IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 26, 2021, accepted August 16, 2021, date of publication August 18, 2021, date of current version August 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106116

Guideline-Driven Medical Decision Support Methods for Family Healthcare

HUAQIONG WANG, (Member, IEEE), AND GUIPING QIAN
College of Media Engineering, Communication University of Zhejiang, Hangzhou, Zhejiang 310018, China

Corresponding author: Huaqiong Wang (wanghuaqiong@cuz.edu.cn)

This work was supported by the Basic Public Welfare Research Project of Zhejiang Province under Grant LGF20H180001 and Grant LGG20F020001.

ABSTRACT Medical guidelines are effective to guide medical practice and improve therapeutic effect. Currently, medical guidelines are primarily used in medical institutions but are still not available to the public. To improve the practicality of guidelines in family practice, this paper proposes a novel approach to assisting family medical decision support using semantic technology and open data analysis. A disease-specific knowledge model is constructed using semantic guideline expressions, which provide standard care plans to the public. A medical text corpus is formed by collecting medical information from open data sources online. Via text mining and sentiment analysis of the medical text corpus, the word frequency and sentiment score of different medical procedures are calculated, which are used to provide detailed instructions for public treatment. A guideline-based knowledge model is constructed to create eczema-specific standard care plans. According to the resulting frequencies, hydrocortisone butyrate cream (HBC) elicited the most concern among hormone drugs (54%), followed by mometasone furoate cream (MFC), accounting for 28% among all hormone drugs. According to the average sentiment score, MFC is more frequently recommended than HBC. For skin care products, YMJ elicited the most concern, while Cetaphil was the most recommended. The results of the word frequency analysis and sentiment analysis are combined to provide detailed and clear recommendations to supplement standard care plans for family medical decision support. The method proposed by this paper is an important supplement and extension to in-hospital data analysis.

INDEX TERMS Medical expert systems, text mining, semantic web, sentiment analysis, family practice.

I. INTRODUCTION

With increasing public health awareness and the need to address aging, family healthcare scenarios are becoming more common [1], [2]. However, the contradiction between the increase in health awareness and a lack of medical knowledge makes it difficult for family practice to become popular on a large scale. This contradiction is also the root cause of over-treatment and increases in medical expenses. How to provide the public with effective medical knowledge through computer software and services is an urgent problem that must be solved in the field of family practice.

Medical practice guidelines, as prescriptive standards of clinical conduct [3], are helpful to provide medical knowledge to the public. Many studies have shown that medical guidelines can improve therapeutic effect and even shorten the clinical process [4]–[6]. The key problem that restricts the practicability of medical guidelines is that the treatment procedures defined in guidelines are too general to direct the treatment practice. For example, the medical guideline of acute appendicitis suggests using antimicrobial agents but does not provide specific drug names or doses. To address this problem, many papers have proposed combining electronic medical records with medical guidelines for more precise implementation [7], [8]. Clinical data are collected and stored by electronic medical record systems and other hospital medical information systems. Previous studies extracted specific treatment details from clinical data to supplement and describe the procedures in medical guidelines, which can effectively improve the practicability of the guidelines [9]. The limitation of clinical data is reflected in its privacy security and information barrier between different medical institutions. Therefore, the results based on in-hospital clinical data analysis primarily serve institution-specific clinical

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

decision support. Conversely, this paper is based on open online data, with the goal of improving the availability of medical guidelines in public applications.

Unlike local services within health care organizations, public services do not have access to clinical data recorded in hospital information systems. Another problem is that the clinical data in electronic medical records systems is well-structured, while open data outside of hospitals, which primarily comes from the Internet, is primarily unstructured. Therefore, the approach to improving the usability of guidelines for family practice is different from that for medical institutions.

With the penetration of Internet technology into daily medical treatments, people are more accustomed to seeking help and sharing clinical experience on the Internet. Therefore, a large amount of open online medical data has accumulated [10], [11]. With the rapid growth of medical social websites, the average number of search results for common diseases such as hypertension in search engines has exceeded 100 million. The scale of online data is huge and is growing rapidly. The openness of online data also makes it more convenient to serve the public [12]. Medical procedures or treatment details can be extracted from open online data, which provides a feasible solution to improve the practicability of medical guidelines in family practice.

This paper presents a novel approach to extract effective medical knowledge from online medical data, and demonstrates the construction of a semantic-based framework to provide medical decision support to make medical guidelines more convenient and effective to the public with regard to family practice. The primary contributions of this study can be summarized as follows: (1) a feasible technical solution is provided to improve the practicality of medical guidelines in family practice; (2) disease-specific standard care plans are generated from guideline-based knowledge model; (3) medical suggestions about detailed treatment procedures are extracted from online data; and (4) family medical decision support is provided by combining standard care plans and detailed treatment procedures. This paper explores extracting disease-specific treatment procedures from online social data and proposes a novel method to provide guideline-driven medical decision support for family healthcare. Family medical decision support can enhance the care of patients in family practice and improve patient outcomes in primary care practice.

II. METHODS

Medical data on the Internet is primarily in text form, and its sources are complex and lack integration [13]–[15]. How to extract effective information from these data and express them as standard knowledge is the key problem to solve. Due to the complexity of medical terms, it is impossible to find key medical information via traditional word segmentation methods. To solve these problems, this paper uses semantic technology to build a guideline-based knowledge model that defines key information elements in treatment procedures to support word segmentation, and uses word frequency statistics and text sentiment analysis to extract treatment suggestions that can provide decision-making support to the public.

A. DISEASE-SPECIFIC KNOWLEDGE MODELING

A semantic web provides a common framework for data sharing and reusing [16]. In a semantic web, online data is expressed as natural language, which is understandable to humans but also easily processed by machines, making information discovery, sharing and integration more intelligent [17]–[19]. Semantic technologies typically use Web Ontology Language (OWL) to describe knowledge terminologies [20], use Semantic Web Rule Language (SWRL) to support semantic reasoning [21], and use Jena Semantic Web framework to complete system architecture [22]. With the rapid development of knowledge graphs in recent years, semantic technology has made remarkable progress in language development, tool development, standard establishment and other aspects, and has been used in industry, logistics, medicine and other fields [23]–[25].

To describe guideline-based standard care plans effectively, a disease-specific guideline model has been constructed with Protégé. In previous work, we designed four super classes called **ClinicalPathway**, **CPElementBase**, **CPEventModel**, and **Patient** in guideline ontology, and summarized 84 guideline-related classes and 98 individuals [26]. For a specific disease, the corresponding classes, instances, and properties must be expressed and added to the guideline model. Considering the chronic disease eczema as an example, **Eczema** is an instance of the class **Disease**. Its ICD10 code is set as L30.902 through the data property **hasICD_10Code**. The Chinese name of eczema is defined by the property **rdfs:comment**.

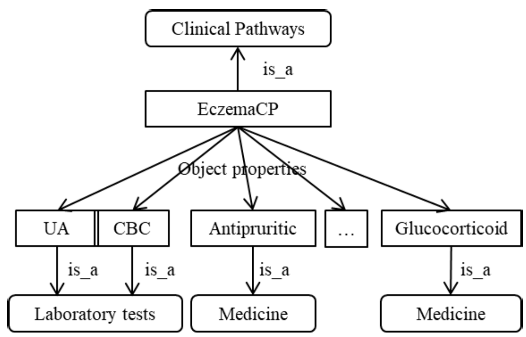

FIGURE 1. Key treatment procedures defined in the instance EczemaCP.

EczemaCP is an instance of the class **ClinicalPathways**, which is used to document the standard care plans of eczema. The routine treatment procedures for chronic eczema are shown in Fig. 1, which includes key procedures, such as urine analysis (UA), cell broadcast center (CBC), antipruritic, glucocorticoid, etc. These treatment procedures are defined in the class **CPEventModel** and each procedure has its own properties to explicit details such as name, dose and operator.

Semantic technology is used to construct disease-specific knowledge model, define the instance of treatment procedures, and refine the key elements during medical guidelines. The role of semantic technology is primarily embodied in two aspects. First, a standard disease-specific care plan is expressed electronically from the disease guideline. Second, these key elements defined in the knowledge model will serve as an important part of customized dictionary for the web text segmentation, providing a basis for extracting key clinical information from the text corpus.

B. DATA ACQUISITION AND PROCESSING

Clinical data can be directly exported from the database of electronic medical record system. With multisource heterogeneous systems, semantic technology can be used for data standardization and unified processing. Semantic technology has been intensively studied in the field of medical data analysis [27], [28]. The process of data acquisition, integration and analysis of online medical data is more complex than in-hospital clinical data analysis. The first step is to identify the reliable online data sources and then actively obtain related information from them. Due to the limited amount of data on a single platform, it is necessary to obtain data from multiple platforms. However, there are marked differences in the structure and organization of web pages in different platforms. The methods for data acquisition are also different. In this paper, the data acquisition process is implemented through a topic-focused crawler based on Python.

The topic-focused crawler is used to obtain relevant medical data for a specific disease. According to the different HTML document structures, website platforms can be divided into two types: static websites and dynamic websites. For different types of platforms, the data-crawling process varies significantly and requires introducing different Python third-party libraries. For static websites, the corresponding URL can be obtained directly using the obtain method of the Request object. The primary difficulty lies in the analysis of the HTML structure of web pages. Python libraries BeautifulSoup and LXML are used to parse web pages and extract the target information. For dynamic websites, the content of web pages changes dynamically with time, environment, or user actions. Selenium is used to simulate the browser's page interactions for dynamic websites. Considering Zhihu website (zhihu.com) as an example, we use Selenium to simulate the pull down operation of the browser; thus, that much more data can be downloaded from the server to the browser. Based on the downloaded results, the BeautifulSoup library is then used for page parsing to extract the target information.

Whether crawling static or dynamic websites, the most important task is to analyze the website type, parse the HTML document structures to obtain as much of the topic-related results as possible and save them locally. In this paper, a MySQL database is used to store the target information, primarily including the page information and the comment information.

C. TEXT SEGMENTATION AND KNOWLEDGE EXTRACTION

Hydrocortisone and cortisone are two different glucocorticoids, and cortisone is 1.25 times more potent as an anti-inflammatory. The text "hydrocortisone" contains the word "cortisone"; therefore, particularly in Chinese, it is not possible to calculate the frequency of drug use by searching and calculating the frequency of drug names appearing on the web pages. This practice is common in medical terminologies. Under such circumstances, it is impossible to extract key treatment information by calculating word frequency directly. The text of the page content must be segmented first. Text segmentation is the process of dividing a document or a sentence into its base components such as lines, words and characters. Chinese text segmentation is an important and widely studies sequence modeling problem. There are many tools or services for implementing Chinese text segmentation. The Jieba library is used in this paper to segment Chinese text.

The Jieba library is a mainstream Python third-party ecological library for text analysis. Its principle of text segmentation is to use a Chinese word library, compare the target content with the library, and find phrases with the highest probability through graph results and dynamic programming methods. In addition to word segmentation, the Jieba library can also use customized Chinese dictionaries to support the identification of medical terms. Three segmentation modes are supported by the Jieba library: (1) precise segmentation mode divides sentences accurately without producing redundant phrases and is suitable for text analysis; (2) all segmentation mode lists all possible phrases in the sentence but includes redundancies; and (3) search engine mode segments long words to improve the recall rate based on the precise segmentation mode. To ensure maximum accuracy and minimum redundancy, this paper creates a customized dictionary using a disease-specific knowledge model and uses the precise segmentation mode to divide the medical text.

After Jieba word segmentation, the word list from medical texts is created. Then, based on the statistical methods, the word frequencies of medical knowledge terms defined in the knowledge model are calculated.

The disease-specific knowledge model accurately describes the definition, attributes and relationships of the key treatment procedures. The customized word dictionary obtained from the model contains a variety of expressions for the key treatment procedures. For example, the common expressions of the drug mometasone furoate cream include "mometasone furoate", "eloson" and "furoic acid". There also may be a variety of treatment procedures for a disease; for example, there are a variety of topical drugs for infants with eczema. During family care and treatment of eczema patients, drugs must be selected, and word frequency statistics are used to indicate the degree of concern associated with drugs. Assuming that n types of hormone drugs are available for disease D, each hormone drug may have m different expressions in the medical text, and the frequency of each expression in the online data is $X_{i,j}$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$. Then, the word frequency of certain hormone drug can be calculated by:

$$\mathrm{WF}_j = \sum_{i=1}^{m} x_{i,j} \tag{1}$$

Among all hormone drugs, the degree of concern of a certain hormone drug is calculated by:

$$\mathrm{p}_j = \sum_{i=1}^{m} x_{i,j} \Big/ \sum_{j=1}^{n}\sum_{i=1}^{m} x_{i,j} \tag{2}$$

D. SENTIMENT ANALYSZE

When people post information on the Internet, personal emotional information is typically included. For example, "It is better to use eloson", and "It is not recommended to use eloson". For the keyword "eloson", the first sentence contains positive sentiment information, while the second contains negative sentiment information. Direct statistics of the word frequency of the keyword "eloson" cannot accurately determine the attitude of the online data to "eloson"; thus, we must analyze the sentiment polarity of the sentence according to the context, and even calculate the sentiment score of the keyword. Sentiment analysis, as a subdomain of opinion mining, is focused on the extraction of emotions and opinions of the people towards a particular topic from textual data. Sentiment scores can describe people's attitudes towards certain keywords. This section focuses on analyzing the sentiment scores of keywords, which can more accurately determine people's underlying attitude towards a given treatment based on online data.

There are two primary methods to analyze sentiment polarity: sentiment dictionaries and machine learning. Due to the limited amount of corpus data, this paper uses a sentiment dictionary to analyze sentiment polarity. A sentiment dictionary combines two types of dictionaries: one is the positive and negative sentiment dictionary provided by Taiwan University, and the other is the positive and negative sentiment dictionary and the degree level dictionary provided by the Zhihu website.

The process of sentiment polarity analysis includes the following steps:

(1) Searching keywords from the corpus;

(2) Extracting the sentence S with its keywords from the text corpus;

(3) Text segmentation with the Jieba library;

(4) Identifying the sentiment words in sentence S with the positive and negative sentiment dictionary. The number of the positive sentiment word is $\mathrm{Np}_j\ (j = 1, 2, \ldots, n)$, and the number of the negative sentiment word is $\mathrm{Nn}_j\ (j = 1, 2, \ldots, n)$.

(5) Identifying the sentiment degree words in sentence S with the degree level dictionary. Assuming that k degree words are identified in sentence S, each degree word has its degree level $\mathrm{D}_{\mathrm{l},j}\ (l = 1, 2, \ldots, k; j = 1, 2, \ldots, n)$, which can be queried in the degree level dictionary. Thus, the degree level of sentence S can be calculated by:

$$\mathrm{D}_j = \sum_{k=1}^{l} D_{k,j} \tag{3}$$

(6) Calculating the sentiment score of the keyword with Formula (4). To avoid the situation where the degree level of sentence S is equal to zero, we add 1 to the degree level of sentence S. Similarly, the sentiment score of the keyword is also increased by 1. For certain hormone drugs, the sentiment scores of m different expressions must be summed up with Formula (5), where S_j stands for the sentiment score of certain hormone drug:

$$\mathrm{Y}_{i,j} = 1 + (\mathrm{Np}_j - \mathrm{Nn}_j) \times (1 + \sum_{k=1}^{l} \mathrm{D}_{k,j}) \tag{4}$$

$$\mathrm{S}_j = \sum_{i=1}^{m} Y_{i,j} \tag{5}$$

(7) Calculating the average sentiment score by dividing the sentiment score with the word frequency yields:

$$E_j = \mathrm{S_j} / \mathrm{WF}_j \tag{6}$$

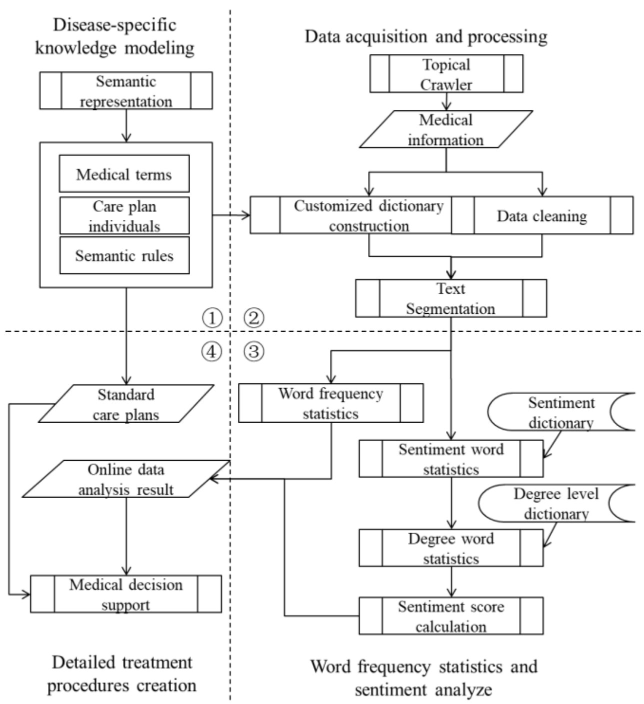

FIGURE 2. Technical architecture of providing family medical decision support.

E. TECHNICAL ARCHITECTURE

The technical architecture of providing family medical decision support is shown in Fig. 2, which includes four phases. First, semantic technologies are used to represent guideline-based knowledge model, which is responsible for customized dictionary construction and standard care plans creation. Second, topical crawlers are created to capture medical information from online text. Via data cleaning and text segmentation, word lists are extracted from the medical information. Third, word frequency statistics and sentiment analysis are both used. Finally, family assistant decision support is provided by combining guideline-based standard care plans and online data analysis results.

III. RESULTS

Eczema is a chronic inflammatory and itchy skin disease that has a high incidence and frequently recurs, and typically requires long-term treatment. In recent years, the incidence of eczema in infants and young children has increased each year, causing serious problems to patients and their families [29], [30].

According to the clinical pathway issued by the Ministry of Health, the urticarial treatment cycle typically lasts for 7 days. External use of glucocorticoids is the primary method that is used to treat mild to moderate eczema. Anti-allergic agents and anti-inflammatory drugs are typically used to assist this treatment: common glucocorticoids include mometasone furoate, hydrocortisone, cortisone and betamethasone; anti-allergic agents include Glycyrrhizin Tablet, Chlorphenamine Maleate Tablet, etc.; and anti-inflammatory topical drugs primarily include Mupirocin Ointment. In these drug terminologies, expressions such as hydrocortisone and betamethasone are the primary components of hormone drugs, and Glycyrrhizin Tablet is the alias name of common drugs. There are many expressions of terms in the medical field. For example, each drug contains its name, primary ingredient, alias name, etc. In most cases, these words are not common words. Therefore, direct use of traditional methods of medical text segmentation cannot effectively identify key information.

Various drugs are involved in the treatment procedures of eczema. Table 1 lists 5 common hormone drugs, 4 anti-allergic agents and 1 antibiotic anti-inflammatory drug. The type, drug names, primary ingredient, and aliases of drugs are given. Because certain drugs have multiple aliases, only one of the most common aliases is listed.

Due to their long duration of use, repeatability and frequent occurrence in infants and young children, the treatment process of eczema depends primarily on family care. Too many types of drugs to choose from is the primary problem in home care for eczema. Therefore, this paper chooses eczema as an experimental case to show how to use medical knowledge models and open online data to extract effective treatment information, generate key treatment elements, and provide decision support for family practice via word frequency statistics and sentiment analysis. The decision-support process of family medical consultation for eczema is shown in Fig. 3. An eczema-specific web text corpus is created via websites crawler, and a guideline-based eczema knowledge model is constructed to generate standard care plan and create related dictionaries for Jieba word segmentation. Via frequency statistics and sentiment analysis of the eczema-specific word list, detailed medical decision support can be provided for family practice.

TABLE 1. List of commonly used drugs for eczema.

Type	Drug Name	Ingredient	Chinese Alias
Hormones	Mometasone Furoate Cream	mometasone furoate	艾洛松
Hormones	Hydrocortisone Butyrate Cream	hydrocortisone	尤卓尔
Hormones	Desonide Cream	desonide	力言卓
Hormones	Triamcinolone Acetonide and Econazole Nitrate Cream	rriamcinolone/econazole	派瑞松
Hormones	Betamethasone Cream	betamethasone	荷洛松
Antiallergic	Compound Glycyrrhizin Tablets	glycyrrhizin	美能
Antiallergic	Chlorphenamine Maleate Tablets	chlorphenamine	扑尔敏片
Antiallergic	Levocetirizine Hydrochloride Tablets	levocetirizine	迪皿
Antibiotics	Mupirocin Ointment	mupirocin	百多邦

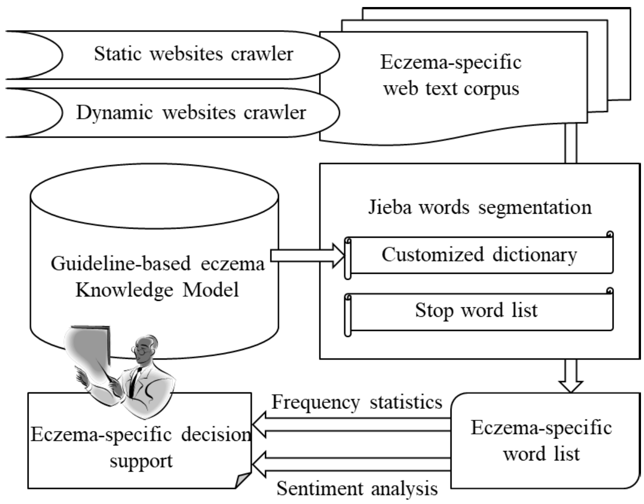

FIGURE 3. Eczema-specific decision support process for family medical consultation.

A. GUIDELINE-BASED ECZEMA KNOWLEDGE MODEL

A good medical knowledge model forms the foundation for effective extraction of key information elements for text analysis. Based on the eczema clinical pathway published by the Ministry of Health, the structure and content of the pathway are analyzed to identify key information elements during treatment. According to the key information elements, a knowledge engineering method is used to construct the eczema-specific ontology model and define the corresponding classes, instances, and properties.

Based on the existing knowledge model, we define the disease instance **Eczema**, construct the care plan instance **EczemaCP**, and explicitly express the key information elements in the treatment procedures of eczema. In the knowledge model of eczema, the association relationship between the care plan instance and the drug instance is shown in Fig. 4, where the prefix **CP** is the abbreviation of clinical pathways and the name of the entire ontology model namespace.

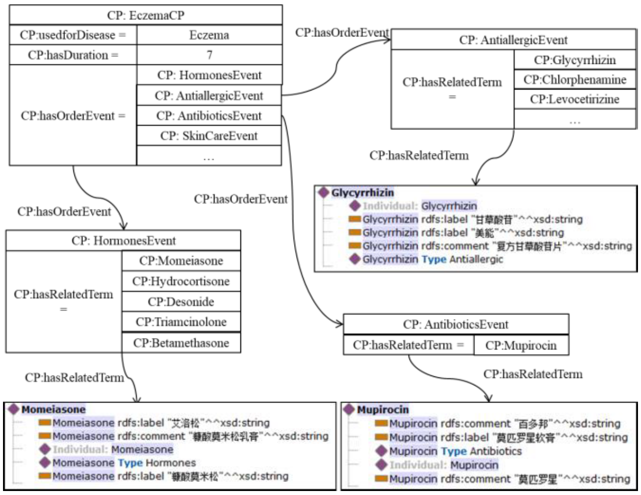

FIGURE 4. **Relationships between the care plan instance EczemaCP and the drug terms in the Eczema knowledge model.**

Fig. 4 lists three important properties and their values of the instance **EczemaCP**. The value of the object property **usedforDisease** indicates that the applicable disease of the care plan instance **EczemaCP** is the disease **Eczema**. The value of the datatype property **hasDuration** represents a 7-day treatment period for this instance. The object property **hasOrderEvert** contains multiple values; each attribute value represents a key treatment procedure during eczema care plan. For example, **HormonesEvent**, **AntiallergicEvent** and **AntibioticsEvent** represent the hormone drug therapy, allergy drug therapy and antibiotics therapy, respectively, which are prescription class instances of doctor's orders. Each doctor's order instance has its own properties associated with the specific drug via the object property **hasRelatedTerm**. Considering hormone drugs as an example, the instance **HormonesEvent** associates with instances of drugs including **Momeiasone**, **Hydrocortisone**, **Desonide**, **Triancinolone** and **Betamethasone**, which correspond to five different types of common hormone drugs listed in Table 1. For each drug instance, its name, primary ingredient name, and alias are set by defining its attributes **rdfs:label** and **rdfs:comment**. On the Protégé platform, the drug instances **Momeiasone**, **Mupirocin**, and **Glycyrrhizin** are also defined as shown in Fig. 4.

B. ECZEMA-SPECIFIC TEXT CORPUS

According to the size of the correlation data, four web forums were selected as data source websites, including DXY (dxy.cn), Zhihu (zhihu.com), Yuer (ci123.com) and 19Lou (19lou.com). The keyword "eczema" was entered into the four platforms to search for eczema-specific information. The search results provided by each platform listed eczema-related articles or consultation information. Due to the marked differences in the web page structures of the different platforms, it is necessary to analyze the HTML structure of web pages and create different web crawlers to collect data from these heterogeneous platforms.

For the platform DXY, Yuer and 19Lou, the search results are organized as static pages. Then, we visit each result page and download the data to the local database. On the Zhihu website, the search results are organized as dynamic pages. The browser drop-down operation is simulated via Selenium to load the result page dynamically and then download detailed data. Due to the download restriction of each website, the number of downloaded eczema-specific pages is limited. The number of pages obtained from DXY, Yuer, 19Lou and Zhihu is 198, 1002, 1248 and 205, respectively. These pages are raw materials that are used to construct an eczema-specific text corpus.

According to these numbers, the number of relevant pages obtained from 19Lou is highest. 19Lou is an online community website that provides a platform for the free exchange of information about local life and consumption. 19Lou is one of the most influential social online media in China. This paper considers the 19Lou forum as an example to show how web page parsing is performed in this study. We analyze the page structure, then use the LXML library to parse HTML webpage files, and finally extract all **div** elements whose class value is '**post-cont**'. The first one of the div elements is the topic of the article, while the others are the reply messages of this article. The basic attributes of the article are then extracted from pages, including article type, posted time, author, etc. Via these operation, the eczema-specific information of 1,130 articles and 7,990 reply messages are extracted from 19Lou and saved into the local MySQL database.

C. MEDICAL DECISION SUPPORT FOR ECZEMA

Using these processes, the guideline-based eczema knowledge model is constructed using semantic technology, and the eczema-specific text corpus is built using web crawlers. This section presents how to provide eczema-specific medical decision support based on the knowledge model and text corpus. The standard care plan for eczema is first created from the guideline-based eczema knowledge model; then the detailed treatment procedures are extracted by the word frequency statistics and sentiment analysis of the eczema-specific text corpus. The entire process can be summarized as follows:

(1) Extract the eczema-specific standard care plan from the knowledge model. The standard care plan can provide basic treatment scheme for the public, while certain procedures may be ambiguous or undefined. The following processes are primarily used to address this problem.

(2) Obtain the attribute values of each drug instance from the eczema knowledge model through SPARQL semantic searching, and create a customized dictionary for word segmentation. The customized dictionary contains the medical term information such as drug name, ingredient and Chinese alias of the drug.

TABLE 2. Frequency statistical results of commonly used drugs for eczema.

Drug Name	DXY	Yuer	19Lou	Zhihu	Total
Mometasone Furoate Cream	17	18	109	69	213
Hydrocortisone Butyrate Cream	33	61	241	82	417
Desonide Cream	9	4	6	27	46
Triamcinolone Acetonide and Econazole Nitrate Cream	14	7	19	34	74
Betamethasone Cream	1	0	0	14	15
Compound Glycyrrhizin Tablets	15	0	0	17	32
Chlorphenamine Maleate Tablets	1	1	1	2	5
Levocetirizine Hydrochloride Tablets	0	0	0	2	2
Mupirocin Ointment	7	26	15	24	72

(3) Query the eczema-related text statements from the MySQL database, including 1,130 articles and 7,990 reply messages.

(4) Segment the text statements using the Jieba library. In this step, the **load_userdict** method of Jieba library is used to load the customized dictionary created in Step (2), the Chinese stop words list is introduced to improve statistical efficiency, and the precise segmentation method called **lcut** is used to segment the eczema-related text obtained in step (3). After text segmentation, a word list including eczema-related keywords is generated.

(5) Calculate the occurrence frequency of eczema-related keywords based on the result word list of text segmentation. The eczema-related keywords primarily include the critical treatment procedures, which are defined in a keyword list extracted from guideline-based eczema knowledge base.

(6) Analyze the sentiment score of each eczema-related keyword. The sentence where the keyword exists is also queried from the MySQL database to execute sentiment analysis. Sentiment dictionaries and degree level dictionaries are developed to calculate the sentiment score of each keyword.

(7) Provide eczema-specific medical decision support by combining the results of word frequency statistics and sentiment analysis.

Word frequency statistics are performed in step (5). Table 2 shows the statistical results of the frequency of commonly used eczema-related drugs. The first column lists the drug instance names defined in the knowledge model. The following four columns list the statistical result of the four different platforms, respectively. The last column lists the aggregated frequency of each drug instance, whose value is equal to the sum of the previous four columns.

Table 2 shows that hydrocortisone butyrate cream elicited the most concern among hormone drugs, and compound glycyrrhizin tablets elicited the most concern among anti-allergy drugs for eczema. According to Formula (2), the proportion of each hormone drug and each anti-allergy drug can be calculated, as shown in Fig. 5 and Fig. 6.

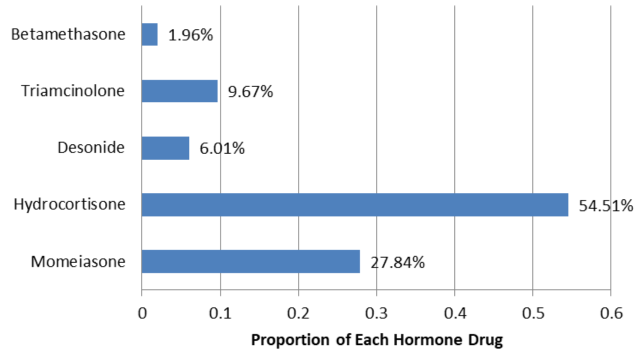

FIGURE 5. Comparison of degrees of concern of different hormone drugs.

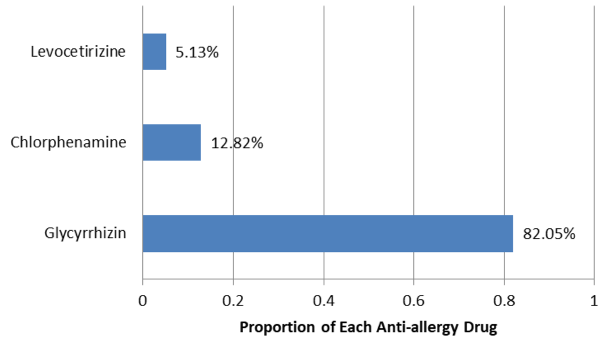

FIGURE 6. Comparison of degrees of concern of different anti-allergy drugs.

Among all hormone drugs, the degree of public concern for hydrocortisone butyrate cream is the highest (417), which accounts for 54% of all hormone drug concerns. Followed by mometasone furoate cream, the frequency is 213, or 28% of all hormone drugs. The frequencies of these two hormone drugs are much higher than those of other hormone drugs.

Among all anti-allergy drugs, the degree of public concern of compound glycyrrhizin tablets is the highest (32), accounting for 82% of all anti-allergy drugs. The word frequencies of all anti-allergy drugs are shown to be lower than those of all hormone drugs. These results make sense because hormone drugs are the primary prescriptions for eczema, while anti-allergy drugs have not received much attention.

Additionally, as the only antibiotic drug, the word frequency of mupirocin ointment is 72. Compared to the anti-allergy drugs, this antibiotic drug is more often used in conjunction with the hormone drugs for eczema treatment.

Sentiment analysis of the drugs is executed in step (6). For each eczema-related keyword, the sentence where it exists is extracted from the articles or reply messages. Via text segmentation of the sentence, the sentiment words and degree level words are identified. Then, the sentiment score of each keyword is calculated with Formula (6). The sentiment analysis results of commonly used drugs for eczema are shown in Fig. 7. The value on the vertical axis indicates the word frequency of each keyword. The size of each bubble, which is also shown next to the bubble, indicates the average sentiment score of each keyword, which is calculated by dividing sentiment score by word frequency.

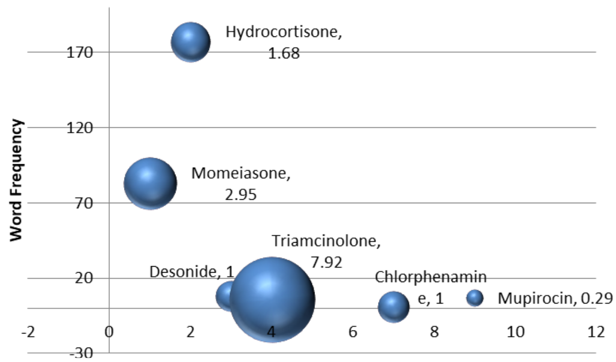

FIGURE 7. Sentiment analysis results of commonly used drugs for Eczema.

According to the sentiment analysis results, the average sentiment score of mometasone furoate cream is 2.95, which is higher than hydrocortisone butyrate cream (1.68). This result demonstrates that mometasone furoate cream is more recommended than hydrocortisone butyrate cream for eczema treatment. Among all drugs, the sentiment score of Triamcinolone Acetonide and Econazole Nitrate Cream is the highest, but its word frequency is relatively low.

The standard care plan, combined with frequency statistical results and sentiment analysis results, will provide intuitive and friendly eczema-specific medical suggestions to people of the public that must chose a drug to treat eczema. This method can provide clinical decision support to help establish a final treatment plan in family practice.

Compared to in-hospital clinical data, open online medical data typically contains more information about routine health care, nursing and other nonprescription treatment details. For eczema, in addition to drug therapy, daily skin care is also the primary treatment and corresponds to the instance **SkinCareEvent** defined in the eczema-specific knowledge model. There are many different skin care products for infants, and the choice of skin care products is important for treatment results. Currently, the skin care brands on the market primarily include Aveeno, California Baby, Johnson and Johnson, Cetaphil and YMJ. Five commonly used skin care products of these brands are defined in the medical terms corresponding to the instance **SkinCareEvent**. The process of frequency statistics and sentiment analysis of different skin care products is the same as that applied to the different drugs above. By calculating the word frequency and average sentiment score of different skin care products, the results are shown in Fig. 8. The bars in Fig. 8 show the word frequencies of five skin care products, and the line points represent their average sentiment scores.

According to the frequencies shown in Fig. 8, the native skin care band called YMJ elicited the most concern and has a word frequency of 439, which accounts for 68% of all skin care products. According to the sentiment analysis results, the skin care band called Cetaphil, with an average sentiment score 5.76, is the most recommended.

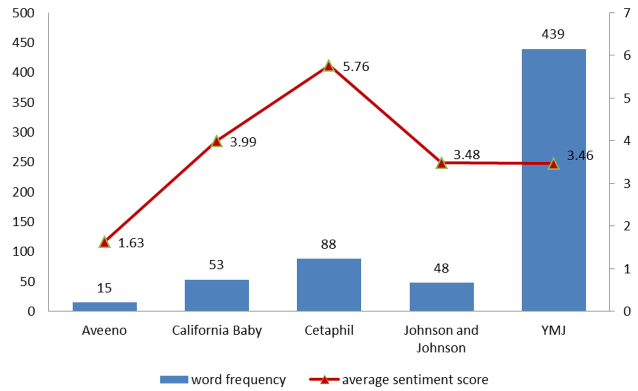

FIGURE 8. Frequencies and sentiment scores of commonly used skin care products for Eczema.

IV. DISCUSSION

The object of this paper is to solve the contradiction between the increase in health awareness and the lack of medical knowledge. Medical guidelines can help address this problem. However, the scope of use of medical guidelines is currently limited to medical institutions. Previous work primarily focused on knowledge base construction and in-hospital clinical data analysis. Patient-specific individual guidelines [26] and hospital-specific customized guidelines [9] are created via semantic technology and clinical data timing, concentrating on the extraction and representation of CP domain terms and rules. Therefore, previous experience in building knowledge models provides a solid foundation for generating standard care plans. The challenge is how to perform text mining and extract effective medical suggestions from online data. Many studies have focused on the analysis of online medical data. Reference [31] analyzed information in online health communities and investigated which factors influence patient decisions to switch from a doctor's online to offline medical service. Reference [32] constructed a novel healthcare monitoring framework based on the cloud environment and an online data analytics engine to store and analyze healthcare data precisely. However, studies about extracting disease-specific treatment procedures from online medical data are rare; thus, this topic is an emerging research direction in medical informatics. Reference [33] shows that few good-quality online decision aids are available for people with knee osteoarthritis or low back pain. Studies about combining the results of online data analysis with medical guidelines are even rarer, and it is data-deficient to compare the results of other studies.

Generally speaking, the family medical decision support provided in this paper includes standard care plans and detailed treatment procedures. Standard care plans extracted from guideline-based knowledge model provide the public with a standard, general treatment scheme. In this process, semantic technologies are introduced for knowledge representation, reasoning and inquiry. Detailed treatment procedures, as a supplement to standard care plans, provide detailed, clear and implementable treatment recommendations to the public. Detailed treatment procedures are obtained via frequency statistics and sentiment analysis of online text corpus.

In the results section, medication recommendations for eczema are considered as an example to show how to provide family medical decision support by combining the disease-specific knowledge model and text corpus based on open online data. A standard 7-day care plan of eczema is generated from the guideline-based knowledge model, and key treatment procedures, such as hormone drugs, anti-allergy drugs and skin care products, are not explicitly suggested. Patients may select hormone drugs and anti-allergy drugs based on the results shown in Fig. 5, Fig. 6, and Fig. 7. The selection of skin care products can be based on the results shown in Fig. 8. With the frequencies and sentiment analysis results, suggestions of detailed treatment procedures are provided to supplement and refine the guideline-based standard care plans.

Compared to other diseases, the treatment procedures for eczema are relatively simple. We chose eczema as an experimental example for two reasons. First, the characteristics of the online data made eczema a good object for research. For diseases with more complex treatment procedures, such as acute appendicitis, key treatment procedures rely on in-hospital interventions, such as operations, where data are primarily recorded in the electronic medical record systems. For these diseases, key treatment procedures are too complex to convey to the public. Thus, the online data for these diseases are lower than those for simpler diseases. Second, the assistant decision support method based on online data is proposed primarily for the family medical scenario. This method is suitable for chronic, long-lasting and recurrent diseases whose treatment procedures depend on long-term drug control and routine care. During long-term treatment, there are often problems with the choice of drugs or equipment. Therefore, the methods proposed in this paper are not limited to eczema but can also provide corresponding auxiliary treatment support for chronic diseases and daily healthcare of the elderly.

Compared to in-hospital clinical data, the disadvantage of online data is primarily reflected in its low quality. Online data rely on the description of the treatment procedures of the disease by patients, who typically lack the knowledge of standard terminology and medical profession. Thus, descriptions of treatment procedures are typically unclear and incomplete. To address this problem, semantic technology is introduced into the text analysis process in this paper, and word frequency statistics and sentiment analysis are also combined to improve the accuracy of text analysis. In future work, patient-specific characteristic data will be integrated to provide personalized treatment procedures. Currently, personalized recommendations are difficult to achieve due to a lack of necessary original data support. In this paper, guideline-based standard care plans and data-based detailed treatment procedures are provided to assist the clinical decision support in family practice.

Compared to in-hospital clinical data, the advantages of online data primarily lie in its openness and increasing quantity. These data are available freely on the Internet, which can be better accessed and used by the public without involving the disclosure of patient privacy. Research based on online data is more focused on statistical results rather than individual records. The method presented in this paper is an exploratory attempt to provide a basic technical framework to assist family medical decision support via Internet-based data extraction. Based on the current amount of eczema-related data on various platforms, it is not sufficient to reach the level of big data currently. With the development of information technology and further popularization of web applications, online medical data will likely continue to increase and reach a higher level of standardization. Open online data will be an important supplement to clinical data and contribute to the development of family practice. For example, the analysis results of skin care products, as mentioned in the experimental results, can support eczema treatment. These nonprescription data can only be obtained through the online platform, which is an important supplement and extension to the clinical treatment plans. Additionally, medical-related data on the Internet may expose public opinions more efficiently, thus contributing to epidemic surveillance and early warnings of depression.

V. CONCLUSION

This paper applies medical guidelines to family practice and proposes a novel approach to family medical decision support. Standard care plans are created based on the guideline-based knowledge model, which serve as standard, general treatment schemes for the public. Detailed treatment procedures are extracted from online text corpus to provide detailed, clear and implementable treatment recommendations. In this paper, semantic technology is used to describe knowledge expression, reasoning and searching. Web crawler technology is used to collect online data. Frequency statistics and sentiment analysis are combined to extract medical suggestions from text corpus. Based on these technologies, this paper proposes a novel family medical decision support method that is an important supplement and extension to in-hospital data analysis. As a portion of medical big-data analysis, the proposed methods will play an important role in promoting family practice.

REFERENCES

[1] G. Charbonneau, "Family medicine professional profile," *Can. Family Physician Medecin De Famille Canadien*, vol. 64, no. 2, p. 157, 2018.

[2] K. Dhillon, D. Jerome, R. Teeluck, and Y. Yu, "Advance care planning in family medicine training," *Can. Family Physician Medecin De Famille Canadien*, vol. 64, no. 5, pp. 394–396, 2018.

[3] S. V. Rajkumar and P. Sampathkumar, "Medical practice guidelines," *Western J. Med.*, vol. 161, no. 6, p. 618, 1995.

[4] R. Vadia and T. Stargardt, "Impact of guidelines on the diffusion of medical technology: A case study of cardiac resynchronization therapy in the UK," *Appl. Health Econ. Health Policy*, vol. 19, no. 2, pp. 243–252, 2021.

[5] J. R. Bach, "A short history of medical expert guidelines and how they pertain to tracheostomy tubes and physical medicine and rehabilitation," *Amer. J. Phys. Med. Rehabil.*, vol. 98, no. 7, pp. 622–626, 2019.

[6] L. Onofa, M. Odunleye, L. Kola, and O. Gureje, ''Reliability and clinical utility of ICD-11 diagnostic guidelines for severe mental disorders in Nigeria,'' *Arch. Med. Res.*, vol. 50, no. 8, pp. 535–542, Nov. 2019.

[7] J. Studnicki, D. V. Schapira, and D. D. Bradham, ''Response to the National Cancer Institute alert: The effect of practice guidelines on two hospitals in the same medical community,'' *Cancer*, vol. 72, no. 10, pp. 2986–2992, 2015.

[8] D. A. Schwartz, L. J. Ghazi, and R. Miguel, ''Guidelines for medical treatment of Crohn's perianal fistulas: Critical evaluation of therapeutic trials,'' *Inflammatory Bowel Diseases*, vol. 21, no. 4, p. 737, 2015.

[9] H.-Q. Wang, T.-S. Zhou, L.-L. Tian, Y.-M. Qian, and J.-S. Li, ''Creating hospital-specific customized clinical pathways by applying semantic reasoning to clinical data,'' *J. Biomed. Informat.*, vol. 52, pp. 354–363, Dec. 2014.

[10] Y. Xu, Z. Zhang, and M. Liu, ''Design of cancer classification and visualization platform based on internet big data,'' *J. Phys., Conf. Ser.*, vol. 1650, Oct. 2020, Art. no. 032084.

[11] M. Reiser, B. Wiebner, and J. Hirsch, ''Neural-network analysis of socio-medical data to identify predictors of undiagnosed hepatitis C virus infections in Germany (DETECT),'' *J. Transl. Med.*, vol. 17, no. 1, p. 94, Dec. 2019.

[12] Y.-J. Mei, Y.-M. Mao, F. Cao, T. Wang, and Z.-J. Li, ''Using internet search data to explore the global public concerns in ankylosing spondylitis,'' *Postgraduate Med. J.*, vol. 97, no. 1144, pp. 93–96, Feb. 2021, doi: 10.1136/postgradmedj-2019-137407.

[13] F. S. Lu, S. Hou, K. Baltrusaitis, M. Shah, J. Leskovec, R. Sosic, J. Hawkins, J. Brownstein, G. Conidi, J. Gunn, J. Gray, A. Zink, and M. Santillana, ''Accurate influenza monitoring and forecasting using novel internet data streams: A case study in the Boston metropolis,'' *JMIR Public Health Surveill.*, vol. 4, no. 1, p. e4, Jan. 2018.

[14] R. Kimmons and G. Veletsianos, ''Public internet data mining methods in instructional design, educational technology, and online learning research,'' *TechTrends*, vol. 62, no. 5, pp. 492–500, Sep. 2018.

[15] E. Velasco, T. Agheneza, K. Denecke, G. Kirchner, and T. Eckmanns, ''Social media and internet-based data in global systems for public health surveillance: A systematic review,'' *Milbank Quart.*, vol. 92, no. 1, pp. 7–33, Mar. 2014.

[16] C. B. Mahmoud, I. Azaiez, and F. Bettahar, ''Multi-intentional description of learning semantic web services,'' *Int. J. Semantic Web Inf. Syst.*, vol. 16, no. 2, pp. 108–125, Apr. 2020.

[17] A. Abello, O. Romero, T. B. Pedersen, R. Berlanga, V. Nebot, M. J. Aramburu, and A. Simitsis, ''Using semantic web technologies for exploratory OLAP: A survey,'' *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 571–588, Feb. 2015.

[18] M. Klusch, P. Kapahnke, S. Schulte, F. Lecue, and A. Bernstein, ''Semantic web service search: A brief survey,'' *Künstliche Intell.*, vol. 30, no. 2, pp. 139–147, 2016.

[19] P. Ristoski and H. Paulheim, ''Semantic web in data mining and knowledge discovery: A comprehensive survey,'' *Web Semantics Sci. Services Agents World Wide Web*, vol. 36, pp. 1–22, Jan. 2016.

[20] P. Patel-Schneider and I. Horrocks, ''OWL 1.1 web ontology language overview,'' World Wide Web Consortium, Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 45, Feb. 2006, pp. 990–996, vol. 63, no. 45.

[21] F. J. G. Clemente, G. M. Pérez, J. A. B. Blaya, and A. F. G. Skarmeta, ''On the application of the semantic web rule language in the definition of policies for system security management,'' in *Proc. OTM Confederated Int. Conferences Move Meaningful Internet Syst.*, in Lecture Notes in Computer Science, 2005, pp. 69–78.

[22] B. McBride, ''Jena: A semantic web toolkit,'' *IEEE Internet Comput.*, vol. 6, no. 6, pp. 55–59, Nov. 2002.

[23] M. Rico, D. Vila-Suero, I. Botezan, and A. Gómez-Pérez, ''Evaluating the impact of semantic technologies on bibliographic systems: A user-centred and comparative approach,'' *J. Web Semantics*, vol. 59, pp. 100500.1–100500.14, Dec. 2019.

[24] M. Smid, ''OnGIS: Geospatial data integration using semantic technologies and query containment,'' *Int. J. Semantic Web Inf. Syst.*, vol. 15, no. 1, pp. 1–20, Jan. 2019.

[25] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, ''A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion,'' *Inf. Fusion*, vol. 63, pp. 208–222, Nov. 2020.

[26] H.-Q. Wang, J.-S. Li, Y.-F. Zhang, M. Suzuki, and K. Araki, ''Creating personalised clinical pathways by semantic interoperability with electronic health records,'' *Artif. Intell. Med.*, vol. 58, no. 2, pp. 81–89, Jun. 2013.

[27] H. H. G. Noriega and F. G. Sanchez, ''Semantic (big) data analysis: An extensive literature review,'' *IEEE Latin Amer. Trans.*, vol. 17, no. 5, pp. 796–806, May 2019.

[28] H. Zhang, Y. Guo, Q. Li, T. J. George, E. Shenkman, F. Modave, and J. Bian, ''An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival,'' *BMC Med. Informat. Decis. Making*, vol. 18, no. S2, p. 41, Jul. 2018.

[29] K. Powell, E. L. Roux, J. P. Banks, and M. J. Ridd, ''Developing a written action plan for children with eczema: A qualitative study,'' *Brit. J. Gen. Pract.*, vol. 68, no. 667, pp. e81–e89, Feb. 2018.

[30] J. Ring, A. Alomar, T. Bieber, M. Deleuran, A. Fink-Wagner, C. Gelmetti, U. Gieler, J. Lipozencic, T. Luger, A. P. Oranje, T. Schäfer, T. Schwennesen, S. Seidenari, D. Simon, S. Ständer, G. Stingl, S. Szalai, J. C. Szepietowski, A. Taïeb, and T. Werfel, ''Guidelines for treatment of atopic eczema (atopic dermatitis): Part I,'' *J. Eur. Acad. Dermatol. Venereol.*, vol. 26, no. 8, pp. pp. 1045–1060, 2012.

[31] Y. Li, Y. Song, W. Zhao, X. Guo, X. Ju, and D. Vogel, ''Exploring the role of online health community information in patients' decisions to switch from online to offline medical services,'' *Int. J. Med. Informat.*, vol. 130, Oct. 2019, Art. no. 103951.

[32] F. Ali, S. El-Sappagh, S. M. R. Islam, A. Ali, M. Attique, M. Imran, and K.-S. Kwak, ''An intelligent healthcare monitoring framework using wearable sensors and social networking data,'' *Future Gener. Comput. Syst.*, vol. 114, pp. 23–43, Jan. 2021.

[33] M. A. Fajardo, B. Durayb, H. Zhong, L. Trevena, A. Traeger, and C. Bonner, ''Online decision aids for knee osteoarthritis and low back pain: An environmental scan and evaluation,'' *Med. Decis. Making*, vol. 39, no. 4, pp. 328–335, May 2019.

HUAQIONG WANG (Member, IEEE) was born in Ningbo, Zhejiang, China, in 1987. She received the Ph.D. degree from the Department of Biomedical Engineering, Zhejiang University, in 2015. She is currently a Lecturer with the Communication University of Zhejiang, China. Her research interests include deep learning, semantic technology, and medical informatics.

GUIPING QIAN was born in 1975. He received the B.S. degree from the Department of Electrical Engineering, Zhejiang University, in 1997, and the Ph.D. degree from the Department of Computer Science and Engineering, Zhejiang University, in 2008. He is currently an Associate Professor with the Communication University of Zhejiang, China. His research interests include digital geometry process, deep learning, and computer vision.