IEEE
Advancing Technology for Humanity

IEEE UKRAINE SECTION

IEEE Central Ukraine chapter

IEEE Kharkiv Joint Chapter

HUAWEI

NOKIA

lifecell

2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics

November 29 – December 3, 2021, Kyiv, Ukraine

IEEE UkrMiCo'2021

Conference Proceedings

Table of Contents

Part Number: CFP21J07-USB
ISBN: 978-1-6654-2651-0

Forecasting of COVID-19 Epidemic Process by Lasso Regression

Dmytro Chumachenko
Mathematical Modelling and Artificial Intelligence department
National Aerospace University "Kharkiv Aviation Institute"
Kharkiv, Ukraine
dichumachenko@gmail.com

Tetyana Chumachenko
Epidemiology department
Kharkiv National Medical University
Kharkiv, Ukraine
tatalchum@gmail.com

Ievgen Meniailov
Mathematical Modelling and Artificial Intelligence department
National Aerospace University "Kharkiv Aviation Institute"
Kharkiv, Ukraine
evgenii.menyailov@gmail.com

Olena Muradyan
Sociology faculty
V.N. Karazin Kharkiv National University
Kharkiv, Ukraine
o.s.muradyan@karazin.ua

Grigoriy Zholtkevych
Mathematics and Informatics faculty
V.N. Karazin Kharkiv National University
Kharkiv, Ukraine
g.zholtkevych@karazin.ua

***Abstract*— The coronavirus epidemic has changed the life of the whole world. Containment of the further development of the pandemic requires the implementation of effective evidence-based control measures. For this, it is advisable to use mathematical modeling. The most accurate predictions are shown by machine learning methods. The article discusses a lasso regression model for predicting the dynamics of a new coronavirus in Ukraine, Great Britain, Germany and Japan. The model shows high accuracy. The disadvantage of this approach is the impossibility of identifying the factors influencing the dynamics of morbidity.**

Keywords—epidemic model, COVID-19, machine learning, lasso regression, infectious diseases forecasting.

I. Introduction

The pandemic of the new coronavirus began in December 2019 in China, and in March 2020, WHO declared it global. To date, more than 250 million people have fallen ill worldwide, and more than 5 million cases have been fatal.

Despite an active campaign to vaccinate the population around the world, in most countries the rate of immunization is insufficient to end the pandemic [1]. Autumn 2021 is characterized by new outbreaks, resulting in new restrictions, lockdowns and border closings.

As vaccination against Covid-19 is gaining momentum in developed countries, the question becomes more and more relevant - what next? When all the borders will be open up? How much can the emergence of new variants of the coronavirus hinder this? Is it far from herd immunity? Will the world be able to return to normal life?

Mathematical modeling can provide answers to these questions. Scientists have been working on approaches to modeling morbidity for more than a hundred years. But the global challenges in the wake of the COVID-19 pandemic have motivated new research teams to take a look at this object. Most models of the COVID-19 epidemic process into several groups:

- Models of other processes extended to epidemiology and public health [2-5]. The disadvantage of such models is that they do not take into account the specifics of epidemic processes, such as transmission routes, population heterogeneity, seasonality, etc.
- Classical compartment models, the states of which have been extended to the new coronavirus [6-9]. The disadvantage of this approach is the high complexity of the models. To make changes caused by new strains and changing virulence of infection, it is necessary to completely rebuild the entire system of differential equations. This leads to a decrease in the accuracy of the forecasts obtained.
- Intelligent Models [10-13]. Such models usually take into account the intellectual behavior of the population and are characterized by high accuracy.

The **aim** of given research is to develop machine learning model of COVID-19 propagation based on random forest method.

Proposed model is a part of complex intelligent information system of epidemiological diagnostics whish concept is described in [14].

II. Analysis of COVID-19 Epidemic Process

To verify COVID-19 epidemic process model we have selected four countries: Ukraine, Great Britain, Germany and Japan. Chosen countries have different rules of COVID-19 propagation, different control and anti-epidemic measures, what influencing the dynamics of epidemic process.

In Ukraine, the coronavirus infection Covid-19 (new type of pneumonia) was first diagnosed on March 3, 2020 in Chernivtsi. On March 13, the first death was recorded as a result of a coronavirus infection. As of October 10, 2021, there were 2,922,302 infected in Ukraine, of which 67729 died. There is no exact data on those infected with COVID-19 in the occupied territories. In the occupied Crimea, according to Johns Hopkins University, as of 09/21/2021, there were

The study was funded by the Ministry of Education and Science of Ukraine in the framework of the research project 0121U109814 on the topic "Sociological and mathematical modeling of the effectiveness of managing social and epidemic processes to ensure the national security of Ukraine".

978-1-6654-2651-0/21/$31.00 ©2021 IEEE

105,610 cases (3573 of them died). There is no reliable information on the occupied regions of Donbass. [15]. On March 12, 2020, the government introduced the first quarantine measures, closing educational institutions. On March 17, 2020, stricter measures were introduced, in particular restrictions on mass gatherings, the suspension of public transport and the closure of "unimportant" businesses, namely restaurants, hotels, beauty salons and non-food stores, as well as the postponement of planned inpatient treatment for people who do not have COVID 19. Crossing the state border, the contact line in the east, and the AG with Crimea have also been restricted, including restrictions by the self-proclaimed "republics" and the Russian Federation. Starting from April 4, 2020, the Government of Ukraine, guided by international practice, introduced restrictions on stay in public places and mandatory self-isolation for people over 60 years old, with some exceptions for working people. At the end of May 2020, nationwide quarantine restrictions were eased. From May 22 to November 12, 2020, the Government introduced an "adaptive quarantine" 7, under which restrictions were applied selectively in different geographic areas depending on the incidence of COVID19 [16]. In June 2020, the World Bank extended a US $ 350 million loan to Ukraine as part of the Modernization of the Social Support System for the Population of Ukraine project, and the International Monetary Fund (IMF) approved an 18-month Reserve Agreement totaling about US $ 5 billion USA. Vaccination against coronavirus in Ukraine is voluntary; they do not plan to make it mandatory. At the same time, experts only urge Ukrainians to be vaccinated, because the more citizens are immunized, the less the need for quarantine will be. In the fall of 2021, access to vaccines is free; nevertheless, Ukraine ranks last place in Europe in terms of the dynamics speed of vaccinated population.

The United Kingdom of Great Britain does not have a unified health care system. There is the National Health Service (NHS) and the health service for Wales, Scotland and Northern Ireland. NHS England is one of the most efficient and powerful healthcare systems in the world. About 90% of people fully support the principles and activities of NHS England. The service has the trust of society like no other institution in the country [17]. Following the first cases of the disease in the UK on 31 January, a public health awareness campaign was launched to advise people on how to reduce the risk of the virus spreading. Public Health England (PHE) has published guidance on infection prevention and control, COVID-19 detection and diagnosis, and daily updates including travel advice. In addition, NHS has opened COVID-19 testing centers in some hospitals. The chief medical adviser to the UK government, Chris Whitty, explained four aspects of the outbreak strategy: containment, containment, research and mitigation [18]. As of October 30, 2021, 9,057,629 cases of Covid-19 coronavirus infection were recorded in the UK. The total number of deaths from coronavirus infection in the UK is 140,632, with a case fatality rate of 1.55%.

In Germany, each region makes the decision to ease quarantine measures independently. In particular, in the state of Baden-Württemberg, all public events were allowed with an attendance of up to 100 people, and in Thuringia, the ban on social contacts was lifted. In Berlin, bars and restaurants will be open after 23:00. Theaters, concert halls and cinemas will open in Bremen from 12 June. It is reported that in connection with the opening of borders within the EU, the German railway concern Deutsche Bahn resumes the movement of passenger trains to other European countries, in particular Switzerland, Austria, Italy, Czech Republic, France, Belgium, the Netherlands, Denmark and Poland. At the same time, it is reported that 247 new infections of SARS-CoV-2 were detected in the country on June 13, the total number of infections registered since the beginning of the epidemic was more than 186 thousand [19]. The highest mortality from covid is observed among the elderly, who have a weakened immune system, as well as chronic and age-related diseases. For this reason, the risk of developing severe covid disease in older people is much higher than in young people. 87 percent of people who have died from coronavirus were aged 70 or older. And the average age of the deceased is 83 years [20]. In Germany, amid rising COVID-19 deaths among those vaccinated in nursing homes, the debate about mandatory vaccination for certain occupations has intensified. As of October 30, 2021, 4,607,958 cases of Covid-19 coronavirus infection were recorded in Germany. The total number of deaths from coronavirus infection in Germany is 96,259.

Japan became one of the first countries after China to face COVID-19, and under rather unusual circumstances. In February 2020, the Diamond Princess cruise ship with 3,711 passengers arrived in Yokohama and was forced to quarantine due to an outbreak of coronavirus on board: a total of 712 people fell ill, 14 of them died [21]. The Japanese authorities saw this as the prototype of their own country in the midst of an epidemic. And the Japanese strategy was seriously different from the strategies of other developed countries. Japanese virologists decided that there was no point in trying to identify and isolate all infected people - in a country with such a population density, this would be almost impossible. In addition, it would require a huge number of tests, which also would not make much sense - the bulk of those infected in Japan were sick either asymptomatically or with very mild symptoms. Therefore, the number of tests for COVID-19 in Japan was several times less than in other developed countries included in the G7: for example, in March 2020, an average of about 100 people per day per million population was tested in Japan, in December 2020 - about 270 [22]. As of October 30, 2021, 1,722,610 cases of Covid-19 coronavirus infection were recorded in Japan. The total number of deaths from coronavirus infection in Japan is 18,261.

When analyzing the data, it should be borne in mind that approaches to registering cases in each country may differ. Therefore, direct comparison of indicators is incorrect. The projected increase in the incidence in the studied countries may be associated with a change in approaches to wearing masks, an increase in the number of daily flights compared to 2020. It should also be taken into account that the number of cases of infection of children has increased, which are not yet subject to vaccination; infection of children is not always accompanied by a pronounced manifestation of the disease. Because Since the number of sources of infection among children has increased, they have begun to play an active role in the spread of the virus. You should also pay attention to the fact that no country has achieved such a level of collective (population) immunity, which would affect the decrease in the virulence of the circulating variant of the virus. The circulation of the highly virulent variant is very easy, a very small dose of the pathogen causes the disease, and an increase in the infectious dose (for example, with prolonged contact with an infected person) leads to more severe cases of the disease and more massive excretion of the pathogen by these

cases, the vaccinated can also become infected and become additional sources infections.

III. Lasso Regression Model

Regression is a model of the dependence of the variable x on one or more other variables (factors, regressors, independent variables) with a dependence function. Linear regression refers to the problem of determining the "line of best fit" across a set of data points and has become a simple precursor to non-linear methods that are used to train neural networks. The simple linear regression formula is:

$$\hat{y} = B_0 + B_1 x + e \quad (1)$$

where $\hat{y}$ is forecasted value, B_0 is the slope, B_1 is the value of x and they are called regression coefficients, and e is an error term that reconciles variances among actual values and forecasted values and corresponds to $y - \hat{y}$ [23].

Lasso regression (LASSO, Least Absolute Shrinkage and Selection Operator) is a variation of linear regression specially adapted for data that exhibits strong multicollinearity (that is, strong correlation of features with each other) [24].

It automates parts of the model selection, such as variable selection or parameter exclusion. Lasso uses shrinkage, which is a process in which data values approach a center point (for example, an average).

The compression process adds several advantages to regression models:

- More accurate and stable estimates of true parameters.
- Reducing sampling errors and out-of-sampling.
- Smoothing of spatial fluctuations.

Lasso can be understood as adding a regular L_1 term based on linear regression. It also limits W so it doesn't get too big. Where $\lambda > 0$, by determining the λ value, the model can be balanced between bias and variance. As λ increases, the variance of the model decreases and the deviation increases.

$$l = \frac{1}{2m}\sum_{i=1}^{m}(y^i - \sum_j w_j x_j^i)^2 + \lambda \sum_j |w_j| \quad (1)$$

Rather than adjusting the complexity of the model to compensate for the complexity of the data, like high variance neural network and decision tree regression methods, the Lasso attempts to reduce the complexity of the data so that it can be handled by simple regression methods by curving the space it lies on. In this process, the Lasso automatically helps to eliminate or distort highly correlated and redundant features in a low variance method.

Lasso regression uses L_1 regularization, that is, weights the errors by their absolute value. Instead of, for example, L_2 regularization, which weights errors by their squared, in order to punish more significant errors more strongly. This regularization often leads to sparser models with fewer coefficients, as some coefficients may become zero and therefore be excluded from the model. This allows it to be interpreted.

IV. Results

We used data on incidence, new cases, deaths and people recovered from COVID-19 distributed by countries. Center of Public Health of MOH Ukraine provided Ukrainian data. Data for other countries was parsed from Coronavirus Resource Center of John Hopkins University and European Center for Disease Control. Data parser was realized as web API by .Net Core 3.1. Two indicators of accuracy were calculated for Lasso regression epidemic model: mean absolute error (MAE) and mean square error (MRE). Figure 1 presents forecast for Ukraine. MAE is 6908.40. MSE is 86410005.5

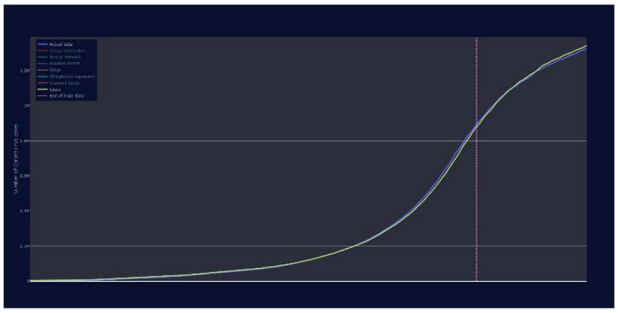

Fig. 1. Forecating of COVID-19 in Ukraine using Lasso regression

Figure 2 shows results of simulation for Germany. MAE is 10612.18. MSE is 223879052.28

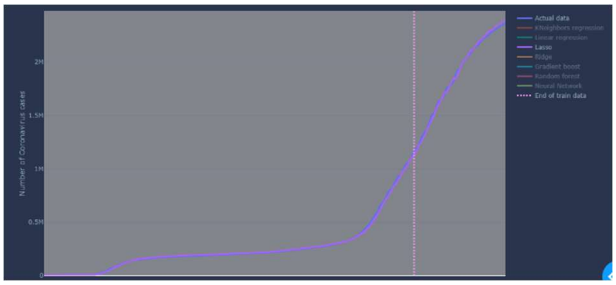

Fig. 2. Forecating of COVID-19 in Germany using Lasso regression

Figure 3 presents forecasted results for Great Britain. MAE is 14112.16. MSE is 465335742.19

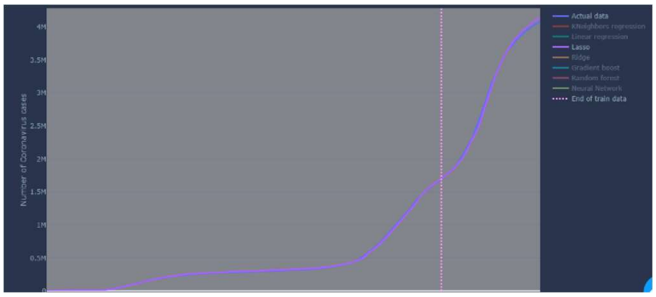

Fig. 3. Forecating of COVID-19 in Great Britain using Lasso regression

Figure 4 shows model results for Japan data. MAE is 1345.84. MSE is 5455667.64.

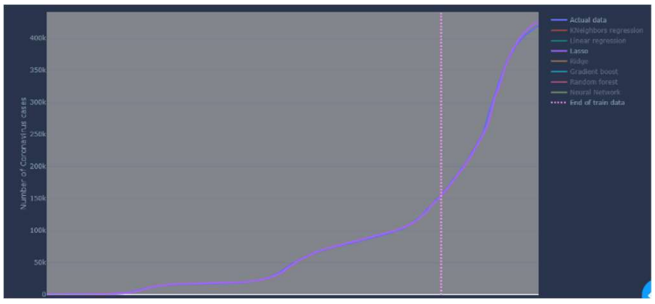

Fig. 4. Forecating of COVID-19 in Japan using Lasso regression

Values of errors are shown in Table 1

TABLE I. FORECASTING ERRORS

Country	*MAE*	*MSE*
Ukraine	6908.40	86410005.5
Germany	10612.18	223879052.28
Great Britain	14112.16	465335742.19
Japan	1345.84	5455667.64

V. CONCLUSIONS

The constructed epidemic model of the spread of COVID-19 makes it possible to calculate the predicted incidence. The model shows more accurate results when forecasting for 10 days or less. When simulating epidemic processes, such a time period is acceptable, since Based on the forecast, decision-makers must implement urgent containment measures that affect the dynamics of the epidemic. At the same time, the forecast for a longer period can be used when planning material resources and their redistribution in certain territories.

The constructed model of the epidemic process can be used not only in the study of COVID-19. Also an important task is emergent infections, which are characterized by the fact that the rules of the spread are not well understood. In this case, machine learning methods are the most effective, because the use of classical compartment models and other approaches is difficult.

ACKNOWLEDGMENT

The study was funded by the Ministry of Education and Science of Ukraine in the framework of the research project 0121U109814 on the topic “Sociological and mathematical modeling of the effectiveness of managing social and epidemic processes to ensure the national security of Ukraine” [25].

REFERENCES

[1] J. Sultana, et. al., “Potential effects of vaccinations on the prevention of COVID-19: rationale, clinical evidence, risks, and public health considerations,” Expert Review of Vaccines, vol. 19, iss. 10, 2020, pp. 919-936. doi: 10.1080/14760584.2020.1825951

[2] M. Ala'raj, M. Majdalawieh, N. Nizamuddin, “Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections,” Infectious Diseases Modelling, vol. 6, 2021, pp. 98-111. doi: 10.1016/j.idm.2020.11.007

[3] S. Hendy, et. al., “Mathematical modelling to inform New Zealand’s COVID-19 response,” Journal of the Royal Society of New Zealand, vol. 51: sup1, 2021, S86-S106, doi: 10.1080/03036758.2021.1876111

[4] M. H. Kabir, M. O. Gani, S. Mandal, M. H. A. Biswas, “Modeling the dispersal effect to reduce the infection of COVID-19 in Bangladesh,” Sensors International, vol. 1, 2020, 100043. doi: 10.1016/j.sintl.2020.100043

[5] M. Locquet, et. al., “A systematic review of prediction models to diagnose COVID-19 in adults admitted to healthcare centers,” Archive of Public Health, vol. 79, iss. 105, 2021. https://doi.org/10.1186/s13690-021-00630-3

[6] Z. Khan, F. Van Bussel, F. Hussain, “A predictive model for Covid-19 spread – with application to eight US states and how to end the pandemic,” Epidemiology and Infection, vol. 148, 2020, E249. doi:10.1017/S0950268820002423

[7] R.A. Brown, “A simple model for control of COVID-19 infections on an urban campus,” Proceedings of the National Academy of Sciences, vol. 118, iss. 36, 2021, e2105292118. doi: 10.1073/pnas.2105292118

[8] A. Leontitsis, et. al., “SEAHIR: A Specialized Compartmental Model for COVID-19,” International journal of environmental research and public health, vol. 18, iss. 5, 2667. doi: 10.3390/ijerph18052667

[9] M. Dashtbali, M. Mirzaie, “A compartmental model that predicts the effect of social distancing and vaccination on controlling COVID-19,” Scientific Reports, vol. 11, 2021, 8191. doi: 10.1038/s41598-021-86873-0

[10] S. Fedushko, T. Ustyinovych, “Operational Intelligence Software Concepts for Continuous Healthcare Monitoring and Consolidated Data Storage Ecosystem,” Advances in Intelligent Systems and Computing, vol. 1247, 2021, pp. 545–557. doi: 10.1007/978-3-030-55506-1_49

[11] R. Vaishya, M. Javaid, I. H. Khan, A. Haleem, “Artificial Intelligence (AI) applications for COVID-19 pandemic”, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, iss. 4, 2020, pp. 337-339. doi: 10.1016/j.dsx.2020.04.012

[12] I. Izonin, et. al., “Predictive modeling based on small data in clinical medicine: RBF-based additive input-doubling method,” Mathematical Biosciences and Engineering, vol. 18, iss. 3, 2021, pp. 2599-2613. doi: 10.3934/mbe.2021132

[13] M. Mahdavi, et al., “A machine learning based exploration of COVID-19 mortality risk,” PLoS ONE, vol. 16, iss. 7, 2021, e0252384. doi: 10.1371/journal.pone.0252384

[14] S. Yakovlev, et. al., “The concept of developing a decision support system for the epidemic morbidity control,” CEUR Workshop Proceedings, vol. 2753, 2020, pp. 265–274.

[15] V. J. M. Quinn, et. al., “COVID-19 at War: The Joint Forces Operation in Ukraine,” Disaster Medicine and Public Health Preparedness, 2021, pp 1-8. doi: 10.1017/dmp.2021.88

[16] S.S. Patel, et. al., “The Landscape of Disinformation on Health Crisis Communication During the COVID-19 Pandemic in Ukraine: Hybrid Warfare Tactics, Fake Media News and Review of Evidence,” JCOM Journal os Science Communication, vol. 19, iss. 5, 2020, AO2. doi: 10.22323/2.19050202

[17] T. Kirby, “New variant of SARS-CoV-2 in UK causes surge of COVID-19,” The Lancet. Respiratory Medicine, vol. 9, iss. 2, 2021, e20-e21. doi: 10.1016/S2213-2600(21)00005-9

[18] K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, “Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020,” Euro Surveillance: European communicable disease bulletin, vol. 26, iss. 1, 2021, 2002106. doi: 10.2807/1560-7917.ES.2020.26.1.2002106

[19] S.C.E. Schmidt, et. al., “Physical activity and screen time of children and adolescents before and during the COVID-19 lockdown in Germany: a natural experiment,” Scientific Reports, vol. 10, iss. 1, 2020, 21780. doi: 10.1038/s41598-020-78438-4

[20] T. Mitze, R. Kosfeld, J. Rode, K. Wälde, “Face masks considerably reduce COVID-19 cases in Germany,” Proceedings of the National Acadademy of Sciences of the United States of America, vol. 117, iss. 51, 2020, 32293-32301. doi: 10.1073/pnas.2015954117

[21] O. Amengual, T. Atsumi, “COVID-19 pandemic in Japan,” Rheumatology International, vol. 41, iss. 1. 2021. pp. 1-5. doi: 10.1007/s00296-020-04744-9

[22] S. Saito, et. al., “First and second COVID-19 waves in Japan: A comparison of disease severity and characteristics,” The Journal of Infections, vol. 82, iss. 4, 2021, pp. 84-123. doi: 10.1016/j.jinf.2020.10.033

[23] S. Kavitha, S. Varuna, R. Ramya, “A comparative analysis on linear regression and support vector regression,” 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp. 1-5, doi: 10.1109/GET.2016.7916627.

[24] R. Muthukrishnan, R. Rohini, “LASSO: A feature selection technique in predictive modeling for machine learning,” 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 18-20, doi: 10.1109/ICACA.2016.7887916.

[25] D. Boyko, et. al., “The Concept of Management Decisions Support System to Mitigate the COVID-19 Pandemic Consequences based on an Assessment of Social and Epidemic Processes using Social Networks,” CEUR Workshop Proceedings, 2021, vol. 3003, pp. 55-64.