Methodological problems of big data and artificial intelligence in the medical specialists training

O V Mikhailenko¹ and G A Dorrer¹,²

¹Siberian Federal University, prospect Svobodny 79/10, Krasnoyarsk, Russia
²Reshetnev Siberian State University of Science and Technology, prospect “Krasnoyarsky rabochy”, 31, Krasnoyarsk, Russia

E-mail: g_a_dorrer@mail.ru

Abstract. The emergence of big data and artificial intelligence firstly in healthcare has caused considerable excitement, stating the need to improve approaches to diagnosis, prognosis, and treatment. Despite enthusiasm, the methodological assumptions underlying the movement of big data and artificial intelligence in medicine are rarely studied. This article outlines the methodological problems facing this movement. In particular, the following topics were considered: the theory of large data congestion, the limits of the algorithms action, and the phenomenology of the disease. These methodological issues demonstrate several important roles for these technologies that must be considered and studied before they are integrated into the healthcare system.

1. Introduction
The emergence of big data and artificial intelligence (AI) firstly in healthcare has caused considerable excitement, stating the need to improve approaches to diagnosis, prognosis, and treatment. Amid this enthusiasm, methodological assumptions underlying the movement of big data and artificial intelligence in medicine are rarely considered. This article outlines some methodological issues facing this movement.

In particular, the following topics were considered: the theory of large data congestion, the limits of the algorithms action, and the phenomenology of the disease. These methodological issues demonstrate several important roles for these technologies that must be considered and studied before they are integrated into the practical healthcare system.

2. Big data congestion theory
Critics of big data have recognized the importance of theoretically sound data collection and interpretation [1]. The main assumption of the movement of big data is that an unbiased, theory-free access to data is possible and that more data will allow learning algorithms to make more accurate forecasts [2]. Data, however, does not exist in a vacuum; rather, they represent carefully selected information based on a priori assumptions and theories. The relationship between theory and data has been a central issue in the philosophy of science over the past century.

Karl Popper [3] tells a joke from the time when he lectured a group of physics students in Vienna in the 1930s. Popper began his lecture with the following words: “Take paper and pencil, observe and describe your observations!” His students asked, of course, what exactly they should observe. Clearly,
a simple instruction: “Watch” is absurd. Observation is always selective. It is necessary to choose an object, a specific task, have some interest, point of view, problem.

This problem, sometimes called the “Observation of Congestion Theory”, has occupied many leading philosophers of science throughout the 20th century. On this example, Popper launched his criticism of the logical positivism that prevailed at the time of the philosophical school, which considered “facts” as independent, valueless “observational statements” about the world [4]. According to the positivist view, only propositions related to scientifically verifiable facts are significant.

Positivism continues to live today in the quantitative epistemologies of medicine, the latest incarnation of which is big data and machine learning [5,6,7,8]. This perspective remains, despite the serious problems posed by modern philosophy. In addition to Popper, other well-known critics have shown how observations always occur in the light of the existing “conceptual scheme”, or showed how empirical data are generated by the theoretical premises of the prevailing “scientific paradigm” [9, 10]. The concepts of theory were expanded to help us understand how other extra-empirical factors shape the production of knowledge, influencing movement in social epistemology [11]. Some (but not all) of these factors are reflected in Popper's statement on the role of “interests, points of view, and problems” in scientific research.

The consequences of these ideas were applied to medicine, especially in criticizing evidence-based medicine [7,12,13]. Although some scholars have a finer understanding of data, the generally positivist point of view remains undeniable in the movement for big data. The relation of big data to causality is the rudiment of logical positivism.

Big data and machine learning approaches often neglect causal reasoning, preferring to interrogate data without reference to a causal structure. This may be due to the nature of some machine learning algorithms, such as neural networks that build black boxes that prevent researchers from identifying cause-effect relationships between inputs and outputs [14]. Critics argue that the lack of a causal structure limits the interpretation of the data and that the theoretical framework linking causes and effects underlies sound clinical reasoning and statistical thought [15]. Proponents of big data and machine learning argue that the “atheistic” perspective is actually a strength of these methods, avoiding the biases that hold back scientific progress. Some scholars have come to the point of declaring the “end of theory” in the era of big data [16].

However, these supporters do not understand that there can be no theoretical perspective. The theory did not die in the era of big data, because, as Popper reminded us, data cannot do without theory - all studies involve “interests, points of view and problems.” This lesson is of particular importance to the movement of big data. What is observed - this is what is considered data - is formed by the epistemic interests of specific researchers and research communities. This is reflected in how data is collected into pipelines for training neural networks and machine learning algorithms.

3. The limits of the algorithms

Any report on the usefulness of artificial intelligence and machine learning in medicine should explain how a finite set of algorithms or bits of a programming language can relate to the complex realities of the phenomenal and biological world. Since algorithms are just a specification of the set of rules that should be followed in a programming language, it would be necessary to completely identify all processes - biological, social, psychological, historical. This is impossible.

To begin with, such events occur outside the programming language and can be translated into this language only after they have occurred, forming a data set for algorithm access. This is a serious inconsistent factor, given that the past empirical success of the program does not guarantee its future success, in particular when it comes to the dynamics of complex systems.

This is partly impossible due to the incompleteness of current empirical explanations of the phenomenon to be explained. Given that the Universe is an open system with the laws of thermodynamics that indicate high degrees of entropy, and that biological systems are subject to complex evolutionary dynamics that are not fully understood and lack consistent teleology for both
cosmological and evolutionary forces, the algorithm cannot find a way out of these forces to give predictions.

Any attempt to build a computer program that can predict the clinical outcome “with perfect accuracy” (as some proponents of big data analysis suggested [2]): will require a full description of this phenomenon, which remains elusive in clinical medicine. Consider a common disease such as exacerbation of asthma, an event that some proponents of the theory claim can be predicted using machine learning algorithms [17]. A growing number of studies have focused on the use of AI to predict asthma events and outcomes [18]. However, attempts to construct algorithms for predicting such extremely complex, conditional results may encounter significant limitations. In some cases, they can lead to illogical and even potentially dangerous results, such as the assumption of one predictive model of machine learning that asthma is a protective factor in patients admitted to a hospital with pneumonia [19]. Such errors arise because these programs remain incomplete and underestimate the complexity of the systems they are trying to simulate. Even with an extended set of explanatory variables, it is unlikely that such models will be able to take into account rare or specific factors, such as thunderstorms and other environmental impacts, which are known to have a significant impact on asthma epidemics. Given these limitations, a more modest perspective is needed, far from any concept of perfect prediction.

4. Phenomenology of the disease
The phenomenology of the disease has become an influential area in the methodology of medicine, which has important applications in clinical practice and medical education. This movement focuses on experiencing a first-person illness, as opposed to a third-person perspective proposed by the naturalistic biomedical model. The phenomenology of the disease is often contrasted with the view of biomedicine on the disease as a biological dysfunction, which, according to phenomenologists, objectifies patients and devalues the subjective experiences of the disease. The leading contemporary authors of this movement are S. Kay Tumb, Javi Karel and Fredrik Sveneus, among others who rely on the work of 20th-century phenomenologists from Edmund Husserl, Martin Heidegger and Hans-Georg Gadamer to Jean-Paul Sartre and Maurice Merlot-Ponti. These authors offer convincing arguments in favor of the value of the phenomenological approach to go beyond the generalizations of third-party biomedicine and to illuminate the first-person experience of the disease, paying attention to its diversity and complexity.

Incarnation is a key concept in the phenomenology of the disease. Body perception is filled with meaning that precedes any isolated "sensory data." Contrary to the logical positivistic view, a person does not perceive the world as isolated sensory data, and then begins to build observational statements with empirical content and meaning.

Phenomenology emphasizes that meaning and context are the main features of experience that are not explained by the depleted sensory data of logical positivism. This approach also illustrates why first-person experience in the health field cannot be embraced by simply looking at context-free statements of individual preferences and values. Such attempts remain tied to atomistic ideas about the input data, which refuse nuances and complexity, failing to arrange perception and experience in a wider world of life.

The first-person view provided by the phenomenology of attention to a living body helps us better understand the experience of the disease. As Tumb points out [20]: “Arthritis is not so much an inflammation of the joints as an ‘inability’ to fasten a shirt, swing a golf club, play tennis.”

Such perspectives provide valuable information to practitioners and can help restructure care so that it best meets the needs of patients. Phenomenological approaches served as the basis for creating patient resources and educational tools for medical workers [21]. Karel offers “phenomenological tools” that allow patients to better articulate their experience and allow clinicians to hone “epistemic feelings and skills” [22]. A key step in phenomenological tools is “bracketing a natural relationship,” which entails an instant rejection of narrow biomedical ontologies to contribute to a richer and more comprehensive understanding of how a disease affects a person’s world.
Karel and Kidd [23] advocate a phenomenological approach as a means of counteracting epistemic injustices that occur in healthcare and can lead to denial or disbelief in the patient's experience. Neglect of big data and machine learning by the phenomenology of the disease can exacerbate epistemic injustice: not only first-person experience is excluded from these quantitative tools, but this knowledge is also deprived of clinical significance in the face of powerful algorithms fueled by data arrays. This echoes the criticism expressed ten years ago regarding the hierarchies of evidence-based medicine, which, preferring certain research methodologies and definitions of evidence, devalue other sources of knowledge, such as the testimonies of patients and doctors [7].

Elsewhere, we have argued that interaction with first-person experience is important for clinical judgment [5,6]. In addition to phenomenology, other methods have been proposed to support this interaction, such as narrative or historical approaches [5]. There are similarities and differences between these approaches, as well as differences that can be made between medical phenomenologists. However, the phenomenology of the disease is a particularly articulated position stemming from a strong philosophical structure that offers a range of clinically useful applications. The phenomenological approach also best underlines the shortcomings of the epistemological approach generated by the movement of big data and machine learning in medicine.

5. Conclusion
The advent of big data and artificial intelligence in healthcare raises important methodological issues that are relevant to clinical medicine. Allegations that doctors will soon be replaced by AI are truly exaggerated; however, the methodological problems highlighted here suggest closer detrimental consequences of these technologies, from the privilege of quantitative data and the exclusion of first-person knowledge to the undetermination of clinical complexity by AI algorithms.

References
[1] Coveney P V, Dougherty E R and Highfield R R 2016 Big Data Need Big Theory Too Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences 374
[2] Sniderman, A D, D’Agostino R B and Pencina M J 2015 The Role of Physicians in the Era of Predictive Analytics Journal of the American Medical Association 314
[3] Popper K 1962 Conjectures and Refutations (London: Routledge)
[4] Carnap R 1919 Logical Syntax of Language. (London: Routledge and Kegan Paul)
[5] Chin-Yee B H and Upshur R E G 2015 Historical Thinking in Clinical Medicine: Lessons from R. G. Collingwood’s Philosophy of History Journal of Evaluation in Clinical Practice 21
[6] Chin-Yee B and Upshur R 2018 Clinical Judgment in the Era of Big Data and Predictive Analytics Journal of Evaluation in Clinical Practice 24
[7] Goldenberg M J 2006 On Evidence and Evidence-Based Medicine: Lessons from the Philosophy of Science Social Science & Medicine 62
[8] Goldenberg M J 2009 Iconoclast or Creed? Objectivism, Pragmatism, and the Hierarchy of Evidence Perspectives in Biology and Medicine 52
[9] Kuhn T S 1962 The Structure of Scientific Revolutions (Chicago: University of Chicago Press)
[10] Quine W V O 1980 From a Logical Point of View: Nine Logico-Philosophical Essays (Cambridge: Harvard University Press)
[11] Longino H E 2002 The Fate of Knowledge (Princeton: Princeton University Press)
[12] Chin-Yee B H 2014 Underdetermination in Evidence-Based Medicine Journal of Evaluation in Clinical Practice 20
[13] Upshur R E G 2005 Looking for Rules in a World of Exceptions: Reflections on Evidence-Based Practice Perspectives in Biology and Medicine 48
[14] Tu J V 1996 Advantages and Disadvantages of Using Artificial Neural Networks Versus Logistic Regression for Predicting Medical Outcomes Journal of Clinical Epidemiology 49
[15] Pearl J 2009 Causality: Models, Reasoning and Inference (Cambridge: Cambridge University
[16] Anderson C 2008 *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* Wired Retrieved from <https://www.wired.com/2008/06/pb-theory/>

[17] Finkelstein J and Jeong I C 2016 Machine Learning Approaches to Personalize Early Prediction of Asthma Exacerbations *Annals of the New York Academy of Sciences* **1387**

[18] Cabitza F R, Rasoini, and Gensini G F 2017 Unintended Consequences of Machine Learning in Medicine *Journal of the American Medical Association* **318**

[19] Sculley D 2014 Machine Learning: The High-Interest Credit Card of Technical Debt *Software Engineering for Machine Learning* Retrieved from <https://se4ml.org/>#

[20] Merleau-Ponty M 1945 *Phenomenology of Perception* (London: Routledge)

[21] Carel H 2016 *Phenomenology of Illness* (Oxford: Oxford University Press)

[22] Carel H and Kidd I J 2014 Epistemic Injustice in Healthcare: A Philosophical Analysis *Medicine, Health Care and Philosophy* **17**

[23] Chin-Yee B, Messinger A and Young L T 2019 Three Visions of Doctoring: A Gadamerian Dialogue *Advances in Health Sciences Education* **24**