Fine-tuned Large Language Models for Recommending Genetic Tests for Neurodevelopmental Diseases

Kai Wang
Children's Hospital of Philadelphia, Pennsylvania, USA

Abstract:

Introduction: Molecular genetic testing has become an integral component in diagnosing rare diseases. Despite technological advances that have lowered sequencing costs, selecting the appropriate type of genetic tests remains complex due to many factors such as accuracy, informatics costs, and logistical requirements. Currently, gene panels and whole-exome/whole-genome sequencing (WES/WGS) are two main diagnostic modalities; however, determining which one to use often relies on human experts' interpretation of ACMG (American College of Medical Genetics) guidelines, a process that can be slow and difficult to standardize.

Methods: Here we develop a novel integrative approach utilizing LLMs to assist clinicians in choosing between gene panels and WES/WGS by analyzing structured phenotypes, ICD-10 codes, and unstructured clinical notes for patients with neurodevelopmental and other clinical phenotypes. We fine-tuned Llama 3.1 8B and 70B models using a self-distillation process with chain-of-thought (CoT) reasoning, generated by a foundation model.

Results: Evaluation on real clinical data show that our top-performing model reached 78.0% accuracy, 78.1% precision, 74.9% recall and 76.5% F1-score, demonstrating a strong potential to facilitate human experts to streamline diagnostic workflows and make informed clinical decisions. Notably, the self-distillation fine-tuning method, incorporating chain-of-thought explanations from the Llama 3.1 70B model, significantly improved the model's ability to provide detailed and thoughtful reasonings before recommending the optimal genetic tests.

Conclusions: We introduce a novel methodology that uses LLM-based model, pretrained on large-scale datasets and further fine-tuned with domain-specific knowledge, to provide accurate molecular genetic recommendations for diagnosing rare diseases. This algorithm should be only used as a supporting tool rather than the primary decision-making tool, as other factors can influence physician's assessment of clinical information and making decisions on genetic tests.