

## БАЙЕСОВ МРЕЖОВ МОДЕЛ В ПОМОЩ НА ДИФЕРЕНЦИАЛНАТА ДИАГНОСТИКА НА САРКОИДОЗАТА

Людмила Димитрова<sup>1</sup>, Олга Голубева<sup>2</sup>

<sup>1</sup> Университет „Проф. д-р Асен Златаров“, Бургас  
e-mail: lyudim@gmail.com

<sup>2</sup> ГБУЗГП №17, Санкт Петербург  
e-mail: golubeva@gmail.com

**Резюме:** Предложен е модел – байесова мрежа за оценка на вероятността на диагнозата „саркоидоза“ в сравнение със заболяванията със сходни симптоми. Параметрите на модела са оценени на базата на обобщени статистически данни за честотата на отделните показатели за диагностициране на заболяването. Моделът може да се използва като помощно средство в лекарската практика.

**Ключови думи:** Байесови мрежи, Медицинска диагностика, Саркоидоза.

### 1 Въведение

Байесовите мрежи са широко използван понастоящем инструмент на изкуствения интелект. Те обединяват теорията на графите и теорията на вероятностите и предоставят ефективен, компактен и интуитивен метод за представяне на знания, характеризирани с висока степен на неопределеност. Настоящата работа илюстрира използването на байесова мрежа в диагностиката на саркоидозата.

Саркоидозата представлява грануломатозно заболяване, което засяга хора по целия свят [1, 2]. Саркоидозата варира значително в различните популации [2, 3]. Характеризира се с неизвестна етиология и голямо разнообразие от клинични и радиологични прояви [4]. Заболяването включва необичайни струпвания на възпалителни клетки (грануломи), които могат да образуват възли в различни органи, като най-често грануломите в белите дробове или свързаните с тях лимфни възли [5–7]. Други често участващи органи и системи включват на кожата, очите, черния дроб, сърцето, нервната, опорно-двигателната и ендокринната система [6, 7]. Тези грану-

ломи не са уникални за саркоидозата и се наблюдават при други заболявания като туберкулоза, гъбични заболявания, берилиоза, карциноми и лимфоми [4, 7].

Саркоидозата засяга средно 20 на 100 000 души от населението и се влияе от пол, раса и възраст [8–12]. Например, болестта се проявява различно при бели в сравнение с чернокожи индивиди и при жени в сравнение с мъже [9–11]. Въпреки че етническите различия не са напълно добре характеризирани [12], според мащабно изследване включващо 736 пациенти със саркоидоза в срок от 6 месеца на диагнозата в Съединените щати, расата е основният определящ фактор за заболяването, като засегнатите органи се различават в зависимост от расата, пола и възрастта [8].

Въпреки че през последните няколко години се наблюдава значителен напредък в разбирането на саркоидозата, като например по-добро характеризиране на етиопатогенезата, диагностиката и лечението на това заболяване, много въпроси остават да бъдат изяснени: какво води до заболяването и какви са ключовите гени които влияят на склонността към заболяването и формират индивидуалните клинични резултати [13]. Диагностиката на саркоидозата е много трудна, защото симптомите и индикаторите на заболяването не са специфични. Само пълен преглед, включващ изследване на цялостната симптоматика позволява да се постави правилна диагноза. Ето защо един от най-важните въпроси касае правилното разпознаване на саркоидозата, което е трудно поради забележителната прилика между клиничните и хистологични особености на саркоидозата и тези характерни за други болести като *туберкулоза и лимфом на Ходжкин* [7, 14–16]. Например, при саркоидозата често са засегнати паратрахеалните и хилусни лимфни възли, но СТ скенера често показва разширяването на лимфните възли в по-рано неподозирани места като предната медиастина. Лимфомът, от друга страна е системно заболяване, при което също често са засегнати медиастиналните лимфни възли, и е лесно да се бърка с саркоидозата [4]. Саркоидозата и туберкулозата имат толкова много клинични прилики, че са считани от някои автори за казуално свързани и дори за две крайности на един и същ процес на болестта [16]. Разграничаването на саркоидозата от тези две заболявания е изключително важно, тъй като различните заболявания имат различна прогноза и се нуждаят от напълно различен подход към лечението.

## **2 Байесови мрежи и приложението им в медицинската диагностика**

Байесовите мрежи (БМ) [17–25] се отнасят към категорията на вероятностните графични модели. БМ представлява направлен ацикличен граф. Възлите на графа са вероятностни променливи, а дъгите описват връзките между тях. Параметри на БМ с дискретни променливи са таблиците на условните вероятности. За входните възли те представляват априорните безусловни вероятности на възможните състояния на дадения възел, а за възлите с родители – условните вероятности на дадено състояние на възела при всички възможни комбинации на състоянията на родителските му

възли. Основната идея на БМ е да използва вероятностния подход за решаване на множеството проблеми в реалния свят, свързани с неопределеност. Това става чрез определянето на апостериорните маргинални разпределения на вероятностите при постъпване на наблюдения за някои от възлите в мрежата. В зависимост от вида на наблюдаваните възли двата крайни варианта на логическия извод с БМ са: прогнозиране – определение на вероятността на дадено събитие при наблюдаване на причини, и диагностициране – определяне на причината за наблюдаваните следствия. В медицинската диагностика се решава най-вече втората задача.

Създадени са редица методи за точен и приблизителен логически извод с байесовите мрежи [17–25].

Байесовата мрежа може да бъде построена въз основа на експертни оценки или на базата на статистически данни. Експертната информация може да се използва както за установяване на връзките между случайните променливи, т.е. структурата на мрежата, така и за оценка на условните вероятности. Както при всички модели, построени на принципите на изкуствения интелект, съществува възможност за автоматично обучение на структурата и параметрите на модела при постъпване на необходимите данни [26–30].

Разработени са много приложения на байесовите мрежи в медицината, в частност за медицинска диагностика на патологията на лимфните възли [31], нервно-мускулни патологии [32], болести на сърцето [33], рака на хранопровода [34], пневмонията в отделенията за интензивно лечение [35], в традиционната китайска медицина [36], сърдечните болести въз основа на данните от ехокардиографията [37], на лимфомите [38], анемията [39], зъбобола [40], белодробната емболия [41], психиатричните болести [42].

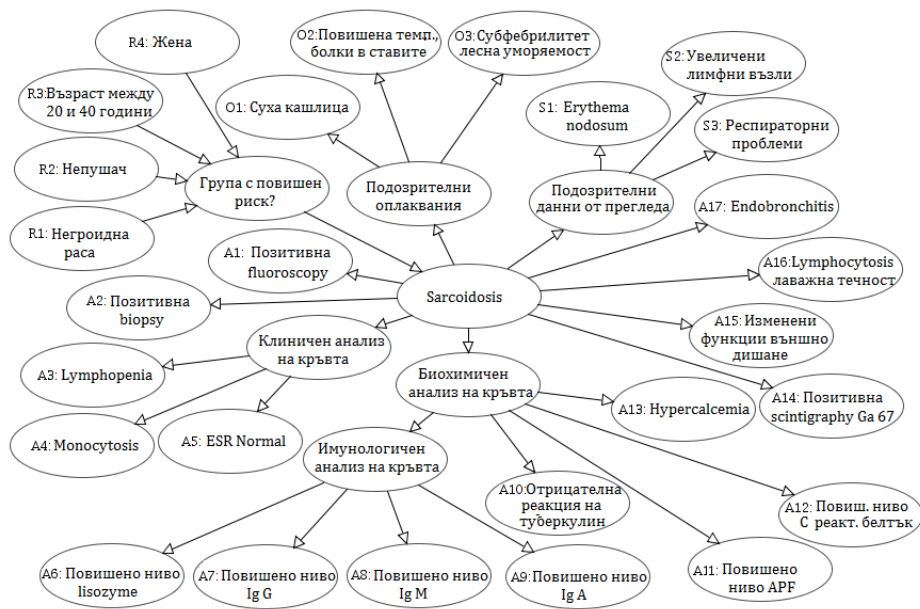
Байесовите мрежи са подходящо помощно средство при диференцирането на саркоидозата от туберкулозата и лимфома на Ходжкин, защото позволяват да се обединят и отчетат всички показатели за диагностиката. В настоящата работа са анализирани множество публикации с проучвания за пациентите и наблюдения от практиката относно честотата на възникване и относителната важност на всеки фактор, който влияе върху диагностицирането на разглежданата група болести. Тези данни от специализираната литература са използвани за задаване на тежестта на дадения фактор при параметризирането на мрежовия модел.

### **3 Описание на модела**

Разработката и логическите изводи с модела са изпълнени с помощта на софтуерния пакет AgenaRisk [43]. Структурата на модела е показана на Фиг. 1.

Предложеният БМ модел включва целевия възел *Sarcoidosis* и 27 входни възли, обхващащи основните диагностични показатели. Те са разделени на 5 групи: рискови фактори (R1–R4), оплаквания (O1–O3), симптоми, открити при медицинския преглед (S1–S3) и данни от медицинските анализи (A1–A17). Всички възли в мрежата са

логически (*Boolean*) със състояния *Yes/No*. Смисълът на възлите е формулиран така, че състоянието *Yes* води до увеличение на вероятността за изследваната диагноза. На дадения етап целта е да се диференцира (т.е. разграничи) саркоидозата от алтернативните възможни диагнози, които се задават със състоянието *No* на целевия възел *Sarcoidosis*.



Фигура 1. Структура на модела

Таблиците на условните вероятности (ТУВ) на основната част от възлите са зададени по стандартния за класификаторите начин. В Таблица 1 е показана параметризацията на възела *A1: Позитивна флуороскопия*. Тя е получена въз основа на данни от специализираната литература, въз основа на които при 92% от болните от саркоидоза се наблюдава двустранно увеличение на вътрешногърдните лимфни възли (състояние *Yes*). При алтернативните заболявания с подобни симптоми такава рентгенографична картина се наблюдава при 20% от пациентите.

	Sarcoidosis = "Yes"	Sarcoidosis = "No"
Позитивна флуороскопия = "yes"	0.92	0.2
Позитивна флуороскопия = "no"	0.08	0.8

Таблица 1. Параметризация на възела *Позитивна флуороскопия*

Таблицы 2–4 показват данните, използвани за параметризацията на останалите входни възли въз основа на информацията в [1–16]. Тъй като целта на модела е да потвърди или отхвърли диагнозата саркоидоза, са подбрани показатели, които се наблюдават в по-голяма степен при болните от саркоидоза. По някои от показателите съществува значителен разброс в статистическите данни, поради което са използвани осреднени значения. За по-голяма гъвкавост на модела са добавени междинните възли *Клиничен анализ на кръвта*, *Имунологичен анализ на кръвта* и *Биохимичен анализ на кръвта* с априорни разпределения на вероятностите на състоянието “Yes” (0.7, 0.4), *Подозрителни оплаквания* (0.2, 0.05) и *Подозрителни данни от прегледа* (0.2, 0.02). Първата цифра съответства на наличието на диагнозата саркоидоза, а втората – на отсъствието ѝ. Ниските значения за вероятността на състояние “Yes” на последните два възела са продиктувани от факта, че оплакванията и рутинният преглед са слабоинформативни за поставяне на разглежданата диагноза.

Възелът *Група с повишен риск* отчита повишената вероятност от заболяване при хора от негроидната раса (възел с име *negro*), жени (*woman*), непушачи (*nosmoker*) и хора на възраст от 20 до 40 години (*age20\_40*). Таблицата на условните вероятности на този възел е зададена чрез функцията *noisyor*[17-19]:

$$\text{noisyor}(\text{woman}, 0.1, \text{age20\_40}, 0.1, \text{nosmoker}, 0.1, \text{negro}, 0.15, 0.3) \quad (1)$$

Априорното разпределение на вероятностите за състоянието “Yes” за възела *Sarcoïdosis* е зададено (0.1, 0.067) – за *Група с повишен риск* = “Yes” и “No”, съответно. Цифрата 0.067 е избрана на базата на приблизителните статистически данни, че на 15 пациента от разглежданата група заболявания се пада 1 със саркоидоза.

	Саркоидоза	Алтернативни заболявания
<i>O1: Суха кашлица</i>	50%	20%
<i>O2: Повишена температура, болки в ставите</i>	20%	5%
<i>O3: Субфебрилитет, лесна уморямост</i>	50%	40%

Таблица 2. Данни за ТУВ на възлите, описващи оплакванията на пациентите

	Саркоидоза	Алтернативни заболявания
<i>S1: Erythema nodosum</i>	20%	2%
<i>S2: Увеличени лимфни възли</i>	20%	15%
<i>S3: Респираторни проблеми</i>	30%	20%

Таблица 3. Данни за ТУВ на възлите - симптоми, открити при прегледа

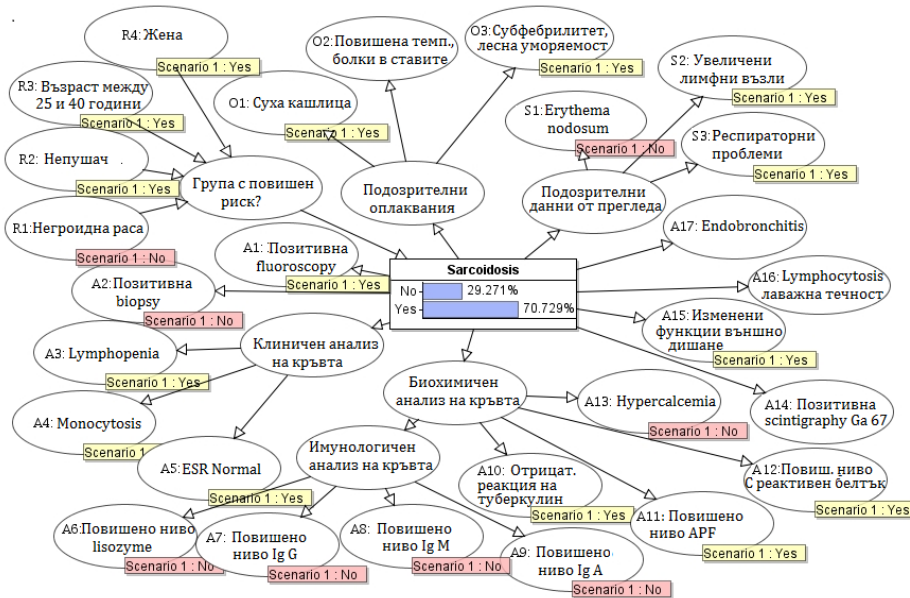
	Саркоидоза	Алтернативни заболявания
<i>A1: Позитивна флуороскопия</i>	92%	20%
<i>A2: Позитивна биопсия</i>	80%	15%
<i>A3: Lymphopenia</i>	35%	15%
<i>A4: Monocytosis</i>	35%	25%
<i>A5: ASR Normal</i>	60%	30%
<i>A6: Повишено ниво lysozyme</i>	25%	10%
<i>A7: Повишено ниво Ig G</i>	50%	40%
<i>A8: Повишено ниво Ig M</i>	50%	30%
<i>A9: Повишено ниво Ig A</i>	60%	40%
<i>A10: Отрицателна реакция на туберкулин</i>	70%	15%
<i>A11: Повишено ниво APF</i>	65%	5%
<i>A12: Повишено ниво C реактивен белтък</i>	30%	10%
<i>A13: Hypercalcemia</i>	15%	3%
<i>A14: Позитивна сцинтиграфия Ga 67</i>	80%	10%
<i>A15: Изменени функции на външно дишане</i>	30%	15%
<i>A16: Lymphocytosis на лаважна течност</i>	70%	20%
<i>A17: Endobronchitis</i>	45%	25%

Таблица 4. Данни за ТУВ на възлите – резултати от медицинските анализи

На Фиг. 2 е показан резултатът от логическия извод с модела при въведени наблюдения, съответстващи на случай от практиката на един от авторите: жена на възраст 38 години, непущачка, оплаква се от слабост, уморяемост, субфебрилитет.

При прегледа се установява задъхване при физическо натоварване и увеличение на лимфните възли. Данните от рентгена са положителни, бронхоскопията – отрицателна. Резултатите от клиничния анализ на кръвта са положителни, от имунологичния – отрицателни. Пациентката има отрицателна реакция на туберкулин, повишени нива на APF и C-реактивен белтък, няма хиперкалциемия.

Моделът дава около 70% вероятност за наличие на заболяването, което предполага извършване на допълнителни анализи за потвърждаване на диагнозата.



Фигура 2. Пример за използване на модела за сценарий от лекарската практика

#### 4 Изводи

Предложен е модел, байесова мрежа, който може да се използва при диференциалната диагностика на саркоидозата спрямо заболявания със сходни симптоми. При задаване на параметрите на модела са използвани обобщени статистически данни, което предполага достоверно предсказване за вероятността за изследваното заболяване. Подобряване на точността на количествената оценка може да се постигне чрез разширяване на модела с допълнителни показатели по диагностиката и обучението на структурата и параметрите му чрез използване на база данни с индивидуални медицински записи на пациентите.

#### Литература

- [1] Hunninghake G.W., Costabel U., Ando M., Baughman R., Cordier J.F., Du B.R., Eklund A., Kitaichi M., Lynch J., Rizzato G. , "ATS/ERS/WASOG statement on sarcoidosis". Sarcoidosis Vasc Diffuse Lung Dis,16, pp.149-173, 1999

- [2] Siltzbach L.E., James D.G., Neville E., Turiaf J., Battesti J.P., Sharma O.P., Hosoda Y., Mikami R., Odaka M. "*Course and prognosis of sarcoidosis around the world*", Am J Med 1974;57:847–852
- [3] Izumi T. "*Population differences in clinical features and prognosis of sarcoidosis throughout the world*" Sarcoidosis, 9:S105–S118, 1992
- [4] Mehrian P., Ebrahimzadeh, S., "*Differentiation between sarcoidosis and Hodgkin's lymphoma based on mediastinal lymph node involvement pattern: Evaluation using spiral CT scan*", Pol J Radiol., 78(3), pp. 15–20, 2013
- [5] King, T.E., Jr. *The Merck Manual Home Edition*. Merck Sharp & Dohme.
- [6] Fauci, A., Kasper, D., Hauser, S., Jameson, J., Loscalzo, J., "*Harrison's Principles of Internal Medicine*", 18 ed.. New York: McGraw-Hill, 2011
- [7] Vagal, A. S., "*Radiological manifestations of sarcoidosis.*", Clin Dermatol, 25(3): pp. 312-325, 2007
- [8] Baughman, R. P., Teirstein A.S., Judson M.A., Rossman M.D., Yeager H. Jr, Bresnitz E.A., DePalo L., Hunninghake G., Iannuzzi M.C., Johns C.J., McLennan G., Moller D.R., Newman L.S., Rabin D.L., Rose C., Rybicki B., Weinberger S.E., Terrin M.L., Knatterud G.L., Cherniak R., "*Clinical characteristics of patients in a case control study of sarcoidosis.*" Am J Respir Crit Care Med, 164(10 Pt 1), pp.1885-1889., 2001
- [9] Maycock RL, Bertrand P, Morrison CE. *Manifestations of sarcoidosis*. Am J Med 1963;35:67–89.
- [10] Longcope W.T., Frieman D.G., "*A study of sarcoidosis based on a combined investigation of 160 cases including 30 autopsies*", Medicine 1952, 31, pp.1–132.
- [11] Hunninghake G.W., Gilbert S., Pueringer R., Dayton C., Floerchinger C., Helmers R., Merchant R., Wilson J., Galvin J., Schwartz D., "*Outcome of the treatment for sarcoidosis*", Am J Respir Crit Care Med, 149, pp.893–898, 1994
- [12] Honeybourne D., "*Ethnic differences in the clinical features of sarcoidosis in south-east London*". Br J Dis Chest, 74, pp.63–69, 1980
- [13] Baughman, R. P., Culver D.A., Judson M.A., "*A concise review of pulmonary sarcoidosis*" Am J Respir Crit Care Med, 183(5), pp. 573-581, 2011
- [14] Gupta, D., "*Tuberculosis and sarcoidosis: The continuing enigma*", Lung India, 26(1), pp. 1–2, 2009
- [15] Perez, R. L., Rivera-Marrero C.A., Roman J. "*Pulmonary granulomatous inflammation: From sarcoidosis to tuberculosis*", Semin Respir Infect 18(1),pp. 23-32, 2003



- [16] Gupta, D., Agarwal R., Aggarwal A.N., Jindal S.K., "Sarcoidosis and tuberculosis: the same disease with different manifestations or similar manifestations of different disorders", *Curr Opin Pulm Med* 18 (5), pp.506-516, 2012
- [17] Darwiche, A., "Modeling and Reasoning with Bayesian Networks", Cambridge University Press, 2009
- [18] Koller D., Friedman N., "Probabilistic Graphical Models. Principles and Techniques", MIT Press, 2009.
- [19] Lauritzen, S. L. , "Graphical Models", Oxford University Press, Oxford, UK, 1996
- [20] Jensen, F. V., "Bayesian Networks and Decision Graphs", Springer, 2001.
- [21] Heckerman, D., "A Tutorial on Learning with Bayesian Networks", *Innovations in Bayesian Networks*, Studies in Computational Intelligence, 156, pp 33-82, 2008
- [22] Kjaerulff, U.B., Madsen, A.L., "An Introduction to Bayesian Network and Influence Diagram", 2005 [Online] <http://people.cs.aau.dk/~uk/papers/pgm-book-I-05.pdf>
- [23] Kjaerulff, U. B. Madsen, A. L., "Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis", *Information Science and Statistics*, 22, 2013
- [24] Korb, K. B., Nicholson , A. E. "Bayesian Artificial Intelligence", Seco Chapman & Hall/CRC Computer Science & Data Analysis, 2010
- [25] Kelly, D. L., Smith, C. L., "Bayesian inference in probabilistic risk assessment—The current state of the art" *Reliability Engineering & System Safety*, 94 (2), pp.628-643, 2009
- [26] Heckerman, D., Geiger, D., Chickering, D. M., "Learning Bayesian networks: The combination of knowledge and statistical data", *Machine Learning* , 20 (3), pp 197-243, 1995
- [27] Heckerman, D., "A Tutorial on Learning with Bayesian Networks", *Innovations in Bayesian Networks*, Studies in Computational Intelligence, 156, pp 33-82, 2008
- [28] Lauritzen, S.L., "The EM algorithm for graphical association models with missing data", *Computational statistics and data analysis*, 19(2), pp.191-201, 1995
- [29] Feelders, A., van der Gaag, L. C., "Learning Bayesian network parameters under order constraints", *International Journal of Approximate Reasoning*, 42 (1-2), pp. 37-53, 2006
- [30] Daly, R., Shen , Q., Aitken, S., "Learning Bayesian networks: approaches and issues", *The Knowledge Engineering Review*, 26(2), pp. 99–157, Cambridge University Press, 2011

- [31] Heckerman D., “*Probabilistic similarity networks*”, Networks, 20, pp. 607–636, 1990
- [32] Xiang, Y., Pant, B., Eiseny, A., Beddoes, M.P., Poolez, D. “*Multiply Sectioned Bayesian Networks For Neuromuscular Diagnosis*”, Artificial Intelligence in Medicine, 5(4), pp.293-314, 1993
- [33] Long, W., “*Temporal Reasoning for Diagnosis in a Causal Probabilistic Knowledge Base*” Artificial Intelligence in Medicine, 8(3), pp.193-215, 1996
- [34] van der Gaag, L.C., Renooij, S., Witteman, C.L.M., Aleman, B.M.P., Taal, B.G., “*Probabilities for a probabilistic network: a case study in oesophageal cancer*”, Artificial Intelligence in Medicine, 25 (2), pp. 123–148, 2002
- [35] Lucas P.J., de Bruijn N.C., Schurink K., Hoepelman A., “*A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU*”, Artif Intell Med. , 19(3), pp.251-79, 2000
- [36] Wang, X., Qu, H. , Liu, P., Cheng, Y., “*A self-learning expert system for diagnosis in traditional Chinese medicine*”, Expert Systems with Applications, 26 (4), pp. 557-566, 2004,
- [37] Diez F.J., Miraa J., Iturraldeb E., Zubillagac S., “*DIAYAL, a Bayesian expert system for echocardiography*”, Artificial Intelligence in Medicine, 10(1), pp. 59–73, 1997.
- [38] Lucas P.J.F, Boot H., Taal B.G., “*Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma*”, Methods Inform Med. , 37(3), pp. 206–219, 1998
- [39] Wiegerincka W.A.J.J., Kappena H.J., ter Braakb E.W.M.T., ter Burgc W.J.P.P., Nijmana M.J., O Y.L., Neijtb J.P., “*Approximate inference for medical diagnosis* , Pattern Recognition Letters, 20(11–13), pp. 1231–1239, 1999.
- [40] Chattopadhyay S., Davis R.M., Menezes D.D., Singh G., Acharya R.U., Tamura T., “*Application of Bayesian classifier for the diagnosis of dental pain*”, Journal of Medical Systems, 36(3), pp.1425-1439, 2012
- [41] Luciani D., Marchesi M., Bertolini G., “*The role of Bayesian Networks in the diagnosis of pulmonary embolism*”, J Thromb Haemost., 1(4), pp.698-707, 2003
- [42] Curiac, D.-I. , Vasile, G. , Baniias, O. , Volosencu, C. , Albu, A., “*Bayesian network model for diagnosis of psychiatric diseases*”, Information Technology Interfaces, ITI '09. Proceedings of the ITI 2009 31st International Conference, 2009
- [43] Bayesian Networks and Simulation Software for Risk Analysis and Decision Support,[online], <http://www.agenarisk.com/>