International Journal of Computer Science and Software Engineering (IJCSSE), Volume 2, Issue 1, November 2014
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 50-61

Bayesian Network-Based Causal Analysis of Injury Risk in Elite Rhythmic Gymnastics

Lyudmila Dimitrova[1] and Kristina Petkova[2]

[1] Department of Computer and Information Technology, University "Prof. As. Zlatarov", 8008 Burgas, Bulgaria

[2] Faculty of Industrial technology, Technical University of Sofia, 1756 Sofia, Bulgaria

[1]lyudim@gmail.com, [2]kristinapetkova.tu@gmail.com

ABSTRACT

Bayesian networks (BN) are a key information technology for dealing with probabilities in artificial intelligence. In the present work we introduce an application of BN as a tool for estimating injury risk in high level rhythmic gymnastics. At this stage we propose the base structure of the model consisting of five subnets, contributing to overall injury risk. Most of the model conditional probability tables are estimated with T Normal functions – a feature included in Agena Risk tool. Sensitivity analysis characterizes the degree of influence of the different input factors, which is consistent with expert knowledge. The model results are satisfactory for the test set of gymnasts in the current competitive season. Quantitative predictions show a significant opportunity for reducing injury rate, but further data collection and research are necessary to improve the precision of the model.

Keywords: *Bayesian network, risk analysis, gymnastics injury.*

1. INTRODUCTION

Physical fitness and sports can play a powerful role in personal development and have multiple benefits to society [2]. Active engagement in sports can help youth build personal attributes that are highly relevant to professional success later in life. Such attributes include self-discipline, courage, confidence, healthy self-esteem, team work and ability to follow rules. Physical activity is increasingly penetrating all spheres of social life, becoming a vital part of human interaction. Sports are a good alternative to leisure activities involving bad habits and addictions and can play an important role in maintaining personal relationships and improving social life. In recent decades an increased popularity of professional sports is observed - this has led to a substantial growth in the industry for supporting sports and entertainment events. At the same time, public health organizations are paying attention to the dark side of professional sport, including the risks for developing injuries [3, 4].

In order to achieve success and recognition, high-level athletes must mobilize all of their physical and mental reserves, which leads to substantial health risks during training and competitions. Nonetheless, most of the risks associated with professional sport can be minimized by the adoption of suitable prevention and control strategies [5]. Recent studies from the fields of sports medicine and epidemiology has investigated opportunities to reduce sports injuries [6,7].

Each sport has its own distinctive injury profile. Rhythmic gymnastics (RG), along with artistic gymnastics (AG), acrobatics and figure skating, form the group of artistic sports. Despite the differences among these sports, all of them require lengthy and intensive training starting at a very early age. Mastering complex gymnastic, acrobatic and jump exercise requires a large number of repetitions and causes an overload of individual muscle groups, tendons, and ligaments, which can lead to injuries.

Depending on their severity, injuries can be classified into severe, moderate and minor [8]. Severe traumas result in adverse impact on overall health and inability to practice sports for more than 30 days. Moderate traumas result in impacts to overall health leading to inability to practice sports for a period between 10 and 30 days. Minor traumas don't result in substantial health consequences. In such cases it is possible to continue training with reduced intensity under appropriate medical supervision.

Depending on the occurrence of the injury, they can be chronic and acute. Acute traumas usually occur suddenly during sports activity as a result of a trauma factor influence. Chronic traumas usually occur after practicing for an extended period of time and are a result from the repetitive action of a trauma factor on one area of the body.

Artistic gymnastics (AG) and in particular female artistic gymnastics, is the most studied with regard to injuries. A large number of prospective and retrospective studies in women's AG have been published [8-11]. They provide summary statistics for injuries - injury rates per season, per 1000 hours of practice or per athletic exposures (AE). The data on the injury risk degree of the various researchers differs widely. For example, studies report injury rates ranging from 0.5 to 5.3 injuries per 1000 hours of practice for female gymnasts [11].

Despite increased popularity of RG (gymnasts from 53 countries of all continents participated in the last World Championship in Turkey), there is not enough statistical data about injuries in RG. Overview of research in this field was presented in our previous study [12, 13]. Table 1 summarizes the available statistical data for injury rates in RG.

Table 1: Statistical data about RG injuries

Study	***Country***	***Study length***	***Number of gymnasts***	***Number of injuries***
Cupistu [14]	Italy	8 months	73	49
Hutchinson[15]	USA	7 weeks	7	474
Zetaruka [16]	National team	1 year	20	108, 74 mild

The above cited data about injury rates differs even more than those in women AG: Cupistu [14] reports 1,08 injuries per 1000 hours of training, whereas Hutchinson's [15] data gives 284.5 injuries of any degree of severity per 1000 hours of training or 56.9 injures with severity 4-10 per 1000 hours (20% of all registered injuries).

On one hand these studies are difficult to summarize and compare because of the various definitions of injury and on the other - due to the fact that many different factors are affecting injury risk. These factors are listed in the literature [17,19] without a quantitative measure for their influence on the injury. For preventive injury measures to be effective, the combined effect of the various risk factors and the mechanisms by which they work should be examined in greater detail.

This paper proposes the application of BN modeling for complex quantity measuring of injury risk. The study is focusing on risks of injury in professional RG for which observational and expert opinion data are available. Different artistic sports have specific physical demands and generalization of quantitative results is difficult. Nonetheless, since the mechanism of influence on the various factors on the risk for injury is similar across the above mentioned sports, this approach can be applied with some modifications for all artistic sports.

2. METHODS AND MODEL

2.1 Bayesian Networks (BN)

Bayesian Networks are graphical models that can efficiently represent and manipulate n-dimensional probability distributions [20-25] This representation has two components that respectively codify qualitative and quantitative knowledge:

1) A graphical structure - directed acyclic graph or DAG, G = (V, E), where the nodes in $V = \{X_1, X_2, ..., X_n\}$ represent the random variables from the problem we are modeling, and the topology of the graph (the arcs in $E \subseteq V \times V$) encodes conditional (in)dependence relationships among the variables.
2) A set of numerical parameters, usually conditional probability distributions (conditional probability tables, CPT, for discrete variables) drawn from the graph structure: For each variable $X_i \in V$ we have a conditional probability distribution $P(X_i|Pa(X_i))$, where $Pa(X_i)$ represents any combination of the values of the variables in $Pa(X_i)$, and $Pa(X_i)$ is the parent set of X_i in G.

The structure or topology, of the network should capture the qualitative relationships between the variables. In particular, two nodes should be connected directly if one affects or causes the other, with the arc indicating the direction of the effect. The experience and knowledge of domain experts can be used to design the initial structure and the initial values of the relationship coefficients (i.e. CPT) between any two causal nodes of the BN. The final structure and parameters can be revised through learning algorithms [26-31]. Generally, learning a Bayesian network from data consists of the induction of its two components:

1) The graphical structure of conditional dependencies (model selection);
2) The conditional distributions quantifying the dependency structure (parameter estimation).

The local Markov property states that each node in BN is independent of its non-descendant given the parent nodes and leads to a direct factorization of the joint distribution of the network variables into the product of the conditional distribution of each variable Xi given its parents. Therefore, the joint probability (or density) of the network variables can be written as:

$$P(X_1, X_2, ...X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)$$

The probability distribution of a particular random variable is determined by marginalizing the joint probability distribution with respect to this random variable. In the presence of evidence (i.e., the state of one or more random variables is known with certainty), the probabilities propagation algorithms [20-25] determine the marginal posterior probability for each variable given the evidence.

Two basic types of inference support of the BN can be used: predictive reasoning for node X_i, based on evidence nodes connected to X_i through its parent nodes (also called top-down reasoning), and diagnostic reasoning for node X_i, based on evidence nodes connected to X_i through its children nodes (also called bottom-up reasoning).

Uncertainty is an inherent feature in nearly all medical problems. For this reason there are many successful application of BN in medicine, in particular as a decision support system in medical diagnosis [32, 43]. Another typical application of BN is risk analysis [44, 47].

Since sports injuries are a medical problem, they have a probabilistic nature. Bayesian networks are a suitable tool for assessing the impact of multiple factors on the likelihood of their manifestation.

In our previous works [12, 13, 48, 49] we presented applications of BN for estimating injury risk in rhythmic and artistic gymnastics. Studies [12, 13] have characterized the conditions leading to injuries in the early career of professional rhythmic gymnasts: (1) poor tissue elasticity, (2) tissues fatigue, (3) incorrect movement, including fall or apparatus hit, and (4) presence of early osteoporosis In [48] we introduced briefly a model for evaluating the injury risk for the national team gymnasts, preparing for top level events in national sports centers. This work provides a comprehensive model presentation as well as injury rate predictions of the model for a test set of 14 national team competitors during the last competitive season.

2.2 Defining the structure of the BN model for estimating injury risk in elite rhythmic gymnastics

For modeling the risk of trauma for professional gymnasts from national teams of various countries, some of the mentioned factors included in the previous model [13] become less applicable. National teams are usually provided with good training and rehabilitation conditions. In addition, top level gymnasts are highly motivated and have strong support from their families. The present Code of Points (COP) does not encourage the flexibility elements which led previously, in the recent past, to high rates of lower back trauma. Thus COP can be ignored as a risk factor at the moment.

On the other hand, the physical and psychological pressures that national team gymnasts are subjected to are substantially higher. The scientific approach in the organization of training, monitoring of gymnast condition, rehabilitation and treatment is critical in trauma prevention. As with every professional sport, a number of factors influence the optional athlete conditions and these factors should be periodically monitored, controlled and optimized. In this study, we use the model presented in Figure 1 for quantitative assessment of the complex influence of these factors to severe injury risk in RG. AgenaRisk software [1] has been used for the model presentation and analysis. The most natural way for building realistic extensive models is to link BNs, describing smaller parts of the modeling problem. Based on the domain knowledge, the network is separated in 5 sub-networks (subnets), through which the influence of the various categories of factors influencing trauma risk can be assessed. First subnet describes how coaching expertise and competitions plan are influencing injury risk. In addition, in the presented model we suggest four mechanisms for occurrence of traumas in elite gymnasts, modelled with sub-networks 2-4 on Figure 1.

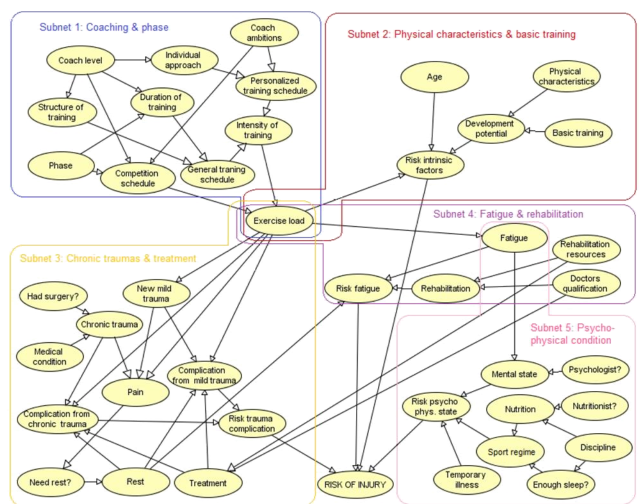

Fig. 1. Overall Bayesian Network (BN) model

2.2.1 Subnet Coaching and Phase

The organization of the training and competition process plays a central role in the management and prevention of gymnastics injury, as demonstrated by subnet 1 in Figure 1. An important input factor is professionalism and competency of the coaches in the training process - node Coach level (high/ very high). This key root node cannot be measured directly but can be indirectly assessed through personal coach indicators such education, experience, pedagogic approach, previous track record and achievements, trauma injury frequency in previous athletes, as well as commitment to professional development.

In order to avoid crowding of the main scheme (Fig.1), the assessment of complex nodes takes places in separate sub-networks. Figure 2 presents the BN for estimating of Coach level probability distribution as means of contributing indicators. The BN implementation of this approach requires additional node Weights, which represents as states, the six indicator nodes weights on which the resulting factor Coach level depends. This approach can be used for other nodes in the model which can be assessed though measurements (controls) or indirect indicators (indicators).

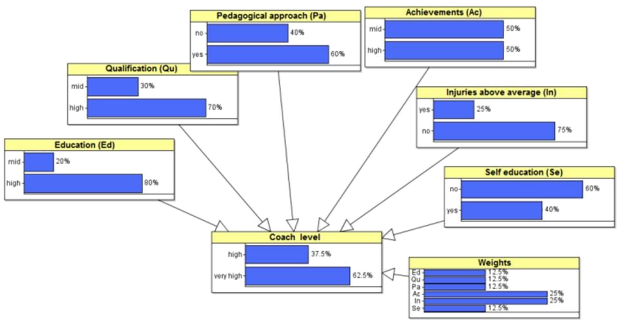

Fig. 2. Estimation of Coach level as weighted average of indicators

Competitions are the goal for each coach and competitor. If the coach pursuits medals at any costs this affects the organization of competitors' training and competing schedule. This is represented by node Coach ambitions (very high/high). The value "very high" has negative influence, because it is often the case when the coach rushes the gymnast to train with higher intense prolonged time; to perform more difficult exercise without having enough time for rest; to fill the schedule with too much competitions.

The node Coach level determines the three major parameters of the training process: Structure of training (non optimal/optimal), Duration of training (long/normal) and Individual approach (no/yes) to competitors. Duration of Training are without a doubt influenced by the stage of training – node Phase (competitive/preparatory). General training schedule depends on the coach ability to organize well the time and structure of the training. It is possible that, even after following the same schedule, outcomes among the different gymnasts differ due to the trainers' individual approach, capabilities and ambition – node Personalized training schedule. Intensity of training for different gymnasts depends on General training schedule, and on Personalized training schedule. Competition schedule (hard/moderate) depends on the stage of training as well as on the appropriate planning by the coach. Intensity of traning and Competition schedule influence together the major injury risk factor – the overall training and competition load - node Exercise load with states very high /high/moderate.

2.2.2 Subnet physical characteristic and basic training

Intrinsic qualities of the gymnast and the physical load she is experiencing during practice also influence the risk – this is defined in subnet 2 by the node Risk intrinsic factors (high/low). The entry nodes Physical characteristics and Basic training, both with states good/excellent, similarly to Coach Level can be assessed as weighted averages of the described in the specialized literature [8] controls and indicators.

2.2.3 Subnet Chronic traumas and treatment

A substantial part of the severe traumas in gymnastics are due to complications resulting from a minor or chronic trauma[11]. This mechanism is illustrated in subnet 3 on figure 1, with end node Risk trauma complication (high/low). The elite gymnast may have undergone surgery - Had sergery surgery (yes/no) or may have trauma-prone body area due to medical issues or physical characteristics - node Medical condition (not quite healthy/healthy). These two factors determine the presence of Chronic trauma (yes/no). New mild trauma (yes/no) can occur depending on the level of physical stress.

Whether a new or existing due to predisposition trauma will worsen and result in a severe trauma depends on the training intensity, proper treatment availability and Rest (shorter with competitions/ shorter/ recommended/ na) The last node assesses if the gymnast is complying with the rest and recovery measures prescribed by a physician. The need for rest in the general network is defined by the node Need rest?(yes/no), which in turn depends on the amount of pain felt – node Pain (intense/bearable/slight or no) If the value of Need rest? is „no" Rest assumes value n/a and doesn't influence the risk caused by complication of chronic or new mild injury.

2.2.4 Subnet Fatigue and rehabilitation

Fatigue is critical risk factor for sports injuries. Its influence is demonstrated with subnet 4 on Figure 1. Underlying factors influencing target node Risk fatigue (high/low) are Exersice load, Rehabilitation level, which is determined by Rehabilitation Resources and Doctors Qualification, and compliance with a prescribed rest regimen.

2.2.5 Subnet Psychophysical factors

The fourth and last risk mechanism for occurrence of severe (time loss>30 days) trauma included in the model is the risk as a result of unfavorable psychophysical condition (subnet 5, figure 1). The athlete condition depends on following an optimal regimen of sleep and nutrition, as well as current physical and mental health

status. The entry nodes for this subnet determine the presence of the appropriate professional - Psichologist?(no/yes), Nutritionist?(no/yes) and athlete health status Temporary illness(yes/no) and Discipline (no/yes). Target node is Risk psychophysical state (high/low).
The overall risk of injury is determined as a result of the joint influence of four analyzed risks, all of which are given equal weight (Fig.1).

2.3 Eliciting parameters of the model

Very convenient feature of the AgenaRisk tool [1,44] is the broad possibility to define CTP (in this product referred as NPT node probability table), with the usage of functions. Possible number of functions depends on the node type - Labelled, Boolean, Continuous Interval, Ranked, Integer Interval or Discrete Real. In the proposed model some of the nodes are defined as Ranked. Ranked nodes are mapped to underlying numerical scale. This means that no matter what the state labels are and how many states a node has, there is an assumption that there is an underlying numerical scale that goes from 0 to 1 in equal intervals.
By defining the ranked scale the user must be sure that the labeling of the states is consistent from worst to best. In our case in the beginning of the scale should be the states of the nodes that lead to the higher risk of injury. For example, to model Risk fatigue in terms of factors Fatigue, Rehabilitation and Rest in consistent way, states of the last three nodes must be ordered from worst to best. Thus, states of the node Risk fatihue should be ordered from highest to lowest. Here this is done in a 2 state scale – high and low.
Due to the underlying numerical scale of the ranked nodes, we can define numerical statistical distribution expressions on it. Especially useful for defining NPT is truncated Normal distribution (TNormal). Unlike the regular Normal distribution TNormal has finite end-points. For ranked nodes whese endpoints are 0 and 1, respectively. Like the Normal distribution, the TNormal is characterized by two parameters: the mean and variance. This enables us to model variety of distribution shapes – the desired distribution is achieved through appropriate setting of mean and variance parameters. In the simplest case the parameter mean is determined as a weighted mean of the parent nodes:

$$WMEAN = \frac{\sum_{i=1\ldots n} w_i X_i}{n} \qquad (5)$$

where $w_i \geq 0$ are weights, and n is number of parent nodes. In AgenaRisk syntax of the function is

wmean(w₁, parent₁, w₂, parent₂,..., w_N, parent_N) (6)

If a simple wighted mean (WMEAN) do not satisfy requirements for the distribution, build in ArenaRisk functions WMIN, WMAX and MIXMINMAX can be used.

The weighted min function is defined as:

$$WMIN = \min_{i=1\ldots n}\left[\frac{w_i X_i + \sum_{i\neq j} X_j}{w_i + (n-1)}\right] \qquad (7)$$

The analogous weighted max function has general form:

$$WMAX = \max_{i=1\ldots n}\left[\frac{w_i X_i + \sum_{i\neq j} X_j}{w_i + (n-1)}\right] \qquad (8)$$

Finally, the function MIXMINMAX is a mixture of two functions WMIN and WMAX with weights w_{min} and w_{max}, respectively

$$MIXMINMAX = \frac{w_{min} WMIN + w_{max} WMAX}{w_{min} + w_{max}} \qquad (9)$$

In AgenaRisk syntax of the function is

$mixminmax(w_{min}, w_{max}, parent_1,\ parent_2, ..,.parent_N)$ (10)

As is known, Normal distribution is widely used in probability theory and statistics [50, 51]. Its incapability to predict vast range of rare events does not affect our model, due to the lack of their usage. TNormal distribution is a very flexible extension of Normal. Equal weights of parent nodes results in symmetric distribution and different weights in biased distribution. Variance parameter reflects the influence of parent nodes uncertainties. As the variance is rising the distributions is getting closer to uniform. Thus TNormal is proper choice for eliciting conditional distributions in our model.

2.3.1 Subnet Coaching and phase

CPT of the input and internal nodes are estimated by experts opinion. The CPT for the node *Exercise load* is calculated as partitioned expression [42]. This is feature of AgenaRisk that allows tocondition CPTs on some of the parent nodes. In our case CPT is conditioned on node *Competition schedule*. If *Competition Shedule*=”*na*”, i.e. in case of a preparatory phase, this node is excluded as parent node in determining the mean as TNormal distribution (Fig. 3). By providing of the probability distribution expressions, the names of the parent nodes– in this case ***it*** (intensity of training) and ***cs*** (competition schedule) are used.

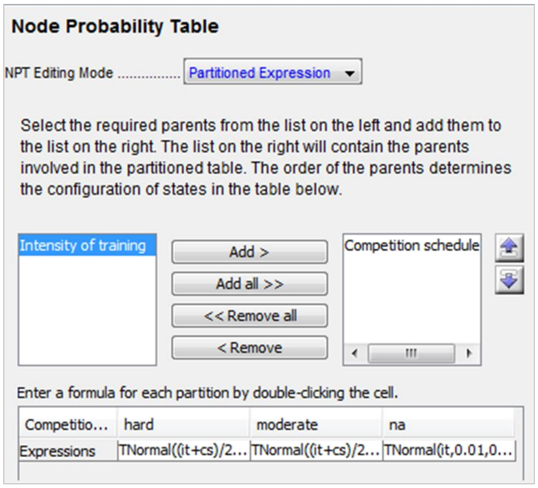

Fig. 3. Defining CPT for the node Exercise load

Prior marginal probability distribution of the subnet is shown in Fig. 4. The tornado graph in Fig.5 is derived by the Sensitivity analysis tool of AgenaRisk [1]. The length of the bars corresponding to each sensitivity node is a measure of the impact of that node on the target node Exercise load.

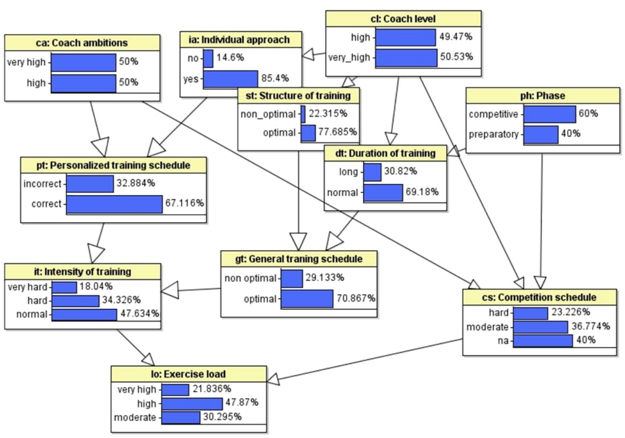

Fig. 4. Prior marginals of the subnet, estimated main risk factor – Exercise load

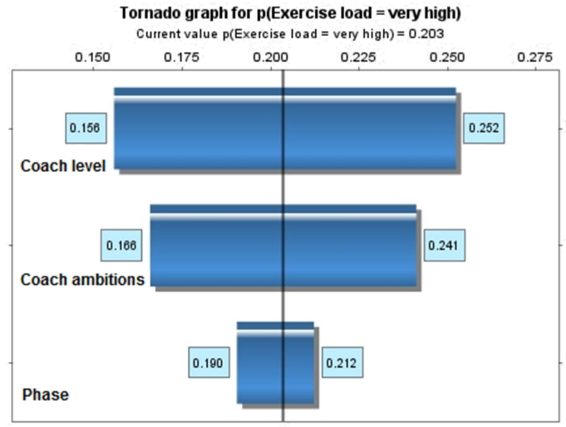

Fig. 5. Impact of the input nodes of subnet 1 on the Exercise load beeng "very high"

As expected, the training and competition intensity is determined by the trainer's qualification to the largest extend followed by the trainer's ambition for achieving high scores and the training phase (Fig.5).

2.3.2 Subnet Physical characteristics and basic training

Eliciting of CPT for this most simple subnet is achieved through EM learning of Samiam tool [52] based on the collected data for injuries in two national teams. In preparing the database, the values for the gymnast's physical abilities, preparation, potential for development and training intensity are provided based on expert assessment. If the gymnast has experienced more than one trauma per year, the risk of trauma is recorded as high. The derived CPTs are approximate because they are derived based on the assumption that the injury depends on the difference in the physical abilities and the initial training of the gymnast. A better assessment will be possible through learning of the overall model.

Fig 6 demonstrates the derived marginal prior probability distribution in the subnet, whereas Fig.7 – the impact of input nodes on target node *Risk intrinsic factors.*

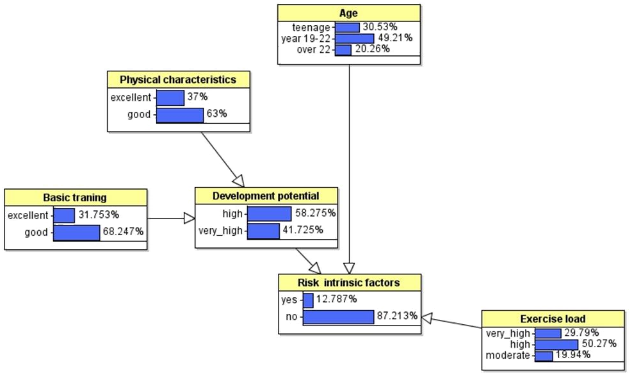

Fig. 6. Prior marginals of the subnet estimated risk due to intrinsic factors

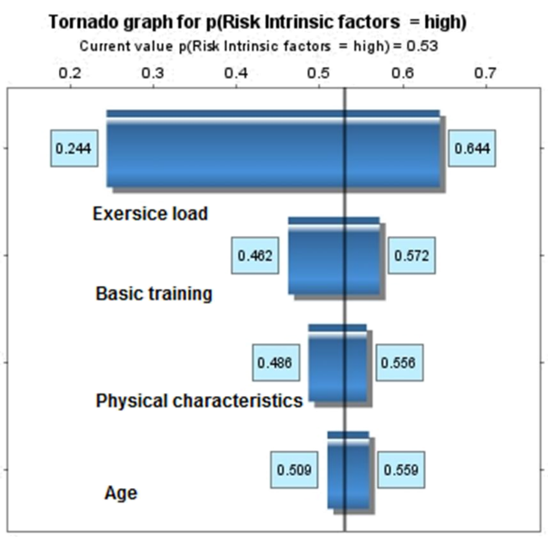

Fig. 7. Impact of the input nodes of subnet 2 on the Risk intinsic factors beeng "high"

2.3.3 Subnet chronic traumas and treatment

CPT for node *Pain* with parent nodes *Chronic trauma* (***ct***), *New mild trauma* (***nmt***) and *Exercise load* (***lo***) is calculated as TNormal with weighted average mean:

TNormal (wmean (1.5, ct, 1.0, nmt, 2.0, lo), 5e-4)

Exercise load has the greatest weight or has the most substantial influence in causing pain, followed by chronic trauma and new mild trauma.

CPT for nodes *Complication from chronic trauma* (***cct***) and *Complication from mild trauma (**cmt**)* are calculated as partitioned exspressions on node *Rest*. For states of *Rest* "*shorter with competitions*", "*shorter*" and "*recommended*" used expressions for ***cct*** and ***cmt*** are:

TNormal (mixminmax (1.0, 2.0, ct, tr, rs, lo), 5e-4)

TNormal (mixminmax (1.0, 2.0, nmt, tr, rs, lo), 5e-4)

For the state *Rest* = "*na*", т.е. rest is not needed and therefore does not present as a parent, CPT of the nodes ***cct*** and ***cmt*** are calculated with expressions:

TNormal (mixminmax (1.0,2.0, ct, tr, lo), 5e-4)

TNormal (mixminmax (1.0,2.0, nmt, tr, lo), 5e-4)

The calculation of the mean of TNormal distribution with *mixminmax* function with `doubled weight of maximum assures a shift of the distribution towards the maximal (favorable) values by the corresponding favorable values of the parent nodes. For example, if *Chronic trauma ="no"*, *Exercise load="moderate"*, *Treatment="excellent"* and *Rest="na"*, the value *Complication from chronic trauma ="no"* is close to 100% - which corresponds to the observed situation.

CPT of the target node *Risk trauma complication* is defined as:

TNormal (wmean (1.0,cct, 1.0, cmt), 5e-4)

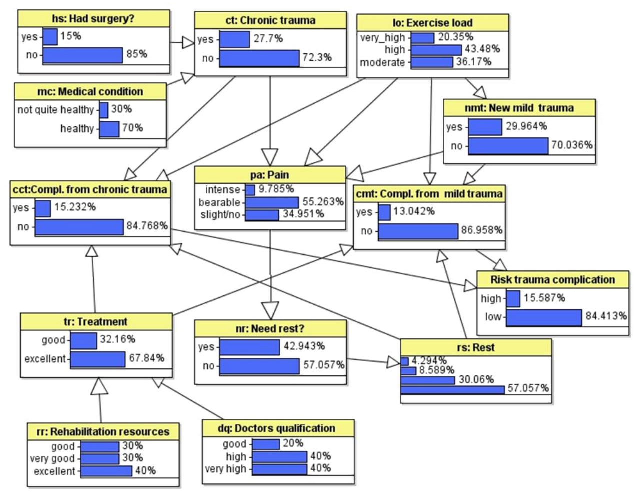

Fig. 8. Prior marginals of the subnet, estimated Risk of trauma complication

Figures 8 and 9 show prior marginals of the subnet and impact of input nodes on the end node *Risk trauma complication*. Accordingly, in this subnet, the main factors influencing risk are the exercise load and the quality of the treatment, defining the variable *Doctors qualification*. Less important factors in order of decreasing importance are the number of previous operations, chronic conditions leading to traumas, and *Rehabilitation resources*.

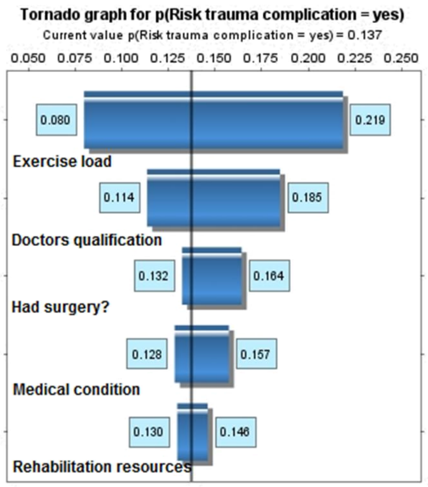

Figure 9. Impact of the input nodes of subnet 3 on the Risk trauma complication beeng "yes"

2.3.4. Subnet Fatigue and rehabilitation

CPT for node *Risk fatigue (rf)* is defined as partitioned exspressions on node *Rest*. For the states of *Rest* "*shorter with competitions*", "*shorter*" and "*recommended*" the used expressions are :

TNormal (mixminmax (1.0,2.0, fa, rh, rs), 5e-4)

and for the state *Rest="na":*

TNormal (mixminmax (1.0,2.0, fa, rh), 5e-4)

where ***fa***, ***rh*** and ***rs*** are its parent nodes *Fatigue*, *Rehabilitation* and *Rest.*

Prior marginal distribution for this subnet is shown on Figure 10. The derived value of around 26% for *Risk fatigue="high"* shows that in the proposed model tissue fatigue poses the most substantial risk.

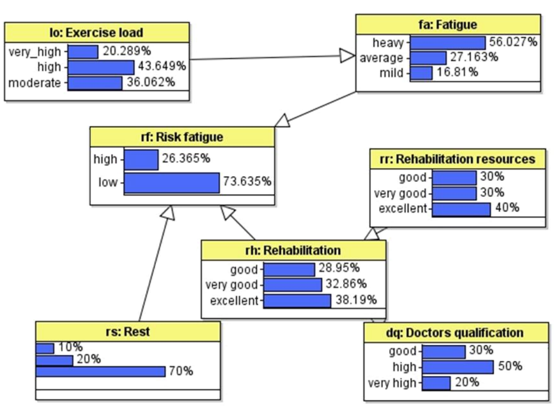

Figure 10. Prior marginals of the subnet, estimated Risk due to fatigue

Figure 11 presents the impact of the input nodes on the target node *Risk fatigue*. The most substantial influence in this subnet have *Rehabilitation resources*, which is consistent with previous studies about the role of rehabilitation in injury prevention[8].

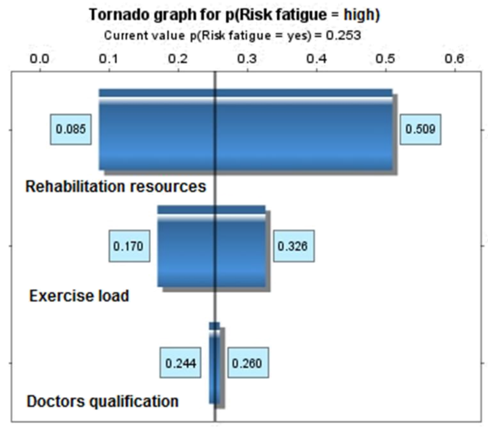

Figure 11. Impact of the input nodes of subnet 4 on the *Risk fatugue* beeng "*high*"

2.3.5. Subnet Psychophysical state

The determination of CPT of the end note Risk psychophysical state takes place in a similar way:

TNormal (mixminmax (1.0,2.0, ms, sr,te), 5e-4)

Prior marginals and impact of the root nodes are given in figures 12-13. The most trauma dangerous factor in the model is the training and competition activity during temporary illness, a situation that occurs often in big competitions because of the assessment that the risk is compensated by a good score. The gymnast discipline, good nutrition and stable psychological state are other important factors.

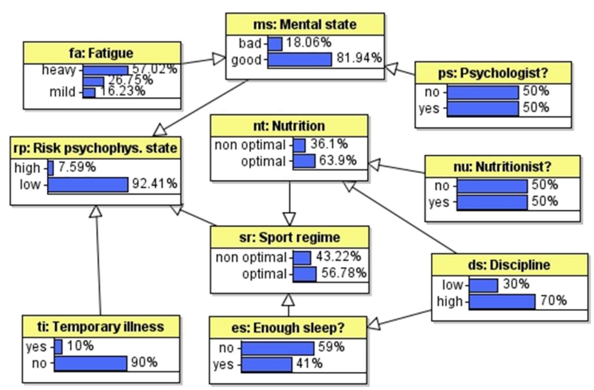

Figure 12. Prior marginals of the subnet, estimated risk of injury due to psychophysical state

The overall risk of trauma - *Risk of injury (high/low)* is calculated as TNormal with mean parameter calculated as weighted mean of the four components – *Risk intrinsic factors* (***ri***) , *Risk trauma complication* (***rtc***), *Risk Fatigue*(***rf***) and *Risk psychophysical state*(***rp***), which participate with equal weights:

TNormal (wmean(1.0, ri, 1.0, rtc, 1.0, rf, 1.0, rp), 5e-4)

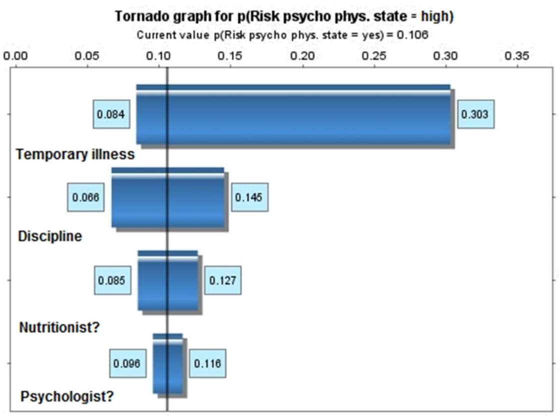

Fig. 13. Impact of the input nodes of subnet 5 on the Risk psychophysical state beeng "high"

3. RESULTS AND DISCUSSION

Figure 14 presents marginal prior probability distribution of overall network. The leaf node *Risk of injury* has value "*high*" with probability 12.82% and "*low*" with probability 87.18%. In accordance with the presented statistical data in the introduction section, these results can be interpreted as percentage of the serious traumas per 1000 hours of training and competitions. The national team rhythmic gymnasts train between 40 and 50 hours per week or approximately 2000 per year (48 training weeks per year). Therefore, depending of the model assessment *for the approximate values* of the input nodes, in average 2 to 3 of every 10 elite gymnasts annually are exposed to risk of more time loss injuries.

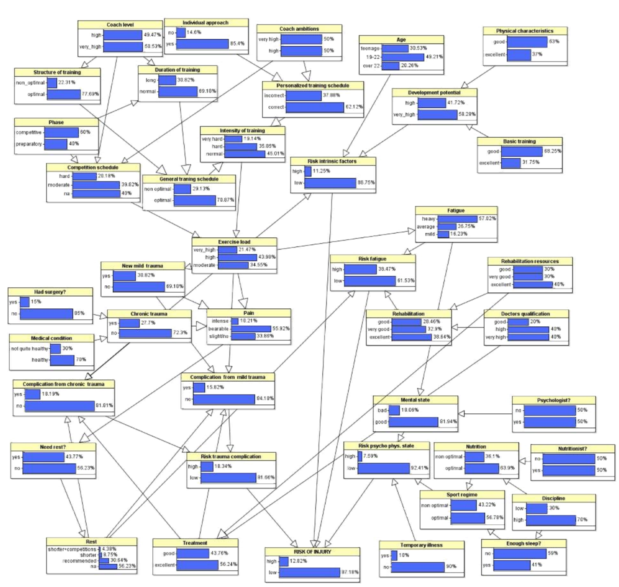

Fig. 14. Prior marginal probability distribution in overal network

Table 2: Observations for three possible scenarios

Sub net	Observations			
	Node	***Scenario 1***	***Scenatio 2***	***Scenario 3***
1	Coach level	very_high	no answer	no answer
1	Coach ambitions	high	very high	no answer
1	Phase	preparatory	competitive	no answer
2	Age	19-22	age 19-22	year 19-22
2	Ph. characteristics	excellent	no answer	no answer
2	Basic training	excellent	no answer	no answer
3	Had surgery?	no	no	no
3	Medical condition	healthy	not quite healthy	no answer
3	Pain	no answer	bearable	no answer
3-4	Rehab. resources	excellent	no answer	good
3-4	Doctors qualification	very_high	no answer	no answer
4	Fatigue	no answer	no answer	heavy
5	Psychologist?	yes	no answer	no
5	Nutritionist?	yes	no answer	no
5	Discipline	high	no answer	no answer
5	Temporary illness	no	no	no

In practice, this number varies greatly for the national teams of different countries which is due mainly to the varied training conditions, or the complex influence of numerous factors in each country. This situation is adequately captured by the proposed model.
Figure 15 demonstrates posterior probability distribution of three scenarios. The provided observations are listed in Table 2.
Scenario 1 corresponds to favorable observations for input nodes (table 2). In this case, the value of *Risk of injury="high"* reduces to 3.52%, which means that give an optimal organization of the training, treatment and rehabilitation throughout the year, time loss injuries would occur in 7 out of 100 gymnasts, which translates to a very low chance for serious trauma for each individual gymnast.
Scenario 2 corresponds to the most often observed situation: gymnast with chronic illness *Medical condition= "not quite healthy"* feels some pain *Pain=„bearable"*, but the importance of the competition or the trainer's ambition *Coach ambitions ="very high"* lead to the gymnast's participation in sport activity. According to the model, trauma risk increases to 15.62%.
Scenario 3 corresponds to a situation in which the gymnast feels substantial fatigue - *Fatigue ="heavy",* the rehabilitation resources are good but not optimal for professional athletes *Rehabilitation resources="good",* the

gymnast doesn't utilize the services of a nutrition specialist *Nutritionist?="no"* and psychologist *Psychologist?="no"*. The model projects a substantial increase of risk by participation in trainings and competitions in such condition - *Risk of injury="high"* increases to 24.52%.

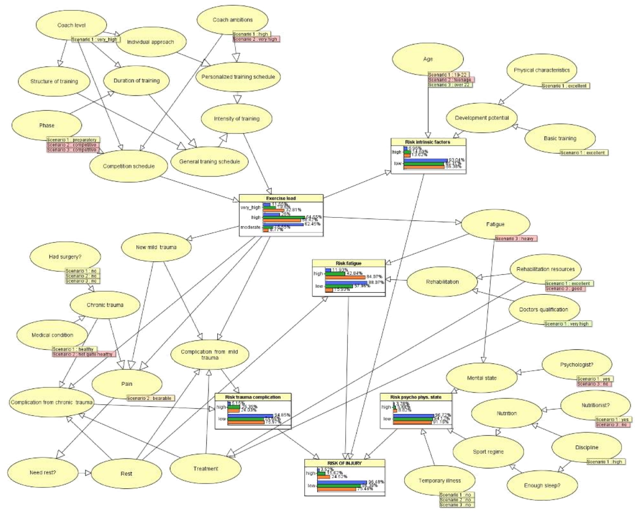

Fig. 15. Posterior marginal probability distribution in overal network for three scenarios

Determined risk components for these three of many possible scenarios and overall injury risk are listed in Table 3.

Table 3: Calculated injury risks

Sub net/ net	Calculated risks , %			
	Node	***Scenario 1***	***Scenario 2***	***Scenario 3***
2	Risk intrinsic factors	6.96	14.69	13,62
3	Risk fatigue	5.15	25.35	24.03
4	Risk trauma complication	11.93	42.04	84.07
5	Risk psychophysical state	3.28	5.88	8.82
full	Risk of injury	3.52	15.62	24.52

The data of fourteen gymnasts from two national teams are used as input to test the model. Comparison between risk estimation derived by the model and the actual injuries occurred during the last competitive season is represented in Table 4. Predictions are approximate due to the missing data for several input nodes. Nevertheless the model gives high risk estimation of injury occurrence for the gymnasts, who actually experienced severe acute or overused injury.. For competitors who receive moderate injuries the model predicts more than 20% probability of serious injury. Only two gymnasts have not experienced any injuries – the risk for them is estimated 5.47%.

The presented results demonstrate that there is a substantial potential for reducing injuries in gymnastics. The proposed model can be used as a decision-support tool by the coaching team of elite rhythmic gymnasts. After considering different variations of causal and evidential reasoning, specialists can use the calculated quantitative assessments of injury risk in decision making regarding the organization of the training and competition process and other components in the preparation of professional athletes.

Table 4: Estimated injury risk compared to real injury occurrence for national team gymnasts

Gymnasts number	***Calculated overall risk***	***Received injuries***		
		Mild	***Moderate***	***Heavy***
1	31.14%	-	-	yes, acute
2	64.43%	-		yes, overused
3	58.80%	-		yes, overused
4	25.66%	-	yes, overused	
5	17.91%	yes, often	-	-
6	22.16%	-	yes,overused	-
7	22.72%	-	yes,overused	
8	23.02%	-	yes,overused	
9	19.03%	yes,often	-	-
10	19.03%	yes,often	-	-
11	5.47%	-	-	-
12	5.89%	yes,rare	-	-
13	6.48%	yes,rare	-	-
14	5.47%	-	-	-

Posterior probability distribution for each case shows the bottleneck factors that caused the injury – components of risk with significant increase compared to default minimal risk (corresponding to optimal training strategy). The latter indicates the preventive cautions proposed for each team member – doctors, coaches, nutritionists, therapists and etc. Further model precision is possible by extending the possible node states (for example, including the actual practice time in hours per week as possible states of node *Exercise Load*) and its learning based on individual records of gymnasts' abilities and underlying risk factors.
An evidence of the need for this tool is the high frequency of injuries among elite rhythmic gymnasts.

4. CONCLUSION

A Bayesian Network model for evaluation of the risk of injury has been proposed for elite, national team rhythmic gymnasts, in which the influence of all trauma-dangerous factors from the literature was taken into account. Quantitative measures were calculated using this model. They show the influence of different factors on the risk of injury. Estimated risk of injury is synchronized with the existing data about injuries and reflects the influence of different factors over it. The BN model could be used by national coaches as a method for analyzing the risk of injury and taking precautions for lowering it. For improving the quantitative assessment of the injury risk additional data on trauma rates among the national teams as well as parameter learning for the model is necessary.

REFERENCES

[1] "Bayesian Networks and Simulation Software for Risk Analysis and Decision Support", http://www.agenarisk.com/
[2] G Scambler, "Sport And Society: History, Power And Culture, McGraw-Hill International", 2005
[3] S. Murphy, "The cheers and the tears: a healthy alternative to the dark side of youth sports today". San Francisco (CA): RJossey Bass Publishers, 1999
[4] C. Finch, E.Cassell, "The public health impact of injury during sport and active recreation", Journal of Science and Medicine in Sport, Vol.9, No.6, 2006, pp 490-497
[5] T.Timpka, C.F. Finch, C.Goulet, T.Noakes and K.Yammine, "Meeting the Global Demand of Sports Safety", Sports Medicine, Vol.38, No.10, 2008, pp. 795-805
[6] S. Aaltonen, H. Karjalainen, A. Heinonen, J.Parkkari, U. M. Kujala, "Prevention of Sports Injuries Systematic Review of Randomized Controlled Trials", Archives of Internal Medicine, Vol.167, No.15, 2007, pp. 1585-1592
[7] R Bahr, T Krosshaug, "Understanding injury mechanisms: a key component of preventing injuries in sport", British Journal of Sports Medicine, Vol.39, No.6, 2005, pp.324-329
[8] V.I.Dubrpvskij, "Sportivnaya medicina", Vlados, 2002
[9] J.P. DiFiori, D.J.Caine, "Gymnastics" in ACSM's Sports Medicine: A Comprehensive Review F.G. O'Conner et al.(eds), McGraw-Hill, 2012
[10]D.J.Caine, L.Nassar, „Gymnastics Injuries", Caine DJ, Maffulli N (eds): Epidemiology of Pediatric Sports Injuries. Individual Sports. Med Sport Sci. Basel, Karger, Vol 48, 2005, pp 18–58
[11]Dennis J. Caine (Editor), Keith Russell (Editor), Liesbeth Lim (Editor), Gymnastics, Wiley-Blackwell, 2013
[12]L.Dimitrova, K. Petkova "A Bayesian Network Application for Estimating the Injury Risk for Pre-Elite Rhythmic Gymnasts", Informatics & IT today, 2 (1), 2014
[13]L.Dimitrova, K. Petkova , "Modeling Injury Risk for Pre-Elite Rhythmic Gymnasts: Bayesian Network Approach", The 2nd Electronic International Interdisciplinary Conference, September 2013, http://www.eiic.cz/
[14]A. Cupisti, C. D'Alessandro, I.Evangelisti, C. Umbri, M. Rossi, F.Galetta, E.Panicucci, S.L.Pegna, and M.Piazza, "Injury survey in competitive sub-elite rhythmic gymnasts: results from a prospective controlled study", Journal of Sports Medicine and Physical Fitness, Vol.47, No.2, 2007, pp. 203-207
[15]M.R. Hutchinson, "Low back pain in elite rhythmic gymnasts", Med Sci Sports Exerc. Vol.31, No.11, 1999, pp.1686-1688
[16]M.N.Zetaruka, M.Violanb, D. Zurakowskic, W.A. Mitchell, and L.J. Micheli, "Injuries and training recommendations in elite rhythmic gymnastics", Apunts Med Esport. Vol.41, No.151, 2006, pp.100-106
[17]R M Daly, S L Bass, and C F Finch "Balancing the risk of injury to gymnasts: how effective are the counter measures?", Br J Sports Med, Vol.35, No.1, 2001, pp. 8–20
[18]D. Miletic, R.Katic, and B.Males, "Performance of Rhythmic Gymnastics Novices", Coll. Antropol. Vol. 28, No. 2, 2004, pp. 727–737
[19]H. F. Freddie, A. David, and M.D. Stone, "Sports injuries: Mechanisms, prevention, treatment. Chapter 26 - Gymnastics", 2nd edition - Lippincott, Williams & Wilkins, 1994
[20]David Heckerman, Bayesian Networks for Data Mining, Data Mining and Knowledge Discovery, Vol. 1, No. 1, 1997, pp 79-119
[21]U.B. Kjaerulff, A.L. Madsen,. An Introduction to Bayesian Network and Influence Diagram, http://people.cs.aau.dk/~uk/papers/pgm-book-I-05.pdf, 2005
[22]U. B. Kjærulff, and A. L. Madsen, "Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis", Information Science and Statistics, Vol. 22, 2013,
[23]A. Darwiche, "Modeling and Reasoning with Bayesian Networks", Cambridge University Press; 2009, 562 p.
[24]Lauritzen, S. L. ,"Graphical Models", Oxford University Press, Oxford, UK, 1996
[25]Jensen, F. V.,"Bayesian Networks and Decision Graphs", Springer, 2001
[26]D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data", Machine Learning , Vol.20, No.3, 1995, pp 197-243
[27]D. Heckerman, "A Tutorial on Learning with Bayesian Networks", Innovations in Bayesian Networks, Studies in Computational Intelligence 2008, pp 33-82
[28]S.L. Laurizen, "The EM algorithm for graphical association models with missing data ", Computational statistics and data analysis, Vol.19, No.2, 1995, pp. 191-201
[29]J.A.Gámez, J. L. Mateo, J.M. Puerta, Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood, Data Mining and Knowledge Discovery, Vol. 22, No. 1-2, 2011, pp 106-148
[30]A.Feelders, A., L. C van der Gaag, "Learning Bayesian network parameters under order constraints", International Journal of Approximate Reasoning, Vol.42, No.1-2, 2006 , pp. 37-53,
[31]R.Daly, Q. Shen , S.Aitken, "Learning Bayesian networks: approaches and issues", The Knowledge Engineering Review, Vol.26, No.2, 2011, pp. 99–157, Cambridge University Press
[32]Heckerman D., "Probabilistic similarity networks", The MIT Press, 1991
[33]Y. Xiang, B.Pant, A.Eiseny, M.P.Beddoes , D.Poolez, "Multiply Sectioned Bayesian Networks For Neuromuscular Diagnosis", Artificial Intelligence in Medicine, Vol.5, No.4, , 1993, pp.293-314
[34]W. Long, "Temporal Reasoning for Diagnosis in a Causal Probabilistic Knowledge Base "Articial Intelligence in Medicine, Vol.8, No.3, 1996, pp.193-215

[35] L.C van der Gaag, S.Renooij, C.L.M Witteman, B.M.P. Aleman, B.G Taal., "Probabilities for a probabilistic network: a case study in oesophageal cancer ", Artificial Intelligence in Medicine, Vol.25 , No.2 , 2002, pp. 123–148

[36]P.J.Lucas, N.C. Bruijn, K. Schurink, A.Hoepelman, "A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU", Artif Intell Med. , Vol.19, No.3, 2000, pp.251-79

[37]X.Wang, H. Qu, P. Liu, Y.Cheng, "A self-learning expert system for diagnosis in traditional Chinese medicine", Expert Systems with Applications, Vol.26, No.4, 2006, pp. 557–566

[38]F.J.Díez , J. Miraa, E.Iturraldeb, S.Zubillagac,"DIAVAL, a Bayesian expert system for echocardiography", Artificial Intelligence in Medicine, Vol.10, No.1, 1997. pp. 59–73,

[39]P.J.F Lucas, H. Boot, B.G. Taal, "Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma", Methods Inform Med., Vol. 37, No.3, 1998, pp. 206–219

[40]S.Chattopadhyay, R.M. Davis, D.D. Menezes, G.Singh, R.U. Acharya , T.Tamura, "Application of Bayesian classifier for the diagnosis of dental pain", Journal of Medical Systems, Vol.36, No.3, 2012, pp.1425-1439

[41]D.Luciani, M.Marchesi, G. Bertolini., "The role of Bayesian Networks in the diagnosis of pulmonary embolism", J Thromb Haemost., Vol.1, No.4, 2003, pp.698-707

[42]D.I. Curiac, G.Vasile, O. Banias, C. Volosencu, A. Albu, "Bayesian network model for diagnosis of psychiatric diseases", Information Technology Interfaces, ITI '09. Proceedings of the ITI 2009 31st International Conference, 2009

[43]W.A.J.J.Wiegerincka, H.J.Kappena, E.W.M.T. ter Braakb., W.J.P.P. ter Burgc, M.J. Nijmana., Y.L.O, J.P. Neijtb, "Approximate inference for medical diagnosis", Pattern Recognition Letters, Vol. 20, No.11–13, 1999, pp. 1231–1239

[44]N. Fenton, M. Neil, Risk Assessment and Decision Analysis with Bayesian Networks, CRC Press, 2012

[45]I. Maglogiannis , E. Zafiropoulos, A. Platis, and C. Lambrinoudakis, "Risk analysis of a patient monitoring system using Bayesian Network modeling", Journal of Biomedical Informatics, Vol.39, No.6, 2006, pp. 637–647

[46]P. Truccoa, E. Cagnoa, F. Ruggerib, O. Grande, "A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation", Reliability Engineering & System Safety, Vol.93, No.6, 2008, 845–856

[47]K. Fong-Rey Liu, Che-Fan Lu, Cheng-Wu Chen, Yung-Shuen Shen "Applying Bayesian belief networks to health risk assessment", Stoch Environ Res Risk Assess, Vol. 26, No.3, 2012, pp. 451–465

[48]L.Dimitrova, K. Petkova, Estimating Injury Risk for Elite Rhythmic Gymnasts with a Bayesian Network Model, Proceedings in conference of informatics and management sciences, 2014

[49]L.Dimitrova, K. Petkova, Modeling Injury Rate in Women Artistic Gymnastics with a Bayesian Network, Gloval Virtual Conference, Vol.2, 2014

[50]M. J. Zaki; W. Meira Jr Data Mining and Analysis, Cambridge University Press, 2014

[51]Peter M. Lee, Bayesian Statistics: An Introduction, John Wiley & Sons, 2012

[52]A. Darwiche, "Samiam: Sensitivity analysis, modeling, inference and more". Software available from http://reasoning. cs. ucla. edu/samiam