



Come What May, Digital Health Technologies Will Never Be Able to Predict the Emergence of Unknown Viruses and Microorganisms with any Degree of Certainty

Arni S. R. Srinivasa Rao^{1,2,3} · Steven G. Krantz⁴

Received: 31 August 2020 / Accepted: 15 October 2020 / Published online: 21 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

When there are a finite number of options to play at any stage of a particular game of standard chess played between two players, a deep-learning algorithm can be easily programmed to act like a digital machine to predict the set of all next possible moves by a player. Similar kinds of prediction rules could also be built for other types of online games played between two or multiple players. Deep-learning (DL) combined with an artificial intelligence (AI) framework becomes a powerful tool in several other useful situations in the real-world [1, 2], and we do not minimize such a potential by machines in the future.

But predicting and identifying accurately new and unknown microorganisms such as bacteria, viruses, etc., before they emerge with the help of digital technologies equipped with DL algorithms or pre-programmed steps seems to be impossible. This is due to an enormously large number of species possibilities and also very large bounds of uncertainties for emergence of newer species that are involved. Naturally we are interested in the meaning and significance

of these ideas in the context of the COVID-19 virus. People seem to be overly optimistic about the efficacy of DL/AI in predicting new viruses.

Accurate identification with certainty for the presence of a known microorganism through a digital machine is only possible when the information fed into the machine, before it was employed to identify such a microorganism, is complete. Besides, there should not exist any possibility of emergence of new information to identify that microorganism in addition to the previously noted rules of identification for which a digital machine was developed.

Digital technologies should use quantification of the data at a precise level. That facilitates establishing deep mathematical structures and relations between the data points within an AI framework. The power of predictability of machines for the occurrence of events of interest increases (in this case the event of interest will be ‘predicting the emergence of a new microorganism’) with the power of the accurate and complete data that the machine has in its functionality. To feed complete information into a digital machine is challenging because the scale of the number of species of the microbe and types of microbial species on Earth is enormous [3, 4]. As per the 2016 study in [4], the number of microorganisms on the Earth could be one trillion, and more than 99% of them are still to be discovered. The number one trillion is a statistical estimate and not an exact number.

Suppose we consider the estimated number of microbes of the study [4] or another estimated number from some other yet-to-be-published study as an accurate estimate and let us assume that all of them were discovered. Suppose we list all possible combinations of these microorganisms such that every future combination has some probability to emerge as a new virus. Now let us enumerate these single microorganisms and all possible combinations of microorganisms. Suppose that information on these enumerated organisms was fed into a digital machine for identifying the prevalence or emergence of new viruses that a human or an animal could contract in the future.

This article is part of the Topical Collection on *Mobile & Wireless Health*.

✉ Arni S. R. Srinivasa Rao
arrao@augusta.edu

Steven G. Krantz
sk@math.wustl.edu

- ¹ Health Economics and Modeling, Medical College of Georgia, Augusta, GA, USA
- ² Laboratory for Theory and Mathematical Modeling, Division of Infectious Disease Methodology, Medical College of Georgia, Augusta, GA, USA
- ³ Department of Mathematics, Augusta University, Augusta, GA 30912, USA
- ⁴ Department of Mathematics, Washington University in St. Louis, Campus Box 1146, One Brookings Drive, MO 63130 St. Louis, USA

There are some uncertainties in the previously explained enumeration because it could take thousands of years to discover all existing microorganisms [5]. That means that, if a digital machine is equipped with a currently known list of microbes and such a list is used to match viruses found in humans and animals, then that would certainly miss new and undiscovered microbes (viruses). These would have otherwise been assumed to be enumerated in the above list of all microbes and their combinations. By this current rate of discovery of microbes any digital machine that exists or is under development based on current knowledge of microbes can at most identify existing viruses successfully. Such a possibility of identification of existing viruses is certain only if complete information on the past viruses is available and machines were trained to identify them. We label here such a digital machine which identifies past discovered viruses successfully but failure to identify a future undiscovered microorganism as an *incomplete digital machine*.

The digital machines we classify broadly here as two types: (i) a wearable one that identifies the presence or absence of microorganisms based on causative symptoms and signs, and (ii) the second type of machines that can be used in blood culture tests. Wearable technologies are not confirmatory and those using blood culture are accurate but suffer the incompleteness of the undiscovered information as explained above. For the new pandemics like COVID-19 (Novel Coronavirus Diseases 2019), caused by the novel coronavirus, those digital machines which use the blood culture tests can identify accurately only if the complete information on the microorganism that causes COVID-19 is fed into the machine. Put in other words, the COVID-19 virus cannot be diagnosed by blood culture. (Fig. 1).

Of course, there will still be issues like the sensitivity and specificity of such machines. Wearable digital machines, on

the other hand, are not as accurate as blood culture-based digital machines because wearable machines use the data of symptoms and signs to identify associated viruses in the body.

Microorganisms, we believe, occupy much more dominant positions than they are perceived to have, and they are better situated than the pace at which digital technology is developing. Moreover, digital health technologies once built might generate data dynamically, but the set of data points that it will predict can never be dynamic. All these quantities have to be accountable and the probability of predicting depends on the event of collecting accurate data for a new epidemic. Even to predict a new epidemic well beforehand with a probability less than one (i.e. the non-perfect probability to predict an epidemic), the data to be fed into digital technology is going to be enormous and mostly not available.

In conclusion, digital health technologies only know that which is fed into them, but they would never be able to predict something which was never seen before in experimental labs. But digital health technologies combined with AI will certainly help in reducing the time taken in identifying viruses at the population level if complete information is available on viruses and in drug discovery trials [6–8]. When outcomes of an experiment are finitely many then deep learning algorithms have greater roles in assisting in digital health-related technologies [9]. Although we are skeptical of the successful role of digital health technology in predicting a new virus with certainty in a population, such kinds of technologies are useful in other health care applications; for example, see [10, 11].

Acknowledgements We thank Dr. Michael Diamond, MD, Medical College of Georgia, Augusta, Dr. Claire Adam Spears, Georgia State University, Atlanta for their helpful editorial comments.

Availability of Supporting Data Not applicable.

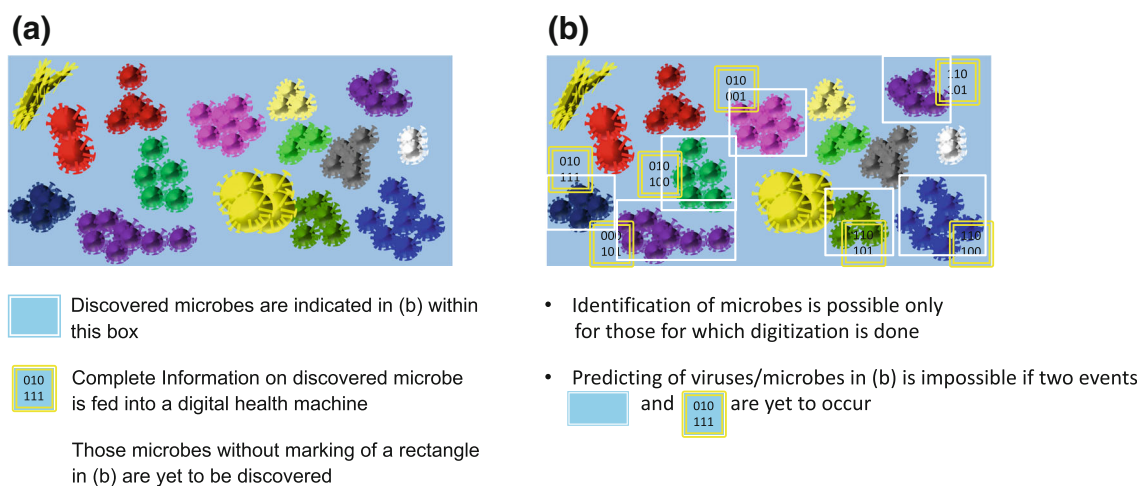


Fig. 1 Identification of viruses through digital technology. (a) All possible microbes on the earth, (b) Discovered and mapped microbes and yet to be discovered microbes

Authors' Contributions Both the authors contributed in writing. ASRS Rao designed the study, developed the methods, created Figure, wrote the first draft. SG Krantz contributed in writing, editing the draft.

Compliance with Ethical Standards

Competing Interests None.

Ethical Approval and Consent to Participate “Not applicable”.

Consent for Publication Not applicable.

Abbreviations DL, Deep-learning; AI, Artificial Intelligence; COVID-19, Novel Coronavirus Disease 2019

References

- Ivanenkov, Y. A., Zhavoronkov, A., Yamidanov, R. S., Osterman, I. A., Sergiev, P. V., Aladinskiy, V. A., et al. (2019). Identification of novel antibacterials using machine learning techniques. *Front. Pharmacol.* 10, 913. <https://doi.org/10.3389/fphar.2019.00913>
- Yang J, Shen C and Huang N (2020) Predicting or Pretending: Artificial Intelligence for Protein-Ligand Interactions Lack of Sufficiently Large and Unbiased Datasets. *Front. Pharmacol.* 11: 69. <https://doi.org/10.3389/fphar.2020.00069>
- Microbiology by numbers. *Nat Rev Microbiol* 9, 628 (2011). <https://doi.org/10.1038/nrmicro2644>
- Kenneth J. Locey, Jay T. Lennon (2016) Scaling laws predict global microbial diversity, *Proceedings of the National Academy of Sciences* May 2016, 113 (21) 5970-5975; <https://doi.org/10.1073/pnas.1521291113>
- A New Way to Find the Tree of Life's Missing Branches <https://www.theatlantic.com/science/archive/2018/01/tree-of-life-microbes/549716/>
- Rao A.S.R.S. and Vazquez J.A. (2020). Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine. *Infection Control & Hospital Epidemiology*, 41(7), 826-830. <https://doi.org/10.1017/ice.2020.61>
- How artificial intelligence is changing drug discovery, <https://www.nature.com/articles/d41586-018-05267-x>
- Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, New York, NY, USA ISBN 1107057132, 9781107057135 (2014).
- Rao, A.S.R.S., Diamond, M.P. (2020). Deep Learning of Markov Model-Based Machines for Determination of Better Treatment Option Decisions for Infertile Women. *Reprod. Sci.* 27, 763–770. <https://doi.org/10.1007/s43032-019-00082-9a>
- Emma Rich & Andy Miah (2017) Mobile, wearable and ingestible health technologies: towards a critical research agenda, *Health Sociology Review*, 26:1, 84-97, <https://doi.org/10.1080/14461242.2016.1211486>
- Gay V, Leijdekkers P Bringing Health and Fitness Data Together for Connected Health Care: Mobile Apps as Enablers of Interoperability. *J Med Internet Res* 2015;17(11):e260 <https://doi.org/10.2196/jmir.5094>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.