Department of Electrical, Electronic and Computer Engineering	Key words	Generative AI, Machine Learning, Deep Learning, Neural Networks, Social Media, Text Mining, Natural Language Processing, Data Science

Doctor (Engineering) / Professor

Ikuo Keshi

Education

Osaka University, Faculty of Engineering, Electronic Engineering
Osaka University, Graduate School of Engineering, Electronic Engineering Master's Course
Nara Institute of Science and Technology, Division of Information Science, Doctor's Course

Professional Background

Sharp Corporation, e-Life Department General Manager/Chief Technical Research Fellow
Artificial Intelligence Laboratory of Massachusetts Institute of Technology, Visiting Scientist

Consultations, Lectures, and Collaborative Research Themes

Collaborative research and consultation on the application of Generative AI in healthcare, nursing, tourism, customer support, and education, along with lectures on 'Practical Examples of Society 5.0' and 'The Appeal and Applications of Cutting-Edge Generative AI.'

e-mail address

keshi@fukui-ut.ac.jp

Main research themes and their characteristics

「Interpretable Disease Name Estimation Using Semantic Representation Learning of Medical Terms」

This research introduces a method for estimating disease names using semantic representation learning, based on a semantic-vector dictionary created from a disease name thesaurus. In 90% of cases, the interpretability of the model is enhanced by the fact that disease feature words representing higher-level concepts carry the greatest weight, making them essential for accurate disease name estimation. The goal is to assist in automatically assigning ICD-10 codes from medical documents, particularly in the Japanese version of Computer Assisted Coding (CAC).

The medical-term semantic-vector dictionary, illustrated in Figure 1, connects 299 disease feature words with disease names, forming the basis for interpretable diagnosis estimation. A neural network with initial weights based on a medical-term semantic-vector dictionary (as shown in Figure 2) processes the first 500 characters of progress summaries in discharge summaries. Patient-specific variables such as gender, age, and department are then added to this vector, which is subsequently used for ICD-10 classification using a linear SVM model. The interpretability of the model is enhanced by visualizing the weight distribution of feature words. For example, the distribution of weights for the disease feature words 'neonatal disorder' is linked to conditions related to newborns through ICD-10 codes starting with 'P,' as shown in Figure 3.

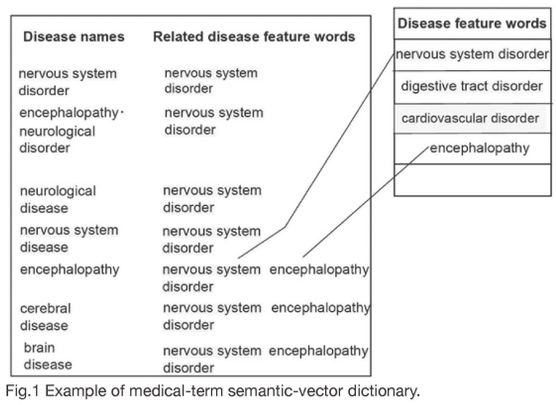

Fig.1 Example of medical-term semantic-vector dictionary.

An evaluation benchmark using Toyama University Hospital's discharge summaries, where the training and test sets differed in data distribution (old vs. new electronic medical records), showed that this approach improved the F1 score for disease name estimation by 10 points compared to conventional methods, reaching 72.4 points. Further fine-tuning with large language models like BERT improved the F1 score by an additional 10 points but highlighted challenges with interpretability. Achieving accurate and easily interpretable estimation of disease names from a patient's chief complaint is a future goal.

Future directions include applying Generative AI to enhance the generalization and adaptability of the method for broader clinical applications.

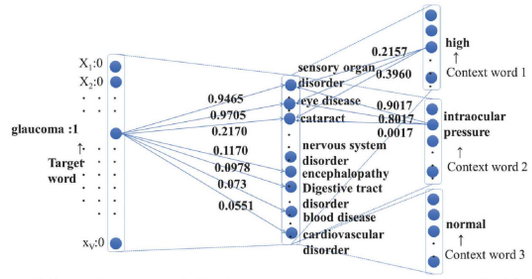

Fig.2 Semantic representation learning process based on the medical-term semantic vector dictionary.

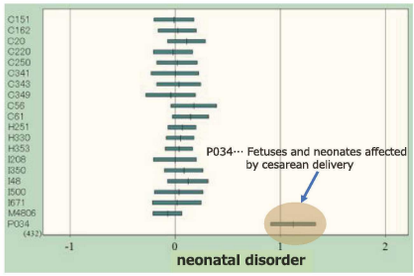

Fig.3 Distribution of weights by ICD-10 code for the disease feature word 'neonatal disorder.'

Major academic publications

I. Keshi, R. Daimon, A. Hayashi,
"Interpretable Disease Name Estimation based on Learned Models using Semantic Representation Learning of Medical Terms", Proc. of the 14th IC3K, 265-272, Oct. 2022.

K. Tsujioka, I. Keshi, H. Nakagawa, A. Hayashi,
"Research on a method for constructing a Japanese version of computer-assisted coding using natural language processing", Health Information Management, 34(1):56–64, 2022.

I. Keshi, R. Daimon, Y. Takaoka, A. Hayashi,
"Integrated Evaluation of Semantic Representation Learning, BERT, and Generative AI for Disease Name Estimation Based on Chief Complaints", Proc. of the 16th IC3K, 2024 (in press).