November-December 2019
ISSN: 0193-4120 Page No. 2239 - 2245

A Study on Analyzing Symptoms of Korean Rheumatoid Arthritis Patients Using Machine Learning

Seo Won Song[1], Mi Kyoung Lim[2], Chunhwa Ihm[3], Min Soo Kang[4*]
[1]*Dept of Medical IT Marketing, Eulji University, Korea, songst32@gmail.com*
[2]*Dept of Rheumatic Internal Medicine, Eulji University Hospital, Korea, mklim@eulji.ac.kr*
[3]*Dept of Laboratory Medicine, Eulji University Hospital, Korea, haneul@eulji.ac.kr*
[4]*Dept of Medical IT, Eulji University, Korea, mskang@eulji.ac.kr (*corresponding author)*

Article Info
Volume 81
Page Number: 2239 - 2245
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *12 December 2019*

Abstract
Rheumatoid arthritis, one of the musculoskeletal disorders, is an autoimmune disease that the immune system attacks itself and causes inflammation of the joint tissue. Rheumatoid arthritis is found in a variety of ages, mainly in 30s and 50s, and is more common in women than in men. Golden time of treatment would be within 6 months after the symptoms of rheumatoid arthritis appear, and the sooner it starts, the better it is for patients. If early diagnosis and treatment are performed, the symptoms may improve sufficiently, but the awareness or diagnosis rate is not high. If significant symptoms associated with rheumatoid arthritis are known, self-diagnosis is possible, and the disease may be detected early. In this study, the decision forest algorithm, which is one of artificial intelligence algorithms, was applied to the EMR data of Eulji University Hospital to analyze the symptoms of rheumatoid arthritis patients. The results of the experiment were compared with the opinion of the doctor and 90% agreement was obtained. If the results of this study are used for self - diagnosis, it will be able to easily diagnose the disease and the quality of life can be improved by early detection and prevention of deterioration.

Keywords*: Early detection, Machine learning, Rheumatoid arthritis, Self-diagnosis*

1. INTRODUCTION

Population aging is taking place all over the world and interest in health care is growing. In 2017, the elderly population over 65years old in Korea made up 14.2% of the total population. Korea entered the aging society due to the increase in average life expectancy and low fertility rate. Korea is aging faster than other countries [1]. The World Health Organization (WHO) and United Nations (UN) defined as the aging society when the proportion of the elderly people over 65 years old is over 7%, the elderly society when over 14%, and the super aged society when over 20%. The population aging of Korea is expected to continue, and a solution to the surge in medical expenses due to population aging and the increase in patients with chronic diseases is required.

1.1 Smart healthcare

Population aging is underway, and interest in ‘healthcare’, which is a concept to manage disease prevention, treatment and health has increased. Recently, with the development of IT technology, there is a strong interest in smart healthcare that can connect ICT technology with medical services and provide healthcare services not limited to time and place. Smart healthcare is a convergence of healthcare and IT technology,

November-December 2019
ISSN: 0193-4120 Page No. 2239 - 2245

which makes possible that people can receive health care in real time. The healthcare paradigm has shifted from the center of the hospital(treatment) to the patient(prevention) after the disease occurred. With the spread of personal medical devices and wearable devices, a large amount of data is being generated and personalized services are emerging. When smart health care is activated, personalized health care is available anytime, anywhere because data generated in real time can be monitored [2]. These health checks not only predict or prevent diseases, but also reduce medical costs. Therefore, it is expected that it will be possible to provide customized medical services to patients and elderly people who have difficulty in going to hospital where telemedicine is needed. According to one of statistics market researches, the market size of the global digital healthcare industry is estimated to be $ 96 billion in 2016, which is expected to grow at a CAGR of 21.1% and to expand to $ 206 billion in 2020 [3]. Research on smart healthcare will be actively conducted

1.2 Rheumatoid arthritis

Rheumatoid arthritis is a chronic inflammatory disease that causes joint destruction. Unlike osteoarthritis, which is caused by degenerative changes, rheumatoid arthritis is an autoimmune disease that causes the immune system to attack itself and cause inflammation of the joint tissue. An abnormality in the body's immune system causes inflammation of cartilage and joints, and damages joints and bones. Rheumatoid arthritis starts with small joints such as the fingers, toes and wrists, symptoms progress, changes the shape of the bones and causes swelling and stiffness. Women are 2-4 times more likely to develop disease than men and are found in various age groups. Golden time would be within 6 months after symptoms appear, and the sooner the treatment starts, the better it is for patients. According to a study by the Korean College of Rheumatology, the initial diagnosis of rheumatoid arthritis in Korea was delayed by 3 to 5 times compared to the patients in Canada, Belgium and Denmark. The later the diagnosis and treatment, the higher the rate of experiencing joint dysfunction [4]. And, according to the American Academy of Rheumatology, patients who underwent early treatment within 6 months had less mortality and disability than those who did not [5]. Symptoms may improve sufficiently if early diagnosis and treatment are performed, but the problem is that the recognition or diagnosis rate is not high. In addition, studies on rheumatoid arthritis are actively performed, but there is a clinical difference according to the characteristics of country and ethnicity, so it is necessary to study the patients with rheumatoid arthritis in Korea.

2. RELATED RESEARCH

Studies on the classification of rheumatoid arthritis have been performed for a long time. The American Rheumatism Association published the classification standard in 1958, but it had a disadvantage that the understanding of the disease was different from the current one and it was difficult to apply it to the actual situation, and the revised standard was announced in 1987. Although this standard is useful for diagnosing advanced rheumatoid arthritis, but it has been difficult to apply for the diagnosis of premature rheumatoid arthritis. The American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) published a new revised rheumatoid arthritis classification standard [6]. The criteria published in 2010 diagnosed as rheumatoid arthritis with a score of 6 or more out of a total of 10 points based on the number of involved joints, serum test, duration of symptoms etc. In Korea, research has been conducted on whether this classification criterion is useful for Koreans. The new 2010 ACR / EULAR classification criteria are more effective in diagnosing early

rheumatoid arthritis, but sensitivity of the diagnosis was low for some patients [7].

Like the classification criteria, studies related to rheumatoid arthritis have been based mainly on physical symptoms such as pain and duration. Kim suggested an approach to patients with rheumatoid arthritis based on history and physical symptoms. Since 80% of information can obtain through history listening and physical examination, so it is important to familiar with the characteristics of disease and the findings of physical examination and proposed systematic items of rheumatoid arthritis [8]. The proposed contents are summarized in Table 1. Research on self-diagnosis that checks individual's health condition on its own is actively conducted. 60 kinds of diseases were selected, and the DB was constructed considering the symptoms of each disease and its relationship with the body parts. The ART2 learning algorithm was applied to classify representative symptoms of each disease and to derive disease. If the user selects the uncomfortable symptoms, the system suggests five possible diseases [9]. And G.B. Kim proposed herbal self-diagnosis disease classification system that allows people who lack expertise in disease to easily understand their health condition. The patient selects representative symptoms to get relevant diseases, select detailed symptoms again for suggested diseases, and the system suggest the final three diseases by separating the necessary symptoms and accompanying symptoms [10].

Table 1. Systematic items of rheumatoid arthritis

System	Symptoms
Systemic symptoms	high fever, weight loss, fatigue
Eyeball	diplopia conjunctivitis, dry eye
Oral	tongue, throat ulcer, periodontitis, dry mouth
Digestive	gastrointestinal ulcer, esophagitis, stomachache, jaundice, melena
Urinary	hematuria, pollakisuria, urinary tract infection
Skin	skin rash, Raynaud's phenomenon, psoriasis, photosensitivity, hair loss
Respiratory	dry cough, dyspnea on exertion, hemoptysis
Cardiovascular	chest pain, angina, arrhythmias, hypertension
Brain-nervous	hemiparalysis, convulsion, dysaesthesia, muscular weakening
Musculoskeletal	arthrodynia, edema, subcutaneous nodule, range of motion

In recent years, research has been actively conducted to combine the medical field with the ICT technology. Research on applying artificial intelligence to the medical field has been actively conducted all over the world and has been introduced into the medical industry to provide medical services. Artificial intelligence is also being introduced in diagnostics, and IBM's Watson and Google predictions of blindness are representative examples. Google applied artificial intelligence to about 130,000 eyeball images and 880,000 diagnostic data to predict the likelihood of blindness to a similar level as ophthalmologist. In China, doctors use AI to diagnose colon polyps in colonoscopy. In a clinical trial, it had shown that there was 20% increase in polyp detection when diagnosed by AI and specialists. It is because they have found exactly the smallest polyps that can be easily missed [11]. IBM's Watson for oncology is an AI software that helps diagnose and treat cancer, collecting, analyzing and updating a large number of newly published papers in real time and suggesting probable disease names and treatments. Therefore, doctors can study up-to-date knowledge and refer to it when making a diagnosis. Watson has updated and

November-December 2019
ISSN: 0193-4120 Page No. 2239 - 2245

trained more than 300 medical journals, 200 medical textbooks, and more than 15 million pages of specialized data at the Memorial Sloan Catering, one of the three major cancer centers in the world. Watson reached a level where he could propose an accurate diagnosis and optimal prescription in 8 seconds.

AI is also being applied to medical platforms. Medical Platform IHP (INFINITT Healthcare platform) is an AI system that integrates and manages patient medical data scattered by each department. IHP has the advantage of making it easier to identify the patient's overall condition and reducing hospital operating costs by unifying data management. In addition, it will be able to utilize the accumulated data to better diagnose and treat various diseases [12].

In this study, AI applied to EMR of rheumatology department of Eulji Hospital of Eulji University and tried to analyze symptoms of rheumatoid arthritis. Experiments and results are presented in section 3 and conclusions are in section 4.

3. EXPERIMENT

Artificial intelligence refers to the intelligence created by a computer or system, and Machine Learning is a field of artificial intelligence that refers to the fields in which algorithms and techniques that computers can learn [13]. It is a technology that can apply algorithms to data to find information or characteristics and predict the outcome of new events. Machine learning can be divided into supervised learning and non-supervised learning depending on the presence or absence of the label. Supervised learning is performed when each object has a label. For the input x, it is given what the label y is. Typically, there is a classification applied when the label y is discrete, and a regression applied when the label y is continuous. If there is no label for the input object, it is non-supervised learning and clustering is one of the representative techniques.

In this study, one of the classification algorithms, the Decision Forest algorithm was applied to the data and Python and artificial intelligence tool of Microsoft Azure ml used. Microsoft's azure ml is a cloud-based solution that does not require complex programming and can be easily modeled with drag and drop. It is convenient because it provides visualized results, and R or Python scripts can be used in module form, so users can implement additional functions. Unlike existing tools, GUI environment is provided for user's convenience, and anyone can easily create and deploy models [14].

Decision forest algorithm is a kind of ensemble learning method used for classification and regression analysis. It gives class (classification) or average predicted value from a large number of decision trees constructed in the training process. If a general decision tree is used, there is a risk of overfitting and generalization is difficult. On the other hand, Random Forest improves generalization performance through the use of bagging and is robust against data containing noise [15].

EMR data of patients who were treated at the Department of Rheumatology of Eulji University Hospital was used. Data includes diagnosis, diagnosis class, diagnosis code, symptoms, symptom code etc., and de-identified. Symptoms associated with rheumatoid arthritis were identified as '1' if the patient had symptoms, '0' if the patient had no symptoms, and unnecessary attributes were removed. The number of variables used in the analysis, the dimension increase, the curse of dimension problem occurs that the amount of data required increases exponentially. In other words, it is important to select features that can represent the meaning of the data because it is difficult to express meaning in a very high dimensional data. To reduce the dimension, Azure ml 'Filter based feature selection' module was used, and one of the classification algorithms, 'Two class Decision

November-December 2019
ISSN: 0193-4120 Page No. 2239 - 2245

Forest' algorithm is applied. The result of applying the algorithm was 73.3 of accuracy and the AUC was 77.4. The top part of the tree is shown in Fig 1 and the entire tree is shown in Fig 2.

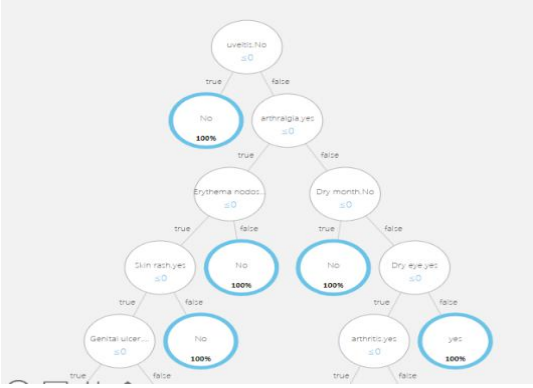

Figure 1: Enlarged top of the tree

In the Fig 2, uveitis, arthralgia, erythema nodosum, dry month, skin rash, dry eye etc. were significant attributes to classify whether patient of rheumatoid arthritis or not. The results of this study were consulted to the doctor. Most of the symptoms that were derived from the analysis using artificial intelligence were significant symptoms and showed about 90% agreement rate with doctor [18-19].

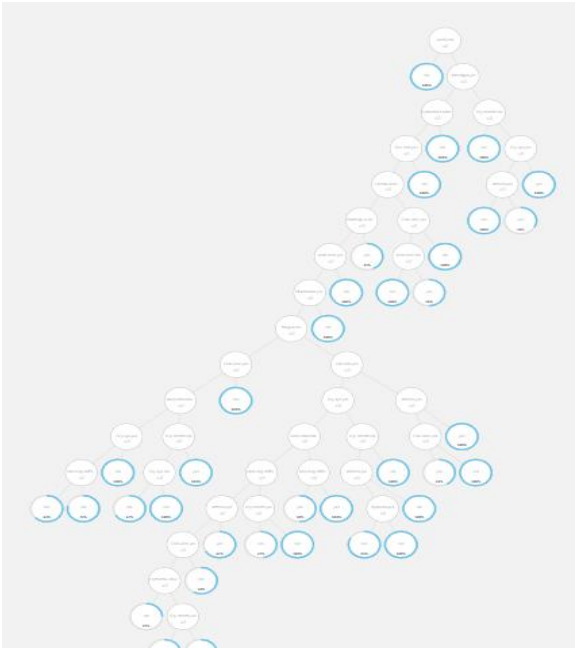

Figure 2. Entire tree

Through this study, identified 18 attributes that are significant in classifying rheumatoid arthritis patients. If these results are used in self – diagnosis(questionnaires), it is possible to prevent deterioration and detect disease early by simple method and improve the quality of life through self-diagnosis of the disease [20].

4. CONCLUSION

As population aging progresses rapidly, there is growing interest in quality of life. Korea has also entered the aging society and is aging faster than other countries. The aging of Korea is rapidly progressing, and a solution to the surge in medical expenses due to population aging and the increase of patients with chronic diseases is required.

Rheumatoid arthritis is an autoimmune disease that causes inflammation of the joint tissues because of abnormalities of the immune system, which can lead to changes in the shape of the bones, swelling and stiffness. Rheumatoid arthritis is one of the diseases that lower the quality of life and it is a disease that does not have a high awareness or diagnosis rate even though it is a disease that occurs in a wide age group. Since the progression of rheumatoid arthritis is fast and the rate of getting joint dysfunction will accelerate as diagnose and treatment are delayed, it is crucial to have early diagnosis. Thus, if early diagnosis of rheumatoid arthritis is easy, it will improve the quality of life.

In this study, the EMR data of the Rheumatology Department of Eulji Hospital of Eulji University were analyzed to find meaningful symptoms in classifying patients with rheumatoid arthritis. The Microsoft artificial intelligence tool, Azure ML was used, and the two-class decision forest algorithm, one of the classification algorithms, was applied. As a result, uveitis, arthralgia, erythema nodosum, dry month, skin rash, dry eye etc. were significant attributes to classification, and the accuracy was 73.3 and the AUC was

November-December 2019
ISSN: 0193-4120 Page No. 2239 - 2245

77.4. 18 symptoms were found to be significant, and 13 symptoms were overlapped when compared with the higher attributes ranked in order of frequency. Comparing the experimental results with the opinions of the doctor, most of symptoms from the analysis using artificial intelligence were significant and showed about 90% agreement rate. If these results are used in self – diagnosis(questionnaires), the simple - rheumatoid arthritis questionnaire can diagnose the disease easily and early. It can reduce patient waiting time and increases patient turnover. In addition, it can improve the quality of life by early detection and prevention of deterioration and reduce medical costs. The results of this questionnaire will be great value for future research in the field of medical IT as well as for rheumatoid arthritis.

ACKNOWLEDGMENT

"This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03034411)."

REFERENCES

1. Population census, *Statistics Korea*, 2017.
2. M.Y. Kang, D.H. Park, G.S. Kim, **Current and future of smart healthcare**, *Issue monitor*, *Samjong KPMG*, No.79 January 2018.
3. **Smart Healthcare Medical Devices Technical and standard strategy report**, *Ministry of Food and Drug Safety*, 2018.
4. D.I. Kwon, Hankook-Ilbo, **Increase Young patients rheumatoid arthritis, early intensive treatment is necessary**, Available: https://www.hankookilbo.com/News/Read/201505081388048525
5. *Herald Corporation, manuscript of health information*, **Arthritis depends on the management**, Available: http://biz.heraldcorp.com/view.php?ud=20190424000291
6. Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., Bingham III, C. O., ... & Combe, B., **2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative**, *Arthritis & Rheumatism*, Vol.62, No.9, pp.2569-2581, 2010.
7. Su-Jin Moon, Chang Hoon Lee, Yun Sung Kim et al., **Usefulness and limitation of 2010 ACR/EULAR classification criteria in Korean patients with early RA**, *The Journal of Rheumatic Diseases*, Vol. 19, No.6, pp. 326-333, 2012.
8. Jin Seok Kim, **Approach to arthritis patients**, *The Korean Journal of Medicine*, Vol.2, pp.368-372, 2017.
9. Kwang-Beak Kim, Young Woon Woo, Ju Sung Kim, **Self-disease diagnosis system enhanced ART2 algorithm**, *The Journal of Korea Institute of Information and communication engineering*, Vo.11, No.11, pp. 2150-2157, 2007.
10. Kwang-Beak Kim, **Self-diagnosing disease classification system for oriental medical science with refined fuzzy ART algorithm**, *The Journal of Korea contents association*, Vol.9, No.7, pp. 1-8, 2009.
11. Seong Gyu Lee, **AI innovates health care services**, *The Science Times*, Available: https://www.sciencetimes.co.kr/?news=ai-%EC%9D%98%EB%A3%8C-%EC%84%9C%EB%B9%84%EC%8A%A4%EB%A5%BC-%ED%98%81%EC%8B%A0%ED%95%98%EB%8B%A4
12. Nahar, A. K., Abdalla, A. N., Jaber, A. Y., & Ezzaldean, M. M. (2017). **PAPR Reduction Using Eight Factors Rotating Phase Shift Technique Based on Local Search Algorithm in OFDM**. Review of Computer Engineering Research, 4(2), 38-53.
13. J.H. Kwon, **AI Doctor is responsible for our health**, *SK Hynix blog*, Available: https://blog.skhynix.com/2501
14. Wikipedia, **Artificial intelligence**, Available: https://ko.wikipedia.org/wiki/%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5
15. M.S. Kang, H.J. Kang, G.B. You, C.H. Ihm, E.S. Choi., **Getting started with machine learning using Microsoft Azure Machine Learning Studio**, *Hantee media*, 2018.

16. Wikipedia, **Random forest**, Available: https://ko.wikipedia.org/wiki/%EB%9E%9C%EB%8D%A4_%ED%8F%AC%EB%A0%88%EC%8A%A4%ED%8A%B8
17. Jung-Soo Song, **New classification criteria for rheumatoid arthritis**, *The Korean Journal of Medicine*, Vol.87, No.4, pp.383-388, 2014.
18. Schouten, M.J. 2019. Undoing gender inequalities: insights from the Portuguese perspective, Insights into Regional Development (2): 85-98. https://doi.org/10.9770/ird.2019.1.2(1)
19. Kumaraswamy, S.; Ebrahim, R.H.; Wan Mohammad, W.M. 2019. Dividend policy and stock price volatility in Indian capital market. Entrepreneurship and Sustainability Issues, 7(2), 862-874. http://doi.org/10.9770/jesi.2019.7.2(5)
20. Jabarullah, N. H., Jermsittiparsert, K., Melnikov, P. A., Maseleno, A., Hosseinian, A., & Vessally, E. (2019). Methods for the direct synthesis of thioesters from aldehydes: a focus review. Journal of Sulfur Chemistry, https://doi.org/10.1080/17415993.2019.1658764 .