

# Diagnosis of Cardiovascular Diseases with Bayesian Classifiers

Alaa Elsayad and Mahmoud Fakhri

Department of Computers and Systems, Electronics Research Institute, 12622 Bohoth St., Dokki, Geza, Egypt

## Article history

Received: 04-12-2013

Revised: 11-05-2014

Accepted: 30-01-2015

## Corresponding Author:

Mahmoud Fakhri  
Department of Computers and  
Systems, Electronics Research  
Institute 12622 Bohoth St.,  
Dokki, Geza, Egypt  
Email: mafakhri@mcit.gov.eg

**Abstract:** Cardiovascular disease or atherosclerosis is any disease affecting the cardiovascular system. They include coronary heart disease, raised blood pressure, cerebrovascular disease, peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. They are treated by cardiologists, thoracic surgeons, vascular surgeons, neurologists and interventional radiologists. The diagnosis is an important yet complicated task that needs to be done accurately and efficiently. The automation of this system is very much needed to help the physicians to do better diagnosis and treatment. Computer aided diagnosis systems are widely discussed as classification problems. The objective is to reduce the number of false decisions and increase the true ones. In this study, we evaluate the performance of Bayesian classifier (BN) in predicting the risk of cardiovascular disease. Bayesian networks are selected as they are able to produce probability estimates rather than predictions. These estimates allow predictions to be ranked and their expected costs to be minimized. The major advantage of BN is the ability to represent and hence understand knowledge. The cardiovascular dataset is provided by University of California, Irvine (UCI) machine learning repository. It consists of 303 instances of heart disease data each having 76 variables including the predicted class one. This study evaluates two Bayesian network classifiers; Tree Augmented Naïve Bayes and the Markov Blanket Estimation and their prediction accuracies are benchmarked against the Support Vector Machine. The experimental results show that Bayesian networks with Markov blanket estimation has a superior performance on the diagnosis of cardiovascular diseases with classification accuracy of MBE model is 97.92% of test samples, while TAN and SVM models have 88.54 and 70.83% respectively.

**Keywords:** Cardiovascular Disease, Bayesian Classifier, Naïve Bayes, Markov Blanket Estimation and Support Vector Machine

## Introduction

Recent trends in the field of data mining have driven to the emergence of expert systems for medical applications. Many computational tools and algorithms have been recently developed to increase the experiences and the abilities of physicians for taking decisions about different diseases. Normally physician acquires knowledge and experience after analyzing sufficient number of cases. This experience is reached only in the middle of a physician's career. However, for the case of rare or new diseases, experienced physicians are also in the same situation as new comers. In fact, humans do not work like statistic computers but as pattern recognition systems. Humans can distinguish patterns or objects very

easily but fail when probabilities have to be allocated for observations (Salim, 2004). Recent paper demonstrated the power of neural networks and evaluated their limits, possible trends, future developments and connections to other branches of human medicine in (Filippo *et al.*, 2013). They concluded that Artificial Neural Networks (ANNs) represent a powerful tool to help physicians perform diagnosis and other enforcements. A study was conducted to demonstrate that machine learning algorithms can help in making correct diagnosis in (Brause, 2001). Their results show that even the most experienced physician can diagnose properly (79.97%) when compared to the diagnosis made with the help of machine learning and expert system (91.1%). A more recent comparisons between trained neural network and

experienced clinicians made in (Sabina *et al.*, 2012) where the diagnostic accuracy of the glaucoma visual fields interpreted by the trained ANN was at least as good as those performed by clinicians who had full access to sophisticated interpretation tools. They noted that ANN offered better sensitivities. However, to build an efficient computer aided diagnosis, the matter requires a set of samples containing all the variations of the disease. A good training samples that efficiently represent different cases of the disease is a must to build an efficient and reliable system. Consequently; designer has to face extra problems including data acquisition, collection and organization.

Computer aided diagnosis systems are widely considered as classification problems and solved by artificial intelligence algorithms where the choice of the classification algorithm is not an easy problem. There is often a trade-off between model accuracy and model transparency. There is no explanation for a decision made by a system like neural network or support vector machine. On the other hand, probability estimation classifiers estimate the prediction with its conditional probability and a direct connection between the predicting attributes and the target class. Such systems are concise and easy to comprehend (Sherly, 2012). Probability estimate classifiers include Naïve Bayes, logistic regression, decision tree and Bayesian network. Naïve Bayes and logistic regression models can only represent simple distributions, whereas decision tree models can represent arbitrary distributions, but they fragment the training dataset into smaller and smaller pieces, which unavoidably yield less reliable probability estimates.

Bayesian networks are the best-known classifier that able to provide the probability distributions concisely and comprehensibly (Witten and Frank, 2005; Darwiche, 2008). They give compact representation of joint probability distributions via conditional independence and handle uncertainty in mathematically rigorous yet efficient and simple method. While they enables efficient uncertainty reasoning with hundreds of variables, they also enables human experts to better understand the modelled domain. Felipe *et al.* (2013) considered BN model with Naive Bayes algorithm is one of the most effective classification algorithms today, that competing with more modern and sophisticated classifiers. BN is a probabilistic model that consists of dependency structure and local probability. BN is drawn as a network of nodes, one for each attribute, connected by directed edges in such a way that there are no cycles; a directed acyclic graph. The major advantage of BN is the ability to represent and hence understand knowledge. Recently, there is increasing attention regarding the application of BN in medical contexts (Linda *et al.*, 2008). BN classifiers have been evaluated as potential tools for the diagnosis of breast cancer using two real-world databases in (Cruz-Ramírez *et al.*, 2007; 2009).

This study evaluates the performance of two different implementations of BN for the diagnosis of cardiovascular diseases; Tree Augmented Naïve Bayes (TAN) and Markov Blanket Estimation (MBE) learning algorithms. Both algorithms use Naïve Bayes classifier as a starting point for the learning procedure. The class attribute is the single parent of each node of a Naïve Bayes network: TAN considers adding a second parent to each node. While MBE ensures that every attribute in the data is in the Markov blanket of the node that represents the class attribute.

The heart is a muscular organ located nearly the middle of chest. It pumps blood to all parts of the human body (Rosendorff, 2013). It consists of two separate pumps: The right one pumps blood through the lungs and the left one pumps blood through the whole body. In turn, each of these parts is a pulsatile two-chamber pump composed of an atrium and a ventricle. Each atrium is a weak primer pump for the ventricle, helping to move blood into the ventricle. While the ventricles provide the main pumping force that propels the blood either through the pulmonary circulation by the right ventricle or through the whole body circulation by the left ventricle. There are four valves to control circulations: The aortic valve, the pulmonary valve, the mitral valve and the tricuspid valve that control the forward and backward flows of the blood through the heart (Steven *et al.*, 2006). The trouble to this circulation of blood can result in serious health problems possibly may come to the death. There are different types of heart diseases, among which the major types are: Atherosclerosis, coronary, rheumatic, congenital, myocarditis, angina and arrhythmia (Khan, 2005). Furthermore symptoms of this disease is not the same from person to person and in most cases, there is no early symptom and the disease is diagnosable only in the advanced stage. Some common symptoms of heart disease are (Khan, 2010): Chest pain (Angina pectoris); strong compressing or flaming sensation in the chest, neck or shoulders; discomforts in chest area; sweating, light-headedness, dizziness, shortness of breath; pain spanning from the chest to arm and neck and that amplifying with exertion; cough; palpitations; fluid retention.

Data mining and knowledge discovery techniques have been used in the diagnosis and analysis of heart diseases. Support vector machine, neural network, decision tree and other classification algorithms have been used in the diagnosis and prediction systems of heart diseases (Ghumbre *et al.*, 2011; Can, 2013; Rani, 2011). A new modified K-means technique is presented in (Nihat *et al.*, 2014) for clustering based data preparation method for the elimination of noisy and inconsistent data and Support Vector Machines is used for classification. Yadav *et al.* (2014) an improved association rule mining of data mining for the detection of Coronary Artery Disease is demonstrated and used.

Nahar *et al.* (2013a; 2013b). She used association rule mining to identify these factors and the UCI Cleveland dataset, a biological database. The organization of the paper is as follows. Section 2 gives preliminaries that provide

some information on the used dataset. Section 3 illustrates principles of BNs techniques. Section 4 presents the experimental results and finally the paper is concluded in section 5 with a short summary and remarks.

Table 1. Attributes of Cleveland heart disease dataset

Attribute	Description	Domain of values
Age	Age in years	29-79
Sex	Sex	0→Female, 1→Male
Cp	Chest pain type	1→Typical angina, 2→Atypical angina 3→Non-angina, 4→Asymptomatic
Trestbps	Resting blood sugar	94 to 200 mm Hg
Chol	Serum cholesterol	126 to 564 mg dL <sup>-1</sup>
Fbs	Fasting blood	sugar >120 mg dL <sup>-1</sup> ; 0→False, 1→True
Restecg	Resting ECG result	0→Normal, 1→ST-T wave abnormality 2→LV hypertrophy
Thalach	Maximum heart rate achieved	71 to 202
Exang	Exercise induced angina	0→No, 1→Yes
Oldpeak	ST depression induced by exercise relative to rest	0 to 6.2
Slope	Slope of peak exercise ST segment	1→Upsloping, 2→Flat, 3→Downsloping
Ca	Number of major vessels colored by fluoroscopy	0-3
Thal	Defect type	3→Normal, 6→Fixed defect, 7→Reversible defect
Num	Heart disease	0-4

## Materials

This study uses the same dataset used by Nahar *et al.* (2013a; 2013b). Which is provided by the University of California, Irvine Machine Learning Repository (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases>). This dataset is a part of the Heart Disease Data Set (the part obtained from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation), using a subset of 14 attributes; 13 numeric input attributes namely age, sex, chest pain type, cholesterol, fasting blood sugar, resting ecg, maximum heart rate, exercise induced angina, old peak, slope, number of vessels colored and thal and one output attribute named Num. The task here is to detect the presence of heart disease in the patient. There are five classes indicating either healthy or one of four sick types. Table 1 contains the detailed description of this dataset with continuous, binomial, nominal and ordinal attributes.

## Bayesian Networks

Several classification algorithms have been developed in the field of data mining information systems. Some of these algorithms are able to produce probability estimates rather than predictions. That is for each class label, they estimate the probability that a given sample belongs to that class. Probability estimates are often more useful than plain predictions. They allow predictions to be ranked and their expected costs to be minimized. BNs among other models are ones of these classification approaches. The benefits of BNs are that they present well-founded methods to represent any arbitrary probability class distributions concisely and

comprehensibly in a graphical manner. BN model is drawn as a network of nodes, one for each attribute, connected by directed edges in such a way that there are no cycles. In other words, a BN is a directed acyclic graph consisting of (Cheng *et al.*, 1998):

- Nodes (or small circles), that stand for random attributes; edges (or arrows), which represent probabilistic relationships among these attributes and
- For each node, there exists a local probability distribution attached to it that depends on the state of its parents

BN consists of a qualitative part (structural model) that presents a visual representation of the interactions among attributes and a quantitative part (set of local probability distributions), which provides probabilistic inference and numerically measures the effect of attributes on each other. The qualitative and quantitative parts mutually determine a unique joint probability distribution over the attributes in a specific problem (Cooper, 1999). The main idea within the structure of BN is that of independence. This idea refers to the case where the instantiation of a specific attribute leaves other two attributes independent of each other. BN model allows the representation of a joint probability distribution in a compact and economical way by making extensive use of conditional independence, as shown in Equation 1:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i / P_a(X_i)) \quad (1)$$

Where:

- $P(X_1, X_2, \dots, X_n)$  = Represent the joint probabilities of attributes  
 $X_1, X_2, \dots, X_n, P_a(X_i)$  = Represents the set of parent nodes of  $X_i$ ; i.e., nodes with edges pointing to  
 $X_i$  and  $P(X_i/P_a(X_i))$  = Represents the conditional probability of  $X_i$  given its parents

Equation 1 shows how to pick up a joint probability from a product of local conditional probability distributions; such representation may be used to solve classification problems (Linda *et al.*, 2008; Cheng *et al.*, 1998). The learning algorithm for BN has to contain two components:

- A function for evaluating a given network (goodness of fit measure)
- A method for searching through the space of possible networks

Normally, the learning algorithm starts with a given ranking of the attributes (i.e., nodes). Then it processes each node in turn and greedily adds edges from previously processed nodes to the current one. In each step it selects the edge that maximizes the network's score. If there is no additional enhancement, attention goes to the next node. The Naïve Bayes (NB) classifier is one of the most effective methods to build BNs (Friedman *et al.*, 1997). However, it works well only for simple distributions. Usually, NB network is used as a starting point for the search. In this study, two learning algorithms have been used to build the BN classifiers starting NB network; Tree Augmented Naïve Bayes (TAN) and Markov Blanket Estimation (MBE) learning algorithms.

#### Markov Blanket Estimation (MBE)

MBE is a learning algorithm to create BN model by identifying the conditional independence relationships among the attributes. This algorithm ensures that every attribute in the dataset is in the Markov blanket of the node that represents the class attribute (Witten and Frank, 2005). A node's Markov blanket includes all its parents, children and children's parents. Hence, if a node is absent from the class attribute's Markov blanket, its value is completely irrelevant to the classification. Using statistical tests, this algorithm finds the conditional independence relationships among the nodes and uses these relationships as constraints to construct a BN structure (Baesens *et al.*, 2002; Frey *et al.*, 2003). This algorithm is referred to as a dependency-analysis-based or constraint-based algorithm. The Conditional Independence (CI) test investigates whether two

attributes are conditionally independent. There are two common methods to compute the CI test; Pearson chi-square test and log likelihood ratio test (Witten and Frank, 2005). The Likelihood Ratio (LR) tests for target-predictor independence by calculating a ratio between the maximum probability of a result under two different hypotheses. While the Pearson Chi-square (CHI) assesses for target-predictor independence by using a null hypothesis that the relative frequencies of occurrence of observed events follow a specified frequency distribution. MBE explores not only the relations between the class target and predictive attributes, but also the relations among these predictive attributes themselves. Both independence tests; Likelihood ratio and Chi-square have been used to predict and diagnose heart diseases.

#### Tree Augmented Naïve Bayes (TAN)

TAN is an improvement over the Naïve Bayes model as it allows for each attribute to depend on another attribute in addition to the target attribute. The class attribute is the single parent of each node of a NB network: TAN considers adding a second parent to each attribute; the predictive attributes are allowed to point to each other (as long as no cycles are introduced). The decision to add these edges between attributes is made on the basis of a specific goodness of fit measure, such as Maximum Likelihood (ML), Bayesian Dirichlet (BD) (Heckerman *et al.*, 1995), Bayesian Information Criterion (BIC) (Grunwald *et al.*, 2005), or Akaike Information Criterion (AIC) (Bozdogan, 2000), among others. If the class node and all corresponding edges are excluded from consideration and assuming that there is exactly one node to which a second parent is not added, the resulting classifier has a tree structure rooted at the parentless node. There is an efficient algorithm for finding the set of edges that maximizes the network's likelihood based on computing the network's maximum weighted spanning tree (Witten and Frank, 2005). This method associates a weight to each edge corresponding to the mutual information between the two variables. The TAN learning procedure is as follows:

- Assume the training dataset  $D, X, Y$  as input
- Build the tree-like network structure over the predictive attribute  $X$  by using the maximum weighting spanning tree
- Add  $Y$  as a parent of every  $X_i$  where  $1 \leq i \leq n$
- Estimate the parameter of TAN (conditional probability of each node given the value of its parents) using ML criterion

When the dataset is small it is preferable to use the BD criterion to prevent the over fitting of the model (Heckerman *et al.*, 1995).

## Experimental Results

The performance of each classification model is evaluated using three statistical measures; classification accuracy, sensitivity and specificity. These measures are defined using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). A true positive decision occurs when the positive prediction of the classifier coincided with a positive prediction of the physician. A true negative decision occurs when both the classifier and the physician suggest the absence of a positive prediction. False positive occurs when the system labels a benign case; a negative one as a positive one (malignant). Finally, false negative occurs when the system labels a positive case as negative (benign). Classification accuracy is defined as the ratio of the number of correctly classified cases and is equal to the sum of TP and TN divided by the total number of cases N:

$$Accuracy = \frac{TP + TN}{N} \quad (4)$$

Sensitivity refers to the rate of correctly classified positive and is equal to TP divided by the sum of TP and FN. Sensitivity may be referred as a True Positive Rate:

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

Specificity refers to the rate of correctly classified negative and is equal to the ratio of TN to the sum of TN and FP. False Positive Rate equals (100-specificity):

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

The dataset contains 303 sample with class distribution: Num = 0 (Healthy) 164 samples, 54%; Num = 1, 55 samples 18.15%; Num = 2, 36 samples; Num = 3, 35 samples, 11.55% and Num = 4, 13 samples 4.29%. The whole dataset is divided for training the models and test them by the ratio of 70:30% respectively. The training set is used to estimate each model parameters, while the test set is used to independently assess the individual models. Three models have been trained to predict the diagnosis of heart disease; MBE, TAN and SVM. These

models are applied again to the entire dataset and to any new data. The predictions are compared to the original classes to identify true positives, true negatives, false positives and false negative values. These values have been computed to construct the confusion matrix. The performance of MBE and TAN is benchmarked with well-known SVM. The component nodes of the proposed stream are shown in Fig. 1. The stream is implemented in SPSS Clementine data mining workbench (SPSS, 2007). A brief description of each component is given in the following.

Cleveland\_14 dataset node is connected directly to Excel file that contains the source data. The dataset was explored for incorrect, inconsistent. Only, the age attribute is normalized and no preprocessing for other attributes. They are ordinal and nominal data types. Type node specifies the field metadata and properties that are important for modeling and other work in Clementine. These properties include specifying a usage type, setting options for handling missing values, as well as setting the role of an attribute for modeling purposes. Select node is used to ensure that every sample has a specified class label and discard all samples with undefined ones. Partition node is used to generate a partition field that splits the data into separate subsets for the training and test the models. In this study, the dataset was partitioned by the ratio 70:30% for training and test subsets respectively.

MBE classifier node is to train and test a Bayesian classifier with MBE learning algorithm and Likelihood Ratio (LR). MBE algorithm selects the set of nodes in the dataset that contain the target attribute's parents, its children and its children's parents. Essentially, MBE identifies all the attributes in the network that are needed to predict the target class. Figure 2 illustrates the network topologies with LR test. It is clear that there is no direct relation between the class attribute and the mass density in both topologies. It could be concluded that mass density attribute is out of the Markov blanket of the severity class. The MBE model with LR conditional probability test is assumed to be more accurate and experimental results shows MBE with, 99, 99, 99, 99 and 100% classification target accuracy for the five classes of Num target attribute.

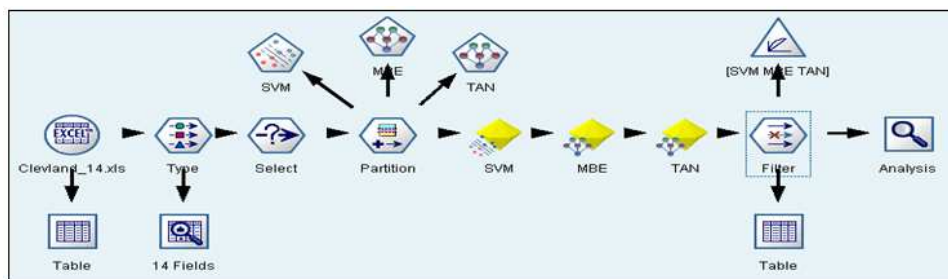


Fig. 1. Data mining stream for the diagnosis of the cardiovascular diseases

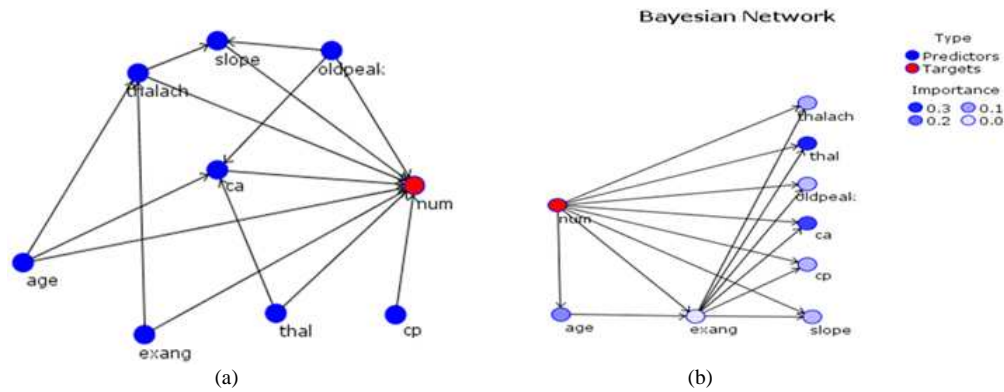


Fig. 2. Topology of the Bayesian networks (a) MBE learning algorithms with Likelihood ratio test (b) TAN learning algorithm with Likelihood ratio test

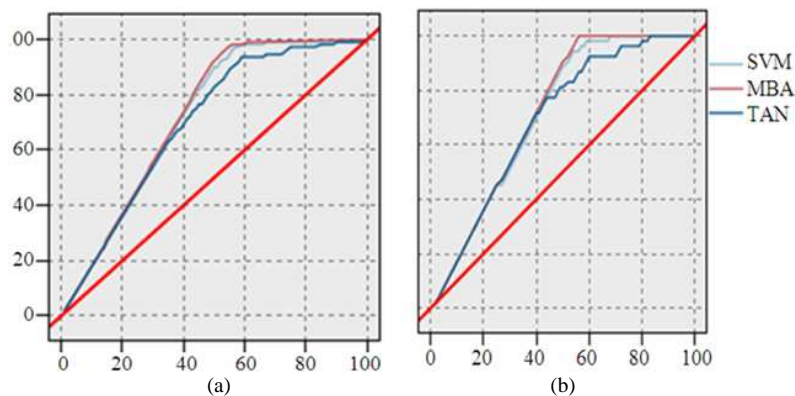


Fig. 3. Gain charts for training and testing datasets (a) training dataset (b) testing dataset

TAN classifier node is to train and test a BN model with TAN learning algorithm where each predictive attributes are allowed to depend on each other in addition to the target attribute, thereby increasing the classification accuracy. In order to prevent over fitting of the classifier, the maximum likelihood is used to control the estimation of the conditional probability for each node given the values of its parents.

SVM classifier node is used to train the well-known SVM with Radial Basis Function (RBF) kernel to benchmark the performance of Bayesian classifiers. There are three parameters that need to be optimized; regularization parameter, regression precision and RBF gamma. Trial and error may be used to find the best values for other parameters (Witten and Frank, 2005; Vapnik, 1998). This study are set to 10, 0.1 and 0.1. These values result in 86.47 and 88.54% prediction accuracies for training and test subsets respectively.

Filter, analysis and evaluation nodes are used to select and rename the classifier outputs in order to compute the performance statistical measures and to graph the evaluation charts.

Table 2 shows the computed confusion matrix, each cell contains the raw number of samples classified for

the corresponding combination of desired and actual model outputs. Table 3 presents the values of the statistical parameters (sensitivity, specificity and total classification accuracy) of the predictive models. Sensitivity and Specificity approximates the probability of the positive and negative labels being true. These results show that the sensitivity, specificity and classification accuracy of Bayesian network with MBE learning method and likelihood ratio test are better than those of the other individual classifiers.

Table 4 shows that the classification accuracy of MBE model is 97.92% of test samples, while TAN and SVM models have 88.54 and 70.83% respectively. The MBE has achieved better sensitivity and specificity also.

Gain charts have been used to compare the performances of different predictive models. The gain curves summarize the utility that can be expected by using the respective predictive models, as compared to using baseline information only. Figure 3a and b show the cumulative gain curves of the two Bayesian models and the SVM for training and test samples. The higher lines indicate better models, especially on the left side of the chart. The higher curves are of the MBE that coincides with the above statistical measures.

Table 2. Confusion matrices of different models for test data partitions

Classifier	Desired results	Real output				
		0	1	2	3	4
SVM	0	52	1	0	0	0
	1	3	13	1	1	0
	2	1	1	7	1	0
	3	0	1	0	10	0
	4	0	1	0	0	3
MBE	0	53	0	0	0	0
	1	1	17	0	0	0
	2	0	0	10	0	0
	3	0	0	1	10	0
	4	0	0	0	0	4
TAN	0	49	2	2	0	0
	1	7	5	4	2	0
	2	1	1	6	2	0
	3	2	3	0	6	0
	4	0	0	2	0	2

Table 3. Values of statistical measures for different predictive models

Classifier	Desired results	Sensitivity	Specificity	Accuracy
SVM	0	0.98	0.91	0.95
	1	0.72	0.95	0.91
	2	0.70	0.99	0.96
	3	0.91	0.98	0.97
	4	0.75	1.00	0.99
MBE	0	1.00	0.98	0.99
	1	0.94	1.00	0.99
	2	1.00	0.99	0.99
	3	0.91	1.00	0.99
	4	1.00	1.00	1.00
TAN	0	0.92	0.77	0.85
	1	0.28	0.92	0.80
	2	0.60	0.91	0.88
	3	0.55	0.95	0.91
	4	0.50	1.00	0.98

Table 4. Total accuracies of different models on the diagnosis of training and testing subsets

Model		Training		Testing	
		# of patterns	Accuracy (%)	# of patterns	Accuracy (%)
SVM	Correct	179	86.47	85	88.54
	Wrong	28	13.53	11	11.46
MBE	Correct	194	93.72	94	97.92
	Wrong	13	6.28	2	2.08
TAN	Correct	150	72.46	68	70.83
	Wrong	57	27.54	28	29.17

## Conclusion

Bayesian network classifiers have three major advantages; they have the ability to deal with missing values, they explicitly provide the conditional probability distributions of the values of the class attribute given the values of the other input attributes and finally they are easy to comprehend. For these fine proprieties, the awareness to apply and use Bayesian network classifiers in the medical context is

increasing. The main goal of this study is to show the effectiveness of these classifiers in the prediction of cardiovascular disease. Two different implementations of Bayesian network have been applied on the mammographic mass dataset; tree augmented Naïve Bayes and Markov blanket estimation learning algorithms. The cardiovascular dataset is provided by University of California, Irvine machine learning repository. This dataset is a part of the Heart Disease Data Set (the part obtained from the

V.A. Medical Center, Long Beach and Cleveland Clinic Foundation), using a subset of 303 instances of heart disease data each having 14 attributes; 13 numeric input attributes and one class output. The performances of Bayesian classifiers are benchmarked against the support vector machine using statistical measures and gain charts. Bayesian network classifier with Markov Blanket Estimation outperformed other classifiers on the prediction of cardiovascular diseases.

## Acknowledgement

The authors would like to express their gratitude to Eng. Mohamed Zohdi from SPSS Egypt office for the software of Predictive Analytics portfolio (PASW) statistical software they granted (formerly called SPSS).

## Funding Information

This study was supported by the computers and systems department, Electronic Research Institute of Egypt. We thank Prof. Hesham Eldeeb, the president of the Institute and all our colleagues in the department for all the hardware and software resources and for the valuable discussion they provide.

## Author's Contributions

**Alaa Elsayad:** Designed the study, collected the datasets from databases and analyzed the data, then prepared the original draft the manuscript.

**Mahmoud Fakhr:** Guided the analysis and the interpretation of the results, validated the experimental results and made discussion, collecting the literature. All authors reviewed and approved the manuscript.

## Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

## References

- Baesens, B., M. Egmont-Petersen, R. Castelo and J. Vanthienen, 2002. Learning bayesian network classifiers for credit scoring using markov chain monte carlo search. Proceedings of the 16th International Conference on Pattern Recognition, Aug. 11-15, IEEE Xplore Press, pp: 49-52.  
DOI: 10.1109/ICPR.2002.1047792
- Bozdogan, H., 2000. Akaike's information criterion and recent developments in information complexity. J. Math. Psychol., 44: 62-91.  
DOI: 10.1006/jmps.1999.1277
- Brause, R.W., 2001. Medical analysis and diagnosis by neural networks. Proceedings of the 2nd International Symposium on Medical Data Analysis, Oct. 8-9, Madrid, Spain, pp: 1-13.  
DOI: 10.1007/3-540-45497-7\_1
- Can, M., 2013. Diagnosis of cardiovascular diseases by boosted neural networks. Southeast Eur. J. Soft Comput., 2: 91-102.
- Cheng, J., B. David and W. Liu, 1998. Learning bayesian networks from data: An efficient approach based on information theory.
- Cooper, G.F., 1999. An overview of the representation and discovery of causal relationships using Bayesian networks.
- Cruz-Ramírez, N., H. Acosta-Mesa, H. Carrillo-Calvet and R. Barrientos-Martínez, 2009. Discovering interobserver variability in the cytodiagnosis of breast cancer using decision trees and bayesian networks. Applied Soft Comput., 9: 1331-1342.  
DOI: 10.1016/j.asoc.2009.05.004
- Cruz-Ramírez, N., H. Acosta-Mesaa, H. Carrillo-Calvet, L. Nava-Fernándezc and R. Barrientos-Martínez, 2007. Diagnosis of breast cancer using Bayesian networks: A case study. Comput. Biol. Med., 37: 1553-1564.  
DOI: 10.1016/j.combiomed.2007.02.003
- Darwiche, A., 2008. Bayesian Networks. In: Handbook of Knowledge Representation, Foundations of Artificial Intelligence, Frank Van Harmelen, Vladimir Lifschitz and Bruce Porter (Eds.), Elsevier, Amsterdam, ISBN-10: 0080557023, pp: 1034-1034.
- Felipe, S.C., M.M.D.S. Pires and S.M. Nassar, 2013. Analysis of bayesian classifier accuracy. J. Comput. Sci., 9: 1487-1495.  
DOI: 10.3844/jcssp.2013.1487.1495
- Filippo, A., A. López, E.M. Peña-Méndez, P. Vañhara and A. Hampl *et al.*, 2013. Artificial neural networks in medical diagnosis. J. Applied Biomed., 11: 47-58.  
DOI: 10.2478/v10136-012-0031-x
- Frey, L., D. Fisher, I. Tsamardinos, C. Aliferis and A. Statnikov, 2003. Identifying markov blankets with decision tree induction. Proceedings of the 3rd IEEE International Conference on Data Mining, Nov. 19-22, IEEE Xplore Press, pp: 59-66.  
DOI: 10.1109/ICDM.2003.1250903
- Friedman, N., D. Geiger and M. Goldszmidt, 1997. Bayesian network classifiers. Machine Learning, 29: 131-163. DOI: 10.1023/A:1007465528199  
ftp://ftp.ics.uci.edu/pub/machine-learning-databases
- Ghumbre, S., C. Patil and A. Ghatol, 2011. Heart disease diagnosis using support vector machine. Proceedings of the International Conference on Computer Science and Information Technology, (SIT, 11), Pattaya, pp: 84-88.



- Grunwald, P., I.J. Myung and M.A. Pitt, 2005. Advances in Minimum Description Length: Theory and Applications. 1st Edn., MIT Press, Cambridge, ISBN-10: 0262072629, pp: 444.
- Heckerman, D., D. Geiger and D.M. Chickering, 1995. Learning bayesian networks: The combination of knowledge and statistical data. Machine Learning, 20: 197-243. DOI: 10.1023/A:1022623210503
- Nahar, J., T. Imama, K.S. Ticklea and Y.P.P. Chenb, 2013b. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst. Applic., 40: 1086-1093. DOI: 10.1016/j.eswa.2012.08.028
- Nahar, J., T. Imama, K.S. Ticklea and Y.P.P. Chenb, 2013a. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Syst. Applic., 40: 96-104. DOI: 10.1016/j.eswa.2012.07.032
- Khan, M.G., 2005. Heart Disease Diagnosis and Therapy: A Practical Approach. 2nd Edn., Springer, Humana Press, ISBN-10: 1588294390, pp: 544.
- Khan, M.G., 2010. Encyclopedia of Heart Diseases. 2nd Edn., Humana Press, ISBN-10: 1607612186, pp: 470.
- Linda, C., S. Renooij, A. Feelders, A. Groote and M. Eijkemans *et al.*, 2008. Aligning bayesian network classifiers with medical contexts. Proceedings of the 6th International Conference on Machine Learning and Data Mining in Pattern Recognition, Jul. 23-25, Leipzig, Germany, pp: 787-801. DOI: 10.1007/978-3-642-03070-3\_59
- Nihat, Y., I. Onur and S.U. Mustafa, 2014. A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases. J. Med. Syst., 38: 48-48. DOI: 10.1007/s10916-014-0048-7
- Rani, K.U., 2011. Analysis of heart diseases dataset using neural network approach. Int. J. Data Mining Know. Manage. Process, 1: 1-8. DOI: 10.5121/ijdkp.2011.1501
- Rosendorff, C., 2013. Essential Cardiology: Principles and Practice. 3rd Edn., Springer Science and Business Media, New York, ISBN-10: 1461467055, pp: 823.
- Sabina, A., A. Heijl, D. Bizios and B. Bengtsson, 2012. Comparison of clinicians and an artificial neural network regarding accuracy and certainty in performance of visual field assessment for the diagnosis of glaucoma. Acta Ophthalmol., 91: 413-417. DOI: 10.1111/j.1755-3768.2012.02435.x
- Salim, N., 2004. Medical diagnosis using neural network.
- Sherly, K.K., 2012. A comparative assessment of supervised data mining techniques for fraud prevention. TIST Int. J. Sci. Technol. Res., 1: 1-6.
- SPSS, 2007. Data mining work bench software. SPSS Inc.
- Steven, M., M.D. Hollenberg and M.D.T. Walker, 2006. Cardiology in Family Practice: A Practical Guide. 1st Edn., Humana Press, ISBN-10: 1588295095, pp: 176.
- Vapnik, V.N., 1998. Statistical Learning Theory. 1st Edn., John Wiley and Sons, New York, ISBN-10: 0471030031, pp: 736.
- Witten, I.H. and E. Frank, 2005. Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edn., Morgan Kaufmann, Amsterdam, ISBN-10: 008047702X, pp: 560.
- Yadav, C., S. Lade and M.K. Suman, 2014. Predictive analysis for the diagnosis of coronary artery disease using association rule mining. Int. J. Comput. Applic., 87: 9-13. DOI: 10.5120/15195-3575