

Cardiovascular disease prediction models on Linear Discriminant Analysis of depression

Junggi Yang¹, Youngho Lee², Un Gu Kang²

¹Dept. of IT Convergence Engineering, Gachon University, Gyeonggi-do, Korea,

²Dept. of Computer Science, Gachon University, Gyeonggi-do, Korea,
{ jgyang, lyh, ugkang }@gachon.ac.kr

Abstract. In depression patients, prediction of cardiovascular diseases is very important and critical for effective treatments and health care. In the present study, we developed the prediction model utilizing data-mining based linear discriminant analysis in order to make early prediction of cardiovascular diseases in depression patients. The objective of the study was to find the optimal method in the clinical decision support system to provide best solution as well as diagnosis in terms of clinical decision support system. In results, the suggested model provided 69% of accuracy which is helpful to predict cardiovascular diseases for depression patients as well as doctors. And eventually, it is considered to be an effective model for depression patients to prevent cardiovascular diseases and self-manage their health via monitoring risks.

Keywords: Depression, Cardiovascular, Coronary Heart Disease, LDA

1 Introduction

Cardiovascular diseases have been considered as a major cause of death worldwide. It gives a large financial burden and suffering for not only patients but also their surroundings; such burdens are not limited to individuals but same for nations. Due to critical risks after occurrence of cardiovascular diseases, it is closely related to early deaths and influences on mental disorders such as depression. It is important to maintain managing patients when it comes to treatments of cardiovascular diseases (e.g., stroke) since they accompany depression in often times. Further, if cardiovascular diseases of depression patients are not diagnosed early enough and appropriate treatments are not given, patients may have serious outcomes.

If it is possible to predict cardiovascular diseases in depression patients, it will be a very effective measure for not only treatments of such diseases but also management plans for patients [1]. Hence, in the present study, we developed the cardiovascular diseases prediction model for depression patients utilizing the Korean database of depression patients.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP)(NRF-2013M3C8A2A02078403).

2 Related Work

Degradation is one of brain diseases that can cause decline in motivation, issues with interpersonal relationships and even may lead to suicide, if serious. However depression can be effectively treated and patients are often fully recovered from it. Counseling as well as psychiatric treatments are required with careful management. Recently, there are a number of studies have been done in regards to depression and cardiovascular diseases; to take an example, it was investigated how education programs for patients with cardiac insufficiency impacted on quality of life as well as depression [2]. In other study, prevalence rate of depression after stroke and predictive factors were investigated in order to find associations between stroke and depression [3]. In addition, many efforts have been actively made in order to develop prediction models for diseases such as depression and cardiovascular diseases that are critical in terms of managing risks and treatments [4]. The prediction model for risks of cardiovascular diseases in general populations [5], and decision-making support services were established for providing individualized monitoring and guidelines for patients via the PEI model [6]. Lastly, a number of studies have been proceeded to provide accurate information for patients utilizing various medical information [7].

3 Prediction method of cardiovascular diseases in depression patients

In the present study, we developed the prediction model for cardiovascular diseases in depression patients utilizing three years-dataset from the Korean National Health and Nutrition Examinations Survey (KNHANES). Given that various risk factors are responsible for cardiovascular diseases, the liner discriminant analysis was done with factoring environmental variables (e.g., smoking, diabetes mellitus, cholesterol levels, age, and blood pressure) in order to develop the prediction model which can be utilized in decision-making support system.

3.1 Korean National Health and Nutrition Examinations Survey (KNHANES)

The data we utilized in the present study was generated by the Korea Centers for Disease Control and Prevention in regards to levels of public health, health consciousness and behaviors, food and nutrition intakes; this has been known to be representative and reliable for national-, cities-, and provinces-levels. Specifically, we used the fifth edition of KNHANES which covers from 2010 through 2012 [8].

3.2 Dataset

The fifth edition of KNHANES includes a total of 25,534 subjects (8,958 subjects in 2010, 8,518 subjects in 2011, and 8,058 subjects in 2012). Three hundred and thirty

five subjects out of 368 depression patients were included for the study as 3 patients and 30 patients did not respond regarding smoke and HDL cholesterol level, respectively. For cardiovascular diseases, hypertension, hyperlipidemia, stroke, myocardial infarction, and angina pectoris were investigated; it was defined as a cardiovascular disease patient if one has more than one of above disorders. The final data set was summarized in the Table 1. Subjects were divided into training data (70%; n = 225) and testing data (30%; n = 100) for the analysis.

Table 1. General characteristics

Attributes	Types	Values	Mean
Sex	Flag	[1= Male, 2= Female]	1.81
Age	Range	[19, 82]	55.37
HDL cholesterol	Range	[13, 132]	51.11
Total cholesterol	Range	[112, 385]	194.94
Systolic blood pressure	Range	[78, 181]	121.95
Smoking	Flag	[0= Non-smoker, 1= Smoker]	0.14
Diabetes	Flag	[0 = Absent, 1 = Diabetes]	0.1
Coronary heart disease	Flag	[0 = Absent, 1 = Present]	0.38

3.3 Result

Utilizing classification function coefficients derived from the linear discriminant analysis, we were able to determine cardiovascular diseases. Patients information was applied into two equations; if the values from the Eq.(1) were higher, then, it was considered as normal health condition otherwise (meaning the values from the Eq.(2) were higher) it was predicted as cardiovascular diseases.

$$Absent : 14.110*(Sex) + 0.174*(Age) + 0.101*(Total\ cholesterol) + 0.207*(HDL\ cholesterol) + 0.400*(Systolic\ blood\ pressure) + -1.163*(Diabetes) + 9.869*(Smoking) - 57.504 \quad Eq. (1)$$

$$Present : 15.082*(Sex) + 0.238*(Age) + 0.099*(Total\ cholesterol) + 0.199*(HDL\ cholesterol) + 0.411*(Systolic\ blood\ pressure) + 0.404*(Diabetes) + 9.617*(Smoking) - 63.704 \quad Eq. (2)$$

Utilizing the identical dataset, we compared the prediction power between FRS, one of the representative prediction model for cardiovascular diseases, and the linear discriminant analysis; in results FRS showed 62.4% of accuracy while suggested prediction model with linear discriminate analysis represented 69% of accuracy.

In order to characterize, classification and misclassification of results data were analyzed; in FRS, the rates of classification and misclassification were 50.8% and 49.2%, respectively. Meanwhile, in the linear discriminant analysis, the rates of classification and misclassification were shown to be 68.6% and 31.4%, respectively. These results might be because the FRS is not formulated for the prediction of Korean patients with cardiovascular diseases and depression. In contrast, the model suggested

in the present study was based on the Korean dataset which may explain better accuracy as well as low error rate. Therefore it seems that the suggested prediction model may be more suitable for prediction of cardiovascular diseases in Korean depression patients.

4 Conclusions

In the present study, we developed the model for risks of cardiovascular diseases in depression using the linear discriminant analysis based on data-mining which allows early prediction. As results, the suggested model provided better accuracy as well as lower error rates than the guidelines of FRS. Therefore, the prediction model in the study may help to predict cardiovascular diseases for not only depression patients but also doctors. And eventually, it is considered to be an effective model for depression patients to prevent cardiovascular diseases and self-manage their health via monitoring risks.

References

1. Yang, Jung-Gi, et al.:Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS-LDA). 1--12, Personal and Ubiquitous Computing (2013)
2. Lim Seojin : The impact of educational program on quality of life and depression in heart failure patients. Korea University (2013)
3. Kang Huiju. :Prevalence and risk factors of Poststroke depression. Chonnam National University (2011)
4. Yang, Junggi, and Youngho Lee : Development of Measurement Model for the Value of QOL as an Influential Factor of Metabolic Syndrome. Wireless Personal Communications, 1--16, (2014)
5. Yang, Jung Gi, et al. : A Study of Cardiovascular Disease Prediction Models Using Discriminant Analysis. Information Science and Applications (ICISA), 2013 International Conference on. IEEE, (2013)
6. Yang, J-G., et al.:Study on a HDSS-based PEI model for chronic disease management. Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication. ACM, (2014)
7. Choi, Sungbin, et al. "Semantic concept-enriched dependence model for medical information retrieval." Journal of biomedical informatics 47. 18--27 (2014)
8. The Fifth Korea National Health and Nutrition Examination Survey (KNHANES V-3), Korea Centers for Disease Control and Prevention, (2012)