Development of a case database including atypical cases of rare diseases with emphasis on symptoms and findings using generative AI

Eisuke Dohi[1)], Itaru Hayakawa[2)], Tomoyasu Matsubara[3)], Toyofumi Fujiwara[4)], Tomoyasu Yamamoto[4)]

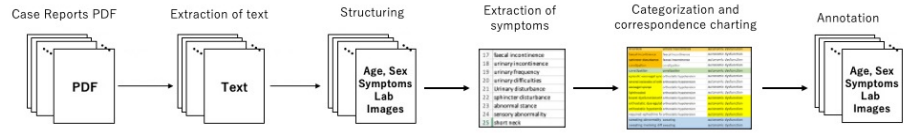
1) National Center of Neurology and Psychiatry, National Institute of Neuroscience, 2) National Center for Child Health and Development, Department of Neurology 3) Tokushima University Department of Neurology, 4) Database Center for Life Science

Background

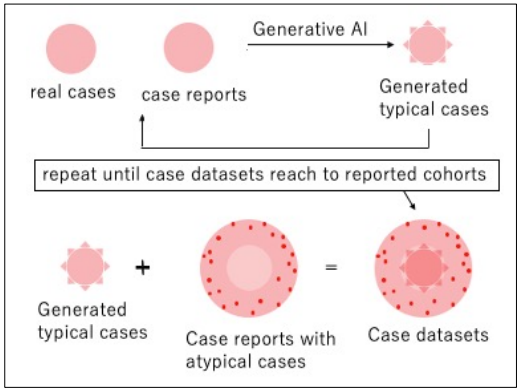
In the case of intractable and rare diseases, the sheer number of diseases and the small number of individual cases prevents medical practitioners from gaining sufficient case experience. Educational resources are also limited to textbooks with typical cases and case reports biased toward atypical cases, and there is a lack of educational resources to grasp the overall picture of the disease. Even if all of the published learning resources were collected, case reports are biased toward atypical cases, making it difficult to create a patient distribution that reflects the image of the disease seen in actual clinical practice.

Therefore, this study aims to create data resources for learning and diagnostic support by constructing a case database that is similar to real clinical cases using generative AI.

Method



1: Case data were extracted from case reports PDF and symptoms and findings were categorized. Case data were annotated with corresponding charts. All the procedure were conducted with generative AI (LLM such as ChatGPT) except making the corresponding charts.



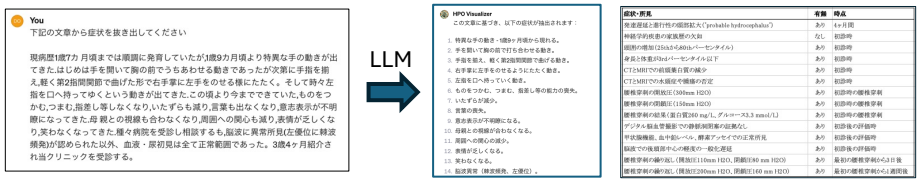
2: To compensate typical cases, typical cases will be generated with generative AI with typical cases in real cases and case reports. The additional generated typical cases will be added to case reports datasets until basic statics reach previously reported cohort studies.

3: Case dataset will be visualized based on their symptoms and findings in individual cases to ease to understand the atypical case among the case datasets.

Results

① PDF → text extraction：Ask your PDF pro works well!

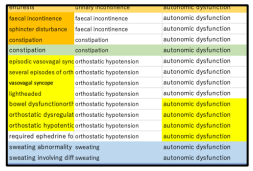
② The extraction of symptoms：LLM works well (+/-, time course)



③ Annotation with LLM did not work with HPO....

④ Annotation with corresponding chart works!

corresponding chart

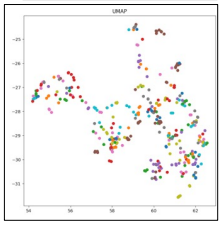


LLM

Annotated clinical text

A 13-year-old boy exhibited cervical kyphosis and deteriorating cognition<"dementia">. He was born normally and achieved developmental milestones without delay. The family history was noncontributory. At the age of 2, he had 2 simple febrile seizure<"seizure">s. At the age of 8, he had a generalized febrile convulsion that lasted about 10 min. Subsequent electroencephalography (EEG) revealed bilateral frontal spikes during sleep. Valproate was started under the diagnosis of epilepsy. Since then, his school performance slowly began to deteriorate<"psychomotor regression">.

visualization pf typical and atypical cases



Current issues and solusions

(1) It is unclear if the HPO does not have term or if the annotation function is inadequate.

(2) Streamlining the creation of correspondence tables.

(3) Stability o results when using LLM

Organizing the HPO to make it easier for clinicians to understand.
Automate each procedure and clarify where expert help is needed.

Acknowledgments and Fundings

Special thanks to Yuka Tateishi, Terue Takatsuki and our mentors, advisors, and colleagues for their guidance and support.

Kakenhi 22K12253, 23K11886, Rois-DS Joint 2023