Expert Systems With Applications 183 (2021) 115333

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Probabilistic modeling approach for interpretable inference and prediction with data for sepsis diagnosis

Shuaiyu Yao [a,*,1], Jian-Bo Yang [b], Dong-Ling Xu [b], Paul Dark [c]

[a] School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Alliance Manchester Business School, The University of Manchester, Manchester M15 6PB, United Kingdom
[c] Division of Infection, Immunity and Respiratory Medicine, NIHR Manchester Biomedical Research Centre, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M15 6JA, United Kingdom

ARTICLE INFO

Keywords:
Sepsis diagnosis
Probabilistic modeling
Rule-based inference
Interpretability
Maximum likelihood evidential reasoning framework

ABSTRACT

Sepsis is a serious disease that can cause death. It is important to predict sepsis within the early stages after the presence of sepsis symptoms. In this paper, a new probabilistic modeling approach is used to establish classifiers for sepsis diagnosis. This approach is characterized by unique strong interpretability, which is reflected in three aspects: (1) evidence acquisition based on likelihood analysis, (2) probabilistic rule-based inference, and (3) parameters optimization using machine learning algorithms. Four-fold cross-validation is used to train and validate classifiers established by the new approach and alternative ones. Results show that in terms of classification capability, the classifier established by the new approach generally performs better than the majority of alternative classifiers for sepsis diagnosis, and close to the best one. As the classifier also features an inherent interpretability, it can be used as a tool for supporting diagnostic decision-making in sepsis diagnosis.

1. Introduction

As per the definition of the international consensus conference, sepsis is a life-threatening organ dysfunction generated by a dysregulated host response to infection (Singer et al., 2016). Each year, it influences more than 1 million patients in the U.S. (Angus et al., 2001; De Backer & Dorman, 2017) and over 30 million around the world (Fleischmann et al., 2016). The associated mortality rates are as high as around 20–30% (Friedman, Silva, & Vincent, 1998; Kumar et al., 2011; Martin, Mannino, Eaton, & Moss, 2003). For the last several decades, the incidence of sepsis has shown a steady increase (Elfeky et al., 2017). Moreover, sepsis imposes a significant financial burden on the U.S. healthcare system. It has been reported that the annual medical expenditures for sepsis patients in the U.S. exceeds 24 billion dollars (Lagu et al., 2012; Torio & Moore, 2006). Another important fact is that sepsis survivors have more chance to suffer from impaired quality of life (Heyland, Hopman, Coo, Tranmer, & McColl, 2000; Perl, Dvorak, Wenzel, & Hwang, 1995) and higher risk to develop major adverse cardiovascular events (Ou et al., 2016).

In a general sense, sepsis is generated by an invasion of microbes from local infection source into the bloodstream, which leads to organ dysfunction (Reinhart, Bauer, Riedemann, & Hartog, 2012). The infection can be typically identified by clinical signs and microbiological findings (Vincent, 2016). Hence, currently a golden standard for sepsis diagnosis is blood culture (Henriquez-Camacho & Losa, 2014). Nonetheless, blood culture has a very low sensitivity, and the culture results of more than half of patients with sepsis symptoms are negative (Jones & Lowes, 1996; Vincent et al., 2006). In addition, blood culture requires multi-step analysis, hence it is costly and time-consuming, and it needs professional operation (Kumar, Tripathy, Jyoti, & Singh, 2019). Since blood culture is laboratory-based, its instrumentation is not appropriate for miniaturization (Kumar et al., 2019). As considerable increase in mortality is associated with delay of treatment, the time of blood culture brings about the early use of broad-spectrum antimicrobial drugs (Kollef, Sherman, Ward, & Fraser, 1999; Kumar et al., 2006). An inevitable consequence is the unnecessary and/or prolonged broad-spectrum antimicrobial use that is related to a series of side effects such as the development of drug-resistant pathogens and Clostridium difficile infections (Dark et al., 2015).

Alternatively, the approach that can overcome the disadvantages of

* Corresponding author.
E-mail addresses: shuaiyuyao@sjtu.edu.cn (S. Yao), jian-bo.yang@manchester.ac.uk (J.-B. Yang), Ling.Xu@manchester.ac.uk (D.-L. Xu), paul.m.dark@manchester.ac.uk (P. Dark).
[1] ORCID: 0000-0002-9360-6882.

https://doi.org/10.1016/j.eswa.2021.115333
Received 18 March 2020; Received in revised form 9 April 2021; Accepted 1 June 2021
Available online 5 June 2021

blood culture, can be the implementation of highly sensitive biosensors for sepsis diagnosis, which offers assistance to the clinicians around the world (Kumar et al., 2019). The handheld point-of-care alternatives can also be developed for lab-based diagnosis to improve healthcare (Kumar et al., 2019). Using different biomarkers, many biosensors have been developed for sepsis diagnosis in recent years (Kumar et al., 2019). Biomarkers are defined as the biological molecule or features acting as an indicator of pathological or physiological process (Kumar et al., 2019). How the levels of sepsis biomarkers increase can be used to answer three clinically important questions: (1) whether the patient is infected; (2) whether the condition is serious; and (3) whether the patient responds to the treatment (Vincent, 2016).

A number of biomarkers that can be used in sepsis diagnosis are introduced in brief as follows. C-reactive protein (CRP) is an acute phase protein (Thompson, Pepys, & Wood, 1999), and according to numerous studies, its level is elevated in sepsis (Matson, Soni, & Sheldon, 1991; Póvoa et al., 2005; Yentis, Soni, & Sheldon, 1995). CRP has some advantages for neonatal sepsis such as rapid results and small amount of blood required (Chiesa et al., 2011; Patil, Dutta, Attri, Ray, & Kumar, 2016). Cytokines can be defined as pleiotropic regulators of immune response that play a part in the complex pathophysiology underlying sepsis (Schulte, Bernhagen, & Bucala, 2013). Interleukin 6 (IL-6) is a cytokine which has an initial response to infection and injury (Yang et al., 2016). As the level of IL-6 gets augmented significantly during early sepsis, IL-6 has been used to diagnose sepsis and predict patient outcome (Bozza et al., 2007; Yende et al., 2008). Interleukin 10 (IL-10) is an important anti-inflammatory cytokine that can be used for sepsis diagnosis and evaluation of current inflammatory response level (Prucha, Bellingan, & Zazula, 2015). High level of IL-10 indicates higher mortality of sepsis patients (Kellum et al., 2007). Procalcitonin (PCT) is a useful sepsis biomarker, as it can distinguish between infectious and non-infectious inflammation with high specificity and sensitivity (Angeletti et al., 2015; Angeletti, Battistoni, Fioravanti, Bernardini, & Dicuonzo, 2013; Spoto, 2018). PCT can be used as an early biomarker for guiding appropriate management, and it is helpful in indicating and excluding sepsis (Wolf, Wimalawansa, & Razzaque, 2019). White cell count (WCC) is a conventional infection indicator (Lam et al., 2011).

Although sepsis has been confirmed to be associated with biomarkers, the clinical presentation of sepsis is complicated, and it is difficult to decide whether a patient has sepsis using biomarkers alone (Chen, Lu, Hu, Liu, Zhao, Yan, & Tan, 2017). Machine learning is a promising tool to identify sepsis patients (Bhattacharjee, Edelson, & Churpek, 2017; Bradley et al., 2007). Appropriate prediction models (classifiers) have been established to accurately predict the onset of sepsis using different approaches such as gradient-boosted tree (Barton et al., 2019; Desautels et al., 2016; Mao et al., 2018), support vector machine (Taneja et al., 2017), random forest (Lamping et al., 2018; Oonsivilai et al., 2018; Ratzinger et al., 2018; Taylor et al., 2016), artificial neural network (López-Martínez, Núñez-Valdez, Lorduy Gomez, & García-Díaz, 2019; Saqib, Sha, & Wang, 2018; Scherpf, Gräßer, Malberg, & Zaunseder, 2019), and logistic regression (Faisal et al., 2018). However, these approaches have some limitations. For example, random forests give up the interpretability which is the major benefit of decision trees (Fischetti, Lantz, Abedin, Mittal, Makhabel, Berlinger, & Daroczi, 2016; Strobl, 2008). The artificial neural network is generally considered as 'black box', and it is very difficult to record how specific classification decision is made (Qiu & Jensen, 2004). The logistic regression may have a low prediction accuracy for the data set where the classes are far from linearly separable (Ng & Jordan, 2002).

To address these limitations, in this paper, we use a probabilistic modeling approach to establish interpretable classifiers for sepsis diagnosis, which is based on the concepts of Dempster-Shafer (D-S) theory (Dempster, 1967, 1968; Shafer, 1976). In D-S theory, a frame of discernment (FoD) is used to include predefined classes, to which basic probabilities are assigned to generate a belief distribution (BD). Basic probabilities are used to measure the extent to which observations of input variables point to different classes or subsets of classes. BD for each observation of an input variable is referred to as a piece of evidence. Dempster's rule may be used to combine multiple pieces of evidence. Nonetheless, the use of Dempster's rule needs the satisfaction of several conditions including the one that any evidence is assumed to be fully reliable. This assumption is rarely practical and often leads to a counter-intuitive issue encountered when Dempster's rule is used to combine highly or completely conflicting evidence (Yang & Xu, 2013). Based on the basic concepts of D-S theory, the evidential reasoning (ER) rule (Yang & Xu, 2013) eliminates this assumption by considering the reliability and relative importance of evidence, while still retaining the desirable features of Dempster's rule. One of the most important characteristics of the ER rule is that it constitutes a unique probabilistic inference process for conjunctive combination of independent evidence. The ER rule has been established to handle problems of discrete probabilistic inference in which both input and output variables are assumed to take categorical or discrete numerical values. However, this assumption can be violated, as discrete and continuous variables can coexist in inference and prediction problems. The new probabilistic modeling approach used in this paper is capable of dealing with both discrete and continuous data, which is demonstrated in a numerical case study.

When machine learning is used to construct classifiers from data, it is insufficient for a classifier to simply make accurate predictions on test data. A wide variety of applications require that the classifiers are interpretable, which indicates that humans are able to easily understand the information the models contain. In the context of machine learning, the term 'interpretability' can be used to represent the ability of explaining or presenting in understandable terms to a human. While there is a general consensus for the need of interpretability in machine learning models, much less agreement has been achieved for what constitutes interpretability. The method investigated in this paper is aimed to augment human intelligence in decision making processes and allow humans to intervene the processes using their intuition and experiences. It is in this context that the interpretability of the proposed method is emphasised. There are many areas, such as clinical decision-making, creative marketing advertisements, personal selling, hiring and monitoring employees and strategic decision-making, where human intervention, managerial intuition, and human–computer symbiosis can enhance work performance (Pavlou, 2018). Human intelligence is still necessary for decision-making related to human-oriented data associated with patients, employees, and customers that include their personalized information (Pavlou, 2018). Human intervention and intelligence generally outperforms artificial intelligence when dealing with ambiguity, vagueness, and incomplete information (Pavlou, 2018). Any intelligent system would have come across a situation where human intervention is essential and human intelligence is necessary for the complete functioning of a system (Hebbar, 2017). This crossover of the worlds is the key to augment both human and artificial intelligence, which is very important in clinical decision-making (Hebbar, 2017). The augmented intelligence will herald the application of artificial intelligence for improved clinical decision-making and predictions of outcomes (Görges & Ansermino, 2020). It will offer unprecedented opportunities to improve patient outcomes and performance of clinical decision-making, and hence reduce costs and impact population health (Görges & Ansermino, 2020). The interpretability plays an important role in the connection between artificial intelligence and clinical decision-making.

In this paper, by the interpretability of models (classifiers) we specifically mean their ability to interpret domain knowledge. In this context, the proposed MAKER-based classifier is inherently interpretable. This inherent interpretability of the MAKER-based classifier can be explained in the following three aspects. Firstly, evidence is acquired directly from patient data by means of likelihood analysis, and combined under the maximum likelihood evidential reasoning (MAKER) framework (Yang & Xu, 2017). The interdependence between multiple

pieces of evidence can also be captured under the MAKER framework. Secondly, evidence can be combined to generate belief rules which formulate a belief rule-base (BRB) for inference. The transparent and interpretable BRB inference is used to deduce the probabilistic relationship between biomarker features of patients and patients' outcomes. In the context of disease diagnosis, the BRB inference has been used for the diagnosis of lumph node metastasis in gastric cancer (Zhou et al., 2013, 2015), cardiac chest pain (Kong et al., 2012), and trauma (Kong et al., 2016). For each given instance of patient, the activated belief rules out of a BRB are combined to generate the predicted output distribution for patients' outcomes. Thirdly, the parameters of the MAKER-based classifiers (e.g., the reliability of each evidential element) established by the approach are statistically meaningful. They can be trained using machine learning algorithms. The inference process of the MAKER-based classifiers can be examined for better clinical decision making, as all the belief rules can be validated by experienced doctors. By using this approach, medical diagnostic knowledge can be learned from patient records more efficiently and effectively.

The rest of this paper is organized as follows. Sections 2 and 3 briefly introduce the methodologies of BRB inference and MAKER framework, respectively. A numerical example of using the new modeling approach to establish the MAKER-based classifiers for sepsis diagnosis is displayed in Section 4. In Section 5, the MAKER-based and alternative classifiers are compared in terms of their prediction performance for the sepsis data set. The inherent interpretability of this approach is also discussed in this section. Some conclusions and suggestions for future research are given in Section 6.

2. BRB inference

A belief rule base (BRB) is capable of capturing incomplete, fuzzy, and ignorant information, and nonlinear causal relationship between antecedent attributes and consequences (Yang, Liu, Wang, Sii, & Wang, 2006). It is composed of a finite number of belief rules, which is defined as follows (Yang et al., 2006).

$$\begin{aligned} &R_k: if\left(x_1 is\, A_1^k\right) \wedge \left(x_2 is\, A_2^k\right) \wedge \cdots \wedge \left(x_{M_k} is\, A_{M_k}^k\right), \\ &then\left\{\left(D_1, \beta_{1,k}\right), \left(D_2, \beta_{2,k}\right), \ldots, \left(D_N, \beta_{N,k}\right)\right\}, \\ &with\ a\ rule\ weight\ \lambda_k\ and\ attribute\ weights\ \delta_1, \delta_2, \ldots, \delta_{M_k}. \end{aligned} \tag{1}$$

In (1), R_k indicates the $k^{th}(k = 1, \ldots, L)$ belief rule. M_k is the number of antecedent attributes in the k^{th} rule, and $x_m(m = 1, \ldots, M_k)$ represents the m^{th} antecedent attribute. $A_j^k(j = 1, \ldots, M_k; k = 1, \ldots, L)$ is the referential value of the j^{th} antecedent attribute in the k^{th} rule. $'\wedge'$ is a logical connective which denotes the relationship of 'AND'. $\beta_{n,k}(n = 1, \ldots, N; k = 1, \ldots, L)$ signifies the belief degree assigned to a consequence D_n that can be initially provided by experts. If $\sum_{n=1}^{N}\beta_{n,k} = 1$, the k^{th} rule is complete. Otherwise, it is incomplete. In this study, a consequence is referred to as a class of the output variable (either "sepsis" or "non-sepsis"). Rule weight "λ_k" and attribute weight $\delta_m\,(m = 1, \ldots, M_k)$ are the relative weights of the k^{th} rule and the m^{th} antecedent attribute in the k^{th} rule, respectively. Both rule weight and attribute weight can be trained using optimization algorithms. The intuitive explanation of a belief rule is exhibited in Fig. 1.

Under the above belief rule base system, we can use a belief rule to represent a functional mapping between inputs of antecedent attribute and output possibly with uncertainties (Chen, Yang, Pan, Xu, & Zhou, 2015). This indicates that belief degree can be used to describe how likely a consequence element D_n in a belief rule is believed to be the consequence, given the logical relationship of a belief rule $if\left(x_1 is\, A_1^k\right) \wedge \left(x_2 is\, A_2^k\right) \wedge \cdots \wedge \left(x_{M_k} is\, A_{M_k}^k\right)$. Compared to a conventional IF-THEN rule, a belief rule provides a more realistic and informative scheme (Chen, Yang, Xu, & Yang, 2013). The belief rules can be established based on experts' knowledge, historical data, and random generation (Chen et al., 2015). If a belief rule base is generated using different ways, the detection methods for information inconsistency should be used to reduce information inconsistency, partiality, redundancy, and etc., (Ma, Lu, & Zhang, 2009). Once a belief rule base is constructed, we can use the knowledge embedded in belief rules to make inference for an instance (Yang et al., 2006).

3. Maximum likelihood evidential reasoning (MAKER) framework

Suppose that $e_{i,l}$ indicates the i^{th} piece of evidence acquired at a referential value $A_{i,l}$ ($A_{i,l}$ denotes the i^{th} referential value of the l^{th} input variable and referential value is a value of input variable where a piece of evidence is defined) from the l^{th} input variable x_l. Each piece of evidence can be partitioned into evidence elements. Evidence element $e_{i,l}(\theta)$ is referred to as an element of evidence that points exactly to a group of classes in the power set of the classes or a proposition which is referred to as an assertion that a group of classes is true or not given several pieces of evidence. The intersection of θ and $e_{i,l}(\theta)$, which is represented by $s_{i,l}(\theta) = \theta \cap e_{i,l}(\theta)$, indicates that θ is supported by evi-

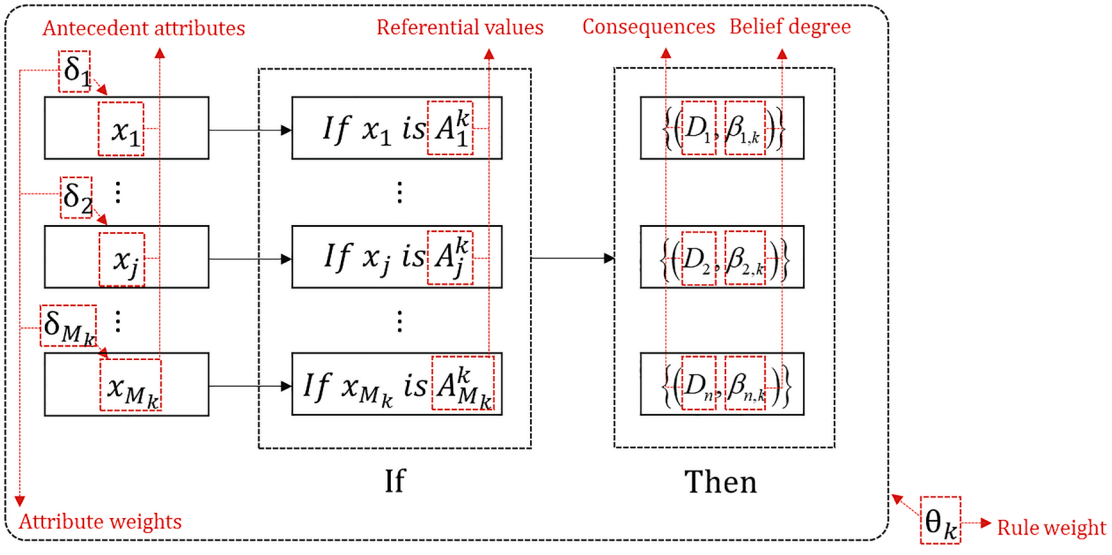


Fig. 1. Explanation of a belief rule.

dence $e_{i,l}$. Under the MAKER framework (Yang & Xu, 2017), multiple pieces of evidence obtained from input variables are combined together to form the joint support for a proposition, by considering three perspectives: basic probability, the reliability and weight of a piece of evidence, and the interrelationship between a pair of evidence.

A basic probability function (Yang & Xu, 2017) is profiled by an ordinary discrete probability function or a set function $p: 2^{\Theta} \rightarrow [0,1]$, if the following conditions are satisfied.

$$0 \leqslant p(\theta) \leqslant 1, \forall \theta \subseteq \Theta \tag{2}$$

$$\sum_{\theta \subseteq \Theta} p(\theta) = 1 \tag{3}$$

$$p(\varnothing) = 0 \tag{4}$$

In (2), (3), and (4), θ represents a subset of classes, which is named a proposition. $p(\theta)$ indicates the basic probability of proposition θ being true. $p(\theta)$ is an atom, which cannot be decomposed into pieces pointing to subsets of proposition θ, but can be assigned exactly to θ (Yang & Xu, 2017). Let $c_{\theta,i,l}$ be the likelihood of the i^{th} referential value of the l^{th} input variable x_l being observed given proposition θ. The basic probability $p_{\theta,i,l}$ can be acquired from a normalized likelihood, which is displayed in Eq. (5).

$$p_{\theta,i,l} = p_l\left(e_{i,l}(\theta)\right) = \frac{c_{\theta,i,l}}{\sum_{A \subseteq \Theta} c_{A,i,l}} \tag{5}$$

Based on Eq. (5), a piece of evidence $e_{i,l}$ can be defined by a basic probability distribution, which is shown in Eq. (6).

$$e_{i,l} = \left\{ \left(e_{i,l}(\theta), p_{\theta,i,l}\right), \forall \theta \subseteq \Theta \; and \sum_{\theta \subseteq \Theta} p_{\theta,i,l} = 1 \right\} \tag{6}$$

In Eq. (6), $p_{\theta,i,l}$ represents the basic probability of evidence $e_{i,l}$ pointing to proposition θ, which is acquired from an input variable x_l. What Eqs. (5) and (6) exhibit is the so-called one-dimensional evidence acquisition process.

Likewise, we can obtain joint basic probability from a joint likelihood function, following the Bayesian principle (Yang & Xu, 2017). Suppose that evidence: $e_{i,l}$ and $e_{j,m}$ are obtained from input variables: x_l and x_m, at $A_{i,l}$ and $A_{j,m}$, respectively. The joint basic probability of $e_{i,l}$ and $e_{j,m}$ pointing to proposition θ can be generated as follows.

$$p_{\theta,il,jm} = \frac{c_{\theta,il,jm}}{\sum_{A \subseteq \Theta} c_{A,il,jm}}, \forall \theta \subseteq \Theta \tag{7}$$

In Eq. (7), $c_{\theta,il,jm}$ indicates the joint likelihood of both $A_{i,l}$ and $A_{j,m}$ being observed given proposition θ. Based on the above discussions, we use an interdependence index 'α' to represent the interrelationship between evidential elements $e_{i,l}(A)$ and $e_{j,m}(B)$, which is defined in Eq. (8).

$$\alpha_{A,B,i,j} = \begin{cases} 0, \; if \; p_{A,i,l} = 0 \; or \; p_{B,j,m} = 0 \\ \dfrac{p_{A,B,il,jm}}{p_{A,i,l} p_{B,j,m}}, \; otherwise \end{cases} \tag{8}$$

In Eq. (8), $p_{A,i,l}$ and $p_{B,j,m}$ are the basic probabilities for single input variables: x_l and x_m, at $A_{i,l}$ and $A_{j,m}$, respectively. $p_{A,B,il,jm}$ denotes the joint basic probability of both x_l at $A_{i,l}$ and x_m at $A_{j,m}$ pointing to proposition $\theta(\theta = A \cap B, \theta \subseteq \Theta)$ simultaneously, which is generated using the same joint table for $p_{A,i,l}$ and $p_{B,j,m}$. The properties of the interdependence index 'α' are shown as follows.

$$\alpha_{A,B,i,j} = \begin{cases} 0, \; if \; e_{i,l}(A) \; and \; e_{j,m}(B) \; are \; disjoint \\ 1, \; if \; e_{i,l}(A) \; and \; e_{j,m}(B) \; are \; independent \end{cases} \tag{9}$$

The above calculations for $p_{A,i,l}$, $p_{B,j,m}$, $p_{A,B,i,j}$, and $\alpha_{A,B,i,j}$ are based on an assumption that all the evidential elements are fully reliable. However, as a matter of fact, seldom is evidence fully reliable. Hence, we need to consider the reliability of evidence, when measuring the degree of its support for proposition θ. Suppose that $r_{\theta,i,l}$ is the reliability of evidence $e_{i,l}$ that points to proposition θ. As shown in Eq. (10), it is defined as a conditional probability of θ being true given that $e_{i,l}$ points to θ, which essentially measures the quality of $e_{i,l}$ (Yang & Xu, 2017). It is associated with how data is generated and how $e_{i,l}$ is acquired from data (Yang & Xu, 2017).

$$r_{\theta,i,l} = p\left(\theta \middle| e_{i,l}(\theta)\right) \tag{10}$$

Based on Eq. (10), we can further generate the reliability of a piece of evidence $e_{i,l}$, which is shown in Eq. (11).

$$r_{i,l} = \sum_{\theta \subseteq \Theta} r_{\theta,i,l} p_{\theta,i,l} \tag{11}$$

In terms of whether $e_{i,l}$ and other pieces of evidence are acquired from the same data source, we can have two scenarios for generating the probability mass of proposition θ being supported by $e_{i,l}$, which are shown as follows.

Scenario 1: If $e_{i,l}$ and other pieces of evidence are acquired from the same data source, the probability mass of proposition θ being supported by $e_{i,l}$ is given in Eq. (12).

$$m_{\theta,i,l} = p\left(s_{i,l}(\theta)\right) = p\left(\theta \middle| e_{i,l}(\theta)\right) p\left(e_{i,l}(\theta)\right) = r_{\theta,i,l} p\left(e_{i,l}(\theta)\right) \tag{12}$$

Scenario 2: If $e_{i,l}$ is acquired from a data source that is different from other pieces of evidence, which is featured by a probability function p_l, the basic probability that $e_{i,l}$ points to θ is given by $p_l\left(e_{i,l}(\theta)\right)$, and the probability mass of $e_{i,l}$ supporting θ (denoted by $m_{\theta,i,l}$ and defined in Eq. (13)) is given by $p_l\left(s_{i,l}(\theta)\right)$ with $s_{i,l}(\theta) = \theta \cap e_{i,l}(\theta)$. According to the likelihood principle, the support from $e_{i,l}$ to θ have the same implication whether $e_{i,l}$ is measured by p or p_l, if $e_{i,l}$ is measured by the likelihood function defined in Eq. (7) (Yang & Xu, 2017). This means $p\left(s_{i,l}(\theta)\right)$ and $p_l\left(s_{i,l}(\theta)\right)$ should be proportional to each other, which can be indicated by $p\left(s_{i,l}(\theta)\right) = \omega_{i,l} p_l\left(s_{i,l}(\theta)\right)$. $\omega_{i,l}$ is a positive scaling factor for all $\theta \in \Theta$.

$$\begin{aligned} m_{\theta,i,l} &= p\left(s_{i,l}(\theta)\right) = \omega_{i,l} p_l\left(s_{i,l}(\theta)\right) \\ &= \omega_{i,l} p_l\left(\theta \cap e_{i,l}(\theta)\right) = \omega_{i,l} p_l\left(\theta \middle| e_{i,l}(\theta)\right) p_l\left(e_{i,l}(\theta)\right) \\ &= w_{\theta,i,l} p_l\left(e_{i,l}(\theta)\right) \end{aligned} \tag{13}$$

In Eq. (13), $w_{\theta,i,l} = \omega_{i,l} p_l\left(\theta \middle| e_{i,l}(\theta)\right)$ indicates the weight of an evidential element $e_{i,l}(\theta)$. It is proportional to the conditional probability of θ being true given that $e_{i,l}$ points to θ, which is measured by a probability function p_l established from data for x_l only (Yang & Xu, 2017). It is important to note that when $p = p_l$, $w_{\theta,i,l} = r_{\theta,i,l}$, which implies that $\omega_{i,l} = 1$.

Based on the above definitions and discussions, we can use a conjunctive MAKER rule to generate the combined probability $p_{\theta,e(2)}$ of proposition θ being jointly supported by two pieces of evidence $e_{i,l}$ and $e_{j,m}$, which is given as follows.

$$p_{\theta,e(2)} = \begin{cases} 0, \; \theta = \varnothing \\ \dfrac{\widetilde{m}_{\theta,e(2)}}{\sum_{C \subseteq \Theta} \widetilde{m}_{D,e(2)}}, \; \theta \subseteq \Theta, \; \theta \neq \varnothing \end{cases} \tag{14}$$

$$\widetilde{m}_{\theta,e(2)} = \left[\left(1 - r_{j,m}\right) m_{\theta,i,l} + \left(1 - r_{i,l}\right) m_{\theta,j,m}\right] + \sum_{A \cap B = \theta} \gamma_{A,B,il,jm} \alpha_{A,B,il,jm} m_{A,i,l} m_{B,j,m} \tag{15}$$

In Eqs. (14) and (15), $\widetilde{m}_{\theta,e(2)}$ represents the combined probability mass of both pieces of evidence $e_{i,l}$ and $e_{j,m}$ that jointly support proposition θ. $\gamma_{A,B,i,j}$ is a nonnegative parameter that reflects the degree of joint support of both $e_{i,l}$ and $e_{j,m}$ for θ, which is relative to the individual support of $e_{i,l}$ and $e_{j,m}$ pointing to propositions A and B, respectively (Yang & Xu, 2017). In the numerical example of this paper, $\gamma_{A,B,i,j}$ is assumed to be 1. Given that $w_{\theta,i,l} = w_{i,l}$ for any A, B, $\theta \subseteq \Theta$, the above conjunctive MAKER rule can be applied to the combination of

independent evidence, which reduces to evidential reasoning rule (Yang & Xu, 2013). The MAKER rule can further reduce to Dempster's rule (Shafer, 1976), given that $w_{i,l} = r_{i,l} = 1$. If there is no ambiguity in data, it can further reduce to Bayes's rule (Yang & Xu, 2017).

4. The new probabilistic modeling approach and its application to sepsis diagnosis

4.1. Problem formulation and data

The original sepsis data set for this study was collected from several hospitals in Northwest England. In line with the experts' knowledge and experience, the biomarkers: CRP, IL-6, IL-10, PCT, and WCC are chosen for classifier establishment. Based on the original data set, we use a series of data preparation techniques to generate the sepsis data set, which include patients' test results of these biomarkers and the associated diagnosis outcomes (sepsis or non-sepsis). The sepsis data set is further partitioned into four subsets for cross-validation, using a method based on stratified random sampling, which guarantees that all the subsets have similar class distribution of samples. Each subset is used for test set, while the remaining subsets are taken as training sets. The combination of each subset and the remaining ones is referred to as a fold for model training and testing.

Here, we take the training set of the first fold of the sepsis data set (henceforth the 'training set') as an example, to illustrate how to use the new modeling approach to establish a MAKER-based classifier for sepsis diagnosis. Suppose that a training input–output data set, which has N instances, is represented by $T = \left\{ \left(x_1^{(1)}, \ldots, x_l^{(1)}, \ldots, x_M^{(1)}, y^{(1)}\right), \ldots, \left(x_1^{(n)}, \ldots, x_l^{(n)}, \ldots, x_M^{(n)}, y^{(n)}\right), \ldots, \left(x_1^{(N)}, \ldots, x_l^{(N)}, \ldots, x_M^{(N)}, y^{(N)}\right) \right\}$. $\left(x_1^{(n)}, \ldots, x_l^{(n)}, \ldots, x_M^{(n)}, y^{(n)}\right)$ denotes an input–output sample pair, and $x_l^{(n)} (n = 1, \ldots, N; l = 1, \ldots, M)$ denotes the n^{th} observation of the l^{th} input variable, and $y^{(n)}$ denotes the class of the output variable for the n^{th} instance (represented by $\left(x_1^{(n)}, \ldots, x_l^{(n)}, \ldots, x_M^{(n)}\right)$). Each instance is characterized by M input variables. There are 272 instances in the 'training set' (the complete 'training set' is provided in Table S1 of the supplementary materials), and N is hence 272 in this example. The 'training set' consists of five input variables (i.e., the five biomarkers), and the value of M is taken as 5. Further, we use $x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, x_4^{(n)}$, and $x_5^{(n)}$ to denote an observation of input variables: CRP, IL-6, IL-10, PCT, and WCC, respectively. These instances are classified in a nominal output variable: $y = \{k | k = 1, \ldots, K\}$ where an integer represents a class of output variable. $y^{(n)}$ is used to represent the class of output variable for each instance. The output variable of the 'training set' is composed of two classes: "sepsis" and "non-sepsis", which are signified by "1" and "2", respectively. The output variable is therefore denoted as $y = \{1, 2\}$. The parameters of a MAKER-based classifier include referential values of observations of input variables and weights of referential values pointing to different classes of output variable. For the purpose of illustration, we use the referential values and weights trained by a machine learning algorithm (henceforth 'trained referential values' and 'trained weights', respectively), to construct a MAKER-based classifier step by step in the remaining part of this section. The remaining part is organized into two subsections including the key components of the new probabilistic modeling approach and the application of the approach to sepsis diagnosis. The components of the new approach involve evidence acquisition from data, analysis of evidence interdependence, belief rule-base inference, prediction for sepsis diagnosis, and training of MAKER-based classifier.

4.2. The key components of the new probabilistic modeling approach

4.2.1. Evidence acquisition from data

The first step for classifier establishment is evidence acquisition which is based on the referential values (Xu, Zheng, Yang, Xu, & Chen, 2017; Yao, 2019). With the referential values, we can transform an observation $x_l^{(n)}$ into a belief distribution over the referential values $A_{i,l}$, which is shown in Eq. (16).

$$S\left(x_l^{(n)}\right) = \left\{ \left(A_{i,l}, \alpha_{n,i,l}^{(k)}\right) | n = 1, \ldots, N; l = 1, \ldots, M; i = 1, \ldots, V_l; k = 1, \ldots, K \right\}$$

where

$$\alpha_{n,i,l}^{(k)} = \frac{A_{i+1,l} - x_l^{(n)}}{A_{i+1,l} - A_{i,l}} \text{ and } \alpha_{n,i+1,l}^{(k)} = 1 - \alpha_{n,i,l}^{(k)}, \text{ if } A_{i,l} \leqslant x_l^{(n)} \leqslant A_{i+1,l};$$

$$\alpha_{n,i',l}^{(k)} = 0. \text{ for } i' = 1, \ldots, V_l \text{ and } i' \neq i, i+1. \tag{16}$$

In Eq. (16), $\alpha_{n,i,l}^{(k)}$ is the similarity degree to which the n^{th} observation of the l^{th} input variable (indicated by $x_l^{(n)}$) matches the i^{th} referential value of the l^{th} input variable (signified by '$A_{i,l}$'), which points to a class (represented by 'k') of output variable.

Matching degrees for each referential value are aggregated for each class to generate an associated total matching degree for each class, as shown in Eq. (17). The total matching degree of a referential value for a class is then treated as the frequency for the referential value matching the class. Table 1 shows all the frequencies of the referential values for an input variable.

$$\alpha_{i,l}^{(k)} = \sum_{n=1}^{N} \alpha_{n,i,l}^{(k)} \tag{17}$$

In Table 1, we have $\sum_{k=1}^{K} \sum_{i=1}^{V_l} \alpha_{i,l}^{(k)} = N$. For each row of Table 1, we can generate a sum of frequency values ($\delta_l^{(k)}$), using $\delta_l^{(k)} = \sum_{i=1}^{V_l} \alpha_{i,l}^{(k)}$. Let $c_{i,l}^{(k)}$ be the likelihood to which the i^{th} referential value of the l^{th} input variable is observed given that the k^{th} class of output variable is known. Likelihood can be calculated as shown in Eq. (18). Table 2 displays the likelihoods calculated from the frequencies in Table 1.

$$c_{i,l}^{(k)} = \begin{cases} \dfrac{\alpha_{i,l}^{(k)}}{\sum_{i=1}^{V_l} \alpha_{i,l}^{(k)}}, \text{ if } \sum_{i=1}^{V_l} \alpha_{i,l}^{(k)} \neq 0 \\ 0, \text{ if } \sum_{i=1}^{V_l} \alpha_{i,l}^{(k)} = 0 \end{cases} \tag{18}$$

It is noted from Table 2 that $\sum_{k=1}^{K} \sum_{i=1}^{V_l} c_{i,l}^{(k)} = K$. Based on $\eta_{i,l} = \sum_{k=1}^{K} c_{i,l}^{(k)}$, the sum of the likelihoods (signified by '$\eta_{i,l}$') can be produced for each column of Table 2. With these likelihoods, the probability for a referential value of an input variable pointing to a class of output variable ($p_{i,l}^{(k)}$) can be generated by Eq. (19). Table 3 shows the probabilities calculated through the normalization of likelihoods in Table 2 by Eq. (19).

Table 1
Frequencies of referential values for an input variable.

$y^{(n)} \backslash x_l^{(n)}$	$A_{1,l}$	⋯	$A_{i,l}$	⋯	$A_{V_l,l}$	*Total*
1	$\alpha_{1,l}^{(1)}$	⋯	$\alpha_{i,l}^{(1)}$	⋯	$\alpha_{V_l,l}^{(1)}$	$\sum_{i=1}^{V_l} \alpha_{i,l}^{(1)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	$\alpha_{1,l}^{(k)}$	⋯	$\alpha_{i,l}^{(k)}$	⋯	$\alpha_{V_l,l}^{(k)}$	$\sum_{i=1}^{V_l} \alpha_{i,l}^{(k)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	$\alpha_{1,l}^{(K)}$	⋯	$\alpha_{i,l}^{(K)}$	⋯	$\alpha_{V_l,l}^{(K)}$	$\sum_{i=1}^{V_l} \alpha_{i,l}^{(K)}$
Total	$\sum_{k=1}^{K} \alpha_{1,l}^{(k)}$	⋯	$\sum_{k=1}^{K} \alpha_{i,l}^{(k)}$	⋯	$\sum_{k=1}^{K} \alpha_{V_l,l}^{(k)}$	N

Table 2
Likelihoods for referential values of input variables pointing to classes of output variable.

$y_n \backslash x_{n,l}$	$A_{1,l}$	⋯	$A_{i,l}$	⋯	$A_{V_l,l}$	*Total*
1	$c_{1,l}^{(1)}$	⋯	$c_{i,l}^{(1)}$	⋯	$c_{V_l,l}^{(1)}$	$\sum_{i=1}^{V_l} c_{i,l}^{(1)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	$c_{1,l}^{(k)}$	⋯	$c_{i,l}^{(k)}$	⋯	$c_{V_l,l}^{(k)}$	$\sum_{i=1}^{V_l} c_{i,l}^{(k)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	$c_{1,l}^{(K)}$	⋯	$c_{i,l}^{(K)}$	⋯	$c_{V_l,l}^{(K)}$	$\sum_{i=1}^{V_l} c_{i,l}^{(K)}$
Total	$\sum_{k=1}^{K} c_{1,l}^{(k)}$	⋯	$\sum_{k=1}^{K} c_{i,l}^{(k)}$	⋯	$\sum_{k=1}^{K} c_{V_l,l}^{(k)}$	K

Table 3
Probabilities for referential values of input variables pointing to classes of output variable.

$y_n \backslash x_{n,l}$	$A_{1,l}$	⋯	$A_{i,l}$	⋯	$A_{V_l,l}$
1	$p_{1,l}^{(1)}$	⋯	$p_{i,l}^{(1)}$	⋯	$p_{V_l,l}^{(1)}$
⋮	⋮	⋮	⋮	⋮	⋮
k	$p_{1,l}^{(k)}$	⋯	$p_{i,l}^{(k)}$	⋯	$p_{V_l,l}^{(k)}$
⋮	⋮	⋮	⋮	⋮	⋮
K	$p_{1,l}^{(K)}$	⋯	$p_{i,l}^{(K)}$	⋯	$p_{V_l,l}^{(K)}$

$$p_{i,l}^{(k)} = \begin{cases} \dfrac{c_{i,l}^{(k)}}{\sum_{k=1}^{K} c_{i,l}^{(k)}}, \ if \sum_{k=1}^{K} c_{i,l}^{(k)} \neq 0 \\ 0, \ if \sum_{k=1}^{K} c_{i,l}^{(k)} = 0 \end{cases} \tag{19}$$

Further, we can acquire a piece of evidence at each referential value of Table 3 under the MAKER framework, which is defined as a belief distribution shown in Eq. (6) of Section 3.

4.2.2. Analysis of evidence interdependence

In order to achieve greater predictive power, it is necessary to combine multiple pieces of evidence to generate predicted probabilities of patients having sepsis or not. In the original evidential reasoning (ER) rule, any two pieces of evidence to be combined are assumed to be independent from each other, which is simplistic. In the MAKER framework, the interdependence between two pieces of evidence is taken into consideration through the introduction of an interdependence index 'α', which is defined in Eq. (8) of Section 3. To generate the interdependence index between two pieces of evidence, we need to first estimate the joint probabilities to which both pieces of evidence point to classes of output variable, using Eq. (20). Let $x_l^{(n)}$ and $x_m^{(n)}$ be the n^{th} observation of the l^{th} and the m^{th} input variable, respectively. The joint belief distribution for the instance of these two observations $S\left(\left\{x_l^{(n)}, x_m^{(n)}\right\}\right)$ is given as follows.

$$\begin{aligned} S\left(\left\{x_l^{(n)}, x_m^{(n)}\right\}\right) &= \left\{\left(\left\{A_{i,l}, A_{j,m}\right\}, \alpha_{n,il,jm}^{(k)}\right) | n = 1, \ldots, N; l, m = 1, \ldots, M; \right. \\ & \left. i = 1, \ldots, V_l; j = 1, \ldots, V_m; k = 1, \ldots, K \right\} where \alpha_{n,il,jm}^{(k)} \\ &= \alpha_{n,i,l}^{(k)} \alpha_{n,j,m}^{(k)}, \alpha_{n,i,l}^{(k)} = \frac{A_{i+1,l} - x_{n,l}}{A_{i+1,l} - A_{i,l}}, and \alpha_{n,i+1,l}^{(k)} \\ &= 1 - \alpha_{n,i,l}^{(k)}, if A_{i,l} \leqslant x_{n,l} \leqslant A_{i+1,l}; \alpha_{n,i',l}^{(k)} = 0, if i' = 1, \ldots, V_l, and i' \neq i, i+1. \end{aligned} \tag{20}$$

In Eq. (20), $\alpha_{n,il,jm}^{(k)}$ indicates the joint similarity degree to which the instance: $\left\{x_l^{(n)}, x_m^{(n)}\right\}$ matches the combination of referential values: $\left\{A_{i,l}, A_{j,m}\right\}$, which points to a class of output variable (indicated by 'k'). After all the instances are transformed into belief distributions over referential values combinations in a similar way, we can generate the aggregated frequency values of referential values combinations associated with a class of output variable ($\alpha_{il,jm}^{(k)}$), using Eq. (21).

$$\alpha_{il,jm}^{(k)} = \sum_{n=1}^{N} \alpha_{n,il,jm}^{(k)} \tag{21}$$

Based on these frequency values, the joint likelihoods of observing referential values given classes of output variable ($c_{il,jm}^{(k)}$) can be generated by Eqs. (22) and (23).

$$\delta_{lm}^{(k)} = \sum_{i=1}^{V_l} \sum_{j=1}^{V_m} \alpha_{il,jm}^{(k)} \tag{22}$$

$$c_{il,jm}^{(k)} = \frac{\alpha_{il,jm}^{(k)}}{\delta_{lm}^{(k)}} \tag{23}$$

With likelihoods, we can use Eq. (24) to generate the joint probabilities of referential values combinations pointing to classes of output variable ($p_{il,jm}^{(k)}$).

$$p_{il,jm}^{(k)} = \begin{cases} \dfrac{c_{il,jm}^{(k)}}{\sum_{k=1}^{K} c_{il,jm}^{(k)}}, \ if \sum_{k=1}^{K} c_{il,jm}^{(k)} \neq 0; \\ 0, \ if \sum_{k=1}^{K} c_{il,jm}^{(k)} = 0. \end{cases} \tag{24}$$

Using Eqs. (19) and (24), we can generate $\alpha_{A,B,i,j}$ representing interdependence indices between a pair of evidential elements indicated by $e_{i,l}(A)$ and $e_{j,m}(B)$, which is shown in Eq. (8) of Section 3.

4.2.3. Belief rule-base inference

Based on the acquired evidence and the interdependence analysis, we are now in a position to construct a belief rule base to infer the likelihood of a patient having sepsis based on their test results of the biomarkers. According to the belief rule described in (1) of Section 2, the antecedent part of the belief rule, which is expressed in the form of '*if* $\left(x_1 is A_1^k\right) \wedge \left(x_2 is A_2^k\right) \wedge \ldots \wedge \left(x_2 is A_{M_k}^k\right)$', in this numerical example should be interpreted as 'if a patient's test results of biomarkers are just equal to the referential values of the respective biomarkers'. The associated consequent part of the belief rule, which is expressed in the form of '*then* $\left\{\left(D_1, \beta_{1,k}\right), \left(D_2, \beta_{2,k}\right), \ldots, \left(D_N, \beta_{N,k}\right)\right\}$', should then be understood as 'the consequents that the patient is diagnosed with sepsis and not diagnosed with sepsis have respective probabilities'.

To obtain the consequences' probabilities of all the belief rules in a belief rule base, all pieces of evidence (each piece of evidence is acquired at a referential value from a different input variable) in each combination of evidence are combined to generate the associated belief rule. This process of evidence combination uses the conjunctive MAKER rule shown in Eqs. (14) and (15), which is based on the 'trained weights' mentioned in Section 4.1. In this process, the reliability $\left(r_{\theta,i,l}\right)$ and weight $\left(w_{\theta,i,l}\right)$ for each evidential element are essentially the same, as all sepsis data in this numerical example were collected from Northwest England, which can be generally considered homogeneous.

4.2.4. Prediction for sepsis diagnosis

Based on the belief rules activated by an instance, we need to calculate the similarity degree to which each instance matches the associated activated combinations of referential values, using Eq. (25). Eq. (25) is for the instances featured by two input variables, and it can be further extended to those featured by five input variables.

$$\alpha_{n,il,jm} = \alpha_{n,i,l}\alpha_{n,j,m}$$

where

$$\alpha_{n,i,l} = \frac{A_{i+1,l} - x_{n,l}}{A_{i+1,l} - A_{i,l}}, and \alpha_{n,i+1,l} = 1 - \alpha_{n,i,l}, if A_{i,l} \leqslant x_{n,l} \leqslant A_{i+1,l}; \quad (25)$$

$$\alpha_{n,i',l} = 0, if\ i' = 1, \ldots, T_l, and\ i' \neq i, i+1.$$

With these similarity degrees, we can calculate the joint similarity degree to which the instance matches the associated activated combination of referential values, by multiplying all the associated similarity degrees of an observation to the activated referential values. A joint similarity degree indicates the degree to which we should use the activated belief rules to predict the probability for each class of output variable associated with an instance.

Having generated the joint similarity degrees, we are able to combine the belief rules activated by an instance to predict the probability for each class of output variable. To combine these belief rules, the reliabilities $(r_{e(L)})$ and weights $(w_{e(L)})$ of each belief rule are needed. They can be initially assigned from expert knowledge or trained using data as shown in Section 4.6. $r_{e(L)}$ is equal to $w_{e(L)}$ in this study.

Based on the joint degree of similarity to which the instance matches the activated belief rules, and the reliability (weight) of each activated belief rule, we can generate the updated reliability (weight) of each activated belief rule for the instance (represented by '$r'_{e(L)}$'), which is displayed in Eq. (26).

$$r'_{e(L)} = \alpha_{n,e(L)} r_{e(L)} \quad (26)$$

In Eq. (26), $\alpha_{n,e(L)}$ indicates the similarity degree to which an instance matches a belief rule, which is generated in a similar way shown in Eq. (25). The updated reliability (weight) helps us consider the degree to which we should use the activated belief rules to generate the predicted probabilities for classes of output variable associated with an instance.

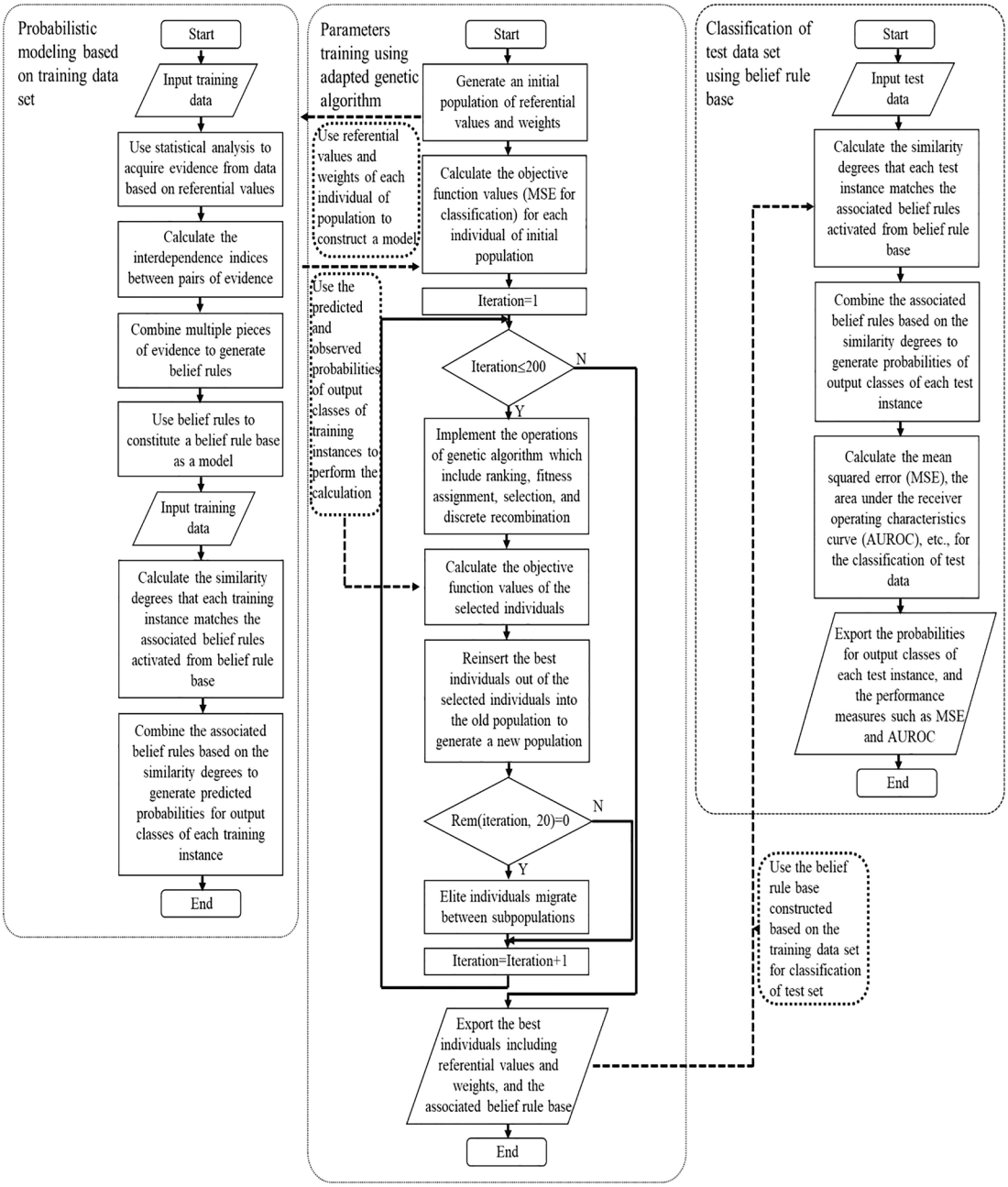


Fig. 2. Probabilistic modeling, training of model parameters, and classification.

With the associated updated reliability (weight), we can combine the belief rules activated by the instance to predict the probabilities for classes of output variable assigned to the instance, using the conjunctive MAKER rule.

4.2.5. Training of MAKER-based classifier

In the above process for MAKER-based classifier establishment, $A_{i,l}$, $r_{\theta,i,l}$, $w_{\theta,i,l}$, etc., are the adjustable parameters assigned for inference. These parameters can be trained based on the sepsis data set, and the trained parameters can be used to establish a MAKER-based classifier to predict the probabilities that patients have sepsis or not. An optimal learning model is then proposed to train the parameters by maximizing the likelihood of true class of output variable, which is shown in Eq. (27).

$$\begin{aligned} &min\ \delta = \frac{1}{2N}\sum_{n=1}^{N}\sum_{\theta\subseteq\Theta}\left(p_n(\theta)-\widehat{p}_n(\theta)\right)^2 \\ &s.t. A_{i,l}, r_{\theta,i,l}, w_{\theta,i,l}, \gamma_{A,B,il,jm} \in \Omega \end{aligned} \tag{27}$$

In Eq. (27), $\widehat{p}_n(\theta)$ and $p_n(\theta)$ are the observed and predicted probability, respectively, of the proposition θ being true, which is given in the n^{th} sample of the sepsis data set. The objective of the optimal learning model is to minimize the mean squared error (MSE) to make $p_n(\theta)$ as close to $\widehat{p}_n(\theta)$ as possible. Ω is the feasible space of the parameters, which include the constraints such as $0 \leqslant r_{\theta,i,l} \leqslant 1$. An adapted genetic algorithm is further used to train the parameters of a MAKER-based classifier, which is based on the optimal learning model. In the algorithm, each chromosome of a population include the referential values that multiple pieces of evidence are located at and the weights for evidential elements. This algorithm is hence appropriate for parallel computing.

Based on the above five key components of the probabilistic modeling approach, Fig. 2 shows the flow diagram describing the probabilistic modeling based on the training data set, adapted genetic algorithm, classification based on BRB for test data set, and their interrelationship. From the above process, it is evident that the MAKER-based classifier is interpretable and transparent, which is established by the new modeling approach integrating statistical analysis, belief rule-based inference, and machine learning.

4.3. An example of application of the new approach to sepsis diagnosis

In this example, the "training set" mentioned in Section 4.1 is used to show how the modeling approach described in Section 4.2 is applied to sepsis diagnosis. Corresponding to the first component of the modeling approach, the first step for classifier establishment is evidence acquisition which are based on the "trained referential values" and the associated boundary referential values. The boundary referential values indicate the minima and maxima of observations of input variables. The referential values of input variables for classifier establishment, which include the 'trained referential values' and the boundary referential values, are displayed in Table 4.

Table 4
The referential values of input variables for classifier establishment.

Input variables	CRP	IL-6	IL-10	PCT	WCC
Boundary referential values (minima)	2.9000	0.8200	0.1400	0.0500	0.0000
Trained referential values	190.2799	101.9637	97.6625	9.8923	5.7066
Boundary referential values (maxima)	690.0000	20971.0100	4563.8700	200.0000	66.0000

Using Eq. (16), an observation of the "training set" can be transformed into a belief distribution over a referential value. For example, an observation of CRP is 86, which is located between two adjacent referential values of CRP: 2.9000 and 190.2799. The similarity degree to which 86 (pointing to sepsis class) matches the referential values: 2.9000 and 190.2799 are generated by $\frac{190.2799 - 86.0000}{190.2799 - 2.9000} = 0.5565$ and $1 - \frac{190.2799 - 86.0000}{190.2799 - 2.9000} = 0.4435$, respectively. Hence, the belief distribution of an observation: 86 over referential values: 2.9000 and 190.2799 is generated as $(0.5565, 0.4435)$. As 86 is not located between 190.2799 and 690.0000, the similarity degree of 86 to referential value: 690.0000 is 0.

After all the observations of each input variable are transformed into belief distributions, we can use Eq. (17) to aggregate the similarity degrees of all the observations in class 'k' to each referential value of an input variable. In this way, we can generate the observed frequency values of referential values of each input variable associated with classes of output variable. Tables 5 and 6 (associated with Tables 1 and 2) display the frequency values of referential values of input variables: CRP and IL-6, respectively. For each row of the tables, we can generate a sum of frequency values ($\delta_l^{(k)}$), using $\delta_l^{(k)} = \sum_{i=1}^{T_l} \alpha_{i,l}^{(k)}$. For instance, the sum of frequency values for the first row of Table 5 is generated as $\sum_{i=1}^{3} \alpha_{i,1}^{(1)} = 7.7714 + 57.9262 + 11.3024 = 77.0000$.

Next, we can generate the likelihood of observing a referential value of an input variable given a class of output variable, using Eq. (18). Tables 7 and 8 display the likelihoods of observing the referential values of input variables: CRP and IL-6, respectively, which point to classes of output variable. An example is that the likelihood of observing the referential value of CRP: 190.2799 given that the output class is 'sepsis' (indicated by '1′) is produced by $c_{2,1}^{(1)} = \frac{57.9262}{77.0000} \approx 0.7523$ (57.9262 and 77.0000 are from Table 5).

Based on the equation $\eta_{i,l} = \sum_{k=1}^{K} c_{i,l}^{(k)}$, the sum of the likelihoods (signified by '$\eta_{i,l}$') can be produced for each column of Tables 7 and 8. For example, the sum of the likelihoods for the column beginning with '2.9000′ is given by $\eta_{1,1} = 0.1009 + 0.3430 = 0.4439$. With these likelihoods, the probability to which a referential value of an input variable points to a class of output variable ($p_{i,l}^{(k)}$) can be generated by Eq. (19). We take the probability of the referential value of CRP: 190.2799 pointing to the class: 'sepsis' (represented by '1′) as an example. It is generated by $p_{2,1}^{(1)} = \frac{0.7523}{1.3298} \approx 0.5657$.

Tables 9 and 10 exhibit the probabilities of the referential values of input variables: CRP and IL-6, respectively, pointing to classes of the output variable. From the probabilities shown in Tables 9 and 10, we can acquire a piece of evidence from each referential value of an input variable (corresponding to Eq. (6) of Section 3). For instance, as shown in Table 9, the probabilities 0.2274 and 0.7726, associated with referential value: 2.9000, indicate that, if the CRP test result for a patient is 2.9000, the probability of this patient having sepsis is 0.2274 and that of this patient not having sepsis is 0.7726. We can hence acquire a piece of evidence ($e_{1,1}$) from the CRP test result: 2.9000, in that it points to the sepsis class with a probability of 0.2274 and the non-sepsis class with that of 0.7726.

Corresponding to the analysis of evidence interdependence shown in Section 4.2, we take the instance: $\{82.0000, 32.1800\}$ as an example to demonstrate how to use Eq. (20) to generate the joint similarity degree

Table 5
The frequency values of the referential values of CRP associated with each class of output variable.

$y_n = k \backslash A_{i,1}$	2.9000	190.2799	690.0000	$\delta_1^{(k)} = \sum_{i=1}^{3} \alpha_{i,1}^{(k)}$
1	7.7714	57.9262	11.3024	77.0000
2	66.8805	112.6240	15.4954	195.0000

Table 6
The frequency values of referential values of IL-6 associated with each class of output variable.

$y_n = k \backslash A_{i,2}$	0.8200	101.9637	20971.0100	$\delta_2^{(k)} = \sum_{i=1}^{3} \alpha_{i,2}^{(k)}$
1	4.8761	68.3554	3.7686	77.0000
2	84.8470	108.3789	1.7741	195.0000

Table 7
The likelihoods of observing referential values of CRP given each class of output variable.

$y_n = k \backslash A_{i,1}$	2.9000	190.2799	690.0000
1	0.1009	0.7523	0.1468
2	0.3430	0.5776	0.0795
$\eta_{i,1} = \sum_{k=1}^{2} c_{i,1}^{(k)}$	0.4439	1.3298	0.2262

Table 8
The likelihoods of observing referential values of IL-6 given each class of output variable.

$y_n = k \backslash A_{i,2}$	0.8200	101.9637	20971.0100
1	0.0633	0.8877	0.0489
2	0.4351	0.5558	0.0091
$\eta_{i,2} = \sum_{k=1}^{2} c_{i,2}^{(k)}$	0.4984	1.4435	0.0580

Table 9
The probabilities of referential values of CRP pointing to each class of output variable.

$y_n = k \backslash A_{i,1}$	2.9000	190.2799	690.0000
1	0.2274	0.5657	0.6488
2	0.7726	0.4343	0.3512

Table 10
The probabilities of referential values of IL-6 pointing to each class of output variable.

$y_n = k \backslash A_{i,2}$	0.8200	101.9637	20971.0100
1	0.1270	0.6150	0.8432
2	0.8730	0.3850	0.1568

to which an instance matches a combination of referential values, for the generation of the interdependence index between two pieces of evidence. 82.0000 and 32.1800 are from input variables: CRP and IL-6, respectively. These two observations can activate two sets of two adjacent referential values: $\{2.9000, 190.2799\}$ and $\{0.8200, 101.9637\}$, respectively. The similarity degree of 82.0000 to 2.9000 is generated as $\frac{190.2799 - 82.0000}{190.2799 - 2.9000} \approx 0.5779$, and that of 32.1800 to 101.9637 is given by $\frac{32.1800 - 0.8200}{101.9637 - 0.8200} \approx 0.3101$. Thus, the joint similarity degree to which $\{82.0000, 32.1800\}$ matches the set of two adjacent referential values $\{2.9000, 101.9637\}$ is generated as $0.5779 \times 0.3101 \approx 0.1792$. Thus we can have the associated belief distributions over the sets of the adjacent referential values.

When all the instances are transformed into belief distributions over referential values combinations in a similar way, Eqs. (21) to (24) can be used to generate the joint probabilities of referential values combinations pointing to classes of output variable. Table 11 displays the joint probabilities of all the combinations of referential values from input variables CRP and IL-6 pointing to classes of output variable.

Based on the probabilities displayed in Tables 9–11, we can use Eq. (8) to generate the interdependence indices between a piece of evidence from CRP and that from the IL-6. For example, the interdependence index between a piece of evidence from CRP at 190.2799 and that from IL to 6 at 0.8200, which points to sepsis class (signified by '1'), can be produced by $\frac{0.1810}{0.5657 \times 0.1270} \approx 2.519$ (0.5657, 0.1270, and 0.1810 are acquired from Tables 6–8, respectively). The interdependence indices between two pieces of evidence from CRP and IL-6 are exhibited in Table 12.

From Table 12, it is clear that the test results for CRP and IL-6 are moderately independent from each other, as the interdependence indices are between 1 and 5. For instance, the interdependence between CRP test result: 190.2799 and the IL-6 test result: 101.9637 is moderate, as the associated interdependence index for the sepsis class (indicated by '1') is 1.8679 and that for the non-sepsis class (represented by '2') is 2.0941.

Based on the acquired evidence and interdependence analysis, a belief rule base can be constructed to infer the likelihood of a patient having sepsis based on their test results of the biomarkers (corresponding to the belief rule-base inference shown in Section 4.2). For example, through calculation of evidence fusion, we can generate probabilities: 0.0086 and 0.9914 for sepsis and non-sepsis classes, given the evidence combination at the referential values combination: {2.9000, 0.8200, 0.1400, 0.0500, 66.0000}. The associated belief rule is expressed as 'if a patient's test results of biomarkers: CRP, IL-6, IL-10, PCT, and WCC are equal to 2.9000, 0.8200, 0.1400, 0.0500, and 66.0000, respectively, the probabilities that the patient has and does not have sepsis are 0.0086 and 0.9914, respectively. As each of the five input variables has a total of three referential values (shown in Table 1) and each referential value is assigned to a piece of evidence, there are altogether $3^5 = 243$ combinations of five pieces of evidence at referential values combinations. With the 243 evidence combinations and the associated probabilities for sepsis and non-sepsis classes, we can hence generate 243 belief rules, which constitutes a belief rule base (which is shown in Table S2).

Each observation of an input variable can activate two adjacent referential values between which the observation is located. Hence, each instance of the 'training set', which has five input variables, can activate $2^5 = 32$ belief rules out of the belief rule base. For example, an instance: {158.0000, 619.4500, 120.1000, 123.8600, 32.5000} from the 'training set' can activate 32 rules from the belief rule base for inference, which are displayed in Table I of Appendix. From the 32 belief rules activated by each instance of the 'training set', we can find the associated probabilities for sepsis and non-sepsis classes, which can be used to further generate the predicted probabilities for each instance.

Based on the belief rules activated by an instance, we can use Eq. (25) to calculate the similarity degree to which each instance matches the associated activated combinations of referential values (corresponding to prediction for sepsis diagnosis in Section 4.2). For example, 158.0000 is an observation of the instance: $\{158.0000, 619.4500, 120.1000, 123.8600, 32.5000\}$, and according to Table 4, the referential values of CRP activated by the observation: 158.0000 are 2.9000 and 190.2799.

Table 11
The joint probabilities of referential values combinations of CRP and IL-6 pointing to classes of output variable.

$A_{i,1}$	$A_{j,2}$	$y_n = 1$	$y_n = 2$
2.9000	0.8200	0.0596	0.9404
2.9000	101.9637	0.3922	0.6078
2.9000	20971.0100	0.8109	0.1891
190.2799	0.8200	0.1810	0.8190
190.2799	101.9637	0.6498	0.3502
190.2799	20971.0100	0.8312	0.1688
690.0000	0.8200	0.2615	0.7385
690.0000	101.9637	0.6614	0.3386
690.0000	20971.0100	0.9089	0.0911

Table 12
The interdependence indices between two pieces of evidence from CRP and IL-6, respectively, pointing to classes of output variable.

$e_{i,1}$ at $A_{i,1}$	$e_{j,2}$ at $A_{j,2}$	$y_n = 1$	$y_n = 2$
2.9000	0.8200	2.0634	1.3943
2.9000	101.9637	2.8052	2.0430
2.9000	20971.0100	4.2298	1.5610
190.2799	0.8200	2.5186	2.1602
190.2799	101.9637	1.8679	2.0941
190.2799	20971.0100	1.7425	2.4792
690.0000	0.8200	3.1723	2.4087
690.0000	101.9637	1.6578	2.5038
690.0000	20971.0100	1.6614	1.6542

Using Eq. (25), the similarity degree of the observation: 158.0000 to the referential value: 2.9000 is calculated as $\frac{190.2799 - 158.0000}{190.2799 - 2.9000} \approx 0.1723$ and that of 158.0000 to 190.2799 is calculated as $1 - \frac{190.2799 - 158.0000}{190.2799 - 2.9000} \approx 0.8277$. These similarity degrees indicate that the observation: 158.0000 matches the referential value: 190.2799 to a high degree and 2.9000 to a low degree. In this way, we can calculate the similarity degrees of each observation of the instance: $\{158.0000, 619.4500, 120.1000, 123.8600, 32.5000\}$ to the associated activated referential values. Based on these similarity degrees, we can obtain the joint similarity degree to which the instance matches the associated activated combination of referential values using the extended version of Eq. (25).

Having generated the joint similarity degrees, we are able to use the conjunctive MAKER rule to combine the belief rules activated by the instance: $\{158.0000, 619.4500, 120.1000, 123.8600, 32.5000\}$ to predict the probability for each class of output variable. For the instance: $\{158.0000, 619.4500, 120.1000, 123.8600, 32.5000\}$, the predicted probability of the sepsis class is 0.9553 and that of the non-sepsis class is 0.0447. It indicates that if a patient's test results for CRP, IL-6, IL-10, PCT, and WCC are 158.0000, 619.4500, 120.1000, 123.8600, and 32.5000, respectively, the probability that this patient has sepsis is 0.9553 and the probability that this patient does not have sepsis is 0.0447.

5. Results and discussion

The plotting of receiver operating characteristic (ROC) curve is generally used to demonstrate the discrimination accuracy of a diagnostic classifier for predicting whether a patient has an illness (Faraggi & Reiser, 2002). An individual is typically evaluated as ill (positive) or healthy (negative), given that the associated test value is greater, less, or equal to a specific threshold value (Kong et al., 2016). The threshold values are associated with probabilities for being true positive (sensitivity) and true negative (specificity) (Kong et al., 2016). For all the possible threshold values, the associated 'sensitivity' and '1-specificity' are plotted to generate an ROC curve (Kong et al., 2016). The area under the ROC curve (AUROC) is the most common global index for diagnosis accuracy (Faraggi & Reiser, 2002).

In this study, we use a number of performance measures including AUROC, sensitivity (SEN), specificity (SPC), and accuracy (ACC), to compare the performance of the MAKER-based and alternative classifiers. The alternative classifiers for comparison consists of a number of groups of candidate sub-classifiers, which include classification trees, discriminant analysis, support vector machine, k-nearest neighbors, ensembles, naïve Bayes, and artificial neural networks. As mentioned, the sepsis data set is partitioned into four folds of training and test sets for cross-validation. The MAKER-based classifier and all the candidate sub-classifiers (displayed in Table 13) are trained based on each training set. Within each group of sub-classifiers, the sub-classifier which has the highest average training accuracy is selected as the group representative classifier. The trained MAKER-based and representative classifiers are then tested based on each test set, and we can hence generate the performance measures (the threshold values for generative classifiers are 0.5) for each test set and the associated average measures. Based on these performance measures, we can provide a comprehensive evaluation for each classifier's classification performance.

Table 13
The alternative classifiers for sepsis diagnosis.

Alternative classifiers	Candidate sub-classifiers	Representative classifier
Decision tree	Simple tree, Medium tree, and Complex tree	Complex tree
Discriminant analysis	Linear discriminant, and Quadratic discriminant	Quadratic discriminant
Support vector machine (SVM)	Linear SVM, Quardratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, and Coarse Gaussian SVM	Fine Gaussian SVM
K-nearest neighbor (KNN)	Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, and Weighted KNN	Fine KNN, and Weighted KNN
Ensemble	Boosted trees, Bagged trees, Subspace discriminant, Subspace KNN, and RUSBoosted trees	Bagged trees, and Subspace KNN
Naïve Bayes	Naïve Bayes	Naïve Bayes
Artificial neural network (ANN)	Feed-forward backpropagation	Feed-forward backpropagation

It is noted that the training for the MAKER-based classifier is terminated based on the stopping criteria proposed by Yao (2019), when there are five trained referential values for each input variable of the training sets, which generally considers the balance between model complexity and accuracy. Among these trained MAKER-based classifiers, we select the ones which have one trained referential value for each input variable, to be the representative for the comparison with the alternative classifiers. The reason is that the selected ones have the highest average training accuracy among all the trained MAKER-based classifiers. The average results of performance measures of all the classifiers for the comparison are exhibited in Table 14 (the complete table for performance measures are shown in Table S3 of the Supplementary materials).

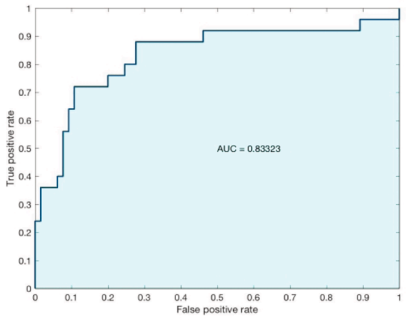
From Table 14, it is clear that, although the MAKER-based classifier does not produce the best performance results among all the classifiers, its performance result is generally near the best in accuracy, and the ones are near to the average in sensitivity and specificity across all alternatives. To be more specific, the MAKER-based classifiers have an average accuracy of 77.34% that is close to the best accuracy of 80.67%. They have an average sensitivity of 45.12%, which is above the average sensitivity of 41.29% among all the alternatives. They have an average specificity of 90% that is very close to the average specificity of 90.87% among all the alternatives. All of this suggests that the MAKER-based classifier has a similar performance with the alternative ones in terms of sensitivity, specificity, and accuracy.

Regarding the AUROC, Fig. 3 presents the ROC curves of the MAKER-based classifier and the best alternative one (with the best AUROC) over each test set. Table 15 displays the AUROCs of all the classifiers for test

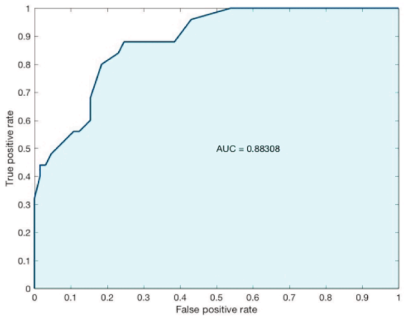
Table 14
Average performance measures of the classifiers for sepsis diagnosis.

Classifiers/Measures	SEN (%)	SPC (%)	ACC (%)
MAKER	45.12	90.00	77.34
Complex tree	62.81	83.08	77.36
Quadratic discriminant	34.27	96.15	78.73
Fine Gaussian SVM	21.58	95.38	74.58
Fine KNN	47.16	83.85	73.49
Weighted KNN	35.35	90.77	75.15
Ensemble: bagged trees	51.04	90.00	79.00
Ensemble: subspace KNN	50.00	85.39	75.43
Naïve Bayes	36.27	98.08	80.67
ANN: feed-forward backpropagation	31.54	92.69	75.42

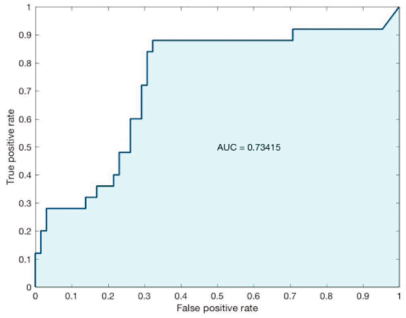
sets. From Table 15, it can be observed that the average AUROC of the MAKER-based classifier is 0.8127, which is the third largest one among all the AUROCs of the 10 classifiers. A general rule of thumb for using AUROC to judge the classification capability of a classifier is that AUROCs that are in the range of 0.7–0.8, 0.8–0.9, and 0.9–1.0 are considered acceptable, excellent and outstanding discrimination, respectively (Hosmer & Lemeshow, 2000). The MAKER-based classifiers can be considered excellent, as its average AUROC is 0.8127, which is



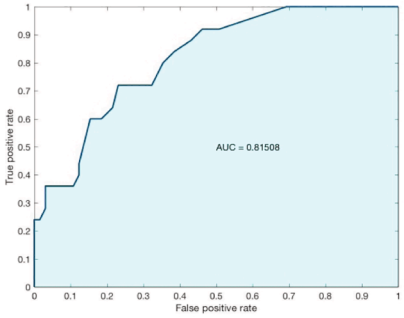
(a) MAKER for the 1st test set



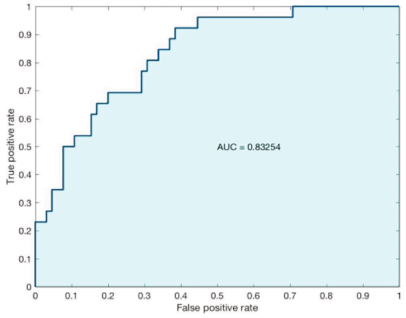
(b) Bagged trees for the 1st test set



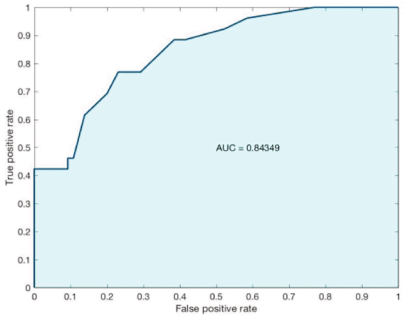
(c) MAKER for the 2nd test set



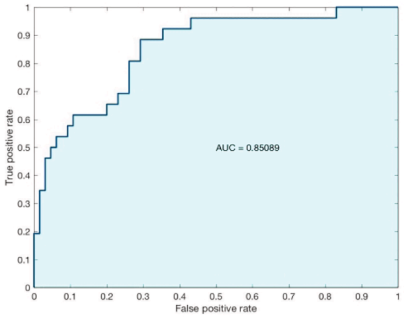
(d) Bagged trees for the 2nd test set



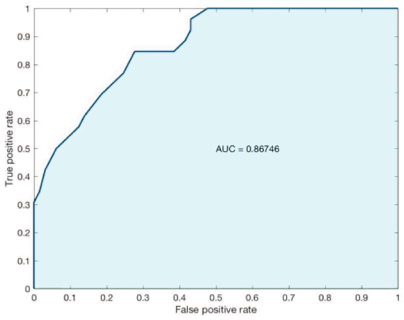
(e) MAKER for the 3rd test set



(f) Bagged trees for the 3rd test set



(g) MAKER for the 4th test set



(h) Bagged trees for the 4th test set

Fig. 3. The ROC curves of the MAKER-based and best classifiers for each test set.

Table 15
The area under the receiver operating characteristic curve (AUROC) of each of the classifiers for the sepsis diagnosis.

classifiers/measures	AUROC				
	1st test set	2nd test set	3rd test set	4th test set	avg.
MAKER	0.8332	0.7342	0.8325	0.8509	0.8127
Complex tree	0.8129	0.7674	0.6962	0.7997	0.7690
Quadratic discriminant	0.8486	0.7905	0.8284	0.7894	0.8142
Fine Gaussian SVM	0.7655	0.7508	0.7781	0.8284	0.7807
Fine KNN	0.6754	0.6908	0.5962	0.6577	0.6550
Weighted KNN	0.7828	0.7443	0.7651	0.7101	0.7506
Ensemble: bagged trees	0.8831	0.8151	0.8435	0.8675	0.8523
Ensemble: subspace KNN	0.8323	0.7120	0.8284	0.7902	0.7907
Naïve Bayes	0.8452	0.7923	0.8249	0.7852	0.8119
ANN: feed-forward backpropagation	0.8782	0.7840	0.8095	0.7675	0.8098

lower than the average AUROCs of discriminant classifier (0.8142), and bagged trees (0.8523).

As shown in the average AUROCs, the MAKER-based classifier for sepsis diagnosis performs better than the majority of alternative classifiers, and close to the best one. In addition to this, the MAKER-based classifier established by the new modeling approach features an inherent interpretability, which is reflected in the three aspects as follows. (1) The evidence acquisition is interpretable. Evidence is acquired from referential values of continuous data directly using likelihood analysis: sample casting and likelihoods normalization. Under the MAKER framework, we can combine multiple pieces of evidence. We can also capture interdependence between a pair of evidence using interdependence index based on marginal and joint likelihood functions, rather than assumes interdependence between a pair of evidence. (2) The inference mechanism is interpretable. The inference with belief rule base (BRB) on which the MAKER-based classifier is based is inherently interpretable from the perspective of the structure of belief rules. (3) The parameters determination is interpretable. Each parameter of the classifier is statistically or preferentially meaningful from the perspectives of probabilistic inference and evidence-based decision-making, and can be optimally trained based on whatever data is available.

More importantly, clinical decision-making (e.g., sepsis diagnosis), as much as art as science, is a final outcome of a complicated process that relies on scientific knowledge and clinical experiences accumulated through years of training and practice (Hans, Shevroja, & Leslie, 2021). It is difficult to assess the interpretability of a classifier without the consideration of specific domain knowledge and experience. The MAKER-based classifier is inherently interpretable from the perspective of supporting clinical decision-making. Specifically, the determination of referential values of specific biomarkers can be combined with the knowledge and experience of doctors. The weights of biomarkers is a reflection of doctors' knowledge and understanding of specific biomarkers in sepsis diagnosis. The adjustment of classifier parameters, i.e. referential values and weights, can be used to better reflect doctors' judgement and opinions about patients' condition. The BRB inference based on the combination of biomarkers is similar to the process of clinical adjudication made by doctors. An instance from sepsis patients' continuous biomarker data can activate multiple pieces of evidence from different biomarkers. It is of great necessity to combine activated evidence to generate belief rules reflecting actual patients' condition. This is similar to the process that doctors integrate information of patients' biomarkers to form the basis of clinical adjudication for sepsis. Each given instance is capable of activating belief rules from BRB, which can be combined to generate a predicted output distribution. This resembles the process that doctors link the evidence-based knowledge and experience to the patients' condition to make clinical adjudication about sepsis for patients. Due to the inherent interpretable structure, BRB can be used to document how changes in biomarkers affect the predicted output distribution to support doctors' clinical adjudication.

Overall, the MAKER-based classifier is designed to enhance doctor's perception and support doctors in clinical decision-making, and learning or planning with the help of artificial intelligence, instead of replacing doctors in the processes automated by artificial intelligence. The interpretability of the MAKER-based classifier is assessed in this paper without unnecessary repetitions, as there are many other papers and books explaining the built-in interpretable structure of BRB inference of the classifier and the interaction between the MAKER-based classifier and doctors' knowledge and experiences (Kong et al., 2012, 2016). As such, the MAKER-based classifier established by the new modeling approach features an inherent interpretability, and hence a recommended tool to help doctors make reasonable diagnostic decisions about sepsis.

Regarding the time of optimization of the classifiers established by the new approach, it can be noted that the offline MAKER-based classifier can be trained based on large quantity of data. The associated online MAKER-based classifier can be established based on the offline one. The inference of online classifier can be very fast as we only need to fine-tune the parameters of the classifier such as the weights of evidence elements to make the classifier fit for a batch of new data.

In addition to the proposed probabilistic modeling approach of this study, Fuzzy rule-based model is also an interpretable rule based approach. Fuzzy rule-based model, which is based on fuzzy sets, is used to express knowledge in a set of fuzzy rules to deal with complex real world problems. It is capable of dealing with uncertainty, imprecision, and non-linearity. Due to its interpretability, it can be interpreted, verified, and modified by human experts. Compared with the fuzzy rule-based model, the MAKER-based model acquires evidence directly from data using statistical analysis. In addition, evidence is combined using the MAKER framework to generate the belief rules representing the probabilistic input–output relationship. The rules in the MAKER-based model are both knowledge-based and data-driven, while those in the fuzzy rule-based models are predefined by specialized knowledge and expertise. The belief rules of the MARKER-based models are developed from the traditional if-then rules that are used in the fuzzy rule-based models. The belief rules are designed with belief degrees embedded in all possible consequents of a rule. The belief rules are hence capable of capturing vagueness, incompleteness, and random nonlinear causal relationship, while the traditional if-then rules can be represented as a special case. The MAKER framework is effective for aggregating both qualitative and quantitative information under uncertainty (especially that caused by ignorance or incomplete information).

This paper focuses on the "interpretability" of the proposed approach in the application of sepsis diagnosis. The importance of "interpretability" has been recognized by many researchers in the field of machine learning, but the studies of quantifiable measures for the quality of interpretability are just at the initial stages. Schmidt and Biessmann (2019) proposed a quantitative measure for the quality of interpretability based on the mutual information between the observations and model predictions. They also developed a quantitative measure of trust in machine learning decisions. Carrington, Fieguth, and Chen (2018) identified criteria for model transparency from literatures and proposed an objective binary measure of model transparency. In the future studies, a new quantitative measure will be developed from the perspectives such as separability, stability, validity, correctness, and compactness of explanations of model predictions.

6. Conclusions

In summary, this paper presents the application of a new probabilistic modeling approach to sepsis diagnosis. A brief introduction to sepsis and its diagnosis is provided at the beginning of this paper. Then we present the methodology of BRB inference and MAKER framework, respectively. Next, a numerical example based on a data set collected from Northwest England is given to describe how the new approach is

used to construct MAKER-based classifiers for sepsis diagnosis, from the perspectives of evidence acquisition, interdependence analysis, BRB inference, prediction, and model training. The MAKER-based and alternative classifiers are trained and validated based on a four-fold cross-validation. Based on the cross-validation classification results, we generate the results of performance measures including sensitivity, specificity, accuracy, and AUROC for these classifiers. The performance comparison shows that the MAKER-based classifier is near-optimal, which outperforms 7 out of 9 alternative ones for sepsis diagnosis.

More importantly, the MAKER-based classifier established by the new modeling approach is characterized by an inherent interpretability. It is reflected in three aspects as follows. (1) The evidence acquisition using likelihood analysis is interpretable. (2) The inference based on BRB is interpretable. (3) The parameters determination using machine learning algorithms is interpretable. Overall, the MAKER-based classifier established using the new approach, is highly transparent and interpretable. It is hence potentially helpful for sepsis diagnosis. Further studies are necessary for establishing a model based on the sepsis data sets with 'unknown' class and handling the high multiplicative complexity for the number of referential values of input variables in the belief rule base (Xu et al., 2017). It is also necessary to conduct research about quantification of "interpretability" in the context of machine learning.

CRediT authorship contribution statement

Shuaiyu Yao: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Jian-Bo Yang:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing - review & editing. **Dong-Ling Xu:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Paul Dark:** Data curation, Investigation, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their sincere thanks for the support from European Union's Horizon 2020 Research and Innovation Programme RISE under grant agreement no. 823759 (REMESH), the Alliance Strategic Research Fund of the University of Manchester, and Natural Science Foundation of China under the grant No. U1709215, No. 71971139, No. 72071056, and No. 71601060.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2021.115333.

References

Angeletti, S., Battistoni, F., Fioravanti, M., Bernardini, S., & Dicuonzo, G. (2013). Procalcitonin and mid-regional pro-adrenomedullin test combination in sepsis diagnosis. *Clinical Chemistry and Laboratory Medicine*. https://doi.org/10.1515/cclm-2012-0595

Angeletti, S., Dicuonzo, G., Fioravanti, M., De Cesaris, M., Fogolari, M., Lo Presti, A., … De Florio, L. (2015). Procalcitonin, MR-Proadrenomedullin, and cytokines measurement in sepsis diagnosis: Advantages from test combination. *Disease Markers, 2015*, 1–14. https://doi.org/10.1155/2015/951532

Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., & Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine, 29*(7), 1303–1310. https://doi.org/10.1097/00003246-200107000-00002

Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., … Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in Biology and Medicine, 109*, 79–84. https://doi.org/10.1016/j.compbiomed.2019.04.027

Bhattacharjee, P., Edelson, D. P., & Churpek, M. M. (2017). Identifying Patients With Sepsis on the Hospital Wards. *Chest, 151*(4), 898–907. https://doi.org/10.1016/j.chest.2016.06.020

Bozza, F. A., Salluh, J. I., Japiassu, A. M., Soares, M., Assis, E. F., Gomes, R. N., … Bozza, P. T. (2007). Cytokine profiles as markers of disease severity in sepsis: A multiplex analysis. *Critical Care, 11*(2), R49. https://doi.org/10.1186/cc5783

Bradley, J. S., Guidos, R., Baragona, S., Bartlett, J. G., Rubinstein, E., Zhanel, G. G., … Tillotson, G. S. (2007). Anti-infective research and development-problems, challenges, and solutions. *Lancet Infectious Diseases, 7*(1), 68–78. https://doi.org/10.1016/S1473-3099(06)70689-2

Carrington, A., Fieguth, P., & Chen, H. (2018). Measures of model interpretability for model selection. *International Cross-Domain Conference for Machine Learning and Knowledge Extraction, 2018*(11015), 329–349. https://doi.org/10.1007/978-3-319-99740-7_24

Chen, M., Lu, X., Hu, L., Liu, P., Zhao, W., Yan, H., … Tan, H. (2017). Development and validation of a mortality risk model for pediatric sepsis. Medicine (United States). https://doi.org/10.1097/MD.0000000000006923.

Chen, Y.-W., Yang, J.-B., Pan, C.-C., Xu, D.-L., & Zhou, Z.-J. (2015). Identification of uncertain nonlinear systems: Constructing belief rule-based models. *Knowledge-Based Systems, 73*, 124–133. https://doi.org/10.1016/j.knosys.2014.09.010

Chen, Y. W., Yang, J. B., Xu, D. L., & Yang, S. L. (2013). On the inference and approximation properties of belief rule based systems. *Information Sciences, 234*, 121–135. https://doi.org/10.1016/j.ins.2013.01.022

Chiesa, C., Natale, F., Pascone, R., Osborn, J. F., Pacifico, L., Bonci, E., & De Curtis, M. (2011). C reactive protein and procalcitonin: Reference intervals for preterm and term newborns during the early neonatal period. *Clinica Chimica Acta, 412*(11-12), 1053–1059. https://doi.org/10.1016/j.cca.2011.02.020

Dark, P., Blackwood, B., Gates, S., McAuley, D., Perkins, G. D., McMullan, R., … Warhurst, G. (2015). Accuracy of LightCycler® SeptiFast for the detection and identification of pathogens in the blood of patients with suspected sepsis: A systematic review and meta-analysis. *Intensive Care Medicine, 41*(1), 21–33. https://doi.org/10.1007/s00134-014-3553-8

De Backer, D., & Dorman, T. (2017). Surviving sepsis guidelines: A continuous move toward better care of patients with sepsis. *JAMA - Journal of the American Medical Association, 317*(8), 807. https://doi.org/10.1001/jama.2017.0059

Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics, 38*(2), 325–339. https://doi.org/10.1214/aoms/1177698950

Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological), 30*(2), 205–232.

Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., … Das, R. (2016). Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Medical Informatics, 4*(3), e28. https://doi.org/10.2196/medinform.5909

Elfeky, S., Golabi, P., Otgonsuren, M., Djurkovic, S., Schmidt, M. E., & Younossi, Z. M. (2017). The epidemiologic characteristics, temporal trends, predictors of death, and discharge disposition in patients with a diagnosis of sepsis: A cross-sectional retrospective cohort study. *Journal of Critical Care, 39*, 48–55. https://doi.org/10.1016/j.jcrc.2017.01.006

Faisal, M., Scally, A., Richardson, D., Beatson, K., Howes, R., Speed, K., & Mohammed, M. A. (2018). Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Critical Care Medicine, 46*(4), 612–618. https://doi.org/10.1097/CCM.0000000000002967

Faraggi, D., & Reiser, B. (2002). Estimation of the area under the ROC curve. *Statistics in Medicine, 21*(20), 3093–3106. https://doi.org/10.1002/(ISSN)1097-025810.1002/sim.v21:2010.1002/sim.1228

Fischetti, T., Lantz, B., Abedin, J., Mittal, H. V., Makhabel, B., Berlinger, E., … Daroczi, G. (2016). R: Data Analysis and Visualization. Packt Publishing Ltd.

Fleischmann, C., Scherag, A., Adhikari, N. K. J., Hartog, C. S., Tsaganos, T., Schlattmann, P., … Reinhart, K. (2016). Assessment of global incidence and mortality of hospital-treated sepsis current estimates and limitations. *American Journal of Respiratory and Critical Care Medicine, 193*(3), 259–272. https://doi.org/10.1164/rccm.201504-0781OC

Friedman, G., Silva, E., & Vincent, J.-L. (1998). Has the mortality of septic shock changed with time? *Critical Care Medicine, 26*(12), 2078–2086. https://doi.org/10.1097/00003246-199812000-00045

Görges, M., & Ansermino, J. M. (2020). Augmented intelligence in pediatric anesthesia and pediatric critical care. *Current Opinion in Anaesthesiology, 33*, 404–410. https://doi.org/10.1097/ACO.0000000000000845

Hans, D., Shevroja, E., & Leslie, W. D. (2021). February 1). Evolution in fracture risk assessment: Artificial versus augmented intelligence. *Osteoporosis International, 32*(2), 209–212. https://doi.org/10.1007/s00198-020-05737-x

Hebbar, A. (2017). Augmented intelligence: Enhancing human capabilities. Proceedings - 2017 3rd IEEE International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2017, 2017-Decem, 251–254. https://doi.org/10.1109/ICRCICN.2017.8234515.

Henriquez-Camacho, C., & Losa, J. (2014). Biomarkers for sepsis. *BioMed Research International, 2014*, 1–6. https://doi.org/10.1155/2014/547818

Heyland, D. K., Hopman, W., Coo, H., Tranmer, J., & McColl, M. A. (2000). Long-term health-related quality of life in survivors of sepsis. Short Form 36: A valid and reliable measure of health-related quality of life. *Critical Care Medicine., 28*(11), 3599–3605. https://doi.org/10.1097/00003246-200011000-00006

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed). New York: Wiley.

Jones, G. R., & Lowes, J. A. (1996). The systemic inflammatory response syndrome as a predictor of bacteraemia and outcome from sepsis. *QJM – Monthly Journal of the Association of Physicians, 89*(7), 515–522. https://doi.org/10.1093/qjmed/89.7.515

Kellum, J. A., Kong, L., Fink, M. P., Weissfeld, L. A., Yealy, D. M., Pinsky, M. R., … Angus, D. C. (2007). Understanding the inflammatory cytokine response in pneumonia and sepsis: Results of the genetic and inflammatory markers of sepsis (GenIMS) study. *Archives of Internal Medicine*. https://doi.org/10.1001/archinte.167.15.1655

Kollef, M. H., Sherman, G., Ward, S., & Fraser, V. J. (1999). Inadequate antimicrobial treatment of infections: A risk factor for hospital mortality among critically III patients. *Chest, 115*(2), 462–474. https://doi.org/10.1378/chest.115.2.462

Kong, G., Xu, D.-L., Body, R., Yang, J.-B., Mackway-Jones, K., & Carley, S. (2012). A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research, 219*(3), 564–573. https://doi.org/10.1016/j.ejor.2011.10.044

Kong, G., Xu, D.-L., Yang, J.-B., Yin, X., Wang, T., Jiang, B., & Hu, Y. (2016). Belief rule-based inference for predicting trauma outcome. *Knowledge-Based Systems, 95*, 35–44. https://doi.org/10.1016/j.knosys.2015.12.002

Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., … Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine, 34*(6), 1589–1596. https://doi.org/10.1097/01.CCM.0000217961.75225.E9

Kumar, G., Kumar, N., Taneja, A., Kaleekal, T., Tarima, S., McGinley, E., … Nanchal, R. (2011). Nationwide trends of severe sepsis in the 21st century (2000–2007). *Chest, 140*(5), 1223–1231. https://doi.org/10.1378/chest.11-0352

Kumar, S., Tripathy, S., Jyoti, A., & Singh, S. G. (2019). Recent advances in biosensors for diagnosis and detection of sepsis: A comprehensive review. *Biosensors and Bioelectronics, 124-125*, 205–215. https://doi.org/10.1016/j.bios.2018.10.034

Lagu, T., Rothberg, M. B., Shieh, M.-S., Pekow, P. S., Steingrub, J. S., & Lindenauer, P. K. (2012). Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Critical Care Medicine, 40*(3), 754–761. https://doi.org/10.1097/CCM.0b013e318232db65

Lam, H. S., Wong, S. P. S., Cheung, H. M., Chu, W. C. W., Wong, R. P. O., Chui, K. M., … Ng, P. C. (2011). Early diagnosis of intra-abdominal inflammation and sepsis by neutrophil CD64 expression in newborns. *Neonatology, 99*(2), 118–124. https://doi.org/10.1159/000311289

Lamping, F., Jack, T., Rübsamen, N., Sasse, M., Beerbaum, P., Mikolajczyk, R. T., … Karch, A. (2018). Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children – A data-driven approach using machine-learning algorithms. *BMC Pediatrics, 18*(1). https://doi.org/10.1186/s12887-018-1082-2

López-Martínez, F., Núñez-Valdez, E. R., Lorduy Gomez, J., & García-Díaz, V. (2019). A neural network approach to predict early neonatal sepsis. *Computers and Electrical Engineering, 76*, 379–388. https://doi.org/10.1016/j.compeleceng.2019.04.015

Ma, J., Lu, J., & Zhang, G. (2009). Information inconsistencies detection using a rule-map technique. *Expert Systems with Applications, 36*(10), 12510–12519. https://doi.org/10.1016/j.eswa.2009.04.040

Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., … Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open, 8*(1), e017833. https://doi.org/10.1136/bmjopen-2017-017833

Martin, G. S., Mannino, D. M., Eaton, S., & Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *New England Journal of Medicine, 348*(16), 1546–1554. https://doi.org/10.1056/NEJMoa022139

Matson, A., Soni, N., & Sheldon, J. (1991). C-reactive protein as a diagnostic test of sepsis in the critically ill. *Anaesthesia and Intensive Care, 19*(2), 182–186. https://doi.org/10.1177/0310057X9101900204

Ng, A. Y., & Jordan, M. I. (2002). *On discriminative vs. Generative classifiers: A comparison of logistic regression and naive bayes*. Advances in Neural Information Processing Systems.

Oonsivilai, M., Mo, Y., Luangasanatip, N., Lubell, Y., Miliya, T., Tan, P., … Cooper, B. S. (2018). Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia [version 1; referees: 2 approved]. *Wellcome Open Research, 3*, 131. https://doi.org/10.12688/wellcomeopenres10.12688/wellcomeopenres.14847.1

Ou, S.-M., Chu, H., Chao, P.-W., Lee, Y.-J., Kuo, S.-C., Chen, T.-J., … Chen, Y.-T. (2016). Long-term mortality and major adverse cardiovascular events in sepsis survivors a nationwide population-based study. *American Journal of Respiratory and Critical Care Medicine, 194*(2), 209–217. https://doi.org/10.1164/rccm.201510-2023OC

Patil, S., Dutta, S., Attri, S. V., Ray, P., & Kumar, P. (2016). Serial C reactive protein values predict sensitivity of organisms to empirical antibiotics in neonates: A nested case-control study. Archives of Disease in Childhood: Fetal and Neonatal Edition. https://doi.org/10.1136/archdischild-2015-309158.

Pavlou, P. A. (2018). Internet of things – Will humans be replaced or augmented? *GfK Marketing Intelligence Review, 10*(2), 42–47. https://doi.org/10.2478/gfkmir-2018-0017

Perl, T. M., Dvorak, L., Wenzel, R. P., & Hwang, T. (1995). Long-term Survival and Function After Suspected Gram-negative Sepsis. *JAMA: The. Journal of the American Medical Association*. https://doi.org/10.1001/jama.1995.03530040066043

Póvoa, P., Coelho, L., Almeida, E., Fernandes, A., Mealha, R., Moreira, P., & Sabino, H. (2005). C-reactive protein as a marker of infection in critically ill patients. *Clinical Microbiology and Infection, 11*(2), 101–108. https://doi.org/10.1111/j.1469-0691.2004.01044.x

Prucha, M., Bellingan, G., & Zazula, R. (2015). Sepsis biomarkers. *Clinica Chimica Acta, 440*, 97–103. https://doi.org/10.1016/j.cca.2014.11.012

Qiu, F., & Jensen, J. R. (2004). Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing, 25*(9), 1749–1768. https://doi.org/10.1080/01431160310001618798

Ratzinger, F., Haslacher, H., Perkmann, T., Pinzan, M., Anner, P., Makristathis, A., … Dorffner, G. (2018). Machine learning for fast identification of bacteraemia in SIRS patients treated on standard care wards: A cohort study. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-30236-9

Reinhart, K., Bauer, M., Riedemann, N. C., & Hartog, C. S. (2012). New approaches to sepsis: Molecular diagnostics and biomarkers. *Clinical Microbiology Reviews, 25*(4), 609–634. https://doi.org/10.1128/CMR.00016-12

Saqib, M., Sha, Y., & Wang, M. D. (2018). Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. https://doi.org/10.1109/EMBC.2018.8513254

Scherpf, M., Gräßer, F., Malberg, H., & Zaunseder, S. (2019). Predicting sepsis with a recurrent neural network using the MIMIC III database. *Computers in Biology and Medicine, 113*, 103395. https://doi.org/10.1016/j.compbiomed.2019.103395

Schmidt, P., & Biessmann, F. (2019). Quantifying interpretability and trust in machine learning systems. *ArXiv*.

Schulte, W., Bernhagen, J., & Bucala, R. (2013). Cytokines in sepsis: Potent immunoregulators and potential therapeutic targets - An updated view. *Mediators of Inflammation, 2013*, 1–16. https://doi.org/10.1155/2013/165974

Shafer, G. (1976). *A Mathematical Theory of Evidence*. In Princeton University Press (illustrate): Princeton.

Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., … Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - Journal of the American Medical Association, 315*(8), 801. https://doi.org/10.1001/jama.2016.0287

S. Spoto E. Cella M. de Cesaris L. Locorriere S. Mazzaroppi E. Nobile … S. Angeletti Procalcitonin and MR-Proadrenomedullin Combination with SOFA and qSOFA Scores for Sepsis Diagnosis and Prognosis: A Diagnostic Algorithm 50 1 2018 44 52 10.1097/SHK.0000000000001023.

Strobl, C. (2008). *Statistical Issues in Machine Learning Towards Reliable Split Selection and Variable Importance Measures*. Cuvillier Verlag.

Taneja, I., Reddy, B., Damhorst, G., Dave Zhao, S., Hassan, U., Price, Z., … Zhu, R. (2017). Combining biomarkers with EMR data to identify patients in different phases of sepsis. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-09766-1

R.A. Taylor J.R. Pare A.K. Venkatesh H. Mowafi E.R. Melnick W. Fleischman M.K. Hall A. Jones Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Driven 23 3 2016 269 278 10.1111/acem.12876.

Thompson, D., Pepys, M. B., & Wood, S. P. (1999). The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure, 7*(2), 169–177. https://doi.org/10.1016/S0969-2126(99)80023-9

Torio, C. M., & Moore, B. J. (2006). National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204. In Healthcare Cost and Utilization Project (HCUP) Statistical Briefs.

Vincent, J.-L. (2016). The Clinical Challenge of Sepsis Identification and Monitoring. *PLoS Medicine, 13*(5), e1002022. https://doi.org/10.1371/journal.pmed.1002022

Vincent, J.-L., Sakr, Y., Sprung, C. L., Ranieri, V. M., Reinhart, K., Gerlach, H., … Payen, D. (2006). Sepsis in European intensive care units: Results of the SOAP study. *Critical Care Medicine, 34*(2), 344–353. https://doi.org/10.1097/01.CCM.0000194725.48928.3A

Wolf, T. A., Wimalawansa, S. J., & Razzaque, M. S. (2019). Procalcitonin as a biomarker for critically ill patients with sepsis: Effects of vitamin D supplementation. *Journal of Steroid Biochemistry and Molecular Biology, 193*, 105428. https://doi.org/10.1016/j.jsbmb.2019.105428

Xu, X., Zheng, J., Yang, J.-b., Xu, D.-L., & Chen, Y.-W. (2017). Data classification using evidence reasoning rule. *Knowledge-Based Systems, 116*, 144–151. https://doi.org/10.1016/j.knosys.2016.11.001

Yang, J. B., Liu, J., Wang, J., Sii, H. S., & Wang, H. W. (2006). Belief rule-base inference methodology using the evidential reasoning approach – RIMER. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*. https://doi.org/10.1109/TSMCA.2005.851270

Yang, J. B., & Xu, D. L. (2013). Evidential reasoning rule for evidence combination. *Artificial Intelligence, 205*, 1–29. https://doi.org/10.1016/j.artint.2013.09.003

Yang, J. B., & Xu, D. L. (2017). Inferential modelling and decision making with data. *IEEE International Conference on Automation and Computing: Addressing Global Challenges through Automation and Computing*.

Yang, Y.i., Xie, J., Guo, F., Longhini, F., Gao, Z., Huang, Y., & Qiu, H. (2016). Combination of C-reactive protein, procalcitonin and sepsis-related organ failure score for the diagnosis of sepsis in critical patients. *Annals of Intensive Care, 6*(1). https://doi.org/10.1186/s13613-016-0153-5

S. Yao Investigation into Rule-based Inferential Modelling and Prediction with Application in Healthcare (University of Manchester) Retrieved from https://www.research.manchester.ac.uk/portal/en/theses/investigation-into-rulebased-inferential-modelling-and-prediction-with-application-in-healthcare (e73ae49a-887e-4305-8973-728c1bbe251e).html 2019.

Yende, S., D'Angelo, G., Kellum, J. A., Weissfeld, L., Fine, J., Welch, R. D., … Angus, D. C. (2008). Inflammatory markers at hospital discharge predict subsequent mortality

after pneumonia and sepsis. *American Journal of Respiratory and Critical Care Medicine, 177*(11), 1242–1247. https://doi.org/10.1164/rccm.200712-1777OC

Yentis, S. M., Soni, N., & Sheldon, J. (1995). C-reactive protein as an indicator of resolution of sepsis in the intensive care unit. *Intensive Care Medicine, 21*(7), 602–605. https://doi.org/10.1007/BF01700168

Zhou, Z.-G., Liu, F., Jiao, L.-C., Zhou, Z.-J., Yang, J.-B., Gong, M.-G., & Zhang, X.-P. (2013). A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer. *Knowledge-Based Systems, 54*, 128–136. https://doi.org/10.1016/j.knosys.2013.09.001

Zhou, Z.-G., Liu, F., Li, L.-L., Jiao, L.-C., Zhou, Z.-J., Yang, J.-B., & Wang, Z.-L. (2015). A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. *Knowledge-Based Systems, 85*, 62–70. https://doi.org/10.1016/j.knosys.2015.04.019