Bayesian Network Learning for Cardiovascular Data

Kate M. Sanders[1], Robert M. Long[2], Nandini Ramanan[3], Greg Terry[4], Jeffrey Carr[5] and Sriraam Natarajan[6]

***Abstract*— Coronary Heart Disease (CHD) is the leading cause of death in the United States. We consider the problem of learning probabilistic models from a cardiovascular study. Specifically, we consider the Coronary Artery Risk Development in Young Adults (CARDIA) study and aim to learn a joint distribution over several risk factors and Coronary Artery Calcification (CAC), an indicator of the risk of CHD present subclinically in an individual. In the artificial intelligence community, Bayesian networks (BN) are considered as popular tools for effectively modeling the joint distribution across several related variables. We employ learning of these BNs on CARDIA data and demonstrate that the resulting BN captures some interesting associations (possibly causations) between risk factors that could lead to developing individualized treatment plans for improving population health.**

I. INTRODUCTION

According to the American Heart Association, an estimated 16.5 million Americans over the age of 20 suffer from Coronary Heart Disease (CHD). Approximately 1 in 7 deaths that occur in the United States are a result of CHD. The most common and deadly cardiac event associated with CHD is Myocardial Infarction (MI), which is more commonly known as a heart attack. Each year, approximately 790,000 MIs occur, which means, on average, an American has an MI every 40 seconds [1]. We consider modeling Coronary Artery Calcification (CAC), a strong indicator that an individual has CHD. Detrano et al. found that doubling CAC levels caused around a 25% increase in the probability of a major cardiac event occurring, a correlation which held true across all races [2].

Many resources have gone into researching the causes of CHD. Traditional studies take years to conduct and only analyze the effects of a very small number of risk factors (typically under 5) on CHD development. With the recent advances in Data Mining algorithms, it is now possible to model the connections between multiple complex and temporal risk factors of CHD. We consider the the data collected from the Coronary Artery Risk Developments in Young Adults (CARDIA)[1]. While the study itself is longitudinal, we consider the data at year 20 corresponding to the calendar year 2005 to learn a model that considers the interaction between multiple risk factors on CAC and correspondingly in cardiovascular health of a patient.

Traditionally, most adaptations of machine learning methods to cardiovascular health focus on *predictive models* [3]–[5]. The key idea behind these methods is that they learn a discriminative model (in machine learning terminology) that aims to discriminate between different class (CHD patients vs normal subjects). While effective and successful, the goal of these methods is limited in that they aim to develop robust models with high predictive performance. We, on the other hand, aim to learn a model that *explains* the underlying interactions between the different risk factors and the CAC levels.

Our learned model is, in general, referred to in machine learning as a *generative model* as it addresses the question of how the data is generated (evolved). This allows one to dive deep into the underlying interplay between the different factors and the target disease. One of the most common generative models is a Bayesian network [6] (BN), that factors a joint distribution among random variables into a product of conditional distributions that can possibly be learned locally. One of the common tasks in BNs is the full model learning aka *structure learning* where the goal is to learn the qualitative influences between random variables along with the quantitative conditional distributions that describe the effect of these relations. One of the major advantages of these models is their interpretability and explainability which allows us to understand the complex interactions between random variables (risk factors and CAC in our setting) more clearly when compared to standard techniques. While not all BNs are necessarily causal models, with some domain knowledge it is possible to tease out the causal relationships inside a BN, an approach that we take in this work.

We make the following key contributions: (1) We consider several risk factors from the highly informative CARDIA study and adapt standard BN learning techniques to learn probabilistic models that explain the data. (2) Our approach can be viewed as a probabilistic knowledge discovery method for CVD data. (3) We consider several standard scoring metrics and identify the most useful metric to tease out relevant knowledge from observational data. (4) The resulting

This work is supported by the National Science Foundation under Grant No. CNS-1560276.

[1]Kate M. Sanders is a Computer Science major at Hendrix College, Conway, AR 42032, USA sanderskm@hendrix.edu

[2]Robert M. Long is a Computer Science major at DePauw University, Greencastle, IN 46135, USA robmlong95@gmail.com

[3]Nandini Ramanan is a graduate student at the School of Informatics and Computing Indiana University, Bloomington, IN 47408, USA nramanan@iu.edu

[4]Greg Terry is a Research Programs Manager, Department of Radiology Radiological Sciences at Vanderbilt University Medical Center, USA james.g.terry@vanderbilt.edu

[5] Jeffrey Carr is the Cornelius Vanderbilt Professor of Radiology and Radiological Sciences and Professor of Biomedical Informatics and Cardiovascular Medicine at Vanderbilt University Medical Center, USA j.jeffrey.carr@vanderbilt.edu

[6]Sriraam Natarajan is an Associate Professor at the School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA natarasr@indiana.edu

[1]http://www.cardia.dopm.uab.edu/

BN clearly captures some known knowledge about the risk factors and CVD and discovers some interesting relationship between variables such as race and education. We present and analyze this knowledge in greater detail.

The rest of the paper is organized as follows: after reviewing relevant work in knowledge discovery and probabilistic models, we present the sub-set of CARDIA data that we consider for learning BNs. We then present the results of the different scoring metrics and analyze the network learned from the best performing one. Finally, we conclude the paper by pointing out some interesting avenues for future research.

II. BACKGROUND

A. Knowledge Discovery in Medicine

One of the key features of Data Mining (DM) is the ability to adapt algorithms to large databases to find interesting patterns. These relationships can then be analyzed and modeled for further understanding. This sort of data exploration is referred to as Knowledge Discovery (KD). The focus of KD is identifying probable hypotheses rather than testing known hypotheses [7].

While several algorithms exist, adapting them for medical domains to perform KD remains an interesting challenge. This is largely because most clinical data sets are large, expensive, temporal, non-reproducible, and incomplete [8]. Also, domain expert knowledge is required for proper interpretation of the exploratory model [9].

Despite these challenges, the rewards of successful KD on clinical data are significant. This unique domain has the potential to extend lives and benefit society as a whole [10]. Because of this, researchers have applied DM techniques to many complex medical databases. Regression, Neural Networks, Decision Trees, and Bayesian Networks are just a few of the DM methods utilized. Applications of Bayesian Networks, in particular, include: studying brain structure-function associations [11], assisting in the diagnosis of pulmonary embolisms [12], and describing relationships between genes [13].

B. Bayesian Networks

BNs [6] have long been one of the most promising approaches to knowledge discovery [7]. A BN factorizes the joint probability distribution over a set of features (called *random variables* (RVs) to denote that they are probabilistic) as a product of conditional distribution over these RVs. A key aspect of these networks is their graphical structure which consists of nodes and edges. The RVs form the nodes of a BN and the existence of a directed edge between two random variables specify the direction of the influence. An edge from a RV A to a RV B denotes that A influences B and A is called as the *parent* of B. While the presence and direction of an edge specify the qualitative influences, the quantitative influences are captured by the conditional probability distributions. Note that a possible interpretation of BN is that of a causal model but this interpretation is not always true. This is due to the fact that a single joint distribution can be factored in multiple ways and one of them can be the causal model. Hence, a causal model can be in many cases represented using a BN but the converse is not true. Not all BNs are causal models.

Conditional independence assumptions are necessary to make learning from large databases tractable. BNs incorporate these independence assumptions by the structure of the directed acyclic graph (DAG). This means that directed edges cannot make loops within the network. Because of this, complex networks can be broken up into local models with fewer dimensions. The resulting structure is easy to summarize and interpret, a primary reason why BNs are frequently used in many domains [7] [15].

C. Structure Learning Algorithms

While effective, full model learning (also called *structure learning*) is a computationally intensive task. Typically, structure learning, a task we address in this paper, requires repeated parameter learning that corresponds to learning the quantitative probabilities given a BN DAG skeleton. And if the data is not fully observed, this requires performing probabilistic inference based on the current model, an NP-complete problem. Approximate methods exist, such as search-and-score techniques, constraint-based methods, or a combination of the two [16]. We will focus on search-and-score structure learning, where, many structures are created from the data and scored. The network with the best score is returned as the optimal model for the data. Typically, after a structure is scored, a local change (addition, deletion or reversal of an edge) is performed and the resulting structure is scored to determine if the change is to be preserved.

In this study, we use a Greedy Local Search algorithm called Hill-Climbing (HC). This algorithm starts with an edgeless network. In each iteration, an edge is created, reversed, or deleted. The resulting network is then scored. If the score improves, the edge operation is kept. This continues until the score no longer improves with new edge operations. At this point, the network is considered to be at the top of a hill.

There are three primary drawbacks of HC algorithms. One is the tendency to get stuck at a local maximum rather than continuing on to the global maximum, the truly optimal network. Another drawback is getting lost in a plateau, where no direction causes a significantly better score. Ridges can also hinder progress; the searcher zig-zags over the ridge while only making slight progress towards the top of the hill. These problems can be minimized by using random restarts throughout the search process to better explore the space [14].

Structure scoring metrics for HC include Log Likelihood (LL), Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Likelihood-Equivalence Bayesian Dirichlet (BDe). Each of these is described in detail below.

1) Information Theoretic Scores: LL is the simplest scoring metric; the top score goes to the BN structure, B that

best fits the data, D. LL is given by:

$$LL(B|D) = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log(\frac{N_{ijk}}{N_{ij}}) \quad (1)$$

where n is number of features, q_i is the total possible configurations of the parents of a particular feature, and r_i is the number of states for a particular feature. N_{ijk} represents the total number of times that a feature takes it's k^{th} value and the the feature's parents are in their j^{th} configuration within the given data. N_{ij} is the value after k is summed out of N_{ijk}.

Using LL alone can lead to an excessive number of parameters which in turn can lead to the classical overfitting issue in standard machine learning. AIC and BIC improve on LL by adding a penalty term for network complexity [16].

AIC was developed by Hirotugu Akaike [17] in the early 1970s. The score can be calculated using the following equation:

$$AIC(B|D) = LL(B|D) - |B| \quad (2)$$

where the penalizing term is $|B|$. This represents the network complexity and can be calculated as $\sum_{i=1}^{n}(r_i - 1)q_i$.

BIC was created in 1978 by Gideon Schwarz [18] as an alteration on AIC. It can be calculated by using:

$$BIC(B|D) = LL(B|D) - \frac{1}{2}\log(N)|B| \quad (3)$$

The sample size of *B* is represented by *N*. The penalizing factor for BIC is greater than that of AIC.

2) Bayesian Scores: A Bayesian approach can also be taken when scoring networks. Bayesian scores are based on Bayes Rule:

$$p(B|D) = \frac{p(D|B)p(B)}{p(D)} \quad (4)$$

where the posterior, $p(B|D)$ is maximized to find the best network structure given the data. $p(D)$ is the marginal probability of the data, $p(D|B)$ is the likelihood of the given data for a particular structure, and $p(B)$ is a prior that changes given the Bayesian scoring metric used.

BDe is a Bayesian scoring metric developed by Heckerman, Geiger, and Chickering in 1995 [19]. After much simplification, BDe can be calculated as:

$$BDe(B|D) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\grave{N}_{ij})}{\Gamma(\grave{N}_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\grave{N}_{ijk} + N_{ijk})}{\Gamma(\grave{N}_{ijk})} \quad (5)$$

where Γ is the Gamma function. Score equivalence is achieved using $\grave{N}$, the equivalent same size structure prior. The parameter prior, $\grave{N}_{ijk}$ when a feature is in it's k^{th} value and the the feature's parents are in their j^{th} configuration within the given data. $\grave{N}_{ij}$ is the value when k is summed out of $\grave{N}_{ijk}$.

D. Prior Probabilistic models of CARDIA Data

Dynamic Bayesian Networks (DBNs) have been used to model the temporal links between behavioral and socioeconomic data and CAC. Using only non-clinical CARDIA data revealed how life-style decisions young adults make influence CAC levels later in life [24]. Our paper is closely related to this work with one major change, instead of temporal modeling the development of CAC based on behavioral data, we consider the problem of modeling CAC levels at year 20 (at which time participants are aged 38 to 50) using some clinical and some non-clinical data.

A different approach to boosting the conditional distribution over CAC by viewing time as a relation was performed by Natarajan et al. [3].The resulting model outperformed standard machine learning algorithms for modeling the CAC level. While successful, the model was itself uninterpretable as it employed a successive approximation of regression trees that cannot be interpreted individually.

TABLE I
FEATURES ANALYZED ALONG WITH THE DATA DIVISIONS.

Feature	Divisions
Sex	Male, Female
Race	Black, White
CAC	Observed, Unobserved
HBP	No, Yes, Not Sure
Smoker	Never, Previously, Currently
Age	Ordinal
Education	Ordinal
Heavy Exercise	Ordinal
Moderate Exercise	Ordinal
BMI	Ordinal
Triglycerides	Ordinal
Cholesterol	Ordinal
LDL	Ordinal
HDL	Ordinal
Glucose	Ordinal
DBP	Ordinal
SBP	Ordinal

III. PROPOSED APPROACH

We now outline our proposed approach for discovering knowledge from the cardiovascular study CARDIA. Before we outline the approach, we present the background about the data.

A. Data

The Coronary Artery Risk Development in Young Adults (CARDIA) study gathered the data used in our analysis. This study followed 5115 subjects from $1985 - 6$ until present. Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA served as centers for data collection. Each location recruited participants in a way that ensured an even distribution of sex, race, education level, and age-group (18-25 or 25-30). Data gathered from participants included physical measurements, clinical tests, and an in-depth questionnaire about lifestyle and socioeconomic status. The specifics of the study procedures are detailed elsewhere [20]. This data set is often used for studying the development of heart disease due to the breadth of recorded features and the study's longitudinal nature and high retention rate.

The data we focus on in this analysis comes from the year 20 check-in, which occurred in 2005-6. At this time,

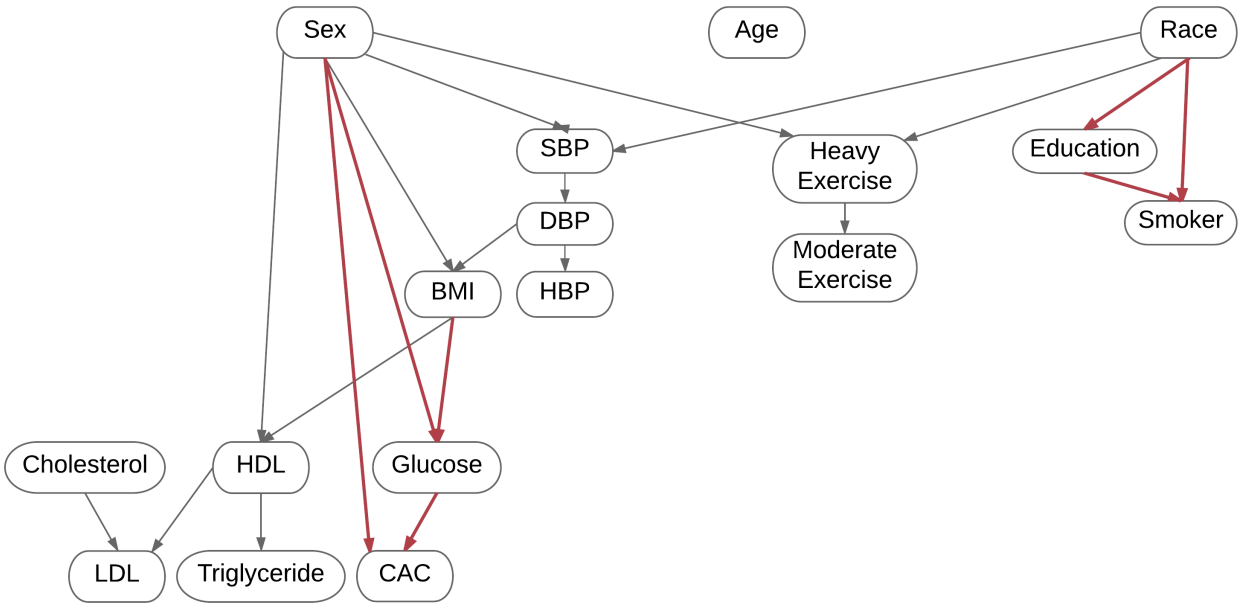

Fig. 1. Intersection of models learned using the AIC, BIC, and BDe scoring metrics in the Hill-Climbing algorithm.

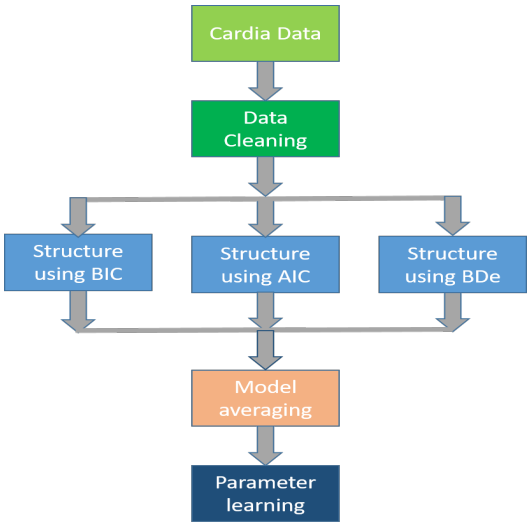

Fig. 2. Flowchart of the proposed model.

the retention rate was 72%. In this check-in, approximately 11% of participants had observed CAC.

The features modeled in our network can be viewed in Table I. We discretized continuous data into Quintiles, five even groups based on frequency.

A participant was excluded from our study if their CAC was not assessed in year 20 or if at least one of their features were not recorded in any year of the study. Otherwise, missing data imputed using the mean of that patient's data in every observed year.

Hence our task can be defined as:

`Given`: The feature set **F** of every patient presented in Table I for year 20

`To Do`: Learn a Bayesian network (BN) that models the qualitative and quantitative influence structures between the different variables in **F**

B. Model Creation

For BN structure learning from the data, we used the bnlearn package in R [22]. Sex, race, and age were defined as having no parents; other than this, no ordering occurred.

We learned BN structures using constraint based algorithms such as Incremental Association, Chow Liu, and Semi-Interleaved Hiton-PC. However, due to insufficient data, these networks were too sparse for effective knowledge discovery. It remains a future research direction to extend these algorithms to learning from sparse data.

The Hill-Climbing algorithm for structure learning yielded more meaningful results. We used the optimized implementation of this algorithm to decrease the number of repeated tests within the learning process [23].

The BIC, AIC, and BDe scoring metrics were used with the Hill-Climbing algorithm to learn three separate models of the year 20 CARDIA data. For knowledge discovery, we focused on the intersection BN, which can be viewed in Figure 1. The idea is that we perform *model aggregation*, the aggregation being the intersection of the arcs learned in all the networks. The intuition is that this will yield insights into the most common influence structures according to the different scoring metrics and would provide a comprehensive overview of the data. Our hypotheses which we verify qualitatively is that such a resulting network would provide interesting insights into the relationships that are hidden in the study data.

A high-level overview of our framework is presented in Figure 2. As can be seen, we use the year 20 data for learning and employ multiple scoring metrics since each of

BMI						Sex
[0, 22. 8]	(22. 8, 25.6]	(25.6, 28.6]	(28.6, 33.2]	(33.2, 154]	Glu	
0.337	0.172	0.12	0.049	0.052	[0, 82]	
0.294	0.249	0.175	0.147	0.113	(82, 88]	
0.174	0.207	0.218	0.184	0.083	(88, 93]	Male
0.14	0.228	0.244	0.27	0.236	(93, 100]	
0.053	0.141	0.241	0.349	0.513	(100, 448]	
0.406	0.327	0.236	0.176	0.131	[0, 82]	
0.267	0.281	0.261	0.242	0.156	(82, 88]	
0.161	0.191	0.198	0.214	0.188	(88, 93]	Female
0.111	0.131	0.184	0.218	0.229	(93, 100]	
0.052	0.068	0.119	0.15	0.294	(100, 448]	

Fig. 3. Conditional probability table depicting the relationship between an individual's BMI, sex, and glucose level.

them could potentially yield a different sparse network. Then we combine the different networks using model intersection as our aggregator. The resulting network is then used for learning the conditional probabilities from the training data. We can then analyze the model and the parameters to deduce meaningful conclusions.

IV. DISCUSSION

Our aim in this work is to explore the use of BN learning algorithms in discovering knowledge about CAC and related risk factors. Our aim is not to evaluate machine learning algorithms quantitatively but to qualitatively analyze the effects of learning BNs on this data. Hence, we do not consider any baseline methods or evaluate using typical machine learning metrics such as accuracy, ROC or precision-recall. Our goal is to perform discovery using BNs and determine if some of these conclusions reflect known knowledge or lead to some interesting insights.

To this effect, we consider the model aggregation network from multiple scoring metrics. The intuition is that if a dependency is captured in every one of these metrics, then there is a high possibility that this is potentially a true dependency obtained from the data.

When looking at our aggregated network (Figure 1), we first identified three clusters of interdependent correlations that capture knowledge that is commonly well-known. The most obvious connection is between the ability to perform moderate and heavy exercise. Unsurprisingly, those capable of performing heavy exercise can also perform moderate exercise. We also saw a monotonic relationship between SBP and DBP, as well as DBP and HBP (the learned parameters reflect this monotonic relationship). This is also expected because an individual with a high systolic blood pressure (SBP) often also has a high diastolic blood pressure (DBP). High blood pressure (HBP) is the diagnosis of individuals with a high SBP and DBP. The cluster of triglycerides, cholesterol, HDL, and LDL was expected as well because all four of these measurements have to do with lipids in the blood stream.

Since Figure 1 captured these known connections, the proposed methodology can be validated. However, on a deeper examination, one could potentially tease out more interesting correlations. We first examined the influence of glucose

Race		
African American	Caucasian	Edu
0.389	0.184	[0, 12]
0.292	0.184	(12, 14]
0.091	0.047	(14, 15.3]
0.142	0.274	(15.3, 16]
0.084	0.31	(16, 40.3]

Fig. 4. Conditional probability table depicting the connection was between race and education level .

Education						
[0 , 12]	(12, 14]	(14, 15.3]	(15.3, 16]	(16, 40.3]	Smoker	Race
0.412	0.558	0.555	0.758	0.748	Yes	African American
0.13	0.129	0.177	0.108	0.132	Quit	
0.456	0.312	0.266	0.132	0.118	Never	
0.31	0.435	0.46	0.643	0.706	Yes	Caucasian
0.223	0.244	0.252	0.241	0.226	Quit	
0.465	0.32	0.286	0.114	0.067	Never	

Fig. 5. Conditional probability table shows that a persons race and education also has an influence on whether that person smokes.

levels and sex on the probability of an individual having CAC. For both sexes, there was a monotonic relationship between glucose levels and the probability of CAC. At each glucose level, it was observed that men were considerably more likely to have CAC than women. It should be noted that, overall, the likelihood of an individual having CAC in year 20 is very low; only approximately 11% of participants had observable CAC at this time.

The risk factors with the most influence on a person's glucose levels were sex and BMI. There was a monotonic relationship between an individual's BMI and glucose levels as shown in Figure 3. As with CAC, men were generally more likely to have high glucose levels. This model also captured some interesting non-clinical interactions as shown in figure 4. The most prominent connection was between race and education level. In year 20, a majority of participants had pursued post-secondary education (more than 12 years of school). However, African American participants were more likely than Caucasian participants to have a total of 12 years of education or fewer. Caucasian participants were more likely to have pursued a graduate degree (over 16 years of education). This striking difference is evidence of the ongoing effects of racism and segregation in America [25]. Another interesting observation is that a person's race and education also has an influence on whether that person smokes. We can see in Figure 5 that, for both races, the likelihood of an individual smoking decreases as their education level rises. Those with 12 or fewer years of education are much more likely to smoke than those who have pursued a graduate degree. African Americans were more likely to smoke than Caucasians at all education levels. This is likely due to the promotional targeting of minority communities by Big Tobacco [26]. Cigarette advertising in minority-specific media pushes menthol cigarettes, which are much more addicting than the

non-menthol variety [27].

V. CONCLUSION

We considered the problem of learning a probabilistic interpretable model such as Bayes net for modeling interactions between risk factors in a cardiovascular study. To this effect, we considered a model aggregator that outputs the intersection of multiple sets of connections from different Bayesian network scoring metrics. Then we learned the parameters of this network to model the quantitative dependencies.

Our findings reveal interesting correlations between both clinical and non-clinical data. These connections warrant further exploration. Particularly, it is imperative to test the generalizability of these observations on other related clinical studies.

We next plan to incorporate more risk factors in our model, focusing on non-clinical data. We will then create a model using data from all available years of the study. Also, we plan to track the progression of risk factors over the course of many years. Eventually, our grand vision is to contribute to a system that helps doctors identify the risk of a patient developing CHD based on a variety of temporal factors. Using this system, young adults at risk of CHD could be identified and made aware of precautionary steps for avoiding this deadly disease.

REFERENCES

[1] Benjamin, Emelia J., et al. "Heart disease and stroke statistics - 2017 update: a report from the American Heart Association." Circulation 135.10 (2017): e146-e603.

[2] Detrano, Robert, et al. "Coronary calcium as a predictor of coronary events in four racial or ethnic groups." New England Journal of Medicine 358.13 (2008): 1336-1345.

[3] Natarajan, Sriraam, et al. "Early Prediction of Coronary Artery Calcification Levels Using Statistical Relational Learning." Workshop on Machine Learning for Clinical Data Analysis. Edinburgh, Scotland. Vol. 30. 2012.

[4] Weiss, Jeremy C., et al. "Statistical Relational Learning to Predict Primary Myocardial Infarction from Electronic Health Records." IAAI. 2012.

[5] Weiss, Jeremy, Sriraam Natarajan, and David Page. "Multiplicative forests for continuous-time processes." Advances in neural information processing systems. 2012.

[6] Pearl, Judea. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, 1998.

[7] Brockmann, Dirk, Lars Hufnagel, and Theo Geisel. "The scaling laws of human travel." arXiv preprint cond-mat/0605511 (2006).

[8] Bellazzi, Riccardo, and Blaz Zupan. "Predictive data mining in clinical medicine: current issues and guidelines." International journal of medical informatics 77.2 (2008): 81-97.

[9] Roddick, J.F., Fule, P. and Graco, W.J., 2003. Exploratory medical knowledge discovery: Experiences and issues. ACM SIGKDD Explorations Newsletter, 5(1), pp.94-99.

[10] Cios, Krzysztof J., and G. William Moore. "Uniqueness of medical data mining." Artificial intelligence in medicine 26.1 (2002): 1-24.

[11] Herskovits, Edward H., and Joan P. Gerring. "Application of a data-mining method based on Bayesian networks to lesion-deficit analysis." Neuroimage 19.4 (2003): 1664-1673.

[12] Luciani, D., M. Marchesi, and G. Bertolini. "The role of Bayesian networks in the diagnosis of pulmonary embolism." Journal of Thrombosis and Haemostasis 1.4 (2003): 698-707.

[13] Friedman, Nir, et al. "Using Bayesian networks to analyze expression data." Journal of computational biology 7.3-4 (2000): 601-620.

[14] Russell, Stuart, Peter Norvig, and Artificial Intelligence. "A modern approach." Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs 25 (1995): 27.

[15] Giudici, Paolo. Applied data mining: statistical methods for business and industry. John Wiley & Sons, 2005.

[16] Koski, Timo JT, and John Noble. "A review of bayesian networks and structure learning." Mathematica Applicanda 40.1 (2012): 51-103.

[17] Akaike, H. "A new look at the statistical model identification." IEEE Transactions on Automatic Control. 1974.

[18] Schwarz, G. "Estimating the Dimension of a Model." The Annals of Statistics. 1978.

[19] Heckerman, D. and Geiger, D. and Chickering, D.M. "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data." Machine Learning. 1995.

[20] Friedman, G.D. and Cutter, G.R. and Donahue, R.P. and Hughes, G.H. and Hulley, S.B. and Jacobs, D.R. and Liu, K. and Savage, P.J. "Cardia: Study design, recruitment, and some characteristics of the examined subjects." Journal of Clinical Epidemiology. 1988.

[21] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3 (1996): 37.

[22] Scutari, Marco. "Learning Bayesian networks with the bnlearn R package." arXiv preprint arXiv:0908.3817 (2009).

[23] Daly, Rónán and Shen, Q. "Methods to Accelerate the Learning of Bayesian Network Structures." Proceedings of the 2007 UK workshop on computational intelligence. 2007.

[24] Yang, Shuo, et al. "Modeling Coronary Artery Calcification Levels From Behavioral Data in a Clinical Study." Conference on Artificial Intelligence in Medicine in Europe. Springer, Cham, 2015.

[25] Blanchett, Wanda J., Vincent Mumford, and Floyd Beachum. "Urban school failure and disproportionality in a post-Brown era: Benign neglect of the constitutional rights of students of color." Remedial and Special Education 26.2 (2005): 70-81.

[26] Robinson, Robert G., et al. "Report of the Tobacco Policy Research Group on marketing and promotions targeted at African Americans, Latinos, and women." Tobacco Control 1 (1992): S24-S30.

[27] Cummings, K. Michael, Gary Giovino, and Anthony J. Mendicino. "Cigarette advertising and black-white differences in brand preference." Public health reports 102.6 (1987): 698.