Accurate Bayesian Prediction of Cardiovascular-Related Mortality Using Ambulatory Blood Pressure Measurements

James O'Neill[1(✉)], Michael G. Madden[1], and Eamon Dolan[2]

[1] College Engineering and Informatics, National University of Ireland, Galway, Ireland
{james.oneill,michael.madden}@nuigalway.ie

[2] Stroke and Hypertension Unit, Connolly Hospital, Dublin, Ireland
eamon.dolan026@indigo.ie

Abstract. Hypertension is the leading cause of cardiovascular-related mortality (CVRM), affecting approximately 1 billion people worldwide. To enable patients at significant risk of CVRM to be treated appropriately, it is essential to correctly diagnose hypertensive patients at an early stage. Our work achieves highly accurate risk scores and classification using 24-h Ambulatory Blood Pressure Monitoring (ABPM) to improve predictions. It involves two stages: (1) time series feature extraction using sliding window clustering techniques and transformations on raw ABPM signals, and (2) incorporation of these features and patient attributes into a probabilistic classifier to predict whether patients will die from cardiovascular-related illness within a median period of 8 years. When applied to a cohort of 5644 hypertensive patients, with 20% held out for testing, a K2 Bayesian network classifier (BNC) achieves 89.67% test accuracy on the final evaluation. We evaluate various BNC approaches with and without ABPM features, concluding that best performance arises from combining APBM features and clinical features in a BNC that represents multiple interactions, learned with some human knowledge in the form of arc constraints.

Keywords: Bayesian network · Ambulatory Blood Pressure Monitoring · Hypertension

1 Introduction

Cardiovascular-related mortality (CVRM) is the top cause of death worldwide, accounting for a third of all deaths worldwide, 81% of which are attributed to coronary heart disease and stroke [1]. High blood pressure plays a significant role in CVRM, untreated hypertensive patients are at the risk of stroke, heart failure, and many other health conditions that increase over time [3]. The current out-of-clinic procedure for determining hypertension involves a patient's ABP

A. ten Teije et al. (Eds.): AIME 2017, LNAI 10259, pp. 86–91, 2017.
DOI: 10.1007/978-3-319-59758-4_10

being measured over a 24-h window, referred to as Ambulatory Blood Pressure Monitoring (ABPM). This work investigates how these ABPM measurements can be used; this is achieved by taking feature transformations and clusters to improve the prediction of CVRM. We also provide a comparison of Bayesian classifiers, with and without ABPM features, and also with and without the incorporation of human knowledge while learning the Bayesian network (BN).

2 Related Research

Dolan et al. [9] previously noted the superiority of ABPM over clinic measurements, highlighting that systolic BP in particular is the strongest predictor of CVRM among all possible ABP measurements in the form of hazard ratios, empirically highlighting the importance of nocturnal hypertension in diagnosis and treatment. However, to date there has been no attempt to incorporate features of the time series data produced during the well-established ABPM protocol; rather, only mean ABP measurements are used in diagnosis and prognosis. Bhatla and Jyoti [5] recently reviewed attempts to classify patients with cardiovascular disease by using k-nearest neighbours, Naive Bayes (NB), fuzzy logic systems, neural networks (NN), decision trees (DT). Soni et al. [4] proposed a weighted associative classifier for heart disease prediction, achieving an 81.51% test accuracy. In that case, the attributes were weighted according to the expertise of doctors to improve performance, similarly we look to incorporate human knowledge in our classification approach. A number of models re-occur in the literature that have performed well for classifying disease which are; fuzzy rule-base classifiers, boosted trees, bagging trees, logistic regression, NB and ANNs [5,6]. However, none of the above studies focus on improvements ABPM may make in both classification and risk scoring of hypertensive patients who are at possible risk of CVRM.

3 Experimental Methodology

Dataset Description and Feature Extraction. The dataset was previously published in Dolan et al.'s [9] study where they gathered data from 5644 hypertensive patients and their respective in-clinic and out-of-clinic ABP measurements over a median period of 8 years from Beaumont Hospital in Dublin, Ireland. We split this dataset into 80% training and 20% for a held out test set. Office ABP, ABPM features, BMI and age are discretized into 10 quartiles while sex, diabetes, cardiovascular history and smoking status are predefined discrete variables. ABPM measurements were recorded for all 5644 patients. Candidate features were extracted from the ABPM time series data and scaled using z-normalization, followed by taking a cubic transformation of the normalized ABPM z-scores. A sliding window was then used over these scaled features for each 3 h 20 window with a 1 h 40 min overlap starting from 8:00 am − 11:00 pm daytime and 11:00 pm − 6:00 am during sleep. This interval length and overlap window allows for a substantial number of cubic sums to be computed to summarize the volatility of blood pressure within a 3 h period. The resulting summed values for each window are then clustered as a final step. These clusters are also considered as features for classification. Furthermore, the maximum value from local peak detection[1] are also considered, in order to account for the envelope of each signal. Feature selection is then carried out on all of the above ABPM features by carrying out a statistical test that indicates the significance of each feature by a p value (small p value indicates strong evidence that the feature is statistically significant). The selected features are then tested using logistic regression as a baseline on the training dataset to determine which features should be kept and concatenated with patient attributes: diabetes mellitus, smoking status, sex, age, body mass index (BMI) office ABP and cardiovascular history for classification.

Bayesian Classification. We use a BNC as human knowledge should be incorporated, the posteriors must be reliably calibrated and the model must be interpretable for informative decision making [8,10]. A hill-climbing (HC) search (score-based algorithm) is employed for learning the BN structure by sequentially adding nodes and directed arcs to an initially empty graph. Bayesian Information Criterion (BIC) is used as a scoring criterion. The BIC score for a graph G and data D is $BIC(G, D) = \sum_{\forall v \in V} log\hat{p}_{v|pa(v)}(X_{v|pa(v)} - {}^1/_2 k_{v|pa(v)})$ where $k_(v|pa(v))$ is the dimension of the parameter $v|pa(v)$ [7]. HC is an effective structure search strategy but it is prone to local minima, hence we introduce restarts to run a number of iterations to avoid this. Three widely used BNs include a GBN, NB and TAN, all of which are compared in this work, with and without the ABPM features. A further comparison is also carried out to assess the performance of the classifiers when arc constraints are applied. For classifiers that include ABPM features, *Nighttime SBP Peak 1*, *Nighttime SBP Peak 2*, *Nighttime SBP Peak 3*, *k-means Nighttime Systolic Cluster* and *Age* nodes are parents of the *CVRM* class node (whitelist) and blacklisted arcs exclude nodes from *CVRM* to *Smoking Status*, *Sex* and *BMI* (blacklist). The whitelisted nodes are chosen based on previous studies that highlight the importance of systolic ABP [9]. For classification without ABPM features, the *Sex* and *Age* nodes are connected to *CVRM* class node and blacklisted arcs exclude *CVRM* being a parent of *Smoking Status* and *BMI*.

4 Results

Table 1 presents the accuracies on the training set using a Logistic Regression model. The features above $p > 0.01$ level are selected from all tested ABPM extracted features. This includes the sum of the cubic values for each sliding window, clusters of these features (k = 5, determined by analyzing the clusters sum of squared distances for varying k) and the maximum value for each sliding window (not the cubic sum). The maximum peaks for systolic-ABP[2] is also

[1] Local peaks considers neighboring measurements when choosing which are peaks.

[2] SBP values are first z-normalized, followed by a cubic transformation.

Table 1. 10-CV accuracy on the training set using Logistic Regression

Features	Accuracy	Features	Accuracy
Original mean ABPM	66.34 %	Sliding window clusters	70.09 %
Sliding window cubic sums	71.56 %	Sliding window clusters & Max peak	72.36 %
Sliding window max	63.78 %	Sliding window clusters & Max peaks clusters	70.82 %

computed. The maximum peaks and sliding window clusters have shown best results with 72.36% accuracy. Using cubic transformations on z-scaled features allows to account for the variability in ABP, while the maximum SBP peaks accounts for the envelope of highest BP throughout ABPM, resulting in a 7 percentage point increase in accuracy on the training set. Based on this analysis, we incorporate these set of features into the BN classifiers and test whether they are useful for CVRM prediction by comparing them with classifiers which do not account for ABPM.

4.1 Bayesian Classification

This section discusses the results of performing classification on a range of BNCs: GBN_{BIC}, GBN_{K2}, NB and TAN. Constraining arcs from nighttime SBP features to the CVRM class increases the test accuracy by 3.96 % points, shown in Table 2. These experiments have shown that the incorporation of systolic ABPM

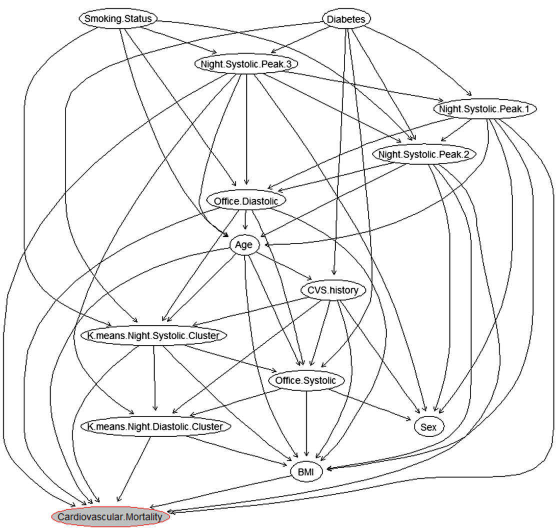

Fig. 1. GBN_{K2} with arcs constraints

Table 2. Bayesian classification results - experimental results with ABPM features

Classifiers	Including ABPM features				Excluding ABPM features			
	Arc constraints		Without arc constraints		Arc constraints		Without arc constraints	
	10-CV	Test	10-CV	Test	10-CV	Test	10-CV	Test
NB	-	-	75.17%	74.45%	-	-	75.38%	73.98%
GBN_{BIC}	97.28%	87.27%	75.16%	75.38%	75.16%	75.38%	75.16%	75.38%
GBN_{K2}	95.01%	**89.67%**	75.16%	75.38%	94.95%	**88.82%**	76.82%	77.29%
TAN	84.28%	84.26%	82.41%	**85.78%**	80.00%	81.74%	80.01%	**81.74%**

features directly influencing the class node plays an important role in finding a different BN structure that achieves better performance. By incorporating these constraints into the network the number of all dependencies has risen, particularly for the class node. The performance improvements can be attributed to the expansion of the Markov Blanket (MB) for the classification node show in Fig. 1, allowing more interaction terms among the ABPM features in classification, forcing the network to learn a network where the class node is dependent on ABPM features has made a significant improvement.

5 Conclusion

The prediction of CVRM using raw ABPM time series data and patient details has shown significant improvement compared to classification without ABPM. The aforementioned BNCs have shown good performance, especially when the proposed ABPM features are incorporated into the models. The learner has recovered a network which produces the best results overall with 89.67% accuracy on a held out test set, TAN also shows good performance with 85.71% accuracy without encoding any constraints on its arcs. This is the first line of work that empirically establishes the importance of using features from ABPM data instead of merely using mean ABP measurements for the purpose of classifying patients. More generally, it is the first piece of work to report an evaluation of ABPM transformed data for improving the accuracy of CVRM outcomes. It also echoes similar empirical findings that systolic blood pressure is a significant predictor of CVRM and ABPM is superior to clinical measurements, not only in the context of applying the CPH model [9] but also for various BNs.

References

1. Finegold, J.A., Asaria, P., Francis, D.P.: Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. Int. J. Cardiol. **168**(2), 934–945 (2013)
2. Kabir, Z., Perry, I.J., Critchley, J., O'Flaherty, M., Capewell, S., Bennett, K.: Modelling coronary heart disease mortality declines in the Republic of Ireland, 1985–2006. Int. J. Cardiol. **168**(3), 2462–2467 (2013)

3. Mozaffarian, D., et al.: Executive summary: heart disease and stroke statistics-2016 update: a report from the AmerIcan Heart Association. Circulation **133**(4), 447 (2016)
4. Soni, J., Ansari, U., Sharma, D., Soni, S.: Intelligent and effective heart disease prediction system using weighted associative classifiers. Int. J. Comput. Sci. Eng. **3**(6), 2385–2392 (2011)
5. Bhatla, N., Jyoti, K.: An analysis of heart disease prediction using different data mining techniques. Int. J. Eng. **1**(8), 1–4 (2012)
6. Austin, P.C., Tu, J.V., Ho, J.E., Levy, D., Lee, D.S.: Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. J. Clin. Epidemiol. **66**(4), 398–407 (2013)
7. Nagarajan, R., Scutari, M., Lèbre, S.: Bayesian Networks in R, vol. 122. Springer, Heidelberg (2013). pp. 125–127
8. Madden, M.G.: On the classification performance of TAN and general Bayesian networks. Knowl.-Based Syst. **22**(7), 489–495 (2009)
9. Dolan, E., Stanton, A., Thijs, L., Hinedi, K., Atkins, N., McClory, S., Den Hond, E., McCormack, P., Staessen, J.A., O'Brien, E.: Superiority of ambulatory over clinic blood pressure measurement in predicting mortality. Hypertension **46**(1), 156–161 (2005)
10. Freitas, A.A.: Comprehensible classification models: a position paper. ACM SIGKDD Explor. Newslett. **15**(1), 1–10 (2014)