A Hybrid Estimation System for Medical Diagnosis using Modified Full Bayesian Classifier and Artificial Bee Colony

Ahmed T. Sadiq[1], Noor T. Mahmood[*2]

[1]Department of Computer Science, University of Technology, Baghdad, Iraq.
[2]Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq.

Abstract

This paper presents a hybrid approach called Modified Full Bayesian Classifier (*M-FBC*) and Artificial Bee Colony (*MFBC-ABC*) for using it to medical diagnosis support system. The datasets are taken from Iraqi hospitals, these are for the heart diseases and the nervous system diseases. The *M-FBC* is depended on common structure known as *naïve Bayes*. The structure for network is represented by *D-separated* for structure's variables. Each variable has Condition Probability Tables (*CPTs*) and each table for disease has Probability. The ABC is easy technique for implementation, has fewer control parameters and it could be easier than other swarm optimization algorithms, so that hybrid with other algorithms to reach the optimal structure. In the input stage, the symptoms and the medical history for the patient are processed through the BNs structures to obtain from Modified Full Bayesian Classifier- Artificial Bee Colony (MFBC-ABC). The proposed system inputs all medical dataset and it filters and extracts the dataset. After the evaluation of the structures, the system can select the best optimal structure by determining the accepted accuracy. The accuracy for M-FBC model is approximately (93%) for heart diseases and approximately (98%) for nervous system diseases. While in The MFBC-ABC model, the accuracy is approximately (100%) for heart diseases and is approximately (99%) for nervous model diseases. The experimental results shown that the results for MFBC-ABC is better than on M-FBC.

نظام تخميني مهجن لأغراض التشخيص الطبي باستخدام مصنف **Bayesian** المطور وخوارزمية مستعمرة النحل الاصطناعية

احمد طارق صادق[1]، نور ثامر محمود[2*]

[1]قسم علوم الحاسبات، الجامعة التكنلوجية، بغداد، العراق،[2]قسم علوم الحاسبات، كلية العلوم، جامعة بغداد، بغداد، العراق.

الخلاصة

البحث يعرض الطريقة الهجينة التي تعتمد على ادوات الذكاء الاصطناعي المتكونة من مصنف Bayesian المطور (M–FBC)وخوارزمية مستعمرة النحل الاصطناعية (ABC) مما يؤدي الى –MFBC (ABC) حيث تستخدم لتدعم نظام التشخيص الطبي. ان قواعد البيانات المعتمدة في البحث لأمراض القلب وامراض الجملة العصبية مأخوذة من المستشفيات العراقية . يعتمد M–FBC على الهيكل العام المعروف naïve Bayes. ويتم تمثل متغيرات هيكل الشبكة بواسطة D–separating. وكل متغير يحتو ي نسبة احتمالية وجودة في الجدول . كما وان كل جدول يتم حساب الاحتمالية له . خوارزمية مستعمرة النحل الاصطناعية هي أسهل من غيرها من خوارزميات السرب التي تستخدم للوصول للحل الامثل لذلك هي سهله

_____

*E-mail: noorthamer@gmail.com

<div dir="rtl">

التهجين مع غيرها من الخوارزميات للوصول الى الهيكل الامثل . في مرحلة الادخال، يتم معالجة الاعراض والتاريخ الطبي للمريض من خلال هياكل BN الناتجة عن MFBC-ABC. النظام المقترح يأخذ كل البيانات المدخلة الطبية وفلاتر واستخراج البيانات . بعد تقيم الهياكل، يمكن للنظام تحديد أفضل هيكل مثالي وذلك بتحديد قيمة الدقة المقبولة . ان دقة الموديل M-FBC كانت تقريبا 93% لأمراض القلب ولأمراض الجملة العصبية تقريبا 98%. ومن القيام باختبار دقة MFBC-ABC لأمراض القلب كانت تقريبا 100% وامراض الجملة العصبية كانت تقريبا 99% والنتائج التجريبية أظهرت MFBC-ABC هو أفضل من M-FBC.

</div>

1 Introduction

In this paper, the proposed system is based on hybrid approach for medical diagnosis, which can help to diagnose the diseases without required all information about patient's medical history and symptoms.

The challenge in this paper lies in obtaining data that is acquired from the patient or the patient's file in order to take the medical history, medical tests of the patient, the symptoms suffered by the patient and entering these information into the system for the purpose of identifying the correct diagnosis and the accurate result for the patient as quickly as possible and with less cost.

The proposed system is used a Bayesian network. BNs are rapidly becoming a leading technology in applied Artificial Intelligence. By combining a graphical representation of the dependencies between variables with probability theory and efficient inference algorithms, BNs provide a powerful and flexible tool for reasoning under uncertainty. Practitioners hold much of their knowledge implicitly, making an appealing target for expert systems development [1].

BNs consist of two parts: the structure and the parameters. The structure is represented by a directed acyclic graph (DAG) and the parameters determine the conditional probability distributions for each node [2].

The approaches in this paper using the Bayesian Networks (BNs) to extract the structure that classify a medical data. That's where these conditions attached are cases of medical data as possible to take advantage of them in the formation of the structure of BNs. Causal and Evidential are most common types of reasoning in BNs, however thanks to evidential reasoning Bayesian networks became the ubiquitous tool for differential diagnosis even though this type requires much more computational power than causal reasoning. On the other hand causality is the fundamental part of Bayesian networks (BNs) and without it there would be no reasoning [3].

In this paper, Artificial Bee Colony (*ABC*) optimization algorithm, which is described the foraging behavior of honey bees for numerical optimization problems [4] based on nature-inspired ideas. Some examples of such algorithms include ant colony optimization, evolutionary algorithm, particle swarm optimization, harmony search etc. Most of these algorithms are metaheuristic-based search techniques and generally referred to as multipurpose optimization algorithms because of their applicability to a wide range of problems. In a similar context, Artificial Bee Colony algorithm (*ABC*) was initially published by Karaboga in 2005 as a technical report for numerical optimization problems. Its development was motivated by simulating the intelligent foraging behavior of honey bees in their colony and its performance was initially measured using benchmark optimization function [5].

The aim of this paper is to present M-FBC model and MFBC-ABC model to obtain high speed of prediction for disease infected by the patients in order to save their life. This system services Iraqi society. All patient data entered into the database to reduce files that written by hand. The patient has not needed to visit more than one doctor to know the disease that he suffered. The system is not needed all information about medical examinations to obtain the correct diagnosis. It needs the patient's symptoms and medical history that mentioning when complained by patients. That save time, effort and money.

The system has been applied in some Iraqi hospitals and it gives results that match with the doctor diagnosis in real world. The special doctors certified these results. The error percentage identified by doctors was approximately (1 - 10) % for most of the cases.

Thus, that should conclude the use of the BNs structure to make a decision regarding the diagnosis through the work of deductive relations between the attributes to reach the appropriate structure of the disease and the disease is diagnosed through this structure of data-based truth until the best results are obtained.

2 Related Work

Ahmed, Noor, 2014 [6],this paper presents a modified Bayesian Network (BN) for Full Bayesian Classifier (FBC) to diagnosis the condition of a patient from some of symptoms and medical history. The conditions are the heart diseases and the nervous diseases. Modified for FBC is depended on common structure known as naïve Bayes. Determining the network structure is D-separated by the variable. Each variable has CPT and each disease (table) has Probability. By modified the equation for find CPT for each variable and Probability for each disease (table) in Modified-FBC (*M-FBC*) structure, the system for diagnosis is designed. Aim of paper is speed of predict for disease infected by the patient in order to save his life. By explaining the patient's symptoms experienced by mentions for medical history is through the introduction of all the information into the system to reach the approximate diagnosis of a patient's disease patient.

The system has been applied in Iraqi hospitals and reached acceptable results and that a certificate therapists doctors in the hospital. Where the margin of error were identified by doctors around 5% for most of the cases. So that it is accurate diagnosis for heart and the nervous system diseases, thus can enter other data set for diseases is entered into the system is up the diagnostic process as well. The experimental resulted show that the successful ratio of heart diseases (93%) and nervous system diseases (98%) approximately.

Beáta, Lehel, 2009[7]. Bayesian networks encode causal relations between variables using probability and graph theory. They can be used both for prediction of an outcome and interpretation of predictions based on the encoded causal relations. In this paper a tree-like Bayesian network learning algorithm optimized has analyzed for classification of data and solutions have given to the interpretation and analysis of predictions. The classification of logical – i.e. binary – data arises specifically in the field of medical diagnosis, where the survival chance has been to predict based on different types of medical observations or the most relevant cause corresponding again have selected to a given patient record. Surgery survival prediction was examined with the algorithm. Bypass surgery survival chance must be computed for a given patient, having a data-set of 66 medical examinations for 313 patients.

Li Hongyan, Wu Zhong, Li Cheng, 2009 [8], this article combines with China's aviation status, conducts a more deep and systematic study of risk analysis and decision-making in overbooking strategy through using qualitative and quantitative method, which contributes a certain practical value. Overbooking refers that an airlines allow confirmed booking beyond the seating capacity for the flight based on reasonable estimation. The airlines are allowed a certain amount of overbooking because cancellations are fairly frequent for many reasons. There are no cancellation charges, only paying a few indemnities even when returning tickets or missing flights after purchasing tickets. Overbooking surely runs a certain risk. Usually, overbooking is estimated by cancellation of reservation and probability of passenger's no-show. Thus, there may be excessive overbooking and the occurrence of super-real, some passengers cannot check-in, thus, extra compensation, operating loss as well as loss of goodwill are created for airlines. Therefore, it is obviously necessary for overbooking to reduce possibilities of vacancy losses. The overbooking is more, the possibility of vacancy lose is smaller. Overbooking is an effective means for airlines to enhance revenue. Today, overbooking has been commonly used by airlines worldwide and already achieved a certain economic benefits.

Silvia, Luis, Juan, Susana, Jose, Jose´ Luis, 2004 [9]. Due to the uncertainty of many of the factors that influence the performance of an emergency medical service, Bayesian networks are proposed using to model this kind of system. Different algorithms are used for learning Bayesian networks in order to build several models, from the hospital manager's point of view, and apply them to the specific case of the emergency service of a Spanish hospital. Real problem includes preliminary data processing, the experiments carried out, the comparison of the algorithms from different perspectives, and some potential uses of Bayesian networks for management problems in the health service.

Stefan, Dr. Lionel, 2011 [10].By using Bayesian networks as the framework, a practical new modeling approach has shown based on the widely studied Wisconsin Breast Cancer Database. Comparable has been prediction accuracy with the results of all known studies on this topic. With BayesiaLab as the software tool, modeling with Bayesian networks becomes accessible to a very broad range of analysts and researchers, including non-statisticians. The speed of modeling, analysis and subsequent implementation make BayesiaLab a suitable tool in many areas of research and especially for translational science.

Jyotirmay, Ognjen, Kwabena, Erol, Yufeng and Ping, 2010 [11].The objective of this paper is to explore the implementation of a Bayesian Belief Network for an automated breast cancer detection support tool. It is intuitive that Bayesian networks are employed as one viable option for computer-aided detection by representing the relationships between diagnoses, physical findings, laboratory test results and imaging study findings. This work brings important entities such as Radiologists, Image Processing Scientists, Data Base Specialists and Applied Mathematicians on a common platform. A brief background concerning causal networks, probability theory and Bayesian networks is given; available computational tools and platforms are described. It is explained that, by exploiting conditional independencies entailed by influence chains, it is possible to represent a large instance in a Bayesian network using little space, and it is often possible to perform probabilistic inference among the features in an acceptable amount of time. The next steps towards realizing a Bayesian Belief Network Implementation are described. Bayesian networks have an unparalleled advantage of being able to exploit the explicit structure of the domain model to derive a graphical representation for learning. The encoding of independencies in the network topology admits the design of efficient procedures for performing computations over the network. For the application of computer-aided detection in mammography, the researchers intend to design an interface between the project's Bayesian network learning algorithm and the radiologists, so that the radiologists can have interaction with the system by labeling only a small number of informative images presented by the active learning algorithm.

3 Bayesian Networks

BNs are triplets *(V, E, P)*, where *(V, E)* is a directed acyclic graph (DAG) with nodes *V*, edges *E*, and a set of probability distributions *P*, called parameters, whose elements are assigned to the nodes of the graph. The nodes represent domain variables and edges mark direct causal relations between these variables [12].

The network encodes a joint probability distribution function representative to the domain:

$$P(X) = \prod_{i=1}^{n} P(X_i | par(X_i)) \tag{1}$$

where *n* is the number of domain variables, X_i is a node from the BN and *par(X_i)* is the set of X_i's parents. The probability of the graph ensures the product to be finite [12].

A tree-like representation has employed for the topology of BN in order to increase efficiency in class variable estimation and interpretation. This algorithm has described, where a tree has constructed in such a way that the root of the tree will be the class variable and the remaining nodes are attributes [12].

Direct causal relations encoded by the BN are interpreted as the maximum of mutual respective conditional mutual information between nodes. Now the necessary information theoretical have presented concepts for our algorithm [12].

The following notations have used: *X* and *Y* are random variables defined on probability spaces Ω_X respective Ω_Y with corresponding distribution functions *p(x)* respective *p(y)*. Their joint and conditional probability functions have used, denoted with *p(x,y)* and *p(x/y)* respectively[12].

Mutual information is the quantity of information two random variables contain about each other, defined as [12]:

$$I(X,Y) = \sum_{x \epsilon \Omega X} \sum_{y \in \Omega y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \tag{2}$$

4 Learning Modified-*FBC* (*M-FBC*)

Completing phase Filter and Feature Extraction, the dataset is stored in the repository Feature DB. All dataset becomes input to M-FBC model. Checked the table in DB is ascertained to be empty and calculate the mutual information for each pair of attributes in the table according to equation (3)

$$M(X;Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}, \tag{3}$$

for purpose of knowledge as a measure of relationship between two random attributes that are in table simultaneously. After that a threshold between each pair of attributes is calculated according to equation (4)

$$\varphi(X_i, X_j) = \frac{\log(m)}{2m} \times T_{ij} , \qquad (4)$$

and the maximum threshold between attributes. Because equation (3) is caused merely by noise, thus a threshold is needed to rule whether the dependency between two variables is reliable. Weight is calculated for each attribute in the table according to equation (5)

$$W(X) = \sum_{Y(Y \neq X)}^{M(X,Y) > Max\ (\varphi)} M(X; Y). \qquad (5)$$

All weights for variables are ordered as descending and select the root. That has maximum weight and the other evidences have less weight of the root. The structure is call naïve Bayes but inverted shown figure-1 [6].
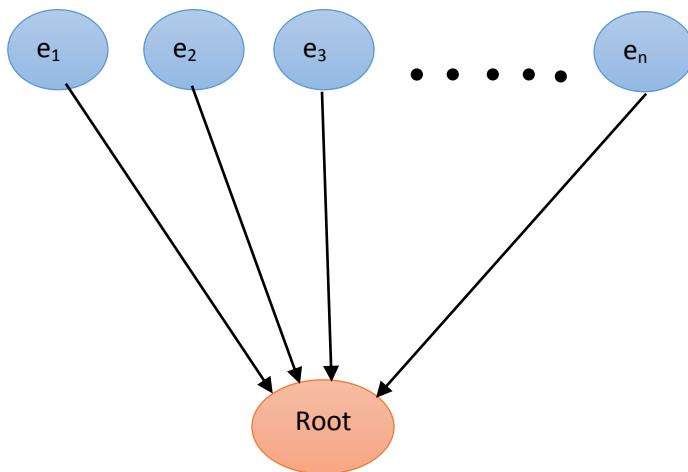

Figure 1- common effect [6].

CPTs is calculated for each node in network structure. The network structure is call D-separate shown in figure-2. When node is positive then selected edge =1 and then calculated (sum no. of node which value =1 / total nodes) or node is negative then selected edge = 0 and then calculated (1- (sum no. of node which value =0 / total nodes)) [6].
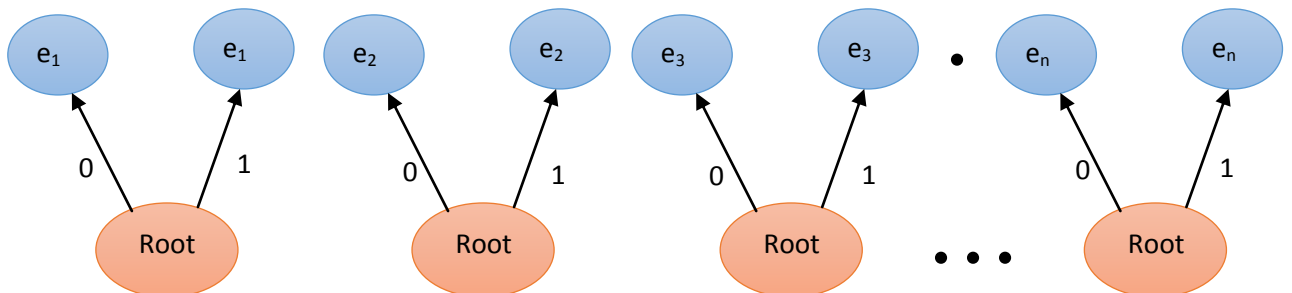

Figure 2- CPTs (common causal) [6].

Algorithm *M-FBC Structure* [6].
Input: Table (i), (i =1,…,n) .
Output: A full BNs **B,** Probability (Prob).
    1.   B = empty.
    2.   If Table (i) is empty then B = empty.
    3.   Calculate mutual information between each pair of attributes base on equation (3).
    4.   Calculate threshold between each pair of attributes base on equation

$$\varphi(X,Y) = \frac{\log{(m)}}{2m} \times T_{xy} \, , \qquad\qquad same\ as\ equation\ (4).$$

$$m = \sum_{i=0}^{no.of\ coloms} no.frecuncy\ in\ all\ rows\ in\ the\ tables$$

$$T_{xy} = |no.of\ variable\ in\ X| \times |no.of\ variable\ in\ Y|$$

    5.   Find the maximum of threshold = Max(φ).
    6.   Calculate weight for each attributes in the Table (i) base on equation (5).
    7.   Sorted all weight for all variables.
    8.   B = the maximum weight is parent and the other are evidences.
    9.   Calculate CPT for each node in structure B.

If nod = positive then $CPT = \left(\frac{\sum_1^r X = yes}{m}\right)$ (6)

Else $CPT = 1 - \left(\frac{\sum_1^r X = no}{m}\right)$ (7)

where, r is no. of row in the table.
    10.   Calculate a probability for Table (i) base on equation Accumulator_Prob
     Accumulator_Prob $= 1 - \left(\prod_{i=1}^n Accumulator\_Prob\ (node\ i)\right)$ (8).
End.

    Taking the structure M-FBC and evaluating it in Evaluation BNs Fitness are evaluated by using some tests data for testing the selected structures and only high evaluation structure BNs are saved.

5 Artificial Bee Colony (ABC)
    The ABC algorithm is a swarm based, meta-heuristic algorithm based on the model first proposed by [13] on the foraging behavior of honey bee colonies [5]. The ABC consists of three groups of artificial bees: employed foragers, onlookers and scouts. The employed bees comprise the first half of the colony whereas the second half consists of the onlookers. The employed bees are linked to particular food sources. In other words, the number of employed bees is equal to the number of food sources for the hive. The onlookers observe the dance of the employed bees within the hive, to select a food source, whereas scouts search randomly for new food sources. Analogously in the optimization context, the number of food sources (that is the employed or onlooker bees) in ABC algorithm, is equivalent to the number of solutions in the population. Furthermore, the position of a food source signifies the position of a promising solution to the optimization problem, whereas the quality of nectar of a food source represents the fitness cost (quality) of the associated solution [5].
The search cycle of ABC consists of three rules: (i) sending the employed bees to a food source and evaluating the nectar quality; (ii) onlookers choosing the food sources after obtaining information from employed bees and calculating the nectar quality; (iii) determining the scout bees and sending them onto possible food sources. The positions of the food sources are randomly selected by the bees at the initialization stage and their nectar qualities are measured. The employed bees then share the nectar information of the sources with the bees waiting at the dance area within the hive. After sharing this information, every employed bee returns to the food source visited during the previous cycle, since the position of the food source had been memorized and then selects another food source using its visual information in the neighborhood of the present one. At the last stage, an onlooker uses the information obtained from the employed bees at the dance area to select a food source. The probability for the food sources to be selected increases with increase in its nectar quality. Therefore, the employed bee with information of a food source with the highest nectar quality recruits the onlookers to that source. It subsequently chooses another food source in the neighborhood of the one currently in her memory based on visual information (i.e. comparison of food source positions). A new food source is randomly generated by a scout bee to replace the one abandoned by the onlooker bees [5].
The major advantages which *ABC* holds over other optimization algorithms include its:

• Simplicity, flexibility and robustness.
• Use of fewer control parameters compared to many other search techniques.
• Ease of hybridization with other optimization algorithms.
• Ability to handle the objective cost with stochastic nature.
• Ease of implementation with basic mathematical and logical operations [5].

6 The Proposed Hybrid System

The learning process is an important tool. It was used in Artificial Intelligence and belonging to the machine learning used in the classification of data which is Bayesian Networks, which was explained and described in section 3. It used mutual information and condition probability table (CPT) to obtain the most appropriate structure for the proposed system and extracting accurate and correct results.The proposed system consists of several phases of the application of the system.

Phase 1: The data are collected from some Iraqi hospitals. The data were acquired from the Statistics Division at Al-Shaheed Ghazi Hariri Hospital for Specialized Surgery and Al-Kindi Teaching Hospital. Patients' files are taken and information is extracted from it. That has been entered in the database. The data consists of medical history, symptoms, laboratory analysis … etc...

Phase 2: The data taken from hospitals is studied for chosen the medical history and the main symptoms. The diseases have main and secondary symptoms. The patient may recall all symptoms for his diseases but the doctors selected only the main symptoms for obtained the batter diagnosis. Also the system entered only the main symptom when required to train. When entered all symptoms may obtain ambiguity in the results.The first step of filtering is to eliminate the redundancy of data by identifying the patient's records that have the same attributes and unified it's in one record. Thus implied to take the 20 cases satisfactory and ratios repeat. Each case depends on the number of similarity that exists between the symptoms and medical history of all patients. The last attribute in each table presents the number of patients that have same attributes. This is the counting process and downsizing at the same time. Second step of filtering is to delete all the attributes that presents secondary symptoms and medical history. These deleted attributes are not useful because if they exist in the system guide it produces wrong diagnosis. All data tables are stored in repository of feature DB as used in section 4.Note, when added all information's patient, these existed in patient's file. The system guided to produce wrong and ambiguity.

Phase 3: After filtering and extracting for data is stored all tables in repository of feature DB see Table (1) and Table (2). These are represented training datasets. can be used in phase 4.

Phase 4: Completing Filter and Feature Extraction, phase the dataset is stored in the repository Feature DB. All dataset becomes input to M-FBC model. Checked the table in DB is ascertained to be empty and calculate (mutual information , threshold, wieght for each variable, modified CPT for each variable and modified Probabilty for structure).

Phase 5: Take each row from tested dataset and make test for BNs(old) by calculate probability base on equation (8) and compare the result from Accumulator_Prop if Accumulator_Prop is greater than Accept-Accuracy, then plus the counter true otherwise plus the counter false. Thus calculate accuracy for the structure by equation (9).

Phase 6: When Complete all results creating fitness array, choose the highest accuracy from it's under in the fitness array can specify the number of columns in the matrix FSM required and taken all the weights in the specify column and make a new structure represents perfect solution (optimized) and calculate CPTs for each node in the structure and store it. Note that each table in dataset has its own structure can see the structure in figure-1.

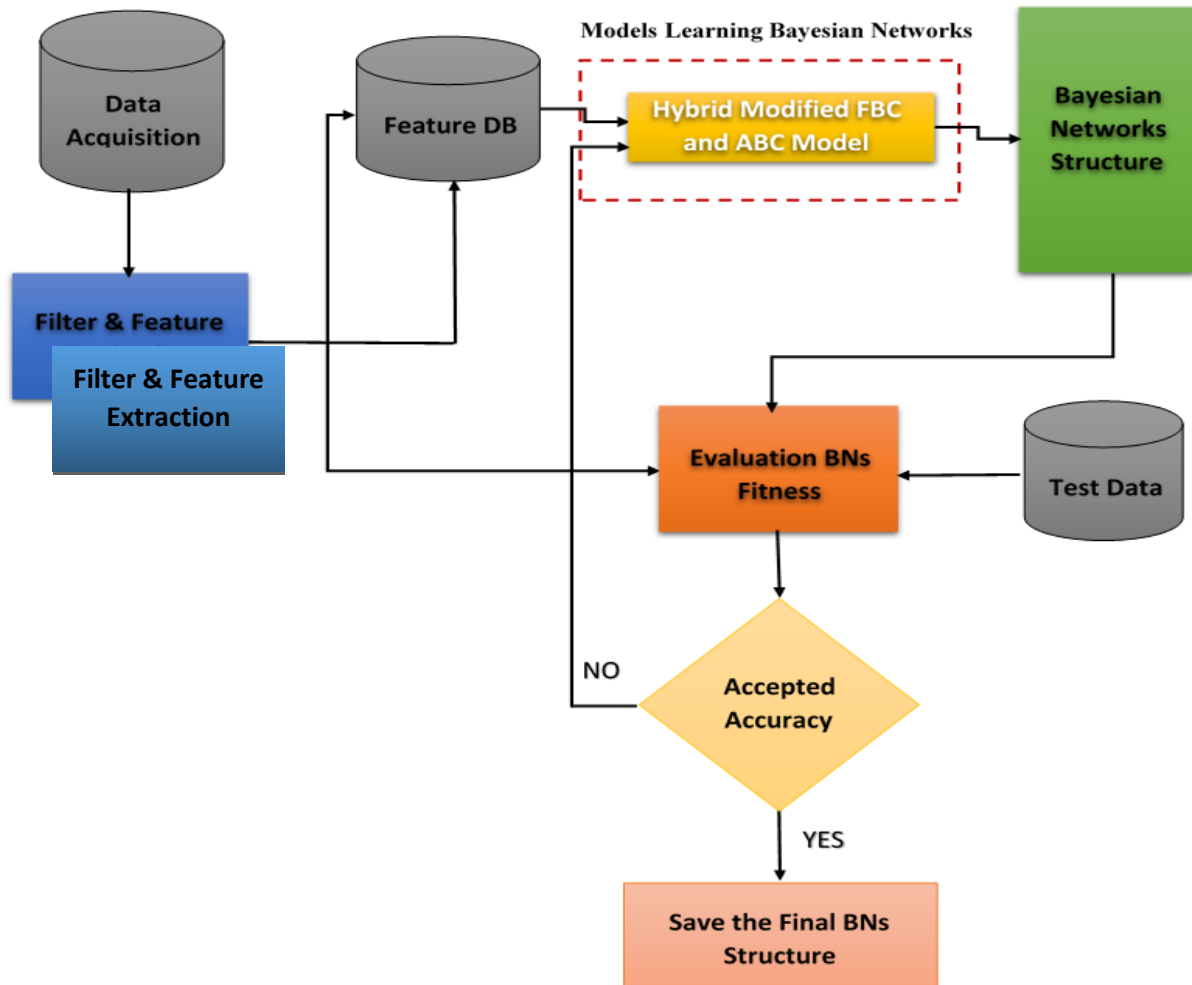Figure-3 represents all phases in block diagram for Hybrid System.

Figure 3- Block Diagram Hybrid System

6.1 Learning Hybrid MFBC – ABC

Calculate weight for each attribute in the table base on equation (5) and take all weights for each variable and make descending order and select the maximum which will be a root and the other are evidence and calculate the CPTs for each node in network structure. All steps in above like M-FBC model.

The structure BNs in M-FBC model takes the training dataset, test dataset and Accept-Accuracy which are represented the input to ABC the other words become the initialize the ABC and problem parameters. Initially ABC is doing to create matrix two dimensions call Food Source Memory (FSM) has size *SN* and *N*. The *SN* represented the number of rows in matrix FSM and *SN* are no. of node from BNs$_{(old)}$. *N* represented the number of columns in matrix FSM and *N* are the number of solutions are supposed to find the best suitable structure. And the ABC is doing to create array one dimension call Fitness has size *N* are represented the number of solutions (No. of bees) are supposed to find the best accuracy by selected the maximum accuracy for Fitness array. After created the FSM matrix and Fitness array take BNs$_{(old)}$ and the development of weight for all node in to FSM$_{[SN,0]}$ after make descending order for all weights. Thus can calculate the fitness first column in FSM, which takes each row from test dataset and make tested for BNs$_{(old)}$ by calculate probability base on equation (8) and compare the result from Accumulator_Prop if Accumulator_Prop is greater than Accept-Accuracy then plus the counter true otherwise plus the counter false. Thus calculate accuracy for the structure by equation (9)

$$Acc(\%) = \left(\frac{True}{True + False}\right) \times 100, \qquad (9)$$

The result is in the first position of the Fitness.
In the next phase the first columns of the matrix FSM is taken so as to generate a neighboring depending on the equation (10).

$$W_{new\,j}(k) = LW_{old\,(j-1)}(k) \pm \left(UW_{old\,(j-1)}(k) + LW_{old\,(j-1)}(k)\right) \times r\,, \qquad (10).$$

Note that $r\sim(0,1)$ generates a uniform random number between 0 and 1.

When new weights are generated new column, a new generated inserted weight and sorted as descending, in matrix FSM calculate a CPTs for each node in BN and recalculate the fitness to generate weight like the first phase in the MFBC-ABC model and put the result in array Fitness.

When Complete of all results creating fitness array, choose the highest accuracy from it's under in the fitness array can specify the number of columns in the matrix FSM required and taken all the weights in the specify column and make a new structure represents perfect solution (optimized) and calculate CPTs for each node in the structure and store it.

Note the model is divided into three phases, in the first phase take the network structure of the M-FBC model, which is one of perfect and the introduction of this structure on the MFBC – ABC model until finding more structure convenient any gives greater accuracy. In the second phase is to create more structures and calculate the accuracy of each structure. In the third phase structure were taken that have the highest accuracy and storage and return of this structure is the optimal among all structures. Note that each table in dataset has its own structure can see the structure in figure-1.

6.2 Algorithm *MFBC – ABC.*

> **Input:** Table (i), Test-Table (i), A Full BNs ***B(i)***, where, (i =1,…, M), M: no. of table, N: no.of columns (no. of Bees), Accept-Accu *.// The input is represented in ABC (**Initialize the ABC and problem parameters**)//*
>
> **Output:** A full BNs ***B(i)*** *new*, Accuracy (i). // *The output is represented in ABC (stop condition are repeated until a stop criterion is met. This is originally determined by the Maximum Cycle Number (MCN) value*//
>
> **Begin:**

For i= 0 *to* M

> **Step 1:** *Initialize* *FSM* [SN, N], *Fitness* [N], where, *SN* = no. of node from full BNs ***B(i)***$_{old}$, *N* = no. of columns
> - Make descending order all Weight (***W***).
> - *FSM*[SN,0] = *Weight Nodes for* ***B(i)*** *old*
> - Calculate *Fitness*
>
> *For* each row in Test-Table (i) *do*
> Tested for ***B(i)*** *old* by calculate Probability base on equation (8):
> *If Accum_Prob > 0.8 then*
>   *True= True +1*
> *Else*
>   *False = False +1*
> *End if*
> *End for*
> - Calculate Accuracy base on equation (9) and put the result in *Fitness[j]*.
>
> **Step 2:** *Employed Bee Phase*
> *For j = 1 to N do*
> - *For* each row ***k*** in the columns make generated ***W***$_{new}$ for each neighboring weight base on equation (10).
>   Note that $r \sim (0, 1)$ generates a uniform random number between 0 and 1.
> *End for*
> 11.   Sorted (descending order) for each ***j*** in *FSM*.
> 12.   Create ***B(i)*** *new* for each ***j*** in *FSM*.
> 13.   Calculate CPTs for each node in ***B(i)*** *new* by equation
> *If* node = positive *then* base on equation (6)
> *Else* base on equation (7)
> 14.   Calculate *Fitness (j)*
> *For* each row in Test-Table (i),

Tested for $B(i)_{new}$ by calculate Probability base on equation (8):
If $Accum\_Prob > 0.8$ **then**
  $True = True + 1$
Else
  $False = False + 1$
 End if
 End for
- Calculate Accuracy base on equation (9) and put the result in *Fitness $_{[j]}$*.
End for

Step 3: *Memorizing the best food source*
   -   Selected the Maximum Accuracy from *Fitness (j)* and determined the position of the Maximum Accuracy.
   -   Select the no. of Column (the position of the Maximum Accuracy) from *FSM*.
   -   Create $B(i)_{new}$ from column (the position of the Maximum Accuracy) from *FSM*.
   -   Calculate CPTs for each node in $B(i)_{new}$ by equation:
    If node = positive **then** base on equation (6)
    Else base on equation (7)
   -   Save $B(i)_{new}$.
    End **For**
   End.

   The model depends on two parameters. The first parameter, N represents number of bees, where has raised a number of experiments to determine the number of bees appropriate in this model. Where has experience almost between the 50 to 200 bees. It model has been reached to the best results through the selection 100 bees. The second parameter, determines the value of (Accept-Accu), the number of experiments are to choose the accepted accuracy in order to access to the best results where experience (0.7, 0.8, 0.9). Through the experience has been selected 0.8 best apical approximate and medium to get the best acceptable outcome.

7 Experimental Results
   The proposed system deals with two types of diseases heart and nervous system for experiment on the model. Each table is stored testing disease cases. This was configured 10 tables, each table represents a particular disease, five tables represent heart disease and the other five represent Diseases of the nervous system. Each table consists of a number of different attributes. Each table consists of a number of different attributes from the other table. A table-1 and table-2 explain all the training dataset in details.
   The system has been tested by inserting symptoms and medical history of each patient. The entered data on the structure and calculates the probability of occurrence of the disease for each patient. This results show the likelihood of disease for each patient, and this is chosen top 2 of the possibilities of these diseases can make sure through the doctor examines exact for patient finds possibilities mentioned are correct and this is achieved the desired results, the performance of the system is also in the table-3, table-4, 22 patients been taken every heart disease, as well as 20 patients been taken every disease of the nervous system.
The performance of the MFBC-ABC shown in Table (3) and Table (4) for each disease. So that the rate of accuracy in the model of M-FBC is approximately (93%) in heart diseases and diseases of nervous system is approximately (98%) [6]. Thus when calculating the percentage accuracy of the structure extracted from the MFBC-ABC model is approximately (100%) in heart diseases and diseases of nervous system is approximately (99%) be higher than from M-FBC model. the result shows the rate of accuracy for the model MFBC-ABC is the best average results.

Table 1- Detail Training Dataset for Heart Diseases

Diseases Code	Diseases Name	Records Number	Attributes Number	Patients Number
I10	High Blood Pressure	20	19	200
I20	Angina pectoris	20	26	330
I21	Myocardial Infarction	20	24	450
I48	Atrial Fibrillation	20	14	390
I50	Heart Failure	20	28	565

Table 2- Detail Training Dataset for Nervous System Diseases

Diseases Code	Diseases Name	Records Number	Attributes Number	Patients Number
G40	Epilepsy	20	15	268
G55	Nerve Root and Plexus Compressions	20	11	238
G91	Hydrocephalus	20	14	188
G94	Head injury	20	18	190
G95	Other diseases of spinal cord	20	14	228

Where,

Disease Code: It represents the statistical number of disease.

Diseases Name: It represents the name of the disease.

Records Number: It represents the number of cases used in one disease.

                                  Attributes Number: It represents the medical history and symptoms of the cases used.

Patients Number: It represents the number of patients who have the disease.

Table 3- Performance Hybrid System for Heart Diseases

Disease Code	Diseases Name	Number of Patients used in testing	Accuracy (%)	
			M-FBC	*MFBC-ABC*
I10	High Blood Pressure	22	100%	100%
I20	Angina pectoris	22	81.8%	100%
I21	Myocardial Infarction	22	95.45%	100%
I48	Atrial Fibrillation	22	100%	100%
I50	Heart Failure	22	86.36%	100%
Average			93%	100%

Table 4- Performance Hybrid System for Nervous System Diseases

No. Of Disease	Name of Diseases	Number of Patients used in testing	Accuracy(%)	
			M-FBC	*MFBC-ABC*
G40	Epilepsy	20	100%	95%
G55	Nerve Root and Plexus Compressions	20	100%	100%
G91	Hydrocephalus	20	100%	100%
G94	Head injury	20	90%	100%
G95	Other diseases of spinal cord	20	100%	100%
Average			98%	99%

The results show that many data large number of patients can be taken get with good result see figure-4. This conclusion is based through access to the data in the Internet where all research is based on a particular disease either in this research is based on 10 diseases and the diseases is specific for heart and nervous system diseases. And not only to mention the existence or non-existence, but also the percentage probability of infection for heart disease and diseases of the nervous system.
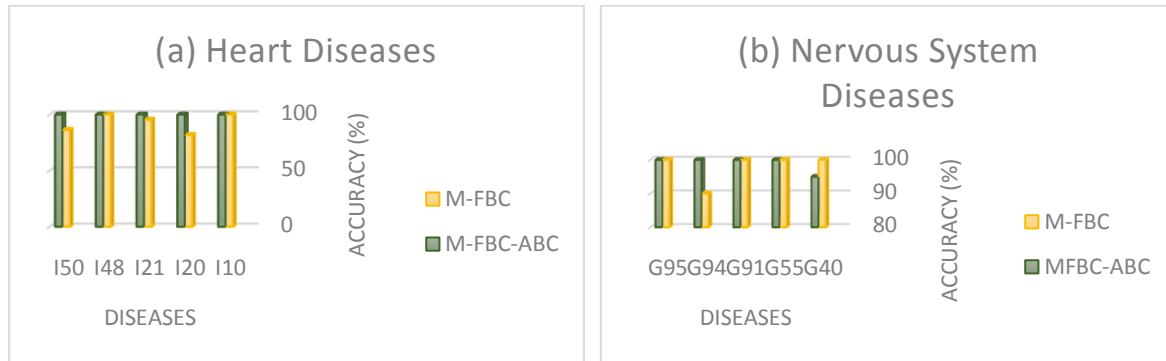

Figure 4- (a) Performance Hybrid System for Heart Diseases, (b) Performance Hybrid System for Nervous System Diseases

8 Conclusion

In this paper, a hybrid approach based on Learning Bayesian Network Structures, is used where the best chosen structure through the hybrid algorithm is found. The hybrid MFBC-ABC model combines two algorithms: the Modified Full Bayesian Classifier algorithm (M-FBC) and Artificial Bees Colony algorithm (ABC). M-FBC is generated ideal structure for each disease for access to be diagnosed, but we want to get into best structure by using ABC algorithm to obtain a more accurate diagnosis for an optimal structure. ABC has been selected because it is developed and based on nature-inspired ideas and it introduces several optimization algorithms. Most of swarm intelligent algorithms are meta-heuristic based search techniques and generally referred to as multipurpose optimization algorithms. The system (MFBC-ABC) is tested by the doctors in Iraqi hospitals. The accuracy for heart diseases is approximately (100%) and for nervous system diseases it is approximately (99%).

The MFBC-ABC is better than M-FBC, because that is the best for each disease as it generates a number of structures, by changing the values of the weights that have been calculated from the model (M-FBC). Re-arranged so that they appear different results. Choosing the best results are achieved on the best structure that is tested to get a better diagnosis of the disease. When tested (M-FBC) the accuracy for heart diseases is approximately (93%) and for nervous system diseases it is approximately (98%).

References

1. Flores M., Nicholson E., Brunskill A., Kevin B., Steven Mascaro, **2011**, "Incorporating expert knowledge when learning Bayesian network structure: A medical case study", *Elsevier B.V. journal for Artificial Intelligence in Medicine* V.53, pp:181– 204.
2. Parviainen P., **2012**, "Algorithms for Exact Structure Discovery in Bayesian Networks", PhD Thesis, Department of Computer Science, P.O., University of Helsinki, Finland, Pekka Parviainen.
3. Kosarzycki B., **2011**, "Design and Implementation of a Computer Systems Diagnosis Tool Using Bayesian Networks", Master's thesis, Poznan University of Technology, Faculty of Computing Science, Institute of Computing Science, Poznan.
4. Karaboga D., Ozturk C., **2011**, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm", *Elsevier B.V. journal for Applied Soft Computing*, pp:652–657.
5. LA'aro A., Tajudin A., Azmi M. and Awadallah M., 20th January **2013** "Artificial Bee Colony Algorithm, ITS Variants and Applications: A Survey", JATIT & LLS *Journal of Theoretical and Applied Information Technology*, 47(2).

6. Alobaid,T. A. Mahmood,T. N. **2014**,"Modified Full Bayesian Networks Classifiers for Medical Diagnosis*","2013 International Conference on Advanced Computer Science Applications and Technologies (ACSAT 2013) ,Sarawak, Malaysia.*", PID3061433, IEEE.

7. Reiz Beáta, Csató Lehel, **2009** ,"Bayesian Network Classifier for Medical Data Analysis", Int. J. of Computers, Communications & Control, CCC Publications, IV(1), pp:65-72.

8. Hongyan Li，Zhong Wu，Cheng Li, **2009** "Research on Risk Decision Making of Overbooking Based on the Bayes Theory", *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, IEEE.

9. Silvia, A. M. de Campos Luis, Juan,M. F. Rodriguez, S. Rodriguez, M. J. Salcedo,L. J. **2004**, "A Comparison of Learning Algorithms for Bayesian Networks: A Case study based on data from an emergency medical service", Artificial Intelligence in Medicine 30, Elsevier B.V., pp:215–232.

10. Stefan, C. Lionel, J. **2011**, "Breast Cancer Diagnostics with Bayesian Networks", Conrady Applied Science, LLC - Bayesia's North American Partner for Sales and Consulting.

11. Gadewadikar, J. Kuljaca, O., Agyepong, K. Sarigul, E. Zheng, Y. and Zhang, P. **2010** ,"Exploring Bayesian Networks for Medical Decision Support in Breast Cancer Detection", African Journal of Mathematics and Computer Science Research Vol. 3(10), pp:225-231, Academic Journals.

12. Reiz, B., Csató, L. **2009** ,"Bayesian Network Classifier for Medical Data Analysis", Int. J. of Computers, Communications & Control, CCC Publications, IV(1), pp:65-72.

13. Tereshko V. and Loengarov A., **2005**, "Collective Decision Making in Honey-Bee Foraging Dynamics", Computing and Information Systems, 9(3), pp:1.