Analysis of Pediatric Foot Disorders using Decision Tree and Neural Networks

J. K. Choi, Y. G. Won and J. J. Kim

***Abstract*—-** Data mining is method to extract hidden predictive information, and it 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 using C5.0 decision tree and neural networks. 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 and neural networks were applied for analysis of predictor importance and classification prediction. Consequently, a major symptom information in 13 diagnosis patterns was confirmed.

***Keywords*—** Pediatric Foot, Foot Disorder, Patter Classification, Decision tree, Neural networks

I. 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 a different from that of adult because pediatric foot has different characteristics in structure and function [4]. For this reason, Muller, Carlsohn, 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 grows up by age 5 and 7 quickly. Since then, they keep 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

J. K. Choi is with the Department of Healthcare Engineering, Chonbuk National University, 664-14, Deokjin-dong 1, Deokjin-gu, Jeonju, Jeolabuk-do, Republic of Korea (e-mail: chaosjoozak@nate.com).

Y. G. Won is with the School of Electronics and Computer Engineering, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, Republic of Korea (e-mail: ykwon@chonnom.ac.kr).

J. J. Kim is with the Division of Biomedical Engineering and Research Center of Healthcare & Welfare Instrument for the Aged, Chonbuk National University, 664-14, Deokjin-dong 1, Deokjin-gu, Jeonju, Jeolabuk-do, republic of Korea (corresponding author to provide phone: 82-63-270-4102; fax: 82-63-270-2247; e-mail: jungjakim@jbnu.ac.kr).

observing relation [17]. 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 [18]-[19]. Neural networks, which are another representative technique in the data mining, are composed of simple elements similar to biological nervous system [20]. 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, and therefore they have been applied to the various field [21].

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, Stoean, Lupsor, 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 [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 the data mining.

II. METHOD

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. 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 record 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.

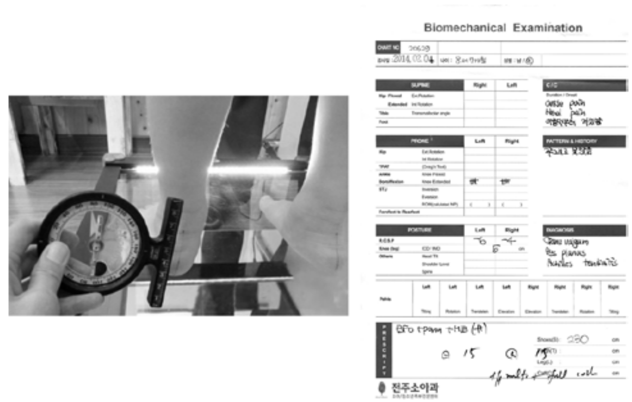

Fig 1. Measurement of RCSP and patient charts

B. 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 I Dependent variable

Class	Disorder
A	Pes planus, Achilles tendinitis
B	Pes planus
C	Pes planus, Intoe gait
D	Pes planus, Intoe gait, Genu valgum
E	Pes planus, Genu valgum

C. Study Procedure

In the study, combination of independent variables meant each of the foot disorder group. Therefore, 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 fig. 2.

Table II Independent variable

Variable	Type	Description
Sex	Nominal	Male, Female
(L) TibiaTMA	Numeric	Angle of the left tibia transmalleolar
(R) TibiaTMA	Numeric	Angle of the right tibia transmalleolar
KneeIMD	Numeric	The knee internal malleolus distance
(L) Talocalcaneal	Numeric	Left angle between the talus and the calcaneus
(R) Talocalcaneal	Numeric	Right angle between the talus and the calcaneus
(L) Cuboidabduction	Numeric	Left angle of the cuboid abduction
(R) Cuboidabduction	Numeric	Right angle of the cuboid abduction
(L) Intermetatarsal	Numeric	Left angle of the metatarsus primus adductus
(R) Intermetatarsal	Numeric	Right angle of the metatarsus primus adductus
(L) Talardeclination	Numeric	Angle of the left talus declination
(R) Talardeclination	Numeric	Angle of the right talus declination
(L) RCSP	Numeric	Left Resting calcaneal stance position angle
(R) RCSP	Numeric	Right Resting calcaneal stance position angle

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.

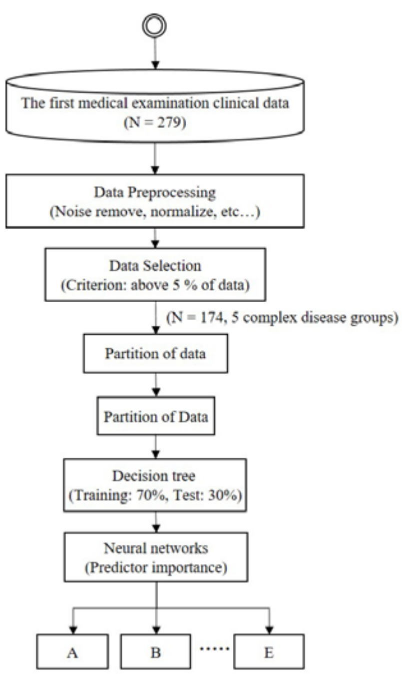

Fig 1. Mining procedure

III. RESULT

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 : (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 below 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°, '(R) Cuboidabduction' was below -2° and '(R) RCSP' was above -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 -6 ° ~ above -8°, then 'A', (4) 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 below 28°, then 'A', (5) If '(L) Tibia TMA' was above -6°, '(R) Intermetatarsal' was above 4°, 'KneeIMD' was above 4 cm, '(L) Cuboidabduction was below 7° and '(R) Intermetatarsal' was above 7°, 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°, '(R) RCSP' was below -8 and '(R) Intermetatarsal' was above 9°, then 'B', (7) 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', (8) 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 'B', (9) If '(L) Tibia TMA' was above -6°, '(R) Intermetatarsal' was above 4°, 'KneeIMD' was above 4 cm, '(L) Cuboidabduction was below 7° and '(R) Intermetatarsal' was below 7°, then 'B', (10) '(L) Tibia TMA' was below -6° and 'KneeIMD' was below 3 cm, then 'C', (11) '(L) Tibia TMA' was above -6° and '(R) Intermetatarsal' was below 4°, then 'C', (12) '(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', as shown in Fig. 3 and Fig. 4.

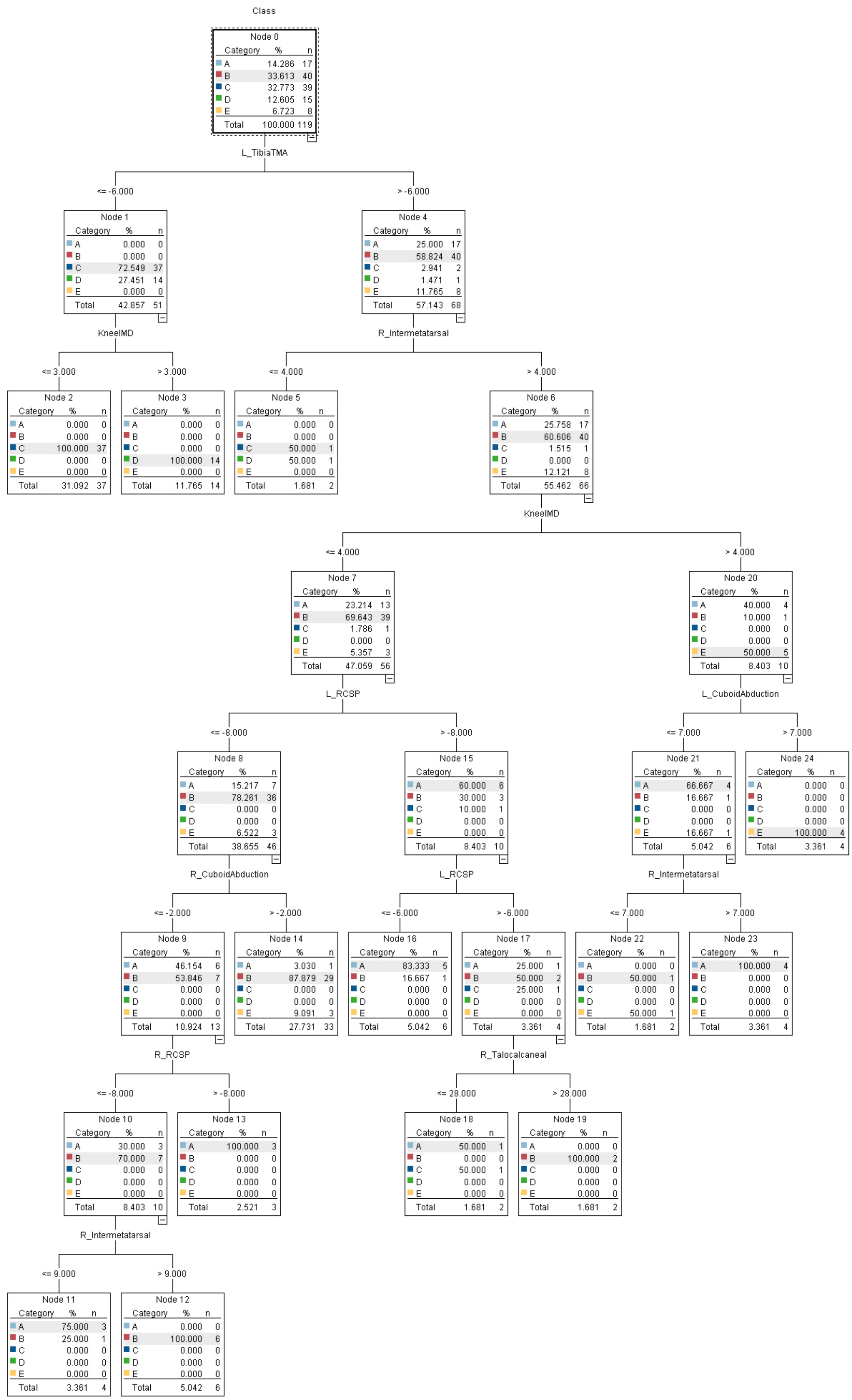

Fig 3. The result of decision tree

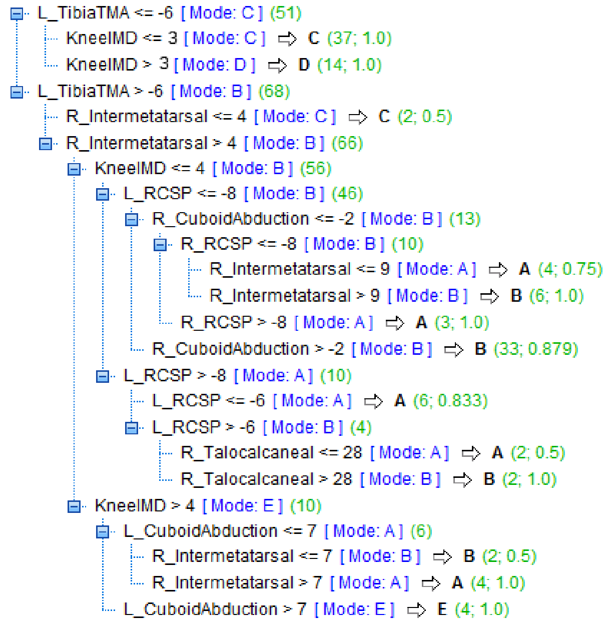

Fig 4. The result of C5.0 decision tree

As a result of neural networks, predictor importance was appeared in the following order : (L) TibiaTMA (0.23), (R) TibiaTMA (0.22), KneeIMD (0.20), (L) Intermetatarsal (0.08), … (L) Cuboidabduction (0.01), as shown in Fig. 5. 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. 6.

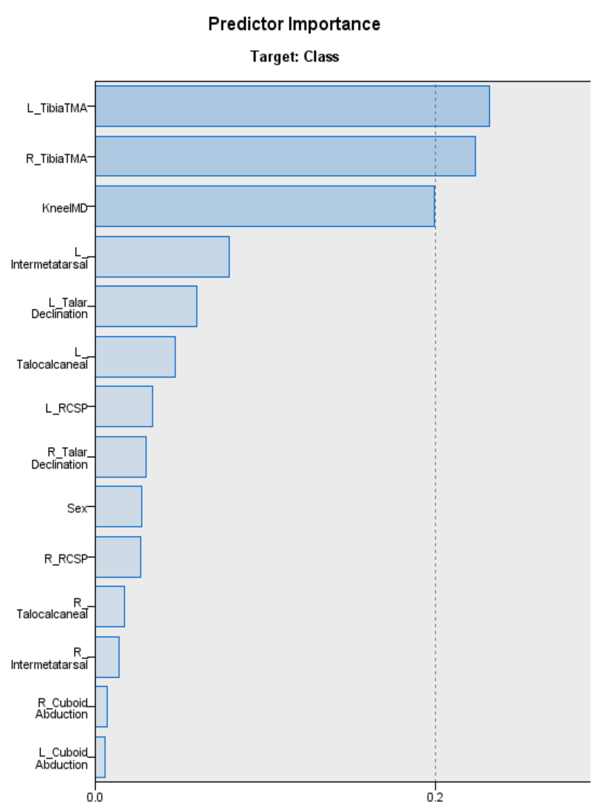

Fig 5. Predictor Importance

Classification for Class

Overall Percent Correct = 79.8%

Observed	Predicted A	Predicted B	Predicted C	Predicted D	Predicted E
A	7.1%	71.4%	0.0%	0.0%	21.4%
B	2.6%	94.9%	0.0%	0.0%	2.6%
C	0.0%	7.3%	90.2%	2.4%	0.0%
D	0.0%	0.0%	6.7%	80.0%	13.3%
E	10.0%	10.0%	0.0%	0.0%	80.0%

Fig 6. Classification prediction rate

IV. 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 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 verified and neural networks were applied for analysis of predictor importance and classification prediction.

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 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°. However, we were able to know that the proportion predicted incorrectly the class A to the class B was high through analysis of neural networks. It meant that classification of achilles tendinitis was difficult from pes planus, in comparison with other disorders.

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 and growth. For these reasons, the error rate of 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 the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2016R1A2B4015 623), Republic of Korea.

REFERENCES

[1] R.A. Donatelli, The Biomechanics of the foot and ankle, The F. A. Davis Company, Philadelphia, 1996.
[2] T. Marasovic, M. Cecic, and V. Zanchi, "Ground Reaction Forces in Gait: Statistical Analysis and Interpretation," 9th WSEAS Int. Conf. SIMULATION, MODELLING and OPTIMIZATION (SMO '09), Budapest, 2009.
[3] D. J. Lott, M. K. Hastings, P. K. Commean, K. E. Smith and M. J. Muller, "Effect of footwear and orthotic devices on stress reduction and soft tissue strain of the neuropathic foot," Clinical Biomechanics, vol. 22, No. 3, 2007, pp. 352–359.
[4] J. H. Chang, S. H. Wang, C. L. Kuo, H. C. Shen, Y. W. Hong and L. C. Lin, "Prevalence of flexible flatfoot in Taiwanese school-aged children in relation to obesity, gender, and age, European journal of paediatrics," vol. 169, No. 4, 2010, pp. 447-452.
[5] S. Muller, A. Carlsohn, J. Muller, H. Baur and F. Mayer, "Static and dynamic foot characteristics in children aged 1–13 years: A cross-sectional study," Gait & posture, vol. 35, No. 3, 2012, pp. 389-394.
[6] M. Anderson, M. M. Blais and W. T. Green, "Lengths of the growing foot," J Bone Joint Surg Am., vol. 38, 1956, pp. 998-1000.
[7] M. Pfeiffer, R. Kotz, T. Ledl, G. Hauser and M. Sluga, "Prevalence of flat foot in preschool-aged children," Pediatrics, vol. 118, No. 2, 2006, pp. 634-639.
[8] R. M. Jay, H. D. Schoenhaus, C. Seymour and S. Gamble, "The Dynamic Stabilizing Innersole System (DSIS): the management of hyperpronation in children," The Journal of foot and ankle surgery, vol. 34, No. 2, 1995, pp. 124-131.
[9] T. L. Lincoln and P. W. Suen, "Common rotational variations in children," Journal of the American Academy of Orthopedic Surgeons, vol. 11, No. 5, 2003, pp. 312–320.
[10] Y. J. Ko and H. W. Kim, "Diagnosis and Conservative Treatment of Common Foot Disorders," Journal of the Korean Medical Association, vol. 47, No. 3, 2004, pp. 247-257.
[11] J. Han, and M. Kamber, Data mining: concepts and techniques, Morgan Kaufmann, San Francisco, 2001.
[12] N. Duru, "An application of apriori algorithm on a diabetic database, Knowledge-Based Intelligent Information and Engineering Systems," Springer Berlin Heidelberg, 2005, pp. 398-404.
[13] T. B. Murdoch, A. S. Detsky, "The inevitable application of big data to health care," Jama, Vol. 309, No. 13, 2013, pp. 1351-1352.
[14] Mckinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, 2011.
[15] A. R. Islam and T. S. Chung, "An Improved Frequent Pattern Tree Based Association Rule Mining Technique," Information Science and Applications (ICISA), 2011 International Conference on. IEEE, 2011, pp. 1-8.
[16] P. PERNER, "Big Data Analysis and Reporting with Decision Tree Induction," 9th WSEAS Int. Conf. COMPUTER ENGINEERING AND APPLICATIONS (CEA '08), Dubai, 2015.
[17] M. Y. Chang, C. C. Shih, D. A. Chiang and C.C. Chen, "Mining a Small Medical Data Set by Integrating the Decision Tree and t-test," Journal of Software, vol. 6, No. 12, 2011, pp. 2515-2520.
[18] J. Y. Song and H. K. Kim, "A Study of Decision Tree in Detecting Intrusions," Journal of the Korean Data Analysis Society, vol. 12, No. 2, 2010, pp. 983-996.
[19] H. Sug, "Improving the performance of minor class in decision tree using duplicating instances," 10th WSEAS Int. Conf. Artificial Intelligence, Knowledge Engineering and Databases (AIKED `11), Cambridge, 2011.
[20] H. Demuth and M. Beale, "Neural Network Toolbox for use with MATLAB," 1993.
[21] D. Enke and S. Thawornwong, "The use of data mining and neural networks for forecasting stock market returns," Expert Systems with applications, vol. 29, No. 4, 2005, pp. 927-940.
[22] J. L. Breault, C. R. Goodall and P. J. Fos, "Data mining a diabetic data warehouse, Artificial Intelligence in Medicine," vol. 26, No. 1, 2002, pp. 37-54.
[23] Y. M. Kim, "A study on analysis of factors on in-hospital mortality for community-acquired pneumonia," Journal of the Korean Data & Information Science Society, vol. 22, No. 3, 2011, pp. 389-400.
[24] R. Stoean, C. Stoean, M. Lupsor, H. Stefanescu, R. Badea, "Evolutionary-driven support vector machines for determining the degree of liver fibrosis in chronic hepatitis C," Artificial intelligence in medicine, vol. 51, No. 1, 2011, pp. 53-65.
[25] K.H. Lim, K.S. Ryu, S.H. Park, H.S. Shon, K.H. Ryu, "Short-term Mortality Prediction of Recurrence Patients with ST-segment Elevation Myocardial Infarction," Journal of The Korea Society of Computer and Information, vol. 17, No. 10, 2012, pp. 145-154.
[26] M. Park, S. Choi, A.M. Shin, H. Chul, "Analysis of the Characteristics of the Older Adults with Depression Using Data Mining Decision Tree Analysis," J Korean Acad Nurs, Vol. 43, 2013, pp. 1-10.
[27] M. H. Huh and Y. G. Lee, Data mining modeling and case 2nded. Hannarae, Seoul, 2008.

Jung-Kyu Choi He received the B. S., M. S. degree from the Chonbuk National University, Republic of Korea, in 2011 and 2013, respectively. He is currently pursuing a doctorate in foot biomechanics at the university. His major research interest is the podiatry, medical informatics and biomechanics.

Yonggwon Won He received the B. S. in Electronics Engineering from Hanyang Univesity in 1987, and M. S. and Ph. D. degrees in Electrical and Computer Engineering from University of Missouri-Columbia in 1991 and 1995, respectively. He worked with Electronics, and Telecommunication Research Institute (ETRI) from 1995 to 1996, and Korea Telecomm (KT) from 1996 to 1999. He is currently a professor in Chonnam National University in Korea, and the director of Korea Bio-IT Foundary Center at Gwangju. His major research interest is the computational intelligence for image analysis, pattern recognition, network and communication security, bio and medical data analysis.

Jung-Ja Kim She received the B.S., M.S. degree in 1985, 1988 and Ph.D. degree from 1997 to 2002, in Computer Science from Chonnam National University 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 an associate professor at Chonbuk National University. Her major research interest is the bio and medical data analysis, database security, and pattern recognition.