Personalization of Infectious Disease Risk Prediction

Towards automatic generation of a Bayesian Network

Retno Aulia Vinarti and Lucy Hederman
School of Computer Science and Statistics
Trinity College Dublin, The University of Dublin
Dublin, Ireland
e-mail: {retnor, hederman}@scss.tcd.ie

Abstract— **Infectious diseases are a major cause of human morbidity, but most are avoidable. An accurate and personalized risk prediction is expected to alert people to the risk of getting exposed to infectious diseases. However, as data and knowledge in the epidemiology and infectious diseases field becomes available, an updateable risk prediction model is needed. The objectives of this article are (1) to describe the mechanisms for generating a Bayesian Network (BN), as risk prediction model, from a knowledge-base, and (2) to examine the accuracy of the prediction result. The research in this paper started by encoding declarative knowledge from the Atlas of Human Infectious Diseases into an Infectious Disease Risk Ontology. Automatic generation of a BN from this knowledge uses two tools (1) a Rule Converter generates a BN structure from the ontology (2) a Joint & Marginal Probability Supplier tool populates the BN with probabilities. These tools allow the BN to be recreated automatically whenever knowledge and data changes. In a runtime phase, a third tool, the Context Collector, captures facts given by the client and consequent environmental context. This paper introduces these tools and evaluates the effectiveness of the resulting BN for a single infectious disease, Anthrax. We have compared conditional probabilities predicted by our BN against incidence estimated from real patient visit records. Experiments explored the role of different context data in prediction accuracy. The results suggest that building a BN from an ontology is feasible. The experiments also show that more context results in better risk prediction.**

Keywords; Bayesian Network, Risk, Personalised Prediction, Infectious diseases

I. INTRODUCTION

Infectious disease is listed as a major cause of human morbidity [1]. However, many infectious diseases are avoidable. There is a need for accurate prediction related to infection risk [2] so that those at risk can take appropriate avoidance precautions. Infectious disease risk is the result of interaction between a person, pathogenic agent[1] and environment [3], The purpose of this research is to provide people with personalised information about their risk of being infected by a disease. The research faces several challenges such as (1) the continuous update of knowledge of predictors of risk, (2) the limited sources of experimental and observational data which are costly to retrieve in full, (3) the need for data, contexts and knowledge to describe a complex situation that resembles the real world about risk of getting an infectious disease, (4) establishing relevant contextual data from a user's details and location.

There has been extensive research in algorithms and techniques for risk prediction, such as fuzzy logic [4,5], Bayesian network [6,7], logistic regression [8-10] or combinations between them [11,12]. In general, a Bayesian network offers flexibility to incorporate dependencies between variables by defining probabilistic relationships [13], works under incomplete data [14] and its results are highly accurate compared to other prediction methods [5].

For this research, using a Bayesian Network approach, knowledge is represented as a network during the knowledge building phase. In this phase, person, season, weather and location risk factors are identified from declarative knowledge sources. Each factor is filled by a parameter value provided by a United Nations Data (UN) API. Then, predictors are transformed into BN nodes and states. After the network is created, the risk is calculated during the runtime phase. In this phase, person-related facts are given by the user. Weather and environment details are also retrieved based on user's location from OpenWeather API and GoogleMaps API. All the retrieved facts are used to yield the personalized infectious disease risk prediction.

A recent innovation of Bayesian Networks to model relationships between variables is Dynamic Bayesian Networks (DBN) [15]. A DBN approach has been used in prediction [16-18] and diagnosing various problems [15] [19] in medical, supply chain and banking cases. Most of the temporal statements in a DBN are the result of machine learning [16-18] while a few of them use manual acquisition from domain experts [15]. A DBN is built by adding temporal dependencies to a static BN [19]. Machine learning approach works well when there is plentiful training data to construct stronger [17] and simpler static BN [18].

In this paper, we adopt the dynamicity concept of a DBN by designing tools that allow refining a BN based on newest information stored in a knowledge-base. However, for infectious disease risk prediction, we have well-established knowledge, encoded in sources such as the Atlas of Human

---

[1] Agent is a microorganism (e.g. fungus, bacteria, virus, prion, parasite or mix) that cause illness.

Infectious Diseases [20], Centres of Disease Control and Prevention [21], Health Protection Surveillance Centre [22] and the Infectious Disease Ontology [23]. Also, we use probability data summarised in the United Nations Statistical Division (UNSD) for demographic parameters and World Health Organization (WHO) for prevalence rates. Thus, machine learning is not appropriate. Instead we seek to directly convert these sources into a BN.

Rules are versatile and have been widely used in health systems [24]. In terms of knowledge-base, rules are usually built on the predefined knowledge structure (e.g. Ontology). This research started by gathering knowledge from various forms of knowledge sources to build an Infectious Disease Risk ontology as base of rules. Thereafter, a knowledge engineer adds disease-centred rules which explain about personal and environmental contribution to the specific infection risk. A tool to auto-create a BN based on the newest rules is then executed. This paper also presents mechanisms and evaluation of the BN by measuring the accuracy of the risk prediction. The risk prediction results are compared to ground truth taken from patient records collected by hospitals in a county in US during a specific time.

The rest of the paper is organized as follows. Section 2 explains an infectious disease risk prediction service where the BN is the key of the service; this is the grand design of the whole research. Section 3 describes the design of tools that are needed to build the system defined in Section 2. The tools are Rule Converter, Joint and Marginal Probability Supplier (JMP) and Context Collector. The Rule Converter and part of the JMP has been created and explained in another related-article [25]. Section 4 presents a process to prepare "ground truth" data used to measure the risk prediction result. Section 5 explains the method to examine prediction accuracy and presents the results. Section 6 reviews conclusions and suggests possible future work.

II. THE PERSONALIZED INFECTIOUS DISEASE RISK PREDICTION SERVICE

The infectious disease risk prediction web service is designed to serve client applications which advise users when and how to protect themselves from infections. The service computes a person's risk of being infected by a specified disease in a specific time frame (may differ for different infections), given their demographic details and location. The service uses the location to find weather, season and environmental features (e.g. swamp, forest, river). This process works at a runtime using the latest version of BN model.

This section explains the components of the service that are needed during knowledge building phase: (1) an ontology that describes Infectious Disease risk; (2) rules to represent the relationships between risk predictors and infectious diseases; the rules are taken from declarative knowledge in Atlas of Human Infectious Diseases (AHID) and represented in the Semantic Web Rule Language (SWRL); finally, (3) a

BN representing the newest knowledge is constructed. The prediction accuracy of this BN is what is evaluated in the third section.

A. Overview of Ontology and Rule Structure

An Infectious Disease Risk (IDR) Ontology (shown in Figure 1) was developed to describe the interaction between human and environment in the context of infectious disease risk. Existing ontologies related to this subject have been studied and reused [23]. Infectious Disease Ontology and Epidemiology Ontology were used as references. Since their focus are not on the *risk prediction* task, the IDR was created.

The relevant concepts were established by studying the AHID, identifying and organising all predictors which might be useful in determining a person's risk[2]. A knowledge-engineer then translated all predictors into classes and sub-classes in the IDR ontology, and rules defined over that ontology. Rules are used to describe predictors' contribution to a disease risk.

In the IDR, Environment is defined by three classes: Location, Climate and Feature details. These explain a person's surroundings, for example the weather and season at the access time, also the location and its feature details (e.g. near woods, river) at a given geocode. The Climate (e.g. weather, season) is retrieved from OpenWeatherMap API while the Feature details of a person's specific location is retrieved from GoogleMaps API.

A person is represented by his demographic details and his surroundings (exact position, nearest features and climate at the access time). In the IDR ontology, this is illustrated by three arrows pointing out of the Person class. An infectious disease risk is mostly determined by the disease prevalence in a specific location, its feature details and the climate condition. Some weather or season may boost or limit certain pathogens [26-30]. In figure 1, this is illustrated by three arrows pointing in to the Infectious Diseases class. Then, a person's risk of being infected by a specified disease today, given their demographic details and environment is illustrated by a line pointing into Person class which is labelled 'risk of'.
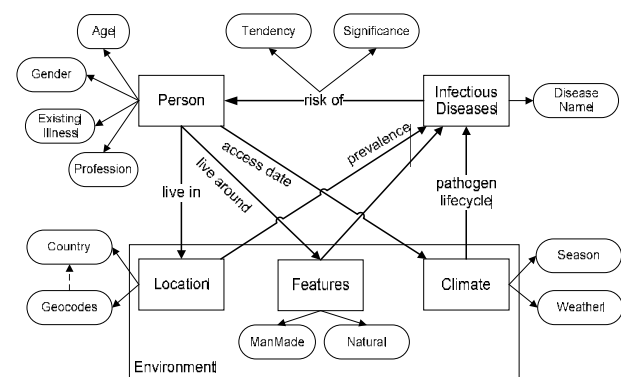

Figure 1 The infectious disease Risk (IDR) Ontology

595

TABLE 1 SAMPLE RULE ENCODING FOR ANTHRAX RISK

Declarative Knowledge	SWRLs
Farmers or people who live close to livestock farms are at high risk. Non-vegetarians are at medium risk.	`Person(?allpeople) ^ hasProfession(?allpeople, Farmers) -> hasRiskOf(Anthrax, high)`
	`Person(?allpeople) ^ liveAround(?allpeople, Farms) -> hasRiskOf(Anthrax, high)`
	`Person(?allpeople) ^ hasHabits(?allpeople, EatingMeats) -> hasRiskOf(Anthrax, medium)`
Anthrax pathogens are dormant during winter and the bacteria are naturally found in soil or grass in US. They are easily spread by the wind.	`Person(?allpeople) ^ liveIn(?allpeople, US) ^ accessDuring(?allpeople, Summer) -> increaseRisk(Anthrax,x)`
	`Person(?allpeople) ^ liveIn(?allpeople, US) ^ accessDuring(?allpeople, Winter) -> setRisk(Anthrax,0)`
	`Person(?allpeople) ^ liveIn(?allpeople, US) ^ accessDuring(?allpeople, Windy) ^ accessDuring(?allpeople, Summer) -> increaseRisk(Anthrax,y)`

Rules facilitate statements in first-order logic about how predictors impact infection risks. The common composition of rules is antecedent/predictor (A), consequent or infectious disease (B) and denoted as (A → B). An antecedent can be formed by individuals (in classes or sub-classes) and object/data properties with AND (^). The narrative knowledge written in AHID determines the individuals and their impact on a disease. The impact includes odds ratios and tendency. A knowledge engineer creates rules to represent all knowledge written in declarative source [20-22]. Table 1 shows sample rules encoding various tendency levels given by certain predictors of Anthrax risk (e.g. high, zero, x, y). For example, `increaseRisk(Anthrax, 200)` would mean the risk of Anthrax is doubled in Summer.

B. Overview of BN

The Bayesian Network is the core risk prediction model for the service. The BN is generated using Rule Converter that takes the newest knowledge in the IDR ontology and turns it into nodes, states and Conditional Probability Tables (CPTs) through several procedures. In figure 2, nodes are represented by a table, whereas states and CPTs are listed each node.
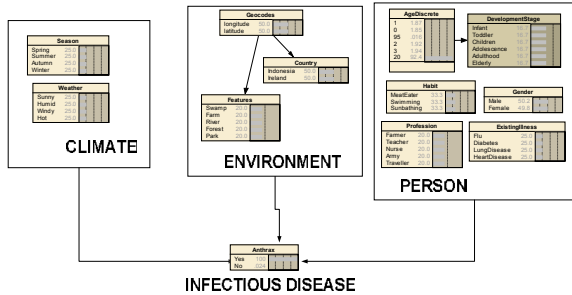

Figure 2. Structure of the Bayesian Network corresponding to Figure 1

To improve accuracy of prediction and as the knowledge develops, more predictors may need to be taken into consideration. In the example, to predict Anthrax risk, the major predictors like Age and Gender need to be taken into the initial risk prediction model. Then, as data and information become more available, more predictors like Profession,

Habit, refine the initial model and make the Anthrax risk prediction become more accurate.

The BN consists of predictors (upper blocks), in three node groups: Person, Climate, and Environment, which contribute to determine the infectious disease(s) risk in a lower block. Nodes are taken from sub-classes, whereas states are taken from individuals in each of the sub-classes. Links between predictors within a group are taken from object or data properties, while links between predictor and infectious disease are taken from SWRL rules. Figure 2 depicts the generated BN from IDR in figure 1.

III. DESIGN OF TOOLS

The Rule Converter tool is used to translate the SWRL rules (Table 1) and IDR ontology (Figure 1) into a BN (Figure 2). Netica-J [31], a Java API for Bayesian Network modeller, is used to calculate the Infectious Disease risks. Netica-J allows a developer to create, modify and connect the package with other APIs. By using this API, the web service will be flexible and easy to be maintained in the future. First, node and states labels are identified by querying sub-class and individuals for each sub-class in the generated RDF[3]. Then, object and data properties are queried to define links between predictor nodes. All properties related to rules are then retrieved to define links between predictors and associated infectious disease. These steps are used to define network properties and are executed using SPARQL[4].

After knowing the network properties, several Netica built-in functions (presented in Script 1 and 2) are used to auto-create the BN structure.

```
Node age = new Node ("Age", "Children, Working, Elderly", net);
Node influenza = new Node ("Flu", "None, Low, High, Decline", net);
Node location = new Node ("Region", "North, South, Central, East, West", net);
Node anthrax = new Node ("Anthrax", "AtRisk, NotAtRisk", net);
```
Script 1 Code Example to generate BN nodes and states from IDR

```
anthrax.addLink(age);
anthrax.addLink(influenza);
anthrax.addLink(location);
```
Script 2 Code Example to generate BN link from IDR

---

[3] RDF (Resource Description Framework) is a model for encoding semantic relationships between items of data so that these relationships can be interpreted computationally.

[4] SPARQL (Simple Protocol and RDF Query Language) is a standard query language to retrieve and manipulate data stored in RDF format.

596

The joint and marginal probabilities are the core of the risk prediction values which will be delivered to the clients. These probabilities are aimed to be populated in each CPT using a Joint and Marginal Probability Supplier (JMP).

To fill in the parent nodes' CPTs, data is downloaded from the United Nations Data web service in the form of SDMX[5]. Then the values are placed in each associated state in a node. Whereas, the child node's CPT is filled by taking numerical arguments and property predicates from the SWRL rules. The probability of a person with their demographic details for the risk of contracting Anthrax (e.g. 0.2, 0.25, 0.35) is assigned in each line in Script 3.

```
anthrax.setCPTable("Children, Low, North", 0.004, 0.996);
anthrax.setCPTable("Working, High, West", 0.0001, 0.9999);
anthrax.setCPTable("Elderly, None, South", 0.0008, 0.9992);
```
Script 3 Code Example to populate CPT

While the Rule Converter and JMP are used for preparing the BN using the newest knowledge, the Context Collector performs the personalization part of the service at runtime. The Context Collector collects facts given by the clients related to demographic and location details.

The collection of clients' facts (e.g. gender, age and current location) are taken as beliefs in the risk calculation process. The collections of a client's facts are exemplified as follows: 3-year old females located at (82.1673907, -168.9778799) are looking for the risk of Anthrax on that day (4th of July 2015). The current weather, season and country name will be automatically retrieved based on access date and time. Then, all the retrieved contexts will be used to calculate the Anthrax risk using Script 4.

```
location.finding().enterState (locationstr);
gender.finding().enterState(keyhumid);
age.finding().enterState(agestr);
double beliefA = anthrax.getBelief ("AtRisk");
```
Script 4 Code Example to enter the known facts into BN

IV. VALIDATION

In order to prove the concept of using a knowledge-base to construct an associated BN, a validation step has been carried out. This research uses patient data to provide the conditional probabilities, whereas the predicted probabilities are generated from the BN. By measuring the distance between conditional probabilities taken from patient visit records and prediction probabilities, the accuracy of the prediction can be observed.

We used two datasets: patients' visit records (whose metadata is given in Table 2) and Anthrax outbreak dates[6] [27,28]. The patient data was taken from Emergency Department of health care units located in Allegheny County, Pittsburgh, US., The datasets present the visit records from

January 2002 until December 2003 with 38,596 in total. The datasets contain demographic details, visit time, and the reported symptoms. These visit records might contain any possible diseases other than Anthrax, thus, another dataset is needed to help identify Anthrax cases within the dataset. The Anthrax outbreak detection project, WSARE[7], supplies this need by giving the estimation of Anthrax occurrences (date). Knowledge taken from an interview with a GP was also needed to determine the Anthrax key symptom (e.g. Rash).

TABLE 2 PATIENT METADATA

Contexts	Field Names	Data Types
Person	Age	child, adult, senior
	Gender	male, female
	Flu	none, low, high
Climate	Weather	hot, cold
	Season	winter, spring, summer, fall
	Date	mmm-dd-yyyy
Location	Region	N, W, C, E, S
Inf. Disease	Reported_symptom	none, respiratory, nausea, rash

Thus, conditional probabilities are estimated from the patient visit records by following these steps: (1) take the Anthrax outbreak data (date) and add the incubation (7 days) and prodromal periods (46 days), (2) select the patients who present to ER within that period (18,527 records), (3) select patients who have 'Rash' as reported symptom (3,112 records), (4) count Anthrax patients for each demographic combination (e.g. elderly females living in South Allegheny County) (5) divide the counts by the corresponding proportion of total population of the region. These probabilities are taken as "ground truth".

The BN for predicting Anthrax risk for this study was built from knowledge in AHID, medical journals, and an interview with a GP. Some marginal probability values (e.g. the probability of female children in a given location) were obtained from WHO and UNSD. To fill the CPTs, a state's value in each node must be acquired (see Figure 1). The Age CPT values were filled by summarizing age structure information from UNSD. Since Influenza was the only illness reported as occurring in conjunction with Anthrax - the risk of Anthrax increases when patients have Influenza – the Existing Illness node only handles Influenza, The basic knowledge about Anthrax risk and Influenza occurrence is acquired using a probability table of reported_symptoms [35-37] and interview with GP. Seasons are represented by Northern Meteorological term (e.g. Winter begins on 1 Dec). Interestingly, the ground truth data used astronomical seasons. To obtain Location context, a deeper investigation of location-related factors needed to be carried out. The marginal probability for the Location at which a Person lives (e.g. N, W, E, S) is obtained by summarizing the population

---

[5] SDMX (Statistical Data and Metadata eXchange) is a standard designed to describe statistical data and normalise their exchange.

[6] Both datasets were downloaded from https://www.autonlab.org/autonweb/15959.html?branch=1&language=2 [Accessed 21/05/2015] but now the web is under construction.

[7] WSARE (What's Strange About Recent Events) is a project to early detection of disease outbreaks by searching a database of emergency department cases for anomalous patterns.

distribution from Allegheny County Information Portal [29]. Another Location factor is the regional distribution of Anthrax (e.g. farms). The marginal probability was obtained by summarizing the number of farms per region and dividing by total farms in Allegheny. This information was manually obtained from Allegheny County Farm Land website [38]. The regional contributions to Anthrax environmental risks are estimated as North (30%), West (13%), South (26%), East (15%), Central (13%).

Infectious disease node (Anthrax) is a child node. Consequently, the child node contains all combinations of states from its parent nodes. Therefore, other information needed to build the Anthrax CPT includes:

a. Children are susceptible to Influenza more than Anthrax, while for a working Adult it is vice versa.
b. Influenza occurs in all seasons, while Anthrax is dormant in Winter and peaks in Summer.

By encoding these rules in the BN, the probability of a person, with his details and contexts, being infected by Anthrax in a specific given time and place, can be estimated.

V. EVALUATION METHOD AND RESULTS

To measure the performance of the BN prediction result, Root Mean Square Error (RMSE) is a common measurement used to assess the prediction accuracy [39]. The lower the errors the better the prediction. In this case, the BN prediction result will be compared against the "ground truth" derived from ED patient visit records (i.e. actual).

$$RMSE = \sqrt{\frac{\sum (prediction - actual)^2}{n}}$$

The RMSEs are categorised by each context in Table 3 below. Thus, n is the number of attribute value combinations.

TABLE 3 PREDICTION ACCURACY GROUPED BY CONTEXTS

Contexts	Context Details	RMSE	Norm RMSE	n
One-context Combination				
Person	Age	0.006006	0.698444	3
	Flu	0.000736	**0.287605**	4
	Age, Flu	0.004341	0.359710	12
Climate	Season	0.000910	0.386974	4
Location	Region	0.007810	0.458550	5
Two-context Combination				
Person – Climate	Age – Season	0.003114	0.314958	12
	Flu – Season	0.001610	0.331685	9
	Age, Flu – Season	0.004270	0.303918	27
Person – Location	Age – Region	0.024840	0.337633	15
	Flu – Region	0.007460	0.316629	20
	Age, Flu – Region	0.018278	**0.196407**	60
Climate – Location	Season – Region	0.004783	0.283449	20
Three-context Combination				
Person – Climate – Location	Age – Season – Region	0.015134	0.243600	60
	Flu – Season – Region	0.066149	0.197595	45
	Age, Flu – Season – Region	0.035340	**0.174622**	135

The RMSEs show that each context detail has good accuracy. But, to conclude which contexts give the best accuracy, the RMSE needs to be compared equally based on each of the context details' range. Range in each context details is not same, for instance, (Age, Flu) ranged from [0.0003, 0.01] and (Age, Flu - Season - Region) ranged from [0, 0.2023]. Thus, RMSE in each context detail (*i*) needs to be rescaled into the range 0-1 with a formula below.

$$NormRMSE_i = \frac{RMSE_i}{(MAX_i - MIN_i)}$$

By having all RMSEs normalized, context details that give best prediction in each context combination can be identified (written in bold in Table 3). For all context combinations, the inclusion of Influenza yields the best accuracy. Also, the Norm RMSEs show that inclusion of more contexts (the bottom of the Table 3) gives better predictions.

VI. CONCLUSION

In this paper, we have introduced an ontology and several disease risk oriented rules: Infectious Disease Risk (IDR) Ontology and described tools which generate a BN from knowledge represented as SWRL rules and using the IDR ontology. We believe that such tools are necessary to allow epidemiologists to refine the prediction model as new data and knowledge of infectious diseases becomes available. As is confirmed by the experiments presented in Table 3, the more factors (contexts) that are taken into consideration, the better the Anthrax risk prediction.

Also in this paper, we have evaluated the accuracy of a BN for personalised Anthrax risk that encodes current knowledge. We compared conditional probabilities taken from patient visit records with BN predictions. We tested different combinations of context details, each of which came out with good accuracy. This suggests that building CPTs from knowledge as rules is a feasible option.

The infectious disease risk prediction system described here will be offered as a web service which advises users when and how to protect themselves from infections. The web service will be linked to several live APIs for supplying the current environmental data to compute a person's risk of being infected by a specified disease in a specific time frame given their demographic details and location.

The further work of this research covers: developing BNs for prediction of risk for other infectious diseases and measuring their accuracy; and prioritizing rules to deal with opposing, partially completed or out-of-date rules. Providing a tool to facilitate the system to auto-extract knowledge and convert it into IDR and rules is also a potential future work.

REFERENCES

[1] A. E. Aiello, A. M. Simanek, M. C. Eisenberg and A. R. Walsh, "Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial," vol. 15, 2016.

[2] E. Fukuda, S. Kokubo and J. Tanimoto, "Risk assessment for infectious disease and its impact on voluntary vaccination behavior in social networks.," *Chaos, Solitons & Fractals,* vol. 68, pp. 1-9, 2014.

[3] L. J. Carpenito, Nursing Diagnosis: Application to Clinical Practice, Philadelphia: Wolters Kluwer Health, 2007.

[4] P. J. Giabbanelli, T. Torsney-Weir and V. K. Mago, "A fuzzy cognitive map of the psychosocial determinants of obesity," vol. 12, no. 12, 2012.

[5] N. Douali, H. Csaba, J. D. Roo, E. I. Papageorgiou and M.-C. Jaulent, "Diagnosis Support System based on clinical guidelines: Comparison between case-based fuzzy cognitive maps and bayesian networks," vol. 113, no. 1, 2014.

[6] H. M. Semakula, G. Song, S. P. Achuu and S. Zhang, "A Bayesian belief network modelling of household factors influencing the risk of malaria: A study of parasitaemia in children under five years of age in sub-Saharan Africa," vol. 75, 2016.

[7] M. Lappenschaar, A. Hommersom, P. J. F. Lucas, J. Lagro and S. Visscher, "Multilevel Bayesian networks for the analysis of hierarchical health care data," vol. 57, no. 3, 2013.

[8] K. Kengkla, N. Charoensuka, M. Chaichana and S. Puangjan, "Clinical risk scoring system for predicting extended-spectrum β-lactamase-producing Escherichia coli infection in hospitalized patients.," *The Journal of hospital infection,* pp. 49-56, 2016.

[9] E. J. Tomayko, B. A. Weinert, L. Godfrey, A. K. Adams and L. P. Hanrahan, "Using Electronic Health Records to Examine Disease Risk in Small Populations: Obesity Among American Indian Children, Wisconsin, 2007-2012.," vol. 13, no. 29, 2016.

[10] Q. Jiang, J. Zhou, Z. Jiang and B. Xu, "Identifying risk factors of avian infectious diseases at household level in Poyang Lake region, China," *Preventive Veterinary Medicine,* pp. 151-160, 2014.

[11] A. S. Fialho, S. M. Vieira, U. Kaymak and R. J. Almeida, "Mortality prediction of septic shock patients using probabilistic fuzzy systems," *Applied Soft Computing,* pp. 194-203, 2016.

[12] G. Koop, C. A. Collar, N. Toft, M. Nielen and T. v. Werven, "Risk factors for subclinical intramammary infection in dairy goats in two longitudinal field studies evaluated by Bayesian logistic regression," vol. 108, no. 4, 2013.

[13] D. Beaudequin, F. Harden, A. Roiko, H. Stratton and C. Lemckert, "Beyond QMRA: Modelling microbial health risk as a complex system using Bayesian networks," vol. 80, 2015.

[14] P. A. Aguilera, A. Fernandez, R. Fernandez, R. Rumi and A. Salmeron, "Environmental Modelling & Software Bayesian networks in environmental modelling," *Environmental Modelling and Software,* vol. 26, no. 12, pp. 1376-1388, 2011.

[15] T. Charitos, L. C. van der Gaag, S. Visscher and K. A. M. Schurink, "A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients," *Expert Systems with Applications,* vol. 36, no. 2, pp. 1249-1258, 2009.

[16] P. Panayiotis, H. X. Simon, A. Denise and B. A. T. Alex, "Prediction of lung cancer incidence on the low-dose computed tomography arm of the National Lung Screening Trial: A dynamic Bayesian network," *Artificial Intelligence in Medicine,* vol. 72, pp. 42-55, 2016.

[17] Q. Xiao, L. Xing and G. Song, "Time series prediction using optimal theorem and dynamic," *Optics,* vol. 127, p. 11063–11069, 2016.

[18] J. J. Dabrowski, C. Beyers and J. P. de Villiers, "Systemic banking crisis early warning systems using dynamic Bayesian," *Expert Systems with Applications,* vol. 62, pp. 225-242, 2016.

[19] H.-Y. Kao, C.-H. Huang and H.-L. Li, "Supply chain diagnostics with dynamic Bayesian networks," *Computers & Industrial Engineering,* vol. 49, pp. 339-347, 2005.

[20] H. F. L. Wertheim, P. Horby and J. P. Woodall, Atlas of Human Infectious Diseases, Oxford: Wiley-Blackwell, 2012.

[21] CDC, "Lesson 1: Introduction to Epidemiology," 18 May 2012. [Online]. Available: http://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section10.html.

[22] HSPC, "Health Protection Surveillance Centre," 2016. Available: http://www.hpsc.ie/NotifiableDiseases/NotifyingInfectiousDiseases/

[23] A. Ruttenberg, A. Goldfain, A. Diehl, B. Smith and B. Peters, "Infectious Disease Ontology," The National Center for Biomedical Ontology, 5 July 2016. [Online]. Available: http://purl.obolibrary.org/obo/ido.owl.

[24] A. Onisko, P. Lucas and M. J. Druzdzel, "Comparison of Rule-Based and Bayesian Network Approaches in Medical Diagnostic Systems," in *8th Conference on Artificial Intelligence in Medicine*, Portugal, 2001.

[25] R. A. Vinarti and L. M. Hederman, "A Performance Analysis of an Algorithm to Auto-create a Bayesian Network," in *Database and Expert System Applications*, Lyon, 2017.

[26] W. H. O. WHO, "Climate Change and Human Health - Risks and Responses Summary," World Health Organization, Geneva, 2003.

[27] X. Wu, Y. Lu, S. Zhou, L. Chen and B. Xu, "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation," vol. 86, 2016.

[28] D. N. Fisman, "Seasonality of viral infections: Mechanisms and unknowns," *American Journal of Preventive Medicine,* vol. 18, no. 10, pp. 946-954, 2008.

[29] T. P. Monath and P. F. C. Vasconcelos, "Yellow fever," *Journal of Clinical Virology,* vol. 64, pp. 160-173, 2015.

[30] H. Yi, B. R. Devkota and J.-s. Yu, "Effects of global warming on mosquitoes & mosquito-borne diseases and the new strategies for mosquito control," *Entomological Research,* vol. 44, no. 6, pp. 215-235, 2014.

[31] Norsys, "Examples," Norsys Software Corporation, 1995-2017. [Online]. Available: https://www.norsys.com/netica-j/examples/SimulateCases.html.

[32] W.-K. Wong, A. Moore, G. Cooper and M. Wagner, "Bayesian network anomaly pattern detection for disease outbreaks," in *Twentieth International Conference on Machine Learning*, Washington, D.C., 2003.

[33] W.-K. Wong, A. Moore, G. Cooper and M. Wagner, "Rule-based anomaly pattern detection for detecting disease outbreaks," in *Eighteenth National Conference on AI*, Alberta, 2002.

[34] A. C. I. Portal, "Allegheny County Information Portal," Pittsburgh, Pennsylvania, US, [Online]. http://www.alleghenycounty.us/.

[35] F. Carrat, A. Tachet, C. Rouzioux and B. Housset, "Evaluation of clinical case definitions of Influenza: detailed investigation of patients during 1995 - 1996 epidemic in France," *Clinical Infectious Diseases,* vol. 28, pp. 283-290, 1999.

[36] A. S. Monto, S. Gravenstein, M. Colopy and J. Scweinle, "Clinical signs and symptoms predicting influenza infection," *Arch International Medicine,* vol. 160, pp. 3243-3247, 2000.

[37] D. N. Kyriacou, A. C. Stein, P. R. Yarnold and D. M. Courtney, "Clinical predictors of bioterrorism-related inhalational anthrax," *The Lancet,* vol. 364, pp. 449-452, 2004.

[38] A. Farmland, "Allegheny County Farmland Preservation Program," [Online]. Available: http://www.alleghenyfarmland.com/.

[39] E. W. Steyerberg, A. J. Vickers, N. R. Cook and T. Gerds, "Assessing the performance of prediction models: a framework," *Epidemiology,* vol. 21, no. 1, pp. 128-138, 2010.

599