LEARNING PATIENTS WITH RHEUMATOID ARTHRITIS BY DEEP PREDICTION CONSIDERING 2-TIME-POINT X-RAYS OF AUTOMATIC FINGER JOINT BONE EROSION SCORE

K. Izumi1,2, K. Suzuki1, M. Hashimoto3, T. Endoh4, K. Doi1, Y. Iwaf5, Y. Kaneko5, M. Jinzaki5, S. Ko5, T. Takeuchi1, 1Keio University School of Medicine, Division of Rheumatology, Department of Internal Medicine, Tokyo, Japan; 2Tokyo Medical Center, Division of Connective Tissue Diseases, Tokyo, Japan; 3Keio University School of Medicine, Department of Systems Medicine, Tokyo, Japan; 4Fujitsu Laboratories Ltd., Kawasaki, Japan; 5Keio University School of Medicine, Department of Radiology, Tokyo, Japan

Background: Medical image analysis using deep learning (DL) has been attracting attention. In previous research, we proposed a DL method for detection of joint region and evaluation for bone destruction at a single point in time in hand X-rays of patients with rheumatoid arthritis (RA) [1-2]. However, in the score of van der Heijde-modified total Sharp scores (mTSS) in X-rays, it is difficult to apply the method as it is. In mTSS, score difference between 2-time points is important, and there is a problem that the score at each time varies depending on the doctor who evaluates.

Objectives: We aimed at developing an mTSS scoring method considering 2-time-point difference with a DL method.

Methods: A total of 104 X-ray image sets of both hands at 2 time points with an interval of ≥1 year were randomly obtained from patients with RA who had visited our clinic in 2015. Well-trained doctors determined the erosion scores of MP and PIP/IP joints of each hand in X-rays according to mTSS. These evaluations of hand joints were performed using our developed annotation software tool. In the learning phase, joint images were randomly divided into five sets for 5-fold cross-validation. We utilized a convolutional neural network model, such as SSD [3], for detecting joint regions and classifying the scores (Fig 1).

The models for classification were designed in consideration of the difference in erosion scores of each patient between the 2-time points of X-rays. The loss function of the DL model was defined bellow;

\[ \text{SCE} = \frac{1}{N} \sum_{i=1}^{N} \text{SCE}(t_i, y_i) \]

Here, y is the coefficient designed to reduce the error for another set of scores with equal differences. The first term of the loss function works to optimize the score at each time point, and the second term works to optimize the score difference at both time points. Thus, our method can be trained without being affected by characteristic training data.

Results: The number of joints with differences in erosion score between the former and latter time points was 1 (±2 points), 9 (±1), 2015 (0), 32 (+1), 17 (+2), and 6 (+3). There were no joints with score changes of −5, −4, −3, +4, and +5 points. As a performance of predicting the difference in erosion score between the 2-time points of each patient’s X-ray, our models presented a mean error of 0.412 per each joint in one set for 5-cross validation as compared with physicians’ evaluation (Fig 2).

Conclusion: Our DL-based models to predict hand joint erosion scores in X-rays were developed with relatively small samples. This suggests that the predictive performance may increase by collecting more training dataset. Next, we will apply our method to the prediction of joint space narrowing score.

References:

Acknowledgments: Izumi and Suzuki are contributed equally.


Figure 1.

The models for classification were designed in consideration of the difference in erosion scores of each patient between the 2-time points of X-rays. The loss function of the DL model was defined bellow;

\[ \text{SCE} = \frac{1}{N} \sum_{i=1}^{N} \text{SCE}(t_i, y_i) \]

Here, y is the coefficient designed to reduce the error for another set of scores with equal differences. The first term of the loss function works to optimize the score at each time point, and the second term works to optimize the score difference at both time points. Thus, our method can be trained without being affected by characteristic training data.

Results: The number of joints with differences in erosion score between the former and latter time points was 1 (±2 points), 9 (±1), 2015 (0), 32 (+1), 17 (+2), and 6 (+3). There were no joints with score changes of −5, −4, −3, +4, and +5 points. As a performance of predicting the difference in erosion score between the 2-time points of each patient’s X-ray, our models presented a mean error of 0.412 per each joint in one set for 5-cross validation as compared with physicians’ evaluation (Fig 2).

Conclusion: Our DL-based models to predict hand joint erosion scores in X-rays were developed with relatively small samples. This suggests that the predictive performance may increase by collecting more training dataset. Next, we will apply our method to the prediction of joint space narrowing score.

References:

Acknowledgments: Izumi and Suzuki are contributed equally.


DOI: 10.1136/annrheumdis-2020-eular.4107