Convolutional Neural Network for the Diagnosis of Pediatric Developmental Dysplasia of the Hip on Conventional Radiography

Yeon Jin Cho¹, Ga Young Choi¹, Seul Bi Lee¹, Young Hun Choi¹, Seung Hyun Lee¹, Jung-Eun Cheon¹, Woo Sun Kim¹, In-One Kim¹, Young Jin Ryu², Jae-Yeon Hwang³, Hyoung Suk Park⁴, Kiwan Jeon⁴

¹Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea

²Department of Radiology, Seoul National University Bundang Hospital, Bundang-gu, Seongnam, Republic of Korea

³Department of Radiology, Pusan National University Yangsan Hospital, Mulgeum-eup, Yangsan-si, Gyeongsangnam-do, Republic of Korea

⁴Division of Medical Mathematics, National Institute for Mathematical Sciences, Daejeon, Republic of Korea
Developmental Dysplasia Of the Hip (DDH):

The most common orthopedic disorder in newborns

Incidence: 1.5 in 1,000

Spectrum of structural abnormalities

(mild acetabular dysplasia ~ dislocation of the femoral head)

Diagnosis and Treatment:

- Modality of choice: Conventional radiography and Ultrasound
- Ossification center of femoral head: 4-6 month of age
- US: < 4-6 months of age
- Conventional radiography: > 4-6 month of age
- Articulation of the femoral head and acetabulum → normal development
- Early diagnosis and treatment are important
Convolutional Neural Network
- tremendous progress; considered to be an emerging technique for the classification of images
- potential of deep learning in the field of lesion detection, classification and image improvement in radiologic image
To evaluate the diagnostic performance of a deep learning algorithm for DDH using conventional radiography.
Materials and Methods

- **Inclusion**
  - Younger than 12 months of age who were suspected of DDH and were undergoing hip AP conventional radiography
  - SNUH: between January 2011 and June 2018
  - SNUBH & PNUYH: between January 2016 and June 2018

- **Exclusion**
  - Inappropriate images for reading
  - Images taken with not proper position
  - Postoperative images
## Dataset
- 2,601 Hip radiographs → 5,202 hip joints images
- Exclusion: 126 inappropriate images
- 5,076 hip joint images → Dataset (Training 80% / Validation 10% / Test 10%)

<table>
<thead>
<tr>
<th>Hospitals</th>
<th>Total</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>DDH</td>
<td>Normal</td>
</tr>
<tr>
<td>SNUH</td>
<td>3433</td>
<td>2406</td>
<td>341</td>
<td>300</td>
</tr>
<tr>
<td>SNUBH</td>
<td>1036</td>
<td>800</td>
<td>32</td>
<td>97</td>
</tr>
<tr>
<td>PNUYH</td>
<td>607</td>
<td>452</td>
<td>19</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>5076</td>
<td>4050</td>
<td>513</td>
<td>513</td>
</tr>
</tbody>
</table>

- **Hospitals**
  - SNUH: 3433
  - SNUBH: 1036
  - PNUYH: 607

- **Dataset**
  - 2,601 Hip radiographs → 5,202 hip joints images
  - Exclusion: 126 inappropriate images
  - 5,076 hip joint images → Dataset (Training 80% / Validation 10% / Test 10%)
- Image Review and Labeling: By two pediatric radiologists in consensus
- Binary Classification: Normal and DDH
- Diagnosis of DDH
  1) high acetabular index (> 30 degree)
  2) abnormal acetabular morphology and delayed femoral head
  3) abnormal femoral head location; out of inferior medial quadrant of acetabulum
  4) disruption of Shenton line
• Image Processing

- Manually cropping: include single hip joint with the femoral head in the center of the cropped image
- To avoid overfitting, the training datasets were augmented (x 10 in DDH, x4 in normal)
- 3,920 DDH and 14,632 normal patches were used for training.
- Training was performed after resizing patch size from 414×414 to 128×128
**Deep Learning Algorithm**

- Tensorflow
- GPU (NVIDIA, Titan Xp. 12GB) system
- Network minimization: using the Adam optimizer
- Learning rate: 0.0001
- Mini-batch size: 16
- Epoch: 100
• **Human Readout**

- Three invited radiologists performed image review
  - Reviewer 1: pediatric experts
  - Reviewer 2: experienced radiologist without experience in pediatric radiology
  - Reviewer 3: inexperienced radiologist without experience in pediatric radiology

- No clinical information, No contralateral Hip image

- **Labeling: 5-point scale**
  - 1, definitely normal; 2, probably normal; 3, indeterminate; 4, probable DDH; and 5, definite DDH
**Materials and Methods**

- **Statistical analysis**
  - **Diagnostic Performance of the deep learning algorithm**
    - construction of 2x2 table
    - calculation of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV)
    - construction of receiver operating curve (ROC) plot and a precision-recall (PRC) plot → calculate area under the curve (AUC)
  - **Comparison with Human Readout**
    - 5-point scale → dichotomization into normal (1, 2) and DΩ (3, 4, 5)
    - calculation of sensitivity, specificity, PPV and NPV
    - McNemar’s test
    - AUC of ROC and PRC plot comparison (algorithm vs. human readout)
<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>AUC of ROC plot</th>
<th>AUC of PRC plot</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deep learning algorithm</strong></td>
<td>94.0 (83.5-98.7)</td>
<td>98.9 (97.5-99.6)</td>
<td>90.4 (79.7-95.8)</td>
<td>99.4 (98.1-99.8)</td>
<td>0.988 (0.974-0.995)</td>
<td>0.979</td>
</tr>
<tr>
<td><strong>Radiologist 1</strong> (p=1.000)</td>
<td>96.0 (86.3-99.5)</td>
<td>99.1 (97.8-99.8)</td>
<td>92.3 (81.9-97.0)</td>
<td>99.6 (98.3-99.9)</td>
<td>0.988 (0.974-0.995)</td>
<td>0.958</td>
</tr>
<tr>
<td><strong>Radiologist 2</strong> (p&lt;0.001)</td>
<td>96.0 (86.3-99.5)</td>
<td>89.0 (85.8-91.7)</td>
<td>48.5 (41.9-55.1)</td>
<td>99.5 (98.1-99.9)</td>
<td>0.959 (0.939-0.975)</td>
<td>0.835</td>
</tr>
<tr>
<td><strong>Radiologist 3</strong> (p&lt;0.001)</td>
<td>84.0 (70.9-92.8)</td>
<td>85.8 (82.2-88.8)</td>
<td>38.9 (33.0-45.1)</td>
<td>98.0 (96.3-98.9)</td>
<td>0.919 (0.892-0.941)</td>
<td>0.618</td>
</tr>
</tbody>
</table>
Results

**ROC curve**

- **Model, AUC = 0.988**
- **Reviewer1, AUC = 0.988**
- **Reviewer2, AUC = 0.959**
- **Reviewer3, AUC = 0.919**

**PR curve**

- **Model, AUC = 0.973**
- **Reviewer1, AUC = 0.958**
- **Reviewer2, AUC = 0.835**
- **Reviewer3, AUC = 0.618**
Conclusion

- The proposed deep learning algorithm provided an accurate diagnosis of developmental dysplasia of the hip on hip AP conventional radiographs, which was comparable to an experienced radiologist.
Thank you for your attentions