

Large Language Learning Models For Perceptual Speech Clinical Data Extraction in 50 Cleft Palate Speech Pathology Notes



Chelsea Yim¹, Amir Ganjouei MD¹, Andi Zhang MD², Jackie Kukoff CCC/SLP¹, Janna Spencer CCC/SLP², Alexander Lin MD¹

¹University of California San Francisco (UCSF), Oakland, CA, USA, ²St. Louis University, St. Louis, MO, USA



Background

- Speech language pathology assessments are crucial for diagnosing and planning treatment for patients with cleft palate
- Manual analysis of speech notes is time-intensive and susceptible to subjective interpretation
- Recent advancements in natural language processing, specifically large language models (LLMs), offer a promising solution for automating and enhancing the extraction of diagnostic information from clinical texts

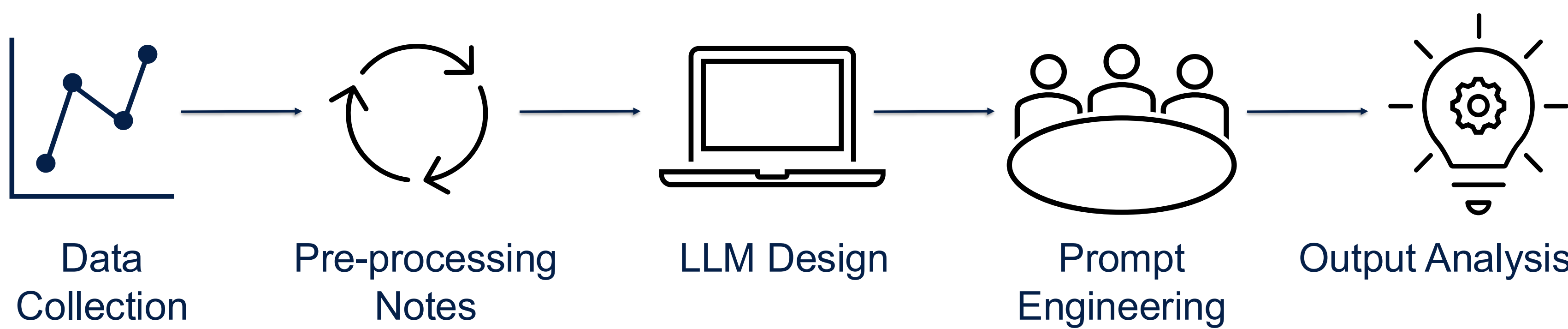
Purpose

Evaluate the efficacy of an in-house version of the GPT-4 model (VERSA) in extracting concepts from speech sample notes written by speech-language pathologists (SLPs)

Methods

- Retrospective analysis of 50 anonymized speech sample notes authored by two SLPs at a single-institution
- Processed using an in-house adaptation of the GPT-4 model (VERSA) to extract key speech and language concepts
- Prompts: patient's diagnosis, any surgical history along with the date of surgery, resonance, nasal airflow evaluations, grimace, and speech acceptability
- Resonance evaluation included hypernasality (0-4 scale) and hyponasality on a (0–2 scale)
- Nasal airflow evaluations assessed audible nasal emission or nasal turbulence (0-2 scale)
- Grimace (0 or 1) and speech acceptability (0-3)

Methods (Continued)



Results

Average patient age: 8.5y ± 4.84
N = 50

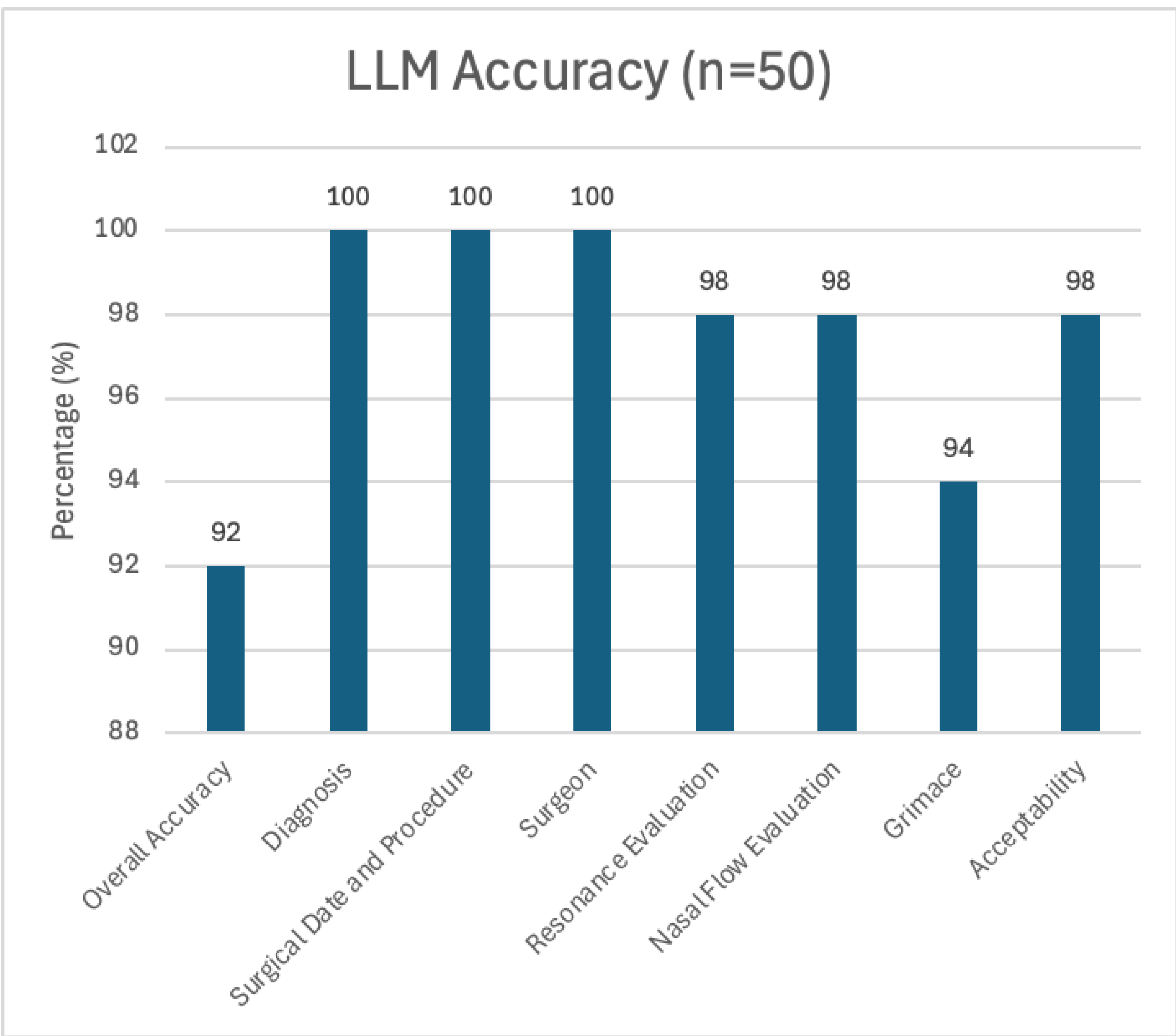


Figure 1: Large Learning Model Output Accuracy per Variable
Large Learning Model Accuracy per Variable

Error Descriptions

- Resonance Evaluation**
Incorrect output
- Nasal Airflow Evaluations**
Partially correct, incomplete data
- Grimace Rating**
Grimace category not present in output data

Discussion

- The LLM had an overall accuracy of 92%, with the lowest performance in the "grimace" category
- These preliminary findings highlight the potential of LLMs in enhancing the efficiency and reliability of clinical assessments in speech-language pathology by streamlining the extraction of diagnostic and evaluative information from pediatric speech pathology notes
- This model reduced the time spent manually retrieving information and opens the potential to aggregate data to see trends and comprehensively extract data for research

Limitations

- Small cohort of 50 anonymized speech notes from two speech language pathologist at a single-institution
- Limited generalizability of LLM at other institutions
- Future studies of this LLM model will focus on expanding the dataset and refining model parameters to further improve accuracy and applicability in diverse clinical scenarios

References

1. Zhang, Andi & Pyon, Rachel & Chen, Kevin & Lin, Alexander. (2023). Speech Analysis of Patients with Cleft Palate Using Artificial Intelligence Techniques: A Systematic Review. FACE. 4. 10.1177/27325016231187985.
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8
3. Melnyk O, Ismail A, Ghorashi NS, Heekin M, Javan R. Generative Artificial Intelligence Terminology: A Primer for Clinicians and Medical Researchers. Cureus. Published online December 4, 2023. doi:10.7759/cureus.49890
4. Jerfy A, Selden O, Balkrishnan R. The Growing Impact of Natural Language Processing in Healthcare and Public Health. Inq J Med Care Organ Provis Financ. 2024;61:469580241290095. doi:10.1177/00469580241290095