Prediction of Osteoporosis Site Using Artificial Intelligence: A Preliminary Study

Jong Soo HAN [1], Yuna YOUN [1], Ju Hee KIM [1], Hye Jin WON [2], Seung Woo HONG [2], Sharry LEE [1], Ha eun LEE [1], Jae Hyeong KIM [1], Jin-Hyeok HWANG [1]*
[1] Seoul National University Bundang Hospital, Gyeonggi, South Korea;
[2] MINDS AND COMPANY, Seoul, South Korea

Abstract

Osteoporosis is a systemic skeletal disease characterized by a decrease in bone density and abnormal microstructure, and as a result, bones become weak and fragile. If the site of bone loss is accurately predicted in advance by non-invasive methods such as bioelectric impedance analysis and 3D scanner, customized exercises for the site can be started earlier. Prior to model development using the 3D scanner we are developing, we studied whether it is feasible with an artificial intelligence (AI)-learned model using bioelectric impedance analysis data.

An AI model was built using retrospective data from 21328 subjects (2892 males, 18436 females). An AI model was built by extracting 23 features from a retrospective dataset of 21328 people (2892 males, 18436 females). The entire dataset was divided into a ratio of 9:1 between the train dataset and the test dataset. Ten percent of the train set was assigned to the validation dataset. As a result of building conventional machine learning models and applying 5-fold validation, the top 5 models were selected. For these 5 models, soft-voting ensemble was applied and the performance of the resulting model was evaluated by AUROC value.

Among the total study subjects, the proportions of the normal group, the spine bone loss group, and the hip joint bone loss group were 59%, 21%, and 20%, respectively. The top five performance models were Logistic Regression, LightGBM, Catboost, AdaBoost, and XGBoost. The AUROC results, which evaluated the performance of the five models using the test dataset, were 0.7508, 0.7479, 0.7455, 0.7365, and 0.7445 in Logistic Regression, LightGBM, Catboost, AdaBoost, and XGBoost, respectively. The performance of the soft ensemble result with the five models was AUROC 0.7544.

Through post hoc analysis such as SHAP value calculation and feature importance, we try to select features that are important for model learning. Through this, it will be possible to obtain which features to use in the 3D scanner we are currently developing to predict the site of bone loss.

Keywords: Osteoporosis, Artificial intelligence, 3D scanner

* jhhwang@snubh.org