Journal of Healthcare Engineering

Data Mining for Biomedicine and Healthcare

Lead Guest Editor: Zhengxing Huang
Guest Editors: Jose M. Juarez and Xiang Li

Data Mining for Biomedicine and Healthcare

Journal of Healthcare Engineering

Data Mining for Biomedicine and Healthcare

Lead Guest Editor: Zhengxing Huang
Guest Editors: Jose M. Juarez and Xiang Li

Editorial Board

Contents

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 7107629, 2 pages
https://doi.org/10.1155/2017/7107629

Editorial

Data Mining for Biomedicine and Healthcare

Zhengxing Huang,[1] Jose M. Juarez,[2] and Xiang Li[3]

[1]*College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*
[2]*University of Murcia, Murcia, Spain*
[3]*IBM Research, Beijing, China*

Correspondence should be addressed to Zhengxing Huang; zhengxinghuang@zju.edu.cn

Received 8 June 2017; Accepted 8 June 2017; Published 20 August 2017

1. Introduction

Artificial intelligence (AI), as a computational method, attempts to mimic, in a very simplistic way, the human cognition capability so as to solve problems that have defined solution using conventional computational techniques. Medicine is unquestionably the main area in which AI tools and techniques have been applied. Medicine plays an essential role in human life to assess and solve problems with guarantees of success. Medicine can be considered a multidisciplinary science by itself, laying on the advances and discoveries of other sciences (biology, physics, statistics, etc.) and using such results as tools to solve diagnosis or prognosis problems. For this reason, AI in medicine has become a great issue. In this regard, developing and applying novel AI techniques and tools, due to their particular strengths, to solve medical problems and provide valuable information might be of interest to healthcare professionals in their decision making [1, 2].

AI-based medical applications can be classified into either knowledge intensive or data driven. The knowledge intensive approaches intend to obtain computational models, extracted from the clinical literature and experts' experience, to represent concepts, their relations, and the mechanisms to enable automatic reasoning to support medical decisions. Unlike knowledge intensive, data-driven approaches, at the center of the vision of learning health systems, extract such knowledge directly from the collected clinical data and this knowledge is traditionally used to explain a patient's current symptoms/vital signs and to predict future disease progression of the patient.

While an increasing amount of data is being produced by various biomedical and healthcare systems, they have not yet fully capitalized on the transformative opportunities that these data provide. Applying data-driven techniques to big health data can be of great benefit in the biomedical and healthcare domain, allowing identification and extraction of relevant information and reducing the time spent by biomedical and healthcare professionals and researchers who are trying to find meaningful patterns and new threads of knowledge.

The major goal of this special issue is to bring together the researchers in healthcare and data mining to illustrate pressing needs, demonstrate challenging research issues, and showcase the state-of-the-art research and development. The selected papers underwent a rigorous extra refereeing and revision process. It is glad to see that the selected papers presented novel methods that were empirically evaluated on medical datasets. In addition, most of the papers in this special issue include healthcare experts as coauthors, which is beneficial to open new approaches to medical experts in this multidisciplinary field. Moreover, the relevance of this special issue is strengthened by the fact that data-driven medicine is also an object of research in different communities. These communities include AI, medicine, medical informatics, decision support, and healthcare management.

2. The Special Issue

Medical activities can be divided into six tasks: screening, diagnosis, treatment, prognosis, monitoring, and manage-

ment. Among these, diagnosis is a particular important task that data-driven approaches have been widely applied to tackle with. Diagnosis is the process of determining which disease or condition explains the patient's symptoms. The information required for diagnosis is typically collected from a patient's medical records. The paper presented by F. Yang et al. studied a problem of esophageal cancer diagnosis. Esophageal cancer is one of the fastest rising types of cancers in China, and the Kazak nationality is the highest-risk group in Xinjiang, China. They propose an effective computer-aided diagnostic system to assist physicians in interpreting digital X-ray image features and improving the quality of diagnosis. The modules of the proposed system include image preprocessing, feature extraction and selection, and image classification for disease diagnosis. They evaluated their system using 300 original esophageal X-ray images, and the experimental results show that their system is promising for the diagnostics of esophageal cancer.

As an irremediable neurodegenerative disorder that causes dementia in elderly people around the globe, Alzheimer's disease has been widely studied in medical informatics. Earlier, the majority of studies were accomplished manually or semi-manually for measuring the a priori region of interest (ROI) of patient images, which are not practicable in hospital. To this end, D. Jha et al. presented an automated approach for Alzheimer's disease diagnosis. Specifically, a dual-tree complex wavelet transform was first adopted for extracting features from magnetic resonance images. Then, the authors applied principal component analysis for feature dimensionality reduction. The reduced feature vector was sent to feed-forward neural network to distinguish Alzheimer's disease and healthy control from the input MR images. The experimental results showed their approach achieved better performance than the state-of-the-art algorithms.

Reusing the data from healthcare information systems can effectively facilitate clinical trials; however, the selection of eligible patients for clinical trial recruitment criteria is still an open problem. In this issue, Y. Zhang et al. presented a computer-aided clinical trial recruitment methodology, based on syntax translation between different domain-specific languages. In their proposal, the clinical trial recruitment criteria are formally represented as general rule that are translated into intermediate query-oriented domain specific languages to map the native database queries. This approach directly uses the underlying database schema as a reference model. In consequence, the system complexity and data-mapping efforts are greatly reduced.

F. Gräßer et al. presented a recommender system for data-driven therapy decision support. They proposed two methods for therapy recommendation, namely, collaborative recommender and demographic-based recommender. Both algorithms aim to predict the individual response to different therapy options using patient data. The authors evaluated their methods using a clinical database incorporating patients suffering from the autoimmune skin disease psoriasis and showed that their methods profit from a combination into an overall therapy recommender system.

A major challenge in medical informatics is to build a regional health information exchange system. J. Lei et al. have faced this challenge in the context of health national reform in China. That is, they study the impact of the large-scale construction of hospitals and the rapid development of health information systems. Specifically, the authors used interviews, focus groups, a filed study, and a literature review to collect insights and analyze data. The case study was conducted in Xinjin region, which was able to build a complete, unified, and shared information system and many electronic health record components to integrate and manage health resources for 198 health institutions in its jurisdiction. Costs and benefits were carefully discussed, and the unique characteristics of the Xinjin case and a comparison with US cases were analyzed.

Acknowledgments

The guest editors of this special issue would like to thank all authors of the submitted papers, as well as all reviewers for their hard work and detailed reviews that led to eight accepted papers.

Zhengxing Huang
Jose M. Juarez
Xiang Li

References

[1] E. H. Shortliffe, "The adolescence of AI in medicine: will the field come of age in the '90s?," *Artificial Intelligence in Medicine*, vol. 5, no. 2, pp. 93–106, 1993.

[2] N. Peek, C. Combi, R. Marin, and R. Bellazzi, "Thirty years of artificial intelligence in medicine conferences: a review of research themes," *Artificial Intelligence in Medicine*, vol. 65, no. 1, pp. 61–73, 2013.

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 9060124, 13 pages
https://doi.org/10.1155/2017/9060124

Research Article

Diagnosis of Alzheimer's Disease Using Dual-Tree Complex Wavelet Transform, PCA, and Feed-Forward Neural Network

Debesh Jha, Ji-In Kim, and Goo-Rak Kwon

Department of Information and Communication Engineering, Chosun University, 309 Pilmun-Daero, Dong-Gu, Gwangju 61452, Republic of Korea

Correspondence should be addressed to Goo-Rak Kwon; grkwon@chosun.ac.kr

Received 30 December 2016; Revised 22 March 2017; Accepted 30 April 2017; Published 21 June 2017

Academic Editor: Jose M. Juarez

Background. Error-free diagnosis of Alzheimer's disease (AD) from healthy control (HC) patients at an early stage of the disease is a major concern, because information about the condition's severity and developmental risks present allows AD sufferer to take precautionary measures before irreversible brain damage occurs. Recently, there has been great interest in computer-aided diagnosis in magnetic resonance image (MRI) classification. However, distinguishing between Alzheimer's brain data and healthy brain data in older adults (age > 60) is challenging because of their highly similar brain patterns and image intensities. Recently, cutting-edge feature extraction technologies have found extensive application in numerous fields, including medical image analysis. Here, we propose a dual-tree complex wavelet transform (DTCWT) for extracting features from an image. The dimensionality of feature vector is reduced by using principal component analysis (PCA). The reduced feature vector is sent to feed-forward neural network (FNN) to distinguish AD and HC from the input MR images. These proposed and implemented pipelines, which demonstrate improvements in classification output when compared to that of recent studies, resulted in high and reproducible accuracy rates of $90.06 \pm 0.01\%$ with a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.6 \pm 0.03\%$ with 10-fold cross-validation.

1. Introduction

Alzheimer's disease (AD) is an irremediable neurodegenerative disorder that causes dementia in elderly people around the globe. It has been predicted that the pervasiveness of AD will double within the next 2 decades and that one out of every 85 people will be afflicted with the disease by 2050 [1]. Therefore, there is a need to identify neuroimaging biomarkers that can grant accurate and early diagnoses of dementia. In addition, to diagnose an AD sufferer clinically at a primitive disease stage, many imaging biomarkers must be identified using different imaging modalities, such as MRI [2], position emission tomography (PET) [3], functional magnetic resonance imaging (fMRI) [4], single-photon emission computed tomography (SPECT) [5], and magnetic resonance spectral imaging (MRSI) [6].

An accurate and early diagnosis of AD and identification of the risk of progression from mild cognitive impairment (MCI) to AD provide AD sufferers with awareness of the condition's severity and allow them to take preventative measures, such as making lifestyle changes and taking medications [7]. Currently, many neurologists and medical analysts have been spending significant time to researching technique to allow for early diagnosis of AD, and encouraging results have been frequently achieved [8]. MRI is an influential, noninvasive brain imaging technique that provides higher-quality information about the shape and volume of the brain than computed tomography (CT), SPECT, and PET scans. It provides superior soft tissue differentiation, high spatial resolution, and better contrast and can even identify tiny irregularities in the brain [9]. Moreover, the diagnostic use of MRI has been tremendously improved due to the automated and precise labeling of MR images, which performs an important role in identifying AD in related patients from healthy and elderly controls (HC) [10].

Earlier, majority of diagnosis work was accomplished manually or semimanually for measuring a priori region of interest (ROI) of MRI, based on the reality that subjects with

AD experience have more cerebral atrophy when compared to HCs [11, 12]. Most of this ROI-based examination focused on the contracting of the cortex and hippocampus and amplified ventricles. Nevertheless, ROI-based approaches are not practicable in hospitals because of few shortcomings: (i) ROI technique needs a priori data and expert knowledge. (ii) The manual diagnosis accuracy is dependent on the knowledge of physicians and interpreter [13]. (iii) The interaction among the voxels was troublesome to enforce. (iv) It was essential to explore other potential areas that may be linked to AD [14]. (v) Automatic segmentation of ROI was not beneficial in practice, and investigator needed to segment the brain using hand [15]. Therefore, automated methods can assist physician in diagnosing diseases from images such as those produced by MRI, for which many slices are extracted from the tissues and long periods of may be necessary for the evaluation of the images.

The aim of this article is to present an automated approach for diagnosing AD by using the "whole brain analysis" method. It has achieved popularity, since it examines entire voxels of the brain. It is not essential to segment the brain as earlier, and it does not require any biomarker for the classification purpose. The main drawback is dimensionality that can be resolved through high-speed computers, which is comparably inexpensive [16]. The whole-brain investigation laboriously relies on true computation, and it can only be finished by a computer researcher after a physician assisted in labeling the input data as either AD or HC. Usually, the whole-brain inspection labels the entire brain as a ROI, where two stages are involved, namely, feature extraction and classification.

Scholars have presented different methods to extract effective features for the detection of AD and other types of pathological brain disease. Additionally, classification models and methods survive; nevertheless, not all of them are suitable for the processing of MR brain images. Based on latest literature, we found two drawbacks with the previous work: (i) The discrete wavelet transform (DWT) is usually utilized for feature extraction. The DWT has better directional selectivity in horizontal, vertical, and diagonal directions and has better image representation than Fourier transform, but its major drawbacks are that it has poor directionality, is sensitive to shifts, and lacks phase information. (ii) Most of the state-of-the-art mechanisms consider only single slice-based detection (SSD) per patient. The obtained slices may not contain the foci of the disease.

To tackle above problems, we suggested two improvements. First, we propose a DTCWT that possesses attractive properties for image processing, including shift invariance and high directionality. Second, we consider multiple slices for each patient unlike previous studies, so that information gain is more consistent, reliable, and accurate. In hospitals, multiple slice-based detection is utilized because of its inexpensiveness. Research has clearly showed that the DTCWT is more suitable than the traditional wavelet domain for feature extraction [17].

Our contribution aims to introduce a novel method for AD detection with higher accuracy than state-of-the-art methods, on the basis of DTCWT, PCA, and ANN technique. Furthermore, we build a computer-aided diagnosis (CAD) system, which can be utilized in the early diagnosis of AD-related brain area and subjects. Our objective is to develop assisting tool for clinicians.

All of the preprocessing methods are used to obtain good results. To show effectiveness of our proposed system, we have evaluated performance measures including accuracy, sensitivity, specificity, precision, and bar plot for the comparison of the proposed method with the existing systems. The paper is arranged as follows. Section 2 offers background knowledge on materials and methods. In Section 3, the experiments, results, and discussion are presented. Finally, conclusion and plan for future studies are presented in Section 4.

2. Materials and Methods

2.1. Materials

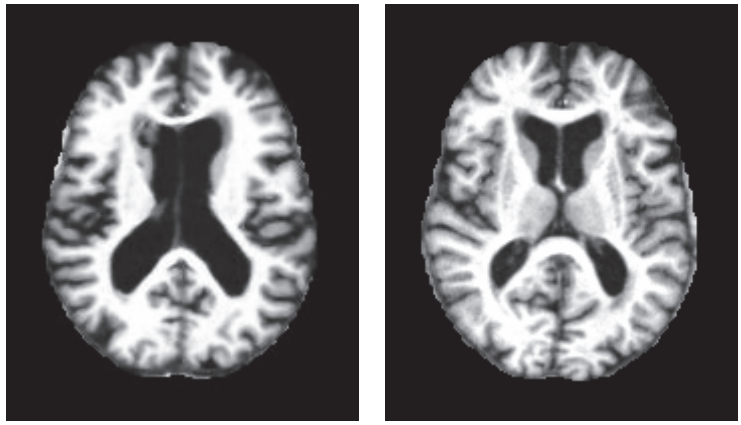
2.1.1. Dataset. In our study, the dataset is accessed from Open Access Series of Imaging Studies (OASIS). OASIS is a project, for compiling and sharing MRI datasets of the brain to make such data accessible to the scientific community. The data are accessible at http://www.oasis-brains.org. A sample of the MR brain image is shown in Figure 1.

OASIS provides two types of data: cross-sectional and longitudinal MRI data. In this study, we used cross-sectional MRI data because we aimed to develop an automatic system for detecting AD, which would not require longitudinal data that had been gathered from AD patients over long periods of time.

The dataset consists of 416 subjects whose ages are between 18 and 96. In our study, we consider 126 samples (including 28 ADs and 98 HCs). Table 1 shows statistical information about the subjects included in the experiment. Only right-handed subjects are included in the study, consisting of both men and women. The exclusion criterion is patients less than 60 years of age or any of their reports are missing. The unbalanced data may cause difficulty in future recognition; we fine-tune the cost matrix to resolve this issue.

The dataset contains information about the patient's demographics. The demographic features contain gender (M/F), age, education, socioeconomic status, and handedness. The mini mental state examination (MMSE) is a short 30-point questionnaire test utilized to monitor for cognitive impairment and dementia. The MMSE test comprises simple questions and problems in numeral areas: the time and place, repeating list of words, arithmetic, language utilization, and comprehension, in addition to basic motor skills. Clinical dementia rating (CDR) is a numeric scale measuring the severity of symptoms of dementia. The patients' cognitive and functional performances were accessed in six areas: memory, orientation, judgement and analytical, community affairs, residence and hobbies, and individual care. The patients' CDR ranks and education level are listed in Tables 2 and 3, respectively.

2.2. Proposed Method. The proposed method consists of three important stages, namely, feature extraction using DTCWT,



(a) Normal (b) Alzheimer's disease

Figure 1: Dataset sample (axial view after preprocessing).

Table 1: Statistical data of the participants.

Factor	HC	AD
No. of patients	98	28
Age (years)	75.91 ± 8.98	77.75 ± 6.99
Education	3.26 ± 1.31	2.57 ± 1.31
Socioeconomic status	2.51 ± 1.09	2.87 ± 1.29
CDR	0	1
MMSE score	28.95 ± 1.20	21.67 ± 3.75
Gender (M/F)	26/72	9/19

Table 2: Clinical dementia rating scale.

CDR	Rank
0	Nondementia
0.5	Very mild dementia
1	Mild dementia
2	Moderate dementia

Table 3: Education codes.

Code	Description
1	Beneath high school graduate
2	Secondary school graduate
3	Some college
4	College graduate
5	Above college

feature dimensionality reduction using PCA, and classification using feed-forward artificial neural network. The overall block diagram of the suggested method is shown in Figure 2. Normalization of image is included in preprocessing section. All of these individual techniques have been proven outstanding, so we strongly believe that the proposed method can also achieve excellent results.

2.3. Image Preprocessing and Normalization. For each patient, each scanning session involves the MR of three or four T1-weighted image scans. In order to add the signal-to-noise ratio (SNR), all indicated MRI scans with the identical protocol of the same individual are motion-corrected and spatially coregistered, to the Talairach coordinate space to produce an averaged image, and then are brain-masked. The motion correction recorded the 3D images of all scans and then developed an average 3D image in initial acquisition space. Also, the scans are then resampled to $1\,\text{mm} \times 1\,\text{mm} \times 1\,\text{mm}$. The obtained image is converted from acquisition space to Talairach coordinate space. Lastly, the brain extraction is achieved.

We used MRIcro software (which can be downloaded from http://www.cabiatl.com/mricro/mricro/) and imported the image from the backup folder and then extracted the 2D MR image slices of each subject. In this paper, we only choose 32 important center slices from each subject manually based on our experience. These slices are used for preprocessing. The reason behind picking center slice from all slices is that it retains more relevant information about the brain tissues as compared to earlier slices and later slices in the group of 1–256 slices. The direction of the slice possibly may be sagittal, coronal, or axial. In this research, we chose axial direction by knowledge. The same process is applied to all the subjects (126 including both ADs and HCs). All images are in PNG format, and the dimensions of the slices are 176×208. The image is resized to 256×256 before being used for further processing.

2.4. Discrete Wavelet Transform. The discrete wavelet transform (DWT) is an image processing method [18] that gives multiscale representation of a stated signal or image [19]. Standard DWT is helpless to shift variance issue and only has horizontal and vertical directional selectivity [20]. Suppose s denotes a particular signal, n symbolizes the sampling point, h and g denote a high-pass filter and low-pass filter, respectively, and H and L depict the coefficients of high-pass and low-pass subbands. We have

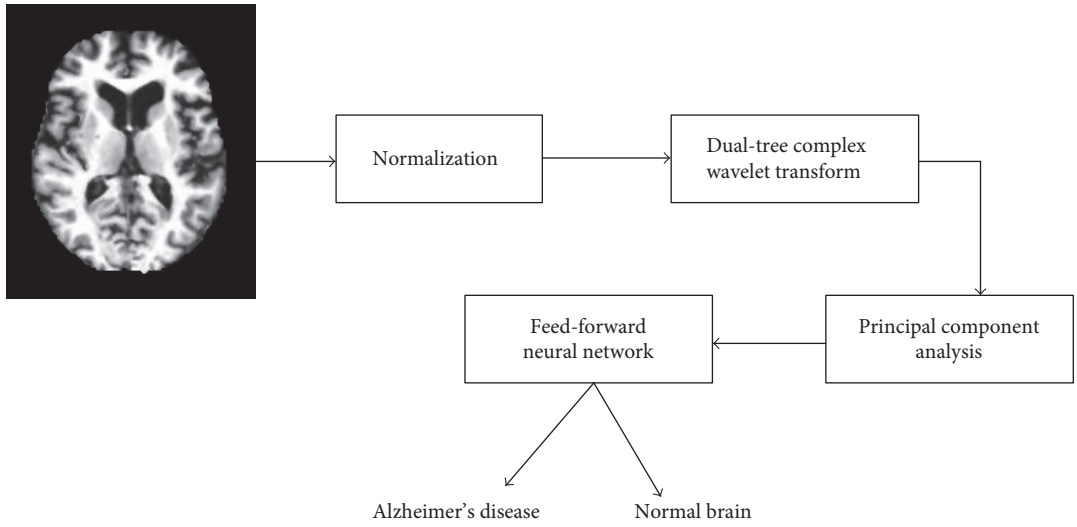


Figure 2: Block diagram of the proposed system.

$$H(n) = \sum_m h(2n-m)s(m),$$
$$L(n) = \sum_m g(2n-m)s(m). \quad (1)$$

The *LH* represents a low-pass filter along *x*-axis and high-pass filter along *y*-axis. *HL* represents a high-pass filter along *x*-axis and low-pass filter along *y*-axis. The *LL* represents low-pass filters along both directions, and *HH* represents high-pass filters along both directions.

Here, the *HL* and *LH* have clear-cut for both vertical and horizontal orientations. For the *HH*, it combines directions of both −45 and +45 degrees jointly, which stems from the utilization of real-valued filters in DWT. This combining also hinders the direction check [21].

2.5. Dual-Tree Complex Wavelet Transform. The dual-tree complex wavelet transform (DTCWT) is a modified version of the traditional DWT. To help boost the directional selectivity impaired by DWT, DTCWT is proposed. The traditional DWT is shift variant because of the decimation operation used in the transform. As a consequence, a small shift in the input signal can create a very dissimilar set of wavelet coefficients formed at the output. It utilizes two real DWTs processing input data in parallel [22]. The first DWT symbolizes the real component of the transform, whereas the second DWT depicts the imaginary component together forming a complex transform.

The DTCWT provides a solution for "shift-invariant problems" as well as for "directional selectivity in two or more dimensions," which are both shortcomings of the ordinary DWT [23]. It obtains directional selectivity by utilizing wavelets that are approximately analytic. It also has the ability to produce a total of six directionally discriminating subbands oriented in the ±15, ±45, and ±75 directions, for both the real (R) and imaginary (I) parts. Figure 3 illustrates the DTCWT. Let $h_i(n)$ and $g_i(n)$ be the filters in the first stage as in Figure 3. Let the new kth stage response of the first filter bank be $H^{(k)}_{\text{new}}(e^{jw})$ and second filter bank be $H'^{(k)}_{\text{new}}(e^{jw})$; we now have the following result as a corollary of Lemma 1.

Corollary 1. *Suppose one is provided with CQF pairs* $\{h_o(n), h_1(n)\}, \{h'_o(n), h'_1(n)\}$. *For* $k > 1$,

$$H^{(k)}_{\text{new}}\left(e^{jw}\right) = H\left\{H'^{(k)}_{\text{new}}\left(e^{jw}\right)\right\}, \quad (2)$$

if and only if

$$h_0^{'(1)}(n) = h_0^{(1)}(n-1). \quad (3)$$

A 2D image $f(x, y)$ is decomposed by 2D DTCWT over a series of dilations and translations of a complicated scaling function and six complex wavelet functions $\varphi^{\theta}_{j,l}$; that is,

$$f(x,y) = \sum_{l\epsilon Z^2} s_{j_o,l^{\phi}j_o}, l^{(x,y)} + \sum_{\theta\epsilon\Theta}\sum_{j\geq j_o}\sum_{l\epsilon z^2} c^{\theta}_j, l^{\varphi^{\theta}_j}, l^{(x,y)}, \quad (4)$$

where $\theta \in \Theta = \{\pm 15^{\circ}, \pm 45^{\circ}, \pm 75^{\circ}\}$ gives the directionality of the complex wavelet function. This is to say that the decomposition of $f(x, y)$ by utilizing the DTCWT creates one complex valued low-pass subband and six complex valued high-pass subbands at every level of decomposition, where every high-pass subband corresponds to one particular direction θ.

A study was carried out in [24] to compare the DTCWT's directional selectivity to that of the DWT. The simulation results showed that the edges detected by the DTCWT had clear contours, and nearly all directions could be detected clearly and perfectly. However, the edges detected by the DWT were discontinuous, and only horizontal and vertical edges could be successfully detected. The results verified the effectiveness of the DTCWT over the DWT. By utilizing DTCWT, we are extracting DTCWT coefficients from the preprocessed images. The additional features include the information about the demographics of the patients such as age, gender, handedness, education, SES, and clinical

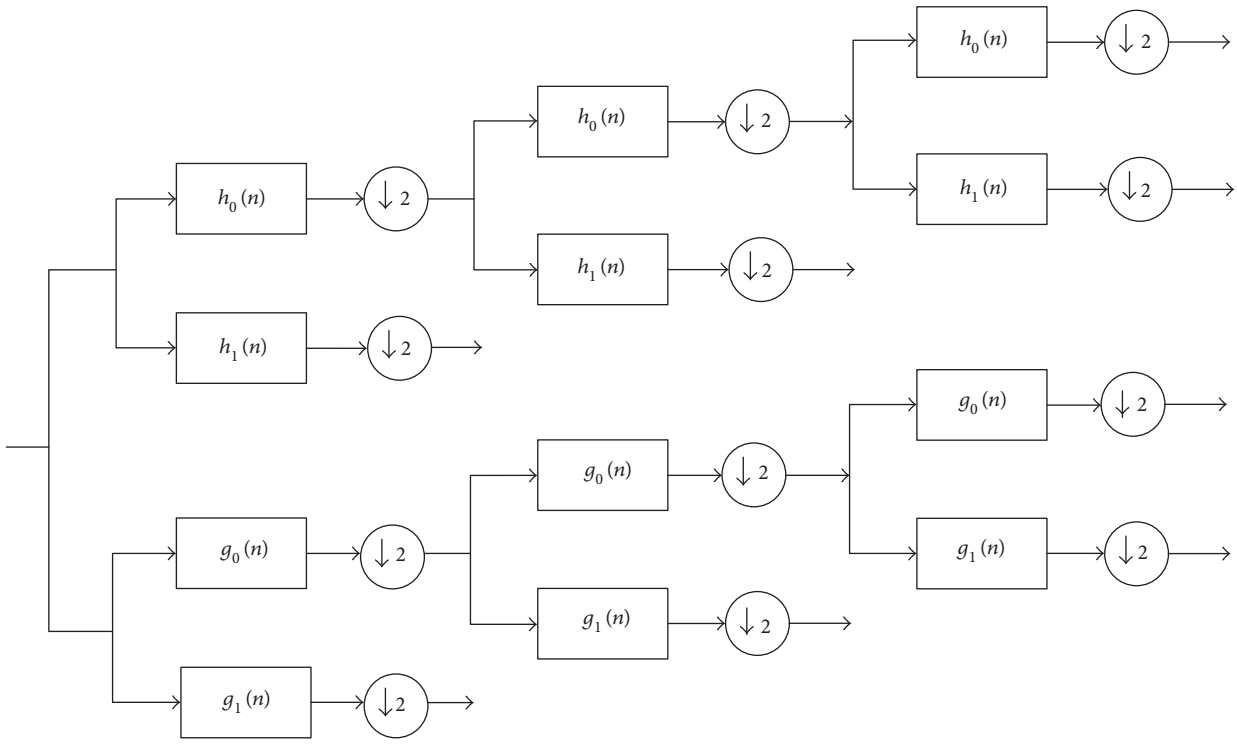


Figure 3: The DTCWT is implemented utilizing two wavelet filter banks functioning in parallel.

Let X be an input text file (X: matrix of dimensions $M \times N$)
Accomplish the following steps:

Step 1. Estimate the empirical mean: $u[m] = \frac{1}{N}\sum_{n=1}^{N} X[\mathrm{m}, \mathrm{n}]$.

Step 2. Compute the deviations from the mean and save the data in the matrix $B[\mathrm{M} \times \mathrm{N}]$: B = X − u.h, here, h is a $1 \times N$ row vector of all 1's: $h[\mathrm{n}] = 1$ for $n = 1, \ldots, N$.
Step 3. Obtain the covariance matrix $C : C = \frac{1}{N} B.B^*$.
Step 4. Get the eigenvectors and eigenvalues of the covariance matrix $V^{-1}CV = D : V$ -the eigenvectors matrix; D -the diagonal matrix of eigenvalues of C, $D[\mathrm{p}, \mathrm{q}] = \lambda_m$ for $p = q = m$ is the *mth* eigenvalues of the covariance matrix C.
Step 5. Rearrange the eigenvectors and eigenvalues: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \ldots\ldots\lambda_N$.
Step 6. Selecting components and developing a feature vector: save the first L columns or V as the $M \times L$ matrix W, $W[\mathrm{p}, \mathrm{q}] = \mathrm{V}[\mathrm{p}, \mathrm{q}]$, for $p = 1, \ldots, M$, $q = 1, \ldots, L$ where $1 \leq L \leq M$.
Step 7. Obtaining the fresh data set: The eigenvectors with the leading eigenvalues are forecasted into space, this projection appears in a vector depicted by fewer dimension ($L < M$) accommodating the essential coefficients.

Algorithm 1: PCA algorithm.

examination. The handedness features are not included in the work since all the patients are right-handedness.

2.6. Principal Component Analysis. The coefficient from the DTCWT enlarges the dimensionality of feature space that makes the classification job more complicated.

Additionally, it leads to excessive computational overhead and enormous memory storage. As a result, it is essential to lower the dimension of the feature set and get the significant features to boost the classification result. Since the last two decades, a method called PCA has earned much more attention for data visualization and reduction of dimensionality. It systematically projects the initial input data to a lower-dimensional space, well-known as principal subspace through an orthogonal transformation while preserving most of the data variations. For a stated set of likely correlated variables, these transformation outcomes in a set of values of linearly uncorrelated variables are called as principal components (PCs). All of the steps to implement PCA are demonstrated in Algorithm 1. The additional information on PCA and its implementations can be viewed in literature [25, 26].

Let us consider a set of data. PCA is employed to find a linear lower-dimensional reduction of the dataset. In this case, the variance of the constructed data is preserved. PCA limits the feature vectors to the component it selects, which

leads to an effective classification algorithm. The main idea behind implementing PCA is reduction of the dimensionality of the DTCWT coefficients, which results in more adequate and accurate classification.

The following algorithm is utilized to obtain the principal components from the input matrix and finally fed to the feed-forward neural network. Now, the input matrix possesses only these PCs. Hence, the size of the matrix is reduced. Therefore, feature extraction is done in two steps: DTCWT extracts the wavelet coefficients, and essential coefficients are later selected by the PCA as described in Algorithm 1.

2.7. Feed-Forward Neural Networks

2.7.1. Structure. Feed-forward neural networks (FNN) are broadly used in pattern classification because they do not need any information regarding the probability distribution or a priori probabilities of distinct classes. Neural networks (NN) harness power from their densely parallel structure and their ability to acquire information from experience. As a result, they can be utilized for accurate classification of input data into different classes, provided that they are pretrained. The architecture of a multilayer feed-forward neural network is shown in Figure 4.

Three factors need to be considered in designing an ANN for a specific application: (i) the topology of the network, (ii) the training algorithm, and (iii) the neuron activation function. A network may have many layers of neurons, and its complete architecture may possess either a feed-forward or a back propagation structure. A multihidden-layer back propagation NN with sigmoid neurons in its hidden layer is chosen. Similarly, linear neurons are selected for the output layer. The training vector is provided to the NN, which is instructed batch mode [27]. The NN is a two-layer network, and its configuration is $N_{\mathrm{I}} \times N_{\mathrm{H}} \times N_0$; N_{I} represents the input neurons, N_{H} is the hidden layer, and N_0 indicates that the brain under observation is either HC or AD.

2.7.2. Training Method. Mathematicians have already proven that a conjugate gradient (CG) algorithm, probing along conjugate gradient directions, produces a faster convergence than the steepest descent directions do. Among CG algorithm, the scaled conjugate gradient (SCG) method is the most powerful [28]. Thus, we utilize the SCG to train our network.

Let ω_1 and ω_2 be the connection weight matrix linking the input layer and hidden layer, and the hidden layer and the output layer, respectively. Later, we can deduce the training process reported by the following equations to improve these weighted values that can be divided into four subsequent steps [29].

(1) The calculation of the outputs of all neurons in the hidden layer is done by

$$y_j = f_{\mathrm{H}}\left(\sum_{i=1}^{N_{\mathrm{I}}} \omega_1(i,j)x_i\right) \quad j = 1, 2, \ldots, N_{\mathrm{H}}. \tag{5}$$

Here, x_i stands for the ith input value, y_j stands for the jth output of the hidden layer, and f_{H} refers to the activation function of hidden layer, commonly a sigmoid function as observed:

$$f_{\mathrm{H}}(x) = \frac{1}{1 + \exp(-x)}. \tag{6}$$

(2) The outputs of all neurons in the output layer are stated as follows:

$$O_k = f_{\mathrm{o}}\left(\sum_{j=1}^{N_{\mathrm{H}}} \omega_2(j,k)y_j\right) \quad k = 1, 2, \ldots, N_{\mathrm{o}}. \tag{7}$$

Here, f_{o} represents the activation function of output layer that is usually a line function. At first, all weights are accredited with random values and amended by the delta rule on the basis to the learning samples.

(3) The error is articulated as the MSE of the distinction among output and target value [30].

$$E_l = \mathrm{MSE}\left(\sum_{k=1}^{N_{\mathrm{o}}} O_k - T_k\right) \quad l = 1, 2, \ldots, N_{\mathrm{s}}, \tag{8}$$

where T_k depicts the kth value of the genuine labels which is already well-known to users and N_{s} denotes the number of samples [31].

(4) Let us consider that there are N_{s} samples; therefore, the fitness value can be written as

$$F(\omega) = \sum_{l=1}^{N_{\mathrm{s}}} E_l, \tag{9}$$

where ω designates the vectorization of the (ω_1, ω_2). The aim is to minimize the fitness function $F(w)$, namely, force the output values of every sample appropriate to equivalent target values.

The hidden layer or the output layer j is depicted in Figure 5. The inputs, weighted sum, and activation function of output layer are shown in Figure 5. The connection weight between the input layer and hidden layer and hidden layer and output layer is shown in Figure 6. The connection weights can also be represented in the matrix form known as connection weight matrix.

3. Experiment, Results, and Discussion

The proposed method is implemented using the 32-bit Matlab 2015b environment on Intel(R) Core (TM) i3-2120, with a processing speed of 3.30 GHz and 2 GB of RAM running Microsoft Windows 7. Readers can repeat our results on any computer with which MATLAB is compatible.

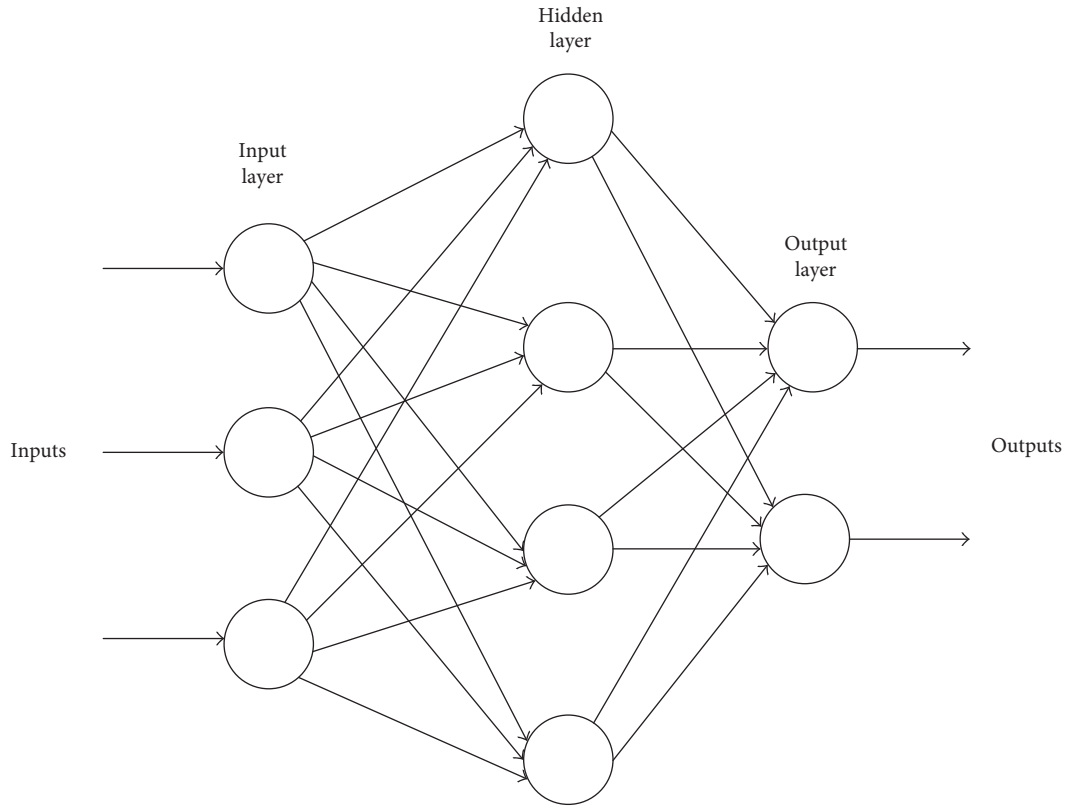


Figure 4: Architecture of a multilayer feed-forward neural network.

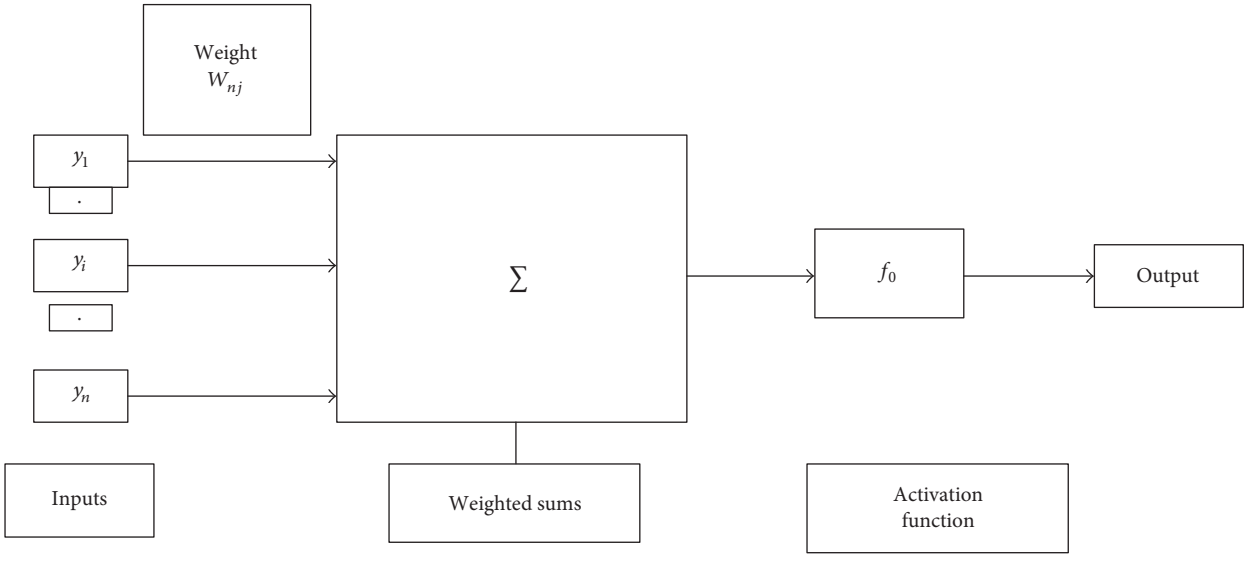


Figure 5: Hidden or output layer j. The input j are outputs from the previous layers. These are multiplied by their corresponding weights to configure a weighted sum. A nonlinear activation function is applied to the net input. [The inputs to input j are labeled $y_1, y_2, \ldots, y_n$. If unit j was in the first hidden layer, then these inputs would correspond to the input tuple $I_1, I_2, I_3, \ldots, I_n$.]

This article aims at developing a CAD of AD brain system with better performance. The pseudocode is listed in Table 4.

3.1. Parameter Estimation for s. It is always a major concern to find the optimum value of decomposition level s. We know that a smaller s provides less information whereas a larger s provides more information to the classifier. In order to avoid overfitting problem, a smaller s is used. Here, we change the value of s from 1 to 5 with increment of 1 and check up the corresponding average accuracies with FNN. The one which gives the highest accuracy is the optimal value of s.

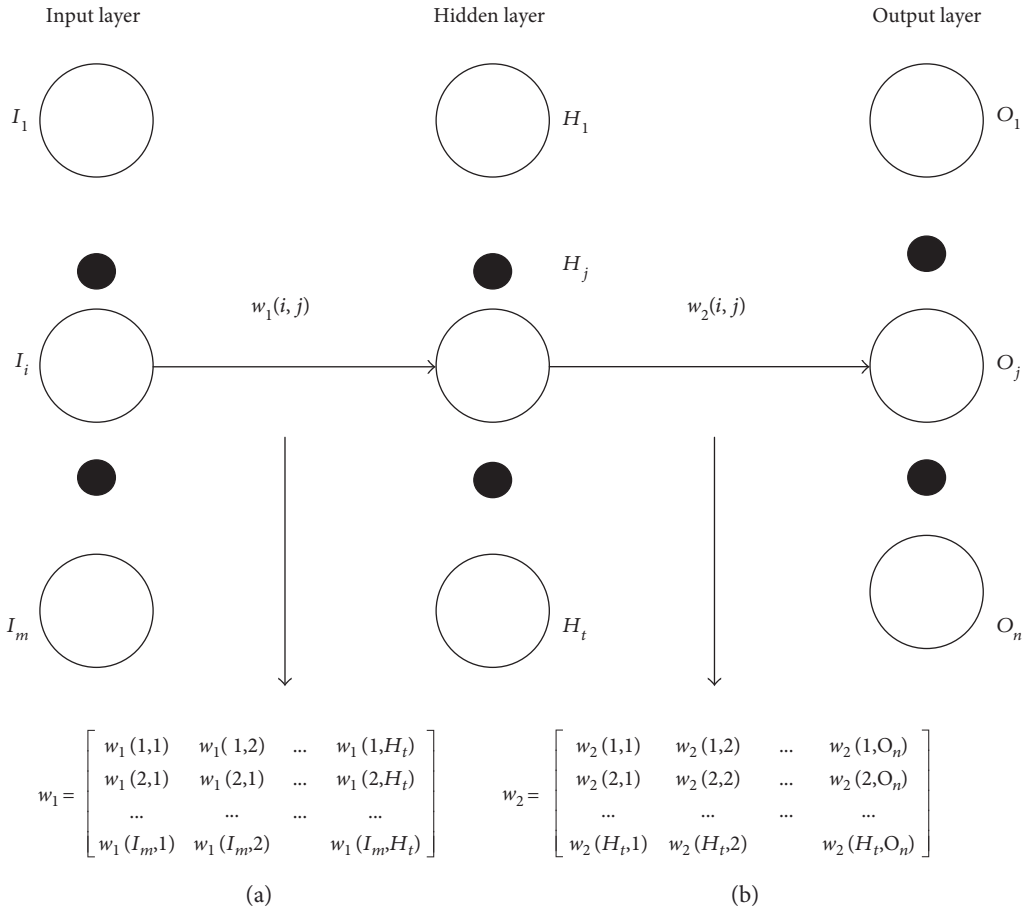


Figure 6: Connection weight matrix between (a) input layer and hidden layer and (b) hidden layer and output layer.

Table 4: Pseudocode of the proposed system.

Step 1: Import.
(a) Import the OASIS dataset.
(b) Ensure MRI as normal or abnormal brain.
Step 2: Resample the image into 256×256.
Step 3: Compute 5-level DTCWT on the preprocessed images.
Step 4: Perform PCA on the obtained matrix. The selected number of principal component (PC) should preserve at least 90% of total variances.
Step 5: Train feed-forward neural network by taking input as reduced set of feature vectors and their corresponding class labels.
Step 6: Evaluation
(a) Obtain the confusion matrix.
(b) Calculate the classification accuracy and other essential parameters.

3.2. Feature Extraction. In this paper, we extract the DTCWT coefficients from the input images. The features of 5th resolution scales are selected because they provide higher classification performance than other resolution level scales. The DTCWT has a multiresolution representation as the wavelet transform does. For disease detection, it is preferable to use a few intermediate coefficient scales as the classifier input. The lowest scales have lost fine signal details whereas the most highly detailed scales contain mostly noise. Therefore, we prefer to choose only a few intermediate scales for the DTCWT coefficients. These obtained coefficients are sent as input to the PCA.

3.3. Feature Reduction. Excessive features increase calculation time as well as memory storage. In addition, they sometimes make classification much more complicated, which is known as curse of dimensionality. In this article, we utilized PCA to decrease the number of features.

Therefore, the extracted feature from DTCWT is sent to the PCA for the feature reduction. For each image, there are 768 features after 5th level of decomposition. As we have employed 32 slices for each patient, the total number of features becomes 32×768. Now, the image is reformed into a row vector of $1 \times 24{,}576$. The row vectors of 126 subjects are arranged into an "input matrix" with dimensions of $126 \times 24{,}576$. It is still too large for calculation. So, the input data matrix is now decomposed into the principal component "score matrix" and the "coefficient matrix." The score matrix size after decomposition is 126×125. Here, the rows and columns of "score matrix" correspond to subjects and components, respectively.

The variance with the number of principal components from 1 to 18 is listed in Table 5. Experimenting with different

Table 5: Detailed data of PCA.

No. of prin. comp.	1	2	3	4	5	6	7	8	9
Variance (%)	63.18	72.15	77.08	80.28	83.05	84.55	85.68	86.59	87.47
No. of prin. comp.	10	11	12	13	14	15	16	17	18
Variance (%)	88.18	88.28	89.41	89.96	90.44	90.86	91.23	91.58	91.91

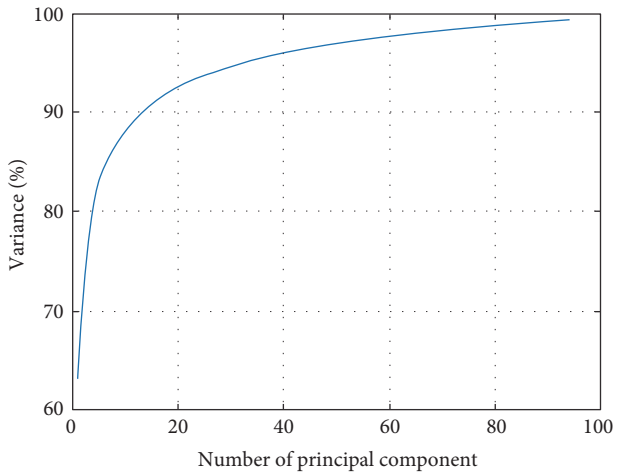


Figure 7: Variances versus number of principal component.

numbers of principal components (PCs) revealed that accuracy with PC = 14 provided the best classification accuracy preserving 90.44% of the total variance. The curve of cumulative sum of variances with the number of principal component is shown in Figure 7. We did not set the energy threshold as 95% because that would cost too many features, along with computational burden.

3.4. BPNN Training. The 14 PCs are directly sent to BPNN. Thus, the number of input neurons N_I is 14. Then, the number of hidden layer neurons (N_H) is determined as 10 according to the information entropy method [32]. Therefore, the architecture of the neural network becomes 14-10-1. The SCG method is employed because it is extremely faster than BP, MBP, and ABP [28].

3.5. Performance Measures. There are several techniques to evaluate the efficiency of classifiers. The performance is calculated on the essence of the overall confusion matrix. It holds the correct and incorrect classification results. Table 6 shows a confusion matrix for binary classification, where TP, TN, FP, and FN depict true positive, true negative, false positive, and false negative, respectively, as illustrated in Table 7.

Here, AD brains are assumed to hold the value "true" and NC ones are assumed to hold the value "false" following normal convention.

The accuracy is the most accepted empirical measure to access effectiveness of classifier. It is formulated by

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{10}$$

Table 6: Confusion matrix for a binary classifier to discriminate between two classes (A_1 and A_2).

True class	Predicted class	
	A_1 (patients)	A_2 (controls)
A_1 (patients)	TP	FN
A_2 (controls)	FP	TN

Table 7: Evaluation indicators.

Indicator	Explanation
TP	True positive, anticipating an AD to AD
FP	False positive, anticipating an HC to AD
TN	True negative, anticipating an HC to HC
FN	False negative, anticipating an AD to HC

Sensitivity is the measure of the proportion of true positives that are correctly classified, and specificity is the measure of the proportion of negatives which are correctly classified. These are calculated by

$$\begin{aligned} \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}. \end{aligned} \tag{11}$$

The precision and the recall are formulated by

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \tag{12}$$

3.6. Statistical Analysis. In order to execute a strict statistical analysis, stratified cross-validation (SCV) is used. We apply a 10-fold CV technique in this experiment because of two reasons: (1) to make balance between reliable estimate and computational cost and (2) for providing a fair comparison because the common convention was to take the value of K equal to 10 [33].

A 10-fold CV means we have to divide our dataset randomly into ten mutually exclusively folds of approximately equal size and almost the same distribution. In each run, 9 subsets will be used for training, and the remaining one will be utilized for the validation. This process is repeated 10 times, in which every subset is utilized for validation once. The 10-fold CV is repeated 50 times; namely, a 50x 10-fold CV is implemented.

The accuracies, sensitivities, and specificities obtained from the 50 runs of 10-fold CV are presented in Table 8.

Table 8: Algorithm performance comparison for MRI brain image.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)
DTCWT + PCA + FNN (proposed)	**90.06 ± 0.01**	**92.00 ± 0.04**	**87.78 ± 0.04**	**89.6 ± 0.03**
VBM + RF [38]	89.0 ± 0.7	87.9 ± 1.2	90.0 ± 1.1	N/A
DF + PCA + SVM [40]	88.27 ± 1.9	84.93 ± 1.21	89.21 ± 1.6	69.30 ± 1.91
EB + WTT + SVM + RBF [56]	86.71 ± 1.93	85.71 ± 1.91	86.99 ± 2.30	66.12 ± 4.16
BRC + IG + SVM [34]	90.00	96.88	77.78	N/A
BRC + IG + VFI [34]	78.00	65.63	100.00	N/A
Curvelet + PCA + KNN [35]	89.47	94.12	84.09	N/A
US + SVD-PCA + SVM-DT [36]	90.00	94.00	71.00	N/A

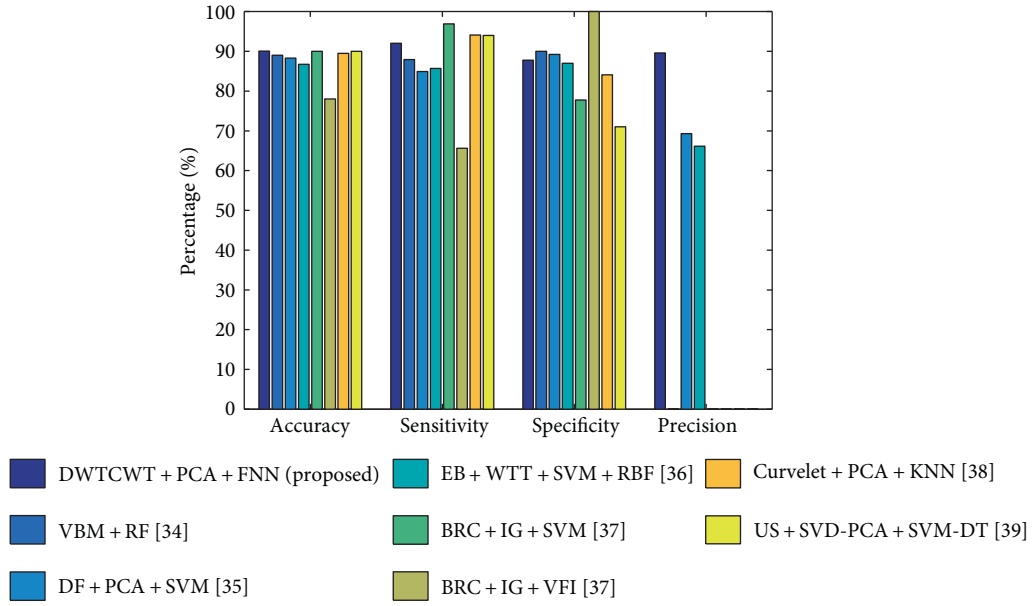


Figure 8: Bar plot of the algorithm comparison ([34–36, 38] did not mention its precision).

Our method achieved an accuracy of 90.06 ± 0.01%, a sensitivity of 92.00 ± 0.04%, a specificity of 87.78 ± 0.04%, and a precision of 89.6 ± 0.03%.

3.7. Comparison to Other State-of-the-Art Approaches. To further determine the effectiveness of the proposed "DTCWT + PCA + FNN," we compared it with seven state-of-the-art approaches in Table 8. Some of these approaches utilized different statistical settings, making direct comparison difficult. The results in Table 8 show that study [34–37] did not present standard deviations (SD) of three standards. The specificities of study [34–36] are lower than those demonstrated by other methods. Therefore, these three methods are not worthy for further study. Similarly, study [37] obtained a classification specificity of 100%. In spite of its high specificity, both the accuracy and sensitivity achieved by this algorithm are poor. Hence, this method is also not considered further for the study. Three other methods reported both mean values and standard deviation values. They also achieved satisfying results. Study [38] obtained promising results because of the voxel-based morphometry (VBM). Indeed, VBM has frequently been employed to study brain changes. Study [37] demonstrated that a taxi driver will normally have a larger back section of the posterior hippocampus. Study [39] concluded that global gray matter decreases linearly with old age but global white matter remains in the same amount. Nevertheless, VBM requires an accurate spatial normalization, or the classification accuracy may decrease significantly. Study [40] was based on a novel approach called the displacement field (DF). This study measured and estimated the displacement field of various slices between AD and HC subjects. There are other methods that have distinguished AD from HC; however, they dealt with images formed by other modalities: PET, SPECT, DTI, and so forth. Hence, they are also not considered in this study.

Finally, the proposed "DTCWT + PCA + FNN" achieved an accuracy of 90.06 ± 0.01%, a sensitivity of 92.00 ± 0.04%, a specificity of 87.78 ± 0.04%, and a precision of 89.60 ± 0.03%. With respect to classification accuracy, our approach outperforms five other methods and is almost equal to the accuracies of the remaining two methods that did not account for means and standard deviations. We also achieved a promising sensitivity and a promising specificity.

Table 9: Acronyms list.

Acronym	Definition
AD	Alzheimer's disease
HC	Healthy control
MR(I)	Magnetic resonance (imaging)
DTCWT	Dual-tree complex wavelet transform
PCA	Principal component analysis
FNN	Feed-forward neural network
DWT	Discrete wavelet transform
OASIS	Open Access Series of Imaging Studies
MMSE	Mini mental state examination
CDR	Clinical dementia rating
SNR	Signal-to-noise ratio
VBM	Voxel-based morphometry
RF	Random forest
DF	Displacement field
SVM	Support vector machine
EB	Eigenbrain
WTT	Welch's t-test
RBF	Radial basis function
BRC	Brain region cluster
IG	Information gain
KNN	k-nearest neighbor
ANN	Artificial neural network
SCV	Stratified cross-validation

Hence, our results are either better than or comparable to those of the other methods. The bar plot of the algorithm comparison is shown in Figure 8. The acronyms list is depicted in Table 9.

4. Conclusions and Future Research

We presented an automated and accurate method for AD identification based on a DTCWT, PCA, and FNN. The results showed that the proposed method achieved an accuracy of $90.06 \pm 0.01\%$, a sensitivity of $92.00 \pm 0.04\%$, a specificity of $87.78 \pm 0.04\%$, and a precision of $89.6 \pm 0.03\%$ and outperformed 7 state-of-the-art algorithms.

We will focus our future research on the following aspects: (i) testing other advanced variants of wavelet such as 3D-DTCWT, wavelet packet analysis, and fractional calculus; (ii) utilizing different feature reduction techniques such as independent component analysis (ICA) [41], linear discriminant analysis (LDA) [42], probabilistic PCA [43], or sparse-autoencoders [44]; (iii) testing our data with least-square techniques [45], kernel support vector machine (k-SVM), such as fuzzy SVM [46], radial basis function neural network (RBFNN) [47], deep learning methods such as convolutional neural network (CNN) [48], and other alternative pattern recognition tool for classification; (iv) utilizing advanced swarm intelligence techniques such as particle swarm optimization [49], artificial bee colony [50], genetic pattern search [51], ant colony optimization [52], and biogeography-based optimization [53] to find the optimal kernel; (v) testing the proposed method on images obtained from different modalities such as computed tomography (CT) [54], ultrasound, spectrum imaging [55], and 3D MRI; (vi) utilizing other advance image preprocessing technique to enhance the classification performance, such as image denoising, image enhancement, and image segmentation; and (vii) classification may be carried out on the sparsity domain.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C7A1046050). And this study was financed by the research fund of Chosun University, 2017.

References

[1] R. Brookmeyera, E. Johnsona, K. Ziegler-Grahamb, and H. M. Arrighic, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[2] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, "Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification," *Neurobiology of Aging*, vol. 32, no. 12, pp. 2322.e19–2322.e27, 2011.

[3] A. Nordberg, J. O. Rinne, A. Kadir, and B. Langstrom, "The use of PET in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 78–87, 2010.

[4] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4637–4642, 2004.

[5] T. Magnander, E. Wikberg, J. Svensson et al., "A novel statistical analysis method to improve the detection of hepatic foci of 111 In-octreotide in SPECT/CT imaging," *EJNMMI Physics*, vol. 3, no. 1, p. 1, 2016.

[6] P. De Visschere, M. Nezzo, E. Pattyn, V. Fonteyne, C. Van Praet, and G. Villeirs, "Prostate magnetic resonance spectroscopic imaging at 1.5 tesla with endorectal coil versus 3.0 tesla without endorectal coil: comparison of spectral quality," *Clinical Imaging*, vol. 39, no. 4, pp. 636–641, 2015.

[7] E. D. Roberson and L. Mucke, "100 years and counting: prospects for defeating Alzheimer's disease," *Science*, vol. 314, no. 5800, pp. 781–784, 2006.

[8] M. Tabaton, P. Odetti, S. Cammarata et al., "Artificial neural networks identify the predictive values of risk factors on the conversion of amnestic mild cognitive Impairment," *Journal of Alzheimer's Disease*, vol. 19, no. 3, pp. 1035–1040, 2010.

[9] B. S. Mahanand, S. Suresh, N. Sundararajan, and K. M. Aswatha, "Identification of brain regions responsible for Alzheimer's disease using self-adaptive resource allocation network," *Neural Networks*, vol. 32, pp. 313–322, 2012.

[10] B. Jeurissen, A. Leemans, and J. SIjbers, "Automated correction of improperly rotated diffusion gradient orientations in diffusion weighted MRI," *Medical Image Analysis*, vol. 18, no. 7, pp. 953–962, 2014.

[11] L. Hamelin, M. Bertoux, M. Bottlaender et al., "Sulcal morphology as a new imaging marker for the diagnosis of early onset Alzheimer's disease," *Neurobiology of Aging*, vol. 36, no. 11, pp. 2932–2939, 2015.

[12] M. Á. Caballero, M. Brendel, A. Delker et al., "Mapping 3-year changes in gray matter and metabolism in Abeta-positive non-demented subjects," *Neurobiology of Aging*, vol. 36, no. 11, pp. 2913–2924, 2015.

[13] G. Yang, Y. Zhang, J. Yang et al., "Automated classification of brain images using wavelet energy and biogeography-based optimization," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 15601–15617, 2016.

[14] E. S. El-Dahshan, H. M. Mohsen, K. Revett, and A. B. Salem, "Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.

[15] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "An MR images classifier system via particle swarm optimization and kernel support vector machine," *The Scientific World Journal*, vol. 2013, Article ID 130134, 9 pages, 2013.

[16] I. Álvarez, J. M. Górriz, J. Ramírez et al., "Alzheimer's diagnosis using eigenbrains and support vector machines," *Electronics Letters*, vol. 45, no. 7, pp. 973–980, 2009.

[17] M. Z. Iqbal, A. Ghafoor, A. M. Siddhiqui, M. M. Riaz, and U. Khalid, "Dual-tree complex wavelet transform and SVD based medical image resolution enhancement," *Signal Processing*, vol. 105, pp. 430–437, 2014.

[18] Y. Zhang, B. Peng, S. Wang et al., "Image processing methods to elucidate spatial characteristics of retinal microglia after optic nerve transection," *Scientific Reports*, vol. 6, article 21816, 2016.

[19] D. K. Shin and Y. S. Moon, "Super-resolution image reconstruction using wavelet based patch and discrete wavelet transform," *Journal of Signal Processing Systems*, vol. 81, no. 1, pp. 71–81, 2015.

[20] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging," *IEEJ Transcations on Electrical and Electronic Engineering*, vol. 10, no. 1, pp. 116-117, 2015.

[21] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurance matrix for detection of breast cancer," *Neurocomputing*, vol. 154, pp. 1–14, 2015.

[22] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–353, 2001.

[23] A. Barri, A. Dooms, and P. Schelkens, "The near shift-invariance of the dual-tree complex wavelet transform revisited," *Journal of Mathematical Analysis and Applications*, vol. 389, no. 2, pp. 1303–1314, 2012.

[24] S. Wang, S. Lu, Z. Dong, J. Yang, M. Yang, and Y. Zhang, "Dual-tree wavelet transform and twin support vector machine for pathological brain detection," *Applied Science*, vol. 6, no. 6, p. 169, 2016.

[25] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.

[27] D. Guo, Y. Zhang, Q. Xiang, and Z. Li, "Improved radio frequency identification method via radial basis function neural network," *Mathematical Problems in Engineering*, vol. 2014, Article ID 420482, 9 pages, 2014.

[28] Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10049–10053, 2011.

[29] M. Manoochehri and F. Kolahan, "Integration of artificial neural network and simulated annealing algorithm to optimize deep drawing process," *The International Journal of Advaned Manufacturing Technology*, vol. 73, no. 1, pp. 241–249, 2014.

[30] S. U. Aswathy, G. G. D. Dhas, and S. S. Kumar, "A survey on detection of brain tumor from MRI brain images," in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pp. 871–877, Kanyakumari, 2014.

[31] A. Poursamad, "Adaptive feedback linearization control of anticlock braking systems using neural networks," *Mechatronics*, vol. 19, no. 5, pp. 767–773, 2009.

[32] H. C. Yuan, F. L. Xiong, and X. Y. Huai, "A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy," *Computers and Electronics Agriculture*, vol. 40, no. 1–3, pp. 57–64, 2003.

[33] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 21–31, 2014.

[34] C. Plant, S. J. Teipel, A. Oswald et al., "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease," *NeuroImage*, vol. 50, no. 1, pp. 162–174, 2010.

[35] D. Jha and G. R. Kwon, "Alzheimer disease detection in MRI using curvelet transform with KNN," *The Journal of Korean Institute of Information Technology*, vol. 14, no. 8, 2016.

[36] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree," *Progress in Electromagnetic Research*, vol. 144, pp. 171–184, 2014.

[37] E. A. Maguire, D. G. Gadian, I. S. Johnsrude et al., "Navigation-related structural change in the hippocampi of taxi drivers," *Proceedings of the National Academy of Sciences*, vol. 97, no. 8, pp. 4398–4403, 2000.

[38] K. R. Gray, P. Alijabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, 2013.

[39] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains," *NeuroImage*, vol. 14, no. 1, Part 1, pp. 21–36, 2001.

[40] Y. Zhang and S. Wang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. 3, article e1251, 2015.

[41] S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 20, pp. 10979–10984, 1997.

[42] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*, pp. 237–280, Springer, New York, 2013.

[43] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[44] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[45] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 482–487, 2008.

[46] S. Wang, X. Yang, Y. Zhang, P. Phillips, J. Yang, and T.-F. Yuan, "Identification of green, oolong and black teas in china via wavelet packet entropy and fuzzy support vector machine," *Entropy*, vol. 17, no. 10, pp. 6663–6682, 2015.

[47] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 697–710, 2002.

[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[49] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22–31, 2014.

[50] Y. Zhang, L. Wu, and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm," *Progress in Electromagnetics Research*, vol. 116, pp. 65–79, 2011.

[51] Y. Zhang, S. Wang, G. Ji, and Z. Dong, "Genetic pattern search and its application to brain image classification," *Mathematical Problems in Engineering*, vol. 2013, Article ID 580876, 8 pages, 2013.

[52] C. Pang, G. Jiang, S. Wang et al., "Gene order computation using Alzheimer's DNA microarray gene expression data and the ant colony optimisation algorithm," *International Journal of Data Mining and Bioinformatics*, vol. 6, no. 6, pp. 617–632, 2012.

[53] S. Wang, Y. Zhang, G.-L. Ji, J.-G. Wu, and L. Wei, "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic ABC and biogeography-based optimization," *Entrophy*, vol. 17, no. 8, pp. 5711–5728, 2015.

[54] Y. Chen, Y. Zhang, J. Yang et al., "Curve-like structure extraction using minimal path propagation with backtracking," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 988–1003, 2016.

[55] Z. Dong, Y. Zhang, F. Liu, Y. Duan, A. Kangarlu, and B. S. Peterson, "Improving the spectral resolution and spectral fitting of H MRSI data from human calf muscle by the SPREAD technique," *NMR in Biomedicine*, vol. 27, no. 11, pp. 1325–1332, 2014.

[56] Y. Zhang, Z. Dong, P. Phillips et al., "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontlier in Computational Neuroscience*, vol. 9, p. 66, 2015.

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 7862672, 9 pages
https://doi.org/10.1155/2017/7862672

Research Article

Computer-Aided Clinical Trial Recruitment Based on Domain-Specific Language Translation: A Case Study of Retinopathy of Prematurity

Yinsheng Zhang,[1] Guoming Zhang,[2] and Qian Shang[3]

[1]*School of Management and E-Business, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China*
[2]*Pediatric Retinal Surgery Department, Shenzhen Eye Hospital, Shenzhen Key Ophthalmic Laboratory, The Second Affiliated Hospital of Jinan University, Shenzhen, Guangzhou 518040, China*
[3]*Management School, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China*

Correspondence should be addressed to Yinsheng Zhang; zhangys@zjgsu.edu.cn, Guoming Zhang; zhang-guoming@163.com and Qian Shang; shangqian_xwz@163.com

Received 10 November 2016; Accepted 5 February 2017; Published 5 April 2017

Academic Editor: Xiang Li

Reusing the data from healthcare information systems can effectively facilitate clinical trials (CTs). How to select candidate patients eligible for CT recruitment criteria is a central task. Related work either depends on DBA (database administrator) to convert the recruitment criteria to native SQL queries or involves the data mapping between a standard ontology/information model and individual data source schema. This paper proposes an alternative computer-aided CT recruitment paradigm, based on syntax translation between different DSLs (domain-specific languages). In this paradigm, the CT recruitment criteria are first formally represented as production rules. The referenced rule variables are all from the underlying database schema. Then the production rule is translated to an intermediate query-oriented DSL (e.g., LINQ). Finally, the intermediate DSL is directly mapped to native database queries (e.g., SQL) automated by ORM (object-relational mapping).

1. Introduction

Clinical trials (CTs) are the building blocks for evidence-based medicine (EBM). Clinical trials are typically performed under the guidance of clinical trial protocols (CTPs), which are standardized guidelines for conducting CT. A CTP contains several components, including purpose, study design, recruitment criteria of subjects, treatment of subjects, assessment of efficacy, assessment of safety, adverse events, quality control and assurance, and ethics. Among them, "recruitment criteria" (or "eligibility criteria") is an essential step for conducting CTs.

The recruitment criteria specify a set of common features that define the population subset of interest. These features include age, gender, habit, diagnosis, stage of disease development, surgery history, and genetic data. The recruitment criteria usually contain both "inclusion" and "exclusion" rules, which define required and unwanted features, respectively. For certain CTs, the criteria could be quite complex, making the patient enrollment very challenging. With the rapid development of healthcare IT (HIT), electronic patient data acquired during the clinical care process offers a new approach to facilitate the recruitment procedure.

Much work has been reported concerning the previously mentioned approach. The traditional way of this approach usually relies on DBA (database administrator) or clinical engineers. The criteria are initially written in natural language form by CT protocol authors or clinical researchers. Then, the engineers translate the narrative criteria into native database query language, such as SQL. Due to the professional barrier and possible ambiguity of natural language, such criteria translation can be inaccurate. Secondly, the involvement of engineers has increased both human resource and communication cost. For these reasons, some researchers chose to develop computer-aided tools or

clinical trial recruitment support systems (CTRSS) to facilitate the recruitment procedure. The following manuscript will introduce some of the related work.

2. Related Work

European researchers built the TRANSFoRm Query Workbench Tool [1] to author, store, and execute clinical data queries to identify potential subjects for clinical studies. This tool can create recruitment criteria in computable representations. Then the criteria are translated into executable queries in institution-specific databases. The Query Workbench Tool uses the Clinical Data Integration Model (CDIM) [2] as an intermediate standard ontology, so it can support queries across multiple heterogeneous data sources. The UK CancerGrid [3] project designs a widely accepted clinical trial model by controlled vocabulary and common data elements (CDEs). Based on this model, a cancer data query system has been developed which supports data sharing across CancerGrid-compliant clinical trial boundaries. Penberthy et al. [4] designed a CT matching system. The system allows users to define the recruitment criteria through a form-based GUI tool. Then the system will perform periodical automatic screening against patients in the information system. Matched subjects will be sent to researchers via emails. The BreastCancerTrials.org [5] project creates a CT matching website whose targeting audience is patients. It uses self-reported patient data collected by web-based forms to match against existing registered cancer CTs and recommend patients with available CTs. Patients can also redirect themselves to participating research sites. BreastCancerTrials.org can increase patients' awareness to participate cancer CTs, and can improve patient data usage among multiple CT research groups. There are also several experimental researches that use more sophisticated methods, such as natural language processing (NLP) [6] and semantic web [7], to facilitate the automatic extraction and comparison of recruitment criteria and patient data.

Generally speaking, much research work focuses on querying across heterogeneous sources, and uses form-alike structured data entry (SDE) techniques to author recruitment criteria. These SDE tools are usually designed according to some intermediate information model (or ontology). Based on this model, the recruitment criteria and referenced data elements can be defined in formal representations, which have the unique advantage in data exchange and semantic interoperability. However, this paradigm requires mapping between the intermediate model and individual databases when performing patient queries. Such mapping work can be quite knowledge-intensive and time-consuming, and due to information granularity and semantic differences, the mapping of certain data elements can be extremely difficult or even unsupported.

With such concern, this paper tries an alternative method that uses direct syntax translation to avoid model mapping. The following manuscript will introduce the method and a corresponding case study in detail.

3. Method

3.1. Knowledge Representation of Clinical Trial Recruitment Criteria. The first step of this study is the knowledge representation of recruitment criteria and relevant patient data. An analysis on knowledge representation will help to determine whether a certain kind of information model or formal language has the ability to represent all related entities (e.g., concepts and rules) in this domain. Following are two typical recruitment criteria used in clinical trials. These two examples are related to retinopathy of prematurity (ROP), which is the target disease in our case study.

Example 1. Consider the following:

[Aim of CT]

Assess the anti-neovascularization activity of intravitreal bevacizumab, as determined by regression of neovascular vessels of retinopathy of prematurity (ROP), in neonates with acute stage 3 ROP in zone I or posterior zone II with plus disease.

[Recruitment criteria]

Birth weight ≤ 1500 grams

Gestational age ≤ 30 weeks

Diagnosis = stage 3 ROP in zone I or posterior zone II

Drug use = bevacizumab (Avastin®)

Without congenital systemic anomaly

Without congenital ocular abnormality.

Example 2. Consider the following:

[Aim of CT]

Assess Pan-VEGF (vascular endothelial growth factor) blockade for the treatment of retinopathy of prematurity.

[Recruitment criteria]

Gestational age ≤ 30 weeks

Gestational age ≤ 36 1/7 weeks

Diagnosis = type 1 pre-threshold ROP

No prior treatment

Without media opacity precluding fundus visualization

Without ocular or periocular infection(s).

As seen from the examples, typical CT recruitment criteria consist of a set of medical concepts (e.g., birth weight, gestational age, and diagnosis) and a rule condition composed of these concepts (e.g., Birth weight ≤ 1500 grams && Gestational age ≤ 30 weeks && (Diagnosis = stage 3 ROP in zone I || Diagnosis = posterior zone II)). In this study, we chose the production rule as the knowledge representation for CT recruitment criteria. In runtime, the data sources should feed the rule with actual patient data for succeeding query or reasoning.

3.2. Patient Query for CT Recruitment by Syntax Translation. Clinical data can be conveyed in two styles. One is natural language (NL), which is extensively used by humans in daily life. The other is formal language (FL) or domain-specific language (DSL), which is usually designed for a specific purpose and contains a limited set of symbols and syntax. Typical FLs/DSLs include the Arabic numbers and arithmetic

symbols used in mathematics, the molecular formula in chemistry, and programming languages. The data carried by NL are usually "unstructured" while the data conveyed by FL/DSL are mostly "structured." Though FL/DSL may lack expressiveness than NL, they are much more suitable for computers to parse and process.

The essence of patient query in CT recruitment is the translation of NL (recruitment criteria from CTP or clinical researchers) to FL/DSL (e.g., SQL). Although there is related work that uses NLP techniques to facilitate the translation of NL to FL/DSL, there is a gap between research and real application. NLP (especially for Chinese NL) is still premature and many studies use the more reliable SDE instead. In these studies, the recruitment criteria are first authored in certain formal representation by SDE. Next, the formally represented criteria need to be converted to database-specific queries (e.g., SQL). As mentioned before, many studies depend on a sharable model or ontology and have to handle the mapping between the sharable model and individual data sources. Such a mapping work can be a quite complicated, knowledge-intensive, and time-consuming task. As an alternative, this paper proposed a method based on DSL syntax translation, other than ontology mapping.

Our method directly uses the underlying database's schema as a reference model for the SDE tool and its generated production rules. That is, all rule variables in the SDE tool come from data fields in the database schema. For example, the SDE tool outputs a production rule "[Birth Weight] < 1.0 && [GestationalAgeInWeeks] < 30". In this rule, the "[BirthWeight]" and "[GestationalAgeInWeeks]" rule variables have corresponding "BirthWeight" and "Gestational AgeInWeeks" data fields in the underlying database schema. The next step is to convert such production rules to query-oriented DSL. The query-oriented DSL is provided in many modern programming languages, such as Java, C#, and VB.Net. Take LINQ (language-integrated query) as an example. In Microsoft .Net Framework, LINQ is a language subset of C# and VB.Net. LINQ supports nearly 40 operators, including "select," "from," "in," "where," and "order by." With LINQ, programmers can directly write SQL-style C# or VB.Net codes to query underlying data providers, such as ORM (object-relational mapping), ODBC (open database connect), or XML (extensible markup language) files. Because our case study is conducted in the context of Microsoft.Net framework and the SQL server relational database, we will use LINQ for demonstration purposes. For other technical platforms, there are also equivalent technologies, such as JINQ, Linq4j, and JaQue, for the Java platform. After the production rule is converted to LINQ, LINQ to native-SQL conversion is automatically supported by the ORM provider, and no extra effort is required.

In the previously mentioned process, the production rule, LINQ, and SQL are all strictly constrained FLs/DSLs, so the syntactic conversion between them is explicit and unambiguous. In addition, these DSLs all use the underlying database schema as the reference model, so there is no need for a complex concept mapping. Such concept mapping is a well-known issue in medical informatics, aka the "curly braces" problem [8], which involves mapping external clinical data to rule variables in the rule expression.

Figure 1 illustrates the comparison of different paradigms. Paradigm A depends on DBA to translate clinicians' narrative recruitment criteria to native database queries. As mentioned before, such manual translation can sometimes be inaccurate due to professional barrier and the ambiguous nature of NL. Paradigm B uses SDE to author formally represented criteria based on a sharable model. The formally represented criteria are then translated to SQL by underlying mappings between the sharable model and individual database. Paradigm C, which is proposed in this paper, also uses SDE to author recruitment criteria in the form of the production rule, but does not involve any data mapping jobs. In the following manuscript, we will introduce a case study to demonstrate our method and give more details on system implementation.

4. Case Study and System Implementation

4.1. Clinical Settings. The case study is conducted in Shenzhen Eye Hospital, which is a 200-bed class III specialized hospital in China. Since 2013, we have been developing a ROP (retinopathy of prematurity) management system for the pediatric retinal surgery department. ROP, aka Terry syndrome, is a common eye disease for prematurely born babies, especially those with low birth weight and early gestation age. ROP is related with disorganized growth of retinal blood vessels resulting in retinal scarring or detachment. Without early screening and timely intervention, ROP can lead to severe visual impairments or even blindness. ROP has become a major reason for children blindness. It is estimated that China has at least 0.2 million ROP newborns each year. ROP has now become a nonneglectable problem in China, whereas ROP-oriented information systems are scarce on the Chinese market. Under this circumstance, we built this ROP management system to help clinicians to manage ROP screening information and track patient disease development.

The ROP management system is also designed as a regional telemedicine system. It covers not only the ROP screening cases in Shenzhen Eye Hospital, but also those from partner hospitals, such as Shenzhen Dapeng-District Maternity Hospital, Meizhou People's Hospital (Guangdong Province, China), and Puning People's Hospital (Fujian Province, China). These partner hospitals can upload ROP-related data (e.g., RetCam imaging, newborn info) to central data repository via the system. Then the ROP medical team in Shenzhen Eye Hospital will help to diagnose these cases. Until now, the system has enrolled more than 22,238 patient cases. The demo version of the system is http://ropd.brahma.pub.

With the advent of "big data era," how to effectively use the collected data has become a great concern for both clinicians and researchers. Supporting clinical trial (CT) is one of the typical applications of clinical data.

4.2. Computer-Aided CT Recruitment. Based on Paradigm C proposed in Figure 1, we developed a computer-aided CT

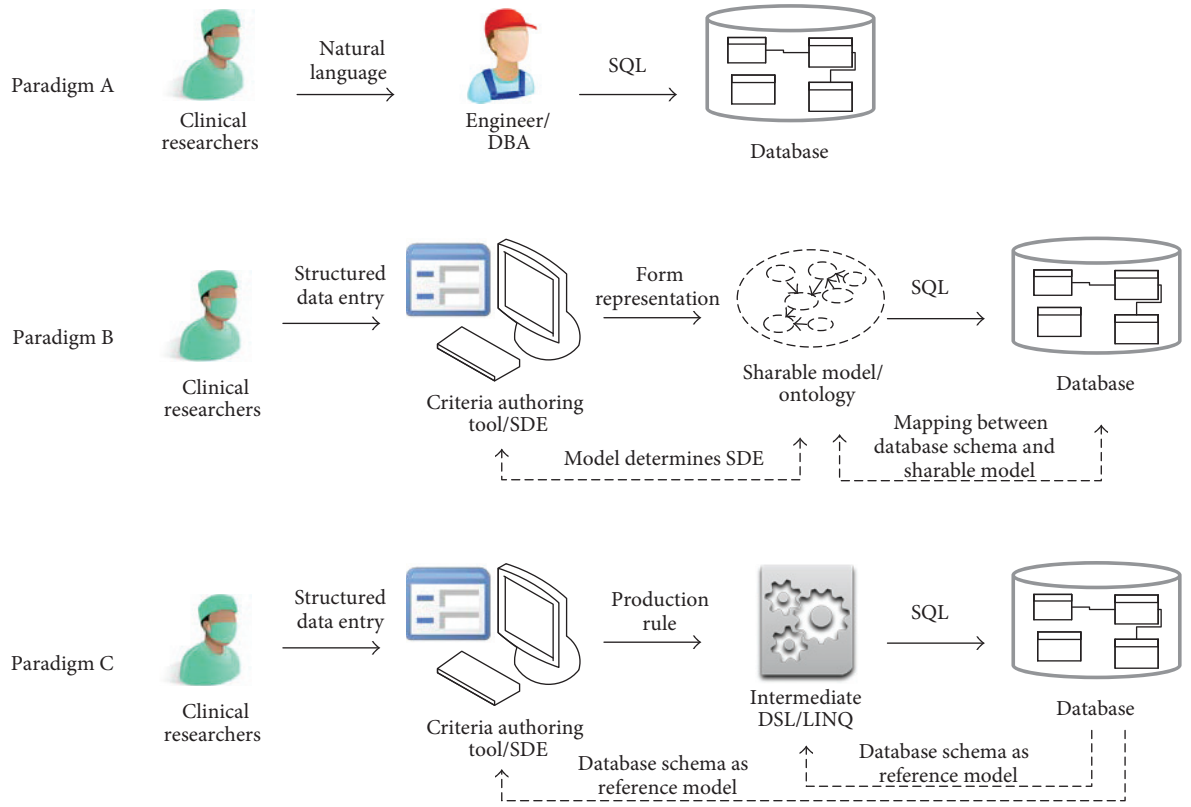


Figure 1: Comparison of the patient query paradigms used in clinical trial recruitment (demonstrated in the case of relational database). Paradigm A depends on clinical engineers or DBA to translate clinical researchers' natural language representation of recruitment criteria into database-specific query language, for example, SQL. Paradigm B uses SDE tools to author formally represented criteria based on a sharable model. The criteria are then translated to SQL by the underlying mapping between the sharable model and individual database. Paradigm C uses SDE to author recruitment criteria in the form of the production rule, which is then translated into LINQ syntax. LINQ to SQL conversion is naturally supported by ORM-like technologies. DBA = database administrator; SDE = structured data entry; DSL = domain-specific language; LINQ = language-integrated query; SQL = structured query language; ORM = object-relational mapping.

recruitment module as a subsystem in the ROP management system. The kernel idea of this system is to reuse the patient data in the central data repository for clinical research purposes. As mentioned before, the CT recruitment criteria vary from study to study. To fulfill different requirements, we developed a versatile SDE tool (Figure 2) by which clinicians can author customized search patterns. The output of the SDE tool are production rules, as the formal representation of CT recruitment criteria.

After the production rule is generated, it needs to be converted to native queries towards the underlying data repository. Figure 3 gives a concrete example of this conversion process. In Figure 3, the production rule is first translated to LINQ lambda expressions (the anonymous function inside the "Where" clause). Then, the LINQ is automatically translated to a corresponding native SQL query by ORM. It should be noted that both the LINQ and SQL code snippets in Figure 3 use only one entity set/data table: "IntegratedView." In reality, patient-related data are often dispersed among multiple data tables. For example, in our ROP management system, there are "patient master table," "patient visit table," and "patient surgery table." Sometimes, users have to make joint queries between multiple tables. In order to reduce complexity, we created a database view "IntegratedView" by joining existing tables. Although "IntegratedView" is a "virtual table," it is a fully functioning searchable object just as physical tables are, and it is equally supported by ORM. Such a view not only integrates data from multiple tables, but also reduces complexity by providing a unified logical schema. This schema is referenced by all modules across the systems, including the following: (1) the data fields (Figure 2) used to compose production rules; (2) the entity set ("domainDBContext.IntegratedView") and entity properties (e.g., "x.Surgery," "x.BirthWeight") in LINQ; and (3) and the final SQL code transformed by ORM.

Based on the DSL-conversion process, the original CT recruitment criteria represented in production rule has been ultimately converted to a native SQL statement on the underlying database. The returned patients will be shown as an html table. Users can export the search result into Excel or csv (comma-separated value) files. They can also check each individual patient, and assign a specific "CT research tag" to this patient.

4.3. Evaluation. The CT recruitment system was first brought on line in August 2015. Until now (November 2016), system logs show there have been 230 user queries. Although we have not yet conducted a large-scale scientific evaluation on

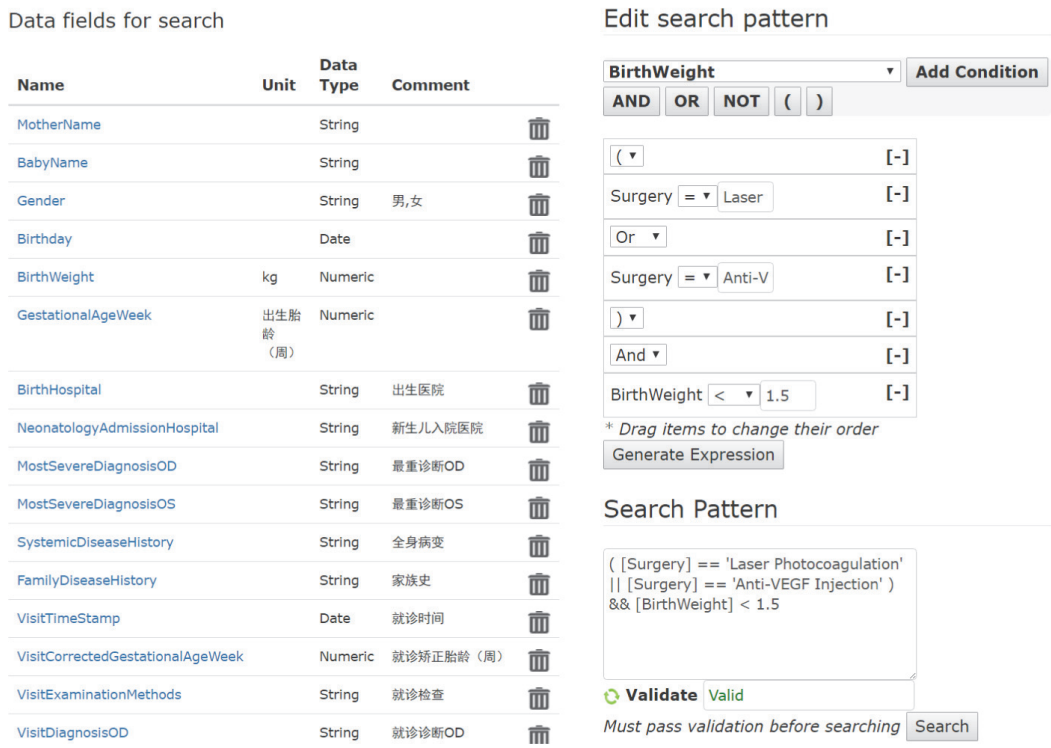


Figure 2: SDE for authoring CT recruitment criteria. The left panel is a list of data fields that can be used as rule variables in the search patterns. These data fields come from the database schema of the central data repository. The right panel is a graphical rule editor. It supports conditional predicates (composed of rule variables and comparison operators, e.g., [BirthWeight] <= 1.5), logical operators (AND, OR, and NOT), and parentheses (specify precedence). For each rule variable, the available comparison operators are strictly confined based on its data type. If the variable is numeric or date type, allowed comparison operators will be ">=, <=, >, <, =, !=". If the variable is string type, the allowed comparison operator will be "=," and the rhs (right-hand side) constant can be a string literal with wild cards ("%"). The sequence of rule components can be adjusted by drag and drop. Furthermore, the graphical editor can generate a rule expression into the textual editor. The textual editor also allows the user to directly edit rule expression with autocomplete (IntelliSense) utilities.

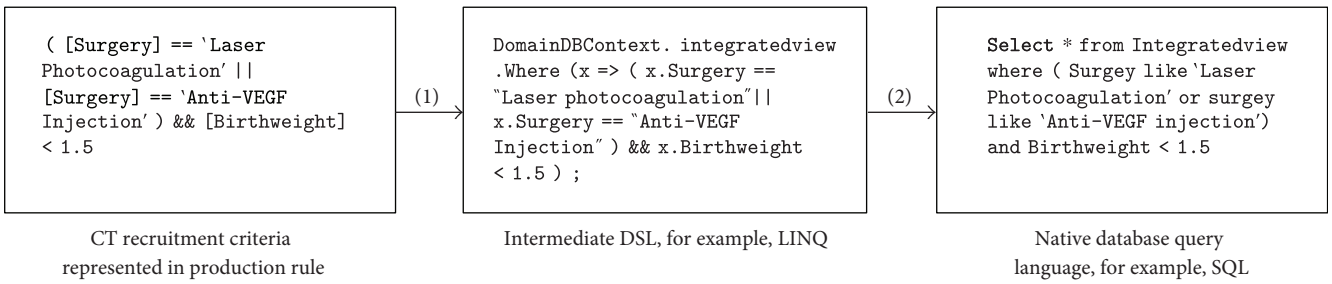


Figure 3: Example of converting production rule to native database query. (1) First, the production rule is translated to an intermediate DSL, such as LINQ. (2) Then, the LINQ DSL is translated to native database query automated by the ORM module.

the system, we have collected plenty of feedbacks from both end users and clinical engineers. Based on these feedbacks, several aspects of the system can be evaluated.

(1) Implementation Cost. As mentioned before, the criteria authoring SDE tools used in many related works are based on a shared information model or ontology, and the mapping of the shared model to native database queries can be quite complicated and time consuming. The method proposed in this paper does not use such an intermediate model, so the mapping effort is saved and the implementation cost is relatively low. In addition, this method uses LINQ as an intermediate DSL. Because both LINQ and the recruitment criteria production rules are strictly constrained DSLs, the conversion between them is unambiguous and easy to achieve. What is more, LINQ can use ORM as the data provider, so the conversion between LINQ and SQL can be automated by ORM.

(2) DSL Expression Power. In this study, the SDE tool uses production rule to author CT recruitment criteria. The expression power of the DSL determines how well the criteria

Table 1: Language primitives of the production rule DSL.

Primitives	Type	Description
&&	Logical operator	Logical AND
||	Logical operator	Logical OR
!	Logical operator	Logical NOT
>	Comparison operator	Greater than
>=	Comparison operator	Greater than or equal
<	Comparison operator	Less than
<=	Comparison operator	Less than or equal
==	Comparison operator	Equal
!=	Comparison operator	Not equal
(	Operator precedence	Left parenthesis
)	Operator precedence	Right parenthesis
[	Variable identifier	Text between "[" and "]" is interpreted as a variable name
]	Variable identifier	
Boolean	Data type	For example, true/false
Numeric	Data type	For example, 1.0
Date	Data type	For example, "2016-12-25"
String	Data type	For example, "Stage 3 ROP in zone I"
%	Wild card	Wild card for string comparison

can be represented and processed by the information system. Table 1 shows the operators and data types used in the production rule DSL. Although the DSL only supports a limited set of operators (e.g., it does not support assignment "=" and arithmetic operators "+, -, ∗, /"), it has satisfied the CT recruitment requirements in 230 query tasks since its first deployment. The expression power of this DSL lies in two aspects: (1) Complex rules can be composed by the combination of fundamental primitives. (2) The wild card supported in this system effectively handles the free-text data fields. In the production rule DSL, "%" represents any text and can be used to compose complicated literal patterns. This wild card is natively supported by the underlying SQL engine.

(3) Performance. A quantitative evaluation of the system performance is carried out on a cloud-computing virtual machine (CPU: Dual Core 2.39 GHz, RAM: 4 GB, OS: Windows Server 2012 R2 Datacenter 64-bit Edition, DB: SQL Server 2012 Express). The total consumed time of a CT recruitment query task contains 4 parts: (1) Rule validation. The system first checks whether the rule expression can be parsed as a well-structured tree (using the ANTLR library). (2) DSL translation: "production rule –> LINQ –> SQL." (3) SQL query: the query execution time by the underlying DB Server. (4) Client processing: time consumed by the web browser, including AJAX calls to the server side and html rendering (show server-returned records in web page).

Three CT recruitment rules in different complexity are tested. As seen from Table 2, the most time-consuming parts are the SQL query and client processing, while the rule validation and DSL translation cost little. It can also been seen that as the rule complexity grows up, the consumed time of the DSL translation process and the underlying SQL server does not scale up dramatically. This shows that the computer-aided CT recruitment system has a good performance curve in regard to rule complexity.

(4) Rule Variable Extensibility. After the first release of the CT recruitment system, one frequent request posed by end users is supporting more data fields (rule variables). In the beginning, we only provided a few basic data fields in the rule authoring tool, such as birthday, gestation age, birth weight, and diagnosis. However, as users conduct more CTs, they continually request to extend more data fields, such as family disease history, drug use during pregnancy, and CPAP (continuous positive airway pressure) treatment.

To meet these frequent requests, we developed the rule authoring SDE tool based on pure web-client technologies (HTML and JavaScript), and the rule authoring tool is decoupled from server-side rule parsing and execution logics. When needing to extend a new rule variable, clinical engineers only need to extend a few lines of HTML and JavaScript codes for the authoring tool. Because the rule variable name and data type are the same as the corresponding data field in the database, the recruitment criteria rule produced by the authoring tool can be directly parsed and executed by the successive modules without any further modification. In this sense, the rule variable extensibility is quite satisfying and the system maintenance workload is greatly reduced.

(5) Query Accuracy and Recall Rate. Accuracy and recall rate are the two common measures for assessing query systems. For our system, we found that these two measures depend on how well the users write the rule expressions. The system is just an "executor" of the search rule. As there is no NLP or fuzzy-logic module in the system that could bring "uncertainty" or "ambiguity," the system itself does not compromise the overall query accurate and recall rate. By the rule

Table 2: Query performance result of the computer-aided CT recruitment system.

CT Recruitment rules in different complexity	Rule description	Rule validation	DSL translation	SQL query	Client processing	Total time
[BirthWeight] <= 1.0 && [GestationalAgeWeek] < 29	Find patients whose birth weight <= 1 kg and gestational age < 29 weeks	<1 ms	<1 ms	151 ms	285 ms	436 ms
[BirthWeight] <= 1.5 && [GestationalAgeWeek] <= 30 && [Birthday] >= 2014-01-01 && ([MostSevereDiagnosisOD] == %retin% || [MostSevereDiagnosisOS] == %retin%)	Find patients born after 2014 and whose birth weight <= 1.5 kg and gestational age <= 30 weeks and who have retina-related diseases	<1 ms	<1 ms	157 ms	289 ms	446 ms
[BirthWeight] <= 1.5 && [GestationalAgeWeek]<= 30 && [Birthday] >= 2014-01-01 && ([NumberOfFetus] >= 2 || [Homozygotic] == true || [MotherAge] >= 35 || [OxygenMethod] == %oxygen% || [SystemicDiseaseHistory] == %Apnea% || ([PregnancyStatus] == %HTN% || [PregnancyStatus] == %DM% || [PregnancyStatus] == %PIH%) || [PregnancyMedication] == %Corticosteroids% || ([MostSevereDiagnosisOD] == %ROP% || [MostSevereDiagnosisOS] == %ROP% || [MostSevereDiagnosisOD] == %retin% || [MostSevereDiagnosisOS] == %retin%) || ([FamilyDiseaseHistory] == %retin% || [FamilyDiseaseHistory] == %fundus%))	Find patients born after 2014 and whose birth weight <= 1.5 kg and gestational age <= 30 weeks and who meet one of the following conditions: homozygotic multiple birth, mother age >=35, oxygen therapy history, OSAS, having at least one of HTN, DM, or PIH during pregnancy, using corticosteroids during pregnancy, having retina-related diseases, or having family disease history of retina or fundus-related diseases	1 ms	<1 ms	159 ms	283 ms	443 ms

The results are averaged on 10 tests. Detailed result can be downloaded from http://ropd.brahma.pub/pages/20170130.xls.

authoring tool, the users can constantly improve query accuracy and recall rate by routine techniques, such as joining multiple variable conditions, increasing or decreasing thresholds, and adding more logical branches.

5. Conclusions and Discussions

The CT recruitment system has been used in Shenzhen Eye Hospital for more than a year. Clinical evaluation and feedback has proven its usability. The system is also well received by system developers and clinical engineers: the development and maintenance cost is significantly low, and the extensibility is also satisfying.

Such positive reviews are mainly due to two features of the system. (1) The system directly uses the underlying database schema as a reference model for all related modules, such as the rule authoring tool, the intermediate LINQ code, and the final SQL query. In this way, the system complexity and data-mapping efforts are greatly reduced. (2) The system uses the query-oriented DSL (e.g., LINQ in C# and VB.Net; JINQ, Linq4j, and JaQue in Java) as an intermediate formal language. Such DSL uses similar syntax as production rules and SQL, and are supported by ORM data providers. In this way, the translation process from recruitment criteria to native database queries is greatly simplified.

Compared to other related work, this study has several limitations and arguments.

(1) Data Integration and Knowledge Sharing. Much existing related work focus on how to achieve data sharing and data query across heterogeneous clinical data sources. The core method of these works is a common ontology or data model. For each data source, a middleware or data mapping module is provided to map the individual database to the shared model. The rule authoring tool also depends on the shared model, so the recruitment criteria are represented in a sharable computable form, and can be distributed as sharable knowledge assets.

On the other hand, because the ROP management system in our case study already has a central data repository that collects and stores data from multiple hospitals (in other words, the data integration work has been shifted down to the underlying data repository), we do not focus on the data integration issue from heterogeneous data sources. Our method does not need a shared ontology or data mapping middleware. However, the disadvantage is also apparent: the recruitment criteria rules created by the authoring tool are tightly coupled with the local database schema. In other words, the rules are "institution-specific" and are difficult to be distributed as sharable knowledge assets.

(2) Extending Rule Variables. As mentioned before, the rule variables provided by the rule authoring SDE tool come from the underlying database schema. In this way, the rule variables referenced in the rule expression can be "recognized" and "processed" by all related modules in the system. Such a design incurred some argument: What if the user wants to add a medical concept/rule variable that does not exist in current database schema? We can analyze this argument in 3 cases: (a) The concept is a high-level or coarse-grained concept, compared to the data fields in the database. For this case, the high-level concept can usually be composed by existing fine-grained items. For example, a user wants to use a "most severe diagnosis" concept, and the database has "most severe diagnosis OD" and "most severe diagnosis OS" data fields (OD = right eye, OS = left eye). A criteria rule of "[most severe diagnosis] = ROP_Stage4A" concept can be represented as "[most severe diagnosis OD] = ROP_Stage4A || [most severe diagnosis OS] = ROP_Stage4A". (b) The concept is a low-level or fine-grained concept. It is usually more difficult and complex to handle this case. For example, the user wants to use "ROP zone" and "ROP stage" as rule variables, but the database only has a string-typed "diagnosis" data field. A rule of "[ROP zone] = 1 && [ROP stage] = 3" can be represented as "[diagnosis] = %Zone1 Stage3% || [diagnosis] = %Stage 3 in zone 1%" (% is a wild card). However, due to the unconstraint nature of natural language, not all eligible patients can be covered by the previously mentioned rule, which will compromise the recall rate. Moreover, for many cases, it can even be impossible to compose an equal representation with coarse-grained data fields. (c) The concept is totally new to the current database. For example, the user wants to do some genotype research, but there are no genetic data fields in the underlying database. For these cases, the only solution in the context of our proposed method is to extend the database schema. In summary, case (a) is well supported in the system, case (b) is partially supported, and case (c) is not supported. For unsupported cases, the only feasible solution is to extend the database schema to add new data fields.

(3) Make Use of Unstructured Clinical Data. In this case study, the data stored in the ROP management system are mostly well formatted and structured. However, some relevant patient information still remains in the non-structured form (e.g., scanned handwritten RetCam report, and eye surgery record). How to use these narrative data is a long-existing problem faced by clinical researchers, and the promising NLP technologies have always been a hot research area. Joint query on both structured and nonstructured data will be a meaningful research topic for our CT recruitment system. Another kind of important nonstructured data is medical imaging. For ROP, the RetCam image is a very important evidence for diagnosis. How to extract morphological or even physiological information from the RetCam image is a challenging and meaningful task. Combining the imaging data will further improve the patient query result.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (71602044), Humanities and Social Sciences Foundation of Ministry of Education of China (15YJC630106), Natural Science Foundation of Zhejiang

Province of China (LQ16G020006), and Shenzhen Science and Technology Innovation Committee, PRC (201602293000443, JCY20140414114853651). This study is also supported by the Research Center of Information Technology and Economic and Social Development, Zhejiang Province, China.

References

[1] T. N. Arvanitis and W. Kuchinke, "TRANSFoRm query workbench," *Journal of Clinical Bioinformatics*, vol. 5, no. 1, article S16, pp. 1-2, 2015.

[2] J.-F. Ethier, O. Dameron, V. Curcin et al., "A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm," *Journal of the American Medical Informatics Association*, vol. 20, no. 5, pp. 986–994, 2013.

[3] R. Calinescu, S. Harris, J. Gibbons, and J. Davies, "Cross-trial query system for cancer clinical trials," in *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering*, Springer, Netherlands, 2007.

[4] L. Penberthy, R. Brown, F. Puma, and B. Dahman, "Automated matching software for clinical trials eligibility: measuring efficiency and flexibility," *Contemporary Clinical Trials*, vol. 31, no. 3, pp. 207–217, 2010.

[5] E. Cohen, J. Belkora, J. Tyler et al., "Adoption, acceptability, and accuracy of an online clinical trial matching website for breast cancer," *Journal of Medical Internet Research*, vol. 14, no. 4, article e97, 2012.

[6] A. Wagholikar, A. Nguyen, and M. Fung, "A method for matching patients to advanced prostate cancer clinical trials," *Electronic Journal of Health Informatics*, vol. 8, no. 1, article e6, 2014.

[7] C. O. Patel and J. J. Cimino, "Semantic query generation from eligibility criteria in clinical trials," in *American Medical Informatics Association Annual Symposium Proceedings*, p. 1070, 2007.

[8] S. Kraus, M. Enders, H.-U. Prokosch, I. Castellanos, R. Lenz, and M. Sedlmayr, "Accessing complex patient data from Arden Syntax Medical Logic Modules," *Artificial Intelligence in Medicine*, 2015, In Press.

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 4620732, 11 pages
https://doi.org/10.1155/2017/4620732

Research Article

Feature Extraction and Classification on Esophageal X-Ray Images of Xinjiang Kazak Nationality

Fang Yang,[1] Murat Hamit,[2] Chuan B. Yan,[2] Juan Yao,[3] Abdugheni Kutluk,[2] Xi M. Kong,[2] and Sui X. Zhang[2]

[1]*Department of Medical Engineering, The Affiliated Tumor Hospital, Xinjiang Medical University, Urumqi 830011, China*
[2]*College of Medical Engineering Technology, Xinjiang Medical University, Urumqi 830011, China*
[3]*Department of Radiology, The First Affiliated Hospital, Xinjiang Medical University, Urumqi 830054, China*

Correspondence should be addressed to Murat Hamit; murat.h@163.com

Received 17 November 2016; Revised 9 January 2017; Accepted 6 February 2017; Published 4 April 2017

Academic Editor: Jose M. Juarez

Esophageal cancer is one of the fastest rising types of cancers in China. The Kazak nationality is the highest-risk group in Xinjiang. In this work, an effective computer-aided diagnostic system is developed to assist physicians in interpreting digital X-ray image features and improving the quality of diagnosis. The modules of the proposed system include image preprocessing, feature extraction, feature selection, image classification, and performance evaluation. 300 original esophageal X-ray images were resized to a region of interest and then enhanced by the median filter and histogram equalization method. 37 features from textural, frequency, and complexity domains were extracted. Both sequential forward selection and principal component analysis methods were employed to select the discriminative features for classification. Then, support vector machine and *K*-nearest neighbors were applied to classify the esophageal cancer images with respect to their specific types. The classification performance was evaluated in terms of the area under the receiver operating characteristic curve, accuracy, precision, and recall, respectively. Experimental results show that the classification performance of the proposed system outperforms the conventional visual inspection approaches in terms of diagnostic quality and processing time. Therefore, the proposed computer-aided diagnostic system is promising for the diagnostics of esophageal cancer.

1. Introduction

Esophageal cancer is the eighth most common malignancy worldwide, with more than 480,000 new patients diagnosed annually. According to the Surveillance, Epidemiology, and End Result (SEER) statistics, the 5-year survival rate for esophageal cancer based on stage at diagnosis (2001–2007) is 17% overall: 37% for local disease; 18% for regional disease; and 3% for distant disease [1]. The World Health Report 2004 ranked esophageal cancer as the highest cause of cancer mortality in China. Among the 446,000 causes of death caused by esophageal cancer worldwide, more than half occurred in China, that is, 288 thousand (WHO, 2004) [2–4]. Xinjiang Uygur Autonomous Region is a high incidence area of esophageal cancer. The mortality rate of esophageal cancer for Kazak nationality is 155.9 out of 100,000, which is significantly higher than the average mortality of 15.23 out of 100,000 in China [5]. Over 80% of esophageal cancer occurs in developing countries, where nearly all cases are esophageal squamous cell carcinoma (ESCC). A number of risk factors for ESCC, including tobacco smoking, alcohol drinking, dietary and micronutrient deficiencies, high temperature of beverage and food consumption, and other miscellaneous factors (such as fast eating habits and polycyclic aromatic hydrocarbon exposure), have been identified over the past few decades [6]. The incipient symptoms of esophageal

cancer are too inconspicuous to be found. Most patients are diagnosed late in the course of the disease, and at this stage, it carries a bad prognosis. X-ray barium technology, as a crucial tool for the detection of esophageal cancer, offers the specialist physician high-quality visual information to identify the disease types [7]. Classically, the X-ray images are examined manually by physicians, and it is inevitability difficult to avoid inconsistent interpretations by interobservers. In some cases, even for experienced radiologists, they may misinterpret images of the esophageal cancer regions and miss smaller lesions. Therefore, the primary preventive strategies and control activities on esophageal cancer should be enhanced in the future, which are potentially effective to reduce the mortality of esophageal cancer and also essential to save lives and resources. In this paper, a computer-aided diagnostic system is developed to assist physicians in classifying the esophageal cancer with specific disease types.

With the rapid development in computer technology, CAD is currently widely used in the diagnosis or quantification of various diseases [8–10]. Many studies have shown that CAD has the potential to increase the sensitivity and the specificity of diagnostic imaging [11, 12]. The merit of CAD of image features lies in the objectivity and reproducibility of the measures of specific features. The conventional paradigm envisions that the CAD output will be used by the physician as a second opinion with the final diagnosis to be made by the physician [13]. Qi et al. developed a computer-aided diagnosis system to assist the detection of dysplasia in Barrett's esophagus. Experimental results showed that the proposed CAD algorithms had the potential to quantify and standardize the diagnosis of dysplasia and allowed high throughput image evaluation for endoscopic optical coherence tomography screening applications [14, 15]. Sommen et al. presented a novel algorithm for automatic detection of early cancerous tissue in HD endoscopic images. Experimental results showed that of 38 lesions indicated independently by the gastroenterologist, the system detected 36 of those lesions with a recall of 0.95 and a precision of 0.75 [16]. Schoon et al. proposed a CAD system to find the early stages of esophageal cancer. The results showed that the proposed system achieved a classification accuracy of 94.2% on normal and tumorous tissue and reached an area under the curve of 0.986 [17]. Esophageal cancer CAD literature published to data mostly focuses on endoscopic images. In addition to our previous study, no other papers have been found in the field of esophageal X-ray images to our best of knowledge.

The algorithms in the published CAD literature included image preprocessing, feature extraction, and pattern classification. Histogram equalization algorithm is one of the most widely used techniques for enhancing image contrast for its simplicity and effectiveness. Shang et al. proposed a Range Limited Peak-Separate Fuzzy Histogram Equalization (RLPSFHE) for enhancing image contrast for its simplicity and effectiveness. The experimental results show that the RLPSFHE can achieve a better trade-off between mean brightness preservation and contrast enhancement [18]. Zohair et al. introduced an ameliorated version of the contrast-limited adaptive histogram equalization (CLAHE) to provide a good brightness with decent contrast for CT images, which provided acceptable results with no visible artifacts and outperformed the comparable techniques [19]. The purpose of feature extraction is to extract the relevant features from the region of interest as the input vectors of the classifiers. Gu et al. proposed a new feature extraction method called adaptive slow feature discriminant analysis (ASFDA) in order to address the weaknesses of the traditional SFDA. Experimental results proved the superiority of ASFDA among some state-of-the-art methods [20]. Mueen et al. extracted three levels of features global, local, and pixel and combined them together in one big feature vector that achieved a recognition rate of 89% [21].

The classification based on multiple image features has the advantage of increasing accuracy via increasing the amount of information used. However, making use of too many image features derived from a limited training data set increases the risk of overfitting, which will decrease the robustness of the system when classifying data outside of the training set [22]. Therefore, it is necessary to select a limited number of image features to balance accurate and robust classification. Gladis et al. applied principal component analysis (PCA) with support vector machine (SVM) to classify the brain MR images by type. The recognition performance of the proposed technique was compared with three other method systems. Experimental results showed the PCA with SVM outperformed the three other methods in terms of classification accuracy [45]. Li et al. utilized the sequential forward selection algorithm (SFS) to figure out the nonunique probe selection problem. The experimental results demonstrate the proposed method outperformed the other greedy algorithms [23]. Techniques such as artificial intelligence and data mining techniques were widely used in the field of medical imaging classification [24]. SVM is a state-of-the-art pattern recognition technique grown up from a statistical learning theory. Papadopoulos et al. implemented artificial neural network (ANN) and a SVM to characterize the microcalcification clusters in digitized mammograms. The results indicated that the classification performance of SVM is superior to the ANN [25]. Zhu et al. employed the SVM to make a distinction within a class of Src kinase inhibitors. The sequential forward selection and sequential backward selection methods were used to remove redundant variables. The results showed that the proposed method could be employed to structure activity relationship modeling with much improved quality and predictability [37]. Katsuyoshi and Alberto detailed the *K*-nearest neighbor method for the application in breast cancer diagnosis. Experimental results showed that the classification accuracy changes with the number of neighbors and also with the percentage of data used for classification [26]. Chen et al. applied the KNN to classify the lung sounds. Experimental results indicated that the error in respiratory cycles between measured and actual values was only 6.8%, illustrating the potential of the detector for home care application [27]. Sharma and Khanna proposed a CAD system to detect abnormalities or suspicious areas in breast X-ray images and classify them as malignant and nonmaligant. Experiments were performed with three texture feature extraction techniques, including Zernike moments, gray-

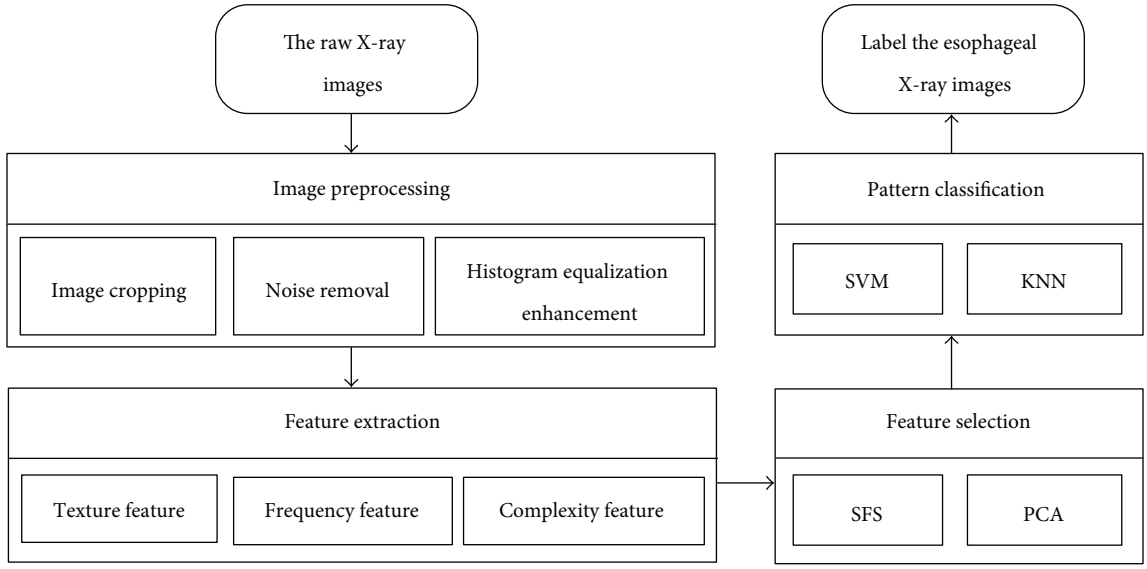


Figure 1: Flow chart of the system design.

level co-occurrence matrix, and discrete cosine transform. Experimental results showed that SVM with Zernike moments attains the optimum performance [28]. Though the literature published has shown the superiority on the recognition performance of the SVM and KNN, the impact of various feature selection algorithms on classification performance has not been fully explored.

This paper presents a computer-aided diagnostic system to classify the medical X-ray images of Xinjiang Kazak nationality esophageal by type. The proposed system consists of (I) image preprocessing, (II) feature extraction, (III) feature selection, and (IV) classification and performance evaluation. Firstly, the original images are resized to a region of interest and then enhanced by the median filter and histogram equalization method. During the feature extraction and selection step, the feature vectors of the classifiers are selected by PCA and SFS among 37 features in the textural, frequency, and complexity domains. The employed classifiers, that is, SVM and KNN, are validated using a 10-fold cross-validation technique that yields an average estimation of classifier performance with 95% confidence intervals. The performances of both classifiers are investigated with and without prior PCA and SFS input feature vector selection. AUC values of the receiver operating characteristic (ROC) curves, accuracy, precision, and recall, are used to evaluate the classification performance.

2. Methods and Techniques

The proposed methodology is applied to 300 raw esophageal X-ray images, of which 100 were classified by a pathologist as normal images and 200 as abnormal images. The abnormal cases were further divided in two categories: 100 fungating type and 100 ulcerative type. These images, which included 221 males (mean age: 65) and 79 females (mean age: 68) with an age range of 45–80 years, were collected from The First Affiliated Hospital, Xinjiang Medical University of China. The proposed algorithms were implemented in the Matlab 2013 platform. The flow chart of the system design is depicted in Figure 1.

2.1. Image Preprocessing. Customarily, preprocessing is a necessity whenever the data to be mined is noisy, inconsistent, or incomplete. Preprocessing significantly improves the effectiveness of data mining techniques [29]. The typical size of the raw images is 1012×974, and almost 50% of the whole image comprised the background with a lot of noise. Moreover, these images are scanned at different illumination conditions, so some images appeared too bright and some are too dark. To circumvent the above-mentioned issue, the first step toward noise removal is pruning the original images with a cropping operation. The images are resized to a region of interest of 140×240 pixels, which can guarantee that all the regions of interest contain the lesion areas meanwhile avoid the useless information. In addition, the median filter is applied to the cropped images in order to further eliminate the image noise. The second step is image enhancement, in particular, the histogram equalization method, which can increase the contrast range in an image by increasing the dynamic range of gray levels, which is utilized to enhance the image for diminishing the effects of over-brightness and over-darkness in images. The preprocessed images are again inspected by a pathologist to ensure that their quality was sufficient for diagnosis. Figure 2 presents the preprocessing results of the abnormal esophageal X-ray images, fungating and ulcerative esophageal X-ray images.

2.2. Feature Extraction. The purpose of feature extraction in this project is to convert a two-dimensional image into a feature vector, which can be further utilized as the input for the mining phase of the classifier. The extracted features should provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into feature vectors. Accordingly, three kinds of

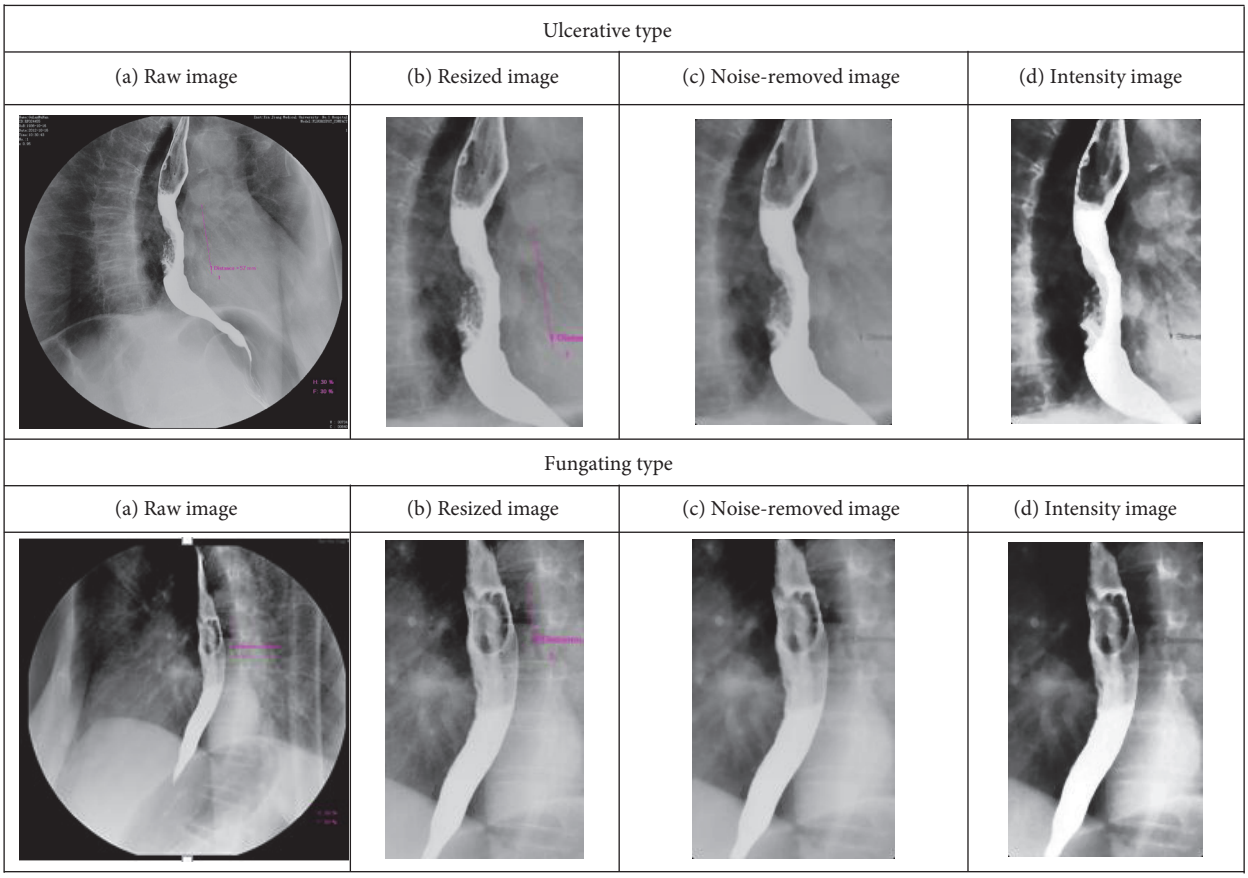


Figure 2: Preprocessing results of the abnormal esophageal X-ray images.

features are extracted to describe the structure information of texture, frequency, and complexity.

2.2.1. Texture Features. Texture contains important information regarding underlying structural arrangement of the surface of an image. Gray-level co-occurrence matrix (GLCM), which describes patterns of gray-level repetition, is a well-known texture extraction method originally introduced by Haralick et al. [30]. The co-occurrence matrix is constructed by getting information about the orientation and distance between the pixels. Assuming that $f(x,y)$ is a two-dimensional image with the size of $M \times N$, the definition of the co-occurrence matrix is as follows:

$$P(i,j \mid d,\theta) = \#\{[(x1,y1),(x2,y2)] \in M \times N \mid d,\theta, f(x1,y1) = i, f(x2,y2) = j\}, \quad (1)$$

where $\#\{\}$ denotes the number of the elements of the set. d and θ are the distance and angle between $(x1,y1)$ and $(x2,y2)$, respectively.

Many texture features can be directly computed from the gray-level co-occurrence matrix. Pourghassem et al. extracted contrast, correlation, energy, and homogeneity from GLCM [31].

$$\begin{aligned}
\text{Contrast} &= \sum_{i=1}^{L-1}\sum_{j=1}^{L-1}(i-j)^2 P(i,j,d,\theta),\\
\text{Correlation} &= \frac{\sum_{i=1}^{L-1}\sum_{j=1}^{L-1} i\cdot j\cdot P(i,j,d,\theta) - \mu_x\mu_y}{\sigma_x\sigma_y},\\
\text{Energy} &= \sum_{i=1}^{L-1}\sum_{j=1}^{L-1}[P(i,j,d,\theta)]^2,\\
\text{Homogeneity} &= \sum_{i=1}^{L-1}\sum_{j=1}^{L-1}\frac{p(i,j,d,\theta)}{1+|i-j|},
\end{aligned} \quad (2)$$

where (μ_x,σ_x) and (μ_y,σ_y) are mean and standard deviation of pixel value in the row and column directions of the GLCM, respectively. For this task, we calculate a gray-level co-occurrence matrix for four different directions $\theta \in \{0°, 90°, 45°, \text{and } 135°\}$ and the distance $d = 1$. As a result, texture feature vector includes 16 elements.

2.2.2. Frequency Features. The discrete wavelet decomposition (DWT) has been widely used as a fast algorithm to obtain the wavelet transform of X-ray medical images [32, 33]. The DWT analyzes the images by decomposing it

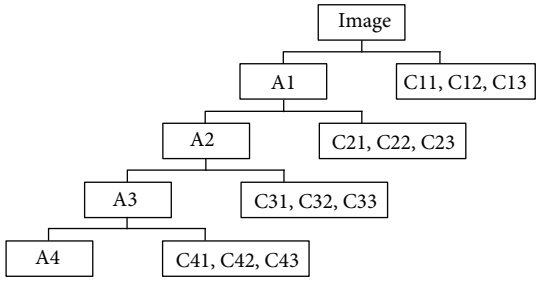


Figure 3: Four-level DWT decomposition process.

into coarse approximation and detailed information representing the low- and high-frequency contents of images, respectively. The approximation can be further calculated to produce the approximation and detailed information at the next level of the decomposition and so on till the required level is reached. Figure 3 depicts the wavelet decomposition process of this work. Specifically, A1–A4, representing the wavelet approximations of four levels, are low-frequency part of the images. C11–C13, C21–C23, and C31–C33, denoting the details of horizontal, vertical, and diagonal directions of four levels, are high-frequency part of the images. Empirically, C11–C13 can be discarded, since they contain little useful information and a lot of noise. And the approximation coefficient A4 at fourth level is used to represent the low frequency of the image. The mean and variance values are further calculated from each coefficient after the DWT is performed on the X-ray images. Therefore, 20 features are extracted from an input image.

2.2.3. Kolmogorov Complexity Features. An image can be converted into a one-dimensional binary sequence via scanning it either horizontally or vertically. The complex value of each row vector can be obtained by evaluating the complexity of each vector in the horizontal direction. The complexity of the complex vector, which is comprised of the complexity of each row, can be calculated as the complexity feature of the image. Kolmogorov [34] proposes to measure the conditional complexity of a finite object x, given a finite object y by the length of the shortest sequence p, that consists of 0 s and 1 s and thus makes it possible to reconstruct x given y. Mathematically, this is explained as follows:

$$K_B(x \mid y) = \min\{l(p) \mid B(p,y) = x\}, \tag{3}$$

where $l(p)$ is the length of the sequence p and $B(p,y)$ is the decoding function, for which there is an algorithm computing its values.

Kolmogorov only gave a general definition of the Kolmogorov complexity. Kasper and Schuster [35] proposed an explicit algorithm to compute the KC measure, which includes two operations, copying and inserting. After the explicit algorithm is applied to the images, one feature is obtained.

2.3. Feature Selection. Feature selection is an optimization technique that, given a set of features, attempts to select a subset of size that leads to the maximization of some criterion function [36]. In this paper, we employ both sequential forward selection (SFS) and principal component analysis (PCA) methods to select the discriminative features among the feature vector.

2.3.1. Sequential Forward Selection. Informally, SFS algorithm can be described as follows [37]: SFS begins with an empty feature set, and all the observation features were marked as nonselected features. At each iteration, one feature from among the nonselected features is added to the feature set, which minimizes the mean square error (MSE). The iterative process could be stopped until the best merit MSE is obtained. MSE can be defined as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x}), \tag{4}$$

where X denotes the random variables. N is defined as the number of samples taken.

2.3.2. Principal Component Analysis. Principal component analysis, which is also known as Karhunen-Loeve (KL) transform, is a projection-based technique that facilitates a reduction in data dimension through the construction of orthogonal principal components that are weighted, linear combinations of the original variables [38–40]. Assuming that a linear transformation mapping the original N-dimensional feature space into an M-dimensional space, where $M < N$, the PCA transform can be denoted as follows:

$$F_D = F_V X_a, \tag{5}$$

where F_V is the so-called eigenvector, whose length depends on the components that we want for expressing the observation feature space. The resultant feature space is the projection of the original data set over the eigenvectors of the covariance matrix. In this study, we applied the PCA for investigating if the reduced set of features can retain significant discrimination of the projected data. Firstly, the original matrix was converted into a standardized matrix. That is, the features were normalized to have zero means and unit variances. Secondly, the covariance matrix, which comprises the weights of each feature in the input space, was calculated. In addition, the eigenvalues and the corresponding eigenvectors of the covariance matrix were computed. The eigenvector with highest eigenvalue was the first principle component that contains the most significant information and accounts for the larger amount of variance in the data. The first few principal components are selected to be the inputs of classifiers when their accumulative contributive rate was 0.9.

2.4. Classification and Performance Evaluation. In this study, two classifiers, that is, K-nearest neighbors (KNN) and support vector machine (SVM) with radial basis function (RBF), were used for classification. SVM seeks the optimal boundary between two classes. The popularity of this method has grown as it provides a powerful machine learning

technique to classify data. KNN is known in the machine learning field as a nonparametric method.

2.4.1. Support Vector Machine (SVM). Support vector machine, a technique derived from statistical learning theory, is the most promising technique for data classification and regression and function estimation [41–44]. The basic idea of applying SVM for solving classification problems can be stated briefly as follows: (a) transform the input space to higher dimension feature space by a nonlinear mapping function and (b) construct the separating hyperplane with the maximum distance from the closest points of the training set [45]. SVM has high classifying accuracy and good capabilities of fault tolerance and generalization. SVM constructs a binary classifier from a set of training samples $(x_1,\ldots,x_n)$, which belongs to a class label. SVM selects the hyperplane that causes the largest separation among the decision function values for the borderline examples of the two classes. The hyperplane decision function can be defined as follows:

$$f(x) = \operatorname{sign}\left[\sum_{i=1}^{n} \alpha_i y_i K(x_i,x) - b\right], \tag{6}$$

where $K(x_i,x)$ is the kernel function. b is the classification threshold. α_i is lagrangian multiplier, which is calculated by quadratic programming problem.

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i\alpha_j y_i y_j K(x_i,x_j) \tag{7}$$

subject to $\sum_{i=1}^{l}\alpha_i y_i = 0$, $0 \le \alpha_i \le C\,(i = 1,\ldots,l)$, $0 \le \alpha_j \le C$ $(j = 1,\ldots,l)$.

There are three parameters in SVM model that we should choose. They make great impact on a model's generalization ability. It is well known that SVM generalization performance depends on a good setting of hyperparameters C, the kernel function, and kernel parameter. For multiclassification problems, there are two general approaches, one-against-one and one-against-all. In the former approach, classifier is calculated from each pair of classes. All classifiers are combined to conclude the final classification by using majority voting scheme. In the latter one, the classifier is calculated from each class versus all classes and then the first object that is classified as a single class is the type of the unlabeled data.

2.4.2. K-Nearest Neighbors (KNN). The K-nearest neighbor classifier is firstly proposed by Cover and Hart in 1968 [46]. It is a nonparametric learning algorithm that is used for classification and regression [47]. KNN is a very simple but efficient algorithm because it is a typical type of instance-based or memory-based learning scheme. The implementation process of the K-nearest neighbor algorithm is as follows [48]:

(I) In the first step, the number of nearest points of test data x against training data K is determined. Euclidean distance is the most commonly used to measure the distance between two instances according to the type of attribute [49]. Assuming there are two points in K-dimensional space, $x = [x_1,x_2,\ldots,x_k]$ and $y = [y_1,y_2,\ldots,y_k]$, the Euclidean distance between the two can be denoted by

$$d(x,y) = \sqrt{\sum_{i=1}^{k} (y_i - x_i)^2}. \tag{8}$$

(II) We can judge that the test data x is a certain category when it has more representatives than a certain category of data.

Generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good k can be selected by cross-validation, running the nearest neighbor classifier on the learning set only. Due to its implementation simplicity and classification effectiveness, KNN has been widely used in pattern recognition. It is also used as a different feature selection algorithm [50, 51] and is integrated into the feature selection framework to evaluate the quality of a candidate feature subset [52–54].

2.4.3. Performance Evaluation. The classifiers are validated using a 10-fold cross-validation technique that yields an average estimation of classifier performance with 95% confidence intervals. In the cross-validation, 90% of samples were used for training and 10% were used for the validation replications. The performances of the classifiers are evaluated in terms of the area under the receiver operating characteristic (ROC) curve (AUC), accuracy, precision, and recall. The ROC analysis is a commonly used approach for classification performance evaluation [55]. The AUC value is the average true positive rates over all possible false positive rates. The accuracy, precision, and recall [56] are given as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{Number of correctly classified images}}{\text{Total Number of images}} \times 100\%,\\
\text{Precision} &= \frac{\text{Number of correctly classified images per class}}{\text{Total number of classified images per class}} \times 100\%,\\
\text{Recall} &= \frac{\text{Number of correctly classified images}}{\text{Total number of expected images in the corresponding class}} \times 100\%.
\end{aligned} \tag{9}$$

3. Results and Discussion

The above-described methodology has been evaluated on a set of esophageal X-ray images collected from The First Affiliated Hospital of Xinjiang Medical University. During the classification stage, performance comparison is divided into three categories: (1) all 37 features; (2) features selected

Table 1: Details of feature selection by SFS for the first-stage classification process.

Features		Feature number			
Texture features (θ, d)	(0°, 1)	*1*	2	3	4
	(45°, 1)	5	*6*	*7*	*8*
	(90°, 1)	*9*	*10*	*11*	12
	(135°, 1)	13	14	15	16
Frequency features		17	18	19	20
		21	22	23	*24*
		25	*26*	*27*	*28*
		29	*30*	31	32
		33	*34*	*35*	36
KC features		*37*			

The numbers in italics are the features selected by SFS.

Table 2: Details of feature selection by SFS for the second-stage classification process.

Features		Feature number			
Texture features (θ, d)	(0°, 1)	*1*	2	3	4
	(45°, 1)	5	*6*	7	*8*
	(90°, 1)	9	*10*	*11*	12
	(135°, 1)	*13*	14	15	16
Frequency features		17	18	19	20
		21	22	*23*	*24*
		25	*26*	*27*	*28*
		29	*30*	31	*32*
		33	34	*35*	36
KC features		*37*			

The numbers in italics are the features selected by SFS.

by SFS; and (3) features selected by PCA. The classification was conducted on a two-stage process. In the first-stage classification process, the X-ray images are classified as normal and abnormal. Then the second-stage classification process continues the abnormal images that are classified as fungating and ulcerative type images. And the classifiers were validated by a 10-fold cross-validation technique. The classification performance was measured by the AUC values of the ROC curves, accuracy, precision, and recall.

Feature selection is carried out using SFS and PCA methods to remove the redundancy due to highly correlated features. During the first-stage and second-stage classification processes, the SFS selected 17 appropriate features out of 37 features, respectively. It means a reduction of computing time and data storage space. The selected features are from the textural, frequency, and complexity domains and all useful for the classification. The results of feature selection of SFS for the two-stage classification process are detailed in Tables 1 and 2. Among the appropriate 17 features selected by the SFS, the higher proportion is $\theta = 45°, 90°$. This result shows that texture of esophageal focus may occur in the particular angle and distance. Each principal component is orthogonal and represents a linear combination of the original variables. The first few principal components typically account for most of the variance in the original data. In this analysis, the first six principal components together explained 90.7% and 92.26% of the variance for the first-stage and second-stage classification processes, respectively. The eigenvalue and the cumulative variance of the first six principal components for the two-stage classification are tabulated in Table 3.

Table 3: Details of feature selection by PCA for the two-stage classification process.

PC	Eigenvalue		Cumulative variance (%)	
	First stage	Second stage	First stage	Second stage
PC1	11.07	15.1	35	45.8
PC2	6.84	6.2	53.4	62.56
PC3	5.57	4.44	68.4	74.57
PC4	3.4	3.24	77.6	83.32
PC5	2.94	1.84	85.6	88.3
PC6	1.91	1.47	90.7	92.26

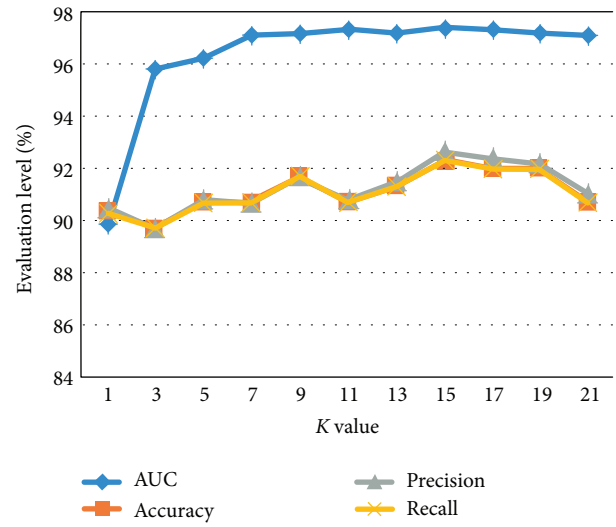


Figure 4: KNN classification results for various choices of K (%).

Figure 4 reports the KNN classification results for values of K ranging from one to twenty-one using 10-fold cross-validation. It can be seen from Figure 4 that KNN classifier achieved the best classification when $K = 15$. It is observed that the KNN classifier has an AUC value of 97.4%, accuracy of 92.33%, precision of 92.7%, and recall of 92.3%.

The radial basis function (RBF) kernel is chosen for SVM classifier. For the training of KNN classifier, the number of the nearest neighbor $K = 15$ and Euclidean distance metric was employed. Based on the result shown in Table 4, Figure 5, and Figure 6, the following conclusions can be drawn:

(a) The step of feature selection not only reduces the dimension of the input vector, but also improves classification performance. This may be due to the elimination of the correlated features from the 37-D feature vector.

Table 4: Classification performance of SVM and KNN classifiers (%).

Parameters		All features		SFS selection		PCA selection	
		SVM	KNN	SVM	KNN	SVM	KNN
AUC	First stage	94.5	93.7	97.4	95.7	95.33	94.5
	Second stage	94	93.4	97	94.67	95.14	94
Accuracy	First stage	92.67	91.3	95	93	93	92.67
	Second stage	91.5	90.14	94.67	92.14	92.5	91.5
Precision	First stage	91	90.4	94.33	92.5	91.4	91.33
	Second stage	90.67	90.33	94.14	92.33	91.67	91.4
Recall	First stage	91	90	94	92.5	91.4	91
	Second stage	90.67	90.33	94.14	92	91.67	91.4

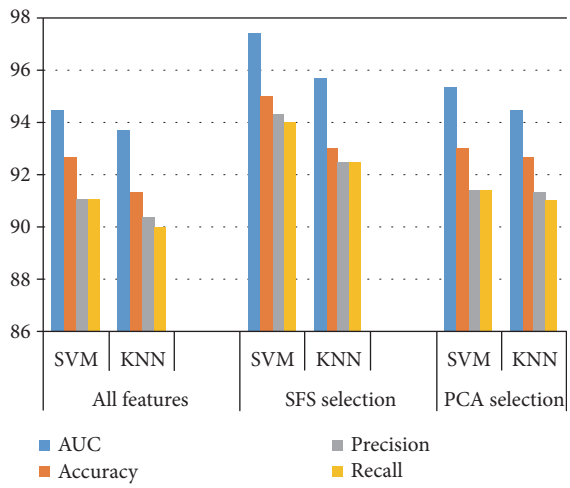


Figure 5: Classification performance of the first classification stage (%).

(b) The SFS outperforms the PCA. In the first-stage classification, for all 37 features used as input vectors, it yields the best AUC value of 94.5%, accuracy of 92.67%, precision of 91%, and recall of 91%. With input features selected by SFS and PCA, the corresponding AUC value, accuracy, precision, and recall are 97.4% and 95.33%, 95% and 93%, 94.33% and 91.4%, and 94% and 91.4%, respectively. In the second-stage classification, it produces the best AUC value of 94%, accuracy of 91.5%, precision of 90.67%, and recall of 90.67% for all the 37 features. With the input vectors selected by SFS and PCA, the corresponding AUC value, accuracy, precision, and recall are 97% and 95.14%, 94.67% and 92.5%, 94.14% and 91.67%, and 94.14% and 91.67%, respectively.

(c) Under either feature selection criterion (no selection, SFS selection, and PCA selection), the performance of SVM is better than the KNN. The highest classification performance was achieved when the SVM classifier and SFS selection are employed.

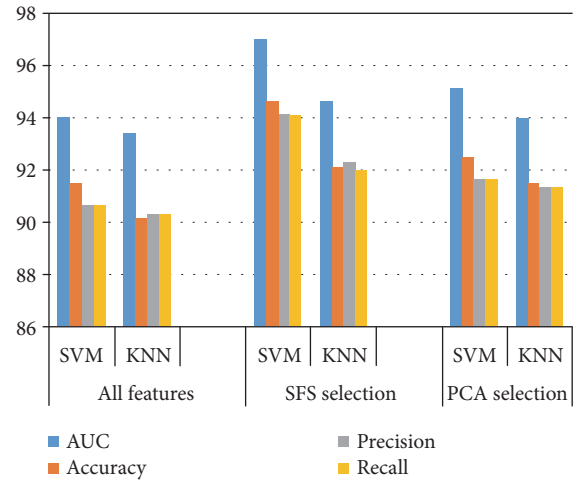


Figure 6: Classification performance of the second classification stage (%).

Table 5: Classification performance of previous studies (%).

Methods	Classification accuracy
GH + Bayes [57]	76.6
WT + Bayes [58]	76.5
GH + GLCM + Bayes [59]	86.7
GLCM + GGCM + PCA + KNN [60]	87.5

GH: gray-level histogram; WT: wavelet-based transform; GGCM: gray-gradient co-occurrence matrix.

In our previous studies, several methods related to computer-aided diagnosis system of esophageal cancer have been developed. The classification performances are tabulated in Table 5. It is observed that single feature reached lower classification accuracy. The classification performance improved in the case of using the comprehensive feature without dimensional reduction algorithm. When the feature extraction methods were utilized, the accuracy obtained the further improvement.

Although the previous works have made some achievements, the classification performance still needs to be improved in order to meet the requirements of esophageal cancer diagnosis. The present study introduced the KC feature extraction and SFS and SVM algorithms, and the high classification performance was achieved by combining with the previous method.

The processing time of the proposed method takes around 14.32 s (11.02 s for image preprocessing, 2.16 s for feature extraction, and 1.14 s for classification) while the manual recognition takes about 37 s. The accuracy of detecting the esophageal cancer via both specialist physicians and the proposed method is 92% and 95%, respectively. And the accuracy of classifying the abnormal images into fungating and ulcerative types reaches up to 90% and 94.67%, respectively. The classification performance of the proposed method outperforms the conventional visual inspection approach by improving the diagnostic quality and processing time.

4. Conclusions

Esophageal cancer has a high mortality in Xinjiang Kazak nationality. X-ray barium technology is more commonly used in the diagnosis of this disease. However, the differences of experience, knowledge, and skills among individual physicians may affect the diagnosis results. This paper presents a computer-aided diagnosis system with image processing and pattern recognition in diagnosing esophageal cancer of Xinjiang Kazak nationality by using X-ray images. The original images, including normal esophageal images, fungating and ulcerative type images, were first resized to a region of interest and then enhanced by the median filter and histogram equalization method. Then, 37 features were obtained from images using three different techniques, which include textural, frequency, and complexity domains. SFS and PCA methods were applied to select the input features for classification. Furthermore, the esophageal cancer images were classified via SVM and KNN classifiers by type. And the classifiers were validated by a 10-fold cross-validation strategy. The classification performance was evaluated in terms of the AUC values, accuracy, precision, and recall, respectively.

A two-stage classification process was carried out for classifying the esophageal cancer by type. In the first-stage classification process, the X-ray images are classified as normal and abnormal. For all 37 features used as input vectors, it yielded the best AUC value of 94.5%, accuracy of 92.67%, precision of 91%, and recall of 91%. With input features selected by SFS and PCA, the corresponding AUC value, accuracy, precision, and recall were increased by 2.9% and 0.83%, 2.33% and 0.33%, 3.33% and 0.4%, and 3% and 0.4%, respectively. Then the second-stage classification process continues the abnormal images that are classified as fungating and ulcerative type images. It produced the best AUC value of 94%, accuracy of 91.5%, precision of 90.67%, and recall of 90.67% for all the 37 features. With the input vectors selected by SFS and PCA, the corresponding AUC value, accuracy, precision, and recall were increased by 3% and 1.14%, 3.17% and 1%, 3.47% and 1%, and 3.47% and 1%, respectively. Experimental results show that the highest classification performance is achieved when the SVM classifier and SFS selection were employed. The accuracy of detecting the esophageal cancer and classifying it by type via specialist physician and the proposed method is 92% and 95% and 90% and 94.67%, respectively. The classification performance of the proposed system outperformed the conventional visual inspection approach by improving the diagnostic quality and processing time.

The proposed method may be limited in the following aspects. First, the regions of interest of the images were selected manually, which result to be time-consuming during the image processing stage. This is because the lesion areas vary greatly from different images, and it is hard to find a unified segmentation method at present. The second important limitation of the study is the lack of comparison with the early esophageal cancer because of the small number of images in early stage. Based on the limitations of the current study, the future perspectives of our work aiming for diagnostic quality improvements may lie in studying more advanced feature extraction model and the segmentation method for esophageal X-ray images. An interesting improvement could be to extend it into the comparison research between the normal esophageal and the early esophageal cancer.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project was performed under the auspices of the Natural Science Foundation of China Grants 81460281, 81560294, 81160182, and 61201125. The authors would like to thank the Department of Radiology, First Affiliated Hospital of Xinjiang Medical University, Urumqi, China.

References

[1] A. H. Maria and A. Katharine, "Image-guided radiotherapy for esophageal cancer," *Imaging in Medicine*, vol. 4, no. 5, pp. 515–525, 2012.

[2] D. M. Parkin, F. I. Bray, and S. S. Devesa, "Cancer burden in the year 2000. The global picture," *European Journal of Cancer*, vol. 37, no. 58, pp. S4–S66, 2001.

[3] J. Y. Guang, W. L. Qian, X. D. Yu et al., "Analysis on the epidemiological characteristics of esophageal cancer in Huai'an area, China from 2009 to 2011," *The Chinese-German Journal of Clinical Oncology*, vol. 11, no. 9, pp. 504–507, 2012.

[4] Y. Z. Xue, F. Z. Da, M. Xin, and D. Jiang, "Esophageal cancer spatial and correlation analyses: water pollution, mortality rates, and safe buffer distances in China," *Journal of Geographical Sciences*, vol. 24, no. 1, pp. 46–58, 2014.

[5] G. Hui, B. D. Jian, Z. Wei, and T. Zhang, "Gene research progress on Xinjiang kazak esophageal cancer," *Basic Medicine and Clinical*, vol. 30, no. 4, pp. 428–430, 2010.

[6] F. Kamangar, W. Chow, C. C. Abnet, and S. M. Dawsey, "Environmental causes of esophageal cancer," *Gastroenterology Clinics of North America*, vol. 38, no. 1, pp. 27–57, 2009.

[7] X. C. Shi, B. L. Xian, and P. C. Hua, "Digital X-ray barium meal in the diagnosis of early esophageal carcinoma," *Practical Journal of Clinical Medicine*, vol. 8, no. 1, pp. 42–44, 2011.

[8] M. B. Nagarajan, P. Coan, M. B. Huber, P. C. Diemoz, C. Glaser, and A. Wismuller, "Computer-aided diagnosis in phase contrast imaging X-ray computed tomography for quantitative characterization of ex vivo human patellar cartilage," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2896–2903, 2013.

[9] X. Yang, Z. Jie, L. N. Li et al., "Computer-aided diagnosis based on quantitative elastographic features with supersonic shear wave imaging," *Ultrasound in Medicine and Biology*, vol. 40, no. 2, pp. 275–286, 2014.

[10] L. J. Meng, T. Z. Shao, S. L. Hong, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783–792, 2015.

[11] R. L. Ellis, A. A. Meade, M. A. Mathiason, K. M. Willison, and W. Logan-Young, "Evaluation of computer-aided detection system in the detection of small invasive breast carcinoma," *Radiology*, vol. 245, no. 1, pp. 88–94, 2007.

[12] F. M. Hall, "Improved sensitivity of mammography with computer-assisted detection on interpretive performance in screening mammography," *American Journal of Roentgenology*, vol. 187, no. 6, pp. 1472–1482, 2006.

[13] M. L. Giger, N. Karssemeijer, and S. G. Armato, "Computer-aided diagnosis in medical imaging," *IEEE Transactions of Medical Imaging*, vol. 20, no. 12, pp. 1205–1208, 2001.

[14] X. Qi, M. V. Sivak, J. E. Willis, and A. M. Rollins, "Computer-aided diagnosis of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," *Journal of Biomedical Optics*, vol. 11, no. 4, p. 044010, 2006.

[15] X. Qi, Y. Pan, M. V. Sivak, J. E. Willis, G. Isenberg, and A. M. Rollins, "Image analysis for classification of dysplasia in Barrett's esophagus using endoscopic optical coherence tomography," *Biomedical Optics Express*, vol. 1, no. 3, pp. 825–847, 2010.

[16] F. V. Sommen, S. Zinger, E. J. Schoon, and P. H. N. de With, "Supportive automatic annotation of early esophageal cancer using local gabor and color features," *Nerocomputing*, vol. 144, pp. 92–106, 2014.

[17] E. J. Schoon, F. V. Sommen, S. Zinger, and P. H. N. de With, "Computer-aided delineation of early Neoplasia in Barrett's esophagus using high definition endoscopic images," *Gastrointestinal Endoscopy*, vol. 77, no. 5, Supplement, p. AB471, 2013.

[18] B. Z. Shang, P. Z. Fu, and A. S. Muhammad, "Range limited peak-separate fuzzy histogram equalization for image contrast enhancement," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 6827–6847, 2015.

[19] A. A. Zohair, S. L. Ghazali, R. Amjad, A. Al-Dhelaan, T. Saba, and M. Al-Rodhaan, "An innovative technique for contrast enhancement of computed tomography images using normalied gamma-corrected contrast-limited adaptive histogram equalization," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–12, 2015.

[20] X. Gu, C. Liu, S. Wang, and C. Zhao, "Feature extraction using adaptive slow feature discriminant analysis," *Neurocomputing*, vol. 154, pp. 139–148, 2015.

[21] A. Mueen, M. S. Baba, and R. Zainuddin, "Multilevel feature extraction and X-ray image classification," *Journal of Applied Science*, vol. 7, no. 8, pp. 1224–1229, 2007.

[22] L. Yu and H. Liu, "Efficient feature selection via analysis of relevant and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.

[23] L. W. Li, N. Alioune, and R. Luis, "Sequential forward selection approach to the non-unique oligonucleotide probe selection problem," *Lecture Notes in Computer Science*, vol. 5265, pp. 262–275, 2008.

[24] J. C. Fu, S. K. Lee, S. T. Wong, J. Y. Yeh, A. H. Wang, and H. K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications," *Computerized Medical Imaging and Graphics*, vol. 29, no. 6, pp. 419–429, 2005.

[25] A. Papadopoulos, D. I. Fotiadis, and A. Likas, "Characterization of clustered micro-calcifications in digitized mammograms using neural networks and support vector machines," *Artificial Intelligence in Medicine*, vol. 34, no. 2, pp. 141–150, 2005.

[26] O. Katsuyoshi and P. P. Alberto, "A detailed description of the use of the KNN method for breast cancer diagnosis," in *The 2014 7th international conference on biomedical engineering and informatics*, pp. 606–610, Dalian, China, 2014.

[27] C. H. Chen, W. T. Huang, T. H. Tan, C. C. Chang, and Y. J. Chang, "Using k-nearest neighbor classification to diagnose abnormal lung sounds," *Sensors (Basel)*, vol. 15, no. 6, pp. 13132–13158, 2015.

[28] S. Sharma and P. Khanna, "Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM," *Journal of Digital Imaging*, vol. 28, no. 1, pp. 77–90, 2015.

[29] R. C. Gonzalez, *N.: Digital Image Processing, 2nd Edn. Addison-Wesley, Reading*, 1993.

[30] R. M. Haralick, K. Shanmugam, and I. Distenin, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[31] H. Pourghassem and H. Ghassemian, "Content-based medical image classification using a new hierarchical merging scheme," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 651–661, 2009.

[32] T. J. Penfold, I. Travernelli, C. J. Milne et al., "A wavelet analysis for the X-ray absorption spectra of molecules," *Journal of Chemical Physics*, vol. 138, no. 1, p. 014104, 2013.

[33] B. C. Ko, S. H. Kim, and J. Y. Nam, "X-ray image classification using random forests with local wavelet-based CS-local binary patterns," *Journal of Digital Imaging*, vol. 24, no. 6, pp. 1141–1151, 2011.

[34] V. V. Yugin, "Algorithmic complexity and stochastic properties of finite binary sequences," *The Computer Journal*, vol. 42, no. 4, pp. 294–317, 1999.

[35] F. Kasper and H. G. Schuster, "Easily calculable measure for the complexity of spatiotemporal patterns," *Physics Review A: Atomic, Molecular and Optical Physics*, vol. 36, no. 2, pp. 842–848, 1987.

[36] S. Rajeswari and J. K. Theiva, "Support vector machine classification for MRI images," *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 3, pp. 1534–1539, 2012.

[37] J. Zhu, W. Lu, L. Liu, and B. Niu, "Classification of Src kinase inhibitors based on support vector machine," *QSAR and Combinatorial Science*, vol. 28, no. 6, pp. 719–727, 2009.

[38] X. G. Rui, A. Mihye, and T. Z. Hong, "Spatially weighted principal component analysis for imaging classification," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 274–296, 2015.

[39] B. C. Yan and S. L. Cheng, "Belnded coal's property prediction model based on PCA and SVM," *Journal of Central South University of Technology*, vol. 15, no. 2, pp. 331–335, 2008.

[40] A. P. Nanthagopal and R. S. Rajamony, "Automatic classification of brain computed tomography images using wavelet-based statistical texture features," *Journal of Visualization*, vol. 15, no. 4, pp. 363–372, 2012.

[41] J. Y. Lin, C. T. Cheng, and K. W. Chan, "Using support vector machines for long term discharge prediction," *Hydrological Sciences Journal*, vol. 51, no. 4, pp. 599–612, 2006.

[42] L. Y. Chuang, C. H. Yang, and L. C. Jin, "Classification for multiple cancer types using support vector machines and outliner detection methods," *Biomedical Engineering Applications, Basis & Communications*, vol. 17, pp. 300–308, 2005.

[43] M. D. Ashanira, M. Z. Azlan, and S. Roselina, "Hybrid GR-SVM for prediction of surface roughness in abrasive water jet machining," *Meccanica*, vol. 48, no. 8, pp. 1937–1945, 2013.

[44] N. H. Chiu and Y. Y. Guao, "State classification of CBN grinding with support vector machine," *Journal of Materials Processing Technology*, vol. 201, no. 1, pp. 601–605, 2008.

[45] V. P. Gladis and S. Palani, "A novel approach for feature extraction and selection on MRI images for brain tumor classification," *Computer Science & Information Technology*, vol. 10, no. 5, pp. 225–234, 2012.

[46] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[47] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Accelerating incremental wrapper based gene selection with k-nearest-neighbor," in *IEEE international conference on bioinformatics and biomedicine (BIBM), IEEE*, pp. 21–23, Belfast, UK, 2014.

[48] H. C. Chin, T. H. Wen, H. T. Tan, C. C. Chang, and Y. J. Chang, "Using k-nearest neighbor classification to diagnose abnormal lung sounds," *Sensors*, vol. 15, no. 6, pp. 13132–13158, 2015.

[49] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Accelerating wrapper-based feature selection with K-nearest-neighbor," *Knowledge-Based Systems*, vol. 83, pp. 81–91, 2015.

[50] K. Moorthy and M. Mohamad, "Random forest for gene selection and microarray data classification," *Bioinformation*, vol. 7, no. 3, pp. 142–146, 2011.

[51] X. Sun, Y. Liu, M. Xu, H. Chen, J. Han, and K. Wang, "Feature selection using dynamic weights for classification," *Knowledge-Based Systems*, vol. 37, pp. 541–549, 2013.

[52] H. L. Chen, B. Yang, G. Wang et al., "A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method," *Knowledge-Based Systems*, vol. 24, no. 8, pp. 1348–1359, 2011.

[53] W. L. Hua, L. Lei, and J. Z. Hui, "Ensemble gene selection for cancer classification," *The Journal of the Pattern Recognition Society*, vol. 43, no. 8, pp. 2763–2772, 2010.

[54] Q. L. Shen, E. H. James, and A. A. Donald, "Random KNN feature selection - a fast and stable alternative to random forests," *BMC Bioinformatics*, vol. 12, no. 1, p. 450, 2011.

[55] J. K. Kim and H. W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms," *IEEE Transactions of Medical Imaging*, vol. 18, no. 3, pp. 231–238, 1999.

[56] D. L. Olson and D. Delen, *Advanced Data Mining Techniques*, Springer, p. 138, 2008.

[57] F. Yang, M. Hamit, A. Kutluk et al., "Feature extraction and analysis on X-ray image of Xinjiang Kazak esophageal cancer by using gray-level histograms," in *2013 IEEE International Conference on Medical Imaging Physics and Engineering*, pp. 61–65, Shenyang, China, 2013.

[58] X. M. Kong, M. Hamit, C. B. Yan, J. Sun, and J. Yao, "Fature extraction on Xinjiang high morbidity esophagus cancer based on wavelet transform," in *Biotechnology and Medical Science: Proceedings of the 2016 International Conference on Biotechnology and Medical Science*, World Scientific, p. 174, 2016.

[59] M. Hamit, F. Yang, A. Kutluk, C. B. Yan, E. Alip, and W. K. Yuan, "Feature extraction and analysis on Xinjiang high morbidity of kazak esophageal cancer by using comprehensive feature," *International Journal of Image Processing*, vol. 8, no. 4, pp. 148–155, 2014.

[60] S. X. Zhang, M. Hamit, C. B. Yan, J. Sun, and J. Yao, "Texture analysis and classification on Xinjiang kazakh esophageal cancer images," in *Biotechnology and Medical Science: Proceedings of the 2016 International Conference on Biotechnology and Medical Science*, World Scientific, p. 297, 2016.

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 1053403, 9 pages
https://doi.org/10.1155/2017/1053403

Research Article

Enabling Health Reform through Regional Health Information Exchange: A Model Study from China

Jianbo Lei,[1] Dong Wen,[2] Xingting Zhang,[2] Jiayu Li,[3] Haiying Lan,[3] Qun Meng,[4] and Dean F. Sittig[5]

[1]*Center for Medical Informatics, Peking University, Beijing, China*
[2]*Peking University Third Hospital, Beijing, China*
[3]*Health Department of Xinjin County, Chengdu City, Sichuan Province, China*
[4]*Center for Statistics and Information, National Health and Family Planning Commission, Beijing, China*
[5]*The University of Texas School of Biomedical Informatics, Houston, TX, USA*

Correspondence should be addressed to Jianbo Lei; jblei@hsc.pku.edu.cn

Received 30 November 2016; Revised 5 February 2017; Accepted 26 February 2017; Published 28 March 2017

Academic Editor: Zhengxing Huang

Objective. To investigate and share the major challenges and experiences of building a regional health information exchange system in China in the context of health reform. *Methods*. This study used interviews, focus groups, a field study, and a literature review to collect insights and analyze data. The study examined Xinjin's approach to developing and implementing a health information exchange project, using exchange usage data for analysis. *Results*. Within three years and after spending approximately $2.4 million (15 million RMB), Xinjin County was able to build a complete, unified, and shared information system and many electronic health record components to integrate and manage health resources for 198 health institutions in its jurisdiction, thus becoming a model of regional health information exchange for facilitating health reform. *Discussion*. Costs, benefits, experiences, and lessons were discussed, and the unique characteristics of the Xinjin case and a comparison with US cases were analyzed. *Conclusion*. The Xinjin regional health information exchange system is different from most of the others due to its government-led, government-financed approach. Centralized and coordinated efforts played an important role in its operation. Regional health information exchange systems have been proven critical for meeting the global challenges of health reform.

1. Background

The global healthcare community is facing many common challenges, such as the growing burden of care for the elderly, obstacles to public access to medical care, and the uneven distribution of medical resources and healthcare quality. In 1999, the Institute of Medicine (IOM) issued a landmark report, "To Err is Human" [1], citing the need for action to prevent medical errors and improve the quality of healthcare through the use of health information technology (HIT). Since then, HIT has received considerable attention around the world, along with substantial policy and funding support [2]. This interest has increasingly focused on health information exchange (HIE) projects that facilitate the access to and sharing of patient information [3, 4]. Over 280 HIE projects have been launched throughout the US, covering almost every state [5]. However, many questions remain regarding the costs, benefits, feasibility, and financing of such initiatives.

In the US market-oriented approach, HIE collaborations, known as "regional health information organizations" (RHIOs), are extremely varied and generally involve diverse stakeholders joining together to plan, finance, and implement systems for sharing electronic health information [6]. Participants may include hospitals, clinics, laboratories, pharmacies, safety net providers, payers, employers, public health departments, quality improvement organizations, and consumers. In 2009, Harvard's Adler-Milstein et al. surveyed all RHIOs in the US (207) and found that only 55 (27%) were operational, with a limited scope of data

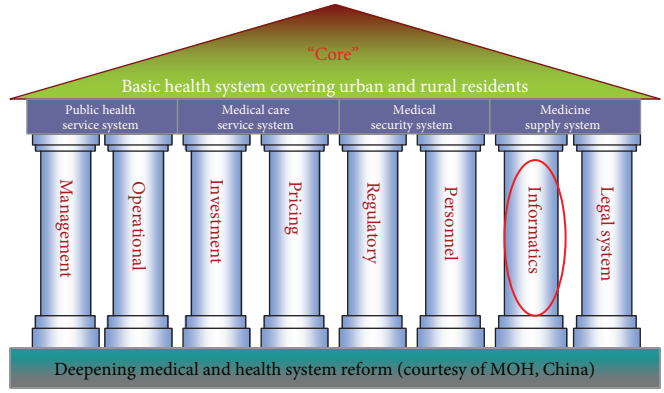


Figure 1: Core foundations (four constructs and eight pillars) of China's health reform.

exchange focused mainly on laboratory test results [7]. The authors also raised concerns about long-term, self-sustaining financial viability. It is not yet known whether the current US market-oriented approach of offering small grants and waiting to see which RHIOs flourish will work [8].

Amid China's rapid economic development and increasing population, its health system is facing enormous challenges, with accessibility, expensive services, and an uneven quality of care among the most discussed [9]. In 2009, the Chinese government launched the second stage of its ambitious healthcare reform [10]. The overall goal is to create a fair, transparent, practical, and efficient healthcare system. A prominent contribution of this reform is its description of HIT as one of the "four constructs and eight pillars," or foundations (see Figure 1) of healthcare reform [10].

This statement has brought considerable attention to HITs and to financial investment. It is estimated that the Chinese government invested more in 2010 than it had during the 60 years since 1949, when the People's Republic of China was established. This policy and financial support have enabled the construction of an electronic medical record system, a telemedicine system, and regional health information exchange (RHIE) systems. China's Ministry of Health developed a national health information development strategy for the 12th Five-Year Plan (from 2010 to 2015) called "3521" [11]. This strategy calls for the establishment of three levels (national, provincial, and municipal/regional) of health information exchange platforms; five applications (public health, medical services, a new rural cooperative medical system, an essential drug management system, and integrated patient management); two databases of electronic medical and health records; and a dedicated health network.

Health information technology has generally been considered the main vehicle of the comprehensive reform of China's primary healthcare institutions. Under the "3 Exchange Platforms" of China's "3521" national plan, regional HIE has been used as the starting point for realizing the key health reform goals: safeguarding the public welfare, stimulating patient motivation, and ensuring sustainability through the establishment of mechanisms such as standardized and orderly operations, open and transparent evaluations, comprehensive compensation and investments, and an efficient pharmaceutical supply chain.

Xinjin County in Sichuan province is one of the pilot locations of China's medical information technology reform. In accordance with the conceptualization of urban and rural centralized planning and regional integration, China has used regional health information construction to facilitate healthcare reform, achieving remarkable success. Relatively quickly and economically, Xinjin has turned itself into a successful case study of the revolutionary transformation and integration of an entire information system for all primary healthcare institutions.

2. Methods

2.1. Setting: Location and Institutional Structures of Xinjin County. Xinjin County is at the third level of China's four-level (provincial, prefecture, county, and town) administrative hierarchy. One of the 20 county-level administrative units under Chengdu city, the capital of Sichuan province, Xinjin County, is located in the western Sichuan Basin, 28 kilometers south of Chengdu city. The county covers an area of 330 square kilometers and has jurisdiction over 11 towns, one township, and 106 administrative villages (communities), with a total population of 307,800 (see Figures 2 and 3 for Xinjin's geographical location and health institutions).

Xinjin County has 198 types of medical and health institution, including eight medical and health institutions at the county level, 11 township public health centers, two community health service centers, 135 village-level medical and health institutions, five private hospitals, and 37 clinics and offices. These facilities employ a total of 2213 medical staff, many of whom lack advanced computer skills (see Figures 4–6 for the structure of Xinjin's health institutions and medical staff).

Xinjin was ranked among the top 25 of China's 100 most investment-worthy middle-to-small cities, 30th among western China's top 100 counties, and 11th in terms of county-level comprehensive economic power in Sichuan province. Xinjin was chosen as one of the first trial counties for national essential drug management system implementation in Sichuan province, as part of the medical and health system reform in Chengdu city.

Figure 2: Bird's eye view of Xinjin County.

Figure 3: Dengshuang town public health center.

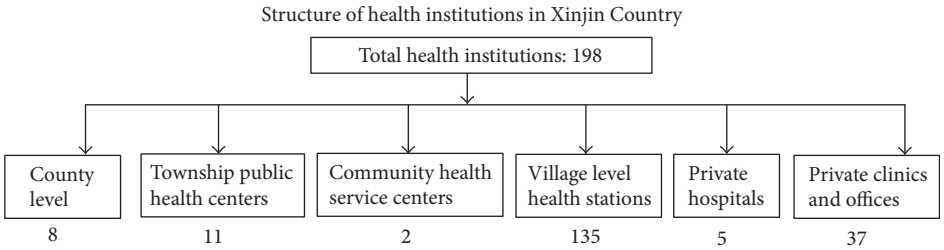


Figure 4: Structure of health institutions in Xinjin County.

2.2. Rationale of and Challenges to Reform. As a healthcare reform trial county, Xinjin is required to conduct experiments on how to achieve the general reform goal of an equal distribution of public health services and consistent quality of basic healthcare. Specifically, Xinjin must implement a standardized and normalized primary care-based system, as it is an established county-level city. Its major aims in this trial include building an advanced health service system at the county, township, and village levels; improving the quality and level of health services; and promoting the development of unified urban and rural health quality. Though Xinjin has made progress, a bottleneck persists with large discrepancies in terms of size, staff competency, and resources across its 198 health institutions. Moreover, the health institutions feature uneven, heterogeneous information technology capabilities. The lack of a unified management, monitoring system,

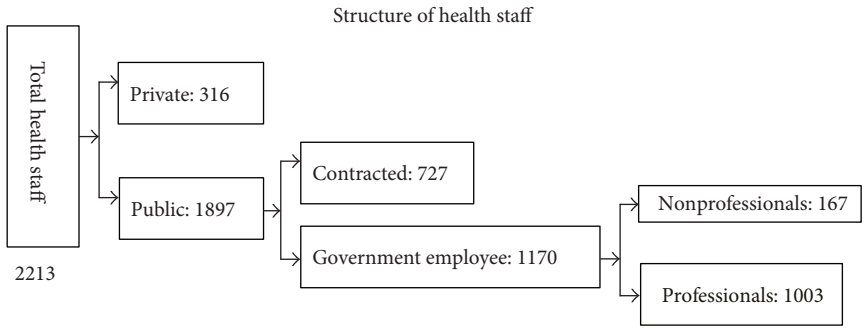


Figure 5: Structure of health staff in Xinjin County.

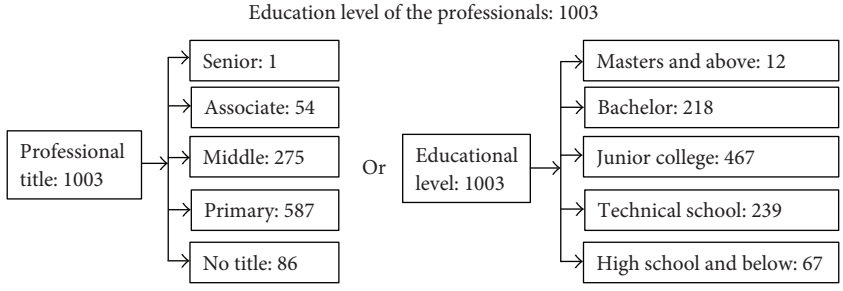


Figure 6: Structure of education level of health staff in Xinjin County.

and information sharing mechanism will make it impossible to achieve the goal of building a unified, efficient, interoperable, and shared healthcare delivery system.

Xinjin has been facing several major difficulties. The healthcare competency of its 1003 professional health staff members is poor. Less than 25% (230) have formal college education. The 198 institutions have financial, technical, and human resources of varying quality. For example, healthcare providers are unable to share patient information and administrative information across the institutions using traditional paper health records or fragmentary information systems. In addition, the level of health services and insurance coverage cannot be monitored, let alone improved. Care quality and service levels across the institutions are uneven, and the institutions are complete silos in terms of both clinical and administrative information. Meanwhile, public needs and complaints are increasing. To resolve these issues, Xinjin County started to build an RHIE with a unified, shared information system for all 198 institutions under the aegis of national health reform. An information system was needed to normalize and standardize the distribution of healthcare, to monitor and improve care quality, and to enhance the capacity of health services. Regional health information exchanges have been considered facilitators in coordinating and deepening the medical and health system reform and ultimately improving public health.

2.3. General Objectives. As a foundational part of the municipal/regional health exchange platform, the Xinjin regional HIE plan, the lowest level of the national three-level information exchange platform, has been framed as the "1131" network (one county-level data center, one health private network, three application systems [medical, public health, and county-level management platform], and one resident health card). The Xinjin regional HIE system will take an existing electronic health record (EHR) as the core and a health insurance/health card as the link to develop a shared exchange platform as the basis for the integration of electronic medical records (EMRs), drug administration, public health, two-way referral, and teleconsultation, while aiming to build a unified, efficient, effective, unblocked, and safe regional health information network. The technical architecture of the HIE system is illustrated in Figure 7.

Ultimately, each of Xinjin's 307,800 residents will receive a unique health card, a smart card issued whenever the resident first accesses health services at any health institution, enabling an unimpeded flow of information when any institution within Xinjin is visited. Patients will be able to make an appointment, register to see a doctor, and pay for their healthcare, and their clinician will be able to order medication using the card. Information on each patient-doctor encounter will be linked to other data in the county data center. Every doctor across the 198 institutions will be able to view complete health records for all patients and receive decision support from the regional HIE system throughout the healthcare process. Health administrators with the proper authority will monitor and manage the drug administration, quantity, and quality of the health services provided every day, or every month, across the whole county in real time.

2.4. Development and Implementation. As no truly successful case of a complete regional HIE system in China exists, we

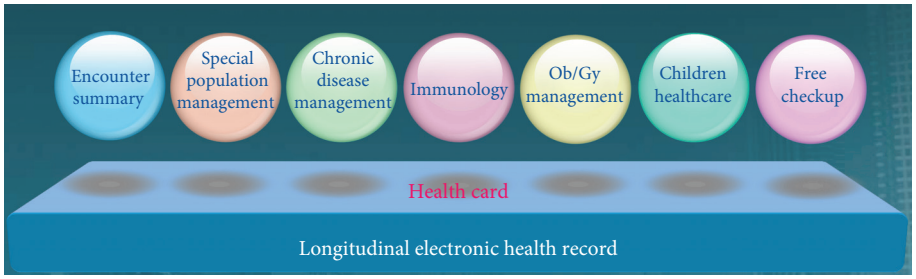


Figure 7: Technical architecture of the regional HIE system.

selected a vendor based on three major criteria. First, the vendor needed a proven record of successfully implementing hospital information systems of various sizes; second, the vendor needed successful experience with similar regional health information projects; third, the vendor needed a team of dedicated professionals who could work onsite to develop and install the systems with the help of local staff with clinical knowledge and health management experience. Through a formal open bidding process, an expert committee consisting of policy makers, informatics professionals, physicians, nurses, and health administrators chose a vendor who had developed a complete proposal for the regional HIE systems suitable for our region and our requirements based on the above three major decision factors. During the implementation period, we created a three-stage plan, in which each stage has its own goals and acceptance criteria:

Stage one (July 2008–April 2009):

(i) Build a preliminary data center, data exchange platform, and overall architecture of the HIE system.

(ii) Concurrently, develop related health information systems, such as systems for urban and rural health management, medical management, pharmaceutical management, chronic disease management, community health, maternal and child health, mental health, infectious disease management, endemic disease management, health education, health supervision, health emergency, electronic prescribing, asset finance, and personnel management.

(iii) Employ them in county and township public health agencies as well as rural community health service stations.

Stage two (June 2009–May 2010):

(i) Improve the regional data center and county-level data management platform.

(ii) Concurrently, develop and employ core systems of the HIE project, such as the health card management system, EMR system, and patient or financial management systems.

(iii) Implement interfaces between systems and deploy all of the stage-one and stage-two systems at all major health institutions.

Stage three (June 2010-2011):

(i) Continue to improve existing system functions by strengthening the data center and county-level management platform.

(ii) Develop and construct systems for drug logistics and distribution, resident self-checkup, performance appraisal, regional digital laboratory test and imaging, medical quality management, hospital queue and calling, and health checkup, as well as the patient portal.

(iii) Extend the upgraded regional health information system to the private hospitals and village health stations to achieve the final goal of integrating the management of the county's urban and rural medical and health institutions.

3. Results

Xinjin County began implementing its regional HIE project in 2008. After the three stages of construction, Xinjin County has built a preliminary regional information system covering the county health bureau and seven county-level medical institutions, 11 village and town public hospitals, two community health service centers, and more than 60 new rural community health service stations, including regional laboratory service and imaging system centers across seven medical institutions.

The operational HIE system consists of a fast, efficient, smooth, and safe regional health information network that takes residents' electronic health records as the center and provides a platform for sharing and exchanging as the base. The major module functionalities include health card management, electronic medical records, computer-based provider order entry (CPOE), regional imaging, laboratory services, drug supply, performance assessment, maternal and child healthcare, health examination, self-checkup services, disease surveillance, chronic disease management, health insurance exchange, administrative management, and e-government.

Patient information was collected at each health institution electronically through various modules and transferred to a county-level regional health exchange data center using a special internal health private network. The workflow of

Table 1: Usage of regional health information exchange system of Xinjin County.

Systems	June 2012			March 2014			
	Number of users	Total size of data (MB)	Increase per day (MB)	Number of users	Total size of data (MB)	Increase per day (MB)	Number of transactions
Outpatient doctor's station	652	12,645	17.6	662	27,904	40.3	4582
Inpatient doctor's station	652	22,694	10.1	675	46,478	43.5	1352
Out/ED billing station	30	9324	11.1	44	18.514	15.4	3458
Inpatient billing station	45	38,797	64.7	48	72,469	82.35	155
Nurse station	766	12,834	18.0	796	30.487	20.5	
Regional imaging sharing	62	10 TB	8 GB	67	20 TB	19 GB	272
Regional lab sharing	68	3.65 GB	22.4	69	10.25 GB	52.2	300/36
Regional EKG sharing	59	6.38	0.11	61	102	0.9	39
Regional quality control system	23	576.2	9.43	28	1088	12	56
Regional public health management		350 GB	10 GB	56	539 GB	52 GB	
Rational use of drugs		9130	0.03		12,500	3.5	5634
Electronic health records	Unknown (352,561 records in total)	14,711	8.23	1512	23,815	10.6	Unknown

Table 2: Comparison of costs in the three RHIO phases between the US [12] and Xinjin.

Stage of development	Costs in US	Costs in Xinjin regional HIE
Planning	\$300,000 to \$1 million	0
Development and implementation	\$3 million to \$10 million	\$2.38 M (14.8 million RMB)
Operations	\$2 million to \$5 million	\$0.16 M (100 M RMB)

each health institution was digitalized and partially optimized. Important health information such as billing and administrative data can be shared among different institutions. Patients can easily create their own health information archives, including an immunization registry and logs for health checkups, doctor visits, billing, and health information queries using a single card.

The growth and usage of the Xinjin County HIE systems have been stable since June 2012 (see Table 1). No significant increases in transactions or uploads that could bring the system down have occurred.

Within three years, Xinjin built a systematic network of services that realized, first, the digitalization of basic health services, then the electronic monitoring and management of the entire health process, and, finally, the centralized exchange of patient and administrative data to assist in accessing, sharing, monitoring, and analyzing health data and health services. The Xinjin regional HIE project is not just an exchange platform but also contains various operating information systems within each of its 198 institutions.

The implementation of the Xinjin HIE project has clearly improved health services and management, while significantly advancing the reform of China's medical and health system.

4. Discussion

4.1. Investment and Unique Organizational Model. During the first and second phases of the project, 6.33 million RMB (\$1.02 million USD) was spent, consisting of 1.23 million RMB (\$200,000 USD) on software, 4.65 million RMB (\$750,000 USD) on hardware, and 0.45 million RMB (\$70,000 USD) on network and other operating costs. Building the county data center costs 1.07 million RMB (\$170,000 USD). A total of 8.47 million RMB (\$1.36 million USD) has been invested in the third phase, comprising 1.28 million RMB (\$210,000 USD) on software, 6.15 million RMB (\$990,000 USD) on hardware, and 1.04 million RMB (\$170,000 USD) on network and other operating costs. An additional 0.77 million RMB (\$120.000 USD) has been reallocated to the data center. All of these investments were made by the local government. Further investment is inevitable, as this is a government-led project. The costs of RHIOs in the US are usually divided into three phases: planning, development and implementation, and operations [6]. Investments in the Xinjin regional HIE are smaller than regular investments in US RHIOs (see Table 2 for a comparison), and further funding should be easy to obtain.

The major difference is that the cost in the planning phase was zero for Xinjin since this was the responsibility of the government, and no external costs were incurred in this phase. In China's unique healthcare system, most of the health institutions/physicians are owned and financed by the government. For instance, 161 of Xinjin's 198 health institutions are public. Therefore, in China's regional HIE project, no complex stakeholders or public-private RHIOs were required. The government is both the owner and user of the HIE project since it owns most of the health institutions, and the users of the HIE system are government employees. The government is also the major stakeholder. The strongest driver of the regional HIE is the public good—in this case, the national health reform.

4.2. Achievements Related to Health Reform

4.2.1. Benefits to the Public: Two Major Problems Addressed. The two major problems the China health reform is intended to solve—difficulties in accessing healthcare and the high costs of doctor visits—have been resolved. Residents of Xinjin County now enjoy more standardized, accurate, simple, and convenient health services because of the digitalized health service process. Residents enjoy better health services provided by upper-level hospitals through telemedicine and regional laboratory and image centers, while paying according to their village station level. Ambulatory and inpatient costs can be reimbursed electronically in real time. Personal health information, including encounter histories, prescriptions, medications, and laboratory tests, are accessible through a patient portal in the electronic health record system. The quality of laboratory tests is monitored by a centralized laboratory service system, and laboratory results are sharable, greatly reducing redundant tests. Therefore, residents have begun to enjoy more convenient, equalized, high-quality, and low-cost health services.

4.2.2. Benefits to Medical Staff: Improved Quality of Care and Health Services. Further standardizing service processes and the clinical behavior of the medical staff have improved the quality of care. The unequal knowledge skills of medical staff across the county are addressed through online training and online knowledge databases such as the most recent diagnostic and treatment knowledge or online technical support provided by senior doctors. Service levels and health service efficiency have been improved via electronic documentation, information sharing, and the prevention of data entry duplication. Staff members' workloads and performance can be measured and evaluated to foster enthusiasm among the medical staff.

4.2.3. Benefits to Health Departments: Improved Health Agency Management and Operational Capability. First, through the EHR management and reporting systems, health departments can monitor the real-time health status of residents in the area, determine a disease spectrum, identify causes of death and health hazards, and develop interventions targeted to health promotion, thus improving health management at the appropriate decision-making level and the health of all residents. Second, through the drug supply chain and prescription statistics, the system-level health department can monitor the proportion of drug use, abnormally large prescriptions, or the inappropriate use of antibiotics to promote a rational use of drugs and reduce medication costs; these were significant issues before the health reform began. Third, through comprehensive data exchange and analysis, health administration departments can now keep track of the working status and health service satisfaction of county health institutions, allowing them to identify problems and find solutions faster. Finally, data sharing and centralized data analysis enable administrators to fully grasp the conditions and usage status of health resources, help to optimize the configuration of health resources, and improve efficiency in the use of government funds.

4.3. Lessons Learned

4.3.1. Strong Executive Support and Efficient Coordination from All Departments. Strong leadership support has been frequently cited as the primary success factor for a complex health IT project, especially a regional HIE [13]. Leadership support is relatively easy to obtain in China's cultural and political context. Under the national health reform strategy, Xinjin County's government attached great importance to the health information exchange project. Major leaders were personally involved in project decision making, and the project was made part of the county's high-priority work, annual strategy, and tangible work plans. A mechanism for coordination between two new working groups, a leadership team and working team, was established specifically for this project. The project leadership team comprised county-level leaders who were heads of the health and other relevant departments of the county government, such as the County Strategy Planning Bureau, Social and Health Insurance Bureau, and Finance Bureau; these guided, coordinated, and supervised the project. Another working group comprising health management, IT professionals, and medical staff was established to manage the execution of the project.

4.3.2. Challenge of the Great Variation in Computer and Clinical Skills within the County: Propaganda, Mobilization, and Training. One of the key challenges in implementing a high-level, technology-oriented health information project starting from grassroots-based health institutions was the uneven health resources across institutions—especially the uneven level of clinical and computer skills among healthcare staff in underdeveloped areas of the county. About one-fifth of the health staff in Xinjin County had limited computer skills. The implementation team also faced enormous resistance from staff members. The project working group thus attached great importance to propaganda, mobilization, and training. A series of comprehensive tactics were applied:

(i) First, mobilization meetings and seminars were held to change mindsets and identify and solve problems.

(ii) Second, training was offered in batches, including for computer operation and software functionality, to improve the application level.

(iii) Third, the training of network administrators was reinforced through intensive sessions and weekly meetings, and employment licenses were evaluated, thus constantly improving the capacity of the daily maintenance of the computers, software, and network of the network management personnel.

(iv) Fourth, competency training for senior work staff was strengthened. One effective way was facilitating onsite observation, learning, and exchanges, which eliminated the fear of using computers among older workers; also, designated personnel were assigned to help senior staff members solve their problems; addressing staff members' fears was especially important during the early period.

These efforts led to most health personnel being able to operate the computers and information systems. The health information technology application level has been significantly improved. The determination and support of the working team produced very positive results.

4.3.3. Design Strategies: Systematic and Centralized Planning and Unified Information System, Ensuring Interoperability and Sharing. The strong executive support enabled unified and centralized planning. The members of the working team investigated all requirements from various types of health-related staff and worked with the vendor to develop a unified infrastructure for the HIE project and subsystems. National and ministerial standards must be developed and applied to ensure interoperability and sharing. All primary, secondary, and tertiary institutions must connect to the county's unified network, and all required modules and subsystems must be provided and supported by the working and implementation team. This approach of centralized planning and unified development, deployment, and support greatly reduced costs, improved efficiency, and prevented the formation of islands of information.

4.3.4. Implementation Strategies: Refining and Decomposing Targets, Step-by-Step Implementation with Regular Supervision and Report. Methodologically, the regional HIE is a huge, complex project of system engineering. The working team decomposed the overall goals, refined them into specific phase-based aims and actions, and then assigned them to responsible entities and personnel along with quality and time requirements. Weekly progress reports and analyses were made. If there were problems, the manner and timing of follow-up actions were defined and later measured. Strong executive leadership support and effective strategies ensured the quality and time requirements for building these complex regional HIE systems.

4.4. Challenges and Lessons Learned

4.4.1. Integration Difficulties with Many Heterogeneous Systems. Health services at the county level comprise primary care, secondary care, and public services. Therefore, many special applications and modules are found across various health institutions, such as maternal and child health modules, a planned immunization module, a health surveillance system, a disease control system, insurance module, and billing module. Integration and interoperability are always a challenge for any regional or cross-departmental information system. We have partially solved this problem through unified planning and executive leadership by replacing most of the old systems with new, unified systems. This approach was feasible for Xinjin, where basic information systems were weak, as the working team was given strong executive support. Complicating the situation, however, is that various upper levels of China's public health system such as national, provincial, and municipal information systems are in their initial construction phases. Docking and integrating county information into upper-level regional information systems to ensure a wider scope of health information sharing is proven problematic. This is especially true for public health reporting and surveillance systems. A variety of data element, data exchange, interface, and even practice standards should be investigated, developed, and enforced by the central government.

4.4.2. Shortage of Professionals Greatly Affecting HIT Usage Level. The county suffers from a severe lack of interdisciplinary talent, as regional health information construction is still in its infancy in China. Very few mature and successful experiences are available to learn from. People need to understand health information technologies, the reality of health work, the nature of health management, and how to combine the above knowledge to generate constructive suggestions and ensure a smooth HIT project path with few detours. China has very few such interdisciplinary education programs. Therefore, Xinjin County, a grassroots county agency, is naturally experiencing an extreme shortage of interdisciplinary talent. Moreover, the county lacks professional computing talent. Several medical and health institutions lack professional system maintenance personnel. An incredible reality in Xinjin is that the role of network manager might be served by a doctor or accounting officer rather than an IT professional; this leaves many problems unresolved.

4.5. Next Plan. The regional HIE project has achieved its preliminary goals. Next, we will pursue the digitalization and centralization of several special health services, such as testing and reading laboratory, electrocardiography (EKG), and imaging results, which require higher-quality clinical skills. After unified and centralized laboratory, EKG, and imaging systems are implemented, the grassroots institutions will mainly be responsible for collecting laboratory specimens, taking images, and conducting EKG exams, while central and high-quality county-level institutions will handle the testing of laboratory specimens, the reading and reporting of images and EKG results, and related teleconsultation services. Meanwhile, we plan to apply wireless and RFID technology to ensure mobile and telecare and reduce medical errors. Finally, one challenging task in the pipeline is to evaluate how well the regional HIE project has been used and what health outcomes it has achieved.

5. Conclusion

An RHIE is an important vehicle for achieving the health reform policy goals of providing safe, effective, convenient, and inexpensive medical and health services. However, there are very few successful RHIE cases because of the cultural, organizational, political, systematic, and technological challenges involved. Further quantitative analyses are needed to evaluate the outcomes of such projects. However, this case study shows that a government-led, non-market-oriented, organization-wide stakeholder approach can be used to implement an RHIE project relatively quickly and with modest funding and that an RHIE helps promote health reform goals. In achieving some of the goals defined by China's second health reform, we have also accumulated experience and lessons that we hope will be valuable to others. China has formally begun the large-scale national construction of hospitals and RHIE projects. We trust that China's experiences will continue to contribute to the global community.

Abbreviations

EHR: Electronic health record
EKG: Electrocardiography
EMR: Electronic medical record
HIE: Health information exchange
HIT: Health information technology
IOM: Institute of Medicine
RHIE: Regional health information exchange
RHIO: Regional health information organization.

Additional Points

Availability of Data and Material. All relevant data are within the paper and its supporting information files.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

The work presented here was carried out in collaboration among all authors. Qun Meng supervised the design and implementation of the project. Haiying Lan, Jiayu Li, and their local team were in charge of the implementation. Jianbo Lei drafted the manuscript. Xingting Zhang and Dong Wen analyzed the data. Dean F. Sittig supervised the study design and made revisions to the manuscript. All authors have attributed to, seen, and approved the manuscript.

Acknowledgments

This study was partly supported by the National Natural Science Foundation of China (NSFC) Grant nos. 81171426 and 81471756.

References

[1] M. S. Donaldson, J. M. Corrigan, and L. T. Kohn, Eds., *To err is human: building a safer health system, Vol. 6*, National Academies Press, 2000.

[2] A. K. Jha, D. Doolan, D. Grandt, T. Scott, and D. W. Bates, "The use of health information technology in seven nations," *International Journal of Medical Informatics*, vol. 77, no. 12, pp. 848–854, 2008.

[3] J. Covich Bordenick, J. Marchibroda, and E. Welebob, *Improving the quality of healthcare through health information exchange*, Report, eHealth Initiative, Washington, DC, 2006.

[4] H. Hyppönen, J. Reponen, T. Lääveri, and J. Kaipio, "User experiences with different regional health information exchange systems in Finland," *International Journal of Medical Informatics*, vol. 83, no. 1, pp. 1–18, 2014.

[5] M. Hagland, "Health information exchange: are we at an inflection point? As health information exchange evolves forward, industry leaders and experts debate what to do about some of the continuing obstacles to granular-level data exchange. Should federal officials intervene?" *Healthcare informatics: the business magazine for information and communication systems*, vol. 30, no. 6, p. 10, 2013.

[6] Department of Health & Human Services (HHS) Office of the National Coordinator for Health Information Technology (ONC), *State HIE Strategic and Operational Plan Emerging Models*, 2011, June 2015, http://www.nationalehealth.org/sites/default/files/onc_state_hie_strategic_and_operational_plan_models_full_study-_feb_2011.pdf.

[7] J. Adler-Milstein, D. W. Bates, and A. K. Jha, "U.S. regional health information organizations: progress and challenges," *Health Affairs (Millwood)*, vol. 28, no. 2, pp. 483–492, 2009.

[8] P. Dullabh, J.-A. Milstein, C. Nye et al., *Evaluation of the State Health Information Exchange Cooperative Agreement Program: Early Findings from a Review of Twenty-Seven States*, 2012, January 2014, http://www.healthit.gov/sites/default/files/pdf/state-health-info-exchange-coop-program-evaluation.pdf.

[9] Q. Meng, L. Xu, Y. G. Zhang et al., "Trends in access to health services and financial protection in China between 2003 and 2011: a cross-sectional study," *The Lancet*, vol. 379, no. 9818, pp. 805–814, 2012.

[10] *CPC Central Committee and State Council Opinions of the CPC Central Committee and the State Council on Deepening the Health Care System Reform*, 2009, June 2015, http://news.xinhuanet.com/newscenter/2009-04/06/content_11138803_3.htm.

[11] Centre for Health Statistics Information, Ministry of Health, People's Republic of China, *The Ministry of Health Will Speed Up the Establishment of Health Information Standard in the '12th Five Year Plan' Period*, 2012, June 2015, http://www.chinadaily.com.cn/dfpd/2012-03/20/content_14870883.htm.

[12] American Hospital Association, *Health Information Exchange Projects: What Hospitals and Health Systems Need to Know*, 2006, June 2015, http://www.aha.org/content/00-10/AHARHIOfinal.pdf.

[13] A. B. Phillips, R. V. Wilson, R. Kaushal, and J. A. Merrill, "Implementing health information exchange for public health reporting: a comparison of decision and risk management of three regional health information organizations in New York state," *Journal of the American Medical Informatics Association*, vol. 21, articl e1, pp. e173–e177, 2013.

Hindawi
Journal of Healthcare Engineering
Volume 2017, Article ID 8659460, 11 pages
https://doi.org/10.1155/2017/8659460

Research Article

Therapy Decision Support Based on Recommender System Methods

Felix Gräßer,[1] Stefanie Beckert,[2] Denise Küster,[2] Jochen Schmitt,[2] Susanne Abraham,[3] Hagen Malberg,[1] and Sebastian Zaunseder[1]

[1]*Institut für Biomedizinische Technik, Technische Universität Dresden, Dresden, Germany*
[2]*Zentrum für Evidenzbasierte Gesundheitsversorgung, Universitätsklinikum Dresden, Dresden, Germany*
[3]*Klinik und Poliklinik für Dermatologie, Universitätsklinikum Dresden, Dresden, Germany*

Correspondence should be addressed to Sebastian Zaunseder; sebastian.zaunseder@tu-dresden.de

Received 11 November 2016; Accepted 23 January 2017; Published 28 March 2017

Academic Editor: Xiang Li

We present a system for data-driven therapy decision support based on techniques from the field of recommender systems. Two methods for therapy recommendation, namely, *Collaborative Recommender* and *Demographic-based Recommender*, are proposed. Both algorithms aim to predict the individual response to different therapy options using diverse patient data and recommend the therapy which is assumed to provide the best outcome for a specific patient and time, that is, consultation. The proposed methods are evaluated using a clinical database incorporating patients suffering from the autoimmune skin disease psoriasis. The *Collaborative Recommender* proves to generate both better outcome predictions and recommendation quality. However, due to sparsity in the data, this approach cannot provide recommendations for the entire database. In contrast, the *Demographic-based Recommender* performs worse on average but covers more consultations. Consequently, both methods profit from a combination into an overall recommender system.

1. Introduction

The large volume of daily captured data in healthcare institutions and out-of-hospital settings opens up new perspectives for healthcare. Due to the amount of that data, its high dimensionality and complex interdependencies within the data, an efficient integration of the available information is only possible using technical aids. In this regard, data-driven clinical decision support systems (CDSS) can be expected to take a major role in future healthcare. Generally, CDSS are designated to assist physicians or other health professionals during clinical decision-making. CDSS are demanded to be integrated into the clinical workflow and to provide decision support at time and location of care [1]. Data-driven CDSS, in particular, make use of data-mining and machine-learning techniques to extract and combine relevant information from patient data, in order to provide assistance for diagnosis and treatment decisions or even to be used in clinical quality control based on large-scale data [1].

While most works related to CDSS deal with diagnosis decision support [2, 3], predicting patient condition [4–8], or determining drug interaction [9], data-driven CDSS for therapy decision support are rare to date. This fact can be partially attributed to traditional data-mining and machine-learning techniques, which have limitations in case of missing and inhomogeneous data and often show an undesired black-box behavior.

Within this contribution, we present a system for therapy decision support based on techniques from the field of recommender systems which originates from E-commerce and has developed considerably over the last years. Recommender systems are able to overcome the aforementioned limitations of traditional data-mining and machine-learning techniques, which render suchlike systems an interesting alternative for therapy decision support. In medicine, however, the application of recommender systems is rather limited. In [10], we proposed two methods for therapy recommendation based on recommender systems' techniques,

namely, *Collaborative Recommender* and *Demographic-based Recommender*. In this work, we extend our previous work by a comprehensive evaluation of recommenders' performance in terms of accuracy and decision support capability and added a systematic comparison of similarity metrics. Additionally, various aggregation algorithms are compared differing in the way how similarity between consultations, that is, patients and their overall therapy response impacts the therapy recommendations. Finally, in extension to [10], a possible fusion approach is introduced for combining both individual recommendation engines. The proposed methods are evaluated using a clinical database incorporating patients suffering from the autoimmune skin disease psoriasis.

2. Background and Basic Taxonomies

2.1. Clinical Decision Support Systems. In general, CDSS can be classified into knowledge-based and data-driven approaches having both advantages and suffering from disadvantages. Knowledge-based systems on the one hand usually rely on manually encoded rule-based expert knowledge (if-then rules) to infer decision support. Applied rules typically represent clinical guidelines and best practice rules providing a reliable decision basis [11–13]. Disadvantage of such approaches is the bottleneck during development and updating on the basis of population-based studies and limited personalization.

Data-driven approaches on the other hand apply methods from data-mining and machine-learning to automatically extract knowledge from clinical data, facilitating more individual recommendations, learning from past experience, and revealing unknown patterns in the available data [14]. Clinical data, however, is characterized by uncertainties as well as heterogeneity (various data types), high dimensionality, and incompleteness (sparsity) [14, 15]. Such structure and characteristics put challenges on conventional machine-learning methods such as support vector machines, artificial neural networks, and decision trees. Though some conventional machine-learning techniques can cope with suchlike data properties, they require application of problem-specific a priori knowledge and highly complex models. Additionally, a crucial disadvantage of such methods in the context of CDSS is the limited interpretability (black-box behavior) of the produced results, which leads to lacking acceptance amongst health professionals. Recommender systems can overcome the limitations of traditional data-mining or machine-learning techniques, which makes them a highly interesting choice even for medical applications.

2.2. Recommender Systems and Intended Use. Recommender system technologies date back to the nineties [16–18] and are primarily intended to make personalized product suggestions based on previously recorded data on users' preferences [19]. Nowadays, recommender systems are an accepted and widespread technology used by many market leaders of various industries (e.g., Amazon (https://www.amazon.com), Netflix (https://www.netflix.com), and Spotify (https://www.spotify.com)).

Over the years, the field of recommender systems has evolved considerably yielding extremely sophisticated and specialized methods depending on domain, purpose, and personalization level [20]. Unlike conventional machine-learning methodologies, recommender systems can be capable of coping with the stated challenges associated with the data to be processed and are additionally able to permit insight into the decision-making process making results interpretable. Surprisingly, recommender systems have not found wide application in medicine so far.

A basic taxonomy of recommendation algorithms differentiates between content-based [21], collaborative filtering [19], and hybrid approaches [22]. All approaches have in common to convert estimations of a user's preference for items into recommendations using explicit or implicit previous ratings as expressions of preference. While the content-based approach links preference to item attributes, the collaborative filtering method considers the ratings of other users in the system to make personalized predictions on an active user's preference. The underlying algorithms for rating prediction and recommendation computation are based on similarity metrics which are capable of processing sparse and inhomogeneous data vectors. Furthermore, by presenting the respected data subset and impact factors, interpretation and explanation of the recommendation can be provided. Finally, the underlying databases of such systems can easily be adapted and extended which correspond to continuous adaption to new impact factors and environment in which the system is applied [20, 23].

2.3. Approaches for Therapy Decision Support. Concerning data-driven therapy or treatment recommendation in general, some scientific works were proposed, ranging from approaches based on majority voting [24], systems based on association rules [25], or applying case-based reasoning [26]. In spite of the named benefits, as stated beforehand, the application of recommender system methods in the medical context is generally limited, and there is mainly work loosely related to the idea of using suchlike methods for therapy decision support. Proposed medical applications of recommender systems are presentation of personalized health record related information for both patient and medical practitioner [27], optimized literature search for radiologists, and disease risk or mortality prediction [6, 28, 29]. Using typical recommender system methodologies for treatment recommendation, two related works are a nursing care plan [30] and an approach recommending wellness treatment [26], both applying collaborative filtering techniques.

In this work, we transfer the idea of collaborative filtering to the domain of CDSS. We present a recommender system which aims at predicting the adequacy of different therapy options for a given patient at a given time. To that end, two methodologies for therapy adequacy estimation, a *Collaborative Recommender* and a hybrid *Demographic-based Recommender*, are compared and finally combined to an ensemble of recommenders aiming at compensating for the individual algorithms' drawbacks. The exemplary recommender system is developed targeting therapy recommendations for patients suffering from the skin disease psoriasis.

Table 1: Patient describing attributes.

Attribute	Scale	Range	Availability %
Patient data			
Year of birth	Interval	1931–1998	100
Gender	Nominal	1, 2	100
Weight	Ratio	50–165	51.40
Size	Ratio	99–204	35.73
Family status	Dichotomous	0, 1	53.02
Planned child	Nominal	1, 2, 3	8.01
Year of first diagnosis	Interval	1950–2014	89.74
Type of psoriasis	Nominal	1, 2, 3, 4, 5, 6	100
Family anamnesis	Ordinal	1, 2, 3	50.95
Comorbidities			
Comorbidity	Nominal	1, 2, 3, ..., 34	—
Status	Ordinal	1, 2, 3	100
Under treatment	Dichotomous	0, 1	100
Disease-free	Dichotomous	0, 1	100
State of health			
PASI score	Ratio	0–43	70.57
Self-assessment severity	Ordinal	1, 2, 3, 4, 5	9.45
Development face	Ordinal	1, 2, 3	6.84
Development feet	Ordinal	1, 2, 3	9.81
Development nails	Ordinal	1, 2, 3	20.97
Development hands	Ordinal	1, 2, 3	12.33
Treatment contentedness	Ratio	0–10	10.62

Table 2: Therapy describing attributes.

Attribute	Scale	Range	Availability %
Systemic therapy history	Nominal	1, 2, 3, ..., 15	—
Effectiveness	Ordinal	1, 2, 3	23.67
ΔPASI	Ratio	−27–18	42
Adverse effect	Dichotomous	0, 1	100
Systemic therapy	Nominal	1, 2, 3, ..., 15	—
Effectiveness	Ordinal	1, 2, 3	98.43
ΔPASI	Ratio	−27–18	42
Adverse effect	Dichotomous	0, 1	100

3. Materials and Methods

3.1. Data. In this work, different recommender system approaches are developed and evaluated based on excerpts from health records provided by the Clinic and Polyclinic for Dermatology, University Hospital Dresden. The data consists of $V = 1111$ consultations from 213 patients suffering from various types of psoriasis. For each of the consultations, patient and therapy describing attributes are at hand, containing demographic data, information on comorbidities, and state of health as well as previous and current local and systemic treatment. The data was manually extracted from health records and stored in a MySQL database. Within a careful revision process, corrupted and invalid data was corrected or eliminated. Overall, the data comprises $A_0 = 20$ patient describing attributes for each consultation and $T = 3$ therapy describing attributes for each of the therapies applied. However, some patient describing attributes were converted into binary features for further processing which extends the patients describing feature space to $A = 123$ features.

The different attributes making up the data are of various levels of measurement ranging from dichotomous to ratio-scaled attributes. Moreover, in spite of data padding in cases where information was missing but could be assumed to be constant over consultations, availability of certain attributes is very limited. As a consequence, the resulting data matrices are characterized by inhomogeneity and sparsity. Patient attributes and therapy information are summarized in Tables 1 and 2 along with scale of measurement, range of values, and relative availability, respectively. Attributes that were present in less than five percent of the available data were neglected.

Previous treatment is the collection of all relevant therapies applied to a patient up to the consultation under consideration, whereas in the current treatment database, all therapies are collected which were applied within the last two weeks preceding the respective consultation. Even though there is information on both local and systemic therapies available, this study focuses on recommending the most effective systemic therapy out of $M = 15$ available therapies of this type for a given consultation. For both previous and applied therapies, up to three outcome indicators are intermittently given: (1) a therapy effectiveness indicator (bad, medium, and good), representing the subjective assessment, (2) an objective health state improvement indicator, and (3) adverse effects (yes, no). The health state improvement indicator relates to the severity of psoriasis quantified by the Psoriasis Area and Severity Index (PASI). PASI ranges from 0 (no disease) to 72 (maximal disease) combining both the skin area affected and the severity of lesions in a single score [31]. Change of PASI between two consecutive consultations (ΔPASI) is assigned as objective health state improvement indicator attribute to all therapies applied between those consultations where the change occurred.

3.2. Methodology

3.2.1. System Overview. The algorithms described in the following aim at recommending the potentially most effective systemic therapy for a given patient and consultation. The collaborative filtering idea is transferred to the therapy recommendation domain, considering therapies as items and therapy response as a user's preference. For representing therapy response, effectiveness, ΔPASI, and absence of adverse effects are incorporated.

In a preceding prediction step, individual therapy outcome is estimated for all available therapies that have not yet been applied to the patient. The outcome estimate is computed based on the therapy response of the nearest neighbors to the consultation under consideration. At this stage, similarity computation between consultation representations

plays an essential role. The two recommender approaches proposed in this work differ in the information used to represent consultations. The applied *Collaborative Recommender* algorithm uses solely outcome from all preceding consultations to represent a consultation. The hybrid *Demographic-based Recommender* approach on the other hand is taking additionally all available patient describing data into account. In the subsequent recommendation step, the therapies are ranked according to their response estimates and the top N-ranked entry or entries are recommended. Both recommender engines suffer from drawbacks depending on the data properties which the other approach is capable of compensating for. Therefore, an ensemble of recommenders is introduced combining both recommender engines. In the following sections, computation of therapy response estimates is detailed, both recommender approaches along with the applied similarity metrics are described and the actual outcome prediction algorithm is presented. Finally, the ensemble model is described and the applied evaluation metrics are introduced.

3.2.2. Affinity Model. In typical recommender system applications, data reflecting a user's preferences is collected from both explicit and implicit input. Where explicit expression of preference usually is provided as item ratings or votes on items, implicit information can be derived from clicked items, items being part of the shopping basket, or visited pages, respectively. Here, the preference to a therapy is derived from the therapy response. The mathematical quantification of therapy response, in the following denoted as *affinity*, is modeled using a weighted sum of three parameters attained from subjective effectiveness ($f_{1,v,m}$), change of the PASI score (ΔPASI), that is, current PASI compared to the PASI of the previous consultation ($f_{2,v,m}$), and adverse effects ($f_{3,v,m}$). The individual components impact on the overall affinity measure can be varied by adjusting weight w_t. As was figured out in own previous analysis, particularly the *Collaborative Recommender* approach benefits from decreasing influence of adverse effects. Consequently, the adverse effect weight was set to $w_3 = 0.25$, whereas both effectiveness indicators have equal weight $w_{1/2} = 1$. As can be seen in Table 2, the three attributes are just intermittently available, meaning that not all applied therapies are provided with all three components. Missing data is respected by normalization with the sum of weights w_t, where $\delta_{t,v,m}$ is set to 0 in case of missing values and set to 1 otherwise. Thus, the affinity of a given patient and consultation $v \in V$ to a therapy $m \in M$ is modeled as

$$r_{v,m} = \frac{\sum_{t=1}^{T} \delta_{t,v,m} \cdot w_t \cdot f_{t,v,m}}{\sum_{t=1}^{T} \delta_{t,v,m} \cdot w_t}, \tag{1}$$

where the two effectiveness related affinity components (subjective effectiveness and ΔPASI score) are mapped to the domain $0 \le f_{t,v,m} \le 1$ according to the following specifications. Effectiveness is factorized with a constant value resulting in the nominal values 0.25 (poor), 0.5 (moderate), and 0.75 (good). The ΔPASI score is mapped by a negative sigmoid function adjusted to the domain 0.1 as shown in Figure 1 facilitating large impact on small PASI score variations declining with increasing absolute value. Finally, the binary adverse effect indicator is mapped to −0.25 if any adverse effect is present and 0 otherwise to penalize the overall affinity measure if adverse effects have occurred. The mapping rules can be summarized as

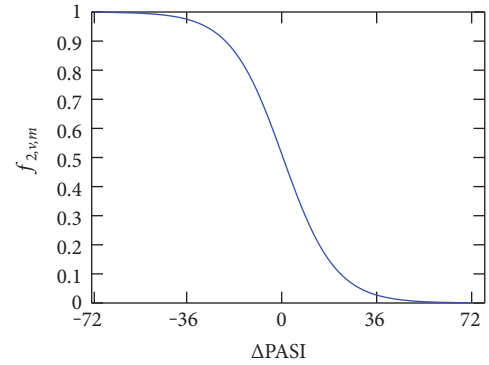


Figure 1: Sigmoid function mapping the ΔPASI score to the domain $0 \le f_{t,v,m} \le 1$.

$$f_{1,v,m} = 0.25 \cdot \text{effectiveness}, \tag{2}$$

$$f_{2,v,m} = 1 - \frac{1}{1 + e^{-0.1 \cdot \Delta\text{PASI}}}, \tag{3}$$

$$f_{3,v,m} = \begin{cases} -0.25 & \text{if adverse effect} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

3.2.3. Similarity Metrics. Both proposed methods, *Collaborative Recommender* and *Demographic-based Recommender*, are related to user-user collaborative filtering [20] which correlate between consultations as users to recommend therapies as items. The basic idea is to make affinity estimations for (a subset of) therapies which were not yet applied to a patient. For affinity estimation, a neighborhood-based algorithm is used [32]. Therefore, similarity computation between consultations has critical impact on the recommender engine's output and is highly dependent on the consultation representation. Similarity $w_{v,k}$ between two consultations $v \in V$ and $k \in V$ is calculated for all consultation pairs considering only therapies $m \in M$ which are corated in both consultations. Depending on the information which is used for consultation representation, a distinction between the two aforementioned recommendation techniques and consequently the similarity metrics applied can be made. In this section, various similarity metrics are detailed which were studied in this work for both, the *Collaborative Recommender* approach and *Demographic-based Recommender* approach.

(1) Collaborative Recommender. In this approach, the consultation under consideration is only represented by the affinity values related to therapies applied up to this consultation. The underlying assumption is that the therapy applied to a given patient within the therapy history and

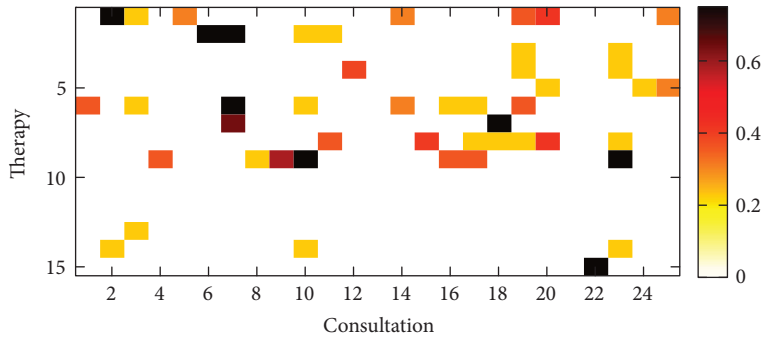


Figure 2: Affinity matrix for 25 randomly selected consultations.

the associated outcome reincorporates information about that respective patient and consultation which can then be transferred to patients with similar therapy history. Here, all attributes respected for similarity computation, that is, affinity entries for previously applied therapies, are of ratio scale and within the same range of values. However, the affinity matrix is characterized by only intermittently available entries, that is, resulting in sparse data vectors to be compared. Three similarity metrics proposed in the recommender system literature [32] are investigated and compared in this work. *Vector Similarity*,

$$w^1_{v,k} = \frac{\sum_{m\in M} r_{v,m}\cdot r_{k,m}}{\sqrt{\sum_{m\in M} r^2_{v,m}}\cdot\sqrt{\sum_{m\in M} r^2_{k,m}}}, \quad (5)$$

which originates from information filtering using vector space models, is widely used in collaborative filtering algorithms. Vector Similarity simply computes the cosine of the angle between two vectors $r_{v,m}$ and $r_{k,m}$ representing two consultations $v \in V$ and $k \in V$, respectively. Furthermore, the degree of linear relationship between two vectors can be quantified using the *Pearson Correlation*,

$$w^2_{v,k} = \frac{\sum_{m\in M}\left(r_{v,m}-\overline{r_v}\right)\cdot\left(r_{k,m}-\overline{r_k}\right)}{\sigma_v\cdot\sigma_k}, \quad (6)$$

derived from a linear regression model. This metric relies on the assumption that a linear relationship must exist and the errors are independent and have a probability distribution with zero mean and constant variance. However, these assumptions are often violated in the context of collaborative filtering data which can deteriorate the outcome accuracy. To overcome these stated model assumptions, the *Spearman Rank Correlation*,

$$w^3_{v,k} = \frac{\sum_{m\in M}\left(\mathrm{rank}_{v,m}-\overline{\mathrm{rank}_v}\right)\cdot\left(\mathrm{rank}_{k,m}-\overline{\mathrm{rank}_k}\right)}{\sigma_v\cdot\sigma_k}, \quad (7)$$

computes a measure of correlation between ranks of the individual therapies' affinity values instead of the affinity values directly.

Figure 2 depicts an exemplary affinity matrix excerpt. Here, the affinity measure derived from the available information on therapy outcome from 25 randomly selected consultations is shown. As can be seen, only a small fraction of the 15 available therapies were applied per patient and consultation. Additionally, only a reduced number of therapy preference parameters are available for those applied therapies. As stated beforehand, this results in an extremely sparse affinity matrix for consultation representation which is relied upon when computing similarity. Additionally, this approach suffers from the so-called *cold start* problem occurring when a new patient is included into the system providing no therapy history at all. However, lacking information makes it difficult or even impossible to find appropriate similar consultations [20].

One approach to address the trust that can be placed in the similarity to a neighboring consultation, depending on the available information, is significance weighting. In case of the *Collaborative Recommender* approach, a significance weight depending on the number of co-rated therapies n (out of N) included in the computation between consultations is applied. Thus, if two consultations to be compared have fewer than b therapies (affinity entries) in common, the similarity weight is decreased, respectively. However, results showed that the *Collaborative Recommender* is not benefiting from significance weighting in this application.

$$w_{v,k} = \begin{cases} \frac{n}{N}\cdot w_{v,k} & n \le b \\ w_{v,k} & \text{otherwise.} \end{cases} \quad (8)$$

(2) Demographic-Based Recommender. To overcome the limitations related to the above described collaborative filtering approach concerning lacking information and cold start, the *Collaborative Recommender* is extended to utilizing all patient describing information summarized in Table 1 to represent a consultation. The straightforward underlying assumption here is that the available patient describing data carries sufficient information for facilitating meaningful comparisons between consultations. However, as already stated in Section 3.1, the patient describing data employed for consultation comparison is not only sparse but the attributes involved into the

similarity calculation are characterized by inhomogeneity, that is, are of various level of measurement (dichotomous, nominal, ordinal, interval, and ratio scaled). The similarity measure utilized in this work facilitating both, handling missing values and varying levels of measurement, is the *Gower Similarity Coefficient* [33]. Here, the level of measurement of the individual attributes is respected for each attribute comparison. Furthermore, the Gower coefficient offers the opportunity to control the individual attribute's impact on overall similarity by assigning specific weights to attributes. Thus, the overall similarity $w_{v,k}$ between two consultations is computed including the individual attribute similarities $\rho_{a,v,k}$ depending on their presence $\delta_{a,v,k}$ and assigned weights w_a as

$$w^4_{v,k} = \frac{\sum_{a \in A} \delta_{a,v,k} \cdot w_a \cdot \rho_{a,v,k}}{\sum_{a \in A} \delta_{a,v,k} \cdot w_a}. \quad (9)$$

Finally, the computed overall similarity is normalized with the sum of weights w_a of present values by setting the $\delta_{a,v,k}$ to 0 in case of missing values and 1 otherwise, respectively. The data type-specific similarity coefficients $\rho_{a,v,k}$ are defined as follows: For similarity computation between ordinal, interval, and ratio-scaled values, the Manhattan distance normalized to the attribute range is utilized, whereas for nominal or dichotomous attributes, simple matching (M-coefficient) or the Jaccard similarity (S-coefficient) coefficients are applied, respectively.

Significance weighting in case of the *Demographic-based Recommender* showed slight improvement over the unweighted similarity between consultations. Here, similarity is penalized in case of fewer than $b = 30$ common attributes from the extended feature space A.

3.2.4. Affinity Estimation. To generate an affinity estimate on appropriate therapies for a consultation under consideration, various methods for computing aggregates of neighboring consultations' therapy response, that is, affinity, are compared. Here, an affinity estimate $r_{v,m}$ for a consultation under consideration $v \in V$ and therapy $m \in M$ is calculated according to three different calculation rules, differing in the way the affinity distribution is accounted for. When computing the affinity prediction, previous consultations of the same patient need to be discarded. The basic way for affinity estimation is computing a simple average over all K most similar consultations $k \in K$

$$p^1_{v,m} = \frac{\sum_{k \in K} r_{k,m}}{|K|} \quad (10)$$

which serves as baseline estimate. To account for the similarity between consultations and to control influence on the outcome, the above formula is extended estimating affinity as weighted average

$$p^2_{v,m} = \frac{\sum_{k \in K} r_{k,m} \cdot w_{v,k}}{\sum_{k \in K} |w_{v,k}|} \quad (11)$$

assuming that all consultation affinity entities have the same distribution. Under the assumption, that therapy outcome for different consultations, that is, patients, is centered around different means, the weighted average of deviations from the neighboring consultations' means is computed. Thus, the affinity for a consultation under consideration and therapy is determined by adding this average deviation across all respected neighboring consultations to the consultation under consideration's mean affinity

$$p^3_{v,m} = \overline{r_v} + \frac{\sum_{k \in K} \left(r_{k,m} - \overline{r_k}\right) \cdot w_{v,k}}{\sum_{k \in K} |w_{v,k}|}. \quad (12)$$

For all introduced approaches, the summations are performed over the K most similar consultations $k \in K$ as predictors for therapy m with $w_{v,k}$ being the weight between consultations v and k, representing similarity, and $r_{k,m}$ being the affinity during consultation k on therapy m. The size of an adequate subset of consultations, that is, the number of nearest neighbors k included in the computation, is crucial and needs to be chosen cautiously. Two approaches are possible for the application at hand. Correlation thresholding on the one hand determines the respected neighborhood size according to a predefined threshold assuming that highly correlating consultation are more valuable predictors. However, in this work, the best K neighbors approach is applied which considers a predefined number K of predictor consultations as consultation subset size. Both approaches suffer from reduction in performance and prediction quality due to noise when too many consultations are included into the estimate computation. In contrast, having too few neighbors for affinity estimation, the coverage of available therapies can be very low.

3.2.5. Recommender Ensemble. As stated beforehand, both proposed recommender approaches have their strengths and weaknesses. The idea of building a recommender ensemble is to generate an overall recommendation which combines both approaches while compensating for the individual recommender engines' drawbacks. Fusing decisions in machine-learning applications, denoted as ensemble learning, have shown to be capable of outperforming basic algorithms [34]. The extensive benefit of decision fusion has also been proven in the Netflix Grand Prize in 2009 where a combination of algorithms finally achieved the largest accuracy improvement (http://www.netflixprize.com). In the contrast to fusion of two competitive recommender engines as proposed in [35] to benefit from as much information as possible, selecting one recommender assumed to be expert under certain constraints is applied here to overcome the stated cold start problem.

3.3. Evaluation. In this work, two different evaluation metrics are considered. On the one hand, the individual recommender engine or system yields to predict the response to specific therapies. If the prediction meets the real therapy response, the system can provide the medical practitioner with a reliable support for his decision-

making based on the estimation. To quantify the difference between estimated response and real response, the root mean squared error (RMSE) for a specific consultation is computed between provided affinity entries and predictions. RMSE reflects the rating error in the same value domain as the actual affinity measure with large errors having more impact [32].

$$\mathrm{RMSE}_v = \frac{1}{M_v}\sqrt{\sum_{m=1}^{M_v}\left(p_{v,m} - r_{v,m}\right)^2}. \quad (13)$$

On the other hand, N top-ranked therapies are usually selected from the affinity predictions for a consultation under consideration and presented to the user. For evaluating recommendation quality, decision support accuracy metrics are commonly utilized as described in [36]. In our application, the objective is to evaluate how effective the recommender engine or system is at helping a medical practitioner to select an appropriate therapy from the set of all therapies which is essential for boosting acceptance of suchlike systems. The challenge in the context of therapy recommendation is the only partially observed ground truth as described in [37]. Thus, an outcome-driven precision metric as additional decision support metric is defined as follows. Initially, all consultations are divided into those cases where at least one of the top N therapies recommended were actually applied in the respective consultation and those where the therapies were not compliant. Furthermore, therapies were considered having good outcome if the affinity assigned to that therapy complies with $r_{v,m} \geq 0.5$ and $r_{v,m} < 0.5$ otherwise which leads to the definitions summarized in Table 3.

The outcome-driven precision describes the ratio of all therapies recommended by the system for a consultation $v \in V$, that is, top N therapies, which were applied and show good response, that is, are considered successful, and is defined as

$$\mathrm{Precision@}N_v = \frac{TP_v}{TP_v + FP_v}. \quad (14)$$

4. Results and Discussion

In the following, the three affinity estimation approaches introduced in Section 3.2.4 are compared for both the *Collaborative Recommender*, employing all three proposed similarity metrics, and the *Demographic-based Recommender*. Figures 3, 4, and 5 demonstrate affinity estimation error and recommendation precision for all proposed affinity prediction algorithms. Both evaluation metrics are averaged over all consultations for each distinct number of the nearest neighbors k being evaluated. In case of the *Collaborative Recommender* approach, the estimated affinity can only be compared, that is, an estimation error computed, if (1) therapy describing data from previously applied therapies is available. Without such information no similarity and consequently no affinity estimate and recommendation can be computed. (2) Therapy describing data for therapies applied in the consultation under consideration are available. The ratio of consultation for which both requirements are met amounts to 76.87% of all consultations in the database. Furthermore, prerequisite for affinity estimate evaluation, that is, RMSE computation, are common therapies for which affinity entries are available in both estimation and applied therapy vector. The relative number of consultations for which this requirement is met and the RMSE is computed from increases with the size of the neighborhood k as demonstrated in Figure 6 for both recommender approaches.

Table 3: Outcome-driven evaluation definitions.

	Good outcome	Bad outcome
Recommendations compliant	*TP*	*FP*
Recommendations not compliant	*FN*	*TN*

Regarding recommendation precision, the ground truth is obtained from all consultations having one or more therapies which showed good response according to the definition described in Section 3.3. As a result, precision can only be computed for 67.24% of all consultations in the database for the *Demographic-based Recommender* and 61.48% for the *Collaborative Recommender* approach.

In both nonnormalized *Collaborative Recommender* approaches, the affinity estimation error quickly declines for all studied similarity metrics with increasing k. Both approaches have a minimum at around $k = 75$ considered nearest neighbors. The *Demographic-based Recommender* affinity prediction error is significantly higher. However, the *Demographic-based Recommender's* performance improves asymptotically with rising k. Concerning recommendation precision, both nonnormalized *Collaborative Recommender* approaches perform very robust on up to around 150 nearest neighbors for cosine similarity and the Pearson correlation. The *Demographic-based Recommender* again provides substantially lower precision with a maximum at a very small neighborhood of around $k = 20$. In case of the *Collaborative Recommender*, the Spearman correlation for computing similarity between consultations is able of having a minor advantage when estimating affinity, but the recommendation precision is only for a very small k as good as the other similarity metrics. Particularly Pearson correlation clearly shows the advantage of weighting the impact of neighbors according to their similarity compared to the nonweighted averaging approach. By considering similarity as weight, the influence of noise introduced by more distant neighbors is reduced resulting in smaller prediction errors and better recommendation precision for large neighborhoods. However, the nonweighted averaging approach approximates asymptotically the mean affinity for each of the therapies as affinity estimate for large k and yields only low precision. Normalization by adding deviations from average affinity for each respected neighbor to the mean affinity for each consultation under consideration is not improving the computed estimate. This is mainly due to the fact that average is computed here for only a very small number of applied therapies having no representative meaning on overall outcome tendency for a consultation under consideration.

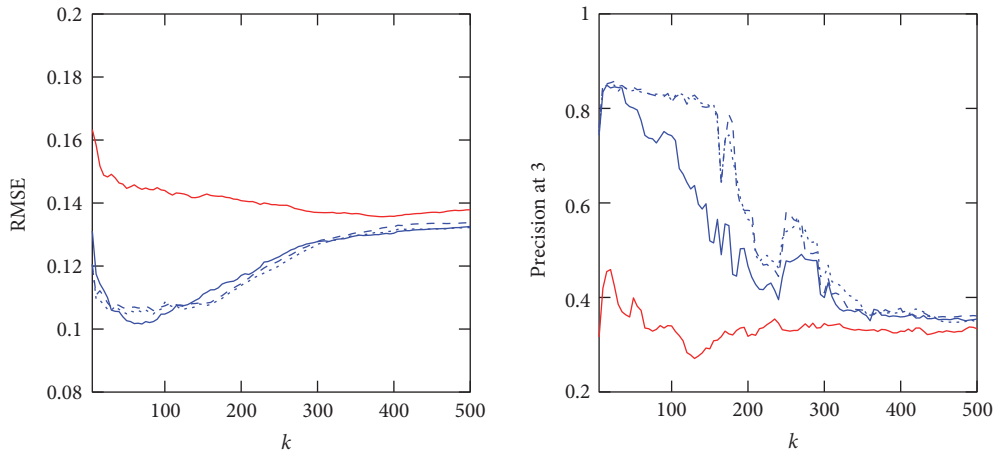


Figure 3: RMSE and precision at 3 over $5 \leq k \leq 500$ of *Demographic-based Recommender* and *Collaborative Recommender* employing Gower (red line), cosine (blue dash line), Pearson correlation (blue dotted line), and Spearman correlation (blue line) for similarity computation and averaging for affinity estimation.

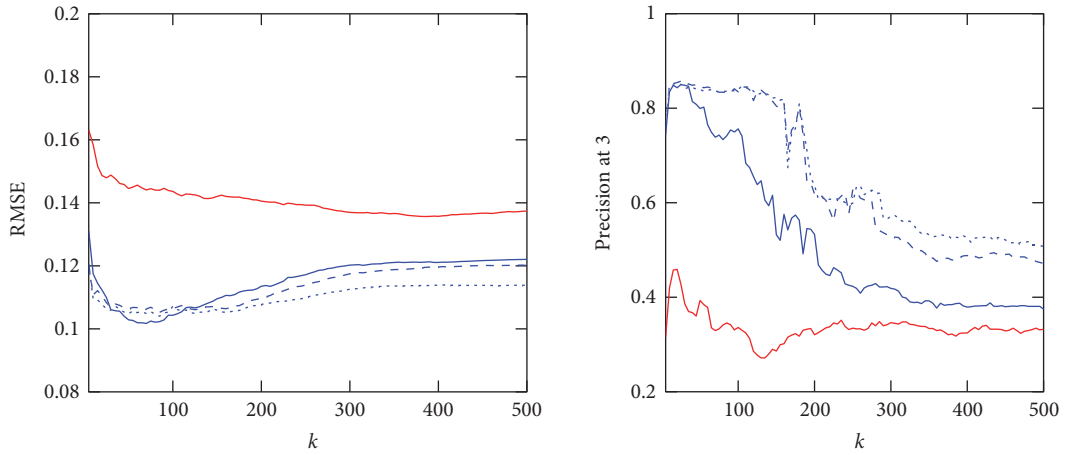


Figure 4: RMSE and precision at 3 over $5 \leq k \leq 500$ of *Demographic-based Recommender* and *Collaborative Recommender* employing Gower (red line), cosine (blue dash line), Pearson correlation (blue dotted line), and Spearman correlation (blue line) for similarity computation and weighted averaging for affinity estimation.

As stated beforehand, for the *Collaborative Recommender* affinity estimation and consequently therapy recommendation can only be generated if therapy describing data from previous therapies is available for neighborhood selection. In the provided data, only 85.33% of all consultations are applied with such data from previous therapies. To overcome this cold start limitation, the *Demographic-based Recommender* can adopt the recommendation task for the remaining 14.67% consultations being not dependent on historical therapies and outcome information. In the following, for all consultations providing previous therapy information, recommendations are computed using the weighted averaging *Collaborative Recommender* approach. For similarity computation, Pearson correlation is employed yielding the best precision performance in the neighborhood of $k = 100$ neighbors. All recommendations which could not be provided by the *Collaborative Recommender* due to missing information were imputed by a weighted averaging *Demographic-based Recommender* showing the best performance with $k = 20$. This approach is capable of compensating the deficiency in recommendations due to lacking information resulting in recommendations for all consultations in the available data. Using this ensemble approach, a mean precision over all consultations of 79.78% could be obtained. In comparison, utilizing average affinity for each therapy in the entire database to determine recommendations to compensate the *Collaborative Recommender* gaps, that is, recommending the therapies ranked by their mean response, overall precision of 78.51% is yielded.

5. Conclusion

In this work, the application of recommender system algorithms in the context of therapy decision support was studied. Even though there is an extensive impact of recommender systems in other domains, application of suchlike

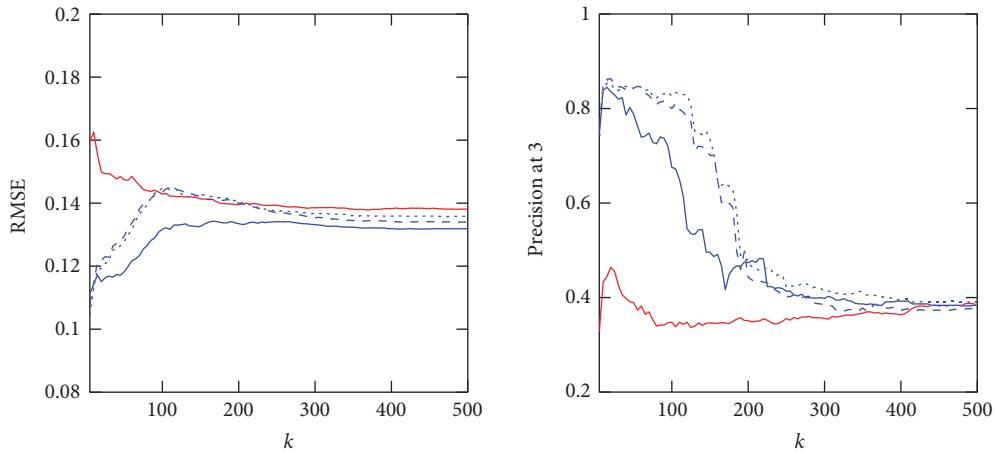


Figure 5: RMSE and precision at 3 over $5 \leq k \leq 500$ of *Demographic-based Recommender* and *Collaborative Recommender* employing Gower (red line), cosine (blue dash line), Pearson correlation (blue dotted line), and Spearman correlation (blue line) for similarity computation and normalized weighted averaging for affinity estimation.

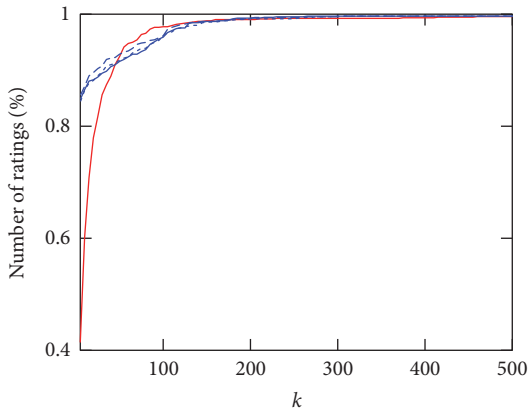


Figure 6: Relative number of consultations for which affinity estimation RMSE can be computed depending on $5 \leq k \leq 500$ for *Demographic-based Recommender* and *Collaborative Recommender* employing Gower (red line), cosine (blue dash line), Pearson correlation (blue dotted line), and Spearman correlation (blue line) for similarity computation.

approaches in healthcare are—to the best of our knowledge—still rare to date. Dependent on the data employed for determining similarity between consultations and therapy outcome estimation, two approaches were compared. For both algorithms, a *Collaborative Recommender* approach and *Demographic-based Recommender*, various variations were studied concerning similarity metric, considering credibility of the computed similarity and aggregation of the respected information for estimating potential therapy response. All algorithms were evaluated using the accuracy of predicting the outcome and according to the precision with which the top 3 recommended therapies meet the ground truth. However, as ground truth only therapies were accepted which have shown good response, therapies with bad outcome were neglected. Therefore, evaluation could only be done on a fraction of the already rather limited database. However, it is assumed that the performance of the approaches studied in this work depend substantially on the amount of available data and will improve considerably if scaled to larger datasets.

The *Collaborative Recommender* utilizing basic collaborative filtering algorithms, considering only therapy outcome from previously applied therapies for consultation representation, outperforms the *Demographic-based Recommender* approach. The weighted averaging *Collaborative Recommender* method taking the similarity weight into account demonstrates better performance than simple averaging over all neighborhood sizes studied in this work. Normalization with respect to deviations from average response for individual consultations performs significantly worse. Concerning similarity metrics, the Pearson correlation shows the best results by exceeding both the cosine similarity and the Spearman rank correlation especially with increasing size of the respected neighborhood.

Applying patient describing data for similarity computation, that is, the *Demographic-based Recommender* approach, does not show a comparable performance to the *Collaborative Recommender*. The similarity computation underlying the *Demographic-based Recommender* is affected unfavorably by less relevant information included into the calculation whereas more important factors have too little effect. Therefore, future work will concentrate on improving the *Demographic-based* method using features selection and weighting methods and learning optimized similarity metrics. Including more information into the recommendation is essential to overcome the limitations in cases where no data on therapy history or just little information is available for a specific patient under consideration as was demonstrated. A simple combination of both recommender approaches was generated which replace each other depending on the available information. Therewith, the cold start problem could be overcome and recommendations provided for consultations having no information on therapy history. In future work, more sophisticated hybrid [22], time-aware

approaches considering feature and preference evolution [38], and recommender ensembles will be studied incorporating information from both approaches into the entire recommendation process.

Further on, beyond the presented comparison of different recommender-based approaches, a comparison of the proposed methods to alternative machine-learning algorithms for generating therapy recommendations, particularly model-based approaches, would be of high interest. However, one of the major reasons to apply recommender methods are their capability to handle heterogeneous and sparse data. The application of typical model-based approaches, in turn, is difficult as structure and characteristics of the clinical data at hand, that is, its high degree of heterogeneity and sparsity, would require extensive feature transformation and preparation (handling of missing data, transformation of non-interval-scaled data). As suchlike preparation is complex and will heavily impact the performance of machine-learning algorithms, the usage of such techniques and their comparison to the recommender approaches exceeds the extent of this work. However, future works will address this issue considering the presented clinical data but also using other data in order to yield a comparative assessment of the proposed methods.

Disclosure

This research article is an extended version of the conference paper presented at the 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The work is part of the project *Therapieempfehlungssystem* which is funded by the *Roland Ernst Stiftung für Gesundheitswesen*. The authors further acknowledge the support by the German Research Foundation and the Open Access Publication Funds of the TU Dresden.

References

[1] P. E. Beeler, D. W. Bates, and B. L. Hug, "Clinical decision support systems," *Swiss Medical Weekly*, vol. 144, no. w14073(5), pp. 1–7, 2014.

[2] H. Alder, B. A. Michel, C. Marx et al., "Computer-based diagnostic expert systems in rheumatology: where do we stand in 2014?" *International Journal of Rheumatology*, vol. 2014, Article ID 672714, 10 pages, 2014.

[3] W. F. Bond, L. M. Schwartz, K. R. Weaver, D. Levick, M. Giuliano, and M. L. Graber, "Differential diagnosis generators: an evaluation of currently available computer programs," *Journal of General Internal Medicine*, vol. 27, no. 2, pp. 213–219, 2012.

[4] A. Agrawal, S. Misra, R. Narayanan, L. Polepeddi, and A. Choudhary, "A lung cancer outcome calculator using ensemble data mining on SEER data," *Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics (BIOKDD '11),* pp. 1–9, 2011.

[5] E. Choi, M. T. Bahadori, and J. Sun, "Doctor AI: predicting clinical events via recurrent neural networks," arXiv preprint arXiv:1511.05942, 2015.

[6] D. A. Davis, N. V. Chawla, N. A. Christakis, and A. L. Barabasi, "Time to CARE: a collaborative engine for practical disease prediction," *Data Mining and Knowledge Discovery*, vol. 20, no. 3, pp. 388–415, 2010.

[7] F. Folino and C. Pizzuti, "A recommendation engine for disease prediction," *Information Systems and e-Business Management*, vol. 13, no. 4, pp. 609–628, 2015.

[8] M. Zhu, W. Chen, J. P. Hirdes, and P. Stolee, "The K-nearest neighbor algorithm predicted rehabilitation potential better than current clinical assessment protocol," *Journal of Clinical Epidemiology*, vol. 60, no. 10, pp. 1015–1021, 2007.

[9] G. Kuperman, A. Bobb, T. Payne et al., "Medication-related clinical decision support in computerized provider order entry systems: a review," *Journal of the American Medical Informatics Association*, vol. 14, no. 1, pp. 29–40, 2007.

[10] F. GraBer, S. Beckert, D. Kuster et al., "Application of recommender system methods for therapy decision support," *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom),* pp. 1–6, 2016.

[11] P. A. De Clercq, J. A. Blom, H. H. M. Korsten, and A. Hasman, "Approaches for creating computer-interpretable guidelines that facilitate decision support," *Artificial Intelligence in Medicine*, vol. 31, no. 1, pp. 1–27, 2004.

[12] E. Kilsdonk, L. W. Peute, R. J. Riezebos, L. C. Kremer, and M. W. Jaspers, "From an expert-driven paper guideline to a user-centred decision support system: a usability comparison study," *Artificial Intelligence in Medicine*, vol. 59, no. 1, pp. 5–13, 2013.

[13] M. Musen, Y. Shahar, and E. Shorliffe, "Clinical decision-support systems," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicin*, pp. 643–674, Springer, New York, 2006.

[14] D. Windridge and M. Bober, "A kernel-based framework for medical big-data analytics," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-Of-The-Art and Future Challenges*, pp. 197–208, Springer, Berlin Heidelberg, 2014.

[15] J. A. Doucette, A. Khan, R. Cohen, D. Lizotte, and H. M. Moghaddam, "A framework for AI-based clinical decision support that is patient-centric and evidence-based," 2012.

[16] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[17] J. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "GroupLens: applying collaborative filtering to Usenet news," *Communications of the ACM*, vol. 40, no. 3, pp. 77–87, 1997.

[18] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth"," *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '95),* pp. 210–217, 1995.

[19] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proceedings*

of the 10th International Conference on World Wide Web, pp. 285–295, 2001.

[20] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 3, no. Section 3, pp. 1–19, 2009.

[21] M. J. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5, pp. 393–408, 1999.

[22] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[23] F. Ricci, L. Rokach, B. Shapira, and P. Kantor, *Recommender Systems Handbook*, Springer, US, 2011.

[24] H. Liu, G. Xie, J. Mei, W. Shen, W. Sun, and X. Li, "An efficacy driven approach for medication recommendation in type 2 diabetes treatment using data mining techniques," *Medinfo 2013: Proceedings of the 14th World Congress on Medical and Health Informatics, parts 1 and 2,* vol. 192, 1071 pages, 2013.

[25] J. H. Chen and R. B. Altman, "Automated physician order recommendations and outcome predictions by data-mining electronic medical records," *AMIA Summits on Translational Science proceedings,* pp. 206–210, 2014.

[26] T. P. Lim and W. Husain, "Recommender system for personalised wellness therapy," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 9, pp. 54–60, 2013.

[27] M. Wiesner and D. Pfeifer, "Health recommender systems: concepts, requirements, technical basics and challenges," *International Journal of Environmental Research and Public Health*, vol. 11, no. 3, pp. 2580–2607, 2014.

[28] S. Hassan and Z. Syed, "From netflix to heart attacks: collaborative filtering in medical datasets," *Proceedings of the 1st ACM International Health Informatics Symposium (IHI'10),* pp. 128–134, 2010.

[29] M. Komkhao, J. Lu, and L. Zhang, "Determining pattern similarity in a medical recommender system," in *Data and Knowledge Engineering*, Lecture Notes in Computer Science vol. 7696, pp. 103–114, Springer, Berlin Heidelberg, 2012.

[30] L. Duan, W. N. Street, and E. Xu, "Healthcare information systems: data mining methods in the creation of a clinical recommender system," *Enterprise Information Systems*, vol. 5, no. 2, pp. 169–181, 2011.

[31] T. Fredriksson and U. Pettersson, "Severe psoriasis oral therapy with a new retinoid," *Dermatologica*, vol. 157, no. 4, pp. 238–244, 1978.

[32] J. L. Herlocker, J. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.

[33] J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, vol. 27, no. 4, pp. 857–871, 1971.

[34] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, 2006.

[35] J. Mei, H. Liu, X. Li, G. Xie, and Y. Yu, "A decision fusion framework for treatment recommendation systems," *Studies in Health Technology and Informatics*, vol. 216, pp. 300–304, 2015.

[36] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *Journal of Machine Learning Research*, vol. 10, pp. 2935–2962, 2009.

[37] J. Mei, H. Liu, X. Li, Y. Yu, and G. Xie, "Outcome-driven evaluation metrics for treatment recommendation systems," *Studies in Health Technology and Informatics*, vol. 210, pp. 190–194, 2015.

[38] P. G. Campos, F. Diez, and I. Cantador, "Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols," *User Modeling and User-Adapted Interaction*, vol. 24, no. 1–2, pp. 67–119, 2014.