Personalized Psychiatry and Neurology

Review

Possibilities of Analyzing the Speech of Patients with Mental Disorders Using Mathematical Methods and Artificial Intelligence Algorithms

Marina A. Barulina[1], Natalia V. Zakharova[2], Regina F. Nasyrova[2]*, Dmitriy M. Ivashinenko[3], Maria A. Dvoryankina[4] and Artyom I. Rakhmatullin[2]

[1] Perm State National Research University, 614990 Perm, Russia; marina@barulina.ru
[2] V. M. Bekhterev National Medical Research Center for Psychiatry and Neurology, 192019 Saint Petersburg, Russia; nataliza80@gmail.com, regina_nmrcpn@mail.ru, mayble13@yandex.ru
[3] N.P. Kamenev Tula Regional Clinical Psychiatric Hospital; 301105 Tula, Russia; divashinenko@mail.ru
[4] Tula State University, 300600 Tula, Russia; dvoryankinamar@yandex.ru
* Correspondence: regina_nmrcpn@mail.ru

Citation: Barulina, M.A.; Zakharova, N.V.; Nasyrova R.F.; Ivashinenko, D.M.; Dvoryankina, M.A.; Rakhmatullin, A.I. Possibilities of analyzing the speech of patients with mental disorders using mathematical methods and artificial intelligence algorithms. *Personalized Psychiatry and Neurology* **2024**, *4 (4)*: 2-10. https://doi.org/10.52667/2712-9179-2024-4-4-2-10

Chief Editor: Nikolaj G. Neznanov, D Med Sci, Professor

Received: 12 November 2024
Accepted: 2 December 2024
Published: 15 December 2024

Publisher's Note: V.M. Bekhterev NMRC PN stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Thinking is a fundamental cognitive process assessed in psychiatric practice through speech and non-verbal signals. While speech is the primary basis for evaluating thought, thinking can occur independently of speech. Disorganized thinking, particularly associated with schizophrenia, involves weakened associations and cognitive disruptions that impact neural networks. Modern definitions of disorganized thinking, rooted in the three-factor model of psychotic disorders, highlight formal disturbances in thought, speech, and behavior. This conceptualization has been substantiated through neuroimaging studies, revealing structural connectivity issues in brain regions responsible for social-emotional processing. Despite its potential, artificial intelligence (AI) has yet to be fully integrated into psychiatric diagnostics, though its success in other medical fields suggests promising applications. AI could enhance psychiatric assessments by analyzing speech, facial expressions, and behavior, offering new diagnostic tools.

Keywords: *disorganized thinking, schizophrenia, artificial intelligence, psychiatric assessment, cognitive disruption, neuroimaging.*

1. INTRODUCTION

In the clinical practice of psychiatrists and research projects studying mental processes, the thinking process is assessed through speech and non-verbal signals (appearance, behavior, or, for example, creativity). It is important to emphasize that speech cannot exist without thinking, but thinking can exist outside of speech. Speech is the basis for the clinical assessment of thinking.

Thinking is a process of mediated and generalized cognition of reality, which is usually classified by the level of generalization (object-active (concrete), visual-figurative and abstract-logical (conceptual or conceptual). In addition, several discrete processes are distinguished (affective, cognitive, and social), which, equally with linguistic mechanisms, make up the diversity of human communication [1]. This seems quite reasonable, since it is harmonized thinking that underlies the adequate development of ideas about oneself and the world, ensuring an effective understanding of what is happening and constructive communications based on these ideas [2].

Disorganization or disruption of thinking, marked by weakening of associations, has been described as a core feature of schizophrenia [3], and has been particularly actively studied since the beginning of the 21st century in the paradigm of the three-factor model of psychotic disorders [4] and the hypothesis of the relationship between individual cognitive dysfunctions and different forms of disorganization of thinking [1]. These circumstances indicate complex interactions of multiple pathological processes that may affect different neural networks. For example, delusion is an extreme distortion of such cognitive processes as reward assessment and prediction assessment, the pathomorphological substrate for which is a violation of the structural connectivity between several areas of the brain responsible for the implementation of the received social-emotional experience and the adequacy of the analysis of what is happening [5,6,7,8,9], which is confirmed by the results of neuroimaging studies [10].

An operationalized modern definition of the dimension of disorganization of thinking and speech was proposed in 1987 within the framework of the three-factor model of psychotic disorders [11]. The dimension of disorganized thinking was then identified during repeated replication on independent samples of the results of factor analysis of the indices of assessment of the patients' condition using the scale of positive and negative symptoms of schizophrenia (PANSS). In this way, the factor of formal disturbance of thinking (in the form of incoherence, tangentiality, paralogism and slippage) was identified, including impoverishment of speech, strange behavior, and inappropriate affect [12,13,14].

Thinking is a complex process that is ensured by the harmonious interaction of several hierarchically organized mental functions (including perception, memory, intellect, affective background, and many others) realized through speech.

Currently, artificial intelligence (AI) methods and algorithms help solve various applied and fundamental problems in medicine. Thus, AI is widely used to diagnose various diseases using images [15,16]. In psychiatric practice, the use of AI has not yet received worthy attention. Whereas AI capabilities make it possible to develop diagnostic software and hardware solutions based on the analysis of the patient's speech, facial expressions, posture, and so on.

The main purpose of the work is to review existing methods and algorithms of artificial intelligence for creating systems for objectifying speech disorders as a reflection of disorganization of thinking.

2. MATERIAL AND METHODS

We searched and analyzed scientific papers published on eLibrary, PubMed, Google Scholar on methods of analyzing the speech using mathematical methods and artificial intelligence algorithms for all time till the middle of year 2024. For this we used keywords: artificial intelligence, psychiatric assessment, neuroimaging, speech analysis.

3. RESULTS

3.1. Analysis of speech as a sound wave

Speech analysis is the subject of audio analysis, a scientific field that deals with the problems of capturing an audio signal (Fig. 1) as a combination of sound waves, its digital processing, determining important characteristics of the audio signal, classification, and recognizing patterns in the audio signal. The emergence of well-known voice assistants Alexa, Siri, Alice, and others became possible thanks to the development of audio analysis technologies and artificial intelligence methods and algorithms.

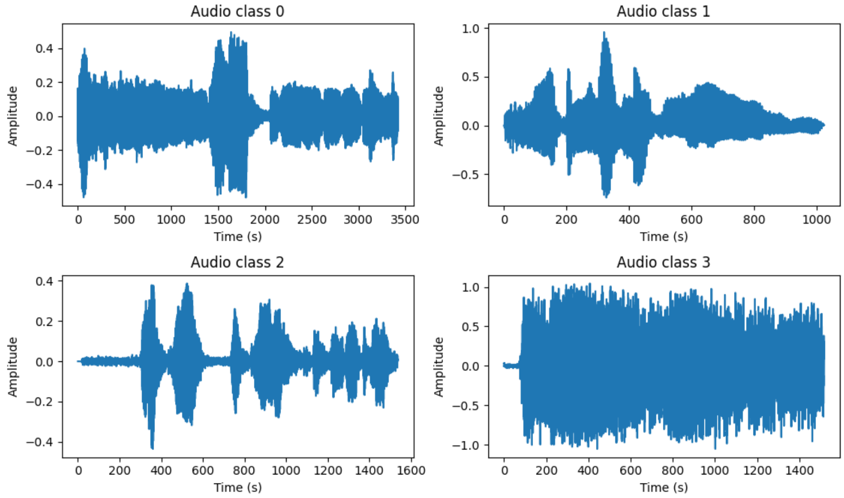
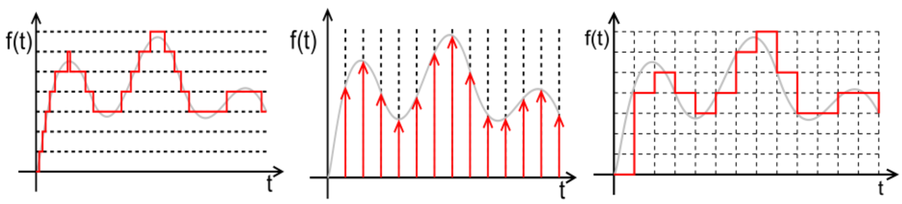


Figure 1. Examples of visualization of human speech audio recordings

In general, human speech is a combination of analog (continuous) sound waves. For speech to be analyzed using algorithmic methods, it must be digitized. For an analog signal to become digital, two operations must be applied to it - quantization and sampling.

During sampling, an analog signal is measured at a given frequency, which is called the sampling frequency. During quantization, the signal is reduced to given signal levels, which are called the sampling depth or bit depth. A quantized signal, a signal after sampling, and a digital signal are shown in Figure 2.



a b

c

Figure 2. Conversion of an analog signal into a digital one a – signal quantization, b – signal sampling, c – digital signal

It is obvious that both the quantization operation and the sampling operation occur with information loss in the audio signal. The greater the sampling depth and the smaller the sampling time interval (and, accordingly, the greater the sampling frequency), the less information we will lose during the conversion. In other words, any conversion of speech into a digital signal leads to information loss, but with a high-quality recording we can reduce these losses. And vice versa, if the recording is of inadequate quality, then information may be lost to the point of making it impossible to work with the resulting digital signal.

3.2. Preprocessing and characteristics of the audio signal

Currently, artificial intelligence (AI) algorithms can solve a wide range of problems - removing noise from a recording, highlighting human speech, and translating it into text, determining the emotional coloring of speech, highlighting semantic contexts, analyzing the conceptual sphere used by a person, and so on. But to train AI models, it is necessary to highlight the properties from the audio recording that are significant for the task being solved. For this, the following operations are performed:

- filtering the signal from noise and unwanted components;
- calculating and analyzing the spectral power of the signal;
- - determining the mel-cepstral coefficients and gamma-tone-frequency cepstral coefficients.

3.3. Filtration

Filtering allows you to cut off parasitic components of the signal. For example, if it was raining when you recorded a person's voice, the sound of the rain will make the signal we are analyzing noisy. Accordingly, we need to remove it. For this, you can use various filters - low-pass filters (removes the signal component whose frequency is below a certain frequency), high-pass filters (removes the signal component whose frequency is above a certain level), BP filters (removes the signal with frequencies within a certain interval). An example of the signal filtering result is shown in Fig. 3 and Fig. 4.

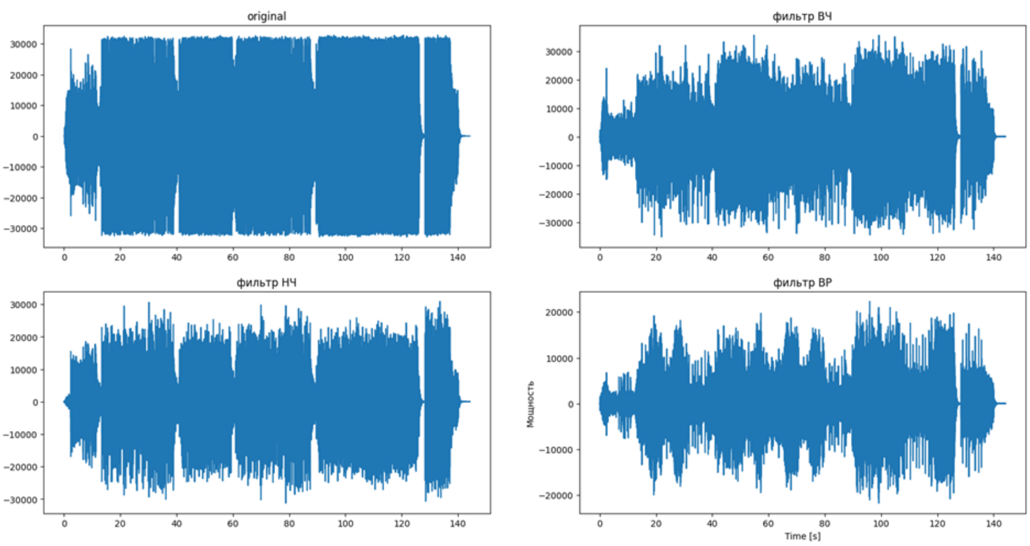


Figure 3. Example of signal filtering using several types of filters

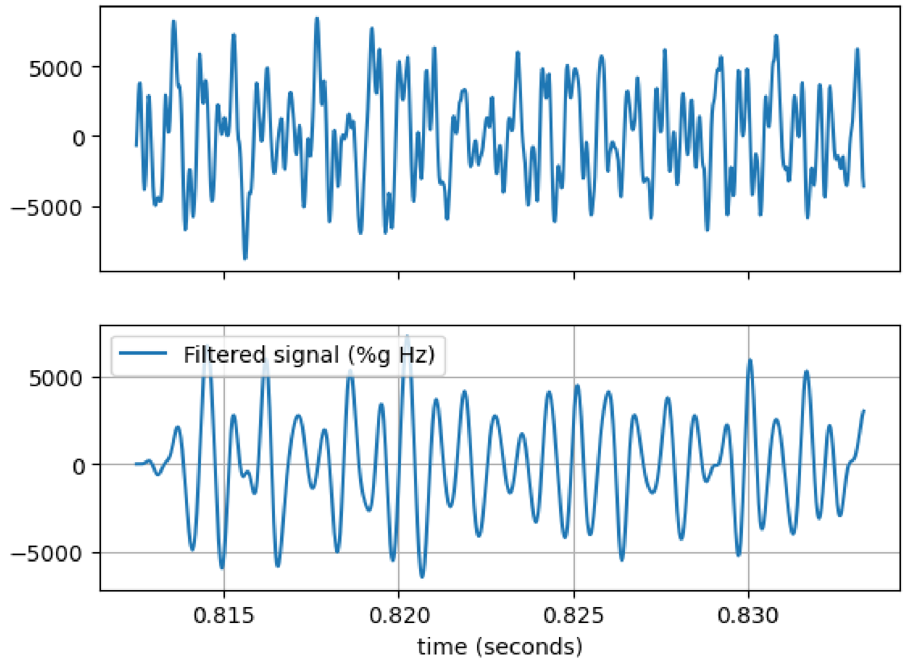


Figure 4. An example of signal filtering to remove noise. At the top is the original signal, at the bottom is after noise removal

3.4. Spectrograms

After filtering the signal, its spectral power can be analyzed. For this, you can use spectrograms (sonogram) - an image showing the dependence of the spectral density of the signal power on time. Fig. 5 shows the spectrograms of the speech of a healthy person and a person with schizophrenia.

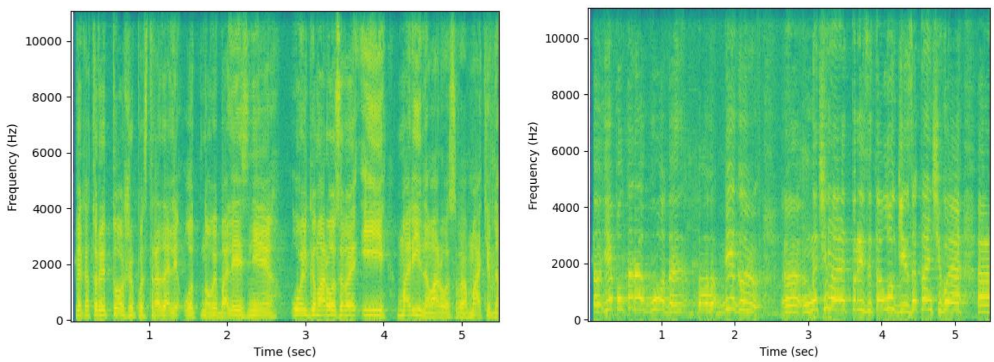


Figure 5. Speech spectrograms of a healthy person (left) and a patient with schizophrenia (right)

As can be seen in Figure 5, the spectrogram of the speech of a healthy person looks more discrete than the spectrogram of a patient. In other words, the speech of a patient with schizophrenia sounded equally intense throughout the entire speech. Whereas the healthy person changed the intensity of his speech, made pauses.

Human emotions can be classified very clearly by spectrograms (Fig. 6). That is, spectrograms in the form of images in some cases can be used to train computer vision models for classifying emotions.

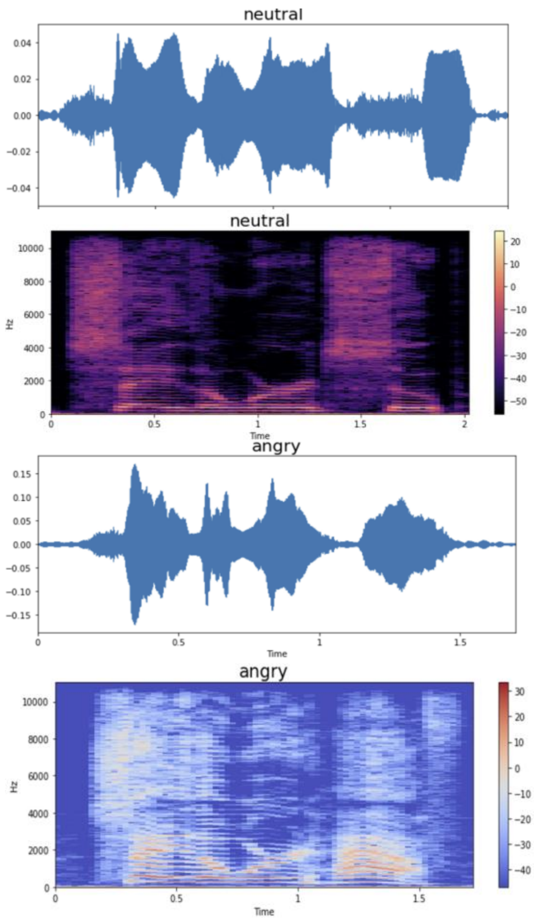


Figure 6. Recording of emotional speech and the corresponding spectrogram. Neutral speech (neutral, top), anger (angry, bottom)

3.5. Mel-cepstral and gammatone-frequency cepstral coefficients

Mel is a unit of pitch based on the perception of this sound by our hearing organs. The cepstrum is an acoustic wave (parameters of the vocal tract, tone signals, noises, and filters) [17].

Mel-cepstral and gamma-tone-frequency cepstral coefficients allow us to obtain some information from an audio recording, which can then be used to train AI models.

Examples of visualization of mel-frequency cepstral coefficients (MFCCs) for anger (class0) and sadness (class1) voices are shown in Fig. 7.

To train AI models, we take coefficients that are statistically distinguishable for the considered classes of audio recordings. In this case, any criterion can be used to calculate the statistical difference, for example, ANOVA or the Mann-Whitney criterion.

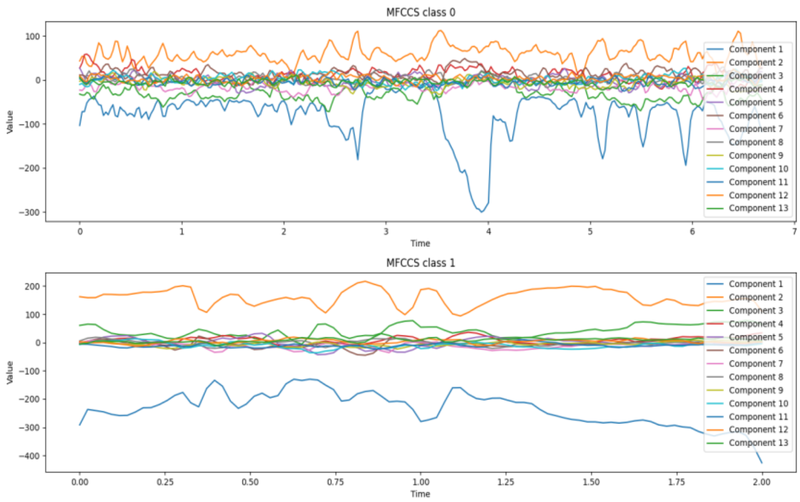


Figure 7. Visualization of mel-frequency cepstral coefficients for anger (class0) and sadness (class1)

3.6. Context of speech constructions

Using natural language processing algorithms, it is possible to analyze the meaning of speech, build the frequency of words used, which phrases are used in speech and how often. The main idea of such algorithms is that words can be represented as a numerical vector. The meaning of words can also be encoded in a numerical vector. Then we can calculate and analyze everything - the order of words, their frequency of words, their proximity, build a concept sphere, and so on.

The concept sphere is a set of concepts that are key elements of speech [18]. These concepts can be words, phrases, or terms that have a significant semantic load. The relationships between concepts reflect their joint use or thematic proximity in human speech. These relationships can be represented as a graph, where the nodes correspond to concepts, and the edges - to their relationships (Fig. 8).

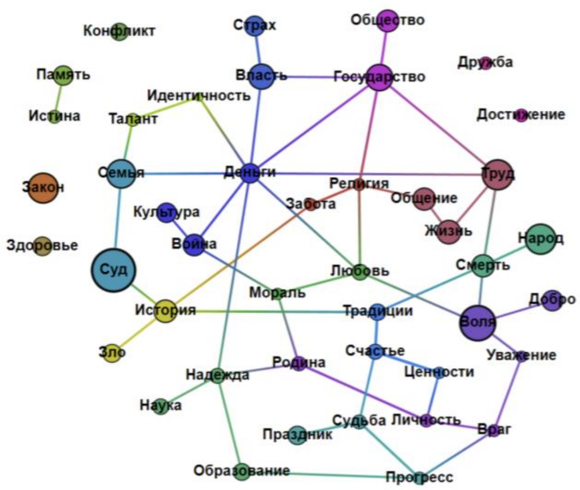


Figure 8. Example of a concept sphere

Fig. 8 shows an example of a concept sphere constructed because of human speech analysis. The relationships between concepts are determined based on their joint occurrence in the patient's speech (or texts). To determine the presence of relationships between concepts, cosine similarity values were calculated. If the cosine similarity value was greater than a certain value, then it was considered that a relationship between the concepts existed, otherwise, it did not exist.

As can be seen from Fig. 8, the Education concept is closely related to the Progress and Hope concepts but is not related to the Holiday concept. The Money concept has the most relationships. But the Law, Health, Conflict, Friendship and Achievements concepts stand apart.

4. DISCUSSION

Thus, the paper provides a brief overview of the most common mathematical methods and algorithms of artificial intelligence that can be used to analyze patient speech. At the same time, the design of clinical trials of these models must be planned to consider several nuances to reduce the risk of systematic errors. Firstly, it is necessary to consider the fact that disorganized speech is a very dynamic characteristic and can be observed in healthy people, for example, under conditions of severe emotional shock (Cohen et al. 2017). Secondly, thinking is a process; therefore, attention should be paid to the features of its course, and not to the effectiveness (that is, speech may not reflect intelligence). Thirdly, speech disorders do not a priori indicate disorders of thinking, and when assessing speech, it is necessary to personalize both the molecular biological and cultural or linguistic characteristics of each person.

5. CONCLUSIONS

Modern development of the mathematical and algorithmic apparatus of training and the use of artificial intelligence models for audio analysis makes it possible to create systems for objectifying the diagnosis of several mental illnesses. These methods can be used both within a separate study and as a stage of preprocessing audio data for their further use in training deep learning models of artificial intelligence.

Author Contributions: Conceptualization, M.A.B. and N.V.Z.; methodology, N.V.Z.; formal analysis, M.A.B.; investigation, N.V.Z., R.F.N., D.M.I AND M.A.D.; data curation, M.A.B. and N.V.Z.; writing—original draft preparation, M.A.B., N.V.Z. ANF M.A.D.; writing—review and editing, M.A.B., N.V.Z and A.I.R.; project administration, R.F.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors are grateful to the administration and management of the V. M. Bekhterev National Medical Research Centre for Psychiatry and Neurology and also thank the Director of the Centre, Scientific Head of the Geriatric Psychiatry Department of the Centre, Doctor of Medical Sciences, Professor N. G. Neznanov for the opportunity to conduct this project.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds ... are available from the authors.

REFERENCES:

1. Cohen, A.S.; Le, T.P.; Fedechko, T.L.; Elvevåg, B. Can RDoC help find order in thought disorder? *Schizophr Bull.* **2017**, 43(3):503-508. https://doi.org/10.1093/schbul/sbx030.
2. Jimeno, N.; Gomez-Pilar, J.; Poza, J.; et al. (Attenuated) hallucinations join basic symptoms in a transdiagnostic network cluster analysis. *Schizophr Res*. **2022**, 243: 43-54. https://doi.org/10.1016/j.schres.2022.02.018.
3. Bleuler, E. Dementia praecox or the group of schizophrenias. *International Universities Press* **1950**.

4. Hardy-Baylé, M.-C.; Sarfati, Y.; Passerieux, C. The cognitive basis of disorganization symptomatology in schizophrenia and its clinical correlates: toward a pathogenetic approach to disorganization. *Schizophrenia Bullet.* **2003**,*29*(3):459–471. https://doi.org/10.1093/oxfordjournals.schbul.a007019
5. Williams, L.M.; Das, P.; Harris, A.W.; et al. Dysregulation of arousal and amygdala-prefrontal systems in paranoid schizophrenia. *Am J Psychiatry* **2004**, 161(3):480-9. https://doi.org/10.1176/appi.ajp.161.3.480.
6. Phillips, M. L.; Senior, C.; Bullmore E.T.; et al. Investigation of the neural correlates of threat perception in paranoid schizophrenia: An FMRI Study. *Schizophrenia Research* **1998**. https://doi.org/10.1016/s092-9964(97)88583-0.
7. Radulescu, A.R.; Rubin, D.; Strey, H.H.; Mujica-Parodi, L.R. Power spectrum scale invariance identifies prefrontal dysregulation in paranoid schizophrenia. *Hum. Brain Mapp.* **2012**, 1582-1593. https://doi.org/10.1002/hbm.21309
8. Rădulescu, A.R.; Mujica-Parodi, L.R. A Systems approach to prefrontal-limbic dysregulation in schizophrenia. *Neuropsychobiology* **2018**. https://doi.org/10.1159/000151731.
9. Murray, J.; Williams, P.; Steiner, J. Paranoid schizophrenia: the radio loves me. *Unimaginable Storms* **2018.** https://doi.org/10.4324/9780429484599-2.
10. Zakharova, N.V.; Mamedova, G.Sh.; Bravve, L.V.; et al. Differential diagnosis of delusional symptoms in schizophrenia: Brain tractography data. *Cognitive Systems Research* **2023**.
11. Liddle, P. F. The symptoms of chronic schizophrenia. A re-examination of the positive-negative dichotomy. *The British Journal of Psychiatry: The Journal of Mental Science* **1987**.
12. Peralta, V.; Cuesta, MJ; de Leon, J. Formal thought disorder in schizophrenia: A factor analytic study. *Comprehensive Psychiatry* **1992**, 105–10.
13. Andreasen, N.C.; Arndt, S.; Miller, D.; Flaum, M.; Nopoulos, P. Correlational studies of the Scale for the Assessment of Negative Symptoms and the Scale for the Assessment of Positive Symptoms: An overview and update. *Psychopathology* **1995**, 28(1):7-17. https://doi.org/10.1159/000284894.
14. Andreasen, N.C.; Arndt, S.; Miller, D.; et al. longitudinal study of symptom dimensions in schizophrenia. Prediction and patterns of change. *Arch Gen Psychiatry* **1995**, 52: 352– 60.
15. Barulina, M.; Sanbaev, A.; Okunkov, S.; et al. Deep learning approaches to automatic chronic venous disease classification. *Mathematics* **2022**, *10*, 3571. https://doi.org/10.3390/math10193571
16. Mohammed, M.A.; Al-Khateeb, B.; Yousif, M.; et al. Novel crow swarm optimization algorithm and selection approach for optimal deep learning COVID-19 diagnostic model. *Comput Intell Neurosci*. **2022**, 1307944. https://doi.org/10.1155/2022/1307944.
17. Xuedong, H.; Acero; A; Hon, H.-W. Spoken language processing: a guide to theory, *Algorithm, and System Development* **2001**.
18. Podvigina, N.B. The concept of concept and concept sphere. *Current Issues in Modern Philology and Journalism* **2007**, 97–105.