

# Big Data, Analyzing and Modelling: New Ways of Health Improvement and Regional Aspects

Vasyl Kopytko<sup>1</sup>, Lyubov Shevchuk<sup>2</sup>, Larysa Yankovska<sup>2</sup>, Zhanna Semchuk<sup>2</sup>

<sup>1</sup> *Lviv Branch of Dnipropetrovsk National University of Railway Transport named after Academician V. Lazaryan*  
12a I. Blazhkevych Street, Lviv, 79052, Ukraine

<sup>2</sup> *Lviv University of Business and Law*  
99 Kulparkivska Street, Lviv, 79021, Ukraine

DOI: [10.22178/pos.37-2](https://doi.org/10.22178/pos.37-2)

JEL Classification: O10

Received 25.07.2018

Accepted 10.08.2018

Published online 31.08.2018

Corresponding Author:  
Zhanna Semchuk  
[semchuk\\_zv1@ukr.net](mailto:semchuk_zv1@ukr.net)

© 2018 The Authors. This article is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) 

**Abstract.** The field of health improvement and life prolonging develops poorly, despite all the advances in medicine, chemistry and genetic engineering. Among the main problems is the difficulty of using new scientific achievements in other industries due to the rapid development of specialized knowledge, the problem of returning costs for the creation of really effective and the problem of aging population in developed countries. There are problems with data for this methods usage with privacy and security on different levels with regional peculiarities. Effective timing of work on health at the personal level can result as a result of increased time and productivity. But it's difficult for people to allocate their intellectual resources for that, so you have to connect artificial intelligence and machine learning. Big Data model with methods and analysis techniques on different levels for health improvement was suggested. The importance of the level of social networks and its regional aspects for the analysis of health improvement data was identified. Big data processing results implementation and levels of interaction with human with request for changes model was proposed. It consists from two levels of interaction with humans by level of quick reaction and discussion with smart personal assistance. Regional aspects from possible AI implementation in undeveloped countries were analyzed on example of personal level big data for health usage.

**Keywords:** big data; analyzing; modelling; health improving; regional aspects.

## INTRODUCTION

The problem of improving health and prolonging life is relevant for everyone. Intelligent devices such as fitness bracelets, toothbrushes, pressure gauges, DNA gene analyzers, and other sensors are already generating a large stream of data that needs to be analyzed and made of predictions and recommendations. The field of health improvement and life prolonging develops poorly, despite all the advances in medicine, chemistry and genetic engineering. People are still ill and dying while an increase in the life expectancy is observed in statistical data, but not as fast and as obvious as it would be desirable. Among the

main problems is the difficulty of using new scientific achievements in other industries due to the rapid development of specialized knowledge, the problem of returning costs for the creation of really effective drugs, because if they really fix the problem, then demand for them is rapidly falling due to the healing [1] and the problem of aging population in developed countries where the states it is necessary to take unpopular measures to increase the retirement age, etc. We also have some sources of data relevant to health improving on different levels from levels of personal mobile accessories to databases of international organizations of health statistics. The main problem is now data interpretation and solutions

predictions on high software level where first general steps must be in the architecture, planning and standardization fields. The privacy, commerciality of the necessary data to analyze and regional aspects remains a serious problem to find effective solutions.

## LITERATURE REVIEW

Researchers such as Groves P., Kayyali B., Knott D., Van Kuiken S. [2], Bhattacharyya O., Khor S., McGahan A., Dunne D., Daar A. S., Singer P. A. [3], Pisani E, AbouZahr C. [4] and others were engaged in research of big data, analyzing and modelling in the field of health improvement. They make predictions of a quick revolution in healthcare with using of big data and are already conducting a regional coordination of the implementation of possible changes in the underdeveloped countries. Also, Ukrainian scientists such as L.Globa, I. Ishchenko, N. Kunieva, V.Kurdecha, A.Zakharчук, work in this area with fuzzy logic, AI (artificial intelligence) and big data methods in health improvement [5, 6]. They are investigating e-Health monitoring system development based on the IoT (Internet of things) technologies as software and hardware infrastructure of the activity with the usage of electronic information resources in the health sector and it ensures rapid access of medical professionals and patients to them. Economy and regional aspects of using big data and AI problems investigated in works such scientists as A. Shevchuk [7], Xiaojun Wang, Leroy White, Xu Chen [8], Aoyun Chen [9]. The authors of these publications note the role of Big Data and AI in the development of regions and their economies and the problem of the massive use of these new technologies and singularity problems of the interaction of scientists from various fields of science to solve problems of this type. These scientific problems are being studied by the United Nations [10], World Health Organization and leading scientific schools from different perspectives such as economy, rationality, privacy and security [11, 12]. This topic is poorly investigated from a multidisciplinary approach, and for successful mass implementation along with a mathematical apparatus and methods of working with large data, a necessary analysis of the economic and regional aspects of the implementation of existing theoretical developments is necessary.

## RESULTS

### Big Data and health improving

Big data is a model of the application and techniques that process with very large source data. The source is often very large and complex for processing it with typical database tools and methods [2, 3, 10]. Source examples for health improvement includes logs from fitness devices, implants, temperature and pressure sensors, daily activity statistics, logs of social networks and work activities, etc. Global big data market size growth up in 2017 to the just under 34 billion U.S. dollars from 7.6 in 2011 [13]. This is one of the fastest growing IT technologies. There are several Big Data analysis techniques that can be combined or used stand alone. Among them association rule learning; genetic algorithms, classification tree analysis, machine learning, regression analysis; sentiment analysis; social network analysis etc. [2, 10] (Figure 1). There are problems with data for this methods usage with privacy and security on different levels with regional peculiarities. Effective timing of work on health at the personal level can result as a result of increased time and productivity. But it's difficult for people to allocate their intellectual resources for that, so you have to connect artificial intelligence and machine learning. There are many fundamental issues with the certification of how it will be assistant to everyone on health then you need to somehow legitimize. Let's consider the main methods of using Big Data to model, analyze and find patterns of indicators and health improvement. Association rule learning is a method that can allow to find correlations between variables data. Correlations finding mechanism must be checked with statistical methods for false conclusions excluding [14]. Such algorithms as Apriori, Eclat, FP-Growth typically are used on this stage for mining frequent itemsets. Another next steps needs to be done with Context Based Association Rule Mining, node-set-based or OPUS search algorithm on the next phase. Connection and correlation definition between series data from different sources is extremely important for the next implementation of solutions and detection of changes. False correlations are a very frequent phenomenon which makes it impossible to develop effective implementations. Also classification tree analysis methods can be used on this stage such as decisions graphs or evolutionary algorithms.

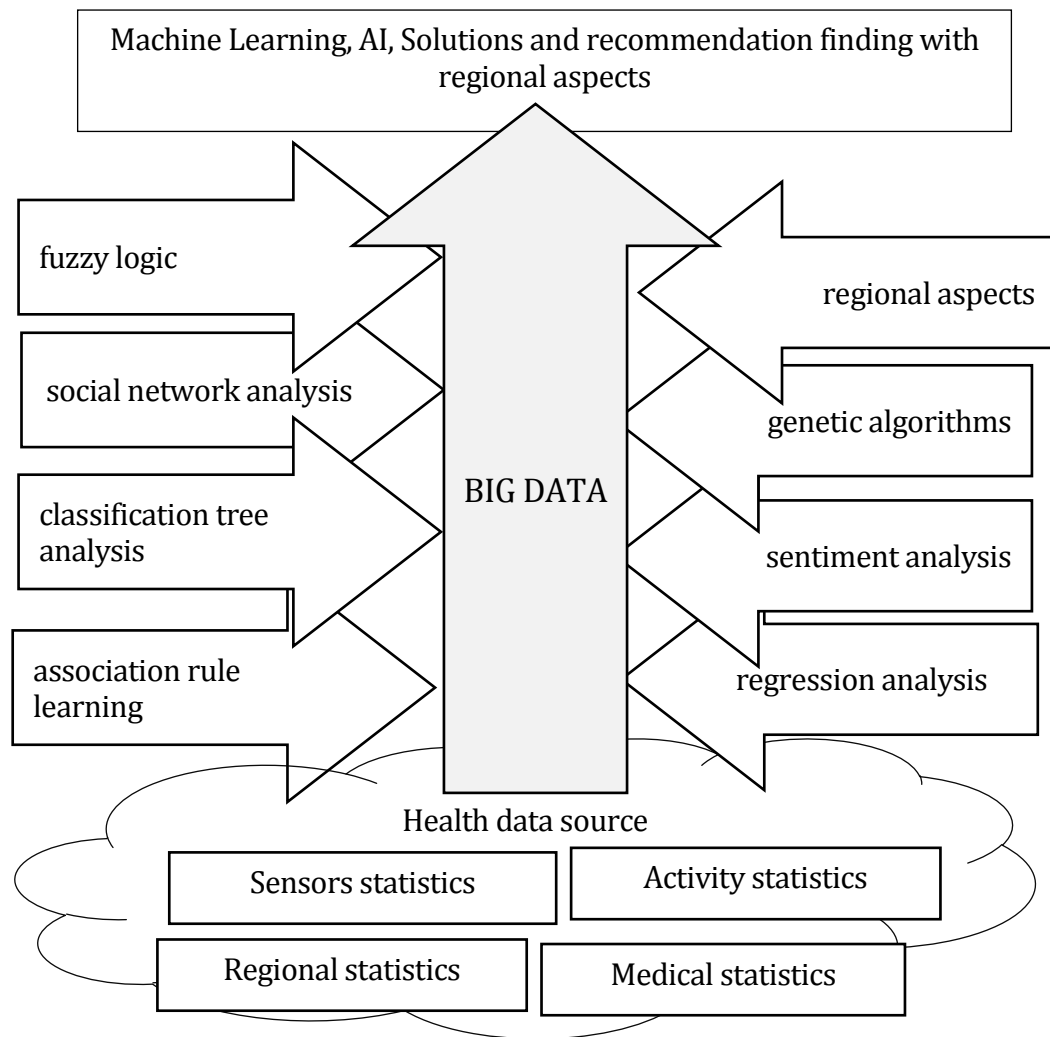


Figure 1 – Big Data methods and analysis techniques for health improvement

The intuitive logic of combining health data or is likely to be misleading in identifying new, important trends. But the regional aspect of the initial processing of data must be taken. All areas of social and socio-economic life have experienced significant changes with the development of information technology, both globally and regionally. Especially tangible information technologies influenced the employment of the population. These changes concerned both the labor process and its organization, the creation of jobs in some places of application of labor and their reduction in other, global and regional labor market. At least one third of the time a typically human is working in our time on PC. That is why it is precisely because of the impact on employment of information technology or not the most affecting the way and quality of life of the population. An important role in the process of integration of information technologies into the employment of the population was played by the mass media,

automation and introduction into the production process of computer technologies and robotics, new means of communication and computer networks, in particular the Internet [15]. So, social networks and activity statistics also can be used as data series for health of person analyzing, especially mental health. Social network analysis can be used on all data series but especially effective results will be with data from concrete social networks like Facebook connection. The problem is in the commerciality and privacy of information from social networking developers, OS and PC manufacturers which is highly segmented and difficult to evaluate. It will also be difficult to assess the likelihood of the results and determine the criteria for success especially from the changes. But this information despite all the disadvantages is extremely important for analysis and modeling and it will have a very pronounced regional aspect. This level is also directly related to the analysis and the positive one is that people

themselves will implicitly provide data for analysis without additional hardware or software. Also for Big data variables we can use regression analysis. It is set of statistical methods and processes for estimating the relationships among variables in the series. Results of this analysis is values changes of fixed data series with variety of another. Methods allow to find an independent and depended variables, estimate the conditional expectation and regression function. Results can help to find the value of the parameters that must be changed to receive changes of another variable for some predefined values [16]. Linear and nonlinear regression, interpolation and extrapolation is typically used to find how we must to change our activity to receive some better results with health. Sentiment analysis operate with biometrics, natural language processing, speech recognition, text analysis and computational linguistics for subjective information identifying, extracting, quantifying [17]. It is important to analyze voice, text and emotions data from people as result for some activity. There are several statistical methods of such as latent semantic analysis, support vector machines, Semantic Orientation Pointwise Mutual Information, "bag of words", etc. Also syntactic methods, deep and machine learning with topic modeling is commonly using. Genetic algorithms can allow to generate solutions and optimization recommendation with mutation, crossover and selection processes and operations [18]. A typical genetic algorithm commonly requires a genetic representation of the solution domain and a fitness function to evaluate the solution domain. Genetic representation can be data from sensors, smart devices and other activities. Its results are important for modelling of implementation solutions with another methods of analyzing to find best practice and detect false ways. And on top is methods of machine learning and artificial intelligence solutions such as neural networks, smart agents and fuzzy logic. This is a core center for solutions on health improving findings based on all previous methods of analyzing. The main problems of their implementation at present are the weak dissemination of the source code for solutions and research and their copyright protection [19].

### **Big Data and health on personal level usage**

Allocating one hour per day to the individual's own health should bring the relevant results.

Only feedback in such interaction is important here. We consider the response in the form of vibrating senses of a smart clock, as an architectural principle for the creation of an appropriate application. Vibes can be identified as certain messages and signaling the occurrence of a sequence of relevant communications that indicate an event. So the event triggers can signal a hazard, an unwanted mode of life or a positive feedback as a continuation of the desired activity. The difficulty here arises unexpectedly in the mental sphere and in the motivation. People dream a lot about eternal life and iron health, but on the other hand they are not able to give time for that. This also indicates that first of all the laws should begin with the psychological scope of activity, the field of motivation and social network. But the feedback can also be in the form of visual images or colors on smart clocks or bracelets. Their introduction in daily use can be carried out with the help of unconditional reflexes and habits. The neural network for a person having analyzed the data should already report the results to execution. Virtualization and modeling of solutions may involve computing power in the clouds, which must be continuously refined by evolutionary generating algorithms. And on high level of interaction is information and discussion with smart assistance by voice like Apple Siri etc. Estimation and feedbacks are important for making changes to the developed schemes and recommendations for changing the lifestyle, additional physical activity, receiving appropriate vitamins or lyre drugs, and applying for a doctor. Applicable methods should take into account the results of previous implementations and decisions. Backpropagation of error than can allow to receive more exact or right solutions. Another manifestation of the influence of technological progress in the form of information technology is the significant intensification of the process of labor, which leads to an increase in the fatigue of the working people of all categories, and this fatigue is not physical, as before but the neuropsychic and it is associated with the change in the nature of labor in the process of technological revolution. The internationalization of large corporations and the middle-aged business with remote employment now gives great advantages in which the sense of patriotism and citizenship of a particular state disappears. The regional aspect still remains individual for everyone, as we live in different climatic zones, settlements of different kinds and sizes, and our conditions around are unique. Concrete decisions and solutions to improve health are needed by everyone in a

shared virtual reality and different environments. The mass media, automation and introduction of computer technologies and robotics, new communication and computer networks, in particular the Internet, played an important role in the integration of information technologies into the employment of the population, which led to the emergence and development of distance work. Such work does not need much physical effort: all the workload in the process of its implementation is transferred, practically, to the mental sphere, which sets new requirements for the formation of a healthy lifestyle (for example, the development of labor standards and their compliance in the work process). Economic efficiency of remote work is achieved by reducing transport and energy (for example, maintenance of office space, electricity costs) costs. Labor productivity and quality of work are not reduced at the same time and sometimes higher. The popularity of remote work is growing due to the uneven workload of the population in its various regions. Remote work is considered as one of the ways to solve employment in agricultural regions. The society derives the following socio-economic benefits from the use of remote work: reducing the severity of transport problems and, as a consequence, environmental pollution; reduction in unemployment: giving people in regions with high unemployment opportunities to access work in any part of the world; access to work of people with disabilities; equal rights to receive quality education using distance learning, regardless of the place of residence of a person. The use of wireless and mobile networks, the mismatch of time zones in remote work with increasing time spent by the population on a PC can negatively affect the health of the population. It also allows for constant monitoring of human health in the workplace, as it constantly interacts with digital devices. The decisions in this case for planning physical activity and compensating for sedentary remote work from analyzing Big data related to monitoring the health status will be extremely important.

## CONCLUSIONS

The field of health improvement and life prolonging develops poorly, despite all the advances in medicine, chemistry and genetic engineering. Among the main problems is the difficulty of using new scientific achievements in other industries due to the rapid development of specialized

knowledge, the problem of returning costs for the creation of really effective and the problem of aging population in developed countries. There are problems with data for this methods usage with privacy and security on different levels with regional peculiarities. Effective timing of work on health at the personal level can result as a result of increased time and productivity. But it's difficult for people to allocate their intellectual resources for that, so you have to connect artificial intelligence and machine learning. Big Data model with methods and analysis techniques on different levels for health improvement was suggested. There are several Big Data analysis techniques that must be used for data series analyzing. Among them association rule learning; genetic algorithms, classification tree analysis, machine learning, regression analysis; sentiment analysis; social network analysis. The top method is machine learning and artificial intelligence solutions such as neural networks, smart agents and fuzzy logic. This is a core center for solutions on health improving findings based on all previous methods of analyzing. The main problems of their implementation at present are the weak dissemination of the source code for solutions and research and their copyright protection. The importance of the level of social networks and its regional aspects for the analysis of health improvement data was identified. Big data processing results implementation and levels of interaction with human with request for changes model was proposed. It consists from two levels of interaction with humans by level of quick reaction and discussion with smart personal assistance. The feedback can also be in the form of visual images or colors on smart clocks or bracelets. Their introduction in daily use can be carried out with the help of unconditional reflexes and habits etc. The neural network for a person having analyzed the data should already report the results to execution. Virtualization and modeling of solutions may involve computing power in the clouds, which must be continuously refined by evolutionary generating algorithms. And on high level of interaction is information and discussion with smart assistance by voice. The decisions from collected data with planning physical activity receiving appropriate vitamins or drugs, applying for a doctor with monitoring the health status can allow to improve health of people on personal level.

## REFERENCES

1. Tae, K. (2018, April 18). *Goldman Sachs asks in biotech research report: 'Is curing patients a sustainable business model?'*. Retrieved from <https://www.cnbc.com/2018/04/11/goldman-asks-is-curing-patients-a-sustainable-business-model.html>
2. Groves P., Kayyali B., Knott D., & Van Kuiken, S. (2013, January). *The 'big data' revolution in healthcare. Accelerating value and innovation*. Retrieved from [https://www.ghdonline.org/uploads/Big\\_Data\\_Revolution\\_in\\_health\\_care\\_2013\\_McKinsey\\_Report.pdf](https://www.ghdonline.org/uploads/Big_Data_Revolution_in_health_care_2013_McKinsey_Report.pdf)
3. Bhattacharyya, O., Khor, S., McGahan, A., Dunne, D., Daar, A. S., & Singer, P. A. (2010). Innovative health service delivery models in low and middle income countries - what can we learn from the private sector? *Health Research Policy and Systems*, 8(1). doi: 10.1186/1478-4505-8-24
4. Pisani, E., & AbouZahr, C. (2010). Sharing health data: good intentions are not enough. *Bulletin of the World Health Organization*, 88(6), 462–466. doi: 10.2471/blt.09.074393
5. Globa, L., Ishchenko, I., & Zakharchuk, A. (2017). Data Processing in E-Health System. *Journal of Communication and Computer*, 14(1). doi: 10.17265/1548-7709/2017.01.006
6. Kurdecha, V., Ishchenko, I., Zakharchuk, A., & Kunieva, N. (2018). *Fuzzy logic usage for the data processing in the Internet of Things networks*. Retrieved from <https://libeldoc.bsuir.by/handle/123456789/30355>
7. Shevchuk, A. V. (2016). *Artificial Intelligence and Intellectualization: New Prospects for Economic Development*. *Scientific Review*, 4(25), 27–36.
8. Wang, X., White, L., & Chen, X. (2015). Big data research for the knowledge economy: past, present, and future. *Industrial Management & Data Systems*, 115(9). doi: 10.1108/imds-09-2015-0388
9. Chen, A. (2017). Big Data and the Development of Regional Economy. *Proceedings of the 2016 International Conference on Modern Management, Education Technology, and Social Science (MMETSS 2016)*. doi: 10.2991/mmetss-16.2017.32
10. UN Global Pulse. (2012). *Big data for development: challenges and opportunities*. Retrieved from <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>
11. London School of Economics and Political Science. (2010, December). *Electronic health privacy and security in developing countries and humanitarian operations*. Retrieved from <http://personal.lse.ac.uk/martinak/ehealth.pdf>
12. World Economic Forum. (2011). *Global health data charter*. Retrieved from [https://share.kaiserpermanente.org/media\\_assets/pdf/ceoletters/2011/downloads/WEFGlobalHealthDataCharter2011.pdf](https://share.kaiserpermanente.org/media_assets/pdf/ceoletters/2011/downloads/WEFGlobalHealthDataCharter2011.pdf)
13. Statista. (2018). *Forecast of Big Data market size, based on revenue, from 2011 to 2026 (in billion U.S. dollars)*. Retrieved July 1, 2018, from <https://www.statista.com/statistics/254266/global-big-data-market-forecast>
14. Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*. doi: 10.1145/170035.170072
15. Freeman, L. (2006). *The Development of Social Network Analysis*. Vancouver: Empirical Press.
16. Freedman, D. A. (2012). *Statistical models: Theory and practice*. New York: Cambridge University Press.
17. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2), 15–21. doi: 10.1109/mis.2013.30

18. Zhang, J., Chung, H. S.-H., & Lo, W.-L. (2007). Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms. *IEEE Transactions on Evolutionary Computation*, 11(3), 326–335. doi: [10.1109/tevc.2006.880727](https://doi.org/10.1109/tevc.2006.880727)
19. AIReligion. (2018, February 20). *The development of artificial intelligence is hampered by programmers who hide their code*. Retrieved from <http://aireligion.org/?p=839>