EasyChair Preprint

№ 4625

Causal Bayesian Networks for Medical Diagnosis: A Case Study in Rheumatoid Arthritis

Ali Fahmi, Amy MacBrayne, Evangelia Kyrimi, Scott Mclachlan, Frances Humby, William Marsh and Costantino Pitzalis

November 23, 2020

Causal Bayesian Networks for Medical Diagnosis: A Case Study in Rheumatoid Arthritis

Ali Fahmi[†], Amy MacBrayne[§], Evangelia Kyrimi[†], Scott McLachlan[†], Frances Humby[§], William Marsh[†], Costantino Pitzalis[§]
[†]School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
[§]The William Harvey Institute, Queen Mary University of London, London, UK
Email: a.fahmi@qmul.ac.uk, a.macbrayne@qmul.ac.uk, e.kyrimi@qmul.ac.uk, s.mclachlan@qmul.ac.uk, f.humby@qmul.ac.uk, d.w.r.marsh@qmul.ac.uk, c.pitzalis@qmul.ac.uk

***Abstract*— Bayesian network (BN) models have been widely applied in medical diagnosis. These models can be built from different sources, including both data and domain knowledge in the form of expertise and literature. Although it might seem simple to depend only on data, this will not be the best approach unless a large dataset is available. In this study, we present a knowledge-based BN modelling approach which we applied for diagnosing the chronic disease of rheumatoid arthritis (RA). We illustrate the process of extracting the relevant knowledge, starting by identifying the BN variables implied by the activities and decision points shown in a model of the caremap for RA diagnosis. To complete this, further medical knowledge is elicited from an expert panel of rheumatologists, the medical literature is investigated, and a data set is used to parameterise the model. We compare the performance of this knowledge-based BN with another BN model learnt entirely from data. The results show that our proposed knowledge-based model outperforms the data-driven one.**

Keywords*— *Bayesian networks, chronic diseases, rheumatoid arthritis, diagnosis

I. Introduction

BNs are a popular technique for decision support in medicine [1], [2] perhaps because they can be built using a combination of data and knowledge elicited from domain experts or literature [3]–[6]. The use of knowledge is an advantage when there is a lack of sufficient good-quality data, but essential when the application requires a causally coherent model. BNs are widely used in medical diagnosis because they can express medical knowledge and uncertainty [6] and deal with incomplete data [7].

Many people suffer from chronic diseases and joint pain caused by inflammatory arthritis (IA). Rheumatoid arthritis (RA) is the most frequent IA, affecting 1% of the UK population [8]. Diagnostic delays may occur due to inadequate evidences at onset of IA [9]. That has led rheumatologists to propose diagnostic criteria, such as American College of Rheumatology (ACR) 1987 [10] and, in 2010, its revision jointly with the European League Against Rheumatism (EULAR) [11]. The 2010 criteria are point-based and consider four domains: joint involvement, serology results, acute-phase reactants and duration of symptoms [12]. Although the 2010 criteria have improved the early identification of RA compared to 1987 criteria, there is still the need to improve the accuracy of RA diagnosis [9].

In this paper we describe a BN model for diagnosing RA from other forms of IA using a combination of experts' knowledge and data. We differentiate RA from other IA conditions by representing causal connections between risk factors, signs, symptoms, and relevant comorbidities. To investigate the value of using knowledge, we compare the knowledge-based BN with an alternative model learnt only from data.

The reminder of this paper is as follows: Section II reviews the literature on BN modelling approaches for medical diagnosis and on models for the diagnosis of RA. In Section III, we explain how the data are processed and how both BN models are built. Section IV includes a comparison of the BN models' performance, a review of the predicted outcomes and an explanation of one RA case. Finally, conclusions are in Section V.

II. Literature Review

Many BN models have been developed for medical diagnosis. Such models can be built using data, knowledge, medical literature, ontologies or other sources of medical evidences, or a combination of these sources. For example, [13] presents a BN model for diagnosing preeclampsia built based on knowledge. In [6], multiple BN models (developed from both knowledge and data) are compared with other classifiers learnt entirely from data to diagnose Alzheimer's disease, dementia, and mild cognitive impairment. For each condition, the performance of knowledge-based BN models, with areas under the ROCs (AUROC) of 0.86, 0.96, and 0.97 respectively, is comparable to that of the best data-driven classifier with AUROCs of 0.90, 0.98, and 0.97 respectively. In contrast, [14] learnt a BN for the diagnosis of Alzheimer's disease from data having first analysed the data to identify the most important variables. Reference [15] presents a BN model to differentiate dengue from other acute febrile illness: the work used knowledge from domain expert and literature, combining this with data. In [16], a BN for diagnosing musculoskeletal disorders of the shoulder was built using expert knowledge and medical literature, together with retrospective data. The expert elicitation used two stages: an initial model was updated following a review by a panel of orthopaedic specialists.

Some studies have suggested extracting medical knowledge from medical dictionaries or ontologies. In [17] a two-level BN model for diagnosis is built by automatically extracting medical knowledge from an ontology; the diagnostic part is supplement by a 'decision network' to evaluate available tests using utility values specified by a physician. In [18], multiple models including a BN model are automatically developed to diagnose mild cognitive impairment using a combination of approaches, including knowledge extracted from the SNOMED ontology.

Some studies use only data to build BN models. For example, [19] combines structured data with information extracted from medical notes (i.e. text) to build a BN model for dementia diagnosis. In another study [20], four data-driven classifier models – rules, logistic regression, tree, and Bayes net – are trained to recognise primary hyperparathyroidism.

To the best of our knowledge, there is no BN model for RA diagnosis. Other types of classifier have been built for RA

diagnosis, mainly by exploiting data. In [21], a combination of data and expert knowledge is used to develop an RA diagnostic tool. The method uses fuzzy logic and neural networks. A team of orthopaedic surgeons and rheumatologists were consulted to determine the relevant variables including symptoms, medical tests and disease severity measurement. Another study [22] used multiple machine-learning algorithms including decision trees (using a variety of algorithms) and support vector machine, combined using a variant of the AdaBoost, showing that their algorithm outperforms others in distinguishing RA from other rheumatic diseases (at highest 81% versus 85% accuracy). Separately, [23] also applied AdaBoost using social and medical data to separate RA from fibromyalgia. In [24], decision trees are trained using personal and medical data of patients, chosen using a feature selection algorithm. The classifier is used to differentiate RA patients from non-RA ones In [24], decision trees are trained using personal and medical data of patients, chosen using a feature selection algorithm. The classifier is used to differentiate RA patients from non-RA ones.

These studies show the lack of a knowledge-based approach and the need for building a BN model for RA diagnosis. A model that not only expresses the causal connection between factors involved in RA development and manifestation learnt from experts and literature, but also can assist clinicians to measure the uncertainty embedded in RA classification from other types of IA.

III. Data processing and modelling

Bayesian networks (BNs) are probabilistic graphical models that represent a set of random variables and their conditional dependencies as a directed acyclic graph (DAG) [25]. The variables and dependencies of the DAG and the parameters that determine the strength of dependencies can be learnt from experts, literature, ontologies, data, or a mixture of them. In this study, we first process and analyse data. Then we create the structure of two BNs: (1) a knowledge-based model elicited from experts and (2) a model learnt from data using structural expectation-maximisation (EM) method. In both cases, the parameters of these models are learnt from the data.

To build the knowledge-based BN, we elicit the rheumatology knowledge from a panel of experts from the Barts Health Rheumatology Department: the second author who is a clinical research fellow, the fifth author who is a consultant rheumatologist, and a postdoctoral researcher. The clinical research fellow is called “main expert” in the rest of this paper since she provided the majority of the advice.

A. Data explanation

We use a dataset collected in the Pathobiology of Early Arthritis Cohort (PEAC) study which has been running since 2009 [26]. In the available data, there are 373 cases that are diagnosed with RA (226 cases), undifferentiated arthritis (UA) (79 cases), psoriatic arthritis (PsA) (49 cases), monoarthritis (12 cases), and others (7 cases).

The dataset consists of 29 variables covering patients’ personal and lifestyle information, demographic, medical background, comorbidities, disease manifestation, diagnosis, disease activity data, and medication. The dataset includes some outliers, beyond clinically possible ranges, and some data values were missing. The extreme outliers were removed and data were cleaned as far as possible. Some variables were discretised using medically meaningful thresholds proposed by the clinical team. After these processes, we remove 6 cases which had two or more missing values out of 29. We also remove 4 cases with no clear diagnosis. In the 363 remaining cases, only 4% of data was missing. The diagnoses of the remaining cases were RA (217 cases), UA (78 cases), PsA (49 cases), monoarthritis (12 cases), and others (7 cases). Since our goal was to classify RA from other arthritic conditions, we grouped all conditions into RA and ‘other’ inflammatory arthritis (IA), which contains UA, PsA, monoarthritis, and others.

B. BN structure

We interview the experts to create a model of the caremap for RA diagnosis using the standardised approach for caremap development [27]. As shown in Fig. 1, the caremap consists of activity nodes and decision points which are represented by rectangular and diamond shapes, respectively. By considering the concepts mentioned in the caremap and the available dataset, we identified the categories of variables needed in the BN.

Once a referral is received, the rheumatologist should collect the patient history which refers to risk factors including personal, demographic, lifestyle, and medical background risks, comorbidities, and intervention.

- Personal risks include age and gender.
- Demographic risks associated with RA diagnosis include ethnicity [28], [29], education, and occupation [30].
- Alcohol [31], smoking [32]–[34], and body mass index (BMI) [35] are three lifestyle risks.
- Relevant medical background includes a family history (FHx) of RA or other IA FHx.
- Relevant comorbidities are those with similar symptoms and serology results. Thyroid autoimmune disease

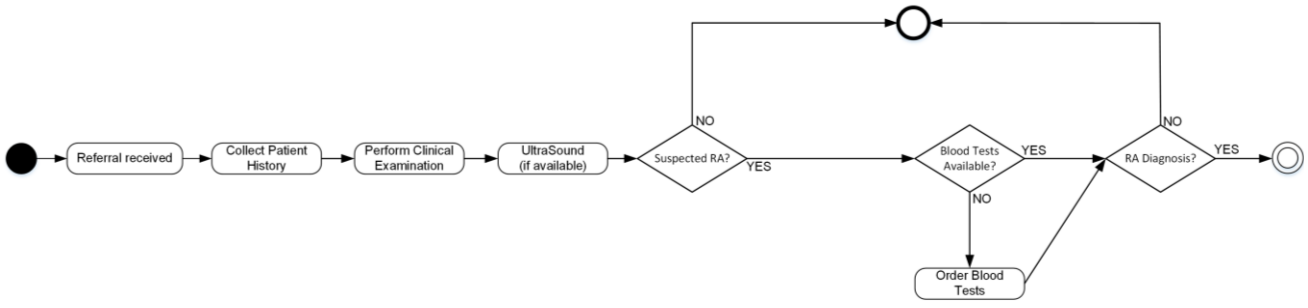

Fig. 1. A caremap for RA diagnosis

(TAD) and skin psoriasis are also considered as possible evidence for a common trigger of other autoimmune diseases such as RA or PsA. Osteoarthritis (OA), skin psoriasis, connective tissue disease (CTD), and crystal arthropathy (CA) cause similar symptoms as RA and CTD and CA also affect blood test results of erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP).

- Intervention refers to the possible prescription of steroids by a primary care doctor to alleviate symptoms and reduce inflammation.

From the caremap node of perform clinical examination, we derived disease manifestations which can be divided into serology results, symptoms, and signs. Serology results include measurements of two cyclic citrullinated peptide (CCP) and rheumatoid factor (RhF), as well as ESR and CRP as the inflammation marker in blood. The most important symptom is morning stiffness for more than 30 minutes. Other symptoms are general malaise, measured by global health scores, joint pain, and fatigue which patients are asked to indicate or provide a number between 0 and 100, where 0 indicates the best and 100 the worst. The count of swollen joints and the count of tender joints reveal the effect of the disease which are measured by the rheumatologists. Another sign is the symmetrical pattern of swollen joints associated with RA. Ultrasound, where available, enables the rheumatologists to have a clearer image of the joints.

At this point, considering all risk factors, comorbidities, and disease manifestations, the rheumatologists may suspect of RA and decide that blood tests are needed. Finally, the rheumatologist diagnoses a person with RA or not.

Apart from the concepts represented in the caremap, our experts present the pathogenesis of RA which is a complex and multifactorial interplay of genetic and environmental factors. Our model considers genetics and the impact on development of pathogenic antibodies (serological pathogenesis), lifestyle, personal and hormonal risk factors. Two known antibodies, CCP and RhF are implicated in RA pathogenesis through a gene-environment interaction. Measured by blood test, CCP is associated with HLA-DRB1 gene (known as the "shared epitope") and smoking [36] and RhF is associated with HLA-DRB1 gene, PTPN22 gene, and smoking [33], [37]. Female sex hormones also play a role in developing RA as female hormonal fluctuations during postpartum [38] and early menopause [39] have been observed to be classical times for the onset of RA [40].

We use the extracted variables to create a BN structure for review by the main expert. The main expert reviewed the variables, the states of each variable and causal / associational dependencies. AgenaRisk software is used to build this model [41], as it is displayed in Fig. 2. The knowledge-based model contains some latent and synthetic variables represented by dashed ovals. The four latent variables are: Menopause, 'Pregnancy and Postpartum', 'HLA-DRB1 Gene', and 'PTPN22 Gene'. Although included in the variables known to be possible risk factors, no values of these variables were recorded in our dataset. Our experts provided prior probabilities for these latent variables to parameterise them. Synthetic variables combine their parent variables and collectively influence their children variables. Ten synthetic variables are 'Demographic Risks', 'Personal Risks', 'Lifestyle Risks', 'Medical Background Risks', Serostatus, 'IA FHx', 'Comorbidity Background', 'Alternative Explanation of Symptoms' and 'Alternative Explanation of CRP and ESR', and 'Female Sex Hormones Effect on RA'. These variables reduce the chances of overfitting of the model by reducing the number of input variables into the Diagnosis variable. These variables do not exist in the dataset, however our experts defined their values by providing a set of rules. Although our experts tried to provide inclusive rules based on the states of the parent variables of each synthetic variable, these rules do not cover all the cases and leave some missing values.

As an alternative to the knowledge-based model, we used the processed data to learn the structure of a BN (see Fig. 3) using bnlearn's structural EM method [42]. This learning method was chosen it is able to deal with missing data. The dataset use was exactly the same one used to learn the

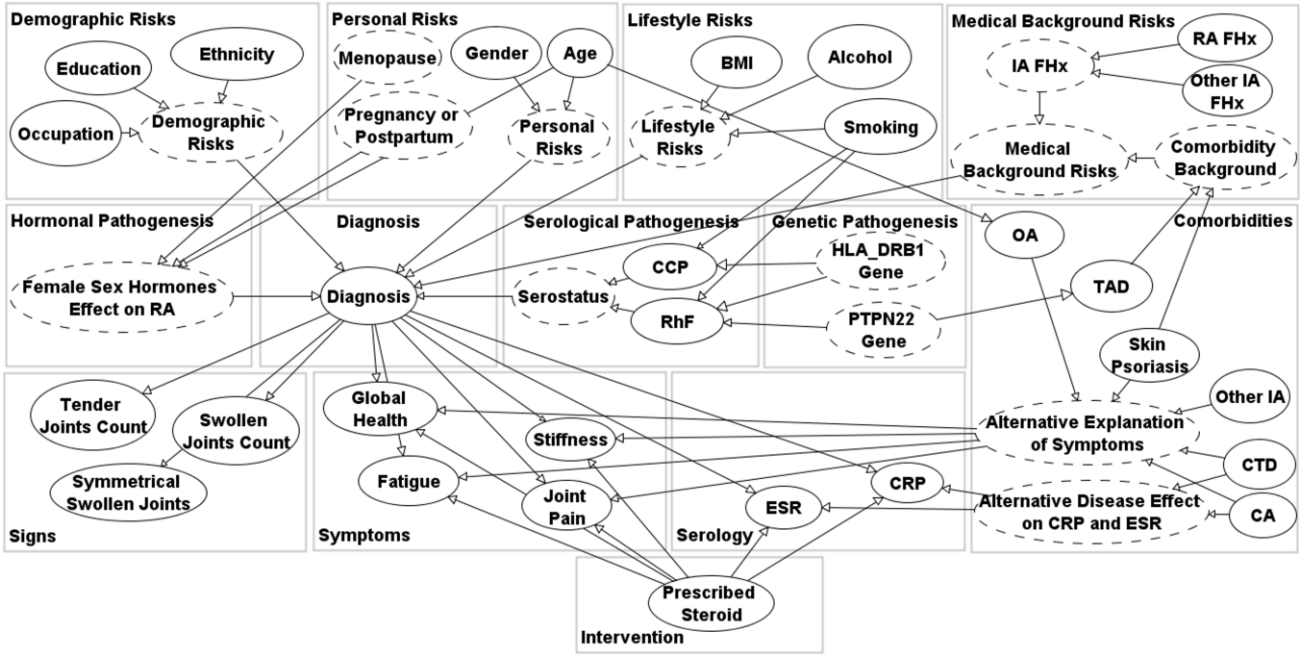

Fig. 2. Knowledge-based BN model for RA diagnosis built from experts' knowledge

parameters of the knowledge-based model, with discretised variables as described in Section III A above.

This BN model is entirely learnt from data and there is no direct intervention by experts in designing the model. However, we made many processes on the data along with the main expert in order to prepare it for structure and/or parameter learning. The data-driven model has no synthetic or latent variables. In this model, it is medically meaningful that the Diagnosis variable is linked to the 'Symmetrical Swollen Joints' and Stiffness. It also make medical sense that Diagnosis is connected in the reverse direction to CCP, RhF, 'Swollen Joints Count', 'Global Health', and indirectly 'Skin Psoriasis' which is a component of 'Comorbidity Background' of the knowledge-based model. The connection between CCP and Smoking reflects the environment-serological pathogenesis mechanism of RA reported by [33]. The reverse connection between ESR and 'Prescribed Steroid' represents the medically meaningful association between them.

The data-driven method does not find any link from or to the 'Other IA' variable, therefore it is not involved in the model.

C. BN parameterisation

We train the parameters of the models using the data of which all variables are discretised. AgenaRisk software is used for parameter learning of both models.

Four latent variables defined by rules have large percentage of missing values. We partly incorporated the EM algorithm proportionate to the missing percentage in order to deal with the large missing percentages of these variables.

IV. Results

A. Cross-validation

We used the results of a 10-fold cross-validation to compare the performance of the models in terms of their discrimination and accuracy. Discrimination measures if the model is able to differentiate between two states of the diagnosis, i.e. RA and 'Other IA', and accuracy investigates whether the predicted outcomes are actually close to the recorded outcomes. For discrimination comparison, we compare the receiver operating characteristics (ROC) curves of each model, sensitivity, and specificity values. The area under the ROC curve (AUROC) of the knowledge-based BN is 0.86 with 95% confidence interval (CI) (0.82-0.90), but the AUROC of the data-driven BN is 0.72 with 95% CI (0.66-0.77). Fig. 4 displays the ROC curves of the knowledge-based BN model (black line) and the data-driven model (grey line). At 90% sensitivity, the knowledge-based BN shows 64% specificity and the data-driven BN results 30% specificity.

To evaluate the accuracy of the two models, we apply the Brier score (BS) and the Brier skill score (BSS). BS is the measurement of the mean squared difference between the predicted results and the real ones. Its value can be between 0 and 1, where 0 refers to the perfect prediction and 1 is the worst outcome. BSS is the improvement of the prediction relative to a reference prediction which is usually the average probability of the prediction event recorded in the real data. Its values can be between minus infinity and 1, where 1 indicate to a perfect prediction and lesser or negative values are the worse outcomes. The BS of the knowledge-based BN model is 0.15, but that of the data-driven model is 0.23. The BSS of the knowledge-based model is 0.38, whereas that of the data-driven model is 0.039.

The evaluation of models in terms of discrimination and accuracy is summarised in Table I.

The confusion matrices of the knowledge-based and data-driven models are shown in Tables II and III, respectively. Considering the proportion of 217 cases of RA out of 363 records, we assumed the threshold probability of diagnosis to separate predicted RA cases from 'Other IA' to be 59.77%. We can see a better prediction accuracy of RA compared to 'Other IA' in both models. The accuracies of RA and 'Other IA' using knowledge-based BN model are 82% and 79%,

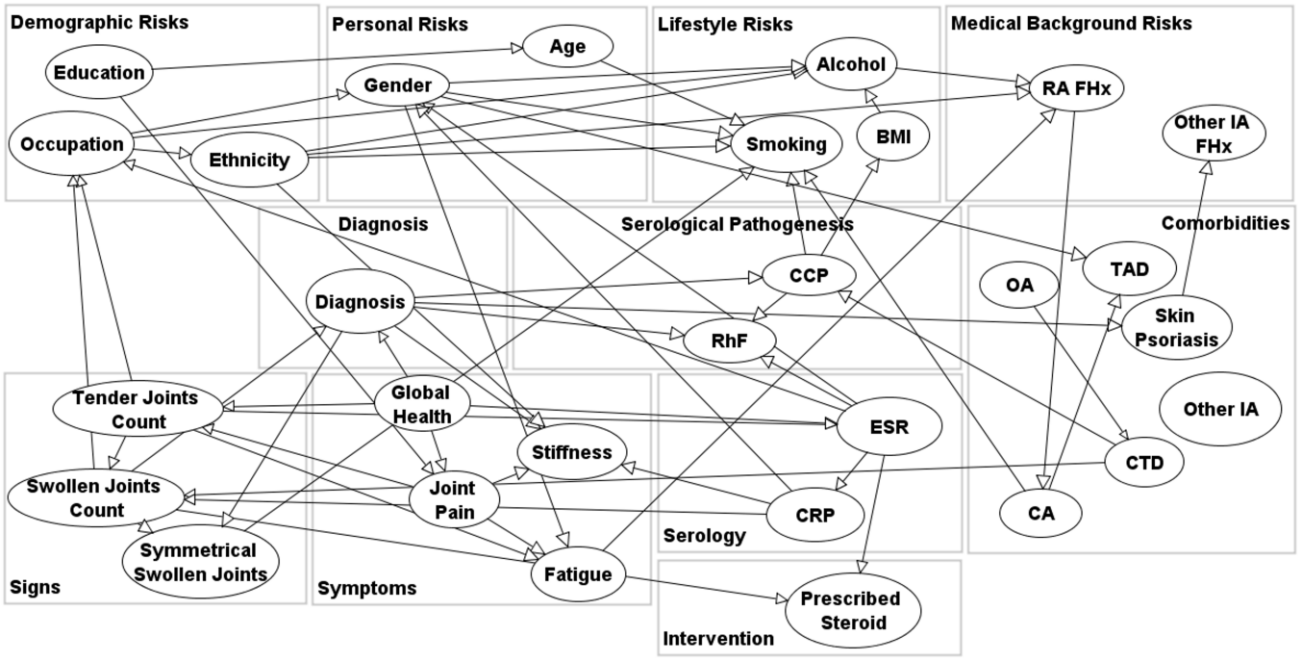

Figure 3. Data-driven BN model for RA diagnosis learnt from structural EM

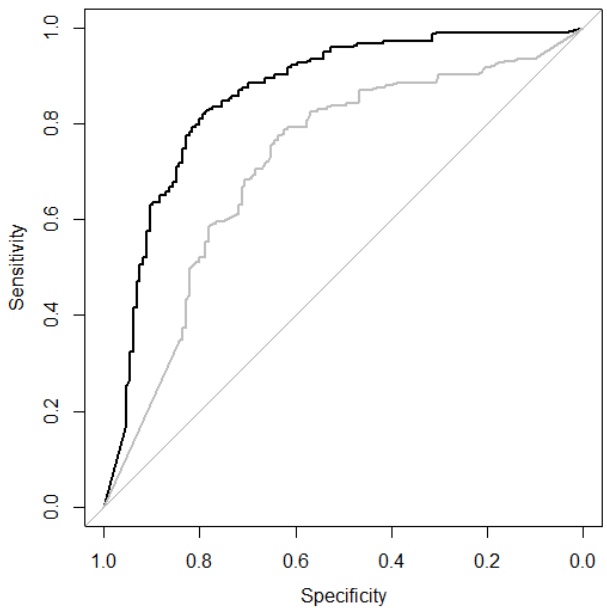

Fig. 4. ROC curves of knowledge-based BN (black) and BN learnt by structural EM (grey)

respectively. Data-driven BN model respectively achieves 73% and 65% in classifying RA and 'Other IA'.

The accuracy of 'Other IA' in both models is relatively lower. We believe that the lower accuracy may result from the relative lack of known risk factors of the 'Other IA' conditions. PsA, one of the other IA conditions, has only two major risk factors in our models: 'Other IA FHx' and 'Skin Psoriasis' comorbidity. For another condition, monoarthritis, 'Other IA FHx' is the only risk factor. Finally, UA, the other main IA condition, refers to those arthritic cases that are not differentiable and these cases have no specific risk factors involved in the models.

We further compared the accuracy of classifying other IA conditions separately. Table IV indicates the number of correctly classified cases of UA (out of 78), PsA (out of 49), monoarthritis (out of 12), and others (out of 7) using both models.

To gain deeper insight into the predicted outcomes and the real diagnosis, we divide the prediction probabilities into five bins and investigate the number of the accurate cases and their percentages. Table V shows the accurate cases and their percentages in each bin for both knowledge-based and data-driven models.

The discrimination and accuracy results show that the knowledge-based BN model performs better than the data-driven model. Although the knowledge-based model is sparser than the data-driven one, both models are susceptible to over-fitting. The Diagnosis variable in the knowledge-based model and two variables of Smoking and Stiffness in the data-driven model have more than four parent variables. They can potentially lead the corresponding models to over-fit considering the number of the parents' states and their big conditional probability tables.

B. Review inaccurate cases

We investigated the inaccurate cases of the knowledge-based model. There are 38 inaccurate RA predictions and 31 inaccurate predictions of 'Other IA'. These inaccurate RA predictions suffer from having a high percentage of missing values in the Serostatus and 'Personal Risks' latent variable, which are respectively 46% and 54%. However, the overall missing percentages of these two variables are respectively 36% and 45%.

17 cases of inaccurate 'Other IA' predictions are actually diagnosed with UA at onset. All these cases have a negative CCP and RhF, except one case that has a low positive RhF. The classifier detects these cases as RA, although two major factors, CCP and RhF, are negative. Further investigation is needed to find out if these cases have been diagnosed with RA in the future or not.

C. Explanation of one RA case

To increase the trustworthiness of the models, we can explain the models by doing reasoning [43]. This needs to pick one real case and enter the evidences of this case into the models to investigate the reasoning. The recorded diagnosis of this specific case is RA. By giving the available evidences of risk factors, comorbidities, and manifestations, we achieve an 85% probability of getting diagnosis with RA using the knowledge-based model, whereas the data-driven model obtains 99%. Both probabilities are greater than the overall

TABLE I. COMPARISON OF MODELS PERFORMANCE

	Knowledge-based BN	**Data-driven BN**
AUROC	0.86	0.72
Specificity (at 90% sensitivity)	0.64	0.30
Sensitivity (at 80% sensitivity)	0.80	0.57
Brier score	0.15	0.23
Brier skill score	0.38	0.04

TABLE II. CONFUSION MATRIX OF KNOWLEDGE-BASED BN MODEL

Diagnosis	**RA**	**Other IA**
59.77 ≤ Prediction	179	31
Prediction < 59.77	38	115

TABLE III. CONFUSION MATRIX OF DATA-DRIVEN BN MODEL

Diagnosis	**RA**	**Other IA**
59.77 ≤ Prediction	159	51
Prediction < 59.77	58	95

TABLE IV. PREDICTION ACCURACY OF OTHER IA SEPARATELY

Other IA conditions	**Knowledge-based BN**	**Data-driven BN**
UA	61/78(78%)	48/78(62%)
PsA	42/49(86%)	36/49(73%)
Monoarthritis	11/12(92%)	9/12(75%)
Others	1/7(14%)	2/7(29%)

TABLE V. ACCURACY OF RA PREDICTION IN 5 BINS

RA prediction bins	Knowledge-based BN		Data-driven BN	
	Cases	Accuracy	Cases	Accuracy
$0.8 < P \leq 1.0$	186/363	162/186(87%)	177/363	136/177(77%)
$0.6 \leq P < 0.8$	24/363	17/24(71%)	30/363	21/30(70%)
$0.4 \leq P < 0.6$	21/363	11/21(52%)	27/363	13/27(48%)
$0.2 \leq P < 0.4$	25/363	15/25(60%)	29/363	16/29(55%)
$0.0 \leq P < 0.2$	107/363	89/107(83%)	100/363	68/100(68%)

ratio of RA cases (59.77%), so that the outcomes match the recorded diagnosis.

We consider the diagnosis probability while removing the evidences one-by-one to check if variables support diagnosis or not. In the knowledge-based model, except BMI and Alcohol all other risk factors support diagnosis, though Smoking shows slight sensitivity which is not favourable. This can be caused by the fact that BMI, Alcohol, and Smoking have the highest percentage of missing values: 16%, 13%, and 13%, respectively. In the data-driven model, however, only Ethnicity supports diagnosis.

In the knowledge-based model, comorbidity of 'Skin Psoriasis' plays a major role in indirectly influencing the probability of diagnosis since it is closely associated with PsA, one of the other IA conditions. Other comorbidities have an expected supportive effect on diagnosis, except CTD. This can be due to the lack of positive CTD cases in the dataset. Similarly, in the data-driven model, 'Skin Psoriasis' supports diagnosis, however other comorbidities have no influence on it.

Although all factors of signs, symptoms, and serology results are sensitive in reasoning by the knowledge-based model, they act differently in the data-driven one. ESR, 'Global Health' and Stiffness support diagnosis in the data-driven model, whereas CRP, 'Joint Pain', and 'Swollen Joints Count' do not. Fatigue, 'Prescribed Steroid', 'Tender Joints Count', and 'Symmetrical Swollen Joints' play no roles in diagnosis reasoning.

V. CONCLUSION

In this paper, we describe a knowledge-based BN model built from experts' knowledge and supported by the established medical literature. The steps of data processing, structure building from knowledge and literature, and parameterisation are all explained.

We develop an entirely data-driven BN model and compare its outcomes with those of the knowledge-based BN. Their performances are analysed in terms of discrimination and accuracy and in both of them the knowledge-based BN outperforms the data-driven model. To compare the inferencing of two models, we pick one specific case and used its records as evidences to contrast the behaviour of the models. It shows that the knowledge-based model behaves more like what the rheumatologists expect, however the data-driven model shows lesser compatibility with the rheumatologist's expectations.

One area for further work is to complete the review of mispredictions in the knowledge-based model (see Section IV B) to improve the definition of synthetic variables, reducing the number of missing values in the parameter learning. Both the models described here are discrete, although the data includes both continuous and discrete values. We could use this data without discretising the continuous values to build and compare knowledge-based and data-driven hybrid models, also comparing their performances with those of the current discrete models. Other structure learning algorithms equipped with imputation to handle data with missing values could also be tested, since this would allow the use of more of the original data.

REFERENCES

[1] A. Abu-Hanna and P. Lucas, "Editorial: Prognostic models in medicine," *Methods Inf. Med.*, pp. 1–5, 2001.

[2] P. J. F. Lucas, L. C. Van Der Gaag, and A. Abu-Hanna, "Bayesian networks in biomedicine and health-care," *Artif. Intell. Med.*, vol. 30, no. 3, pp. 201–214, 2004.

[3] K. Orphanou, A. Stassopoulou, and E. Keravnou, "DBN-Extended : A Dynamic Bayesian Network Model Extended With Temporal Abstractions for Coronary Heart Disease Prognosis," vol. 20, no. 3, pp. 944–952, 2016.

[4] Z. Shojaei Estabragh, M. Mohammad, R. Kashani, F. J. Moghaddam, and S. Sari, "Bayesian Network Model for Diagnosis of Social Anxiety Disorder," *2011 IEEE Int. Conf. Bioinforma. Biomed. Work.*, pp. 639–640, 2011.

[5] S. H. Chen and C. A. Pollino, "Good practice in Bayesian network modelling," *Environ. Model. Softw.*, vol. 37, pp. 134–145, 2012.

[6] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci, and D. C. Muchaluat Saade, "A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment," *Comput. Biol. Med.*, vol. 51, pp. 140–158, 2014.

[7] X. Wang, B. Zheng, W. Good, J. King, and Y. Chang, "Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network," *Int. J. Med. Inform.*, vol. 54, no. 2, pp. 115–126, 1999.

[8] M. Schneider and K. Krüger, "Rheumatoide Arthritis—Early Diagnosis and Disease Management," *Dtsch. Arztebl. Int.*, vol. 110, no. 27–28, pp. 477–484, 2013.

[9] B. Heidari, "Rheumatoid arthritis: Early diagnosis and treatment outcomes," *Casp. J. Intern. Med.*, vol. 2, no. 1, pp. 161–170, 2011.

[10] F. C. Arnett *et al.*, "The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis," *Arthritis & Rheumatism*, vol. 31, no. 3. pp. 315–324, 1988.

[11] D. Aletaha *et al.*, "2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative," *Arthritis Rheum.*, vol. 62, no. 9, pp. 2569–2581, 2010.

[12] J. Kay and K. S. Upchurch, "ACR/EULAR 2010 rheumatoid arthritis classification criteria," *Rheumatol.*, vol. 51, no. SUPPL. 6, pp. 5–9, 2012.

[13] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, R. F. Ramos, and K. Saleem, "A preeclampsia diagnosis approach using Bayesian networks," in *2016 IEEE International Conference on Communications, ICC 2016*, 2016, pp. 1–5.

[14] Z. Yin, Y. Zhao, X. Lu, and H. Duan, "A Hybrid Intelligent Diagnosis Approach for Quick Screening of Alzheimer's Disease Based on Multiple Neuropsychological Rating Scales," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–13, 2015.

[15] C. Sa-ngamuang, P. Haddawy, V. Luvira, W. Piyaphanee, S. Iamsirithaworn, and S. Lawpoolsri, "Accuracy of dengue clinical diagnosis with and without NS1 antigen rapid test: Comparison between human and Bayesian network model decision," *PLoS Negl. Trop. Dis.*, vol. 12, no. 6, pp. 1–14, 2018.

[16] N. Farmer, "An update and further testing of a knowledge-based diagnostic clinical decision support system for musculoskeletal disorders of the shoulder for use in a primary care setting," *J. Eval. Clin. Pract.*, vol. 20, no. 5, pp. 589–595, 2014.

[17] G. Bucci, V. Sandrucci, and E. Vicario, "Ontologies and Bayesian networks in medical diagnosis," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1–8, 2011.

[18] X. Zhang, B. Hu, X. Ma, P. Moore, and J. Chen, "Ontology

driven decision support for the diagnosis of mild cognitive impairment," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 781–791, 2014.
[19] L. B. Moreira and A. A. Namen, "A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia," *Comput. Methods Programs Biomed.*, vol. 165, pp. 139–149, 2018.
[20] Y. R. Somnay *et al.*, "Improving diagnostic recognition of primary hyperparathyroidism with machine learning," *Surg.*, vol. 161, no. 4, pp. 1113–1121, 2017.
[21] J. L. Salmeron, S. A. Rahimi, A. M. Navali, and A. Sadeghpour, "Medical diagnosis of Rheumatoid Arthritis using data driven PSO–FCM with scarce datasets," *Neurocomputing*, vol. 232, no. September 2016, pp. 104–112, 2017.
[22] Z. Shiezadeh, H. Sajedi, and E. Aflakie, "Diagnosis of Rheumatoid Arthritis Using an Ensemble Learning Approach," pp. 139–148, 2015.
[23] B. Garcia-Zapirain, Y. Garcia-Chimeno, and H. Rogers, "Machine Learning Techniques for Automatic Classification of Patients with Fibromyalgia and Arthritis," *Int. J. Comput. Trends Technol.*, vol. 25, no. 3, pp. 149–152, 2015.
[24] S. Shanmugam and J. Preethi, "Improved feature selection and classification for rheumatoid arthritis disease using weighted decision tree approach (REACT)," *J. Supercomput.*, vol. 75, no. 8, pp. 5507–5519, 2019.
[25] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.
[26] PEAC, "What is PEAC?," 2018. [Online]. Available: http://www.peac-mrc.mds.qmul.ac.uk/what.php. [Accessed: 06-Sep-2019].
[27] S. McLachlan, E. Kyrimi, K. Dube, and N. Fenton, "Clinical Caremap Development: How can caremaps standardise care when they are not standardised?," 2019.
[28] T. D. Spector, "Rheumatoid arthritis," *Rheum. Dis. Clin. North Am.*, vol. 16, no. 3, pp. 513–537, 1990.
[29] C. A. Peschken and J. M. Esdaile, "Rheumatic diseases in North America's indigenous peoples," *Semin. Arthritis Rheum.*, vol. 28, no. 6, pp. 368–391, 1999.
[30] G. S. Cooper, "Occupational exposures and risk of rheumatoid arthritis: Continued advances and opportunities for research," *J. Rheumatol.*, vol. 35, no. 6, pp. 950–952, 2008.
[31] Z. Jin, C. Xiang, Q. Cai, X. Wei, and J. He, "Alcohol consumption as a preventive factor for developing rheumatoid arthritis: A dose-response meta-analysis of prospective studies," *Ann. Rheum. Dis.*, vol. 73, no. 11, pp. 1962–1967, 2014.
[32] F. Wolfe, "The effect of smoking on clinical, laboratory, and radiographic status in rheumatoid arthritis," *J. Rheumatol.*, vol. 27, no. 3, pp. 630–637, 2000.
[33] L. Padyukov, C. Suva, P. Stolt, L. Alfredsson, and L. Klareskog, "A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis," *Arthritis Rheum.*, vol. 50, no. 10, pp. 3085–3092, 2004.
[34] E. W. Karlson, I. M. Lee, N. R. Cook, J. A. E. Manson, J. E. Buring, and C. H. Hennekens, "A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals," *Arthritis Rheum.*, vol. 42, no. 5, pp. 910–917, 1999.
[35] X. Feng *et al.*, "Body Mass Index and the Risk of Rheumatoid Arthritis: An Updated Dose-Response Meta-Analysis," *Biomed Res. Int.*, vol. 2019, 2019.
[36] M. Pedersen *et al.*, "Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides," *Arthritis Res. Ther.*, vol. 8, no. 4, 2006.
[37] A. B. Begovich *et al.*, "A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis," *Am. J. Hum. Genet.*, vol. 75, no. 2, pp. 330–337, 2004.
[38] A. Silman, A. Kay, and P. Brennan, "Timing of Pregnancy in Relation to the Onset of Rheumatoid Arthritis," *Arthritis Rheum.*, vol. 35, no. 2, pp. 152–155, 1992.
[39] M. Pikwer, U. Bergström, J. Å. Nilsson, L. Jacobsson, and C. Turesson, "Early menopause is an independent predictor of rheumatoid arthritis," *Ann. Rheum. Dis.*, vol. 71, no. 3, pp. 378–381, 2012.
[40] D. Alpízar-Rodríguez, N. Pluchino, G. Canny, C. Gabay, and A. Finckh, "The role of female hormonal factors in the development of rheumatoid arthritis," *Rheumatol.*, vol. 56, no. 8, pp. 1254–1263, 2017.
[41] Agena, "AgenaRisk." 2018.
[42] M. Scutari, "Structure learning from missing data," 2020. [Online]. Available: https://www.bnlearn.com/documentation/man/structural.em.html. [Accessed: 18-Feb-2020].
[43] E. Kyrimi, S. Mossadegh, N. Tai, and W. Marsh, "An Incremental Explanation of Inference in Bayesian Networks for Increasing Model Trustworthiness and Supporting Clinical Decision Making," *Artif. Intell. Med.*, vol. 103, no. January, p. 101812, 2020.