Predictive validity of radiographic signs of complete discoid lateral meniscus in children using machine learning techniques

Chul-Won Ha¹ | Seong Hwan Kim² | Dong-Hoon Lee³ | Hyojoon Kim⁴ | Yong-Beom Park³

¹Department of Orthopedic Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
²Department of Orthopedic Surgery, Hyundae General Hospital, Chung-Ang University College of Medicine, Namyangju-si, Gyeonggi-do, South Korea
³Department of Orthopedic Surgery, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, South Korea
⁴Department of Computer Science, Princeton University, Princeton, New Jersey

Abstract
The diagnostic utility of radiographic signs of complete discoid lateral meniscus remains controversial. This study aimed to investigate the diagnostic accuracy and determine which sign is most reliably detects the presence of a complete discoid lateral meniscus in children. A total of 141 knees (age 7-16) with complete discoid lateral meniscus and 141 age- and sex-matched knees with normal meniscus were included. The following radiographic signs were evaluated: lateral joint (LJ) space, fibular head (FH) height, lateral tibial spine (LTS) height, lateral tibial plateau (LTP) obliquity, lateral femoral condyle (LFC) squaring, LTP cupping, LFC notching, and prominence ratio of the femoral condyle. Prediction models were constructed using logistic regressions, decision trees, and random forest analyses. Receiver operating characteristic curves and area under the curve (AUC) were estimated to compare the diagnostic accuracy of the radiographic signs and model fit. The random forest model yielded the best diagnostic accuracy (AUC: 0.909), with 86.5% sensitivity and 82.2% specificity. LJ space height, FH height, and prominence ratio showed statistically large AUC compared with LTS height and LTP obliquity (P < .05 in all). The cut-off values for diagnosing discoid meniscus to be <12.55 mm for FH height, <0.804 for prominence ratio, and >6.6 mm for LJ space height when using the random forest model. On the basis of the results of this study, in clinical practice, LJ space height, FH height and prominence ratio could be easily used as supplementary tools for complete discoid lateral meniscus in children.

KEYWORDS
children, complete discoid lateral meniscus, diagnosis, machine learning, radiograph

1 | INTRODUCTION

Complete discoid lateral meniscus is clinically symptomatic and problematic due to its thicker and larger morphology,¹ with 0.4% to 17% of incidence, which is the most common anatomical variation of the knee joint in pediatric patients.²⁻⁴ A symptom or tear of complete discoid lateral meniscus often occurs in children,⁵ and resulting deteriorated cartilage degeneration if it is not addressed properly.⁶⁻⁸ Therefore, early screening or diagnosis of complete discoid lateral meniscus is beneficial. Currently, complete discoid lateral meniscus is confirmed by magnetic resonance imaging (MRI); however, MRI is costly...
and is sometimes hard to perform on children because it requires them to lie motionless and because they can be fearful in the MRI space.9

The diagnostic accuracy of each radiographic sign for complete discoid lateral meniscus still remains unclear. Several radiographic signs for discoid lateral meniscus in knees have been reported10-13; however, very few studies have evaluated the diagnostic accuracy of radiographic signs in children.13-15 One recent study reported that a prediction equation for radiographic signs could improve their diagnostic value for complete discoid lateral meniscus in children (sensitivity = 79.6% and specificity = 90.4%).14 Despite the clinical utility of a prediction equation for children, it is nevertheless worthwhile to determine which radiographic signs have greater reliable diagnostic value for practical clinical use.

Previous studies have identified meaningful factors using several statistical methods,7,13,16-18 including logistic regression analysis, which was previously the methodological standard,19 but machine learning models are becoming increasingly popular in the medical sciences, especially for prediction or classification problems.20-23 Machine learning models can solve classification problems, develop prediction models, and identify high-risk patients. Moreover, when various machine learning models were previously compared with conventional alternatives, several had significantly better predictive power than conventional analysis using field-specific datasets.20,22,24,25 However, to the best of our knowledge, no previous study has employed machine learning to predict complete discoid lateral meniscus in children.

The purpose of this study was to investigate the diagnostic accuracy of radiographic signs for complete discoid lateral meniscus, and determine which radiographic signs were more valuable for detecting the presence of complete discoid lateral meniscus in children using logistic regression analysis, machine learning models and comparisons among receiver operating characteristic (ROC) curves. It was hypothesized that combination of the radiographic signs and demographic data using machine learning models showed improved diagnostic accuracy for complete discoid lateral meniscus in children compared with that of the conventional analysis.

2 | METHODS

This is a retrospective study of children with complete discoid lateral meniscus. We included children (<18 years old) with arthroscopy-confirmed complete discoid lateral meniscus. From January 1995 to February 2016, 164 knees from 138 consecutive patients were arthroscopically diagnosed with complete discoid lateral meniscus. Among the diagnosed knees, we excluded 13 (12 patients) that were not accompanied by full radiographs and 10 from 7 patients who were <7 years old and did not appear to have a fibular ossification center. Consequently, 141 knees (119 patients, range: 7s, <18) with complete discoid lateral meniscus were included in this study.

Normal control subjects were randomly sampled from our institution’s patient pool by age and sex matching with subjects in the study group. Control subjects were confirmed to have normal lateral meniscus based on arthroscopy or MRI regardless of other intraarticular pathologies. Consequently, 141 knees (141 patients) were included in the normal control group.

The study protocol was approved by the Institutional Review Board of investigational institution (No. 2015-07-111).

2.1 | Radiographic analyses

We obtained weight-bearing anteroposterior, nonweight-bearing lateral, and tunnel view radiographs. We measured the following radiographic signs for complete discoid lateral meniscus using previously described methods.10,14-15 Figure 1 shows lateral joint (LJ) space height, fibular head (FH) height, lateral tibial spine (LTS) height, lateral tibial plateau (LTP) obliquity, lateral femoral condyle (LFC) squaring, LTP cupping (LTP cupping), LFC notching, and prominence ratio of the femoral condyle (condylar cut-off sign).

LJ space was the shortest distance between the tibial joint line and the femoral condylar joint line (Figure 1A). FH height was the distance between the parallel line to the tibial joint line at the tip of the FH and the tibial joint line (Figure 1A). LTS height was the distance between the tibial joint line and the tip of the LTS (Figure 1A). LTP obliquity was the angle between the tibial joint line and the line drawn along the lateral slope of the LTS (Figure 1B). LFC squaring was defined as positive when the straightened length of the LFC articular surface was greater than 10 mm (Figure 1B). LTP cupping was defined as positive when LTP depression from the tibial joint line was greater than 1 mm (Figure 1B). LFC notching was defined as positive if the distance from the tangential line of the articular surface of the lateral condyle to the notch was greater than 1 mm (Figure 1C).

The prominence ratio (condylar cut-off sign) was defined as the ratio of prominence of the lateral and the medial femoral condyles adjacent to the intercondylar notch (Figure 1D).11 Lines A (A') were drawn through the outermost points of the femoral condyles medially and laterally. Line B was drawn through the distal lowest points of the femoral condyles. Points a and a' were the points where lines A (A’) meet line B. Point c was the highest point in the intercondylar notch. Lines C (C’) were lines through points c and a (a’). The prominences P (P’) were measured as the longest vertical distance between the prominences of the condyles and lines C (C’).

The prominence ratio was defined as P/P’, ratio of the prominence of LFC to the prominence of medial femoral condyle. The two observers were blinded to the state of lateral meniscus. Measurements were repeated with an interval of 4 weeks. All radiographic signs were measured by two orthopedic surgeons using GE Centricity PACS (General Electric, Milwaukee, WI).

2.2 | Statistical analyses

We used SPSS 19.0 (Chicago, IL), R version 3.4.5 (R Foundation) to run statistical analyses, and we conducted power analyses for our
statistical tests using the G*Power program (version 3.1.5). The R packages we used were caret, pROC, ROCR, rpart, rpart.plot, tree, party, randomForest, caTools, devtools, and repreTree. Patient demographics between the groups were compared using the Shapiro-Wilk test for normality, followed by the $t$ test. We used the $\chi^2$ test to analyze categorical variables. ROC curves and area under the curve (AUC) were plotted and compared with evaluate the diagnostic accuracy of the radiographic measurements for discoid lateral meniscus.

To estimate the predictive ability of each model, the complete dataset was randomly divided into two subsets, a training sample (80%, 221 cases) and a test sample (20%, 61 cases). We performed stepwise logistic regression analysis with backward elimination to evaluate the predictive variables for discoid lateral meniscus using the AIC values.

For our decision tree analyses, the unpruned decision tree was first grown using the Gini impurity score with the Classification and Regression Tree method, then the pruning was performed with x-error value. To assess the predictive ability of the final decision tree, we used the test set to assess accuracy and variable importance.

For the random forest analysis, each tree grows on a bootstrap sample (a random sample selected with replacement, "bagging"). Individual trees are also unpruned classification/decision trees grown using the Gini impurity score, and ensemble
the trees as on tree model, which defines as random forest. Random forest analysis estimates the error rate based on out-of-bag (OOB) data. To assess the predictive ability of the final decision tree, the test set was also used to assess accuracy and variable importance using ROC analysis. ROC curves and AUC were also plotted to comparatively evaluate the diagnostic accuracy of machine learning and logistic regression models. The inter- and intraobserver reliability of each variable’s measurements was analyzed by intraclass correlation coefficients (ICCs) for continuous variables and kappa’s coefficient for nominal variables.

3 | RESULTS

3.1 | Subject characteristics

A total of 141 knees in the complete discoid group and 141 knees in the normal control group were analyzed (Table 1). LJ space in discoid group was significantly larger than that of normal group, but the FH, LTP obliquity and condylar cut-off sign was significantly smaller than that of normal group. Moreover, LFC squaring was significantly found more frequent in discoid group than normal group.

3.2 | Logistic regression analysis

The presence of LFC squaring was most strongly associated with complete discoid meniscus (odds ratio, 2.499; 95% confidence interval, 1.09-5.729; P = .031) (Table 2). The decreased value of FH (high FH clinically), decreased LTS height and condylar cut-off sign, and increased LJ space are also associated with the presence of complete discoid meniscus.

3.3 | Comparisons between machine learning models and logistic regression model

For the decision tree model, we established a pathway after tree pruning to avoid overfitting by the results of x-error values (Figures 2A and S1). The FH and condylar cut-off sign were the most important variables in the decision tree model (Figure 2B).

For the random forest model, the model with the following hyperparameters was established by a grid search (Table 3; Figure S2). The error rate by OOB was estimated to be 20.6%. In the importance plot for the random forest model, LJ space, FH, and condylar cut-off sign were important factors (Figure 3A). In our partial dependence plots, around 6 mm of LJ space (Figure S3A), a ratio of around 0.8 for the condylar cut-off sign (Figure S3B), and around 13 mm of FH (Figure S3C) were in the cut-off value ranges when using the random forest model to predict discoid meniscus presence. The ensemble tree of a random forest model was shown in Figure 3B.

To compare the final performance of all the models, we calculated accuracy, sensitivity, specificity, kappa, and AUC (Table 3). The random forest model yielded the best performance in terms of accuracy, sensitivity, specificity, and the largest AUC (P < .000). The logistic regression model performed better than the decision tree model (Table 3; Figure 4), worse than the random forest model. Ranked by AUC, model performance was as follows: random forest, logistic regression, and decision tree, and all AUC comparisons were significantly different (P = .019, logistic vs decision tree; P < 0.000, RF vs all others).

3.4 | Comparison of the ROC curves for radiographic measurements and criterion values

AUCs were calculated to estimate the diagnostic accuracy of each radiographic sign (Table 4; Figure 5). LFC squaring, LTP cupping, and LFC notching was not statistically significant. Comparing ROC curves, FH, condylar cut-off sign, and LJ space were not statistically different from each other (P > .05) but were statistically different from the curves for LTP obliquity and LTS height (P < .05; Figure 5).

The ICC value for inter- and intraobserver reliability for continuous radiographic variable measurements was >0.8 (range,
0.846-0.982), indicating that all measurements had excellent inter- and intraobserver reliability. The kappa value for inter- and intraobserver reliability for nominal radiographic variable measurements (LFC squaring, LTP cupping) was >0.6 (range, 0.655-0.668), indicating good agreement, except for LFC notching, which yielded poor agreement (kappa, 0.122; \( P < .001 \)).

### TABLE 2
The predictive variables for the discoid lateral meniscus using logistic regression analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Odds ratio</th>
<th>B value ± SE</th>
<th>95% Confidence interval</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The risk of the presence of discoid meniscus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral joint space height (mm)</td>
<td>1.58</td>
<td>0.457 ± 0.08</td>
<td>1.352-1.847</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Height of the fibular head (mm)</td>
<td>0.773</td>
<td>-0.257 ± 0.05</td>
<td>0.701-0.853</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Height of the lateral tibial spine (mm)</td>
<td>0.689</td>
<td>-0.372 ± 0.112</td>
<td>0.553-0.859</td>
<td>.001</td>
</tr>
<tr>
<td>Condylar cut-off sign</td>
<td>0.000</td>
<td>-9.43 ± 1.637</td>
<td>0.000-0.002</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Squaring of the lateral femoral condyle</td>
<td>2.499</td>
<td>0.916 ± 0.423</td>
<td>1.09-5.729</td>
<td>.031</td>
</tr>
<tr>
<td>Constant term</td>
<td>...</td>
<td>10.413 ± 1.853</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

### DISCUSSION

Several radiologic findings have been reported to indicate the presence of discoid lateral meniscus, but their diagnostic utility remains unclear. In addition, there was no consensus on which radiographic sign could be a valuable screening tool for the presence of the discoid lateral meniscus. The predictive variables for the discoid lateral meniscus using logistic regression analysis are presented in Table 2. The decision tree, as shown in Figure 2, indicates that the presence of discoid lateral meniscus can be predicted using the risk of the presence of discoid meniscus, lateral joint space height, height of the fibular head, height of the lateral tibial spine, condylar cut-off sign, squaring of the lateral femoral condyle, and constant term. The decision tree shows that the most significant factors are the fibular head (FH) and the condylar cut-off sign, consistent with the decision tree.
of complete discoid lateral meniscus in children. In this study, the machine learning models and comparison of ROC curves suggested that the combination of the radiographic signs and demographic data could show higher diagnostic values in children. The most important finding of this study was that LJ space, FH, and condylar cut-off sign were significant predictors of the complete discoid lateral meniscus in skeletally immature patients according to the random forest model and ROC curve analysis.

According to the random forest model, LJ space, FH, and condylar cut-off sign were the most important predictors of complete discoid meniscus; combined use of these measurements could increase diagnostic accuracy compared with those of other prediction models. Moreover, <0.8039 of the condylar cut-off sign, <12.55 mm of FH and >6.6 mm of LJ space could be used as cut-off values for complete discoid meniscus according to ROC analysis (Table 4). These results were confirmed by the partial dependence plot using the random forest model, of which the ROC curves were within the significant range of coefficients. However, the final ensemble tree of a random forest model (Figure 3B) seems to have too many pathways, the clinical relevance of this tree of a random forest model is questionable when assessing each variable. Nonetheless, when using this random forest model along with the computer science technique, surgeons can predict the probability of the presence of discoid lateral meniscus to determine whether further MRI evaluation or surgery would be needed or not. Here, we present the web-based prediction application for discoid lateral meniscus in children by the random forest model (http://discoidpred.com).

Moreover, although the decision tree model's performance was relatively poor in terms of the AUC (Table 3), it can be applied easily in clinical practice to detect complete discoid lateral meniscus in cases with a <12.58-mm FH and <0.7821 ratios for a condylar cut-off sign, if someone wants to use simple variables. No variables could be entirely independent clinically, although that is an assumption of the statistical models; thus, significant factors in conventional logistic regression might not be significant in machine learning models.

This study revealed that LJ space, FH, and the condylar cut-off sign were significantly more accurate than the other radiographic signs. In this study, the AUC of LJ space, FH, and the condylar cut-off sign were significantly larger than those of LTP obliquity and LTS height. In addition, LFC squaring, LTP cupping, and LFC notching was inappropriate for predictive variables because of their low ICC, which meant that these variables could induce a subjective assessment. Some previous studies have evaluated the utility of radiographic signs for the diagnosis of the complete discoid lateral meniscus in children.13-15 One found that LJ space, FH, and LTP obliquity were helpful for screening for discoid lateral meniscus.13 Another study found that the condylar cut-off sign was useful for detecting complete discoid lateral meniscus.15 The other study reported that an equation combining LJ space, FH, LTS height, and the condylar cut-off sign could improve the diagnostic accuracy for complete discoid lateral meniscus.14 Considering these previous studies' results alongside our results here, using only one radiographic sign yields limited diagnostic capability for complete discoid lateral meniscus. However, a combination of two or three radiographic signs in a clinical settings is time-consuming. Therefore, we compared the AUCs of the radiographic

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Performance of logistic regression, decision tree, and random forest model with process of establishing model</th>
<th>Prediction error for test set</th>
<th>Description on hyperparameter</th>
<th>AUC (95% CI)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy (95% CI)</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression model</td>
<td>0.848 (0.804-0.891)</td>
<td>70.7%</td>
<td>80.8%</td>
<td>0.809 (0.734-0.874)</td>
<td>0.424</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision tree model</td>
<td>0.791 (0.741-0.842)</td>
<td>66.7%</td>
<td>80.9%</td>
<td>0.397</td>
<td>20.6%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random forest model</td>
<td>0.909 (0.875-0.943)</td>
<td>86.8%</td>
<td>82.2%</td>
<td>0.789 (0.721-0.863)</td>
<td>0.413</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AUC, area under the curve; CI, confidence interval.
signs to identify valuable predictors in clinical practice; we found that LJ space (cut-off value, 6.6 mm), FH (cut-off value, 12.5 mm), and prominence ratio (condylar cut-off sign, cut-off value, 0.8) were most effective for complete discoid lateral meniscus detection in children.

Our study has several limitations. First, differences in anthropometry and skeletal maturation between children could influence distinct features of the condylar cut-off sign in knees with complete discoid lateral meniscus. However, as mentioned above, condylar cut-off sign is the only parameter that uses the index of ratios in radiographic diagnoses; thus, this sign should not be significantly affected by anthropometric differences. Second, machine learning models can be difficult to interpret, especially the random forest model. Although the random forest model usually yielded high predictive performance, the interpretation and implication for clinical practice were very challenging due to the nature of the analysis and assembly of trees. Therefore, we analyzed the ROC curves for individual covariates to find the cut-off

**FIGURE 3** A. The importance plot from the random forest model: lateral joint space, condylar cut-off sign, and fibular head were the important factors for detection. B. The final ensemble tree of random forest was plotted ("0" was set as complete discoid patient, "1" was set as normal subject). There were many pathways to predict complete discoid meniscus

**FIGURE 4** The comparison of ROC curves for prediction performance of the models: the random forest model yielded the largest area under the curve. ROC, receiver operating characteristic

[Color figure can be viewed at wileyonlinelibrary.com]
values. Third, the measured values might be changed according to the height or limb length, although the age was adjusted in this study. Fourth, the accuracy of the model might be changed according to the sampling of the train set. If the validation was performed with external data, the results of prediction model would be more stable. Finally, the poor agreement for LFC notching could have affected the logistic regression results, even though those results were not significant. However, the machine learning models can overcome poor agreement issues by the nature of their analysis. Moreover, the machine learning results also indicated that LFC notching was not helpful for predicting complete discoid meniscus.

5 | CONCLUSION

Prediction by the random forest model combined with radiographic signs yielded much higher diagnostic value than by decision tree model or conventional logistic regression. Among several radiographic signs, LJ space height, FH height, and prominence ratio of the femoral condyle were the best predictive factors for detecting complete discoid lateral meniscus in children. On the basis of the results of this study, these three radiographic signs could be used as supplementary diagnostic tools for complete discoid lateral meniscus in children.

ACKNOWLEDGMENTS

This study was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant number: HI14C3484). The funding...
sources were not involved in the study design, collection, analysis or interpretation of the data, writing of the manuscript, or in the decision to submit the manuscript for publication.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

All authors have read and approved the final submitted manuscript. Contributions to authorship are as follows: C-WH was responsible for the design of the study, and preparing the initial drafting of this manuscript. SHK was responsible for analysis and interpretation of data, drafting this manuscript, and revising this manuscript. D-HL was responsible for the acquisition data, and data processing. HKU was responsible for analysis data, and predictive model development. Y-BP contributed to the initial concept for this study, analysis and interpretation of data, drafting this manuscript, and revising this manuscript.

ORCID

Chul-Won Ha https://orcid.org/0000-0001-5123-6513
Seong Hwan Kim https://orcid.org/0000-0001-5471-5747
Dong-Hoon Lee https://orcid.org/0000-0003-4129-1855
Hyojoon Kim https://orcid.org/0000-0002-0650-5339
Yong-Beom Park https://orcid.org/0000-0002-3741-2311

REFERENCES

27. Grajski KA, Breiman L, Viana Di Prisco G, Freeman WJ. Classification of EEG spatial patterns with a tree-structured methodology: CART.
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.