INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Automated Neural Nursing Assistant (ANNA): An Over-The-Phone System for Cognitive Monitoring

Jacob Solinsky[1], *Raymond Finzel*[1], *Martin Michalowski*[2], *Serguei Pakhomov*[1]

[1]College of Pharmacy, [2]School of Nursing, University of Minnesota

solin020@umn.edu, finze006@umn.edu, martinm@umn.edu, pakh0002@umn.edu

Abstract

ANNA is a telephony-based cognitive assessment tool designed to aid nurses in caring for patients who require close monitoring for the development of confusion or neurological impairment. Of particular concern is the treatment of Immune Effector Cell-Associated Neurotoxicity Syndrome (ICANS), a condition which occurs quite frequently as an adverse outcome of Chimeric Antigen Receptor-T (CAR-T) cancer immunotherapy. ANNA employs both traditional verbal tests for cognitive impairment and novel linguistic methods which identify abnormalities in the patient's speech during ordinary conversation. To collect ordinary speech it uses a lightweight instance of the Facebook's Large Language Model BlenderBot to engage the patient in a partially unscripted conversation. ANNA is designed with easy employment by healthcare providers in mind, being sufficiently lightweight to run on consumer-grade hardware and needing access only to a patient's phone number to interact with them.

Index Terms: cognitive linguistics, spoken dialogue and conversational AI systems, technologies and systems for new applications

1. Introduction

ICANS is an iatrogenic brain inflammatory condition that occurs as an adverse outcome of CAR-T cancer immunotherapy in over 40% and up to 90% of patients [1]. It has a highly variable presentation from patient to patient, including symptoms such as cognitive, linguistic, and motor impairments. Anti-inflammatory interventions such as corticosteroid administration are quite effective if this condition is caught early, but if treatment is delayed, ICANS can progress to irreversible brain damage and even death [2]. With chronic labor shortages in hospitals around the United States and worldwide [3], daily administration of cognitive tests to patients undergoing immunotherapy often exceeds what a hospital's nursing staff can provide. In addition, the verbal cognitive assessments typically used as standard of care to diagnose ICANS, such as the Immune Effector Cell Encephalopathy (ICE) questionnaire, tend to be insensitive to mild manifestations of cognitive impairment, only identifying advanced stages of confusion, delirium, or dementia. We are developing ANNA as a tool to enable caregivers to monitor patients intensively multiple times per day for early signs of cognitive difficulties brought about by ICANS. Because ICANS exhibits such a variety of symptoms and remains fairly poorly studied, we have incorporated a diverse set of cognitive impairment metrics into ANNA, in hopes that some subset of them will prove effective in identification of early ICANS. To this end we have designed a clinical study to in collaboration with the University of Minnesota Masonic Cancer Center to determine the feasibility and effectiveness of ANNA for early diagnosis of ICANS.

2. System Description

ANNA is a web and telephony application which, through its phone number, it can call and be called, and through its web portal, a control page through which an administrator can schedule new phone calls and view the transcripts and test results of previous calls. By far the largest part of the application is devoted to managing ongoing conversations with patients and calculating various diagnostic metrics (e.g., word list recall, verbal fluency scores and a range of speech and language fluency characteristics) from both transcripts of the patient's ordinary conversation and the results of the verbal cognitive tests ANNA administers. We provide ANNA as an easily deployed set of Docker images which can be deployed on an onsite server when configured with a phone number, web address, and a GPU device with at least 20Gb of VRAM.

2.1. Conversation Manager

ANNA's architecture can be split into two detachable components, a conversation manager and a cognitive assessment test battery. As the conversation manager goes through the script of the phone call, it transcribes the patient's speech, responds with synthesized speech, plays audio, and listens for pauses and cue words from the patient's speech to allow ANNA to take turns in conversation in a natural manner. For speech transcription it employs OpenAI's state-of-the-art Whisper transcriber (large-v2 model), which we have found to produce transcriptions with enough accuracy to support subsequent analysis of audio recordings from even the lowest end consumer phones. For speech generation we use Mozilla's Tacotron-2 "Judy" voice, which was most highly rated among a variety of open-source voice systems by an M-Turk study conducted in 2020 [4], being preferred even over one of said study's human voice actors. When the conversation manager is required to generate unscripted text to the patient in open- ended conversation, it uses the 400 million parameter version of Facebook's BlenderBot large language model (blenderbot- 400m-distill).

Though this frontend is not state-of-the-art, it is sufficiently lightweight to be operated on infrastructure available in clinical settings, which is crucial for protecting sensitive health information. Twilio currently provides telephony services to ANNA, however, the conversation manager can easily be reconfigured to accept input from and produce output for other audio recording and playback devices, allowing us to reuse it in our group's other voice application projects.

Table 1: *An Overview of ANNA's Script*

1. Introduction
2. Ask about patient's condition (2 turns)
3. Open ended conversation (2 turns)
4. Word List Memory test (initial)
5. F-initial words COWAT test
6. Animal names COWAT test
7. Word List Memory recall

3. Cognitive Measures

3.1. Cognitive Testing

ANNA's cognitive testing backend combines both classical cognitive assessments and novel methods developed by our own group and collaborators. On each phone call, it administers three traditional tests. The first is the Word List Memory (WLM) test, in which six randomly chosen common words are read aloud to the patient, who is asked to repeat them, an approach which has shown some utility in identifying mild cognitive impairment [5]. It then administers two variants of the Controlled Oral Word Association Test (COWAT), requesting the patient to name as many distinct words starting with the letter F as they can in the space of 30 seconds, and then repeating the same test but asking for as many distinct names of animals as the patient can produce, also found to be sensitive to mild cognitive impairment [6]. Finally, the patient is asked again to list as many of the six previously selected words as they can remember, finishing the WLM test.

3.2. Novel Conversational Metrics

An individual's fluency in ordinary conversation provides a wealth of information about their cognitive status. A variety of mental disorders, such as schizophrenia, bipolar disorder, and Alzheimer's disease, are characterized by linguistic abnormalities in spontaneous speech [7], [8]. ANNA incorporates two linguistic measures of cognitive impairment.

The first, perplexity, is a metric of how "difficult" it is for a large language model to predict a segment of text. Large language models such as OpenAI's GPT family are trained to extend and predict text. Text with conventional grammar, word choice, and narrative structure has low perplexity, whereas text lacking in these qualities, such as transcripts of speech from patients experiencing varying degrees of dementia, has higher perplexity. Using perplexity we achieved state-of-the-art results on distinguishing patients with and without dementia based on a transcription of their description of an image in the "Cookie Theft" description corpus [9].

Our second metric examines a variety of measures of syntactic complexity. Cognitive impairments, by reducing working memory, limit the syntactic complexity of sentences patients can produce [10]. Our group has previously developed a toolkit for measuring syntactic complexity [11], which we reuse in ANNA.

4. Future Research

To determine the feasibility and efficacy of ANNA for ICANS identification we have designed a longitudinal clinical study in partnership with the University of Minnesota's Masonic Cancer Center to have a cohort of patients undergoing immunotherapy speak with ANNA on a three times daily basis over a period of several weeks post immunotherapy medicine infusion. The caregivers will monitor patients for ICANS using conventional methods such as the ICE instrument. In this study, we will determine the extent to which ANNA can identify early indicators of ICANS using the aforementioned manual ICE assessments as the reference. We also plan to identify a set of novel measures consisting of language perplexity and complexity features indicative of early ICANS from patients' interactions with the conversational part of ANNA's assessments.

5. Acknowledgements

We would like to thank the University of Minnesota's Masonic Cancer Center for the collaboration which has made the development of ANNA possible. We would also like to the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing the computational resources required to develop ANNA.

6. References

[1] C. Belin et al., "Description of neurotoxicity in a series of patients treated with CAR T-cell therapy," *Scientific Reports*, vol. 10, no. 1, p. 18997, Nov. 2020, number: 1 Publisher: Nature Publishing Group.

[2] E. C. Morris, S. S. Neelapu, T. Giavridis, and M. Sadelain, "Cytokine release syndrome and associated neurotoxicity in cancer immunotherapy," *Nature Reviews. Immunology*, vol. 22, no. 2, pp. 85–96, 2022.

[3] L. M. Haddad, P. Annamaraju, and T. J. Toney-Butler, "Nursing Shortage," in *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2023.

[4] J. Cambre, J. Colnago, J. Maddock, J. Tsai, and J. Kaye, "Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, Apr. 2020, pp. 1–13.

[5] J. R. Bock, J. Russell, J. Hara, and D. Fortier, "Optimizing Cognitive Assessment Outcome Measures for Alzheimer's Disease by Matching Wordlist Memory Test Features to Scoring Methodology," *Frontiers in Digital Health*, vol. 3, 2021.

[6] K. Bauer and M. Malek-Ahmadi, "Meta-analysis of Controlled Oral Word Association Test (COWAT) FAS performance in amnestic mild cognitive impairment and cognitively unimpaired older adults," *Applied Neuropsychology: Adult*, vol. 0, no. 0, pp. 1–7, Aug. 2021.

[7] J. Jerónimo, T. Queirós, E. Cheniaux, and D. Telles-Correia, "Formal Thought Disorders–Historical Roots," *Frontiers in Psychiatry*, vol. 9, 2018.

[8] D. Kempler and M. Goral, "Language and Dementia: Neuropsychological Aspects," *Annual review of applied linguistics*, vol. 28, pp. 73–90, Jan. 2008.

[9] C. Li, D. Knopman, W. Xu, T. Cohen, and S. Pakhomov, "GPT-D: Inducing Dementia-related Linguistic Anomalies by Deliberate Degradation of Artificial Neural Language Models," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1866–1877.

[10] S. Kemper, E. LaBarge, F. R. Ferraro, H. Cheung, H. Cheung, and M. Storandt, "On the preservation of syntax in Alzheimer's disease. Evidence from written sentences," *Archives of Neurology*, vol. 50, no. 1, pp. 81–86, Jan. 1993.

[11] S. Pakhomov, D. Chacon, M. Wicklund, and J. Gundel, "Computerized assessment of syntactic complexity in Alzheimer's disease: a case study of Iris Murdoch's writing," *Behavior Research Methods*, vol. 43, no. 1, pp. 136–144, Mar. 2011.