



Contribution ID: 117

Type: Sectional reports

## Increasing the accuracy of the diagnosis of mental disorders based on heterogeneous distributed data

*Friday, 9 July 2021 11:00 (15 minutes)*

The medical field, and especially diagnosis, is still an extremely poorly formalized field. This is especially true in the study of diseases associated with changes and disorders in the activity of the brain. In order to improve the results of medical research in this area, various methods of analyzing the condition of patients are used. These include both instrumental methods (MRI, EEG) and traditional medical and psychological research methods (blood tests, interviews, psychological testing, etc.)

For each of these studies, certain conclusions are made about the patient's condition and his diagnosis. However, a qualitative combination of these conclusions can only be made by an experienced physician, who most often makes a diagnosis on the basis of his own experience and uses research data only as arguments "for" or "against". The use of mathematical methods for combining and analyzing heterogeneous data makes it possible to formalize the conclusions via those sources and increase the accuracy of the diagnosis. However, on the way of applying the exact methods of mathematics and computer science, various problems of both objective and subjective nature arise.

Firstly, all the mentioned data is in a different format, even if it is presented in digital form. In addition, they are most often distributed across various nodes and this data needs to be consolidated for general processing. Secondly, in accordance with the law on the protection of personal data, especially related to medical information, simple data consolidation is not sufficient in this case. A procedure for anonymizing data is required for further statistical processing. Moreover, this procedure itself has its own characteristics in comparison with the developed general methods of information depersonalization.

Thirdly, the use of a large number of information sources leads to an increase in dimension of data processed. This, in turn, necessitates a dramatic increase in the sample size of patients for reliable statistical analysis. In practice, it may be impossible to achieve a significant increase in volume for various reasons. Therefore, the question arises of how to use these distributed heterogeneous data of small volume in the number of patients to improve the accuracy and validity of the conclusion about the patient's condition.

The report presents the results of dividing a relatively small sample of patients into stable classes with their refinement based on additional studies.

### Summary

**Primary authors:** DEGTYAREV, Alexander (Professor); BOGDANOV, Alexander (St.Petersburg State University); Dr ZALUTSKAYA, Natalya (V.M.Bekhterev National Research Medical Center for Psychiatry and Neurology); Ms GOMZYAKOVA, Natalya (V.M.Bekhterev National Research Medical Center for Psychiatry and Neurology); Mr BELAVIN, Sergey (St.Petersburg State University); Ms KHOKHRYAKOVA, Anastasia (St.Petersburg State University)

**Presenters:** DEGTYAREV, Alexander (Professor); BOGDANOV, Alexander (St.Petersburg State University)

**Session Classification:** Data Management, Organization and Access

**Track Classification:** 6. Data Management, Organisation and Access