Etiological Evaluation of Seminal Traits Using Bayesian Belief Network

Muhammad Naeem

Université Lumière Lyon 2 France
E-mail: Muhammad.Naeem@univ-lyon2.fr

Abstract

This study has modeled numerous confounding parameters of seminal quality for the purpose of digging out the hidden relationship between these seminal parameters using Bayesian Belief Network (BBN). The data source for this study was retrieved from UCI machine learning repository. Etiological patterns were derived out of complex relationship of nine related attributes. We have shown that as compared to conventional statistical measures, BBN is quite useful in analysis of seminal quality as well as classifying an unknown instance. The outcome is composed of a predictive probabilistic model which can classify any new instance whether the seminal quality is altered or not. The observed accuracy of the model is highest (91%) whereas the previous highest accuracy was reported to be 86% only.

Keywords: Smoking, Social Factors, Machine Learning, Male Infertility, Fecundity, Seminal Quality.

1. Introduction

It has been reported that seminal quality has been deteriorated in last two decades [1]. Same has been reported with particular emphasis on Western and Asiatic counties in association with environmental exposures [2]. There is escalating evidence that public with infertility complaints are looking for treatment in numerous countries. Poor fertility is a matter of profoundly and private condition of concern. This phenomenon has been investigated across various domains of interest including psychosocial and bio-behavioral phenomenon. The immediate reason behind infertility is testicular malfunctioning resulting into production of unhealthy sperm or sperm in non sufficient quantity.

The interactions between biological vs. behavioral pattern and psychological vs. social pattern have been discovered with cause and effect paradigm in the underlying reason of infertility. Furthermore researchers in machine learning have also shown that intelligent heuristics can predict the level of fertility [1]. Human Fertilization and Embryology Authority has pointed out an increase in medically assisted reproduction (MAR) for infertility. The public awareness has turned infertility related matters into public health concern. In this study we have proposed an etiological computational model for establishing the relationship between the parameters governing the infertility in men. The research may be useful in identifying research priorities comprehensively along with its underlying significant pragmatic implications. There are numerous factors responsible for infertility which includes the unwanted effects of treatments for cancer, environmental and behavioral habit such as large number of hours spent in sitting, smoking, use of alcohol and dietary behaviors etc. Some of these causes have permanent effect resulting into prevention less problem; however some of the causes are temporary in their nature and a switch in lifestyle may ameliorate the problem to a great length. In fact, the temporary causations are mostly related to poor reproductive health outcomes. It was reported that a substantial milieu exist between the socio cultural aspect and treatment for human infertility. It was demonstrated that revisions in life style can substantially effects in improving the fertility and it is required that psychosocial

and behavioral elements must be integrated.

A similar work to this study was carried out in [1] where it was elucidated that multilayer perceptron and support vector machines are the most accurate in predicting seminal quality. The reported accuracy was only 86%. Moreover, it was suggested that both of these techniques can be used together with the decision tree. However, while arguing over it, we have demonstrated that BBN is better than this approach. Firstly, according to our experiment, it delivers accuracy up to 91% as if Hill Climbing is selected for its search beam. Secondly, its concrete output is a Directed Acyclic Graph (DAG) which can accommodate very simple to much complicated interaction between the features of underlying dataset. While using its DAG, we can easily deduce on the probabilistic inference between the features and the corresponding designated class .

The remaining paper is organized into four sections including materials and methods, seminal quality, result and discussion.

2. Materials and Methods

Numerous analyses of seminal attributes have been introduced in literature; however in all of these existing analyses, conventional statistical measures have been employed. With the advent of advanced computational techniques, more sophisticated heuristics have been brought forward. Machine learning or data mining is one of such techniques aimed towards analysis of observational data to dig out unsuspected useful information in shape of patterns or data models. Statistical techniques usually look for exhaustive searches; however such methods embrace their limitation as data size expands horizontally or vertically. Data mining has the potential to cope up this limitation in the array of more prohibitive and computationally exhaustive problems. This motivates us to analyze a clinical dataset using one of the state of the art data mining techniques known as Bayesian Belief Network (BBN). BBN is a notable technique for graphical representation of probabilistic relationship between random variables [3]. Statistical models typically show up the tendency of deviation as they face the data loss phenomenon; however BBN has proved its resilience nature in this phenomenon. BBN deals in this problem by summing up or integrating the probabilities of possible data values. BBN modeling has its advantages to its peer modeling techniques with following well-known distinctiveness [4].

- Once the BBN modeling is formulated, novel features can be easily added in perspective of the established structure.
- The outcome model is suitable for handling incomplete data with uncertainty.
- The most prominent aspect of BBN is its capability of being a robust reasoning model apart from its expressive probability model.

The graphical representation of BBN is represented by a Directed Acyclic Graph (DAG). The nodes in DAG are realized by the attributes in a given data [4]. Each node is conditioned with its parental node attributes while it is conditionally independent of the subset of descendent attributes. The casual representation between any two node attribute is denoted by the arcs, a reason by virtue of which BBN is known to be a casual network.

3. Seminal Quality

World Health Organization has standardized the procedure of semen analysis. They have introduced four nomenclatures which are as below.

- Normozoospermia, : all values of semen parameters are on the upper level of the lower limit.
- Asthenozoospermia: A fraction of progressively motile spermatozoa below the lower limit.
- Oligozoospermia: All of spermatozoa below the lower limit.
- Teratozoospermia: A fraction of morphologically normal spermatozoa below the

lower limit.

In this study, 38% of samples bear altered semen. A further break up of these altered semen included asthenozoospermia with 47.37%, oligozoospermia with 21.05%, astenoteratozoospermia and oligoasthenozoospermia appeared to combined percentage of 10.53%, The least frequent alteration was teratozoospermia with 2.63% of the altered samples.

4. Result

The fertility diagnosis dataset comprised of 100 cases (records) and 10 attributes (features). It is a mandatory requirement for BBN to designate one of the data attributes as its outcome class with which an inference upon any of the other attribute is possible. In this study the outcome class is a binary attribute 'Diagnosis' with two possible states normal semen or altered semen. Among rest of 9 attributes, 3 are Boolean in nature which includes childish diseases (such as chicken pox, measles, mumps and polio), accident or serious trauma and surgical intervention.
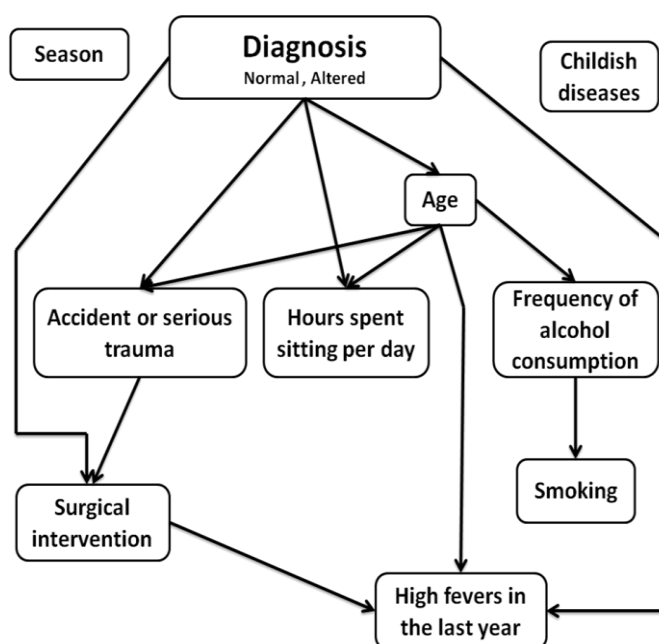

Fig. 1: Probabilistic Model Perceived from Attributes for Etiological Evaluation of Seminal Traits using BBN

Two attributes 'Age' and 'number of hours spent sitting per day' (NHS) were numeric. In order to increase the probabilistic explanation of these two attributes in the learnt model, we discretize both of them into two bins. The quality of final learnt structure is also a factor of discretization [5], a detail of which can be found in [6]. Rests of the attributes were categorical having distinct states of 3 or more. The outcome of BBN is a learnt structure in form of a DAG. Each node in DAG contains its information of reasoning and induction in a structure known as Conditional Probability Table (CPT) which holds a probabilistic detail of its relationship with the other attributes.

Fig. 1 indicates that variable NHS is conditioned with age and diagnose variable. It means NHS receives causations from age and diagnosis. While translating first record in this CPT, the sample space of male with normal seminal quality and age bracket of 18 to 23 years is distributed between two states of NHS. A probability of 75% is observed in favors of 6 to 16 hours sitting hours value for NHS. Another significant value of 75% is for the male of age group of 18 to 23 years with altered seminal quality for NHS range of 6 to 16 hours.

5. Discussion

The model has distinguished five variables which receives direct causations from diagnosis node. All of these five attributes exhibit simple two dimensional CPT. whereas the attributes which are linked to more than a single variable, generate relatively large CPT albeit the complexity of explanation remains simplistic which is a peculiar characteristics of this modeling systems.

Table 1. Conditional Probability Distribution Table for Number of Hours Spent in Sitting per Day

Output	Age (years)	Less than 6 hours	6 to 16 hours
Normal	18 to 23	0.25	0.75
Normal	24 to 36	0.648	0.352
Altered	18 to 23	0.25	0.75
Altered	24 to 36	0.292	0.708

5.1 Age

Age variable is distributed in two bins (see table 1 for detail). In literature, the association between age of men and their semen quality has been investigated by numerous researchers. Some researchers show partial relationship between these two parameters where it was described that sperm count and age has a negative correlation while no substantial relationship was found between age and other parameters of semen such as rate of fertilization, birth rate and rate of pregnancy [7]. It was pointed out that age has diminutive influence over male reproductive potential [7]. A general belief about link of age towards seminal quality is negative correlation. The seminal quality includes semen volume, concentration, motility and morphology [8]. The significance of outcome of age, smoking, obesity and intercourse timing on fertility was analyzed and the results were incorporated into educational campaign in Australia. Two opinions have been established. In first opinion, decline in men's power of fertility has been reported with increasing age. Other opinion is contradicting the previous one [9]. Other than typical statistical measures such as standard deviation, mean, correlation; some more advanced techniques were also used such as multiple regression analysis to tweak out the relationship between fertility and age factors. According to [9] children of elderly father indicated a relatively higher degree of risk of autosomal dominant disease. Our model (learnt structure) indicates a single significant result according to which a probability of 88.5% was observed in case of 24 to 36 years old male given the altered semen attributes.

5.2 Smoking

Smoking is a worldwide habit rather a part of culture. It is reported that smoking can cause carcinogenetic effect as well as somatic cell mutagenic effects or Y-chromosome deletion [10]. A literature survey was carried out; according to which it was reported that majority of research pointed out smoking trim down sperm production, fertilization and its motility by means of altering its normal form with evidence of increased seminal oxidative stress and chromosomal aberrations [11]. Other researchers also provided evidence in support of association between smoking and its lethal effect on seminal quality resulting into disomic sperm and diminution in semen quality in particular aspects [10-11]. Although a few report pointed out trivial relationship of smoking on somatic cells [11]. A limited relationship between smoking and integrity of DNA of spermatogenic cells was reported in [12] according to which a statistically significant difference was observed between regular smokers and never smokers.

Reference [13] presented their analysis of smoking on spermatozoa mobility and Ca2+-ATPase activity. Ca2+ is regarded an essential regulator of sperm motility, capacitation and

in the commencement of the acrosome reaction course. According to [13] cigarette smoking cast its affects on Ca2+-ATPase activity and motility of the spermatozoa.

The smoking habit is conditioned with alcohol consumption as shown by the fig. 1. The probability distribution indicates one strong causation rule that never smoker akin to never or hardly ever drinker (with 73.5% probabilistic certainty). Apart from this, in the light of learnt model from the given data, it can be illustrated that the given data do not conclusively demonstrate any significant effect of smoking habit in conjunction with alcohol consumption. The only feeble sense realized from the model was that both smoking and alcohol can be regarded as risk factor for deteriorating seminal quality.

5.3 Hours Spent Sitting

The duration of sitting during a sedentary nature of job is also investigated into this study in perspective of its effect on seminal quality. It has been reported that the temperature and humidity accumulated around scrotal skin has its relation towards semen production. An evaluation report was presented in [14] in which an elevated temperature up to a difference of 3 degree centigrade was reported when male used non saddle chairs. This increase in scrotal temperature was termed as one of the reason in reducing seminal quality [14]. Later on excessive usage of laptop also become a rudimentary part of working culture. This again motivated the researchers to investigate the scrotal hyperthermia in laptop users. The issue is aggravated with closely approximated legs [15]. Reference [16] introduced a recent study; where a direct relationship between hyperthermia and spermatozoa was speculated; albeit an in vitro study is still required to validate this speculation.

Reference [17] investigated the scrotal temperature with emphasis on specific postures during sleeping hours. The authors divided the results into two general categories, the first was period of physical activity and the second was period of physical inactivity. During the first category, it was noted down that a feeble relationship of high scrotal temperature was observed during left or right side postures during sleep. However in general, there was no close correlation observed between scrotal temperature and semen quality [17].

Prior to this research, it was demonstrated that there is a fragile and contradictory relationship observed between sedentary positions and seminal quality [18]. Moreover a negative correlation was reported between high scrotal temperature and sperm count. It was concluded that sedentary work position during common working hours does not cast any serious effect on semen quality [18].

Regarding accident or serious trauma, men with age range from 24 to 36 and suffering from accident or serious trauma were having the probability of 79.2% with altered semen whereas in the lower bracket of age group, trauma has little effect on seminal quality.

The variable 'surgical intervention' was found correlated with diagnose class and accident or serious trauma. This variable was conditionally independent of all other variables. The distribution of altered seminal quality along with both of these variables was such that surgical intervention with value of 'yes' and accident or serious trauma value as 'no' leads to a probability of 87.5%. Whereas if we swap the states of both of these variables such that surgical intervention with state of 'no' and 'accident or serious trauma' with state of 'yes' then probability was 75%.

5.4 Disease

This study also demonstrates the marked effects of high fever on the integrity of seminal parameters. Reference [19] pointed out that the febrile tenure is significant enough to be considered during counseling of infertile couples. The same was elucidated in [20] according to which they enlarged the canvass of the problem in perspective of other related factors such as stress and psychological factors. Earlier in 1999, it was argued over the relationship of feverish conditions towards fertility and illustrated that gonorrhea can characterize the secondary subfertility [21].

The variable 'High Fever in last Year' was observed conditional with three variables: Age, Diagnosis and Surgical intervention (see figure 1 for casual relationship). Its CPT overall does not provide any hard verdict in favor of any outcome rule of these four variables. However, generally it can be concluded that surgical intervention and high fever in last year plays its role for older age group in increasing the risk for deteriorating male fertility.

5.5 Alcoholism

Another important parameter with detrimental effect on seminal quality is alcoholism. Reference [22] reported adverse effects of chronic alcoholism on male fertility hormones whereas a prior research conducted in [12] shows that there is no relationship between sperm quality and alcohol consumption or intake of coffee. Our model also demonstrates that there is no sound relationship of alcoholism with respect to semen quality amelioration or deterioration. However, this attribute was found correlated (conditioned) with age.

An interesting research in the effect of alcohol was demonstrated in [23] according to which the relationship between semen quality, its underlying reproductive hormonal system and alcohol was tweaked. It was pointed out that the association is only observed if higher intake of alcohol with consecutive days is carried out with tendency of lower semen quality [23].

Two variables 'season' and 'child disease' were observed to be independent of all other variables including the diagnosis class. This shows that both of these two variables have least rather no impact on seminal quality. The fact was realized as no data correlation was observed for these two variables with respect to diagnosis and any other variables.

6. Conclusion

There are numerous factors identified which have been individually discussed, however it is required to model all of these confounding parameters adequately to dig out the hidden relationship between these variables. Motivated by this fact, we investigate the best possible data modeling technique BBN which can identify the most realistic trend among these variables. Bayesian Belief Network due to its strong probabilistic modeling capability is quite useful heuristics to predict the seminal quality from individual profile dataset. The study has shown that BBN can outperform other notable techniques in machine learning. Moreover, its visual graphical representation is easy to interpret the underlying relationship between the psychological, behavioral and surgical attributes responsible for the infertility. It has been shown that in comparison to conventional statistical measures, BBN is quite useful in analysis of seminal quality as well as classifying an unknown instance.

References

[1]   D. Gil, J. L. Girela, J. De Juan, M. J. Gomez-Torres and M. Johnsson, "Predicting Seminal Quality with Artificial Intelligence Methods", Expert Systems with Applications, vol. 39, no. 16, (2012), pp.12564-12573.

[2]   G. Ji, Y. Xia, A. Gu, X. Shi, Y. Long, L. Song, S. Wang and X. Wang, "Effects of non-occupational environmental exposure to pyrethroids on semen quality and sperm DNA integrity in Chinese men", Reproductive Toxicology, vol. 31, no. 2, (2011), pp. 171–176.

[3]   M. Naeem, and S. Asghar", A Parameter Free BBN Discriminant Function for Optimum Model Complexity versus Goodness of Data Fitting", Journal of Applied Research and Technology, vol. 12, no. 4, (2014), pp. 734-749.

[4]   M. Naeem and S. Asghar, "Structure learning via non-parametric factorized joint likelihood function," Journal of Intelligent and Fuzzy Systems, vol. 27, no. 3, (2014), pp. 1589-1599.

[5]   M. Naeem and S. Asghar, "An Information Theoretic Scoring Function in Belief Network", The International Arab Journal of Information Technology, vol. 11, no. 5, (2014), pp. 459-467.

[6]   M. Naeem and S. Asghar, "A novel mutual dependence measure in structure learning", J. Natl. Sci. Found. Sri Lanka, vol. 41, no. 3, (2013), pp. 203-208.

[7]   R. J. Paulson, R. C. Milligan and R. Z. Sokol, "The lack of influence of age on male fertility," American Journal of Obstetrics and Gynecology, vol. 184, no. 5, (2001) , pp. 818-24.

[8]   S. A. Kidd, B. Eskenazi and A. J. Wyrobek, "Effects of male age on semen quality and fertility: a review of

the literature", Fertility and sterility, vol. 75, no. 2, **(2001)**, pp. 237-248.

[9] E. Plas, P. Berger, M. Hermann and H. Pflüger, "Effects of aging on male fertility?" Experimental Gerontology, vol. 35, no. 5, **(2000)**, pp. 543-551.

[10] M. F. Vine, "Smoking and male reproduction: a review", International Journal of Andrology, vol. 19, no. 6, **(1996)**, pp. 323-337.

[11] T. Mostafa, "Cigarette smoking and male infertility", Journal of Advanced Research, vol. 1, no. 3, **(2010)**, pp. 179-186.

[12] S. Horak, J. Polanska, P. Widlak, "Bulky DNA adducts in human sperm: relationship with fertility, semen quality, smoking, and environmental factors", Mutation Research/Genetic Toxicology and Environmental Mutagenesis, vol. 537, no. 1, **(2003)**, pp. 53-65.

[13] T. A. Kumosani, M. F. Elshal, A. A. Al-Jonaid and H. S. Abduljabar, "The influence of smoking on semen quality, seminal microelements and Ca2+-ATPase activity among infertile and fertile men", Clinical Biochemistry, vol. 41, no. 14, **(2008)**, pp. 1199–1203.

[14] R. Koskelo, N. Zaproudina and K. Vuorikari, "High scrotal temperatures and chairs in the pathophysiology of poor semen quality", Pathophysiology, vol. 11, no. 4, **(2005)**, pp. 221-224.

[15] Y. Sheynkin, R. Welliver, A. Winer, F. Hajimirzaee, H. Ahn and K. Lee, "Protection from scrotal hyperthermia in laptop computer users", Fertility and sterility, vol. 95, no. 2, **(2011)**, pp. 647-651.

[16] C. Avendano, A. Mata, C. A. Sanchez Sarmiento and G. F. Doncel, "Use of laptop computers connected to internet through Wi-Fi decreases human sperm motility and increases sperm DNA fragmentation", Fertility and sterility, vol. 97, no. 1, **(2012)**, pp. 39-45.

[17] A. Jung, J. P. Hofstötter, H. C. Schuppe, and W. B. Schill, "Relationship between sleeping posture and fluctuations in nocturnal scrotal temperature", Reproductive Toxicology, vol. 17, no. 4, **(2003)**, pp. 433–438.

[18] N. H. I. Hjollund, L. Storgaard, E. Ernst, J. P. Bonde and J. Olsen, "Impact of diurnal scrotal temperature on semen quality", Reproductive Toxicology, vol. 16, no. 3, **(2002)**, pp. 215–221.

[19] M. Sergerie, R. Mieusset, F. Croute, M. Daudin and L. Bujan, "High risk of temporary alteration of semen parameters after recent acute febrile illness", Fertility and sterility, vol. 88, no. 4, **(2007)**, pp. 970-e1.

[20] E. Hall and V. K. Burt, "Male fertility: psychiatric considerations", Fertility and sterility, vol. 97, no. 2, **(2012)**, pp. 434-439.

[21] O. Akre, S. Cnattingius, R. Bergström, U. Kvist, D. Trichopoulos and A. Ekbom, "Human fertility does not decline: evidence from Sweden", Fertility and sterility, vol. 71, no. 6, **(1999)**, pp. 1066-1069.

[22] K. R. Muthusami, and P. Chinnaswamy, "Effect of chronic alcoholism on male fertility hormones and semen quality", Fertility and sterility, vol. 84, no. 4, **(2005)**, pp. 919-924.

[23] M. L. Hansen, A. M. Thulstrup, J. P. Bondeb, J. Olsenc, L. B. Håkonsen and C. H. Ramlau-Hansen, "Does last week's alcohol intake affect semen quality or reproductive hormones? A cross-sectional study among healthy young Danish men", Reproductive Toxicology, vol. 34, no. 3, **(2012)**, pp. 457-462.