Application of Decision Tree and Neural Network for Diagnosis and Prescription of Pediatric Foot Disorders

Jungkyu Choi, Hee-Sang Lee, and Jung-Ja Kim

Abstract—Decision trees and neural networks are typical data classification methods in data mining methods. Decision trees are so fast and interpretable, but these can be misclassified. Neural networks are slow while more reliable algorithms. In this paper, we have studied an intelligent system that diagnose and prescribe patients with pediatric foot disorder using decision tree and neural network. The object of this study was to discover meaningful knowledge between the foot disorder and biomechanical parameters related to symptoms using C5.0 decision tree and neural network. The first medical record data of 174 pediatric patients were 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 and neural networks were applied for analysis of predictor importance and classification prediction. Consequently, a major symptom information in 13 diagnosis patterns were confirmed.

Keywords—Pediatric foot, foot disorder, data mining, decision tree, neural network.

I. INTRODUCTION

Currently, the medical paradigm is changing with the era of the fourth industrial revolution. Especially in the field of clinical medicine, it is changing the direction of preventive medicine and personalized treatment through intelligent data analysis technology in the determination of the diagnosis and treatment methods for patients. In modern medical fiend, the explosive increase of clinical data has happened by development of computer information technology and tools [1-2]. 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) [3]. 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 million dollars is expected [4]. For a large amount of clinical data, data mining, and extraction method of hidden predictive information has been recognized by many studies [5]. 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 [6]. 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 [7]. This method is a model of decisions and a special form of tree structure, so it has advantage to understand analysis process and results easily [8-9]. Neural networks, are composed of simple elements similar to biological nervous system [10]. This methodology has a character that a pre-specification is not required during the modeling process because it independently learn the relationship inherent in the variables. Neural networks also offer the flexibility of numerous architecture types, learning algorithms, and validation procedures. Therefore, these methods have been applied to the various fields [11].

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 daily life, if normal [12]. 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 [13]. 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 [14]. In case of children, the level of deformation is quite a
different from that of adult because pediatric foot have different characteristics in structure and function [15]. For this reason, previous research 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 years infants and children based on a cohort of 7788 subjects [16]. 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 [17]. 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 varam, genu valgum and normality [18]. 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. It was 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 [19]. Out-toe gait was observed in children with pes planus, and its pattern may result from abnormal conditions of the hip, tibia and femoral region [20]. There are close connections between abnormal shape and cause of various disorders in the lower limbs [21]. 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.

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 [22]. 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 [23]. Stoean, Lupson, 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 [24]. Lim, Ryu, Park, and Ryu the logistic regression and neural networks 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 [25]. 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 [2]. 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 disorders is important.

In medicine and nuclear fields, domain experts should validate a model before it can be used in production. We also focus in the data-mining task with decision trees to generate descriptive patterns based on a set of characteristics selected by our domain expert. The artificial neural network outcome predictions may be used as an additional source of information to assist physicians and patients in selecting the treatment that provides the best quality of life for the patient and reduces treatment decision regret.

In this paper, we have studied an intelligent system that diagnoses and prescribes patients with pediatric foot disorder using decision tree and neural network. The object in this study was to discover meaningful knowledge between the foot disorder and biomechanical parameters related to symptoms using C5.0 decision tree and neural network.

II. METHODS

A. 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, Republic of Korea. To diagnose disorder, total 34 attributes, Resting Calcaneal Stance Position (RCSP), the Tibia Trans Malleolar 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.

B. Experimental Procedure

A dependent variable in the study consists of five complex
disorder group such as A: Pes planus and Achilles tendinitis, B: Pes palnus, 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 processed 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.

C. Data Analysis

In this study, combination of independent variables meant each of the foot disorder group. Accordingly, 14 independent variables were inserted into C5.0 algorithm of decision tree at the same time. Data analysis was performed by IBM SPSS statistics 18 (SPSS Inc., Chicago, IL, USA) and IBM SPSS Modeler 14.2 (SPSS Inc., Chicago, IL, USA). For generating an ideal model, it was effective to develop a couple of the prediction model and perform comparison analysis [26]. Consequently, the entire data was partitioned into training data (70 %) and test data (30 %) at random, and the prediction models of data were developed by C5.0 algorithm, as shown in Fig. 2. The prediction rate was verified through the analysis node after development of model. The prediction model of tree-structured decision tree is comprised of organization as ‘If A, then B. Else B2’ [27]. After development, neural networks were applied for analysis of predictor importance and classification prediction based on the prediction model.

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>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male, Female</td>
</tr>
<tr>
<td>(L) TibiaTMA</td>
<td>Angle of the left tibia transmalleolar</td>
</tr>
<tr>
<td>(R) TibiaTMA</td>
<td>Angle of the right tibia transmalleolar</td>
</tr>
<tr>
<td>KneeIMD</td>
<td>The knee internal malleolous distance</td>
</tr>
<tr>
<td>(L) Talocalcaneal</td>
<td>Left angle between the talus and calcaneus</td>
</tr>
<tr>
<td>(R) Talocalcaneal</td>
<td>Right angle between the talus and calcaneus</td>
</tr>
<tr>
<td>(L) Cuboidabduction</td>
<td>Left angle of the cuboid abduction</td>
</tr>
<tr>
<td>(R) Cuboidabduction</td>
<td>Right angle of the cuboid abduction</td>
</tr>
<tr>
<td>(L) Intermetatarsal</td>
<td>Left angle of the metatarsus primus adductus</td>
</tr>
<tr>
<td>(R) Intermetatarsal</td>
<td>Right angle of the metatarsus primus adductus</td>
</tr>
<tr>
<td>(L) Talardeclination</td>
<td>Angle of the left talus declination</td>
</tr>
<tr>
<td>(R) Talardeclination</td>
<td>Angle of the right talus declination</td>
</tr>
<tr>
<td>(L) RCSP</td>
<td>Left resting calcaneal stance position angle</td>
</tr>
<tr>
<td>(R) RCSP</td>
<td>Right resting calcaneal stance position angle</td>
</tr>
</tbody>
</table>

III. RESULTS

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. Neural networks were also applied for analysis of predictor importance and classification prediction based on the prediction model.

As a result of analysis on five complex disorder groups by using decision tree, 13 rules were discovered as shown in Fig. 3 and Fig. 4. Fig. 5 shows the 13 diagnostic rules that are explored intuitively and can be explained as follows. This can be used as useful and incidental diagnostic information in the professional decision-making process. (1) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was below -8° and ‘(R) Intermetatarsal’ was below ‘(R) Cuboidabduction’ was below -2°, ‘(R) RCSP’ was below -8° and ‘(R) Cuboidabduction’ was below -2° and ‘(R) RCSP’ was below -8°, 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°, ‘(R) Cuboidabduction’ was below -2° and ‘(R) RCSP’ was below -8°, then ‘A’, (3) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (4) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (5) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (6) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (7) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (8) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (9) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (10) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (11) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (12) 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° and ‘(R) RCSP’ was below -8°, then ‘A’, (13) 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° and ‘(R) RCSP’ was below -8°, then ‘A’.
Fig. 3. The result of C5.0 decision tree
Fig. 4. If-then rule types of decision tree

If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was below 4 cm, ‘(L) RCSP’ was above -6° and ‘(R) Talocalcaneal’ was above 28°, then ‘C’, (12) If ‘(L) Tibia TMA’ was below -6° and ‘KneeIMD’ was above 3 cm, then ‘D’, (13) If ‘(L) Tibia TMA’ was above -6°, ‘(R) Intermetatarsal’ was above 4°, ‘KneeIMD’ was above 4 cm and ‘(L) Cuboidabduction’ was above 7°, then ‘E’.

Fig. 6 shows the result of a neural network, and the predictor importance rank was appeared in the following order as shown in Fig. 7: (L) Tibia TMA (0.23), (R) Tibia TMA (0.22), KneeIMD (0.20), (L) Intermetatarsal (0.08), … (L) Cuboidabduction (0.01). In addition, classification prediction rate per class was confirmed as follows: classification for ‘A’ was 7.1 %, classification for ‘B’ was 94.9 %, classification for ‘C’ was 90.2 %, classification for ‘D’ was 80 %, and classification for ‘E’ was 80%, as shown in Fig. 8.
IV. CONCLUSION

In this paper, we have studied an intelligent system that diagnoses and prescribes patients with pediatric foot disorder using decision tree and neural network. The object in this study was to discover meaningful knowledge between the foot disorder and biomechanical parameters related to symptoms using C5.0 decision tree and neural network. 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 relatively high. Therefore, detailed preprocessing and analysis is necessary continually to improve accuracy of classification.

ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2016R1A2B4015623) and (Grands No. C0505552) by Business for Academic-Industrial Cooperative establishments funded Korea Small and Medium Business Administration on in 2017.

REFERENCES


Fig. 8. Classification prediction rate

**Jungkyu Choi** He received the B. S., M. S., and Ph. D. degree from the Chonbuk National University, Jeonju, Republic of Korea, in 2011, 2013, and 2017, respectively. He is currently working on a spatial data analysis project with big data processing methods in the Samah Co., LTD. His major research interest is the image processing, data processing, data mining and deep learning.

**Hee-Sang Lee** He received the B. S. degree from the Chonbuk National University, Jeonju, Republic of Korea, in 2011. His major research interest is the postural balance, gait analysis, data mining, and rehabilitation.

**Jung-Ja Kim** She received the B. S., M. S., and Ph. D. degree from the Chonbuk National University, Gwangju, Republic of Korea, in 1985, 1988, 2002, respectively. She worked with electronic telecommunication Laboratory at Chonnam National University from 2002 to 2004, and Korea Bio-IT Foundry Center at Gwangju from 2004 to 2006. She is currently a professor at Chonbuk National University. Her major research interest is the big data analysis (bio and medical data analysis), artificial intelligence, and rehabilitation.