Forecasting socially significant epidemics based on deep learning of incomplete data

Zyatkov N.Yu.[1], Krivorotko O.I.[1], Kaminskiy G.D.[2]
[1]*Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[2]*State Health Organization Tula Regional Center for Control and Prevention of AIDS and Infectious Diseases, Tula, Russia*

Keywords: neural network, forecasting, infectious disease, mathematical modeling
TL;DR: The article proposes a deep learning algorithm for describing and predicting the dynamics of a socially significant disease in the regions of the Russian Federation with insufficient quantity and quality of statistical information.

Although tuberculosis without HIV infection is declining in Russia, our country is included in the group of countries with a high burden of tuberculosis combined with HIV infection. Annual statistics on the prevalence of tuberculosis and HIV (risk groups) in the regions of the Russian Federation have been known since 2009. These data are insufficient to develop a mathematical model for predicting the spread of infection, the development of resistant forms of the disease, and the dynamics of spread with a given accuracy.
The paper proposes to use the machine learning approach, which allows generating missing data on similar known wave processes. At the second stage, a neural network will be developed and trained on historical complete data in order to build short-term forecasts (for 3 years ahead) of the spread of tuberculosis morbidity and mortality in the Russian Federation with a given error. The regions of the Russian Federation will be classified by tuberculosis morbidity and mortality, with socio-economic sensitive indicators for each class. For each class of regions, a neural network will be adapted using the characteristics of the data.

This work is supported by the Government research assignment for Sobolev Institute of Mathematics SB RAS, project FWNF-2024-0002.