Expert Systems With Applications 201 (2022) 117026

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Improving chronic disease management for children with knowledge graphs and artificial intelligence

Gang Yu [a,b,f], Mohammad Tabatabaei [d], József Mezei [d,e,f,*], Qianhui Zhong [d,f], Siyu Chen [d], Zheming Li [a,b,f], Jing Li [a,b,f], LiQi Shu [c], Qiang Shu [b,**]

[a] Department of IT Center, The Children's Hospital, Zhejiang University School of Medicine, 3333 Binsheng Road, Hangzhou, 310052, China
[b] National Clinical Research Center for Child Health, 3333 Binsheng Road, Hangzhou, 310052, China
[c] Department of Neurology, The Warren Alpert Medical School of Brown University, 593 Eddy St, Providence, RI 02903, United States
[d] Avaintec Oy, Itämerenkatu 1, Helsinki, 00180, Finland
[e] Åbo Akademi University, Tuomiokirkontori 3, Turku, 20500, Finland
[f] Sino-Finland Joint AI Laboratory, 3333 Binsheng Road, Hangzhou, 310052, China

ARTICLE INFO

Keywords:
Chronic disease management
Artificial intelligence
Health care application
Big data
Machine learning

ABSTRACT

Chronic diseases for children pose serious challenges from a health management perspective. When not implemented in a well-designed manner, an inefficient management platform can have a significant negative impact on patients and the utilization of health care resources. Innovations of recent years in information technology, artificial intelligence and machine learning provide possibilities to design and implement knowledge-based systems and platforms that follow-up, monitor and advise child patients with a chronic disease in an automated manner. In this article we propose the Artificial Intelligence Chronic Management System that combines artificial intelligence, knowledge graph, big data and internet of things in a platform to offer an optimized solution from the perspective of treatment and utilization of resources. The system includes patient and hospital clients, data storage and analytic tools for decision support relying on AI-based services. We illustrate the functionality of the system through different situations frequently occurring in pediatric wards. To assess the feasibility of the AI component, we utilize real life health care data from a hospital in China to develop a classification model for patients with asthma. To provide a more qualitative assessment at the same time, we discuss how the Artificial Intelligence Chronic Management System conforms to the requirements set forth by the standard Chronic Care Model.

1. Introduction

The US Centers for Disease Control and Prevention (CDCP) defines chronic diseases as "illnesses that are prolonged in duration, do not resolve spontaneously and are rarely cured completely" (World Health Organization, 2005a). According to the CDCP report, chronic diseases are responsible for two thirds of deaths worldwide, and result in significant costs for the public health system. Furthermore, the burdens attributable to chronic diseases are higher in low-income and middle-income countries than in high-income countries (World Health Organization, 2005b). As predicted by Miranda, Kinra, Casas, Davey Smith, and Ebrahim (2008), by 2030, non-communicable diseases (of which most are chronic) would be accounted for 69% of all global deaths, 80% if which would occur in low and middle-income countries.

Chronic conditions in children have become a greater public health challenge in industrialized countries, despite the increasing knowledge about and resources investment in disease management. As pointed out by Van Cleave, Gortmaker, and Perrin (2010), from 1988 to 2006, the prevalence of pediatric chronic conditions, specifically obesity, asthma, other physical illness, and behavioral or learning disorder varied from 12.8% to 26.6% in the USA. As mentioned in the report of State Health and Family Planning Commission of the People's Republic of China (State Health and Family Planning Commission of the People's Republic of China, 2015), about 10% to 20% of children in China suffer from chronic diseases. Among these children, about 2% to 4%

* Correspondence to: Tuomiokirkontori 3, Turku, 20500, Finland.
** Correspondence to: No. 3333 Binsheng Road, Hangzhou, Zhejiang, 310052, China.
E-mail addresses: yugbme@zju.edu.cn (G. Yu), mohammad.tabatabaei@avaintec.com (M. Tabatabaei), jmezei@abo.fi (J. Mezei), qianhui.zhong@avaintec.com (Q. Zhong), siyu.chen@avaintec.com (S. Chen), 6513103@zju.edu.cn (Z. Li), lijing6@zju.edu.cn (J. Li), liqi_shu@brown.edu (L. Shu), shuqiang@zju.edu.cn (Q. Shu).

https://doi.org/10.1016/j.eswa.2022.117026
Received 16 October 2020; Received in revised form 20 December 2021; Accepted 27 March 2022
Available online 6 April 2022

suffer from severe chronic diseases, requiring continuous care and supervision.

The number of chronic diseases, illnesses, and deaths is increasing, and the burden of chronic diseases on the public health system is continuously growing. The complexity and comprehensiveness of the factors influencing chronic diseases determine the long-term and arduous task of prevention and treatment (Kong, 2017), typically requiring the development of effective chronic disease management systems. Functionality of a chronic disease management system depends on and is determined by the effective cooperation among relevant organizations. In China, the lack of such a cooperation and the consequences of health care reforms poses challenges in the existing systems.

The treatment and rehabilitation of children suffering from chronic diseases is a long-term process. After the completion of a treatment in hospitals, most children require regular visits and follow-up (Liu, Lu, & Wang, 2009). Through the analysis of follow-up data, doctors can timely and comprehensively understand the treatment and recovery of patients after discharge to improve the quality of the treatment. At present, however, medical and public healthcare resources are limited in China to finance the cost associated with personal visitation-based follow-up chronic disease management. At hospitals, doctors encounter an enormous number of face-to-face visits with patients every day. It is timely and financially challenging to provide detailed personalized consultation, diagnosis and other health services for different patients (State Health and Family Planning Commission of the People's Republic of China, 2018). To tackle this problem, the rapid development and penetration of computer and information technologies and artificial intelligence (AI) provide new opportunities for the innovation of effective automated chronic disease management systems.

In recent years, with the development of cloud computing, Internet of Things (IoT), big data, and computing algorithms, the potential of applicability of artificial intelligence in various fields, such as healthcare, has been improved. Already in 2013, Murdoch and Detsky (2013) highlighted the importance and necessity of big data in health diagnosis. Information technology platforms, regional medical and public health cooperation as well as personal electronic health records are in place to establish the core elements for AI-based services for chronic disease management systems. In this paper, we present and discuss a chronic disease management system as proof of concept. The system can provide intelligent, comprehensive, timely, active and continuous chronic disease management for chronically ill children and could result in significant savings in medical resources to improve the general viability of the public health care services. In the proposed system, many tasks of chronic disease management that typically require costly human resources, are automated and solved by various AI tools. As the basis of the system, electronic healthcare records of patients are analyzed to extract informative patterns that can assist doctors in effective diagnosis and treatment planning. As it has been illustrated in the literature numerous times, electronic health records data can be utilized in building high performance predictive models (Hoang & Ho, 2019; Zhang, Xu, Guo, & Gao, 2017). The system facilitates the process of follow-up for patients after discharging from hospital: when patients leave hospitals, their relevant data are monitored through wearable devices, and then analyzed continuously in an automated manner. In case of any deficiency or failure regarding the health status of patients, the system informs doctors and patients for immediate proper actions. A key component of the proposed model is the knowledge-base represented as a knowledge graph (Singhal, 2012). In the core of the system presented in the main part of the article, knowledge graphs created based on health-related databases are used and combined with AI-based tools. Similar approaches utilizing knowledge networks have been shown to offer excellent results in the health care domain (Jiang, Li, Zhao, Guan, & Yu, 2017) .

According to this discussion, in this research we aim to present a *preliminary design for an AI-assisted system that covers the complete lifecycle of childhood chronic care management*, from triage to home care. In order to address this general research objective, a set of research questions will be answered. By answering these research questions, we can shed light on how to deal with the most critical issues related to AI-assisted chronic care management systems. The research questions are as follows:

- RQ1: How can we represent and store the wide variety of data available from the monitoring and treatment of children with a chronic disease?
- RQ2: What are the most important use scenarios of chronic care management in which the introduction of AI-assisted tools can deliver significant improvement and savings of significant (human) resources?
- RQ3: How can we build machine learning models to assess in identifying children with chronic disease, with a special focus on one of the most frequent diseases, asthma?
- RQ4: How can we ensure that an AI-assisted chronic care management systems conforms to the requirement of established standards, such as the industry-standard Chronic Care Model?

As the system is currently in development, our aim here is to offer both quantitative and qualitative results illustrating the potential of the approach. The validation is structured into two steps:

- a quantitative assessment: an experiment using real data of asthma patients is presented to assess the feasibility of developing AI-based systems that can alert doctors when the status of the patient needs extra attention;
- a qualitative assessment: we evaluate how the implementation of the proposed system would conform to the basic requirements in the industry-standard Chronic Care Model.

Both the quantitative and qualitative analysis have been performed in cooperation with the Children's Hospital, Zhejiang University School of Medicine, specifically the pediatric department, as we focus on infant, children and adolescent asthma patients.

This paper is organized as follows. After the Introduction, we present a Literature Review on developing chronic disease management systems and the use of AI and knowledge graphs in related decision support systems. Afterwards, we propose the AI Chronic Management System (AICMS) and describe the components and their interaction in more details. We show the effectiveness of the system through several usage scenarios of varied complexity. After the example use cases, we present the results of analyzing real-life health care data and information collected from health care personnel to empirically illustrate the potential of our proposal. Finally, we present some conclusions and future research directions.

2. Literature review

In this section, we discuss literature relevant to our research: designing a new Chronic Disease Management (CDM) system aimed particularly for children. We will discuss (i) chronic care management theories and practices, (ii) the role of AI in chronic disease management and (iii) knowledge graphs in health care problems.

2.1. Chronic disease management

A typical lens through which academicians and practitioners look at chronic disease management is the Chronic Care Model (CCM, proposed by Wagner et al. (2001)). CCM emphasizes the power of joint intervention of patients, medical workers and medical policies. The main components of the model include: (i) the design of health service deliver; (ii) support for patients to improve self management; (iii) decision support involving patients and different hierarchical levels in the medical system; (iv) data management and clinical information systems; (v) community resources and policy support; (vi) health care system design. Among many different CDMs building on the main idea of CCM, one can mention Innovative Care for Chronic Conditions (ICCC) proposed by the World Health Organization (Pruitt & Epping-Jordan, 2002). ICCC expand and extend CCM by including three levels: the micro level (patients and family members), intermediate level (health care institutions and communities) and macro level (policy and financial resource mobilization). Several of the eight essential elements specified in ICCC could better be addressed by information technology and specifically AI-based tools relying on monitoring and interaction with (child) patients and guardians (Zwar et al., 2017):

- paradigm shift to move away from traditional acute episodic care (Weeks, George, Maclure, & Stewart, 2016)
- information sharing across all levels of the environment (Barr, Vania, Randall, & Mulvale, 2017)
- effective use of health care personnel (Milani & Lavie, 2015)

In the literature review, and the later CDM system proposal, we focus on these highly relevant components to make use of AI in managing chronic diseases.

2.2. AI in chronic disease management

In the literature, researchers have developed systems based on AI to manage chronic diseases (Liu, Hu, Chen, Yu, & Liu, 2016). Incorporating data mining and case-based reasoning techniques for prognosis and diagnosis of chronic diseases have been studied by Huang, Chen, and Lee (2007). An expert system is proposed by Abuel-Reesh and Abu-Naser (2017) to aid doctors in diagnosing and describing some common causes of shortness of breath in infants and children. Liang et al. (2019) discussed a proof of concept for implementing an AI-based system to aid physicians in tackling large amounts of data, augmenting diagnostic evaluations, and to provide clinical decision support in cases of diagnostic uncertainty or complexity. A general-purpose AI framework is discussed by Bennett and Hauser (2013) to predict optimal treatments in which minimizing side effects, medical errors and costs are considered.

Several reviews of the application of AI in developing and developed countries are available in the literature (Guo & Li, 2018; Pérez-Ardanaz et al., 2019; Samb et al., 2010; Wahl, Cossy-Gantner, Germann, & Schwalbe, 2018). As discussed by Guo and Li (2018), there are some issues in this area, such as: (1) infrastructures, (2) training, (3) professionalism, (4) relationship between patients and service providers. In Wahl et al. (2018), ideas are suggested on how these AI developments in developed countries can be integrated into developing countries.

Utilizing wearable devices is considered as an element in chronic diseases monitoring. In Jalaliniya and Pederson (2012), a wearable device is developed to monitor the health of children. In Sendra, Parra, Lloret, and Tomás (2018), an AI system is proposed to monitor children chronic illness remotely by using wearable devices and sensors embedded in smartphones. In this system, as soon as a pre-defined parameter exceeds a threshold, an intelligent decision is made about informing patients with an emergency alarm.

Reviews on how AI has been applied in pediatrics are presented in several articles (Ng et al., 2018; Shu, Sun, Tan, Shu, & Chang, 2019). In Reynolds et al. (2018), chronic disease management systems in adults are reviewed. Moraitou, Pateli, and Fotiou (2017) present a review on smart home application for caring elders and their chronic diseases. Based on the findings from the listed articles, we can also state that interaction between decision support systems and patients can assist in gaining the patients' trust in systems. For instance, patients can ask questions regarding their symptoms and receive relevant answers from the system. In emergency situations, patients are required to visit a hospital, and in these situations providing a smart decision based on the location of patients and capacities of hospitals can help patients to receive necessary care in a decent time interval. Diet recommendations can also contribute to well-being of patients with chronic diseases. These are some of the aspects which have not been considered in the systems reviewed above.

One of the most significant trends in recent years in the application of machine learning is the wide utilization of various deep learning models (Faust, Hagiwara, Hong, Lih, & Acharya, 2018). One of the most important application areas of deep learning concerns unstructured data sources, e.g., textual data and audio or video files. As it will be described in our proposed model, these data sources, additionally to electronic health records, would constitute the core of an automated, AI-based chronic disease management system.

An important example of chronic diseases is asthma, and as such it has been the topic of various studies in the literature. As in this article asthma is chosen for validation purposes, we discuss some of the relevant studies briefly here to offer a comparison to our approach. Asthma is a chronic (long-term) lung disease that inflames and narrows the airways (Toren, Brisman, & Järvholm, 1993). The typical, easily recognizable symptoms include recurring periods of wheezing, chest tightness, shortness of breath, and coughing. The most widely researched problem is the early prediction of asthma, with various studies considering this problem with different methods relying on a wide variety of data types. In Prasadl, Prasad, and Sagar (2011), the authors identify the two main data sources that are typically used for diagnosing and predicting asthma: (i) questionnaires, and (ii) clinical data. In the article, they consider developing expert systems for diagnosing asthma based on the mix of data sources, and compare various machine learning models, including context sensitive auto-associative memory neural network, backpropagation model, Bayesian networks and Particle Swarm Optimization. The results show that it is possible to reach approx. 80% accuracy in the prediction.

In Emanet, Öz, Bayram, and Delen (2014), the sound recorded from the chest of the patient is used to build various classification models, and it is found that ensemble models can significantly over-perform the individually best performing artificial neural network model. Recent contributions make use of the rapid development of deep learning models. For example, in Do, Son, and Chaudri (2017), TensorFlow and multilevel databases are utilized with the relevant variables including age, admission date, hospital state, number of procedures, number of chronic conditions, gender, length of stay. Based on preliminary experiments on a few hundreds of patients, the authors found that reasonable accuracy can be achieved across all severity levels of asthma. Regarding the continuous monitoring of asthma patients and predictions based on the collected data, the main focus of the approach presented later in this article, in Kocsis et al. (2017) it is stated that self-management tools have the potential to support personalized patient guidance. The authors develop myAirCoach, a personalized asthma monitoring system to help patients manage their asthma condition and increase their awareness of their clinical state. The sensors monitor several physiological, behavioral and environmental factors, which are further processed and aggregated to provide tools for clinical decision support. To test and evaluate the approach, the authors make use of various machine learning models, including random forests, support vector machines and AdaBoost. They found that 80% accuracy can be achieved in two main classification tasks: (i) short-term prediction of asthma control level for daily and real-time personalized patient guidance; and (ii) long-term prediction of exacerbation risks, to support clinical decision making and alert medical personnel.

2.3. Knowledge graphs in health care

In recent decades, many important tools, models, and algorithms has gained popularity in a large number of scientific domains and industrial applications from biology to economics; this trend can be seen as the diffusion of business analytics (Holsapple, Lee-Post, & Pakath, 2014). Most of the resulting innovations are based on developments of machine learning algorithms, the availability of large datasets and rapid increase in computing power. One of the important recent developments, combining text analytics, network analytics and predictive analytics is the concept of a Knowledge Graph (Singhal, 2012).

The knowledge graph consists of nodes (entities) and labeled edges (the relationships between entities), creating a semantic network, closely resembling the idea of an ontology. To build a knowledge graph (Pujara, Miao, Getoor, & Cohen, 2013), typically a variety of strategies is used to generate candidate facts (mainly referring to relationships, but also relevant entities and their attributes) from the documents. This process relies on extracting information focusing on both syntactic, lexical, and structural features.

Since its introduction, the knowledge graph has been tested and utilized in various academic experiments and real-life applications. First, regarding the construction of knowledge graphs, a good example is presented by Rotmensch, Halpern, Tlimat, Horng, and Sontag (2017). In the article, a knowledge graph is developed by extracting relationships between diseases and symptoms from electronic medical records. The approach utilizes basic NLP methods to extract positive mentions of diseases and symptoms from structured and unstructured data. In order to create the final knowledge graph, the authors experiment with various machine learning techniques, such as logistic regression and naïve Bayes. Second, a typical utilization of knowledge graphs is to create new knowledge. Nickel, Murphy, Tresp, and Gabrilovich (2015) provide an overview of the statistical methods that are developed to analyze data in the form of a knowledge graph focusing on deriving new knowledge based on existing links in the graphs, and do so in a scalable manner. The authors identify two main approaches for this task: (i) latent feature models (tensor factorization, multilayer perceptron, latent distance models), and (ii) graph feature models (rule-mining, path ranking). Malas et al. (2019) present a more specific example focusing on drug prioritization and repurposing. The authors extract drug and disease concepts from a knowledge graph to be used as features in a classification model. The model is trained using data available on successful and failed proposed drug-disease combinations. Thirdly, an important component of creating and improving knowledge graphs is the utilization of expert opinions and domain knowledge. Sharma et al. (2019) consider the problem of natural language inference aided by knowledge graphs. The authors propose to fuse embeddings obtained from knowledge graph relying on contextual word embeddings with the most important aspect of the approach being to incorporate expert knowledge into the process.

Lei et al. (2020) describe a study that is closest to the model we present in the later sections of this work. The authors propose a new model to identify diseases based on some features related to where the disease is occurring in a certain region. The utilized data includes structured and unstructured sources, combining medical records, family information, textual and video information from different sources. The approach combines knowledge graphs and classification models, and provides a prediction together with an explanation/interpretation of the results. In this article, instead of focusing on an individual model, we describe the design of a chronic disease management system that can incorporate several different features, such as the model in Lei et al. (2020). As it will be described in the following section, the system combines artificial intelligence, internet of things and knowledge graphs in a comprehensive health care solution connecting patients with doctors and increasing the efficiency of the system.

3. The AI chronic management system (AICMS)

In this section, we present the structure of the AI Chronic Management System (AICMS) as shown in Fig. 1. Our premise for developing this model, additionally to the motivational discussion presented above, is that, as identified in Do et al. (2017), the main objectives of asthma care are monitoring symptoms and progression of the disease while avoiding asthma triggers and minimize asthma attack. Our proposed system consists of three components, i.e., central unit, patient, and hospital clients. We first describe an overall picture of each component and how they communicate together. We then provide a detailed discussion of each component.

In AICMS, the patient and hospital clients collect relevant data from children and doctors, respectively. The data is then transferred to the central unit. The central unit analyzes the data and provides reports upon which some decisions are made. We first describe the patient and hospital clients, then continue with describing the functionality of the central unit.

At the core of the proposed system is a set of connected databases: (i) a traditional one, including data from electronic health records, and (ii) a knowledge graph-based one, established and inferred from connections between concepts relevant to children chronic diseases. The knowledge graph database platform integrates structured and unstructured data from the electronic health records, community and general demographic information and data collected from the patients in the client of AICMS. In the following subsections, at each point we discuss what is the relevant data that is collected, and in the next section, an example of creating a knowledge graph database from electronic health records and patient data is provided. An important component of a system, such as the one proposed, is the continuous update of the databases when faced with datastreams of large quantity and varying quality data from the patients and the healthcare system. We do not discuss this specific feature in this proposal in detail, but we note that methods available for maintaining and updating dynamic knowledge graphs are present in the literature (Tay, Luu, & Hui, 2017; Trivedi, Dai, Wang, & Song, 2017) that address constructing knowledge graphs from an ever-growing stream of evidence, that are fully applicable to AICMS.

3.1. Patient and hospital clients of AICMS

In this section, we first discuss the structure of the patient client followed by the hospital client. As can be seen in Fig. 1, the patient client includes two modules, i.e., monitoring and intelligent communication modules. The monitoring module is responsible for collecting data from patients by using wearable devices. The type of data is defined by nurses and/or doctors. Examples of such data are moving coordinates, heart rate, or body temperature. The intelligent communication module acts as an interface between patients and the central unit. Once the data is collected, it is transferred into the central unit by the intelligent communication module.

By using the intelligent communication module, patients can also ask questions. The questions are analyzed by the central unit, and proper answers are provided. We discuss the process of analyzing and answering questions in the AI module descriptions in the following section.

The hospital client includes the intelligent communication module, community doctor module and specialist modules. This client facilitates the communication among the central unit, community doctors and specialist doctors at hospitals. The community doctors or specialist doctors may require receiving information regularly regarding the health status of some patients to promptly understand their health status. In such cases, the community doctor and specialist modules provide reports of the information demanded by doctors.

3.2. Central unit of AICMS

As mentioned earlier, the core part of AICMS is the central unit. This unit consists of the following modules: intelligent communication, analysis, and risk assessment. The intelligent communication module collects relevant data from different sources and passes it into the analysis unit. The analysis unit analyzes the data and generates reports. The reports are then evaluated by the risk assessment module for decision making. In what follows, we describe how each module functions in the central unit.

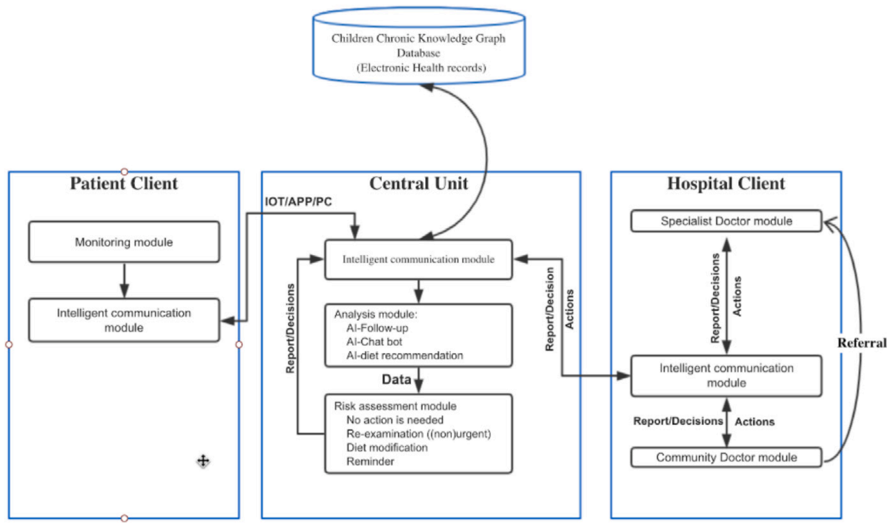

Fig. 1. **AICMS** The structure of the AI Chronic Disease Management System.

The intelligent communication module is responsible for transferring data and information among the database, central unit, patient and hospital clients. In some cases, when an appointment is required for patients, this module provides some appointments to patients to be chosen from.

The analysis module consists of three AI models, i.e., AI-follow-up, AI-chat bot and AI-diet recommendation. The input and output of this module are the data passed from the intelligent communication module and the reports produced by the AI-models, respectively.

AI-follow-up analyzes the health data collected from patients constantly by using data stream mining algorithms. After the analysis, this AI produces reports describing the health status of patients. The AI-diet recommendation model analyzes the diet of patients and generates reports, including some possible new adjustments in the diet.

Patients or their guardians can interactively ask questions through the intelligent communication modules in the patient client and the central unit. Questions are analyzed by the AI-chat bot model and proper answers are suggested to them. Once the data passed into the analysis module is analyzed by the AI models, reports, including relevant information, are produced and passed into the risk assessment module. Risk assessment module is responsible for evaluating the reports and provides smart decisions based on predefined rules specified by doctors and/or nurses who are expert in the domain. Decisions are categorized as:

- Decision 1. no action is required
- Decision 2. a reminder should be sent
- Decision 3. diet adjustment is required
- Decision 4. a non-emergency re-examination is required
- Decision 5. an emergency re-examination is required

Based upon one or some of the above-mentioned decisions, the risk assessment module sends information to the intelligent communication module. The intelligent communication module then acts according to the information received. Decision 1 requires no action. In Decision 2, a reminder regarding a near future action, such as laboratory test or doctor's appointments, is sent. Based on Decision 3, a notification including the diet recommendation is suggested to patients to improve their health status. As far as Decision 4 is concerned, the intelligent communication module extracts relevant information from the database to find some suitable appointment date and time. The appointments and the reason of the non-emergency re-examination are sent to patients. Once a suitable appointment is chosen by patients, a confirmation notification is sent to the patients and doctors.

Regarding Decision 5, the reports are sent to the community doctors. This is done by passing the reports from the central unit to the intelligent communication module embedded in the hospital client. The community doctor module receives the reports and shows them to the community doctors. They then can evaluate the reports to conduct proper actions. Examples of actions can be asking patients to visit immediately the nearest hospital, to visit a community doctor or a specialist doctor. In case of asking patients to visit a specialist doctor, the community doctor provides a referral letter which is sent to the specialist doctors from the community doctor module to the specialist doctor module. The intelligent communication module identifies an emergency appointment and informs the patients to visit a hospital when necessary.

4. Selected components and possible use scenarios of the AICMS

In this section, we present a more detailed description of some features of AICMS that offer functionalities not typically found in traditional chronic disease management systems. We included the description of two example features: (i) the process of creating a health knowledge graph, and (ii) health follow up and smart return visit. Furthermore, we discuss two typical complex situations that constitute crucial parts of the system. First, we describe the process of conducting a Q&A session, in the situation when the guardian of a child with a specific disease performs a (potentially multi-stage) conversation with the systems. Second, we describe how the follow-up data collection and advanced AI algorithms in AICMS contribute to an early warning system for children chronic diseases.

4.1. Health knowledge graph

To illustrate how a Knowledge Graph can be created in the healthcare domain, we present a simple illustration. In the example, the graph construction is based on information extracted from Electronic Health Records.

The basic unit of information extraction is an 'admission': a patient is admitted to a hospital, and the hospital visit can range in duration depending on the criticality of the associated diagnoses. An admission has various associated attributes, such as medications or notes from the medical practitioner presenting pre-diagnosis. Admission data can be combined with information regarding various diseases and related symptoms. For example, migraine is associated with medical symptoms such as headache, nausea or sensitivity to light. To create a knowledge

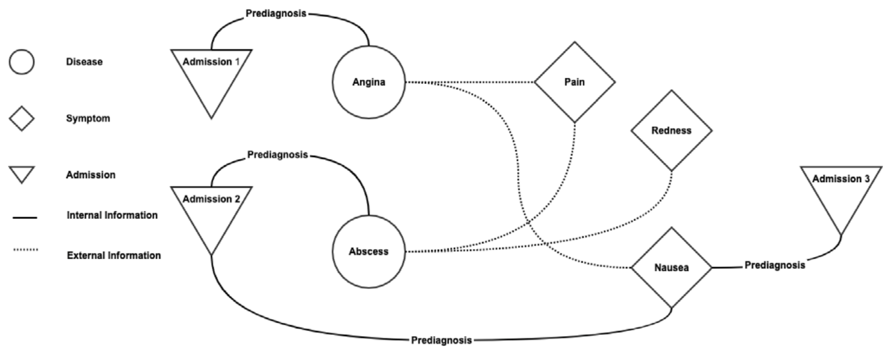

Fig. 2. Health Knowledge Graph A Knowledge Graph, Simplified.
Source: Adapted from Patra (2019).

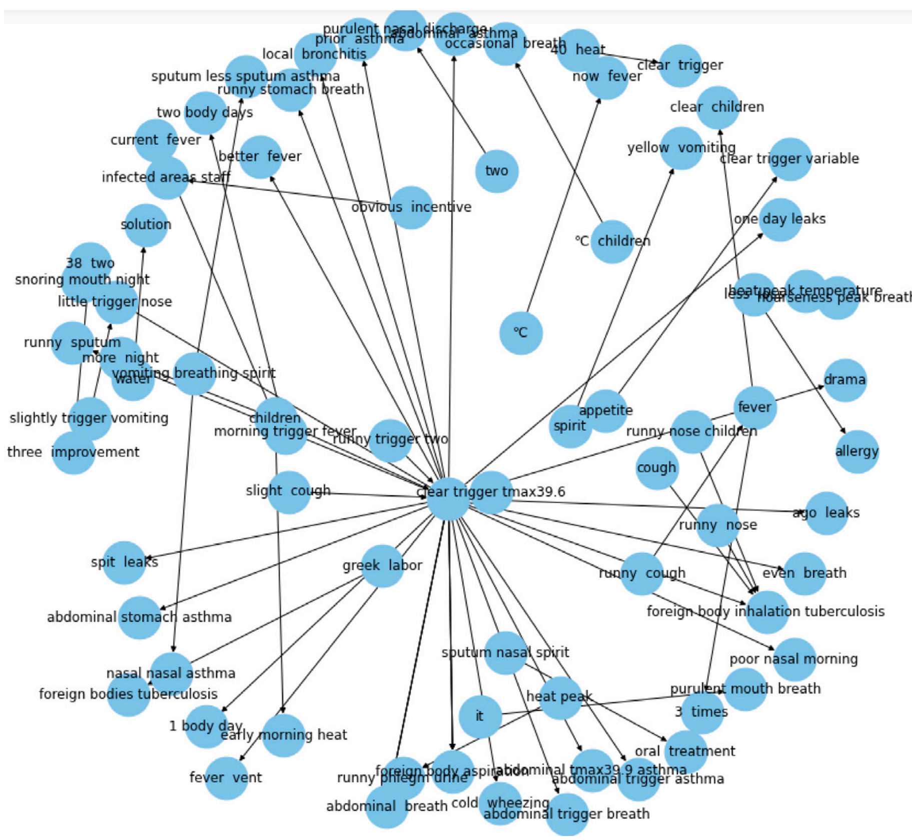

Fig. 3. Sample knowledge graph A Knowledge Graph depicting relations 'improved' and 'fever'.

graph based on these data sources, the mentioned pre-diagnosis can be connected to disease information using extracted keywords.

By completing this process, a knowledge graph, such as the one presented in Fig. 2, can be constructed. The entities of the graph include "Disease", "Symptom" and "Admission". In this knowledge graph, the relationships are either "Admission has similar pre-diagnosis as either a disease or a symptom" or "Disease has the following symptoms". To compute the similarity, we can employ a combination of n-grams with TF-IDF (Term Frequency – Inverse Document Frequency) to link Admissions with diseases or symptoms. In the figure, 'Internal information' refers to data collected from the medical records whereas 'External information' is the information aggregated from a Disease-Symptom knowledge graph (Patra, 2019).

In addition to this simple example, we have performed an initial Knowledge Graph creation process on the dataset we had access to from the Children's Hospital, Zhejiang University School of Medicine. As it will be described in more detail in a later section, the data set has information about 1276 children who are suspected to have asthma. There are several text components in the data, and in creating the knowledge graph example, we focused on the description of recent medical history of the patients. The original Chinese text was translated into English, and we made use of the Python programming language, specifically the spacy and networkx libraries, to construct the knowledge graph.

After basic text preprocessing, we extracted subject and object pairs from all the sentences in the dataset. Typical examples included ['oral antipyretics','sore throat'] or ['cough', 'appetite']. After the entity pairs are extracted, we determined the relation between entities using the sentences available. The relations extracted include, for example 'denied', 'appear' or 'improved'. In the analysis, the algorithm extracted 350 different relations. As it is difficult to fully comprehend a knowledge graph even on a moderately small scale like this, we just selected two of the most frequently occurring relations, 'improved' and 'fever', and created a sub-graph of the full knowledge graph as depicted in Fig. 3. While there are still obviously data-specific relations (noise) in the graph that would disappear as more and more data is collected, a detailed analysis of the whole knowledge graph already reveals important connections established between different symptoms (in the presented graph different types of fever, runny nose, etc., appear) and some corresponding diseases (such as asthma and tuberculosis) as instance of the 'improved'; by extracting the treatments that correspond to the instances, the doctors can obtain useful information for specific new patients.

4.2. Health follow-up and smart return visit

The presented system structure allows for a constant AI follow-up that can achieve uninterrupted and full coverage mitigating the problems resulting from the excessive cost of manual follow-up and low work efficiency. In addition, the AI follow-up collects and processes patient follow-up information and performs statistical analysis to provide effective data support for clinical and scientific work, and finally integrates the data in the knowledge graph databases. At the same time, AI follow-up can also be used to assess patient satisfaction for continuous improvement of the quality of care for chronic disease care. AI-follow-up is mainly used in smart return visits, intelligent SMS push, intelligent medication tips, intelligent early warning and other daily occurring events in the life of children with chronic disease.

As an important example, smart return visit is a human–computer interaction module of the system that aggregates return visit path, service discourse and knowledge base of chronic diseases in children. Several AI technologies, such as speech recognition, semantic understanding and speech synthesis, are combined to design follow-up questions for AI voice communication with child patient's guardians, automatically collect feedback information, complete data acquisition of return visits, generate data reports and other related operations.

The smart return visit module defines a variety of chronic disease return visit paths, which are jointly managed by the smart return visit engine. The return visit engine integrates the question-and-answer library, the knowledge graph, and the return visit path library, and constructs the context in which the return visit path is dominant and the user answers the return visit. Return engine first combines the types of chronic diseases of users, uses the template of return questions stored in the corresponding return path library to initiate a request for targeted return visits to users, then sets up one-to-one return calls with users through external phone calls, and clarifies the content and intention of communication. After the user connects to the return visit call, the user answers the question raised by the return visit engine. The voice recognition model in the system identifies the use's answer and outputs it to the engine in the form of natural language text, then the return visit engine receives the user reply and combines the context. User intent recognition ensures that the user does not answer content other than the return visit. After the user completes the return visit, the return visit engine collects and archives the information replied by the user and investigates and archives the satisfaction of the user.

4.3. Healthcare Q&A

Fig. 4 depicts the processes taking place in AICMS when the guardian of children consults the client and inquires information on, e.g., asthma medications, complications, or diet recommendations. The base setting of the system in case of disease-related information is to follow a single turn conversation mode: after receiving the user's questions, the AI chat-bot processes and analyzes the question and returns the answer to the user. The AI diet recommendation module adopts a multiple turn conversation mode: after receiving the guardian's inquiry and offering recommendation, it can also connect to a food ordering API to complete the task according to the user's ordering needs.

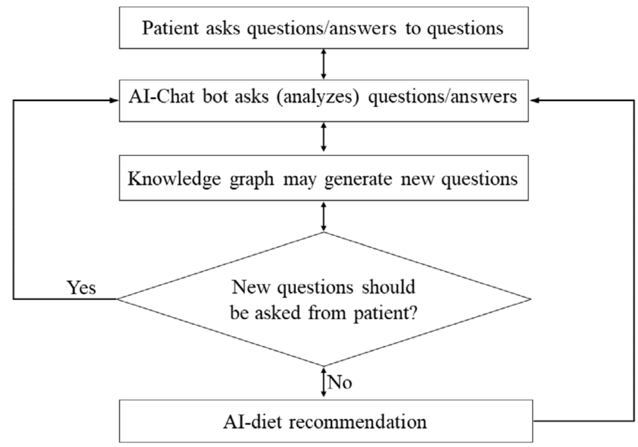

Fig. 4. Healthcare Q&A Process The process of a Q&A session in AICMS.

In the single turn conversation mode, the AI chat bot analyzes the questions received by the system, and extracts the related knowledge stored in the knowledge graph. In processing the user's question, the system performs word segmentation, e.g., utilizing word2vector approach. In the next step, the important entities of the knowledge graph from the user's question are identified. For example, in the case of a child with asthma, the system classifies 'asthma' as a disease entity. An intention recognition model classifies user intentions and combines this with extracted entities to extract information from the knowledge graph. For this extraction, different AI techniques can be utilized as discussed in the overview of knowledge graphs presented in the previous section: one possibility is to use the concept of recommendation spreading. For example, an identified example of user intention can be "asking for complications". Based on this intention, combined with the identified "asthma" entity, the AI engine retrieves related knowledge by also considering the data stored in the system about the specific child. The retrieved answers are converted into natural language and output to the user.

The more advanced option is to perform multiple turn conversations to deal with specific tasks based on users' needs, such as ordering food. For example, when the user inquires about the AI chat-bot and AI diet recommendation engine for patients with asthma, the system will not only communicate the recommended food to the user, but also actively ask whether to order the relevant food. In case of a positive answer from the user, the system becomes activated. The AI chat-bot enters a multiple conversation mode and requires filling in some information regarding the order by the user. In case some of the information slots are filled in a way that is not related to the actual question, the system asks the user to update the answers. When all the slots are completed, the system repeats the collected information to the user for confirmation, and call the API to perform the corresponding tasks.

4.4. Chronic disease warning

Probably the most important functionality of our proposed AICMS is the possibility to systematically perform follow-up management in an automated manner: collecting data, monitoring the patient status with AI and taking actions when deemed necessary. When traditional management systems are used for chronic diseases, follow-up is difficult as the data on the patient's physical condition cannot be traced continuously.

A cornerstone of this module of AICMS is the availability of IoT that allows for tracking the patient not only in the hospital but also at home (or on the move when smart devices with sensors are in use). The devices can be used to collect multiple physical indicators of the patient and store the physical indicators in the cloud. The information that can be automatically collected includes blood pressure data, sleep data, exercise data, heart rate data, and weight data. Additionally to this automatically collected information, guardians of patients can manually enter data about the child, such as food consumption or the medications taken during the day. The collected data is stored in the cloud and after preprocessing, incorporated into the Knowledge Graph and health records of the patient, and can be used at once in all the analysis tasks.

The AI module of AICMS utilizes deep learning for classifying early warning signs related to chronic diseases. As the input for the model, historical data and real time physical measurements are utilized in a text format. Different models appropriate for NLP, such as LSTM, BERT, etc. are then utilized to analyze all data and determine the final output of the model: the obtained likelihood of a problem related to chronic diseases is compared to a predefined threshold value. For emergencies and non-emergencies, follow-up treatment can be carried out by classification, such as reminding doctors and patients' guardians.

Through the early warning service of chronic diseases, patients can be diagnosed actively, remotely and automatically to avoid utilizing the large amount of human resources needed for routine testing. When the model learns from a large number of high-quality medical data, it can optimize predictions for early warning, which in turn can liberate medical resources and improve the quality of early warning services for patients with chronic diseases.

4.5. Additional use cases of interest

Additionally to the examples presented above, there are numerous additional use cases that could be considered in order to improve the proposed structure and utilization of AICMS. As the main motivation in utilizing AI, similarly to other domains, is to offer personalized services to the patients in an automated manner, several use cases may focus on replacing the human component from the interaction of the patient and health care provider as much as possible without imposing extra risks. These extra functionalities can range from purely treatment focused, such as reordering medications when the patient seems to run out of those, to more well-being oriented improvements, such as health activity recommendations personalized to the current status of the patient.

From another perspective, the collected data can be utilized by health care professionals and researchers in deepening the understanding of various chronic diseases. The central unit of AICMS, and the created knowledge graph, when trained on increasing number of cases, will offer doctors the possibility for, e.g., identifying different sub-types of diseases that have not been clearly differentiated in the past in lack of appropriate database and computational models. Another important outcome will be to understand different stages of diseases, e.g., create a more refined understanding of the complete 'patient journey', allowing for increased level of personalization in the existing components as discussed above. These higher level use cases would significantly help in advancing the treatment and chronic management in a more rapid pace that has been possible before.

Finally, a topic that was marginally touched upon, i.e., the resource management of hospitals, is something that can be greatly improved and monitored using existing and possible future functionalities of AICMS. By understanding the utilization of financial, human and other resources and the impact AI-based methods can bring, hospital management and financing can be greatly improved, and this can be crucial in situations or countries when existing resources are not sufficient at the current stage. To further improve the administrative and resource management component, in the future data bases can be connected to other stakeholders in the health care ecosystem, such as insurance companies, which would further improve processing time and accuracy.

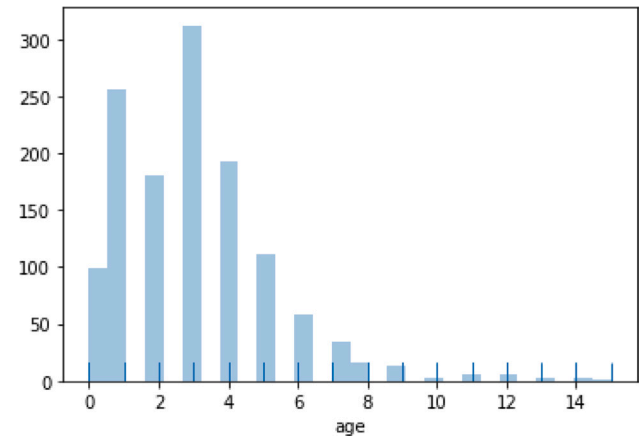

Fig. 5. Age distribution of patients.

5. Validating the proposed approach

In this section, we present some preliminary results on the feasibility of the proposed AICMS. We present the results of quantitative (building a machine learning model to predict patient status) and qualitative (assessing the Chronic Care Model with respect to AICMS) assessment. The validation process has been performed in cooperation with the pediatric department of the Children's Hospital, Zhejiang University School of Medicine, as we focus on infant, children and adolescent asthma patients.

5.1. Building a classification model for asthma patients

In this section, we present the steps of building a classification model for asthma patients. The data is collected in the pediatric department of the hospital on patients aged under 18 years old. The original dataset included data on 86,052 patients, with the total amount of visits being 128,976, i.e., 1.5 visits per patient in average. After discussion with the doctors of the hospital and evaluating the importance of different chronic diseases, we made the decision to focus on patients with asthma as the relevant example of chronic diseases. Formally, we focused on patients who have disease with code J45 according to the classification International Classification of Diseases (ICD-10). After filtering the patients, we identified 1276 datapoints of interest.

Regarding the basic demographic of the patients, the gender distribution is 41%–59%, and the age distribution is depicted in Fig. 5. As we can observe, most of the patients of interest are under 5 years old. Among the visits, 217 are associated with asthma, which is approx. 17% of all the datapoints, implying that we need to work with a moderately imbalanced dataset, which is taken into consideration in model building and evaluation. The main goal of the experiment in this stage was to assess whether we can build a machine learning model of sufficiently good performance to automatically predict whether the patient currently experiences an abnormal health issue related to the chronic disease condition of asthma. This problem is a classical binary classification task (Friedman, Hastie, & Tibshirani, 2001). In the following, we briefly present the data pre-processing and model building. In all the tasks, Python programming language was utilized with different libraries built for machine learning and natural language processing.

The data used in the analysis, additionally to gender and age, includes unstructured information, i.e., textual data. While the text data is in Chinese in the original form, in the analysis we worked with the English translation. In the analysis, we focused on features that are either available already in the electronic health records of the patient, or most likely can be inputted into a monitoring system by experienced parents/guardians taking care of the children with asthma. The three main set of variables are the following:

- complaint: the main reason why the patient visited the hospital. Examples include “more than 10 days of cough, fever five days”, or “cough, fever, two days with dizziness”.
- previous history: description of previous information available about the patient, already stored in the database that can be used by a machine learning algorithm.
- current history: detailed description of conditions and symptoms observed regarding the patient. An example can be “’fever, tmax38.9, 10 days with cough, phlegm, night does not play, reduced appetite, no vomiting, no diarrhea”. All information of this level could be observed and provided by a guardian in a remote system for analysis.

The free-text information was pre-processed in different ways to allow for more detailed experiments. We utilized the following widely known approaches frequently used in the literature (Seol et al., 2020):

- Term frequency–inverse document frequency (TFIDF) (Wu, Luk, Wong, & Kwok, 2008): this approach makes use of term frequency (TF), i.e., the frequency of a word in a document (in our case medical condition descriptions), while Inverse document frequency (IDF) of a word is the measure to help in accounting for the significance of a term in a whole text corpus. Using this approach, we will not simply consider words that occur frequently as relevant (for example, as most of the patients have symptoms related to coughing, this occurs frequently, but not useful in differentiating asthma cases).
- Word2vec (Mikolov, Chen, Corrado, & Dean, 2013): models belonging to this class are used to create so-called word embeddings. Typically, neural networks are constructed to model the linguistic contexts of the words. In plain words, in the produced numerical vector representation, the words that appear in similar contexts will have associated vectors that are located close to each other.

After the text data was pre-processed using the different approaches, we created training and test sets from the original data, with the test set being 20% of the dataset. Additionally for keeping the class distribution the same in both training and test sets, we considered the fact that several patients with multiple visits can be found in the data: in order to avoid any bias, all observations to a given patient were assigned to either the training or the test set. In order to evaluate the created models, we used the performance measure AUC (Area under the Receiver operating characteristic (ROC) curve) as it is common in problems with imbalanced outcome class distribution (Huang & Ling, 2005). The results of the experiments are presented in Table 1.

In order to test different variable combinations, we set up 12 different combinations of data sources to be included. Information about current and previous history of the patient is included in all of the models (but pre-processed in different ways as indicated by ‘yes’ in the Text Features columns), and we built models for each possible combinations of including chief complain, gender and age. We note here that the focus in this article is to provide a conceptual description of the component of a chronic disease management system, consequently our main goal is to show that we can achieve results that are acceptable from the perspective of the practitioners and are on par with similar approaches from the literature. Nevertheless, we naturally follow best practices in machine learning and constructed models making use of the most widely used approaches as identified from the literature. While we experimented with several other methods (such as neural networks), here we present the detailed results for two models: logistic regression and random forest. Logistic regression is presented as it is the model that achieved the highest performance, while random forest is the method that has achieved the best result in almost all the cases in the literature, offering a worthy comparison. We made use of advanced hyperparameter tuning using the combination of random and grid search: penalty and regularization parameters in logistic regression, the number of trees, maximum number of features and maximum number of levels in random forests. Performing and evaluating different machine learning models with optimized parameters also increases the overall reliability of the modeling process. Regarding the execution time of tuning and building the machine learning models, we have not encountered any combination of model type and data set (as presented in Table 1) that would have required more than a few seconds.

When looking at the results, one can make some important observations:

- performance: the best results we can achieve is 0.761 for the test set using logistic regression, which is approximately on par with similar studies as it will be discussed in the following subsection
- demographic characteristics and chief complaints: in general, including these variables improves the performance of the models, but only very slightly
- text features: utilizing TFIDf results in better performance for every possible combination of independent variables. We can also observe slightly more significant over-fitting using this approach: the AUC for the training set is typically more than 20% more than for the test set.

5.2. Evaluating classification performance in relation to previous research

To assess and put in context the quality of performance we achieved, it is important to summarize the results previous studies achieved in predicting asthma. In Goto, Camargo Jr., Faridi, Yun, and Hasegawa (2018), logistic regression, random forest, gradient boosted decision trees and deep neural networks are compared to predict asthma in both critical care and in hospitalization. The authors found random forests and gradient boosted decision trees to have the best performance, with AUC of 0.8 and 0.83 in critical care and hospitalization, respectively. The authors utilized demographic information, vital signs (temperature, pulse rate, systolic and diastolic blood pressure, respiratory rate, and oxygen saturation), and common chief complaints (e.g., dyspnea, cough, chest pain); these predictors are included in our data as part of the diagnosis text. Spathis and Vlamos (2019) utilized demographic profile, medical and special lung measurements, habits and associated symptoms to identify patients with asthma. While they did not present the value of AUC, they found random forest to be the superior model (over logistic regression and neural networks among other methods), with F-score of 0.81, with a very small and balanced dataset.

Kaplan et al. (2021), considering chronic pulmonary diseases in general, observe the potential for utilizing AI for identifying patients with asthma among other respiratory diseases, citing the study of Topalovic et al. (2019), in which the authors construct an AI-based (unspecified) machine learning model that achieves accuracy of 74%. However, as it is also found by (Topalovic et al., 2018), this model largely outperforms the expert evaluation of pulmonologists, with 53.5 (+-6) % mean accuracy.

Specifically considering pediatrics, a vital component of the presented process is identifying patients in home care who would require (immediate) readmission to the hospital. In the context of pediatrics, this has been addressed with machine learning approaches for example by Wolff, Graña, Ríos, and Yarza (2019). They focused on mainly respiratory diseases to predict the need for hospital readmission to the pediatrics department, and found Naivë Bayes to outperform Support Vector Machines by achieving 0.66 AUC.

To compare our results to the discussed studies, we note here that the presented applications from the identified articles in all the cases differ slightly either on the specifications of the children patient population considered or the disease types analyzed. Moreover, we could not find any single classification performance measure that was used in all the papers (except for accuracy, which is not a correct choice in case of imbalanced data that we have). As the main observation, we can see that the results in the literature are typically around 0.8 (being

Table 1
Data selection and results for different models (LR: logistic regression, RF: random forest).

Conditions					Text features			AUC value			
Text features			Structured								
Chief complaint	*Current history*	*Previous history*	*Age*	*Gender*	*TFIDF*	*Word2vec (word level)*	*Word2vec (character level)*	*Training set (LR)*	*Training set (RF)*	*Test set (LR)*	*Test set (RF)*
Yes	Yes	Yes			Yes			0.939	0.947	0.758	0.758
Yes	Yes	Yes	Yes	Yes	Yes			0.939	0.947	0.761	0.739
Yes	Yes	Yes				Yes		0.821	0.964	0.671	0.684
Yes	Yes	Yes	Yes	Yes		Yes		0.824	0.99	0.687	0.694
Yes	Yes	Yes					Yes	0.843	0.988	0.750	0.760
Yes	Yes	Yes	Yes	Yes			Yes	0.843	0.992	0.752	0.754
	Yes	Yes			Yes			0.945	0.950	0.735	0.702
	Yes	Yes	Yes	Yes	Yes			0.945	0.825	0.736	0.696
	Yes	Yes				Yes		0.820	0.985	0.634	0.638
	Yes	Yes	Yes	Yes		Yes		0.822	0.997	0.644	0.645
	Yes	Yes					Yes	0.835	0.884	0.688	0.713
	Yes	Yes	Yes	Yes			Yes	0.836	0.964	0.691	0.693

it either AUC, area under precision recall curve, recall, or F1 score), with the highest value being approximately 0.83, while some articles reporting performance below 0.7. Our final best result, 0.76, is in the middle of the range that is found in the academic articles, with our research having (at least one) unique feature: the patient population. As we analyze data of children aged less than 3 years in average, compared to teenagers or older patients suspected of asthma, we can make use of significantly less historical information as per the very young age. Still, with this limited individual information we managed to obtain performance that is on par with most of the studies. Furthermore, the performance of our models is still significantly higher in comparison to expert evaluations as found in other academic articles and mentioned above.

5.3. AICMS and the chronic care model

In the following, we will discuss how our proposed system, the AICMS, satisfies the criteria specified by the Chronic Care Model (CCM). The main premise of CCM is to emphasize that patients' need to be well-informed, and the healthcare system to be always prepared. This clearly overlaps with the main structure of AICMS by design: the proposed platform keeps both patients/guardians and health care professionals informed at all times and serves as an intelligent intermediary facilitating the interaction when it is required. This general aim is clearly expressed in a more pointed manner in the six components of CCM and the corresponding elements/modules of AICMS. In the following, we present the assessment of the proposed AICMS with respect to the main criteria of the CCM, as worked out through discussions with employees in the pediatric department of Children's Hospital, Zhejiang University School of Medicine, the intended users of the system.

- Delivery System Design (DSD): the hospital client clearly separates the roles of community doctors and specialists. The adaptive role of the AI modules clearly assigns tasks for different health professional groups, pro-actively making recommendations and sending information to the appropriate layer of the healthcare system. Furthermore, follow-up and monitoring, the basis of a well-optimized delivery system, is the core of AICMS.
- Self-Management Support (SMS): the support for self-management is provided mainly by the AI chat bot. This system answers to the inquiries of guardians regarding any aspects of symptoms, treatment, medication, etc. related to the chronic disease. The guardians can also check all the data recorded in the monitoring system to acquire a sense of typical measurement patterns and the child's reaction to different changes. Additionally, a specific module, AI diet recommendation, aids families to follow an optimal selection of daily food sources that can help in mitigating the effects of the disease. Finally, the AI module is designed to continuously evaluate the condition of the patient and provide feedback or contact a healthcare professional when necessary.
- Decision Support (DS): this is the core service provided by the different AI components of AICMS. The database components store all the historical data and are continuously updated as the basis of AI-guided decision support. The modules provide advice to guardians, create and send reminders, offer diet recommendation etc. On the health care side, the result of AI analysis can draw the attention of healthcare professionals to the more urgent cases and also suggest optimal treatment plans. Moreover, while doing so, the system automatically collects data on the results of accepting or rejecting the recommended actions by patients and doctors and continuously improves by learning from the mistakes.
- Clinical Information Systems (CIS): as the databases in AICMS collect and store all patient related data from basic measurements taken at home to hospital visits and doctor's diagnosis, they essentially form a basis of building a CIS. The three important roles of computer information systems are the following: reminder system to improve compliance with guidelines, feedback on performance measures and registries for planning the care for chronic disease patients.
- Community Resources (CR): in our system, this component is, to a large extent, replaced by the AI modules. The communication modules, in particular the AI chat bot, can provide all the necessary knowledge to guardians that was only possible previously in classes or by making use of professional home care assistance. The AI can provide diet recommendations or assess the quality of sports/physical activity performed by the patient and help in improving the quality of life for patients and their families.
- Health Care Organization (HCO): this component is not addressed directly in the article, however, the AI system makes use of information on organizational and hierarchical structures in the healthcare system.

Additionally to considering the generic requirements of the Chronic Care Model, the proposed model, AICMS, addresses the situation, resources and generic health environment around managing children's chronic diseases in China. For this reason, the proposed AICMS framework and its main features have been discussed in the Children's Hospital, Zhejiang University School of Medicine, China, several times through the system design process. The medical personnel participating in these discussions included employees, whose present daily work could be greatly aided and their workload reduced (and to be utilized more efficiently in other tasks) as the consequence of properly implementing a chronic disease management system that is automated to a large extent, such as AICMS.

6. Conclusions

This paper investigates children's chronic disease and child chronic disease management in detail, with special attention paid to the current state of this issue in China. We found that, while children's chronic disease management cycle has several time-consuming components, in practice there is a clear misalignment between resources that need to be continuously invested in children's chronic disease management and the current shortage of medical resources.

In this article, we propose a new model, the AI Chronic Management System (AICMS), that combines state-of-the art artificial intelligence technologies, internet of things and knowledge graphs to form the basis of a new approach to chronic disease management. The proposed system can provide timely, active and efficient management of chronic diseases for child patients. As the answer to RQ1, we identified the knowledge graph as a possible tool that can help in optimizing the operations of the management process. By integrating data from electronic health records, patient reporting and monitoring data in the structure of a knowledge graph, one can construct tools that can aid the doctors and offer real time recommendations when an important change in the patient's status is predicted to happen. We identified several important use scenarios in which the proposed system can significantly ease the load of doctors and nurses, and improve the quality of the provided service for the child patients. To answer RQ2, we specifically looked at the case of recommendations to re-visit the hospital, Q&A component of the remote monitoring component and a chronic disease warning module. To address RQ3, we have utilized anonymized data collected at the pediatric department of Children's Hospital, Zhejiang University School of Medicine, focusing on children who are suspected of having asthma. By making use of state-of-the-art natural language processing tools and comparing various machine learning models, we found that we can achieve results that are superior to expert evaluation performance as reported in other academic works. The presented models also offer performance similar to other proposals, with the presented research being the first documented analysis in academic literature, to the best of our knowledge, focusing on predicting asthma specifically for infants and children aged below 5 years. In order to assess how the model conforms to the industry-standard Chronic Care Model and to answer RQ4, we presented an evaluation of the main CCM criteria with respect to the features of the proposed model. As assessed by the intended users of the system, the employees of the pediatrics department, AICMS satisfies every requirement that ensures timely and correct information flow and appropriateness of interventions.

The structure of the system enables hospitals, community doctors and child guardians to manage children with chronic diseases in a timely, intelligent and efficient manner. Through the management process designed in this paper, one can assist the treatment process of children with chronic diseases, improve their health status, answer knowledge questions in the management procedures, improve the management experience of children with chronic diseases, and save valuable medical and health resources.

In this paper we presented the design of a platform aimed at chronic disease management. We presented the design of the platform, some example use cases, and some preliminary quantitative and qualitative analysis on the feasibility and features of the proposal. The analysis and discussion show that we can achieve good prediction performance for the case of asthma patients and that the proposed system conforms to and improves upon the requirements of the traditional chronic care model. The natural following of this work is an actual implementation of the system. The work will start with mainly focusing on the examples described in more detail in this article, such as database design, follow-up activities, but also on diet recommendation and quality-of-life improvements for patients and guardians. An important component of the implementation will be to test and identify the appropriate machine learning models for each specific problem to be addressed by the platform.

CRediT authorship contribution statement

Gang Yu: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Mohammad Tabatabaei:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **József Mezei:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Qianhui Zhong:** Conceptualization, Project administration, Writing – original draft. **Siyu Chen:** Methodology, Formal analysis. **Zheming Li:** Conceptualization, Funding acquisition, Resources. **Jing Li:** Conceptualization, Funding acquisition, Resources. **LiQi Shu:** Conceptualization, Funding acquisition, Resources. **Qiang Shu:** Funding acquisition, Supervision, 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.

Acknowledgment

The authors are grateful to the editor and referees for their valuable comments and suggestions for improving the paper. This paper was partially supported by the National Key R&D Program of China (2019YFE0126200), the National Natural Science Foundation of China (62076218), and the Zhejiang Province Public Welfare Technology Application Research Project (LGF18H260004).

References

Abuel-Reesh, J., & Abu-Naser, S. (2017). A knowledge based system for diagnosing shortness of breath in infants and children. *International Journal of Engineering and Information Systems*, *1*, 102–115.

Barr, N., Vania, D., Randall, G., & Mulvale, G. (2017). Impact of information and communication technology on interprofessional collaboration for chronic disease management: a systematic review. *Journal of Health Services Research & Policy*, *22*(4), 250–257.

Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, *57*(1), 9–19.

Do, Q., Son, T. C., & Chaudri, J. (2017). Classification of asthma severity and medication using TensorFlow and multilevel databases. *Procedia Computer Science*, *113*, 344–351.

Emanet, N., Öz, H. R., Bayram, N., & Delen, D. (2014). A comparative analysis of machine learning methods for classification type decision problems in healthcare. *Decision Analytics*, *1*(1), 6.

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, *161*, 1–13.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning (Vol. 1)*. Springer Series in Statistics New York.

Goto, T., Camargo Jr., C. A., Faridi, M. K., Yun, B. J., & Hasegawa, K. (2018). Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *The American Journal of Emergency Medicine*, *36*(9), 1650–1654.

Guo, J., & Li, B. (2018). The application of medical artificial intelligence technology in rural areas of developing countries. *Health Equity*, *2*(1), 174–181.

Hoang, K. H., & Ho, T. B. (2019). Learning and recommending treatments using electronic medical records. *Knowledge-Based Systems*, *181*, Article 104788.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130–141.

Huang, M.-J., Chen, M.-Y., & Lee, S.-C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert Systems with Applications*, *32*(3), 856–867.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310.

Jalaliniya, S., & Pederson, T. (2012). A wearable kids' health monitoring system on smartphone. In *Proceedings of the 7th nordic conference on human-computer interaction: Making sense through design* (pp. 791–792). ACM.

Jiang, J., Li, X., Zhao, C., Guan, Y., & Yu, Q. (2017). Learning and inference in knowledge-based probabilistic model for medical diagnosis. *Knowledge-Based Systems*, *138*, 58–68.

Kaplan, A., Cao, H., FitzGerald, J. M., Iannotti, N., Yang, E., Kocks, J. W., et al. (2021). Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *The Journal of Allergy and Clinical Immunology: In Practice*.

Kocsis, O., Arvanitis, G., Lalos, A., Moustakas, K., Sont, J. K., Honkoop, P. J., et al. (2017). Assessing machine learning algorithms for self-management of asthma. In *2017 E-health and bioengineering conference (EHB)* (pp. 571–574). IEEE.

Kong, L.-Z. (2017). China's medium-to-long term plan for the prevention and treatment of chronic diseases (2017–2025) under the healthy China initiative. *Chronic Diseases and Translational Medicine*, *3*(3), 135.

Lei, Z., Sun, Y., Nanehkaran, Y., Yang, S., Islam, M. S., Lei, H., et al. (2020). A novel data-driven robust framework based on machine learning and knowledge graph for disease classification. *Future Generation Computer Systems*, *102*, 534–548.

Liang, H., Tsui, B. Y., Ni, H., Valentim, C. C., Baxter, S. L., Liu, G., et al. (2019). Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nature Medicine*, *25*(3), 433.

Liu, H., Hu, H., Chen, Q., Yu, F., & Liu, Y. (2016). Application of the clinical decision support systems in the management of chronic diseases. In *2016 3rd international conference on systems and informatics* (pp. 482–486). IEEE.

Liu, Y., Lu, W., & Wang, Z. (2009). Current situation and enlightenment of home care services abroad. *Chinese Journal of Nursing*, *44*(7), 645–646.

Malas, T. B., Vlietstra, W. J., Kudrin, R., Starikov, S., Charrout, M., Roos, M., et al. (2019). Drug prioritization using the semantic properties of a knowledge graph. *Scientific Reports*, *9*(1), 6281.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. ArXiv preprint arXiv:1301.3781.

Milani, R. V., & Lavie, C. J. (2015). Health care 2020: reengineering health care delivery to combat chronic disease. *The American Journal of Medicine*, *128*(4), 337–343.

Miranda, J. J., Kinra, S., Casas, J. P., Davey Smith, G., & Ebrahim, S. (2008). Non-communicable diseases in low-and middle-income countries: context, determinants and health policy. *Tropical Medicine & International Health*, *13*(10), 1225–1234.

Moraitou, M., Pateli, A., & Fotiou, S. (2017). Smart health caring home: A systematic review of smart home care for elders and chronic disease patients. In *GeNeDis 2016* (pp. 255–264). Springer.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, *309*(13), 1351–1352.

Ng, C. Y., Thomas-Uribe, M., Yang, Y. A., Chu, M. C., Liu, S.-D., Pulendran, U. P., et al. (2018). Theory-based health behavior interventions for pediatric chronic disease management: A systematic review. *Jama Pediatrics*, *172*(12), 1177–1186.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, *104*(1), 11–33.

Patra, R. (2019). Introduction to knowledge graphs in healthcare. URL: https://blogs.oracle.com/ai/introduction-to-knowledge-graphs-in-healthcare.

Pérez-Ardanaz, B., Morales-Asencio, J. M., García-Piñero, J. M., Lupiáñez-Pérez, I., Morales-Gil, I. M., & Kaknani-Uttumchandani, S. (2019). Socioeconomic status and health services utilization for children with complex chronic conditions liable to receive nurse-led services: A cross-sectional study. *Journal of Nursing Scholarship*, *51*(5), 518–525.

Prasadl, B., Prasad, P., & Sagar, Y. (2011). An approach to develop expert systems in medical diagnosis using machine learning algorithms (asthma) and a performance study. *International Journal on Soft Computing (IJSC)*, *2*(1), 26–33.

Pruitt, S., & Epping-Jordan, J. (2002). *Innovative care for chronic conditions: building blocks for action: Global report*. World Health Organization.

Pujara, J., Miao, H., Getoor, L., & Cohen, W. (2013). Knowledge graph identification. In *International semantic web conference* (pp. 542–557). Springer.

Reynolds, R., Dennis, S., Hasan, I., Slewa, J., Chen, W., Tian, D., et al. (2018). A systematic review of chronic disease management interventions in primary care. *BMC Family Practice*, *19*(1), 11.

Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., & Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. *Scientific Reports*, *7*(1), 5994.

Samb, B., Desai, N., Nishtar, S., Mendis, S., Bekedam, H., Wright, A., et al. (2010). Prevention and management of chronic disease: a litmus test for health-systems strengthening in low-income and middle-income countries. *The Lancet*, *376*(9754), 1785–1797.

Sendra, S., Parra, L., Lloret, J., & Tomás, J. (2018). Smart system for children's chronic illness monitoring. *Information Fusion*, *40*, 76–86.

Seol, H. Y., Rolfes, M. C., Chung, W., Sohn, S., Ryu, E., Park, M. A., et al. (2020). Expert artificial intelligence-based natural language processing characterises childhood asthma. *BMJ Open Respiratory Research*, *7*(1), Article e000524.

Sharma, S., Santra, B., Jana, A., Santosh, T., Ganguly, N., & Goyal, P. (2019). Incorporating domain knowledge into medical NLI using knowledge graphs. ArXiv preprint arXiv:1909.00160.

Shu, L.-Q., Sun, Y.-K., Tan, L.-H., Shu, Q., & Chang, A. C. (2019). Application of artificial intelligence in pediatrics: past, present and future. *World Journal of Pediatrics*, *2*(15), 105–108.

Singhal, A. (2012). Introducing the knowledge graph: things, not strings. *Official Google Blog*, *5*.

Spathis, D., & Vlamos, P. (2019). Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics Journal*, *25*(3), 811–827.

State Health and Family Planning Commission of the People's Republic of China (2015). Report on nutrition and chronic diseases of Chinese residents. URL: http://www.chinadaily.com.cn/m/chinahealth/2015-06/15/content_21008408.htm.

State Health and Family Planning Commission of the People's Republic of China (2018). Chinese family health big data report. URL: https://howtobehealthyforme.wordpress.com/2018/09/29/chinas-big-data-report-on-family-health-2-0-1-7/.

Tay, Y., Luu, A. T., & Hui, S. C. (2017). Non-parametric estimation of multiple embeddings for link prediction on dynamic knowledge graphs. In *Thirty-first AAAI conference on artificial intelligence* (pp. 1–10).

Topalovic, M., Das, N., Burgel, P.-R., Daenen, M., Derom, E., Haenebalcke, C., et al. (2019). Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *European Respiratory Journal*, *53*(4).

Topalovic, M., Das, N., Burgel, P.-R., Daenen, M., Derom, E., Haenebalcke, C., et al. (2018). Artificial intelligence on pulmonary function tests improves respiratory disease diagnosis. In *D28. Respiratory Disease Diagnosis: Pulmonary Function Testing and Imaging* (p. A6390). American Thoracic Society.

Toren, K., Brisman, J., & Järvholm, B. (1993). Asthma and asthma-like symptoms in adults assessed by questionnaires: a literature review. *Chest*, *104*(2), 600–608.

Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th international conference on machine learning-volume 70* (pp. 3462–3471). JMLR. org.

Van Cleave, J., Gortmaker, S. L., & Perrin, J. M. (2010). Dynamics of obesity and chronic health conditions among children and youth. *JAMA*, *303*(7), 623–630.

Wagner, E. H., Austin, B. T., Davis, C., Hindmarsh, M., Schaefer, J., & Bonomi, A. (2001). Improving chronic illness care: translating evidence into action. *Health Affairs*, *20*(6), 64–78.

Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Global Health*, *3*(4), Article e000798.

Weeks, G., George, J., Maclure, K., & Stewart, D. (2016). Non-medical prescribing versus medical prescribing for acute and chronic disease management in primary and secondary care. *Cochrane Database of Systematic Reviews*, *11*.

Wolff, P., Graña, M., Ríos, S. A., & Yarza, M. B. (2019). Machine learning readmission risk modeling: a pediatric case study. *BioMed Research International*, *2019*.

World Health Organization (2005a). Chronic disease fact sheet. URL: http://www.who.int/chp/chronic_disease_report/contents/part1.pdf.

World Health Organization (2005b). Preventing chronic disease: a vital investment. URL: https://www.who.int/chp/chronic_disease_report/contents/foreword.pdf?ua=1.

Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)*, *26*(3), 1–37.

Zhang, J., Xu, W., Guo, J., & Gao, S. (2017). A temporal model in electronic health record search. *Knowledge-Based Systems*, *126*, 56–67.

Zwar, N., Harris, M., Griffiths, R., Roland, M., Dennis, S., Powell Davies, G., et al. (2017). A systematic review of chronic disease management. *The University of New Soth Wales Sschool of Public Health & Community Medicine*.