

Olena Skakalina

National University «Yuriy Kondratyuk Poltava Polytechnic», Poltava, Ukraine

CREATION AND RESEARCH OF FORECASTING MODELS OF THE SPREAD AND DEMOGRAPHIC CONSEQUENCES OF COVID-19

Abstract. The outbreak of "pneumonia of unknown etiology" marked the beginning of a new era in global health care and public life. The disease quickly spread throughout the world, turning into a pandemic that posed a serious threat to the international community. Ukraine, like many other countries, has been seriously affected by the consequences of this pandemic. Effective prediction of the spread of COVID-19 is a key task for informing health management and making strategic decisions to minimize the consequences of the pandemic. Many studies have been conducted using different methods and approaches to predict the spread of COVID-19. However, most of these methods do not adequately take into account all aspects that may affect the spread of the disease. This paper analyzes the known methods of forecasting COVID-19, evaluates the impact of various arguments on the spread of the disease, and develops a mathematical model for forecasting. For this, data on the spread of COVID-19 in Ukraine was used.

Keywords: COVID-19, GMDH, forecasting the spread, pandemic.

Introduction

The World Health Organization announced the end of the pandemic status of the coronavirus disease COVID-19. WHO Director-General Dr. Tedros Adhanom Ghebreyesus announced this on May 5, 2023.

According to the WHO, the coronavirus no longer poses such a danger and therefore cancels the global health emergency. At the same time, many countries of the world canceled the state of emergency covered by the pandemic.

The WHO considers the cancellation of the pandemic status of COVID-19 as a key moment in the evolution of humanity's relationship with the coronavirus. Meanwhile, the disease retains the status of a pandemic, like HIV [1].

Due to COVID-19, the Cabinet of Ministers of Ukraine has extended quarantine restrictions on the entire territory of Ukraine until June 30, 2023, as well as the state of emergency in the unified state system of civil protection [2]. As of early May 2023, 5.5 million cases of COVID-19 and more than 110,000 deaths from this disease have been recorded in Ukraine. The first cases of the disease in the country were recorded in March 2020.

The spread of the coronavirus in Ukraine took place in a wave-like manner. The largest wave of cases was in the fall of 2020, when the daily number of new cases exceeded 18,000. In the spring of 2021, the number of new cases decreased to 2-3 thousand per day, but then there was a second wave, and in July the number of new cases exceeded 2,000 daily.

The most infected areas are located in the western and central parts of the country, especially Lviv, Khmelnytskyi, Kyiv, Chernivtsi and Poltava regions. These regions have seen the highest incidence and highest number of deaths from COVID-19. According to WHO data, at the end of April 2023, more than 560 million cases of COVID-19 and more than 7 million deaths from complications of this disease have been confirmed in the world [5]. These numbers could be much higher because not all cases of COVID-19 have been diagnosed and reported. The global spread of the coronavirus has

significantly affected economic and social life. In many countries, this has manifested itself in a decrease in the rate of economic growth, significant delays in solving social problems, and an increase in unemployment.

Literature analysis. Approaches to solving forecasting problems can be classified into the following categories:

- method of group consideration of arguments;
- simulation modeling;
- methods of machine learning;
- epidemiological models;
- statistical methods of forecasting time series.

A comparative analysis of modeling and forecasting methods is important for choosing the best method for predicting the spread of the coronavirus. The advantages and disadvantages of the methods are presented in the Table 1.

These methods have their advantages and disadvantages, and the choice of method depends on the specific situation and task. Some methods, such as simulation modeling and machine learning, require more computing power and access to large amounts of data, while statistical methods for PD prediction may be simpler to apply but less flexible to account for complex dependencies. Epidemiological models are useful for analyzing the distribution of diseases, but may have limited accuracy, depending on the information available and the quality of the data. Each method has its own strengths and weaknesses, and their use should be considered and adapted to the specific situation and research needs.

Article [9] is a study that offers an overview of current prediction models for the disease COVID-19, as of June 2020, when more than 7 months have passed since the beginning of the pandemic.

This article examines the various approaches and methods used to predict the spread of COVID-19. The authors provide a detailed analysis of various methodologies, including statistical models, machine learning techniques, epidemiological and simulation modeling. They review the basic principles of each method and explain their advantages and disadvantages in the context of COVID-19 forecasting.

Table 1 – Advantages and disadvantages of different methods of creating predictive models

Method	Advantages	Disadvantages
GMDH	Easy to use; May contain many factors; Processes a wide range of data;	It is necessary to clearly understand the relationship between the arguments; Processing noisy data can lead to serious errors;
Simulation modeling	Can simulate various scenarios; Conducting research in conditions that are impossible in reality; Conducting research without risk to humans;	Building a model requires a lot of time and resources; May not take into account all factors affecting the system;
Methods of machine learning	Receives accurate forecasts; Processes large volumes of data; Used in various industries;	Training models requires large amounts of high-quality data; They are difficult to use for people without special training;
Epidemiological models	Can understand and predict the spread of disease; Determine optimal disease control and prevention strategies;	It is necessary to understand in detail the features of the disease and other factors that affect the spread of the disease; In some cases, they have insufficient accuracy;
Statistical methods of forecasting time series	Takes into account the dynamics of changes over time; Identifies trends and periodicity in data; Used to make predictions based on past observations	Dependencies are assumed to be stable, but in real life this is not always the case; Unusual events or significant changes in data cannot be accurately predicted;

It also focuses on various aspects of forecasting, including the assessment of parameters such as trends, rates of spread and peak values, which help to better understand the dynamics of the pandemic. The authors consider the practical application of each method in the context of health care management and decision making. They identified the following problems related to forecasting the future situation:

- observation of infected persons and those who came into contact with them;
- long incubation period;
- lack of necessary data;
- model retraining;
- excessive cleaning of data.;
- large volumes of data;
- errors in the selection of algorithms and input functions;
- complexity of the selected model.

A division of the various forecasting methods into main categories has been made, which involve the processing of large amounts of data from various sources such as WHO, National Data Sources and social media. This classification, without a doubt, can be useful for researchers, inspire them with new ideas for further forecasting.

Group method of data handling

GMDH can be used to predict the spread of the coronavirus. The method involves statistical analysis of data to search for factors affecting the spread of the virus.

In order to apply GMDH to coronavirus, it is necessary to collect data on past cases of infection, their distribution, time and place of occurrence, etc. MGUA reveals complex interrelationships between factors influencing the spread of the virus that are not easily observed. The method is also useful for predicting the risk of infection in the future and allows timely countermeasures.

The article [10] presents the results of modeling the process of the COVID-19 pandemic in Ukraine using official statistics on confirmed cases of the disease. The main goal of the article is to define a dynamic model of the process, which is expressed in the form of a disaster.

Four methods of constructing an autoregressive predictive differential model were used in the study. It is assumed that the basic forecasts are obtained by conventional autoregressive methods. The modified predictions obtained from the baseline predictions are compared using combinatorial genetic algorithms GMDH COMBI-GA and the Lasso method. The research is useful for predicting the further spread of coronaviruses in order to identify trends in time and implement specific protection measures. Forecasts using combinatorial genetic algorithms turned out to be the most meaningful.

Simulation modeling

The article [11] describes a detailed study of the spread of COVID-19 in the city of Daegu, located in South Korea. The authors used an individual-oriented simulation model to understand and predict the dynamics of the spread of the virus among the city's population.

The model is based on an individual modeling approach where each agent in the model corresponds to an individual with its unique characteristics, behavior and interactions with other agents. When developing a model of viral transmission, various factors are taken into account, such as geographic location, population density, contact network and activity level of people.

Researchers run simulations with different scenarios and control strategies, such as quarantine, social distancing and mass testing, to assess their effectiveness and impact on the spread of the virus. They also make forecasts to assess the future development of the pandemic and develop recommendations for strategic decision-making in crisis management.

The main advantage of this work is that the simulation model for individuals can be used to reproduce in detail the reality of the spread of the virus in the population. This allows researchers to analyze different scenarios and control strategies, evaluate their effectiveness and make predictions about the further development of the pandemic.

The research also has some drawbacks. They are characterized by the limited generality with which the

simulation model developed for the city of Daegu can be applied to this particular context. Due to different geographical, sociocultural and epidemiological characteristics, the results may not be applicable to other places. As with any model, there are some assumptions and uncertainties that can affect the accuracy and reliability of the results. The model is simulation-based, so it may have limited ability to predict long-term trends and complex evolutionary changes in a pandemic. Although it can be useful for short-term forecasts and for evaluating the impact of specific strategies, it should be remembered that long-term forecasts may be less accurate.

In [12], it is proposed to use a simulation model to analyze possible scenarios and strategies for controlling the spread of COVID-19 in universities.

The authors use mathematical modeling and agent simulation to reproduce real situations of the spread of the virus in the university environment, among students, teachers and other employees. Factors such as population density, frequency of contact, implementation of safety measures, testing and quarantine are also considered.

The authors use simulation modeling to analyze different control scenarios, such as social distancing, mass testing, vaccination, and the introduction of safety protocols. The impact of these strategies on the spread of the virus, the number of infected individuals, and the ability to control the pandemic in a university environment is being investigated.

The results of this study highlight the importance of combining different control strategies, such as social distancing, mass testing, vaccination, and implementation of safety protocols, to effectively manage the spread of COVID-19 in the university community.

If we talk about simulation modeling in general, we can conclude that it is an important approach for understanding the spread of COVID-19. The only drawback is the lack of statistical data. Currently, we can only artificially simulate all the parameters defined by the authors of previous works.

Methods of machine learning

In [13], machine learning methods are used to create forecasts about the number of infected, recovered and dead people up to 10 days later. The data was obtained from Hopkins University.

Research has shown that linear regression will provide better results than other methods in this case. The authors noted that improved forecasts will allow governments to take timely measures and make decisions to prevent crisis situations related to COVID-19.

The following article [14] uses data obtained from the Ministry of Health of Israel. The dataset includes the daily primary registration of all residents tested for COVID-19 across the country. The researchers developed a model to predict test results using eight binary variables: sex, age (60 years or older), presence of contact with an infected person, and severity of five clinical symptoms. Initially, the database contained 51,831 records of people who had been tested. The

model was built using decision trees with gradient boosting. The method proved effective and showed good accuracy, but required data on people's tests and symptoms, not just a single time series.

Epidemiological modeling

The work [15] used the logistic growth curve model for short-term forecasting; SIR models for predicting the maximum number of active cases and peak times; and a time-lagged regression model to estimate the impact of quarantine and other interventions.

A logistic growth curve model accurately predicts the short-term scenario for India and high-incidence states. Forecasting through the SIR model can be used to plan and prepare health care systems. The study also suggests that there is no evidence that there is a positive effect of quarantine on reducing new cases.

In the article [16] for August 2020, researcher Ihor Nesteruk analyzes the new wave of the COVID-19 pandemic in Ukraine and its dynamics. Epidemic waves are caused by a number of factors, including quarantine conditions, pathogen activity, and compliance with quarantine restrictions. In this work, the classic SIR model is used to forecast the outbreak waves: four waves are distinguished and calculations are made for the fifth and sixth waves. Taking into account the forecasting period, the final estimate of the number of cases is worse, which is quite reasonable given the beginning of the holiday season and non-compliance with the mask regime.

Statistical forecasting methods

The article [17] analyzes the growth model of the COVID-19 pandemic in India from March 4 to July 11 using regression analysis (exponential and polynomial), ARIMA, and exponential smoothing and the Holt-Winters model. The study revealed periods of geometric regression, quadratic regression and linear growth in prevalence. After analyzing each forecasting method, it was concluded that ARIMA gave the best results with an accuracy of 97.38%. The Holt-Winters model took second place with an accuracy of 97.11%. When estimating the number of patients in India, it does not make sense to compare scenarios with other countries due to different demographics. The next step was to split the series and use ARIMA to forecast for a shorter period of time, resulting in a forecast accuracy of 99.86%.

The following article [18] describes the application of a logistic model for forecasting.

A generalized logistic equation is used to interpret data on the COVID-19 epidemic in several countries: Austria, Switzerland, the Netherlands, Italy, Turkey, and South Korea. Model coefficients are calculated: growth rate and expected number of infected, as well as exponential indicators in the generalized logistic equation. The dependence of the number of infected persons on time is on average well described by a logistic curve with a coefficient of determination exceeding 0.8. At the same time, the dependence of the number of infected people per day on time has an uneven character

and is approximately described by a logistic curve. Therefore, it is necessary to take into account the dependence of the model coefficients on time or on the total number of cases. The spectra of the variability of the coefficients have characteristic peaks with a period of several days, which corresponds to the observed serial intervals. A stochastic logistic equation was used to estimate the number of probable disease peaks.

In the end, the authors note that the real variability of daily virus testing data is related to uncontrolled conditions of administrative actions and social behavior of people. The best approximation for such uncontrolled factors is approximation by random functions.

In [19] a study was conducted to forecast the situation in Pakistan using ARIMA. Three projections were made for confirmed cases, recoveries and deaths. It was concluded that the choice of methods as well as accurate data from the Ministry of Health is important for evaluation. It was also noted that there is always uncertainty due to the insufficient amount of historical data, which makes it impossible to identify specific patterns based on this data.

Object, subject and methods of research

The object of the study is the spread of COVID-19 in Ukraine. The subject of the study is the development of a methodology using GMDH for forecasting the spread of COVID-19 in Ukraine. The method of group consideration of arguments is an approach that allows taking into account various factors that influence the spread of a disease, such as socio-economic, demographic and medical indicators. Statistical models for analysis of collected data and development of a mathematical model for forecasting the spread of COVID-19 on the basis of GMDH were used as research methods.

Results

Group reasoning is used for deep learning, knowledge discovery, data prediction and analysis, optimization and pattern recognition in many fields such as economic, environmental, medical, demographic, meteorological, military and many others. GMDH's inductive algorithms automatically discover relationships in the data and select the best models and network structures to improve the accuracy of existing systems.

This unique approach to self-organization differs from the deductive method used for modeling. It is inductive and seeks the best solution from a set of possible alternatives. GMDH aims to minimize the influence of the author on the simulation results by ranking different network solutions. The computer finds the optimal model

structure and law to apply to the system. GMDH is a set of various algorithms for solving various tasks. It consists of parametric algorithms, clustering, analog binding, library and probabilistic algorithms.

This inductive approach ranks increasingly complex models and selects the optimal solution based on the minimum externality criterion. Nonlinear, stochastic functions, clustering, and polynomials are used as basic models [20].

Formation of the sample

The following factors were used to describe the factors used to predict the spread of the virus:

- Infected
- Fatalities
- Recovered

3 algorithms from the GMDH Shell platform are used to build models: combinatorial (Combi), neural network (Neuro), step-by-step mixed (Mixed). The first two are traditional GMDH algorithms. Another is a regression analysis algorithm, but it uses model generation based on the principles of GMDH.

This study used official statistics on cases of COVID-19 among patients in Ukraine for the period from March 3, 2020 to February 1, 2022. Since after the deployment of Russian troops on the territory of Ukraine, the accuracy of statistical data decreased, information about recovery became unreliable and in some cases completely absent. However, this does not indicate the disappearance of COVID-19, since the next update of statistics was carried out in the month of September, where the number of cases is not decreasing. In order to achieve greater forecast accuracy, this time period was taken into account for the analysis.

In accordance with the task, the time series of the listed indicators were constructed, and we also forecast the data of the listed indicators for the cut-off of the forecast horizon with a step of 12 combinatorial GMDH, neural network GMDH and stepwise mixed GMDH.

The results of the validation of the forecasting model based on the combinatorial algorithm of GMDH

After applying the algorithm creation parameters, the following results were obtained. Variable "Infected" (Fig. 1, Tabl. 2).

Results of the validation of the predictive model based on the GMDH neural network algorithm

After applying the algorithm creation parameters, the following results were obtained. The "Infected" variable (Fig. 2, Tabl. 3).

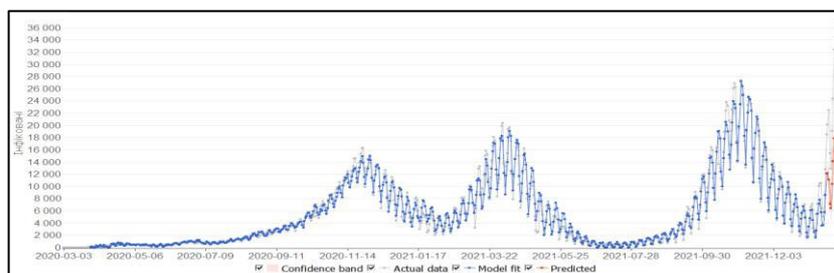


Fig. 1. Combinatorial MGUA for the variable "Infected"

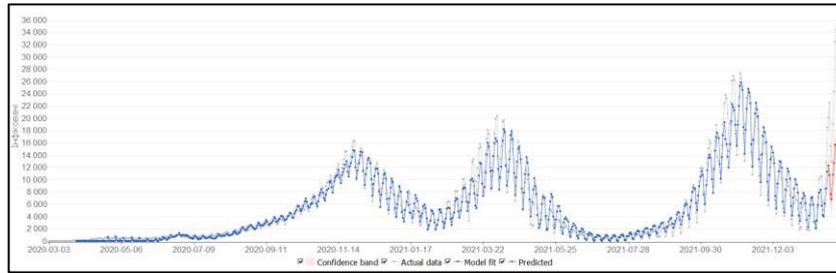


Fig. 2. Neural network MGUA for the variable "Infected"

Table 2 – Actual values "Infected" combinatorial MGUA

Results of final processing	Teaching	Exam
Number of observations	664	12
Max. negative deviation	-4,8932%	-6,4797%
Max. positive deviation	5,1722%	8,2834%
Mean Percent Error (MAPE)	2,73033%	5,4899%
Root Mean Square Deviation in Percent (RMSPE)	4,49645%	7,5847%
The sum of deviations	25,5627%	-35,988%
Standard deviation of residuals	4,49613%	4,9339%
Coefficient of determination (R2)	0,955499	0,918739
Correlation	0,978103	0,945353

Table 3 – Actual "Infected" values of neural network GMDH

Results of final processing	Teaching	Exam
Number of observations	456	12
Max. negative deviation	-26,799%	-81,404%
Max. positive deviation	22,8915%	0%
Mean Percent Error (MAPE)	5,60189%	47,339%
Root Mean Square Deviation in Percent (RMSPE)	7,42653%	50,347%
The sum of deviations	426,961%	-568,07%
Standard deviation of residuals	7,36727%	17,1282%
Coefficient of determination (R2)	0,887649	-2,56757
Correlation	0,943923	0,924417

Step-by-step mixed algorithm

After applying the algorithm creation parameters, the following results were obtained. The "Infected" variable (Fig. 3, Tabl. 4).

Comparative analysis of varieties of GMDH

To compare the obtained models, a table of indicators of the obtained descriptive statistics was created. It was created for all the factors that were investigated. In this work, the results are given for the "Infected" variable. Comparing the values in the tables, we see that the combinatorial GMDH has a smaller average absolute error, so we conclude that the combinatorial GMDH is more accurate than the others (Tabl. 5).

Table 4 – Actual "Infected" values of step-by-step mixed GMDH

Results of final processing	Teaching	Exam
Number of observations	665	12
Max. negative deviation	-24,786%	-75,205%
Max. positive deviation	19,5493%	0%
Mean Percent Error (MAPE)	2,7967%	45,345%
Root Mean Square Deviation in Percent (RMSPE)	4,38279%	47,6518%
The sum of deviations	-37,758%	-544,15%
Standard deviation of residuals	4,38243%	14,642%
Coefficient of determination (R2)	0,957717	-2,19637
Correlation	0,979502	0,946733

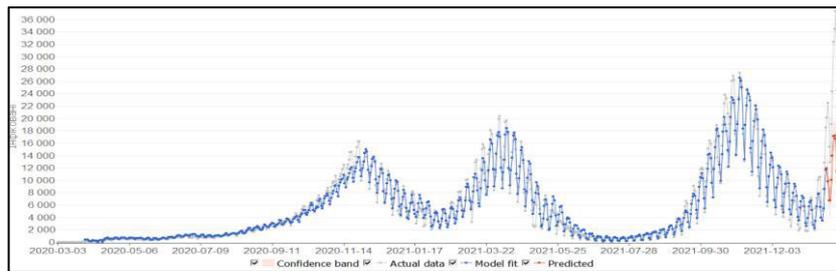


Fig. 3. Step-by-step mixed GMDH for the variable "Infected"

Table 5 – Indicators of descriptive statistics of GMDH models of the "Infected" variable

	Combinatorial		Neural network		Turn-by-turn mixed	
	Training	Foresight	Training	Foresight	Training	Foresight
Number of observations	664	12	456	12	665	12
Maximum negative error	-4,8932%	-6,4797%	-26,799%	-81,404%	-24,786%	-75,205%
Maximum positive error	5,1722%	8,2834%	22,8915%	0%	19,5493%	0%
Mean absolute percentage error(MAPE)	2,73033%	5,4899%	5,60189%	47,339%	2,7967%	45,345%
Root Mean Square Percentage Error (RMSPE)	4,49645%	7,5847%	7,42653%	50,347%	4,38279%	47,652%
The sum of deviations	25,5627%	-35,988%	426,961%	-568,07%	-37,758%	-544,15%
Standard deviation of residuals	4,49613%	4,9339%	7,36727%	17,1282%	4,38243%	14,642%
Coefficient of determination (R2)	0,955499	0,918739	0,887649	-2,56757	0,957717	-2,19637
Correlation coefficient	0,978103	0,945353	0,943923	0,924417	0,979502	0,946733

In order to verify the reliability of the obtained results, let's compare the distribution of the obtained residuals (Fig. 4). Calculation formulas for GMDH

algorithms were obtained (Fig. 5 – 7). Similar mathematical predictive models were obtained for the variables Recovered and Deaths.

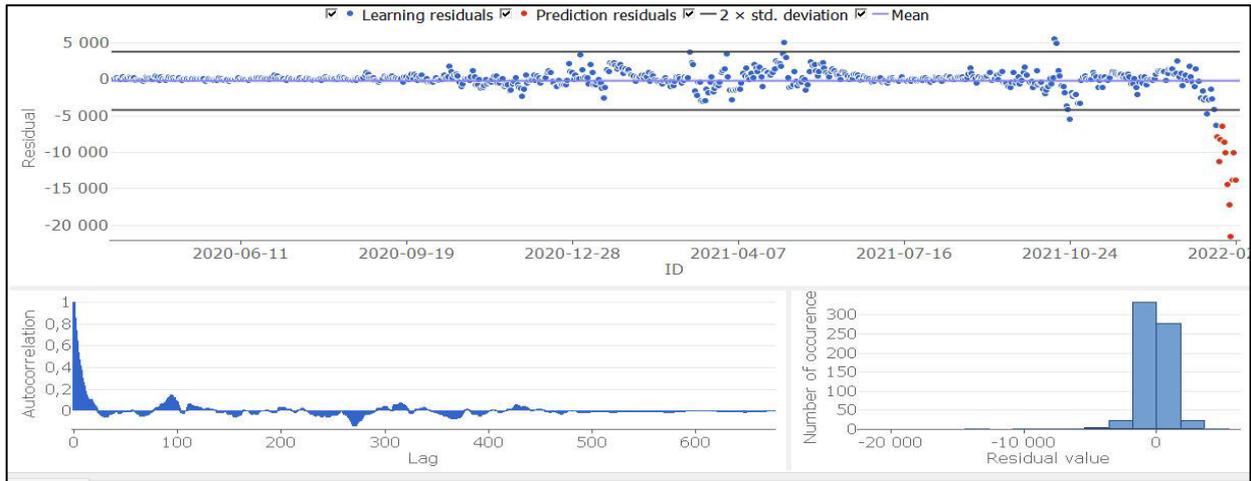


Fig. 4. Distribution of residuals of the series of the "Infected" variable of the combinatorial GMDH

$$Y1[t] = -1618.89 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-1], \text{cubert}^{**}0.0433502 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-8] * (-9.29957e-06) + \text{Інфіковані}[t-1] * \text{cycle} * 4.01134e-05 + \text{Інфіковані}[t-6] * \text{Інфіковані}[t-1], \text{cubert}^{**}(-0.00302266) + \text{Інфіковані}[t-1], \text{cubert}^{**}(-38.2424) + \text{Інфіковані}[t-8], \text{cubert}^{**} * \text{Інфіковані}[t-9], \text{cubert}^{**}13.4263 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-9], \text{cubert}^{**}(-0.0307405) + \text{Інфіковані}[t-1], \text{cubert}^{**} * \text{Інфіковані}[t-6], \text{cubert}^{**}18.5437 + \text{Інфіковані}[t-2] * \text{cycle} * 7.4818e-06 + \text{Інфіковані}[t-2], \text{cubert}^{**} * \text{Інфіковані}[t-8], \text{cubert}^{**}(-8.9343) + \text{Інфіковані}[t-9] * \text{cycle} * 3.2044e-06 + \text{cycle} * 0.26896 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-7] * (-3.12521e-05) + \text{Інфіковані}[t-6] * \text{Інфіковані}[t-11] * 2.70739e-06 + \text{Інфіковані}[t-2], \text{cubert}^{**} * \text{Інфіковані}[t-3], \text{cubert}^{**}6.26657 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-8], \text{cubert}^{**}(-0.0393959) + \text{Інфіковані}[t-9] * \text{Інфіковані}[t-3], \text{cubert}^{**}(-0.039301) + \text{Інфіковані}[t-3] * \text{Інфіковані}[t-3], \text{cubert}^{**}0.00569815 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-11], \text{cubert}^{**}0.050938 + \text{Інфіковані}[t-2], \text{cubert}^{**} * \text{Інфіковані}[t-10], \text{cubert}^{**}12.1116 + \text{Інфіковані}[t-9] * \text{Інфіковані}[t-6], \text{cubert}^{**}0.00676234 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-9] * 3.80441e-05 + \text{Інфіковані}[t-7] * \text{cycle} * (-1.55226e-05) + \text{Інфіковані}[t-11] * \text{Інфіковані}[t-11], \text{cubert}^{**}(-0.0630896) + \text{Інфіковані}[t-11] * \text{Інфіковані}[t-10], \text{cubert}^{**}0.0392643 + \text{Інфіковані}[t-10] * \text{Інфіковані}[t-6], \text{cubert}^{**}(-0.0102642) + \text{Інфіковані}[t-2], \text{cubert}^{**} * \text{cycle} * (-0.0221279) + \text{Інфіковані}[t-2] * \text{Інфіковані}[t-11], \text{cubert}^{**}(-0.021685) + \text{Інфіковані}[t-8], \text{cubert}^{**} * \text{Інфіковані}[t-10], \text{cubert}^{**}(-18.9774) + \text{Інфіковані}[t-8] * \text{Інфіковані}[t-8], \text{cubert}^{**}0.00861705 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-11], \text{cubert}^{**}(-0.0406456) + \text{Інфіковані}[t-11] * \text{Інфіковані}[t-2], \text{cubert}^{**}0.0507985 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-8], \text{cubert}^{**}0.0453702 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-8] * (-1.16938e-05)$$

Fig. 5. The mathematical model of the time series was obtained using the combinatorial GMDH for the variable "Infected"

ID моделі: +1 #1
 Складність моделі: 30 з 326 Значення критерію: 0

$$Y1[t] = 128.663 + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-8] * (-2.37025e-05) + \text{Інфіковані}[t-8] * \text{cycle} * 1.57124e-05 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-5] * (-1.71477e-05) + \text{Інфіковані}[t-7] * \text{Інфіковані}[t-11] * 4.55765e-05 + \text{Інфіковані}[t-11] * \text{Інфіковані}[t-12] * (-4.4346e-05) + \text{Інфіковані}[t-10], \text{cubert}^{**} * \text{cycle} * (-0.0124289) + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-1], \text{cubert}^{**}0.0881802 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-5], \text{cubert}^{**}(-0.122514) + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-4] * (-1.25223e-05) + \text{Інфіковані}[t-5] * \text{Інфіковані}[t-8], \text{cubert}^{**}(-0.037494) + \text{Інфіковані}[t-5] * \text{Інфіковані}[t-1], \text{cubert}^{**}0.138385 + \text{Інфіковані}[t-10] * \text{Інфіковані}[t-12] * 4.2451e-05 + \text{Інфіковані}[t-10] * \text{Інфіковані}[t-7], \text{cubert}^{**}(-0.0370063) + \text{Інфіковані}[t-6], \text{cubert}^{**} * \text{Інфіковані}[t-10], \text{cubert}^{**}16.5001 + \text{Інфіковані}[t-5] * \text{Інфіковані}[t-11], \text{cubert}^{**}(-0.0460126) + \text{Інфіковані}[t-8] * \text{Інфіковані}[t-12] * (-7.5576e-06) + \text{Інфіковані}[t-5], \text{cubert}^{**} * \text{Інфіковані}[t-9], \text{cubert}^{**}(-19.6297) + \text{Інфіковані}[t-11] * \text{Інфіковані}[t-7], \text{cubert}^{**}0.0271007 + \text{Інфіковані}[t-9], \text{cubert}^{**} * \text{cycle} * 0.00654018 + \text{Інфіковані}[t-2] * \text{Інфіковані}[t-5] * (-7.09562e-06) + \text{Інфіковані}[t-8], \text{cubert}^{**} * \text{Інфіковані}[t-11], \text{cubert}^{**}(-9.29727) + \text{Інфіковані}[t-8] * \text{Інфіковані}[t-6], \text{cubert}^{**}0.0129638 + \text{Інфіковані}[t-6] * \text{Інфіковані}[t-10], \text{cubert}^{**}(-0.0290325) + \text{Інфіковані}[t-6] * \text{Інфіковані}[t-9], \text{cubert}^{**}0.0250301 + \text{Інфіковані}[t-5], \text{cubert}^{**} * \text{Інфіковані}[t-10], \text{cubert}^{**}14.4743 + \text{Інфіковані}[t-1], \text{cubert}^{**} * \text{Інфіковані}[t-8], \text{cubert}^{**}3.76744 + \text{Інфіковані}[t-5] * \text{Інфіковані}[t-10], \text{cubert}^{**}0.0134076 + \text{Інфіковані}[t-1], \text{cubert}^{**} * \text{cycle} * 0.00441033 + \text{Інфіковані}[t-1] * \text{Інфіковані}[t-3] * 1.64818e-06$$

Fig. 6. Mathematical model of TS obtained using a step-by-step mixed GMDH for the variable "Infected"

$$Y1[t] = 7.47396 - N7 * 0.0770928 + N2 * 1.07613$$

$$N2[t] = 17.2623 - N5 * 0.707717 + N3 * 1.70549$$

$$N3[t] = 119.582 - \text{Видужали}[t-12] * 0.0473431 + N4 * 1.03226$$

$$N4[t] = -95.8494 + \text{Видужали}[t-6] * 0.0891133 + N5 * 0.922716$$

$$N5[t] = -104.351 - \text{Видужали}[t-7] * 0.373977 + N6 * 1.38976$$

$$N6[t] = -76.1713 + |id, \text{dayofweek}_4 * 626.253 + N7 * 0.998175$$

$$N7[t] = 78.3491 - \text{Видужали}[t-8] * 0.368717 + N8 * 1.36057$$

$$N8[t] = 203.121 + \text{Видужали}[t-1] * 0.338078 + \text{Видужали}[t-7] * 0.631654$$

Fig. 7. Mathematical model of TS obtained with the help of neural network GMDH for the variable "Recovered"

Conclusions

After reviewing the obtained mathematical models with the help of the combinatorial algorithm of GMDH,

the neural network algorithm of MGUA, and the step-by-step mixed algorithm of GMDH, it can be firmly stated that neural network models are more complex than combinatorial ones, and step-by-step mixed models, in turn, are more complex than neural network models. Since the results demonstrated by neural network models have a greater deviation from the average, and therefore are of lower quality for predictions, from the point of view of practical implementation, they are much more expensive. As a result, it can be concluded that the GMDH combinatorial algorithm is more suitable for building a TS based on this problem.

Since the official data on the development of the coronavirus infection after February 24, 2022 are inaccurate, nevertheless, they occasionally coincide with the predicted data of the GMDH combinatorial algorithm.

Which indicates the accuracy and reliability of this algorithm.

The study carried out forecasting of the spread of COVID-19 in Ukraine using GMDH. The following tasks were performed:

➤ a review of the subject activity was carried out, including the history of the emergence of COVID-19 and comparison with other pandemics. In particular, the importance of research and analysis of the COVID-19 pandemic was considered.

➤ a comparative analysis of various modeling and forecasting methods was carried out, including GMDH, simulation modeling, machine learning methods, epidemiological models and statistical methods of TS forecasting.

This study shows that GMDH is an effective method for predicting the spread of the COVID-19 pandemic. GMDH can be used to make accurate forecasts and reasonable recommendations for further management of the situation.

Thus, the results of this study demonstrate the importance of GMDH as a forecasting and management tool in the context of the COVID-19 pandemic, which can be used for further research and decision-making in the field of public health and crisis management.

REFERENCES

1. Yaroslava, Pradyd. (2023, 15 november). *WHO canceled the global pandemic status of the coronavirus*. https://zaxid.net/vooz_skasuvata_status_pandemiyi_cherez_covid_19_n1563000.
2. Ukrinform. (2023, december). *WHO canceled the status of the pandemic of COVID-19*: <https://www.ukrinform.ua/rubric-society/3705279-vooz-skasuvata-status-pandemii-covid19.html>.
3. Corona virus - *Statistics by countries*. (2023, november). <https://index.minfin.com.ua/ua/reference/coronavirus/geography.html>.
4. G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, J. Chaki, and A. E. Hassanien, *Forecasting Models for Coronavirus Disease (COVID-19)*. (2020). A Survey of the State-of-the-Art, *SN Computer Science*, vol. 1, no. 4. <https://doi:10.1007/s42979-020-00209-9>.
5. Shakhovska N., Medykovsky M.O. (eds.). (2021). *GMDH-Based Discovering Dynamic Regularities of the Ukraine Covid-19 Pandemic Process, in Advances in Intelligent Systems and Computing V*, Cham, 456–470. https://doi:10.1007/978-3-030-63270-0_30.
6. W.-S. Son and RISEWIDs Team. (2020). *Individual-based simulation model for COVID-19 transmission in Daegu, Korea*, *Epidemiol Health*, vol. 42, p. e2020042. <https://doi:10.4178/epi2020042>.
7. N. Ghaffarzadegan. (2021). *Simulation-based what-if analysis for controlling the spread of Covid-19 in universities*, *PLOS ONE*, vol. 16, no. 2, p. e0246323. <https://doi:10.1371/journal.pone.0246323>.
8. R. K. Mojada, A. Yadav, A. V. Prabhu, and Y. Natarajan. (2020). *Machine learning models for covid-19 future forecasting*, *Mater Today Proc*. <https://doi:10.1016/j.matpr.2020.10.962>.
9. Y. Zoabi, S. Deri-Rozov, and N. Shomron. (2021). *Machine learning-based prediction of COVID-19 diagnosis based on symptoms*, *npj Digital Medicine*, vol. 4, no. 1, Art. no. 1. <https://doi:10.1038/s41746-020-00372-6>.
10. B. Malavika, S. Marimuthu, M. Joy, A. Nadaraj, E. S. Asirvatham, and L. Jeyaseelan. (2021). *Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models*, *Clinical Epidemiology and Global Health*, vol. 9, 26–33. <https://doi:10.1016/j.cegh.2020.06.006>.
11. I. Nesterukю (2020). *New waves of the COVID-19 pandemic in Ukraine* [“New waves of COVID-19 pandemic in Ukraine”].
12. V. K. Sharma and U. Nigam. (2020). *Modeling and Forecasting of COVID-19 Growth Curve in India*, *Transactions of the Indian National Academy of Engineering*, 1–14. <https://doi:10.1007/s41403-020-00165-z>.
13. Logistic equation and COVID-19 (2023, december). *PubMed Central (PMC)*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7444956>.
14. M. Yousaf, S. Zahir, M. Riaz, S. M. Hussain, and K. Shah. (2020). *Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan*, *Chaos, Solitons & Fractals*, vol. 138. <https://doi:10.1016/j.chaos.2020.109926>.

Received (Надійшла) 17.09.2024

Accepted for publication (Прийнята до друку) 13.11.2024

Створення та дослідження моделей прогнозування поширення та демографічних наслідків COVID-19

О. В. Скакаліна

Анотація. Спалах «пневмонії невідомої етіології» ознаменував початок нової ери у світовій охороні здоров'я та суспільному житті. Хвороба швидко поширилася по всьому світу, перетворившись на пандемію, яка становила серйозну загрозу для міжнародної спільноти. Україна, як і багато інших країн, серйозно постраждала від наслідків цієї пандемії. Ефективне прогнозування поширення COVID-19 є ключовим завданням для інформування керівництва охорони здоров'я та прийняття стратегічних рішень щодо мінімізації наслідків пандемії. Було проведено багато досліджень з використанням різних методів і підходів для прогнозування поширення COVID-19. Однак більшість із цих методів не враховують належним чином усі аспекти, які можуть вплинути на поширення захворювання. У цій роботі аналізуються відомі методи прогнозування COVID-19, оцінюється вплив різних аргументів на поширення хвороби та розробляється математична модель для прогнозування. Для цього були використані дані про поширення COVID-19 в Україні.

Ключові слова: COVID-19, GMDH, прогнозування поширення, пандемія.