Unbiased Quality Assessments and Full-Automatic Diagnosis for the Developmental Dysplasia of the Hip with Hybrid Machine Learning Algorithms on Infantile Ultrasound Images

Hiokazu Shimizu[1], Tomohiro Shimizu[2], Hidenori Koyano, Ken Enda, Daisuke Takahashi[3], Takuya Ogawa, Shinya Tanaka[4], Norimasa Iwasaki[5]
[1]Department of Orthopaedic Surgery, [2]Department of Orthopaedic Surgery, Faculty of Medi, [3]Graduate School of Medicine, Hokkaido University, [4]Department of Cancer Pathology, [5]Hokkaido University School of Medicine

INTRODUCTION: Screening with ultrasound (US) images is recommended for detecting the developmental dysplasia of the hip (DDH) in infants because of the safeties and real-time acquisition at anyplace. While it is vital to obtain a standard plane of US images, it takes some time for physician to evaluate whether the images are standard. In clinic, those are often validated by well experienced ones, which means that the process is subjective and biased. To simplify screening, we developed hybrid machine learning system combining a convolutional neural networks (CNNs) and image processing for the diagnosis of DDH with unbiased quality assessments (Figure.1). In the first step, CNNs can extract a discriminative point, the apex point. In the next step, image processing search the ridge line of ilium. Finally, alpha angle was produced after the qualified images were selected by evaluating the rotation of the ridge line. The aim is to investigate the accuracy of this system.

METHODS:

We collected a total of 1,099 images based on Graf's method in infants under 6 months old. DDH is defined if this alpha angle is less than 60 degrees. The dataset was randomly divided into training and test set (7:3). First, CNNs learn the apex point as training data and estimate it in the test set. Second, the image processing searched the ridge line from the estimated point. The images are qualified if the ridge line was close to vertical (87-93 degrees). Finally, alpha angle was produced by the same methodology. The accuracy of our system was evaluated using the test set.

RESULTS: There are 258 qualified images in a total of 330 images in the test set. The qualified images had the significantly lower error of inter-rater measurements than the disqualified images (3.91±0.16; 5.31±0.24; $p<0.01$) (Figure. 2). The automatically produced alpha angle revealed that the area under curve of DDH showed 91.8% (Figure. 3). Whereas an orthopaedic surgeon scored sensitivity of 77.2% and specificity of 84.4%, this system scored sensitivity of 81.4% at specificity of 84.4%.

DISCUSSION AND CONCLUSION: Our model showed the unbiased quality assessments for infantile US images. The standard plane was evaluated by the rotation of the ilium while its quality was validated by the significant difference of the error of inter-rater measurements. This meant that it automatically rendered the images quality whereas it is clinically concerned especially in US images. In addition, the automatic diagnoses were made with excellent performance among the dataset which was considered close to clinic since our subjects contained the images which had rotated ilium, mimic structures to the apex point, and noise around bone. The apex point was successfully detected by CNNs, then ilium was described by detecting ridge line of bone regardless of noise. These technologies made the accurate analyses possible. In conclusion, we presented the hybrid machine learning algorithms for unbiased quality assessments as well as the diagnosis of DDH using the infantile US images.