast, Richard M Stern, Gerald G Cano, and Stanley A Briller. An approach to cardiac arrhythmia analysis using hidden markov models. *IEEE Transactions on biomedical Engineering*, 37(9):826–836, 1990.

[8] Rodrigo Varejão Andreão, Bernadette Dorizzi, and Jérôme Boudy. Ecg signal analysis through hidden markov models. *IEEE Transactions on Biomedical engineering*, 53(8):1541–1549, 2006.

[9] Markos G Tsipouras, Dirnitrois I Fotiadis, and D Sideris. Arrhythmia classification using the rr-interval duration signal. In *Computers in Cardiology, 2002*, pages 485–488. IEEE, 2002.

[10] TP Exarchos, MG Tsipouras, D Nanou, C Bazios, Y Antoniou, and DI Fotiadis. A platform for wide scale integration and visual representation of medical intelligence in cardiology: the decision support framework. In *Computers in Cardiology, 2005*, pages 167–170. IEEE, 2005.

[11] Eduardo José da S Luz, Thiago M Nunes, Victor Hugo C De Albuquerque, João P Papa, and David Menotti. Ecg arrhythmia classification based on optimum-path forest. *Expert Systems with Applications*, 40(9):3561–3573, 2013.

[12] Rahime Ceylan and Yüksel Özbay. Comparison of fcm, pca and wt techniques for classification ecg arrhythmias using artificial neural network. *Expert Systems with Applications*, 33(2):286–295, 2007.

[13] Shivajirao M Jadhav, Sanjay L Nalbalwar, and Ashok A Ghatol. Modular neural network based arrhythmia classification system using ecg signal data. *International Journal of Information Technology and Knowledge Management*, 4(1):205–209, 2011.

[14] Gary Hals and Chloe McCoy. Supraventricular tachycardia: A review for the practicing emergency physician. *Emergency Medicine Reports*, 36(13), 2015.

[15] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[16] Simona Petrutiu, Alan V Sahakian, and Steven Swiryn. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace*, 9(7):466–470, 2007.

[17] Jiapu Pan and Willis J Tompkins. A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236, 1985.

[18] Hooman Sedghamiz. Matlab implementation of pan tompkins ecg qrs detector. 03 2014.