Application of Decision Tree Analysis for Pediatric Foot Disorders

Jungkyu Choi1, Hwa-In Kim2, Yonggwan Won3, and Jung-Ja Kim4,5, *

1Department of Healthcare Engineering, Graduate school
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
Republic of Korea
chaosjoozak@nate.com

2The Foot Clinic of Jeonju Pediatrics
485-42, Junghwasan-dong 2, Wansan-gu, Jeonju, Jeollabuk-do
Republic of Korea
mssy1016@yahoo.com

3School of Electronics and Computer Engineering
Chonnam National University
77 Yongbong-ro, Buk-gu, Gwangju, 500-757
Republic of Korea
ykwon@chonnom.ac.kr

4Division of Biomedical Engineering
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
Republic of Korea

5Research Center of Healthcare & Welfare Instrument for the Aged
Chonbuk National University
664-14, Deokjin-dong 1, Deokjin-ku, Jeonju, Jeolabuk-do, 561-756
Republic of Korea
jungjakim@jbnu.ac.kr

Abstract: - In modern medical field, the explosive increase of clinical data has happened by development of computer information technology and tools. For a large amount of clinical data, data mining, the extraction method of hidden predictive information, has been recognized by many studies. The object in the study was to discover meaningful knowledge between the foot disorder and biomechanical parameters related to symptom by generating a prediction model of the decision tree. The first medical record data of 174 pediatric patients was extracted for analysis, in total 279 records, and they were diagnosed with a complex foot disorder. The dependent variable consists of five complex disorder groups, and 14 independent variables related to disorder groups were selected by importance, in 34 variables. The extracted data was separated to generate an ideal prediction model. After development of the prediction model, the prediction rate was verified. Consequently, a major symptom information in 13 diagnosis patterns was confirmed. After then, the detailed preprocessing and analysis will be performed to improve the accuracy of the classification.

Key-Words: Pediatric Foot, Lower limb, Disorder, Pattern classification, Decision tree.

1 Introduction

The bipedalism, including walking, running and jumping, is the most fundamental human activity, and a natural behavior that anyone is able to perform easily in everyday life, if normal [1]. For this movement, the lower limbs, including foot, are such an important organ of the body. However, the close collaboration of different skeletal muscles, joint and nervous system matter for a well-stabilizing gait from one point to another [2]. That is why the foot has a highly complex structure composed of 28 bones, 55 small joint and 23 muscles, although it makes up just 5% of the whole body surface. When a human walks for 1 km, there is about 15 t weight-bearing increase on the foot. In addition, the weight-bearing with push-off exercise of gait causes stress or soft tissue strain on the lower limbs. This problem deforms the leg and foot shape,
and the abnormal lower limbs have a bad influence on the balance of the spine and pelvis [3]. In case of children, the level of deformation is quite different from that of adult because pediatric foot have different characteristics in structure and function [4]. For this reason, Muller, Carlssohn, Muller, Baur, Mayer performed the study to acquire static and dynamic foot characteristics in childhood, and to establish data for age groups of a population of 1–13 year-old infants and children based on a cohort of 7788 subjects [5]. Although there is a difference by a variety of causes, the foot usually grow up by age 5 and 7 quickly. Since then, they keeps growing at a constant rate by age 10 and 14[6]. During that time, the foot shape is changing from pes planus to normality, and the leg shape also alters in the same order as follows: genu varum, genu valgum and normality [7]. If the shape of the lower limb does not become normality until about 12 years old, it is more likely to be cause of the adult foot disorder. In addition, the treatment of non-invasive method is only efficient in childhood. Jay, Schoenhaus, Seymour, Gamble confirmed that there was significant improvement in the resting calcaneal stance position (RCSP) of children, aged 20 months to 14 years with pes planus, who were prescribed with a custom-made orthosis [8]. Lincoln, Suen noted that out-toe gait was observed in children with pes planus, and its patterns may result from abnormal conditions of the hip, tibia and femoral region [9]. There are close connections between abnormal shape and cause of various disorders in the lower limbs [10]. However, the disorders appear complexly and symptoms are not clear, on the average. Accordingly, more intelligent analysis is necessary to figure out pattern of symptoms.

In modern medical field, the explosive increase of clinical data has happened by development of computer information technology and tools [11]-[12]. The clinical data contains quantitative data (eg, laboratory values), qualitative data (eg, text-based documents and demographics), and transactional data (eg, a record of medication delivery) [13]. When utilization of medical big data, value production of 330 billion dollars is expected every year on the US medical field. If effective treatment method by analysis data of diagnostic pattern, prognosis, cost, etc., direct effect of about 165 billion dollars is expected [14]. For a large amount of clinical data, data mining, the extraction method of hidden predictive information, has been recognized by many studies [15]. It is a method to handle large data and to find out desired important and meaningful knowledge with utilizing pattern recognition technology, statistics technique or mathematic algorithm [16]. Decision-tree, which is a key issue of representative technique in the data mining, is an algorithm to classify or predict a couple of subgroup from interested object group by modeling rule and observing relation [17]. This method is a model of decisions and a special form of tree structure. Therefore, it has advantage to understand analysis process and results easily [18]-[19]. According to previous studies, data mining was adopted for the analysis of medical data. Breault, Goodall and Fos applied Classification and Regression Trees (CART) of data mining for the analysis of diabetic data warehouse. They figured out that the most important variable associated with bad glycemic control was younger age, not the comorbidity index or whether patients had related disorders [20]. Kim presented that age, associated disorder, pathology scale, course of hospitalization, respiratory failure and congestive heart failure were related to dangerous factors on death of pneumonia by using data mining for analysis of death factor on pneumonia patient [21]. Stoean, Stoean, Lupsoc, Stefanescu, and Badea reported that the evolutionary-driven support vector machine of data mining was utilized to anticipate stage of hepatic fibrosis that determine hardness degree of liver or operation in chronic hepatitis C [22]. Lim, Ryu, Park and Ryu the logistic regression and neural network were applied to extract attribute and perform learning based on widespread clinical data of acute myocardial infarction for forecast short-term relapse mortality of ST-segment elevation myocardial infarction (SEMI) patients. Through this study, the model to foresee short-term mortality of SEMI patients was suggested [23]. In addition, the four decision tree algorithms were used to analyze postoperative status of ovarian endometriosis patient under different conditions. This study reported new meaningful information about recurrent ovarian endometriosis [12]. However, most previous study about the lower limbs just noted simple comparison analysis based on quantitative values. In addition, a study in respect of the pediatric foot disorder is insufficient, though the symptoms are commonly complicated. More integration analysis with data mining technique is necessary, and the interpretation of interconnection between several clinical parameters and the foot disorder is important. Accordingly, the purpose of this study was to find out significant knowledge between the foot disorder groups and biomechanical parameters related to symptom on the basis of the pediatric clinical data in the Foot clinic by developing a prediction model of the decision tree.
2 Method

2.1 Subjects
The first examination clinical data of total 279 pediatric patients diagnosed with complex disorder, including pes planus basically, was used from the Foot Clinic of Jeonju Pediatrics. To diagnose disorder, total 34 attributes, Resting Calcaneal Stance Position (RCSP), the Tibia TransMalleolar Angle (Tibia TMA), the Knee Internal Malleolus Distance (Knee IMD), etc., were measured and patient charts were made up by a podiatrist, as shown in Fig. 1. 64 patients records with missing values were excluded, and complex disorder groups above 5% of data were selected for analysis. Analysis data was composed of 174 patient records with five groups for the complex disorder.

![Fig 1. Measurement of RCSP and patient charts](image)

2.2 Variables
A dependent variable in the study consisted of five complex disorder group such as A: Pes planus and Achilles tendinitis, B: Pes planus, C: Pes planus and Intoe gait, D: Pes planus, Intoe gait and Genu valgum and E: Pes planus and Genu valgum, as shown in Table 1. An independent variable was preprocessed through statistical validity and importance analysis. Therefore, 14 of 34 independent variables related to disorder closely were selected and optimized, as shown in Table 2.

**Table 1. Dependent variable**

<table>
<thead>
<tr>
<th>Class</th>
<th>Disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Pes planus, Achilles tendinitis</td>
</tr>
<tr>
<td>B</td>
<td>Pes planus</td>
</tr>
<tr>
<td>C</td>
<td>Pes planus, Intoe gait</td>
</tr>
<tr>
<td>D</td>
<td>Pes planus, Intoe gait, Genu valgum</td>
</tr>
<tr>
<td>E</td>
<td>Pes planus, Genu valgum</td>
</tr>
</tbody>
</table>

**Table 2. Independent variable**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Nominal</td>
<td>Male, Female</td>
</tr>
<tr>
<td>(L) TibiaTMA</td>
<td>Numeric</td>
<td>Angle of the left tibia transmalleolar</td>
</tr>
<tr>
<td>(R) TibiaTMA</td>
<td>Numeric</td>
<td>Angle of the right tibia transmalleolar</td>
</tr>
<tr>
<td>KneeIMD</td>
<td>Numeric</td>
<td>The knee internal malleolus distance</td>
</tr>
<tr>
<td>(L) Talocalcaneal</td>
<td>Numeric</td>
<td>Left angle between the talus and the calcaneus</td>
</tr>
<tr>
<td>(R) Talocalcaneal</td>
<td>Numeric</td>
<td>Right angle between the talus and the calcaneus</td>
</tr>
<tr>
<td>(L) CuboidAbduction</td>
<td>Numeric</td>
<td>Left angle of the cuboid abduction</td>
</tr>
<tr>
<td>(R) CuboidAbduction</td>
<td>Numeric</td>
<td>Right angle of the cuboid abduction</td>
</tr>
<tr>
<td>(L) Intermetatarsal</td>
<td>Numeric</td>
<td>Left angle of the metatarsus primus adductus</td>
</tr>
<tr>
<td>(R) Intermetatarsal</td>
<td>Numeric</td>
<td>Right angle of the metatarsus primus adductus</td>
</tr>
<tr>
<td>(L) TalarDeclination</td>
<td>Numeric</td>
<td>Angle of the left talus declination</td>
</tr>
<tr>
<td>(R) TalarDeclination</td>
<td>Numeric</td>
<td>Angle of the right talus declination</td>
</tr>
<tr>
<td>(L) RCSP</td>
<td>Numeric</td>
<td>Left Resting calcaneal stance position angle</td>
</tr>
<tr>
<td>(R) RCSP</td>
<td>Numeric</td>
<td>Right Resting calcaneal stance position angle</td>
</tr>
</tbody>
</table>
The first clinical data of 174 pediatric patients with complex disorder, including pes planus basically, was utilized for the study. The prediction model was created to analyze the pattern of the foot disorder group by applying the C5.0 algorithm. The measured prediction rate was Correct: 92.44 % and Wrong: 7.56 % in the training data, and Correct: 74.07 % and Wrong: 25.93 % in the test data.

As a result of analysis on five complex disorder groups by using decision tree, 13 rules were discovered: (1) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was below -8°, ‘(R) Cuboidabduction’ was below -2°, ‘(R) RCSP’ was below -8 and ‘(R) Intermetatarsal’ was above 9°, then ‘A’, (2) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was below -8° and ‘(R) Cuboidabduction’ was above 2°, then ‘B’, (3) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was below -8° and ‘(R) Cuboidabduction’ was below 2°, then ‘B’, (4) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was below -8° and ‘(R) Cuboidabduction’ was below 2°, then ‘B’, (5) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Cuboidabduction’ was below 7° and ‘(R) Intermetatarsal’ was above 7°, then ‘A’, (6) ‘(L) Tibia TMA’ was above -6°, ‘(R) Cuboidabduction’ was below 7° and ‘(R) Intermetatarsal’ was above 7°, then ‘B’, (7) ‘(L) Tibia TMA’ was above -6°, ‘(R) Cuboidabduction’ was below 7° and ‘(R) Intermetatarsal’ was above 7°, then ‘B’, (8) ‘(L) Tibia TMA’ was above -6°, ‘(R) Cuboidabduction’ was below 7° and ‘(R) Intermetatarsal’ was above 7°, then ‘B’, (9) ‘(L) Tibia TMA’ was above -6°, ‘(R) Cuboidabduction’ was below 7° and ‘(R) Intermetatarsal’ was below 7°, then ‘B’, (10) ‘(L) Tibia TMA’ was below -6° and ‘(R) KneeIMD’ was below 3 cm, then ‘C’, (11) ‘(L) Tibia TMA’ was above -6° and ‘(R) KneeIMD’ was below 3 cm, then ‘C’, (12) ‘(L) Tibia TMA’ was below -6° and ‘(R) KneeIMD’ was below 3 cm, then ‘D’, (13) ‘(L) Tibia TMA’ was above -6° and ‘(R) KneeIMD’ was below 7°, then ‘E’, as shown in Fig. 3 and Fig. 4.
Fig. 4. The result of decision tree
4 Conclusion

The purpose in the study was to find out meaningful knowledge between the foot disorder groups and biomechanical parameters related to symptom on the basis of the pediatric clinical data by developing a prediction model of the decision tree. The first examination clinical data of 174 pediatric patients diagnosed with complex disorder including pes planus basically was used for analysis. The dependent variable consisted of five groups, and the 14 independent variables were selected by importance. The analysis data was partitioned into training data and test data to generate an ideal prediction model. After developing the prediction model by C5.0 algorithm, the prediction rate was verified.

In conclusion, we were able to confirm that the variable of each node was a key diagnosis factor to discriminate the foot disorder. As follow the rules of result, major symptom pattern information of disorder was confirmed as follows. The class A had two patterns; (a) the left tibia transmalleolar angle above -6°, the right intermetatarsal angle above 4°, the knee internal malleolus distance below 4 cm, left resting calcaneal stance position angle below -8°, the right cuboid abduction angle below -2° and right resting calcaneal stance position angle above -8°, (b) the left tibia transmalleolar angle above -6°, the knee internal malleolus distance above 4 cm, the left cuboid abduction angle below 7° and the right intermetatarsal angle above 7°. The class B also had two patterns; (a) the left tibia transmalleolar angle above -6°, the right intermetatarsal angle above 4°, the knee internal malleolus distance below 4 cm, both resting calcaneal stance position angle below -8°, the right cuboid abduction angle below -2° and the right intermetatarsal angle above 9°, (b) the left tibia transmalleolar angle above -6°, the right Intermetatarsal angle above 4°, the knee internal malleolus distance below 4 cm, left resting calcaneal stance position angle above -6° and the right talocalcaneal angle above 28°. The class C had one pattern; (a) the left tibia transmalleolar angle bellow -6° and the knee internal malleolus distance bellow 3 cm. In case of the class D, it had one pattern; (a) the left tibia transmalleolar angle bellow -6° and the knee internal malleolus distance above 3 cm. The class E also had one pattern; (a) the left tibia transmalleolar angle above -6°, the right intermetatarsal angle above 4°, the knee internal malleolus distance above 4 cm and the left cuboid abduction angle above 7°.

The symptom of the foot disorder was commonly complicated, not obvious. In case of children, especially, classification of disorder was more difficult than adult due to the soft bones or growth. For these reasons, the error rate of the prediction rate was relatively high in test data. Therefore, detailed preprocessing and analysis will be performed to improve accuracy of classification. Also, other decision tree algorithms will be applied to develop additional model and carried out comparison analysis for an ideal model from now on.

Acknowledgments

This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013R1A2A04016782)

References:


