Sensors & Transducers, Vol. 234, Issue 6, June 2019, pp. 9-15

www.sensorsportal.com

Sensors & Transducers
Published by IFSA Publishing, S. L., 2019
http://www.sensorsportal.com

The Use of Standard Symptom Descriptions in the Macrocura Diagnosis Assistant System for Traditional Chinese Medicine Primary Care Practices

[1] Shasha Wu, [2] Qiang Su, [3] Wenli Fan

[1] Associate Professor of Computer Science, Spring Arbor University, Spring Arbor, MI, United States
[2] Chief Medical Scientist, Macrocura, Ltd., Hangzhou, Zhejiang, China
[3] Vice President of R&D, Macrocura, Ltd., Hangzhou, Zhejiang, China
Tel.: + 1-517-750-6698
E-mail: shasha.wu@arbor.edu

Received: 15 May 2019 /Accepted: 15 June 2019 /Published: 30 June 2019

Abstract: Traditional Chinese medicine (TCM) has been practiced by millions of people over two thousand years in China and surrounding countries like Japan and Korea. It is still widely accepted as a formal health care practice in these countries. The large amount of patients and long history of application helped TCM to accumulate significant medical knowledge both in practice and in theory. Although active in daily practices, the missing of modern scientific technology and analysis methods significantly limited TCM's utilization and improvement. Historically, the TCM diagnosis is restricted by the methods a TCM doctor can use when modern diagnosing technologies like microscope and X ray were not available. As the result, TCM doctors have to build their diagnosis theory on the human detectable symptoms like tongue textures and pulse patterns. Although this historical choice gives TCM diagnosis significant advantages in cost and patient comfort over modern medical diagnosis methods, the subjective sign and symptom descriptions used in TCM diagnosis can cause many ambiguity and accuracy problems in rigorous medical researches. It also makes the TCM clinical data difficult to store, retrieve, and analyze in the modern medical information systems. In this article, we proposed the Macrocura TCM diagnosis assistant system [1] with standard TCM symptom code to regulate the TCM diagnosis data in clinical practices. Through the system, we try to promote the use of standard symptom descriptions to define the TCM clinical data and demonstrate how modern information technology and machine learning algorithms can be used to support the study and research of TCM diagnosis in primary care practices.

Keywords: Medical diagnosis, Clinic data processing, Health information system, Macrocura system, TCM diagnosis.

1. Introduction

Traditional Chinese Medicine (TCM) has been widely practiced in China and surrounding countries like Japan and Korea for over two thousand years. It still plays an important role in the modern Chinese health care system. In the year of 2017, more than 588 million patients received medical treatment in 54,243 clinics by 448,716 certified Chinese medicine clinicians [2]. Many of these clinicians provide primary care or rehabilitation services to the patients. In recognition to the widespread use of TCM, World

Health Organization (WHO) added a new chapter to code traditional medicine sicknesses and disorders in the 11th revision of International Classification of Diseases (ICD 11) [3] in June, 2018.

Human body is an extremely complicated system that consistently interacts with the exterior environment. Historically, the TCM diagnosis is restricted by the methods a TCM doctor can use when modern diagnosing technologies like microscope, Magnetic resonance imaging (MRI), and X ray were not available. As the result, TCM doctors have to build their diagnosis theory mostly on top of human detectable symptoms like appetite, tongue shape and texture, pulse pattern, and descriptions of pain, etc., to match the patients with appropriate treatments and build a rational to explain it. This historical choice gave TCM diagnosis some advantages over modern medical diagnosis. For example, it is very economic because it does not need any expensive medical equipment or chemicals. TCM diagnosis also causes minimal damage or discomfort to the patient because no needles, radiation, or chemicals are involved. Those are all very appreciated for primary care as the first step to screen the patients.

Primary Care Practitioners (PCPs) manage a wide range of complex and diverse conditions through one or more relatively brief encounters [4]. The average Chinese medicine primary care encounter takes less than 10 minutes. Within that time, the clinician was trained to follow a set of sophisticated rules and principles to diagnose the patients based on the patient's description, human reasoning, observation, and clinician's intuition. The result of the diagnosis will lead to one or several established TCM prescriptions as the treatment. TCM uses different combinations of over 1,500 medical herbs and minerals to form about 300 established prescriptions that were experimentally proven in history to be effective for specific sicknesses and presenting symptoms.

2. Challenges and Opportunities

In spite of the widespread use of TCM in Asia and the West, rigorous scientific evidence of its effectiveness is still limited. TCM can be difficult for researchers to study because its treatments are often complex and are based on ideas very different from those of modern western medicine [5]. For example, the TCM medicine uses a lot of natural materials like herbs and minerals as treatments. These materials are difficult to manufacture, evaluate, measure, and control the quality. Modern chemistry and biology were introduced in recent years to improve the situation. One successful story of that is the year 2015 Nobel Prize winner, Youyou Tu's work on extracting the biologically active component of Artemisinin, a TCM herb, and clarify how it worked as an effective malaria cure. However, the discussion of the medicine goes beyond the scope of this article. Our research is to use modern digital technology and machine learning algorithms to assist TCM clinicians in their daily primary care clinical practice, promote the evidence-based medicine (EBM) [6] in TCM diagnosis, and learn through the TCM clinical data to verify the TCM diagnosis theory. To achieve these goals, TCM clinical data collection and standardization is the first step and the foundation of other researches.

Although huge amount of TCM primary care are practiced every day, most of the practices are still recorded by paper and pen with unstructured plain text descriptions. Analogies and poems are widely used in daily diagnosis and TCM literature. Qualifying words such as "a bit," "average," and "very" are common statements to describe the intensity of pain or other symptoms. All those descriptions are subject to the doctor's interpretation. Furthermore, few of the clinical practices resulted in feedback from the patients for the clinicians to review the correctness of the diagnosis and the effectiveness of the treatment. The lack of centralized, high quality, objective, and measurable clinical data and the miss of methods to collect timely feedback information from the patient largely prevented the scientific study of TCM.

Beyond that, most of the rules and principles of TCM are based on Chinese philosophical concepts, which are abstract and difficult to measure. That makes those rules and principles difficult to simulate and hard to verify. However, the patient's signs and symptoms, prescribed medicine, and the result of the treatment are solid and measurable evidences. If we can record those information accurately, it will offer us the opportunity to study the mechanisms and effectiveness of TCM treatments and promote the evidence-based medicine in TCM with the help of modern health information technology.

Health information technology (HIT) systems have the potential to reduce delayed, missed, or incorrect diagnoses. Unfortunately, progress in diagnostic HIT has been slow and incremental with few significant "game-changing" approaches having emerged in the last decade [7]. With recent advances in artificial intelligence, cloud computing, mobile devices, and wearable devices, we believe it's the time to change the situation. For example, mobile applications can be used to collect patients' feedback for researchers to evaluate the effectiveness of TCM diagnoses and treatments. Image processing and convolutional neural networks (CNN) can be used to detect abnormalities through patients' face or tongue images for disease screening. Decision tree algorithms can help clinicians decide whether to rule in or rule out specific TCM treatments. Deep neuronal networks can be trained to classify a set of presenting symptoms to a set of established TCM prescriptions.

3. Implementation

The four major assessments of traditional Chinese medicine are history taking (inquiry), listening, smelling, and palpation (pulse taking). We proposed the Macrocura TCM diagnosis assistant system [1, 8] that can process all four assessments to help the TCM clinicians in their daily practices and accumulate high quality modern clinic data to facilitate information technology researches in the areas of signal processing, machine learning, and medical analysis.

The Macrocura system provides a procedure of six steps to process TCM primary care activities. Fig. 1 shows the whole clinical process defined in the system. More auxiliary functions like appointment scheduling and payment system are beyond our discussion of TCM medical diagnosis assistant feature in this article. Therefore, they are not mentioned in this diagram.

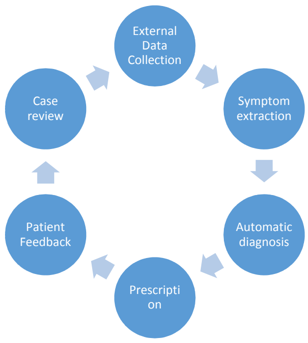

Fig. 1. TCM primary care steps in Macrocura system.

Among the six steps shown in Fig. 1, external data collection and symptom extraction and representation is the main focus of this article. They are the starting point and the foundation of other steps. The succeeding four steps will be mentioned as the direct output of making use of the data collected in the first two steps. If we treat the automatic diagnosis function in the system as the engine that drives the car, symptom data is like the gas that fuels the engine. Low quality of the symptom data is as damaging as dirty gas to a car engine.

3.1. External Data Collection

In the first step, the Macrocura system provides ways to collect and process two kinds of clinical data: multimedia data that needs additional processing steps to label them so that they can be used for efficient storage and automatic diagnosis, and sign or symptom description data that can be directly used in diagnosis. Fig. 2 shows the main screen of the *clinical visit* page, which is the user interface for external data collection and symptom extraction in the Macrocura system.

In Fig. 2, there are four major regions in the screen. The first region on the left most of the screen is the navigation area that allows the user to access six sub systems, which are listed from the top to bottom as: *patient management*, *clinical visit*, *patient feedback*, *personal case collection*, *statistics*, and *custom service*. In Fig. 2, the *clinical visit* button was selected to generate this screen shot. The second region is for multimedia and external data collection. Patient's images, pulse pattern, and external medical lab results can all be uploaded and managed from here. The third region is for the doctor to collect the patient's symptoms. Functions provided in this region will be discussed in Section 3.2. The fourth region on the rightmost of the screen is to display the selected symptom data for reference and adjustment.

Fig. 2. Screen shot of one user interface of Macrocura TCM diagnosis assistant system for primary care.

Fig. 3. Photo taking user interface in the Macrocura system.

A graphic user interface shown in Fig. 3, which is located in Region 2 of Fig. 2, is for the clinician to take normalized pictures from the patient's face, body, and tongue. It also displays the snapshots of those pictures for the doctor to have a quick review. Dedicated buttons are used to upload pictures from the file system or take pictures from a connected camera or other external equipment. Different image data will be fed to different engines to be processed and labeled. This design allows the Macrocura system connect to external data collecting equipment that can improve the quality of input imagery data. For example, the tongue image user interface provides the connection that can use an external 2D/3D tongue diagnosis system (TDS) [9] to collect high quality tongue image data. Fig. 4 shows one of the equipment that was developed by Kim, *et. al.* [10] to collect and process high quality tongue image data.

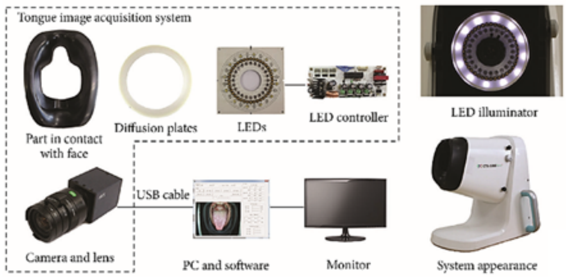

Fig. 4. Components of a tongue diagnosis system (TDS) [10].

TCM pulse diagnosis is one of the four major assessments in TCM consultation. Significant part of TCM diagnosis theory was built on top of the pattern of pulse palpation at three locations, i.e. cun, guan and chi, on both wrists. General health condition of a person and status of a particular organ can be recognized by an experienced TCM doctor through pulse diagnosis. By combining clinical data collected from pulse assessment and other clinical assessments, a TCM doctor can prescribe treatments to his patient and monitor his prognosis. However, mastering pulse diagnosis requires long-term experience and remains subjective up to a certain degree even in an advanced stage of practice [11]. The pulse condition is felt and defined by a TCM doctor through his/her fingers, and represents the subjective judgment of that doctor [12]. Automating TCM pulse diagnosis by means of technological aids is an active research field [11]. Among all of the researches, Fig. 5 shows one of the pulse sensor implementation [13] that can be used to improve the quality of the pulse data. The Macrocura system provides a way to import the result of those objective pulse measurements and integrate them with other patient symptoms for further analysis and study.

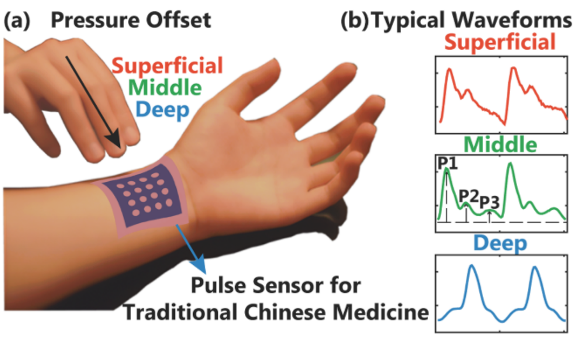

Fig. 5. A TCM pulse sensor [13].

Collecting and digitizing above TCM diagnosis data is only the first step. A well experienced TCM doctor can classify patient's facial color, tongue coating and texture, pulse patterns, voice characteristics in a few minutes and use all of them to make an initial judgement. The whole process is smooth and effortless to the doctor and the patient. However, the descriptions of those symptoms, which are the basis of the TCM diagnosis, are commonly documented by a lot of analogies, poems, and qualifying words that are subject to the interpretation of the doctor. Statements like "string-like pulse pattern" or "dark face" are widely used to describe the doctor's findings. The subjective descriptions in TCM diagnosis can cause many uncertainties when applying them in the TCM clinical practices. It is difficult to verify the correctness of the theory with all those uncertainties in the analysis. To prevent the subjective judgements and promote rigorous scientific studies, it is necessary to label these digital data to a standard format. To achieve this goal, sophisticated image processing algorithms may be implemented for color correction and feature segmentation, a group of TCM medical experts should be recruited to label the data with a standard format, and advanced machine learning algorithms and models can be trained and tested by the labeled data. Many research opportunities can be found in this area.

3.2. Symptom Extraction

After the initial data collection and processing, the doctor will have a conversation with the patient to understand his/her problem and discover more specific symptoms. Macrocura system designed a set of standard sign and symptom descriptions for the doctor to describe the patient's status. Every predefined symptom description is assigned a unique alphanumeric code and can be defined as one of the three states: positive, negative, and neutral. For example, *thirsty* is a symptom description that can be defined as one of the three states: positive (want to drink a lot or very frequently), negative (drink very few), or neutral (drink normally). Fig. 6 shows the three options a doctor can choose to define the *thirsty* symptom and the change of its associated symptoms on the top, which is a software feature that will be explained in the next two paragraphs.

The main reason to use standard symptom descriptions with three states instead of free style text or voice data for the Macrocura system is to avoid the uncertainty problems mention in the end of Section 3.1. The description of medical symptoms is the foundation of the TCM medical diagnosis. Natural language processing is fast and convenient, but the difficulty of handling analogy or poem kind of statements is still a big obstacle to make it a practical solution in the automatic TCM medical diagnosis. The new Traditional Medicine (TM) chapter of the CDC-11 [3] defined a similar coding system for the TM diseases. The difference between the two coding systems is that the CDC-11 code is for diagnosed diseases and the Macrocura code is for observable signs and symptoms.

The Macrocura system predefined more than a thousand commonly seen primary care signs or symptoms for clinicians to choose from. The design of the standard symptom description facilitates machine processing and supports multilingual environment.

The ultimate goal for this design is to make the predefined descriptions specific enough to manage the standard situations, broad enough to encompass the common exceptions, and flexible enough to allow separate decisions for the rare [14]. Based on the standard symptom list, patient's symptoms can be recorded precisely for future analysis and comparison. It also helps to prevent possible data discrepancy caused by doctor's subjective preferences in writing. The downside of this design is the significant incensement of the user interface complexity. Special training may be needed for the doctors to record patient signs and symptoms quickly and appropriately in the Macrocura system.

Fig. 6. Three states of the *thirsty* symptom and the associated symptom list for each state in the Macrocura system.

To address the complexity issue, three user interfaces were designed in the software to help the doctor finding appropriate symptom descriptions quickly and precisely in the Macrocura system. The first way is to navigate the symptom list based on the body part or behavior group it is associated with through the tabs on the left of the panel, as shown in the Area 1 of Fig. 7. For example, *thirsty* symptom can be found in the *food* behavior tab and *inflamed tonsils* symptom can be found from the *neck and throat* tab.

The second choice is the dynamically generated associated symptom list shown in the Area 2 of Fig. 7. For example, if the doctor picked the *cough (positive)* as a symptom from the screen, *running nose* and *sore throat* will be displayed on the top panel as the result of this choice. The doctor can select the two related symptoms directly from the fingertip without exploring the whole symptom list to find them. As another example, three different groups of associated symptoms were provided depend on the *thirsty* symptom state is positive, negative, or neutral shown in Fig. 6. Just like what a web search engine needs to do in the web search process, how to find the most relevant symptoms based a selected symptom is the key to this feature. Two methods were used to generate the associated symptom list appropriately. One is to recruit a group of medical experts to define the related symptoms as clustered symptom groups if those symptoms tends to happen together. This method is immediately usable, but expensive and limited in scope. Another method is making use of unsupervised machine learning algorithms such as Support Vector Machine (SVM) [15], Naïve Bayes [16], and k-Nearest Neighbors (kNN) [17] behind the screen to analyze how symptoms are clustered with each other in the real life clinical cases. This method is cost efficient and has unlimited scope. But it requires large amount of data to generate statistically stable results. Current Macrocura system implemented both methods to make the system immediately useful to the doctor users, while at the same time the system can continuously improve the performance with more clinical data collected.

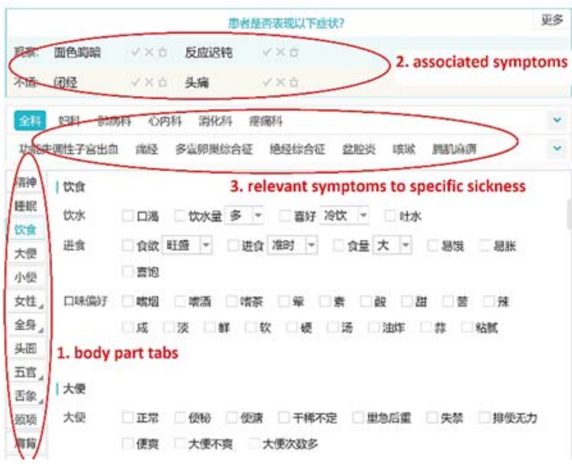

Fig. 7. User interface for symptom input in the Macrocura system.

The third way to improve the efficiency is the *relevant symptoms to specific sickness* user interface design. If the doctor can preliminarily define the sickness of a patient or the patient was already diagnosed to have a specific sickness, the doctor can pick the sickness name from a list in the system and see all the relevant symptoms in the selection panel. The layout of this design is shown in Area 3 of Fig. 7. Current Macrocura system uses medical experts to define the relevant symptoms for about 36 common sicknesses such as pneumonia or functional gastrointestinal disorders (FGIDs). In the future, new machine learning algorithms and researches may be implemented to analyze the clinical data to verify or adjust the relevant symptoms for existing and new sicknesses.

Finally, if a given symptom is not defined in the Macrocura system, doctor can use the *custom symptom* function to define new symptoms. However, customized symptoms will only serve as documented descriptions for given clinical cases and will not be processed automatically by the machine learning algorithms for automatic diagnosis analysis right now.

3.3. Diagnosis Assistant and Prescription

In step three, a hybrid TCM medical analysis engine that combines decision tree algorithm and neural network decision model is used to analyze the normalized symptoms collected from the previous steps. If the engine successfully matches a pattern, it will display the recommended TCM treatments for the doctor to consider. Region A of Fig. 8 listed three treatment recommendations as the result of the automatic diagnosis analysis for a given case. Top two recommendations are labeled with 5 stars to show that the engine is pretty confident on them to treat the symptoms listed in Region C, which were carried from the previous steps mentioned in Section 3.1 and 3.2.

Fig. 8. User interface for automatic treatment recommendation and prescription adjustment in the Macrocura system.

In step four, the doctor can make a decision on picking one or more recommended treatments as the basis for the final prescription. Selected treatments will be imported to region B of Fig. 8. The doctor can then adjust the doses of the medicine or add/remove individual medicine from the list to generate the final prescription. If the doctor does not like any one of the recommended treatments, he can also make his own prescription all from scratch in region B. System's recommendations, doctor's final choices and adjustments will all be recorded for future medical study and machine learning experiments.

3.4. Patient Feedback and Case Review

In step five, a patient can revisit the clinic or run the Macrocura mobile app (Fig. 9) on his/her smart phone to report his/her progress and feedback to the clinic. Patients can upload feedback to define overall effect of the treatment, give feedback on each symptom recorded in the previous visit, or add new symptoms as the direct result of the treatment. Those valuable feedback information will be collected for the doctor to review the case. It is also the key information to improve the performance of the system's machine learning algorithm.

Fig. 9. Screen shot of Macrocura patient feedback mobile application.

In step six, the clinician will review the progress and feedback report in the system to confirm or adjust the original diagnosis and treatment. If necessary, a revisit appointment can be scheduled for the patient to get a different treatment. That will lead to the first step of this cycle as shown in Fig. 1. All information will be saved in a cloud storage for future analysis and research.

4. Outputs and Future Work

The Macrocura TCM diagnosis assistant system for primary care is proposed to assist TCM clinicians in their daily primary care clinical practice, promote the evidence-based medicine (EBM) in TCM diagnosis, and learn through the TCM clinical data to verify the TCM diagnosis theory. A set of standard sign and symptom representations is defined in the system to normalize the description of the TCM clinical symptoms, diagnosis and treatment. The normalized digital images and standard representations are suitable for data processing and machine learning. With the deployment of the system in primary care clinics, it produces large amount of high quality labeled clinical data, which are valuable resources for EBM practice and other researches, such as image processing, machine learning, and medical analysis.

The Macrocura TCM diagnosis assistant system was first adopted by two sponsoring Chinese medicine clinics in April 2018. It was officially purchased and deployed by three hospitals and health care providers for their daily Chinese Medicine primary care practices in July 2018. By the mid of November 2018, the system recorded more than 30,000 clinic visits for more than 20,000 patients. The automatically generated TCM prescription recommendations reached an 80 % truth positive rate when compared with human TCM expert's prescription in an internal test. This rate is a promising result that encourages the TCM doctors to take a serious look at the recommended treatment in their daily practice. With the quick growth of high quality clinical data to support the machine learning algorithm, further improvement in the correctness rate is expected in the near future.

The Macrocura system shows the potential to help clinicians in their daily practices. The sturcturized real life clinic data provides a lot of potentials for information technology research and medical analysis. Although the system was designed for TCM clinics, the method it demonstrated and the data it collected are meaningful to all medical analysis and researches.

References

[1]. S. Wu, Q. Su, W. Fan, Macrocura: A Medical Diagnosis Assistant System for Traditional Chinese Medicine Primary Care Practices, in *Proceedings of the 1st International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2019)*, Barcelona, Spain, 2019, pp. 150-152.

[2]. China Statistical Yearbook of Chinese Medicine, National Administration of Traditional Chinese Medicine, 2017. [Online]. Available: http://www.satcm.gov.cn/2017tjzb/start.htm. [Accessed 11 2018].

[3]. WHO, ICD-11: International Classification of Diseases 11th Revision, *World Health Organization*, 6 2018. [Online]. Available: https://icd.who.int/en/. [Accessed 15 6 2019].

[4]. M. M. Hardeep Singh, M. M. Traber Davis Giardina, P. Ashley N. D. Meyer, M. M. D. Samuel N. Forjuoh, M. Michael D. Reis, M. M. Eric J. Thomas, Types and Origins of Diagnostic Errors in Primary Care Settings, *Journal of the American Medical Association Intern Medicine*, Vol. 173, Issue 6, 2013, pp. 418-425.

[5]. Traditional Chinese Medicine: In Depth, *U.S. Department of Health and Human Service*, [Online]. Available: https://nccih.nih.gov/health/whatiscam/chinesemed.htm. [Accessed 11 2018].

[6]. Evidence-Based Medicine Working Group, Evidence-based medicine. A new approach to teaching the practice of medicine, *JAMA*, Vol. 268, Issue 17, 1992, pp. 2420-2425.

[7]. R. El-Kareh, O. Hasan, G. D. Schiff, Use of health information technology to reduce diagnostic errors, *BMJ Quality & Safety*, Vol. 22, 2013, pp. 40-51.

[8]. Macrocura Homepage, Macrocura Ltd., [Online]. Available: https://www.macrocura.com/index.html. [Accessed 11 2018].

[9]. C. J. Jung, Y. J. Jeon, J. Y. Kim, K. H. Kim, Review on the current trends in tongue diagnosis systems, *Integrative Medicine Research*, Vol. 1, Issue 1, 2012, pp. 13-20.

[10]. J. Kim, J. Son, S. Jang, D.-H. Nam, G. Han, I. Yeo, S.-J. Ko, J.-W. Park, B. Ryu, J. Kim, Availability of Tongue Diagnosis System for Assessing Tongue Coating Thickness in Patients with Functional Dyspepsia, *Evidence-based Complementary and Alternative Medicine*, 2013.

[11]. R. Velik, An objective review of the technological developments for radial pulse diagnosis in Traditional Chinese Medicine, *European Journal of Integrative Medicine*, Vol. 7, Issue 4, 2015, pp. 321-331.

[12]. A. C. Y. Tang, Review of Traditional Chinese Medicine Pulse Diagnosis Quantification, *Complementary Therapies for the Contemporary Healthcare*, 2012.

[13]. Y. Chu, J. Zhong, H. Liu, N. Liu, Y. Song, X. Zang, Y. Dong, X. Wang, L. Lin, Self-Powered Pulse Sensors with High Sensitivity to Reveal Sinus Arrhythmia, in *Proceedings of the 31st IEEE International Conference on Micro Electro Mechanical Systems (MEMS'18)*, Belfast, Northern Ireland, UK, 2018.

[14]. A. R. Feinstein, An Analysis of Diagnostic Reasoning, III. The Construction of Clinical Algorithms, *Yale Journal of Biology and Medicine*, Vol. 1, 1974, pp. 5-32.

[15]. C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning*, Vol. 20, Issue 3, 1995, pp. 273-297.

[16]. S. J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, *Prentice Hall*, 3 ed., 2009.

[17]. N. S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, *The American Statistician*, Vol. 46, Issue 3, 1992, pp. 175-185.