

Prediction of Depressive Tendency from Multidimensional Health Data Collected through Crowdsourcing

Shiori Yamaguchi¹, Hiroki Tanaka¹, Hayato Maki¹, Shigehiko Kanaya^{1,2}, Nobutaka Suzuki³, Satoshi Nakamura^{1,2}

Abstract—This paper investigates whether some aspects of health data are related to depressive tendency. A limitation of previous studies is using only a few aspects of health data, which motivated us to utilize a multidimensional health data. 987 participants, recruited through crowdsourcing answered 253 questions asking about their lifestyle and depressive tendency. Classification between high and low depressive tendency was performed using the health data as features. A random forest based classifier reached 0.97 accuracy, which outperformed a logistic regression based classifier. Important features were extracted from several categories including constitution and personality.

I. INTRODUCTION

Depression is one of the mood disorders, having symptoms, lost of interest or pleasure and depressed mood [1].

Many depression risk factors may exist, leading researchers to investigate the factors related to depression. It was suggested that the level of satisfaction regarding human relationship was related to depressive tendency [2]. Another study reported the strong relationship between eating alone and depression [3].

These past studies focused on few aspects of health data when examining depression risk factors. In addition, generalization ability for unknown data was not discussed in spite of creating predictive models for depressive tendency. This is because their main interest was to investigate the relationships between depression and personal life features as opposed to predicting depressive tendency.

The purpose of the current study is predicting depressive tendency from the multidimensional health data, and finding more efficient questions for the purpose.

II. COLLECTION AND ANALYSIS OF MULTIDIMENSIONAL HEALTH DATA

A. Questionnaire

We created a questionnaire, which includes six questions [4] indicating depressive tendency (K6) and 247 questions about a health data, divided into ten categories (constitution, personality, demographic, communication, smoking, exercise, health condition, sleeping and dietary habits, interaction with neighbors).

¹Shiori Yamaguchi, Hiroki Tanaka, Hayato Maki are with Graduate School of Information Science, Nara Institute of Science and Technology, Japan. yamaguchi.shiori.yk7@is.naist.jp

²Shigehiko Kanaya and Satoshi Nakamura are with Data Science Center, Nara Institute of Science and Technology, Japan.

³Nobutaka Suzuki is with Kanazawa University, Japan.

TABLE I
IMPORTANT QUESTIONS FROM SEVERAL CATEGORIES.

Rank	Categories	Questions
1	Constitution	Are you likely to be scared or anxious?
2	Constitution	Do you have energy?
6	Personality	I am relaxed, handle stress well.
7	Personality	I get nervous easily.
13	Demographic	How old are you?
26	Dietary habit	How many times do you order take-away?

B. Classification

We created predictive models that classify 372 participants, recruited from crowdsourcing as high or low depressive tendency using their health data as features. The participants with K6 score 13+ were labeled as “high depressive tendency” whereas the rests were labeled as “low depressive tendency”. The participants were divided into three groups, train, validation, and test (7:1:2), and logistic regression (LR) as well as random forest (RF) were employed for the classification.

III. RESULTS AND CONCLUSION

The classification accuracy of LR and RF were 0.95 and 0.97, respectively. RF made one false positive and one false negative predictions. Table I describes efficient questions in RF, which include constitution, personality describing neuroticism, and dietary habits.

In conclusion, this work proposed a new approach, using a multidimensional health data and crowdsourcing. The experimental result showed the possibility that depressive tendency is predicted by individuals’ health data.

REFERENCES

- [1] B.W. Donald and G.E. Jon, DSM-5 guidebook : the essential companion to the diagnostic and statistical manual of mental disorders, fifth edition. Washington, District of Columbia ; London, England : American Psychiatric Publishing, 2014.
- [2] S.K. Davidson, C. F. Dowrick, and J. M. Gunn, Impact of functional and structural social relationships on two year depression outcomes: A multivariate analysis., *Journal of Affective Disorders*, 2016.
- [3] Y. Kimura, T. Wada, K. Okumiya, Y. Ishimoto, E. Fukutomi, Y. Kasahara, W. Chen, R. Sakamoto, M. Fujisawa, K. Otsuka, and K. Matsubayashi, Eating alone among community-dwelling Japanese elderly: Association with depression and food diversity, *Journal of Nutrition, Health and Aging*, 2012.
- [4] R.C. Kessler, G. Andrews, L.J. Colpe, E. Hiripi, D. K. Mroczek, S. L. Normand, E. E. Walters, and A. M. Zaslavsky. Short screening scales to monitor population prevalences and trends in non-specific psychological distress., *Psychological Medicine*, 2002.