ATLANTIS PRESS
Advances in Economics, Business and Management Research, volume 100
International Conference of Organizational Innovation (ICOI 2019)

Prospect of the Objectification of Pulse Diagnosis in Traditional Chinese Medicine in the Age of Big Data

Xiao-tao Wang (Department of Traditional Chinese Medicine, Ehu Branch of Xishan People's Hospital, Wuxi, Jiangsu, China)

Kai-jun Yu (Library, Shanghai University of Medicine ＆ Health Science, Shanghai, China)

Yu-zhou Luo (School of Medical Instruments, Shanghai University of Medicine ＆ Health Science, Shanghai, China)

Ming Zhong (Ruihua Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China.)

Email: luoyuzhouluo@126.com

Abstract—As one of the most distinctive diagnostic methods of traditional Chinese medicine (TCM), objectification of pulse diagnosis has a wide range of practical applications. Though the studies on objectification have been conducted for decades, due to small sample sizes, the results are satisfactory and clinical promotion is still in the embryonic stage. The arrival of the era of big data has not only promoted the development of biomedicine, but has also opened up new ways for the objectification of pulse diagnosis in TCM. In this review, we describe theories and methods of big data that may be applied to the objectification of pulse diagnosis in TCM. We propose to establish electronic health records for TCM and create a predictive model for pulse diagnosis and prospect the challenges and opportunities involved.

Keywords—pulse diagnosis, objectification, traditional Chinese medicine, big data

I. Introduction

Traditional Chinese medicine(TCM), originating from ancient China, is an experience-based medicine created by the Chinese. The unique diagnosis of TCM is based on gathering and integrating comprehensive information of the human body which is generally acquired by four distinct diagnostic methods of inspection, auscultation,olfaction, inquiry and pulse-feeling. As an important method for the clinical diagnosis of diseases in TCM, pulse diagnosis is generally considered to be the most difficult one to master because of its characteristic of being "easy in heart but hard to describe." To achieve the objectification of pulse diagnosis in TCM, scholars have put in half a century of scientific and technological efforts, but there are very few recognized results(1). With the rapid development of science and technology in contemporary society, TCM has been constantly evolving and seeking breakthroughs.

Big data is a term which describes data sets incorporating large volumes of information, sometimes of such complexities that traditional data processing methods cannot meet the practical requirements(2) . Currently, "Big Data" does not have an exact definition currently, but its characteristics are the so-called 5Vs, which are Volume, Velocity, Variety, Verification/Veracity, and Value(3). Big data is gaining world-wide acceptance in different fields. Since the Obama Administration announced the "Big Data Research and Development Initiative" in March 2012, it's almost well-known news: big data will innovate medicine(4, 5).

Many scholars have described their views on the relationship between big data and clinical medicine with a futuristic perspective. Limited articles have associated big data with TCM, especially pulse diagnosis. Here, in this paper, we discuss the relevant methods of big data that may be applied to the objectification of pulse diagnosis in TCM, and prospect the challenges and opportunities in the future.

II. CURRENT STATE

At present, the papers related to the objectification of pulse diagnosis mainly rely on the pulse diagnosis instruments with built-in pressure sensors. By measuring the pulse wave, the pressure-pulse diagram is drawn, and then the pulse signal is acquired, analyzed, and processed. After filtering the pulse signal, periods of the pulsation are intercepted, and analysis of the extracted parameters is performed. The parameters involve various amplitudes, time parameters, angles and frequencies. The analysis methods used in the current pulse diagram are mainly time domain analysis, frequency domain analysis and time-frequency analysis. The latter is the most commonly used(6).

Xu et al.(7) innovatively introduced the simultaneous transient wave intensity (WI) ultrasound technique into the study of objectification of a hypertensive pulse, and

Sheng et al.(8) used color Doppler spectrum to observe the hemodynamic changes of the radial artery during the exploration of the objectification of pulse manifestations. All of these researches provide feasible methods for the study of the objectification of pulse diagnosis, and provide new ideas for the introduction of new parameters.

However, at present, there is no uniform standard for the method of pulse diagnosis and the ingestion of information generated in the clinic, and operator biases among doctors are inevitable, leading to the occurrence of different interpretations of the same pulse. The objectification of the pulse diagnosis is to reduce these subjective biases as much as possible. In scientific studies,

because the identification of pulse information only focuses on single feature extraction, a study performed analysis of correlated features is rare. Big data technologies may be available to solve these problems.

III. How should the data be managed?

A. *Collecting and managing data*

Pulse conditions can demonstrate different physiological and pathological information in different individuals, and they can also be influenced by diet, mood and disease. Therefore, when using the existing pulse-recording technologies to collect the pulse conditions of a patient with clear diagnosis, it is necessary to record the characteristic parameters of the pulse information as comprehensively as possible, including pulse shape, pulse position, pulse power, frequency, rhythm, trend, fluctuation amplitude and even the hemodynamic parameters of blood flow in the arteries of both hands.

By collecting the comprehensive manifestation of the pulse characteristics, the three mentioned analysis methods may be used to capture the slightest change in the pulse signal, and different methods can be used to extract the energy distribution characteristics of different pulse diagrams.

For developing prognostic and predictive models in radiation oncology, Lambin et al. have described detailed features which should be considered and integrated. Clinical features (the patient status, organ function and grade, questionnaires, blood test results, etc.), treatment features (drug dosage, surgery and additional therapies), imaging features(CT, MRI, ultrasound, etc.) and other features play a key role in such models(9). So, drawing on the experience of radiation oncology, we should extract this kind of data as much as possible and integrate it into the clinical database. Under the circumstances, the volume of the data grow rapidly. To ensure the quality of data, the terms used in every record of data should be standardized, and independent double data entry and extraction, and verification by a third data checker or curator should be adopted.

B. *Establish TCM electronic health records (EHRs)*

There are usually no black-and-white answers in TCM, even in pulse taking. Appropriate answers are often based on relatively ambiguous or even contradictory theories and literature, and colored by individual experiences or intuition, what is called as “The Art of Medicine”(10). During the TCM diagnosis, pulse conditions ingested by clinicians using the fingertips are a type of fuzzy information. A diagnosis is made combining theory with personal experience.

As pulse diagnosis has such ambiguity, during the development of the patient's disease, the signal of the pulse will have subtle changes that are difficult to detect by pulse-taking that is carried out using fingertips. Therefore, we have to clarify the pulse diagram information of the patient at various periods during the course of the disease and conduct a digital transformation. As EHRs have been widely implemented, extracting data from EHRs for scientific research is quite feasible. By constructing an EHR for TCM, we may also explore the relationship between the nuances of the pulse and the disease with big data technologies(11, 12).

Since healthcare is becoming a new frontier for big data, objectification of the pulse diagnosis in TCM should incorporate information about the patients (including pulse information that changes with medical history, treatment process, etc.), research results and theory into large databases. For example, CancerLinQ, developed by the American Society of Clinical Oncology, will capture cancer data on 100% of patients with cancer to launch a large database, which will make it more likely to discover new knowledge and information compared to traditional clinical trials(10, 13, 14).

In the era of big data, information regarding pulse diagnosis in TCM will not be limited to a small sample of clinical trial analysis, but can be extended to the population level, using EHRs in TCM to build a large database. Only in this way can we fundamentally mine the pulse information at the population level and explore pulse diagnosis in TCM from a new perspective.

C. *Create a predictive model*

Our aim in building up EHRs in TCM is to develop predictive models, with which we may find a combination of factors that could accurately anticipate an individual patient’s outcome(9, 15) . These factors will include, but not be limited to the patient's diagnostic information or pulse information.

The procedure of finding a combination of factors correlated with the outcome should be identified, and whether it is predictive in independent datasets should be tested. While selecting data, potential features should first be reviewed, ideally by an expert panel.

As we wish to evaluate the performance of pulse-taking models in predicting, the current diagnosis of the patients will be defined as the “gold standard”. The primary outcome should be the area under the curve (AUC) and 95% confidence intervals (CIs), obtained from receiver operating characteristic (ROC) analysis which include sensitivity, specificity and accuracy. AUC is the most commonly used performance measure, which values from 0 to 1, where 1 represents the perfect model and 0.5 represents no discrimination. Once predictive factors have been identified, they should be evaluated on a second dataset for external validity.

IV. WHAT RELEVANT METHODS OF BIG DATA MAY BE APPLIED IN TCM?

While exploring some unrelated predictive factors for predicting qualitative outcomes, we may take logistic regression into consideration(9, 11). But when it comes to fuzzy data, big data mining should be considered. To create a model with a high performance, big data mining technologies may play a key role, as it is of more convenience in integrating data compared to traditional statistical methods.

A. *Fuzzy and rough set theory*

Fuzzy theory is used to mine the relevance of related factors based on certain rules. Fuzzy modeling, used to generate decisions from various factors, requires an approach that learns from experience, such as clinical data collected in advance. Fuzzy control, based on fuzzy theory, has been widely used to deal with nonlinear systems. Kuo et al.(16) applied a fuzzy neural network to the prognosis system of prostate cancer, and they found that it can learn the relationship between the clinical features and the prognosis of prostate cancer. In this way, the prognosis of a patient can be predicted once the clinical data are entered(17).

Rough set theory (RST) is also a method used to deal with imprecise information, as sensitivity or subgroup analysis cannot explain why there is heterogeneity between trials once being detected. RST manages to solve the problem. There are scholars who have applied a novel methodology based on RST on a sample dataset containing 1,111 patients from 9 randomized trials to evaluate the effect of two transplantation procedures of hematologic malignancies. The result shows that the method can divide subgroups with remarkably low statistical heterogeneity values(18).

As acquired from pulse-taking, the diagram of a pulse diagnosis in TCM contains a certain amount of noise and redundant information. In such cases, there are some similarities between the TCM diagnosis and the characters of big data theories.

B. *Support vector machine*

Support vector machine(SVM) classification is a useful tool in dealing with a large number of variables that cannot be separated linearly. It chooses similar functions of data, performs a transformation, and then picks “support vectors.”(11) In this way, groups with a new combination of vectors are classified. Several studies have applied this method to help predict or diagnose diseases.

Akanksha et al. used a SVM classifier to help diagnose schizophrenia with non-linear features. Results indicated that the method achieves maximum classification accuracy(19). Sandrine et al. predicted primary progressive aphasias according to multi-center structural magnetic resonance imaging data with an SVM approach. It was found that the whole brain SVM classification enabled a very high accuracy (91~97%) for identifying specific PPA subtypes, compared to healthy controls(20). SVM classification enables the prediction of subtypes of diseases with a very high accuracy. To study the objectification of pulse diagnosis, Gong et al. tried to classify liver cirrhosis with SVM. The result indicated that the method they applied can discriminate cirrhosis patients from healthy ones effectively(21).

C. *Artificial neural network*

An Artificial neural network (ANN) is a computing system vaguely inspired by the biological neural networks that constitute the animal brain. An ANN is based on a collection of connected units or nodes(“artificial neurons”), which loosely model the neurons in a biological brain. Each “neuron” can transmit signals from one artificial neuron to another. An artificial neuron receives data from the previous one can process it and then signal the output to the next neuron. In this way, the system “learns” to classify data, similar to the human brain. In medicine, ANNs have been used in different fields and made remarkable progress. For example, an ANN model has been used to predict the survival of patients with advanced carcinoma of the head and neck treated with irradiation(22). Sunkaria et al. developed a multi-layer ANN based heart rate variability(HRV) classifier to evaluate the health of the hearts of patients by electrocadiography(ECG). The result demonstrated that the HRV tests of the classifier closely matched with the opinions of the experts(23).

D. *Machine learning*

Machine learning plays a more and more indispensable role in big data analysis for it can mine or integrate large-scale and heterogeneous data and solve complex problems(2, 4). The different types of machine learning algorithms differ in their approaches, data types, tasks or problems that they deal with. Usually, supervised learning (SL), unsupervised learning (UL), semi-supervised learning are included, while there are various algorithms and methods that can be applied to enhance the current machine learning process.维基百科 DL 名下 7，8 .

SL algorithms build a mathematical model to reach a best match with the desired output by analyzing a training dataset. Then, the model is generalized to other unknown datasets and used to predict the output associated with new inputs.

UL, however, only has the inputs of a given test dataset. The algorithms learn from the whole unlabeled, unclassified data to find the commonalities. So UL may identify new correlations between data vectors which would not be discovered by human(11). [维基和 4].

Semi-supervised learning is an algorithm that learn both labeled and unlabeled data. It can use available unlabeled data to improve SL tasks when labeled data are lacking or expensive, which makes semi-supervised learning of great importance in machine learning and data mining(2).

In biomedicine, machine learning is able to predict the structure and function of protein from genetic sequences and consequently discern optimal diets for patients from their clinical and microbiome profiles(4). In TCM, it is to find the commonality between the patient's clinical features and his/her pulse conditions. Machine learning is likely to detect nuances in clinical features that have not been discovered as predictive factors by traditional human methods. The current researches reveal that machine learning methods will undoubtedly open up new and vast possibilities in TCM diagnosis.

V. DISCUSSION

A. *Why should we choose pulse diagnosis in TCM to objectify?*

Nowadays, the diagnosis of diseases in clinical medicine is based mostly on large-scale clinical trials, especially randomized controlled trials(RCTs). But with the rapid development of science and technology, both clinical and biological items and parameters of diagnosis technology are increasing substantially which makes it impossible to design such a large number of RCTs.

Pulse diagnosis is one of the most distinctive diagnostic methods for Chinese medicine. For thousands of years, Chinese medicine practitioners have created a profound Chinese sphygmology through clinical diagnosis and treatment. Pulse conditions comprehensively reflect the process of energy, metabolism and even a specific disease of human body. Moreover, for the other diagnostic methods of Chinese medicine, the image of the tongue can be misleading if the patient has eaten before diagnosis, and the diagnosis based on inspection of face can be difficult even if simple makeup has been applied. Pulse information is relatively stable and human-induced interference of pulse condition is comparatively small. Thus, pulse diagnosis can reduce the fraud and forgery of clinical information.

For a long time, pulse diagnosis relied on intuition, experience, analogy and symbolism instead of rationality and logic. Objectifying the pulse information and establishing objective parameter standards, excluding the bias of individual subjective perceptions, can help to standardize diagnosis and assessment of diseases.

In terms of pre-diagnosis of most diseases, patients can only rely on annual medical examinations. However, the accuracy of medicine is not only because it can diagnose diseases, but also because of the safe pre-diagnosis and monitoring of disease dynamically as early as possible, which greatly improves the management of diseases. One prominent goal of precision medicine is to provide personalized surveillance measures and therapies, and to significantly delay or even prevent the onset of diseases(24).

TCM holds the theory of "preventive treatment of disease",which can be traced back to ancient times. The prospective and non-invasive nature of pulse diagnosis is a feature that has existed since the very beginning. Therefore, we believe that the objectification of the pulse diagnosis is very valuable.

B. *Why should we use big data to objectify of TCM pulse diagnosis?*

In a certain sense, TCM is an empirical medicine. The theoretical basis of pulse diagnosis in TCM is summarized by a large amount of empirical data, obtained from the experience of ancient doctors. Although the method used in the actual process is mostly the inductive summary with the deductive mode, tracing it back to its origin, it is an empirical analysis of the correlation of big data. Also, big data itself refers to new tools offered by the evolving technology that allow managing and processing of diverse large-scale data(3, 25). Thus, by the means of big data, the exploration and analysis of the pulse image may result in an expression of pulse information in an objective form.

We believe that maintaining EHRs of the pulse diagnosis in Chinese medicine is a possible step towards its objectification. If big data of pulse image in TCM is constructed, the digital format of the TCM medical records from individual patients and the image mode of the pulse can be extended to complex high-dimensional data. At the population level, big data creates the possibility of exploring large-scale clinical outcomes to reveal hidden correlations. At present, the proposed TCM diagnosis and treatment EHRs can more or less enable us to overcome the problems and challenges we have had before, such as the analysis of rare conditions, and TCM specific data elements(3).

In routine clinical practice, an objective diagnosis of the disease is collected from different data streams, such as blood tests, imaging, pathology, electrophysiology, and even genomics. However, more insights, that can be gleaned from routinely obtained clinical data, into the pathogenesis and mechanisms of diseases remain to be discovered and further studied(3). In the near future, many of the characteristics of big data obtained from TCM clinical databases from millions of patients in clinics, hospitals or large-scale researches will certainly reflect subtle differences that conventional clinics cannot detect.

Using fuzzy and rough set theory, pulse diagnosis in TCM may further improve the ability of diagnosing the disease and judging the prognosis. The use of SVM or ANN, can certainly make the classification of pulse diagnosis in Chinese medicine more precise. In addition, the current research is limited to the study of a single pulse, and there is still a lack of relevant in-depth study on composite and concurrent pulses. Machine learning can create correlations between the pulse and the disease at the big data level, and find more commonalities from unclassified pulse clinical data.

Of course, if we rely entirely on data analysis, algorithms may “overfit” predictions to spurious correlations, which may lead to an overestimate of the pulse diagnosis ability(4). So we propose to build a predictive model and use diagnostic research to assess the accuracy. The accuracy of the pulse diagnosis was evaluated by testing the model of the true independent validation data set, and the reliability and validity were evaluated.

Pulse diagnosis in TCM pays great attention to the "floating, medium, sinking" of pulse, and big data technologies can refine the pulse of the three parts of the floating and sinking, subdividing the pulse position, and finally deepen the research on the pulse.

The current process of objectification still has many limitations. There are many studies on the objectification of pulse that lack systematicity. Many studies only use a certain paragraph of the process to study or optimize, but the methods for the other steps are not considered well, and ultimately affect the experimental results. From instrumentation to acquisition to analysis, pulse information extraction and acquisition techniques urgently need to establish a standard model. After obtaining the pulse diagnosis information, the difficulty and focus of the objectification of pulse diagnosis lies in performing data preprocessing, feature extraction, parameter analysis and data mining. How to extract the features existing in pulse diagnosis information completely and truly is a challenge in the current research in this field.

Therefore, it is imperative to formulate a complete and comprehensive collection of pulse diagnosis information and analysis rules, to build a specific pulse research platform in an industry-recognized format, and to establish the EHRs of pulse diagnosis in Chinese medicine.

In the era of big data, pulse diagnosis in TCM does not only consider a small sample for data analysis, but collects the data of the overall EHRs as much as possible, which can fundamentally reflect the details at the population level and explore the pulse diagnosis in TCM from a new perspective. We are confident that they can promote the objectification of pulse diagnosis in TCM.

ACKNOWLEDGMENT

This work was completely supported by Wuxi Municipal Commission of Health and Family Planning Research Project (Q201773)

REFERENCES

[1] Wang NY, Yu YH, Liu J , et al. Consideration on objective research of pulse taking in TCM. China Journal of Traditional Chinese Medicine & Pharmacy. 2015;30(8):2655-57.

[2] Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. Proteomics. 2016;16(5):741-58.

[3] He KY, Ge D, He MM. Big Data Analytics for Genomic Medicine. Int J Mol Sci. 2017;18(2).

[4] Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-9.

[5] Cui M, Li H, Hu X. Similarities between "Big Data" and traditional Chinese medicine information. Journal of traditional Chinese medicine = Chung i tsa chih ying wen pan. 2014;34(4):518-22.

[6] Chen C, Zhou LY, Liu J , et al. Research progress of pulse information analysis method. China Medical Herald. 2018;15(23):34-6.

[7] XU YL, Xiao HS, XU F. Pulse research based on the wave intensity technology in patients with primary hypertension of different pathological stages. Chinese imaging journal of integrated traditional and western medicine. 2016;14(5):537-9.

[8] Li S, Zheng L. Exploration on the Objectivity of Pulse Manifestations Evaluated by Doppler Spectrum Based on the Features of Hemodynamics. Western Journal of Traditional Chinese Medicine. 2016;29(2):141-2.

[9] Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. Nature Reviews Clinical Oncology. 2012;10(1):27-40.

[10] Kantarjian H, Yu PP. Artificial Intelligence, Big

Data, and Cancer. JAMA Oncology. 2015;1(5).

[11] Bibault JE, Giraud P, Burgun A. Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett. 2016;382(1):110-7.

[12] Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present and future. Genetics in Medicine. 2013;15(10):761-71.

[13] Rubinstein SM, Warner JL. CancerLinQ: Origins, Implementation, and Future Directions. 2018(2):1-7.

[14] Sledge GW, Jr., Miller RS, Hauser R. CancerLinQ and the future of cancer care. American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting. 2013:430-4.

[15] Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Annals of internal medicine. 2012;157(1):29-43.

[16] Kuo R-J, Huang M-H, Cheng W-C, Lin C-C, Wu Y-H. Application of a two-stage fuzzy neural network to a prostate cancer prognosis system. Artificial Intelligence in Medicine. 2015;63(2):119-33.

[17] Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in Clinical Medicine. Chinese medical journal. 2016;129(6):731-8.

[18] Gil-Herrera E, Tsalatsanis A, Kumar A, Mhaskar R, Miladinovic B, Yalcin A, et al. Identifying homogenous subgroups for individual patient meta-analysis based on Rough Set Theory. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2014;2014:3434-7.

[19] Juneja A, Rana B, Agrawal RK. A novel fuzzy rough selection of non-linearly extracted features for schizophrenia diagnosis using fMRI. Computer Methods and Programs in Biomedicine. 2018;155:139-52.

[20] Bisenius S, Mueller K, Diehl-Schmid J, Fassbender K, Grimmer T, Jessen F, et al. Predicting primary progressive aphasias with support vector machine approaches in structural MRI data. Neuroimage Clin. 2017;14:334-43.

[21] Gong SJ, Chen MR. Research on Cirrhosis Diagnosis Based on Pulse Signal Processing. Computer Applications and Software. 2014;31(6):69-70.

[22] Bryce TJ, Dewhirst MW, Floyd CE, Jr., Hars V, Brizel DM. Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck. International journal of radiation oncology, biology, physics. 1998;41(2):339-45.

[23] Sunkaria RK, Kumar V, Saxena SC, Singhal AM. An ANN-based HRV classifier for cardiac health prognosis. International journal of electronic healthcare. 2014;7(4):315-30.

[24] Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. Genome Med. 2016;8(1):134.

[25] Peters SG, Buntrock JD. Big Data and the Electronic Health Record. Journal of Ambulatory Care Management. 2014;37(3):206-10.