Impact of Demographics on Short-term Heart Rate Variability for Detecting Hypertension

Muhammad Usman[1], Pradeep Rajagopalan[1], Aryel Beck[1], Jennifer Nathania[2], Tony Li[2], Toon Wei Lim[2]

[1] Singapore Technology Centre (SGTC), Panasonic Industrial Devices Singapore
[2] National University Heart Centre Singapore (NUHCS), Singapore

Abstract

The relationship between heart rate variability (HRV) and hypertension is well established in multiple studies. However, there is a lack of investigation on the impact of demographics and other diseases related to cardiovascular health on the performance of HRV based hypertension detection models. This study aims to address these issues by determining the efficacy of such models in an unconstrained setting. 24 hours long ECG were recorded for 1377 subjects. HRV features from time, frequency and non-linear domains were extracted from 1 minute long R-peak to R-peak intervals (RRIs). Demographic factors of age, gender and body mass index (BMI) were added one by one as additional features into logistic regression models. The performance of the models was analysed with respect to different age groups. The results show that inclusion of age into the HRV model increased its accuracy from 71.7% to 77.6%. However, the model's predictions were mostly similar to the ones that would be obtained with an age based threshold. This is due to the natural age bias in the data which makes age a confounder for HRV based hypertension detection. This highlights the importance of naturally occurring demographics imbalance and how this must be carefully considered when developing HRV models for hypertension.

1. Introduction

Hypertension is a well established substrate of many critical cardiovascular conditions [1]. In many cases, it remains undetected for long periods of time [1]. It has been shown that the autonomic nervous system is involved in the development of hypertension [2] and it can therefore be captured through HRV. Medical studies have confirmed the efficacy of HRV for hypertension detection [2]. Moreover, these findings have also been established by studies testing wearable sensor such as waist belt [3] or movement sensitive mattress [4].

The popularity of wearable sensors such as smart watches have enabled researchers to collect physiological data at large scale [5]. The demographic bias that naturally exists in conditions such as hypertension [1] is inevitable whenever the data collection happens in large uncontrolled environments. Demographic factors such as age, gender and BMI have also been shown to impact HRV [6,7] which makes them confounders for HRV based models. Although some researchers have discussed the effects of confounders on general medical classification models [8], there is a lack of understanding in regards to models developed for diseases such as hypertension. In some studies [3] [4], the normotensive and hypertensive groups were balanced with respect to demographic factors. However, the exclusion criteria for data collection was designed so that the normotensive group included fully healthy subjects only. This is potentially an issue as HRV can be affected by multiple physiological conditions simultaneously. Consequently, an evaluation on data excluding these conditions is not sufficient for a model that has to be developed for an unconstrained user population.

This study investigates the effect of naturally occurring demographic bias, in the data collected in an unrestricted environment, on HRV based models for hypertension. This aids in determining the physiological accuracy of the model as well. Moreover, a more realistic evaluation is performed for an HRV model that has to be deployed in an unrestrained environment. This is done by including data from subjects that suffer from other primary cardiovascular health conditions [1] that can be expected in the field.

2. Methods

2.1. Dataset

24 hours long Holter ECG recordings were obtained for 1377 subjects. 1238 were collected by the National University Hospital, Singapore and 139 were extracted from

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.084

the PhysioNet SHAREE database [9]. During the recordings, subjects were allowed to perform their normal daily activities. Table 1 summarizes the population distribution. The main motivation of this study was to investigate the performance of HRV based hypertension detection in uncontrolled environments. Therefore, subjects suffering from diseases other than hypertension were included in the normotensive group and subjects with comorbidities in the hypertensive group. In the context of this study, the subjects included cases of hyperlipidemia, diabetes mellitus and coronary artery disease. The prevalence of these diseases as a percentage of the normotensive and hypertensive subject count is shown in Table 1.

Table 1. Data Statistics.

Statistic	Normotensive	Hypertensive
Count	722	655
Age Mean/Std Dev	40.3/18.2	66.4/12.6
BMI Mean/Std Dev	23.7/4.9	25.7/5.1
Male/Female	405/317	373/282
Diabetes Mellitus	3.5%	30.8%
Hyperlipidemia	12.9%	50.5%
Coronary Artery Disease	4.7%	32.1%

2.2. HRV Processing

There were multiple processing steps involved to extract HRV features from raw Holter ECG. At first, the sampling frequency of all ECG signals was standardized to the same value using spline interpolation. The R-peaks of the standardized ECG signals were then detected by the Hamilton QRS complex method. This was implemented with the open source Python library BioSPPy [10]. This peak detection method was selected because it can tackle various issues including baseline wander and noise. The detected R-peaks were then differentiated to get RRIs. The whole RRI series for each subject was then split into 1 minute non-overlapping segments. To get rid of RRIs corresponding to ectopic beats and noise, outliers were detected in each 1 minute RRI segment using the physiologically normal RRI range between 400 msec to 2000 msec [11]. If the number of outliers was greater than 20% then the segment was discarded, otherwise outliers were replaced with normal values through linear interpolation.

HRV features from time, frequency and non-linear domain were then extracted from these filtered RRI segments. Total features extracted are listed in Table 2. Standard time domain features of average (AVNN), median (MDNN), coefficient of variation (cvNN) and standard deviation (SDNN) of RRIs were calculated. Full range of the pNNxx family of indices from 5 msec to 200 msec was included as it has been shown to exhibit promising results [12]. The distributional characteristics of RRIs and their absolute differences were also extracted. In the non-linear domain, apart from typical features, various entropies, Fisher information and Petrosian fractal dimension were also calculated. Frequency domain features included values and ratios of the four spectral power density bands; very low frequency (VLF: 0.003-0.04 Hz), low frequency (LF: 0.04-0.15 Hz), high frequency (HF: 0.15-0.4 Hz) and very high frequency (VHF: 0.4-0.5 Hz). Features selection was implicitly done by the weights learned by the machine learning model.

Table 2. HRV Features.

Time	Frequency	Non-linear
AVNN	VHF	DFA
SDNN	HF	Shannon Entropy
RMSSD	LF	Sample Entropy
MDNN	VLF	Multiscale Entropy
cvNN	Total power(TOT)	SVD Entropy
Mode of RRI	LF/HF	Spectral bands Entropies
RRI Skewness	VLF/HF	Fisher information
RRI Kurtosis	VHF/TOT	Petrosian Fractal dimension
RRI difference Skewness	HF/TOT	SD1
RRI difference Kurtosis	LF/TOT	SD2
17 pNNx indices	VLF/TOT	SD1/SD2

2.3. Modelling and Analysis

Each logistic regression model in this study classified a given 1 minute RRI segment as either belonging to a hypertensive or normotensive subject. Moreover, the impacts of age, sex and BMI on the HRV model performance were assessed by successively incorporating them into the model. These were added one by one as features. Each feature was standardized with respect to the mean and standard deviation of that feature in the training data. The validation approach for each model was 5-fold leave one group out in order to ensure that there is no overlap of subjects between the training and the testing data.

3. Results

Table 3 shows the accuracy and area under the receiver operating characteristic curve (AUC) of the models with different combinations of features. The results show that incorporating gender and BMI into the model had negligible impact on the model's performance. Including age improved accuracy from 71.7% to 77.6% and AUC from 0.80 to 0.87. Since the features were standardized, the coefficients of the logistic regression model were analysed to determine the relative order of importance. This indeed revealed that age was in the top 10 features. Moreover, the dataset showed an age difference between the normotensive and hypertensive groups (Table 1). This is why, the model's performance was further analysed by splitting the data into five distinct age groups. The performance metrics for the model including age as additional feature were then calculated separately for each of the five age groups and compared against the model without age feature. As can be seen in Table 4, the model with age included (termed as Model 1) classified almost all RRI segments belonging to subjects with age below 40 as normotensive. Moreover, the RRIs segments belonging to subjects older than 60 were primarily classified as hypertensive. On the other hand, the selectivity and specificity of the model with HRV, gender and BMI (termed as Model 2) is comparatively more balanced across all age groups.

Table 3. Performance of models with different combinations of HRV and demographic features.

Features incorporated	Accuracy	AUC
HRV	71.7%	0.80
HRV, Gender	71.8%	0.80
HRV, BMI	71.9%	0.80
HRV, Age	77.6%	0.87
HRV, Gender, BMI	72.0%	0.80
HRV, Gender, BMI, Age	77.6%	0.87

Table 4. Age group-wise performance of models.

Model	Age group	Sensitivity	Specificity
HRV, Age	<20 years	0.0%	100%
	20-40 years	0.1%	99.8%
	40-60 years	31.8%	74.8%
	60-80 years	94.6%	12.6%
	>80 years	100%	0%
HRV, Gender, BMI	<20 years	36.9%	89.2%
	20-40 years	21.0%	89.8%
	40-60 years	46.0%	63.0%
	60-80 years	80.6%	38.5%
	>80 years	82.5%	44.0%

4. Discussion

The results presented in Section 3 indicate that age is one of the most pro-eminent feature when incorporated to the model. As shown in Table 4, Model 1 is very close to a fixed age threshold. This is further suggested by the subjects' age distribution. Figure 1 shows that on this dataset, the model can attain a high accuracy by learning a simple age threshold. To check for any anomalies in the distribution, the ratio of hypertensive to normotensive subjects in various age groups shown in Figure 1 was compared with another study that enrolled 10215 subjects in Singapore [13]. The average difference in the ratio per age group was 3.3% which showed that the age distribution of the data in this study is applicable to the target population of Singapore. Whilst achieving higher overall performance, including such a confounding feature of age in the model can be problematic. Firstly, the resulting model behaves very similar to a simple age threshold. This would make it highly sensitive to data drift if the target population becomes different than that of the study population. Secondly and more importantly, the model is not physiologically accurate as it is not giving considerable importance to the HRV features which actually capture the changes in autonomic nervous system due to hypertension.

Even if age was not explicitly included as a feature, the impact of age on HRV could still be aiding the model [6]. This was indeed observed by dividing the population per age and analysing the Model 2 predictions per age group. Table 4 shows that whilst the model still predicts more higher age subjects' segments as hypertensive, there is no simple polarizing threshold. This bias is learnt by the model and is not specific to this study, it is a natural phenomenon. Whether such bias in the model is acceptable depends on the use case. The extent to which these models can be allowed to rely on such bias needs to be tuned for specific use cases. For example, if the model is to be deployed within a casual sensing or general well-being context, then this is fully acceptable. However, if the model were to be used for clinical diagnosis then this bias should be isolated and controlled.

5. Conclusion

The study has highlighted some important points about the impact of demographics, especially age on the performance of HRV based detection of hypertension. More

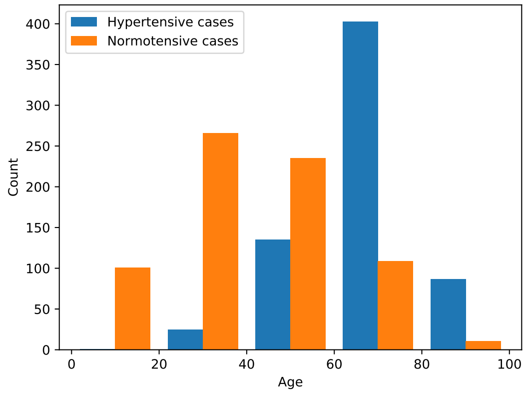

Figure 1. Age distribution of normotensive and hypertensive subjects.

precisely, it was shown that the explicit inclusion of Age (Model 1) can give a performance boost but the resulting model decides primarily based on age threshold. On the other hand, whilst still biased for age, Model 2 was still able to detect both conditions in the youngest and oldest groups ('$<$ 20' and '$>$ 80' in Table 4). Similarly, better performance for Model 2 could have been obtained by selecting only healthy subjects in the normotensive group and purely hypertensive in the other. However, this evaluation gives a more realistic picture of the capability of HRV based hypertension detection in an uncontrolled environment.

Finally, it should be noted that these results can be similar for other medical conditions that become more prevalent with age such as hyperglycaemia and hyperlipidemia. It is important to analyse the performance of these HRV based models with respect to their confounding factors in the data so that their physiological accuracy is not undermined in pursuit of better overall numerical performance.

References

[1] Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN, et al. Heart disease and stroke statistics—2020 update: a report from the american heart association. Circulation 2020;141(9):e139–e596.

[2] Schroeder EB, Liao D, Chambless LE, Prineas RJ, Evans GW, Heiss G. Hypertension, blood pressure, and heart rate variability: the atherosclerosis risk in communities (aric) study. Hypertension 2003;42(6):1106–1111.

[3] Ni H, Cho S, Mankoff J, Yang J, et al. Automated recognition of hypertension through overnight continuous hrv monitoring. Journal of Ambient Intelligence and Humanized Computing 2018;9(6):2011–2023.

[4] Liu F, Zhou X, Wang Z, Cao J, Wang H, Zhang Y. Unobtrusive mattress-based identification of hypertension by integrating classification and association rule mining. Sensors 2019;19(7):1489.

[5] Ballinger B, Hsieh J, Singh A, Sohoni N, Wang J, Tison G, Marcus G, Sanchez J, Maguire C, Olgin J, et al. Deepheart: semi-supervised sequence learning for cardiovascular risk prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32. 2018; .

[6] Voss A, Schroeder R, Heitmann A, Peters A, Perz S. Short-term heart rate variability—influence of gender and age in healthy subjects. PloS one 2015;10(3):e0118308.

[7] Windham BG, Fumagalli S, Ble A, Sollers JJ, Thayer JF, Najjar SS, Griswold ME, Ferrucci L. The relationship between heart rate variability and adiposity differs for central and overall adiposity. Journal of obesity 2012;2012.

[8] Ferrari E, Retico A, Bacciu D. Measuring the effects of confounders in medical supervised classification problems: the confounding index (ci). Artificial intelligence in medicine 2020;103:101804.

[9] Melillo P, Izzo R, Orrico A, Scala P, Attanasio M, Mirra M, De Luca N, Pecchia L. Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PloS one 2015;10(3):e0118504.

[10] Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, Fred A, et al. BioSPPy: Biosignal processing in Python, 2015–. URL `https://github.com/PIA-Group/BioSPPy/`.

[11] Rosenthal L, Gracia EA. Normal electrocardiography (ecg) intervals, May 2020. URL `https://emedicine.medscape.com/article/2172196-overview`.

[12] Villa MM, Niño CA, Mojica E, Acevedo J. Heart rate variability parameters and the pnnx family as discriminators between athletes and sedentary people. Journal of Physical Education and Sport 2016;16(4):1316.

[13] Liew SJ, Lee JT, Tan CS, Koh CHG, Van Dam R, Müller-Riemenschneider F. Sociodemographic factors in relation to hypertension prevalence, awareness, treatment and control in a multi-ethnic asian population: a cross-sectional study. BMJ open 2019;9(5):e025869.

Address for correspondence:

Muhammad Usman
3 Bedok South Road, 469269, Singapore
muhammad.usman@sg.panasonic.com