The Predictive Modeling of Old-age High Blood Pressure in Korea using Data-mining

Haewon Byeon1, SungHyoun Cho2*

1 Dept. of Speech Language Pathology & Audiology, Nambu University, Gwangju, South Korea, byeon@nambu.ac.kr
2 Dept. of Physical Therapy, Nambu University, Gwangju, South Korea, Corresponding Author: geriatricpt1@naver.com

Abstract. This study developed old-age high blood pressure prediction model by using epidemiological data and, based on the developed model, provides empirical data which can be used in health care programs and education in local community. Data were from the A Study on the Seoul Welfare Panel Study 2010. Subjects were 2,111 elderly (879 male, 1,232 female) over the age of 60 in South Korea. A prediction model was developed by the use of a QUEST Algorithm. As the result of QUEST algorithm, factors having significant effect were subjective health status, education level, regular exercise, smoking, alcohol consumption, household income. The most preferentially involved predictor was the subjective health status.

Keywords: Hypertension, Decision tree, Predictive Modeling, QUEST

1 Introduction

Korea established 'National Health Promotion Plan 2020' and has been making efforts to reduce prevalence rate of high blood pressure from 28.0% in 2005 to 23.0% by 2020 and accordingly, studies on risk factors for high blood pressure have also been actively conducted. Numerous studies conducted so far have reported that lifestyle such as smoking, drinking and regular exercise [1-3] and environmental and psychological factors such as family history [4], old age [5] and stress [6] are risk factors of high blood pressure. Still, these preceding studies have the limitation that they were conducted on the patients visiting medical institutions or focused on the relationship between high blood pressure and individual factors.

To build a successful prevention system of high blood pressure, development of a tailored prediction model is required which considers complex risk factors affecting high blood pressure. Recently, with the spread of big data across society, machine learning algorithm which performs various patterns of recognition or prediction is being used in various areas.

This study established prediction model for old-age high blood pressure based on QUEST algorithm using reliable epidemiological data.
2 Methods

2.1 Subjects

This study analyzed 2,111 senior citizens (879 males, 1,232 females) over the age of 60 among 7,761 people who completed Seoul Welfare Panel Study.

2.2 Measures

Dependent variable was defined as prevalence of hypertension. Explanatory variables were included as sex, education level, household income, smoking, drinking, subjective health status, regular exercise.

2.3 Statistical analysis

QUEST (Quick, Unbiased, Efficient, Statistical Tree) [7] is a data-mining algorithm to select classification standard first by choosing significant variables to reduce variable selection bias and, based on selected variables, then by executing quadratic discriminant analysis. As methods to select variables based on classification standard, it uses ANOVA F-statistics for continuous variables while it chooses variables with smallest significant probability in chi-square test of contingency table for categorical variables [7]. During the process, it finally selects threshold by executing quadratic discriminant analysis on selected categorical variables.

All analysis used MINITAB version 16 (Minitab Inc., State College, Pennsylvania, USA).

3 Results

As the result of chi-square test, prevalence of hypertension has statistically significant difference in smoking, drinking, subjective health status (p<0.05).

As the result of constructing statistical classification model using QUEST algorithm, factors having significant effect were subjective health status, education level, regular exercise, smoking, alcohol consumption, household income. The most preferentially involved predictor was the subjective health status (Figure 1).
4 Conclusion

In order to enhance the level of management on high blood pressure, it is more effective to focus on vulnerable groups and intensively implement proper management measures on them rather than on the overall group.

References