

# Chi-Square Automatic Interaction Detection Modeling for Predicting Depression in Multicultural Female Students

Haewon Byeon

Department of Speech Language Pathology  
Honam University  
Gwangju, Republic of Korea

**Abstract**—This study developed a depression prediction model for female students from multicultural families by using a decision tree model based on Chi-squared automatic interaction detection (CHAID) algorithm. Subjects of the study were 9,024 female students between 12 and 15 years old among the children of surveyed marriage immigrants. Outcome variables were classified as presence of depression. Explanatory variables included sex, residing area, experience of career counseling, experience of social discrimination, experience of Korean language education, experience of using a multicultural family support center, Korean reading, Korean speaking, Korean writing, Korean listening, Korean society adjustment education experience, needs of Korean society adjustment education, needs of Korean language education, and rejoined entry. In the CHAID algorithm analysis, female students from multicultural families who experienced social discrimination within the past one year and had ordinary Korean speaking skill posed the highest risk of depression. It is necessary to pay social level interests to the mental health of adolescents from multicultural families for achieving successful social integration based on the results of this study.

**Keywords**—CHAID; data mining; multicultural family; risk factors; depression

## I. INTRODUCTION

The number of children from multicultural families is rapidly increasing in South Korea due to the increase of international marriage. The number of children from multicultural families was 100,000 in 2010, doubled in 2014, and is expected to exceed 300,000 in 2020 [1], [2]. Particularly, the low birth-rate has become lower than the population replacement rate, 2.1 children per female, so the proportion of students from multicultural families will increase steadily [1].

Nevertheless, the policies for multicultural families in South Korea mainly focus on employment or welfare and there are not enough studies about the health of multicultural families [3], [4]. Moreover, the previous studies on adolescents from multicultural families mainly aimed at their academic performance and school adjustment and only a few studies evaluated the emotional characteristics of them, including depression [5].

Adolescents from multicultural families have a high probability to experience negative emotions due to social

prejudice and discrimination [5]. A previous epidemiological survey showed that adolescents growing in low socioeconomic status (e.g., household income) had higher possibility to experience depression [6]. Considering that the proportion of multicultural families is higher in the rural area than in the urban area and only 9.7% of household had an income higher than 30 million KRW in the rural area, it is anticipated that adolescents from multicultural families are more likely to be exposed to negative emotions [7]. Byeon [8] reported that approximately 15% of adolescents (age between 19 and 23) from multicultural families experienced social discrimination. Moreover, Tienda, and Haskins [9] also stated that children from internal marriage families, who were born in South Korea as well as immigrated after birth, had a hard time to adapt to the society due to the social characteristics (e.g., multicultural family) and rapid changes during the adolescent period. As shown, adolescents from multicultural families are very likely to have factors associated with emotional aspects so they are more vulnerable to depression than the other adolescents. However, the previous studies on the emotional aspect of adolescents from multicultural families mainly used methods comparing the characteristics of adolescents from multicultural families and those from ordinary families in order to identify the individual risk factors [10]. Additionally, the previous studies on multicultural families focused on welfare so statistics on health and healthcare status are lacking.

Emotional problems are induced by multi-dimensional factors such as environment, social support, and stress. Therefore, it is necessary to conduct a multiple risk factor analysis in order to accurately identify the characteristics of depression. Recently, data mining techniques (e.g., artificial neural network and decision tree) are frequently used to establish a prediction model for multiple risk factors [11], [12].

This study developed a depression prediction model for female students from multicultural families by using a decision tree model based on Chi-squared automatic interaction detection (CHAID) algorithm. The study is composed as follows. Section 2 explains the study subjects and CHAID algorithm. Section 3 discusses the results and the power of the developed CHAID based prediction model. Section 4 presents conclusions and future study directions.

## II. METHODS

### A. Study Participants

This study used the raw data of 2012 Nationwide Multicultural Family Status Survey, which was conducted for multicultural families living in South Korea by Ministry of Health, Welfare, and Family Affairs, Ministry of Justice, and Ministry of Gender Equality. Multicultural Family Status Survey was carried out to understand the living conditions and the welfare needs of multicultural families in order to develop customized policies for multicultural families [2]. The items of this nationwide survey were composed of general characteristics, economic level, employment, health and health care, and marriage. Multicultural Family Status Survey was conducted between July 20 and Oct 31, 2012. The subject of the survey was 154,333 people, all married immigrants at the time of survey according to the alien resident status and the basic multicultural family status data of Ministry of Public Administration and Security. The selection criteria of multicultural families are based on Multicultural Family Law, as follows. First, it targeted families composed of an immigrant(s) and a Korean citizen(s). Second, it was defined as a family composed of a foreigner(s) who acquired Korean citizenship through report or naturalization and Korean(s) who acquired nationality by birth, report, and naturalization. This study targeted 9,024 female students between 12 and 15 years old among the children of surveyed marriage immigrants.

### B. Measurements

Outcome variables were classified as presence of depression (yes or no). Explanatory variables included residing area (rural or urban), experience of career counseling (yes or no), experience of social discrimination (yes or no), experience of Korean language education (yes or no), experience of using a multicultural family support center (yes or no), Korean reading (good, intermediate, and poor), Korean speaking (good, intermediate, and poor), Korean writing (good, intermediate, and poor), Korean listening (good, intermediate, and poor), Korean society adjustment education experience (yes or no), needs of Korean society adjustment education (necessary, average, and not necessary), needs of Korean language education (necessary, average, and not necessary), and rejoined entry (come to Korea after living in a foreign country or born and grown up in Korea).

## III. STATISTICAL ANALYSIS

### A. Exploring Predictors

General characteristics were presented with mean and percentage by using descriptive statistics. The difference between groups due to the depression was analyzed by Chi-square test.

### B. Chi-Squared Automatic Interaction Detection

CHAID is an algorithm that performs the multiway split\* by using Chi-square or F-test [13]. CHAID algorithm uses Pearson's Chi-square when a target variable is categorical and uses likelihood ratio Chi-square statistic as a separation

reference when a target variable is continuous [14]. Chi-square is calculated from the  $r \times c$  partition table composed of observations ( $f_{ij}$ ). The function of Pearson's Chi-square statistic is shown as (1).

$$x^2 = \sum_{l,j} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad (1)$$

The function of likelihood ratio Chi-square statistic is shown as (2).

$$x^2 = 2 \sum_{l,j} f_{i,j} \times \log_e \left( \frac{f_{ij}}{e_{ij}} \right) \quad (2)$$

The Chi-square statistic, very smaller than the degree of freedom, implies that the distributions of the target variables for each category of the predictor variable are the same. Therefore, it can be concluded that the predictor variable does not affect the classification of target variables. The magnitude of the chi-square statistic for the degree of freedom can be expressed as a p-value. When the chi-square statistic is smaller than the degree of freedom, the value of p increases. As a result, using Chi-square statistic as a separation reference means that the child node is formed by the predictor variable with the smallest p value and the optimum separation.

This study treated all explanatory variables including outcome variables as categorical variables in order to minimize the convenience of CHAID algorithm as much as possible [15]. In the model of this study, the separation and merge criterion of the decision rule for CHAID algorithm was set as 0.05 and the numbers of parent nodes, child nodes, and branch were limited to 250, 150, and 4, respectively [16]. The validity of the model was assessed by using a 10-fold cross-validity test and the degrees of model's risk were compared [17].

## IV. RESULTS

### A. General Characteristics of Participants

Among the total of 9,024 female students subjects, 2,627 subjects (29.1%) experienced depression during the past year (Table I). The results of chi-square test showed that there were significant ( $p < 0.05$ ) differences in rejoined entry, Korean speaking level, Korean listening level, Korean reading level, Korean writing level, career and consulting education experience, and social discrimination between subjects with depression experience and those without depression experience. Depression experience rate was high for female students who entered South Korea through rejoined entry (32.6%), could speak Korean at an intermediate level (42.6%), could listen Korean at a poor level (38.9%), could read Korea at an intermediate level (37.6%), could write Korean at poor level (40.4%), had experienced in Korean education (42.9%), experienced career counseling, and experienced social discrimination (53.0%).

TABLE I. CHARACTERISTICS OF SUBJECTS BASED ON EXPERIENCE OF DEPRESSION, N (%)

Variables	Depression		p
	No (n=6,397)	Yes (n=2,627)	
Residing area			0.436
Urban	4,294 (71.1)	1,742 (28.9)	
Rural	2,102 (70.3)	886 (29.7)	
Rejoined entry			0.003
Come to Korea after living in a foreign country	882 (67.4)	427 (32.6)	
Born and grown up in Korea	5,514 (71.5)	2,200 (28.5)	
Korean speaking			<0.001
Good	6,009 (71.7)	2,375 (28.3)	
Intermediate	265 (57.4)	197 (42.6)	
Poor	123 (69.1)	55 (30.9)	
Korean listening			0.002
Good	6,062 (71.3)	2,440 (28.7)	
Intermediate	279 (64.7)	152 (35.3)	
Poor	55 (61.1)	35 (38.9)	
Korean reading			<0.001
Good	5,999 (71.4)	2,401 (28.6)	
Intermediate	335 (62.4)	202 (37.6)	
Poor	63 (72.4)	24 (27.6)	
Korean writing			<0.001
Good	5,898 (71.6)	2,335 (28.4)	
Intermediate	405 (64.0)	228 (36.0)	
Poor	93 (59.6)	63 (40.4)	
Experience of using a multicultural family support center			0.135
Do not even know that such education exists	2,553 (72.0)	992 (28.0)	
Know the education but never used it before	2,713 (70.4)	1,140 (29.6)	
Not only know the education but also have used it before	1,131 (69.6)	495 (30.4)	
Experience of Korean language education			<0.001
No	6,106 (71.7)	2,408 (28.3)	
Yes	291 (57.1)	219 (42.9)	
Needs of Korean language education			
Necessary	754 (11.8)	420 (16.0)	
Average	965 (15.1)	430 (16.4)	
Not necessary	4,678 (73.1)	1,777 (67.6)	
Needs of Korean society adjustment education			0.053
Necessary	363 (67.2)	177 (32.8)	
Not necessary	6,033 (71.1)	2,450 (28.9)	
Experience of career counseling			<0.001
No	5,737 (73.0)	2,119 (27.0)	
Yes	660 (56.5)	508 (43.5)	
Experience of social discrimination			<0.001
No	5,885 (74.2)	2,049 (25.8)	
Yes	512 (47.0)	578 (53.0)	

TABLE II. GAINS CHART OF PREDICTOR VARIABLE BY CHAID ALGORITHM

Node no	Node n (%) <sup>1</sup>	Gain n (%) <sup>2</sup>	Response % <sup>3</sup>	Gain Index % <sup>4</sup>	Characteristics
7	91 (1.0)	91 (3.5)	100	343.9	Experience of social discrimination=yes; Korean speaking=Intermediate
14	116 (1.3)	83 (3.2)	71.6	246.0	Experience of social discrimination=no; Experience of career counseling=yes; Needs of Korean society adjustment education=yes
9	313 (3.5)	193 (7.3)	61.7	212.0	Experience of social discrimination=no; Experience of career counseling=no; Needs of Korean society adjustment education=yes (average)
15	567 (6.3)	324 (12.3)	57.1	196.5	Experience of social discrimination=yes; Korean speaking=Good or Poor; Needs of Korean language education=Not necessary
12	56 (0.6)	28 (1.1)	50.0	171.9	Experience of social discrimination=no; Korean speaking=Poor; Needs of Korean society adjustment education=no
16	172 (1.9)	81 (3.1)	47.1	161.9	Experience of social discrimination=yes; Korean speaking=good or poor; Needs of Korean language education=Necessary
13	881 (9.8)	308 (11.7)	35.0	120.2	Experience of social discrimination=no; Experience of career counseling=no; Needs of Korean society adjustment education=yes
17	258 (2.9)	81 (3.1)	31.4	108.0	Experience of social discrimination=yes; Korean speaking=good or poor; Needs of Korean language education=Average
8	1,378 (15.3)	343 (13.1)	24.9	85.6	Experience of social discrimination=no; Experience of career counseling=yes; Needs of Korean society adjustment education=yes (average)

10	5,082 (56.3)	1,088 (41.4)	21.4	73.6	Experience of social discrimination=no; Needs of Korean society adjustment education=no Korean speaking=good
11	116 (1.3)	6 (0.2)	5.2	17.8	Experience of social discrimination=no; Needs of Korean society adjustment education=no Korean speaking=intermediate

<sup>1</sup> Node n(%); node number, % to 9,024

<sup>2</sup> Gain n(%); gain number, % to 2,627

<sup>3</sup> Response (%): The fraction of the depression in subjects

<sup>4</sup> Gain index (%):=343.9 in total 10 node

**B. Results of Prediction Model of Female Students from Multicultural Families based on CART Algorithm**

The depression prediction model of female students from multicultural families based on CART algorithm is shown in Fig. 1. The established depression prediction model revealed that the experience of social discrimination, Korean society adjustment education requirement, Korean speaking level, career consulting experience, and Korean education needs were important predictor variables in the order of magnitude.

Table II presents the profit chart of the depression prediction model of female students from multicultural

families. Among 11 paths, 4 paths were confirmed to predict depression effectively. The first path was “female students from multicultural families who experienced social discrimination within the past one year and had ordinary Korean speaking skill”. The profit index was 343.0%. The second path was “female students from multicultural families who did not experience of social discrimination with the past one year, had career consulting experience, and needed to Korean society adjustment education”. The profit index of this path was 246.0%. The third path was “female students from multicultural families who did not experience of social discrimination with the past one year and have career consulting experience and expressed average needs for Korean society adjustment education”. The profit index of this path was 212.0%. The fourth path was “female students from multicultural families who experience of social discrimination with the past one year and considered Korean language education was unnecessary even though their Korean speaking level was good or poor”. The profit index of this path was 196.5%.

The results of the 10-fold cross validity test showed that the predictive accuracy of the model was 74.2%, the risk index of the cross classification model was 0.258, misclassification orate was 25.8%, and standard error was 0.005, which agreed with risk index of 0.259, misclassification rate of 25.9%, and standard error of 0.005 of the predictive model Fig. 2.

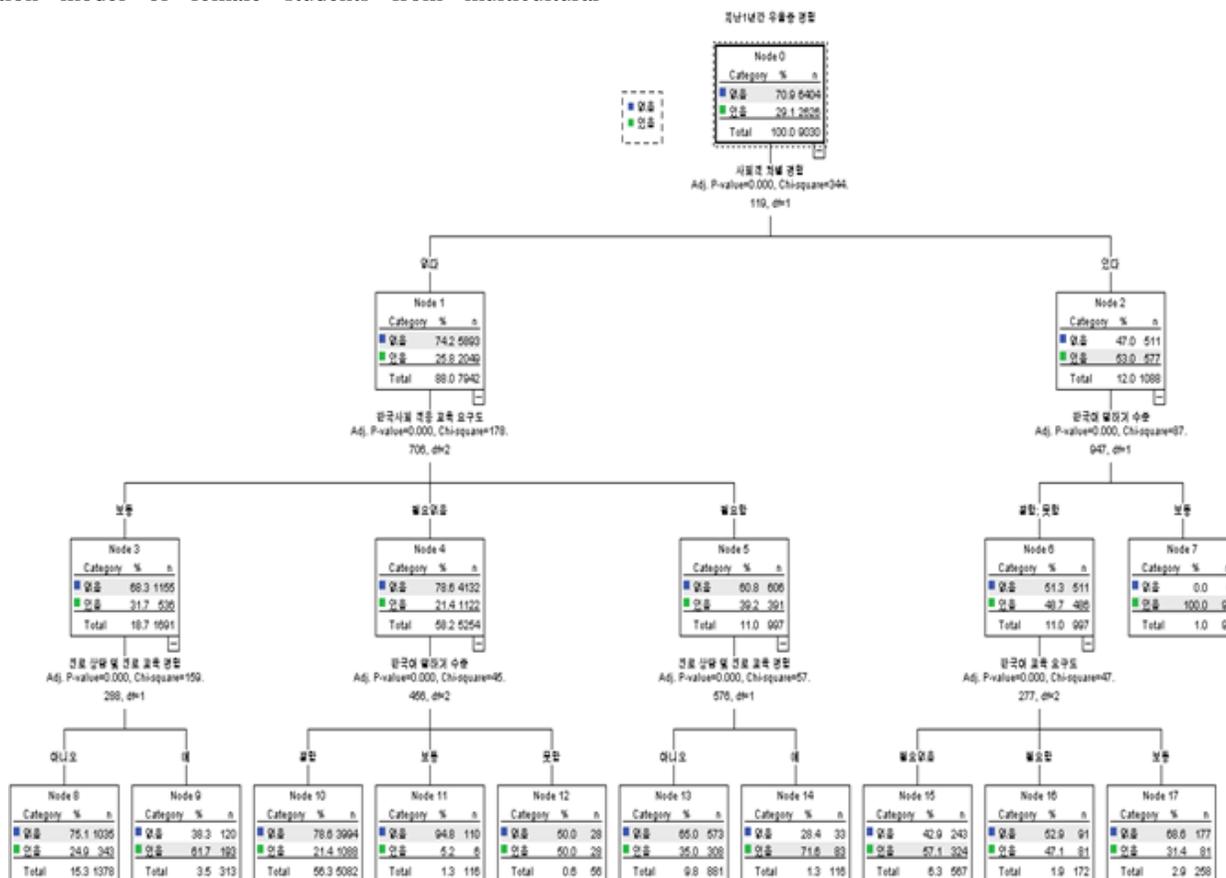


Fig. 1. Prediction model for experience of depression symptoms in children in multi-cultural families.

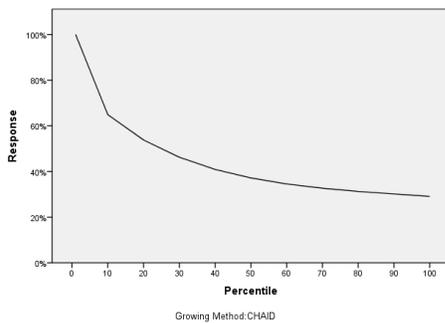


Fig. 2. The results of the validity test.

## V. CONCLUSION

It will be necessary to identify the difficulties experienced by children from multicultural families and provide a systematic program aiding their social adjustment for successful social integration.

This study developed a depression prediction model for children from multicultural families by using CHAID algorithm and found that the experience of social discrimination is the most critical factor affecting depression. Although it is hard to compare the results of this study directly, the previous studies evaluating the relationship between social discrimination and mental health reported that the economic discrimination and the discrimination against a specific group (e.g., the elderly group) were significant predictor variables negatively influencing mental health [18]. Therefore, it is necessary to establish a legal system and pay social level interests to overcome the discrimination and prejudice against adolescents from multicultural families based on the results of this study.

Another finding of this study was that “female students from multicultural families who experienced social discrimination within the past one year and had ordinary Korean speaking skill” posed the highest risk of depression. It has been repeatedly reported that children from multicultural families had lower language level than those from ordinary families in the language development [19]. It was also reported that the difference in the language development between those from multicultural families and those from ordinary families disappeared as they became older [20]. However, it is known that children from multicultural families still experienced difficulties in learning Korean even after adolescence. Since the immaturity in Korean during the adolescence period has a decisive influence not only on academic achievement but also on social adjustment [21], continuous Korean language education for adolescents from multicultural families is necessary to form successful social integration and prevent mental illness.

It is necessary to pay social level interests to the mental health of adolescents from multicultural families for achieving successful social integration based on the results of this study.

## ACKNOWLEDGMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A02068326).

## REFERENCES

- [1] Y. Kim, H. Choi, K. Kim, and S. Sung, The changes of multicultural families and social welfare policies, The Korea Institute for Health and Social Affairs, Seoul, 2012.
- [2] Ministry of Gender Equality & Family, A Study on the National Survey of Multicultural Families, Ministry of Gender Equality & Family, Seoul, 2012.
- [3] J. H. Shin, A study on comparative home environmental factor effect of delinquency in multi-cultural youth. *Journal of Korean Public Police and Security Studies*, vol. 11, no. 2, pp. 1-20, 2014.
- [4] H. M. Kim, and C. H. Cho, Multicultural Policy for Social Integration: Case of Germany and Korea. *Advanced Science and Technology Letters*, vol. 46, pp. 179-186, 2014.
- [5] W. Shim, Racial Identity and Psychological Attitudes of Multicultural Students. *The Journal of Elementary Education*, vol. 22, no. 4, pp. 27-47, 2009.
- [6] Ministry of Health & Welfare, Korean youth obesity survey, Ministry of Health & Welfare, Seoul, 2009.
- [7] D. S. Park, S. J. Ma, and I. H. Kwon, Social adaptation of rural multicultural families and policy tasks, Korea Rural Economic Institute, Seoul, 2011.
- [8] H. Byeon, Development of prediction model for endocrine disorders in the Korean elderly using CART algorithm. *International Journal of Advanced Computer Science and Applications*, vol.6,no.9,pp.125-129,2015.
- [9] M. Tienda, and R. Haskins, Immigrant children: introducing the issue. *The Future of Children*, vol. 21, no. 1, pp. 3-18, 2011.
- [10] D. H. Jwa, A Research on the Psycho-social Adaptation Characteristics of Immigrant Youths based on Nationality and Acculturation Types. Korea Youth Research Association, vol. 21, no. 1, pp. 97-126, 2014.
- [11] H. Byeon. Exploring influence factors for peer attachment in korean youth based on multi-layer perceptron artificial neural networks. *Journal of the Korea Convergence Society*, vol. 8, no.10, pp. 209-214, 2017.
- [12] H. Byeon, H. Jin, and S. Cho, Development of Parkinson's disease dementia prediction model based on verbal memory, visuospatial memory, and executive function. *Journal of Medical Imaging and Health Informatics*, vol. 7, no.7, pp. 1517-1521, 2017.
- [13] A. A. Saa, Educational Data Mining & Students' Performance Prediction. *International Journal of Advanced Computer Science & Applications*, vol. 1, pp. 212-220, 2016.
- [14] H. Wimmer, and L. Powell, A Comparison of the Effects of K-Anonymity on Machine Learning Algorithms. *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 11, pp. 155-160, 2014.
- [15] H. Byeon, and S. Cho, The factors of subjective voice disorder using integrated method of decision tree and multi-layer perceptron artificial neural network algorithm. *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, pp. 112-116, 2016.
- [16] H. Byeon. The risk factors of laryngeal pathology in Korean adults using a decision tree model. *Journal of Voice*, vol. 29, no.1, pp. 59-64, 2015.
- [17] H. Byeon. The prediction model for self-reported voice problem using a decision tree model. *Journal of the Korea Academia-Industrial cooperation Society*, vol. 14, no.7, pp. 3368-3373, 2013.
- [18] C. Kang, Y. Cho, The effect of social discrimination on suicide ideation. *Korea Journal of Population Studies*, vol. 35, no.3, pp. 1-27, 2012.
- [19] E. Jung, A study on the language development and environment of children from internationally married couple living in an agricultural village. *Journal of Speech-Language & Hearing Disorders*, vol. 13, no. 3, pp. 33-52, 2004.
- [20] F. Genesee, E. Nicoladis, and J. Paradis, Language differentiation in early bilingual development. *Journal of Child Language*, vol. 22, pp. 611-631, 1995.
- [21] S. Kim, and C. Lee. The effect of linguistic ability and social support on adjustment among women of multicultural families in Korea. *Korean Journal of Family Welfare*, vol. 15, no. 1, pp. 5-20, 2010.