DETECTION OF COGNITIVE IMPAIRMENT FROM SPEECH USING ARTIFICIAL INTELLIGENCE METHODS

Karimova Maftuna Bahodir qizi
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Assistant, Department of Robotics and Intelligent Systems
m.karimova@tuit.uz

Annotatsiya. *Uşbu tezisda nutq signallari va ularning matnli transkripsiyalarini sun'iy intellekt usullari yordamida tahlil qiliş orqali kognitiv buzilişlarni, xususan yengil kognitiv buziliş va Altsgeymer kasalligini erta aniqlaş masalasi körib çiqilgan. Akustik, temporal va lingvistik belgilar guruhlari tavsiflangan, an'anaviy maşinaviy öqitiş usullari hamda oldindan öqitilgan nutq va til modellari (wav2vec 2.0, Whisper, BERT) asosidagi yondaşuvlar qiyosiy tahlil qilingan. Akustik va lingvistik belgilarni birlaştiruvçi skrining tizimining umumlaşgan sxemasi taklif etilgan, özbek tili uçun patologik nutq korpuslarining mavjud emasligi asosiy muammo sifatida körsatilgan.*

Kalit sözlar: *kognitiv buziliş, Altsgeymer kasalligi, nutq tahlili, maşinaviy öqitiş, akustik belgilar, til modellari, skrining.*

Аннотация. *В тезисе рассматривается задача раннего выявления когнитивных нарушений, в частности умеренного когнитивного расстройства и болезни Альцгеймера, на основе анализа речевых сигналов и их текстовых транскрипций методами искусственного интеллекта. Описаны группы акустических, временных и лингвистических признаков, проведён сравнительный анализ подходов на основе классического машинного обучения и предобученных речевых и языковых моделей (wav2vec 2.0, Whisper, BERT). Предложена обобщённая схема системы скрининга с объединением акустических и лингвистических признаков; отсутствие корпусов патологической речи на узбекском языке выделено как ключевая проблема.*

Ключевые слова: *когнитивные нарушения, болезнь Альцгеймера, анализ речи, машинное обучение, акустические признаки, языковые модели, скрининг.*

Abstract. *This thesis addresses the early detection of cognitive impairment, in particular mild cognitive impairment and Alzheimer's disease, through the analysis of speech recordings and their transcripts using artificial intelligence methods. Groups of acoustic, temporal and linguistic features are described, and approaches based on classical machine learning and on pretrained speech and language models (wav2vec 2.0, Whisper, BERT) are compared. A generalized scheme of a screening system that fuses acoustic and linguistic features is proposed, and the absence of*

pathological speech corpora for the Uzbek language is identified as the key challenge.

Keywords: *cognitive impairment, Alzheimer's disease, speech analysis, machine learning, acoustic features, language models, screening.*

Introduction. Dementia is one of the leading causes of disability and dependency among older people. According to the World Health Organization, more than 55 million people live with dementia worldwide, and nearly 10 million new cases are registered every year [1]. Alzheimer's disease accounts for the majority of these cases. It is preceded by a stage of mild cognitive impairment (MCI), in which cognitive decline is measurable but daily functioning is largely preserved [2]. Identifying patients at the MCI stage is clinically important, because this is the period when lifestyle interventions, treatment of risk factors and emerging disease-modifying therapies are expected to be most effective.

Standard screening relies on neuropsychological tests such as the Montreal Cognitive Assessment (MoCA) [3]. These instruments require a trained specialist and face-to-face time, their scores depend on education level and language, and in practice they are rarely administered at the primary-care level. As a result, a large share of patients are diagnosed only at the stage of moderate dementia.

Speech is a promising alternative source of diagnostic information. Producing coherent speech simultaneously engages memory, attention, lexical retrieval, planning and motor control, so cognitive decline is reflected in how a person speaks and what they say, often years before a formal diagnosis. Recording speech is cheap, non-invasive and can be done with an ordinary smartphone. Modern artificial intelligence (AI) methods make it possible to extract hundreds of objective markers from such recordings and to classify them automatically.

The aim of this thesis is to analyze current approaches to speech-based detection of cognitive impairment and to propose a generalized scheme of a screening system adapted to the conditions of Uzbekistan.

1. Speech as a marker of cognitive decline. Speech for analysis is usually collected through standardized tasks: description of a picture (most often the "Cookie Theft" picture from the Boston Diagnostic Aphasia Examination [4]), semantic and phonemic verbal fluency (for example, naming as many animals as possible in one minute), retelling of a short story, reading aloud, and spontaneous conversation. The most widely used data source is the Pitt corpus of DementiaBank, collected in a longitudinal study of Alzheimer's disease [5].

Typical changes in the speech of people with cognitive impairment include longer and more frequent pauses, a lower speech rate, more hesitations and filled

pauses, word-finding difficulties, reduced lexical richness, overuse of pronouns and generic words instead of specific nouns, simplified syntax, repetitions and reduced informativeness of the description. Fraser et al. extracted 370 linguistic and acoustic features from picture descriptions and showed that they group into four factors — semantic impairment, acoustic abnormality, syntactic impairment and information impairment; a classifier built on these features reached an accuracy of about 81% in separating patients with Alzheimer's disease from healthy controls [6].

2. Feature groups. The features used in the literature can be divided into six groups, summarized in Table 1. The first two groups are computed directly from the audio signal and are largely language-independent; the next three require a transcript; the last group consists of learned representations from pretrained neural networks.

Table 1. *Groups of speech features for detecting cognitive impairment*

Group	Examples of features	Typical tools
Acoustic	fundamental frequency (F0) and its variability, jitter, shimmer, harmonics-to-noise ratio, MFCC, energy	openSMILE (eGeMAPS set [7]), Praat
Temporal	speech rate, articulation rate, number and duration of pauses, phonation-to-pause ratio	voice activity detection, forced alignment
Lexical	type-token ratio, Honoré's statistic, noun-to-pronoun ratio, word frequency, filler words	tokenizers, POS taggers
Syntactic	mean length of utterance, parse-tree depth, share of subordinate clauses	dependency parsers
Semantic	number of information units, idea density, coherence between utterances	sentence embeddings
Deep representations	embeddings of audio and text produced by pretrained models	wav2vec 2.0, Whisper encoder, BERT

3. Classification methods. The classical pipeline consists of extracting hand-crafted features, normalizing them, selecting the most informative ones and training a classifier — support vector machines, logistic regression, random forests or gradient boosting. Such models are interpretable and work reasonably well on the small datasets typical of clinical research. König et al. showed that vocal features extracted from simple tasks such as counting backwards and sentence repetition can distinguish healthy individuals, MCI and Alzheimer's disease patients with high accuracy [8].

A major problem of early studies was the lack of comparable evaluation: results were reported on different subsets of data, often unbalanced by age and sex.

The ADReSS challenge addressed this by providing a subset of 156 speakers from the Pitt corpus, balanced by diagnosis, age and sex, with a fixed train/test split; the official baseline reached 75% accuracy on the test set [9]. On this benchmark, fine-tuning the BERT language model on manual transcripts achieved 83.3% accuracy [10], and encoding pauses and disfluencies directly in the transcript before fine-tuning raised accuracy to 89.6% [11].

In parallel, self-supervised speech models have appeared. wav2vec 2.0 learns representations of raw audio from large amounts of unlabeled speech [12], and Whisper, trained on 680 thousand hours of multilingual data, performs robust speech recognition with timestamps [13]. Their use removes the need for manual transcription, which is the most expensive step in building a practical system.

A systematic review of the field [14] notes that most studies use small samples, lack standardized recording protocols, are conducted almost exclusively in English and rarely include longitudinal follow-up. Across studies, linguistic features from transcripts are generally more informative than purely acoustic ones, but acoustic features are less dependent on the language, which is important in a multilingual environment. The best results are usually obtained by combining both modalities.

4. Proposed scheme of a screening system. Based on this analysis, we propose a generalized scheme of a speech-based screening system consisting of six stages.

(1) Data acquisition. A mobile or web application records the user's responses to standardized tasks — picture description, semantic fluency and story retelling — in Uzbek or Russian, together with age, sex and years of education.

(2) Preprocessing. Noise reduction, voice activity detection, loudness normalization and segmentation of the recording into utterances.

(3) Automatic transcription. A Whisper model fine-tuned on Uzbek speech produces the transcript with word-level timestamps, from which pauses are also measured.

(4) Feature extraction in two branches. The acoustic branch computes the eGeMAPS set and wav2vec 2.0 embeddings; the linguistic branch computes lexical and syntactic features and embeddings of a multilingual language model such as XLM-R [15].

(5) Fusion and classification. Features are combined either by concatenation (early fusion) or by weighted averaging of the probabilities of separate classifiers (late fusion). The output is one of three classes — normal, possible MCI, suspected dementia — together with a continuous risk score.

(6) Interpretation. Feature contributions are estimated with SHAP values [16], so that the clinician sees which markers (for example, pause rate or lexical diversity) drove the decision.

The system is intended for screening, not for diagnosis: a positive result serves as a reason to refer the person to a neurologist for full neuropsychological assessment.

5. Challenges and prospects for Uzbekistan. The main obstacle is the absence of open corpora of pathological speech in the Uzbek language. Creating such a corpus requires cooperation with clinics, informed consent of participants and compliance with legislation on personal data. Uzbek is an agglutinative language, so lexical measures developed for English, such as type-token ratio, must be computed at the level of lemmas or morphemes. Frequent Uzbek–Russian code-switching and differences in education level must be taken into account in the design of tasks, and picture stimuli should be culturally adapted. A realistic first step is a pilot corpus in which every recording is accompanied by a MoCA score, which will allow validation of the proposed features against an established clinical instrument.

Conclusion. Speech analysis with AI methods is a non-invasive, inexpensive and scalable approach to the early detection of cognitive impairment. The combination of acoustic and linguistic features with pretrained speech and language models currently gives the best results on international benchmarks. The proposed scheme of a screening system can serve as a basis for developing a national tool, provided that Uzbek-language speech data are collected and the system is clinically validated.

FOYDALANILGAN ADABIYOTLAR

1. World Health Organization. Dementia: fact sheet [Electronic resource]. – Geneva: WHO, 2023. – URL: https://www.who.int/news-room/fact-sheets/detail/dementia.

2. Petersen R.C. Mild cognitive impairment as a diagnostic entity // Journal of Internal Medicine. – 2004. – Vol. 256, No. 3. – P. 183–194.

3. Nasreddine Z.S., Phillips N.A., Bédirian V. et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment // Journal of the American Geriatrics Society. – 2005. – Vol. 53, No. 4. – P. 695–699.

4. Goodglass H., Kaplan E. The assessment of aphasia and related disorders. – 2nd ed. – Philadelphia: Lea & Febiger, 1983. – 102 p.

5. Becker J.T., Boller F., Lopez O.L., Saxton J., McGonigle K.L. The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis // Archives of Neurology. – 1994. – Vol. 51, No. 6. – P. 585–594.

6. Fraser K.C., Meltzer J.A., Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech // Journal of Alzheimer's Disease. – 2016. – Vol. 49, No. 2. – P. 407–422.

7. Eyben F., Scherer K.R., Schuller B.W. et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing // IEEE Transactions on Affective Computing. – 2016. – Vol. 7, No. 2. – P. 190–202.

8. König A., Satt A., Sorin A. et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease // Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. – 2015. – Vol. 1, No. 1. – P. 112–124.

9. Luz S., Haider F., de la Fuente S., Fromm D., MacWhinney B. Alzheimer's dementia recognition through spontaneous speech: the ADReSS challenge // Proc. Interspeech 2020. – 2020. – P. 2172–2176.

10. Balagopalan A., Eyre B., Rudzicz F., Novikova J. To BERT or not to BERT: comparing speech and language-based approaches for Alzheimer's disease detection // Proc. Interspeech 2020. – 2020. – P. 2167–2171.

11. Yuan J., Bian Y., Cai X., Huang J., Ye Z., Church K. Disfluencies and fine-tuning pre-trained language models for detection of Alzheimer's disease // Proc. Interspeech 2020. – 2020. – P. 2162–2166.

12. Baevski A., Zhou H., Mohamed A., Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 12449–12460.

13. Radford A., Kim J.W., Xu T. et al. Robust speech recognition via large-scale weak supervision // Proc. 40th International Conference on Machine Learning (ICML). – 2023. – P. 28492–28518.

14. de la Fuente Garcia S., Ritchie C.W., Luz S. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: a systematic review // Journal of Alzheimer's Disease. – 2020. – Vol. 78, No. 4. – P. 1547–1574.

15. Conneau A., Khandelwal K., Goyal N. et al. Unsupervised cross-lingual representation learning at scale // Proc. 58th Annual Meeting of the ACL. – 2020. – P. 8440–8451.

16. Lundberg S.M., Lee S.-I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774.