Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Emergence of new disease – how can artificial intelligence help?

Yurim Park, Daniel Casey, Indra Joshi, Jiming Zhu, Feng Cheng

PII: S1471-4914(20)30125-8
DOI: https://doi.org/10.1016/j.molmed.2020.04.007
Reference: TRMOME 1569

To appear in: Trends in Molecular Medicine

Please cite this article as: Y. Park, D. Casey, I. Joshi, et al., Emergence of new disease – how can artificial intelligence help?, Trends in Molecular Medicine (2020), https://doi.org/10.1016/j.molmed.2020.04.007

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.
Emergence of new disease remains a critical parameter in human health and society. Advances in artificial intelligence (AI) allow for rapid processing and analysis of...
massive and complex data. In this forum article we review recent applications across disease prediction and drug development with a focus on the COVID-19 pandemic.

**Principles of artificial intelligence and machine learning**

Evolution is inevitable and unpredictable. Infectious, rapidly-evolving viruses can inflict a significant burden of morbidity and mortality. Despite huge advances in diagnosis, treatment and prevention of infectious diseases, emergence of new diseases poses an urgent challenge to the global population, stressing the need for better countermeasures. A current example is the novel coronavirus, COVID-19, which had the first outbreak in Wuhan, China and became a global pandemic rapidly[1].

To minimize the negative impact of an outbreak, it is clear that rapid processing and analysis of relevant data is critical. This data can be fractionated into molecular, patient, population and society levels, all of which contribute to successful treatment and prevention. Yet, this task has been very difficult due to its enormity and complexity.

However, the growing power of artificial intelligence (AI) can provide a solution in obtaining biological insights into a new viral strain and managing new outbreaks. AI is a broad scientific discipline in which human intellectual processes are mimicked by a computer-based system[2]. Machine learning, a sub-discipline of AI, develops systems with the ability to learn from examples in data using statistical models without explicit programming[3]. Moreover, deep learning is a subset of machine
learning that employs artificial neural networks. These are computer systems which model neuronal bundles. As neuronal connections strengthen with reinforcement, the machines learn by processing training examples and find associations in data.\(^1\) Al’s great advantage is the ability to process the enormous amount of medical and biological data being produced daily, a task which otherwise would not be possible due to sheer scale and complexity. By strategically applying this technology, humans can prepare and respond to the threat of newly emergent diseases.

**Applications in emerging diseases**

The power of these approaches has been harnessed in a wide range of applications across public health, disease prediction and drug development. From a public health perspective, these methods have been used in several domains. Applications include analysing real-time data for disease detection, ML-based disease risk models and increasing the efficiency of health systems as a whole by modelling human behaviour[4].

Some of the most interesting work focuses on predicting viral mutation before a new strain has even emerged. A program developed by Salama’s group could predict nucleotide substitutions in primary RNA sequences of the Newcastle virus (Avian pneumoencephalitis) using a rough set gene evolution technique[5]. Rough set theory is a theory of data analysis which has emerged as an important tool for processing inconsistent and imprecise information, especially in AI applications[6].

Training examples were produced, with RNA sequences of one generation of a virus acting as inputs and the RNA sequences of the next generations acting as outputs.
The algorithm was developed to predict RNA sequences of successive generations and then the results compared to actual RNA sequences. The group applied the algorithm to datasets of the Newcastle disease virus in South Korea and China, and the accuracy in predicting the mutated nucleotides was about 70%[5].

When a viral sequence from predicted or occurred mutations is available, the next task is to identify potential viral proteins from the sequence which can predict pathogenic consequences for the host. Representing an important advance in structural biology, computer scientists at DeepMind, a UK technology company, developed a new protein structure prediction system called AlphaFold using AI methodology. The program trains a neural network to predict the distances between protein residues within a sequence. It constructs a mean potential force that gives average forces on the residues depending on the predicted distances. Then, after applying an optimization algorithm, the system folds the protein into its calculated structure[7]. Such structure predictions could lead to a much wider availability of structural information compared to older techniques, and may be especially useful where no experimentally determined homologous protein structures are available[8, 9].

When viral structures and their functions are inferred, drug development is the next important task. Drug discovery is expensive and time-consuming, but AI can shorten the process and increase efficiency using neural networks[10]. A novel method for de novo compound design combines deep learning with reinforcement learning, which estimates the statistical relationship between possible actions and outcomes. Neural networks are trained to generate chemically feasible compounds and predict
their chemical properties. Then, using reinforcement learning, the program becomes biased towards compounds with desired physical and biological properties[11]. Taken together, these advances are changing pharmaceutical medicine (figure 1). AI is poised to change the entire pharmaceutical lifecycle, from regulatory procedures to pharmacovigilance.

**Current paradigms in relation to COVID-19**

How can AI be applied to the current pandemic of COVID-19 and in the future to prepare for the next pandemic? The first main strand has been to mine existing data for insights into how to treat the disease. Benevolent, a drug discovery company, used their AI system to search through an enormous quantity of medical datasets and identify already approved drugs that can be used for the current epidemic.

The second has been in accelerating our understanding of viral structures. AlphaFold recently released structures of proteins associated with COVID-19. Understanding structure can accelerate the process of drug development. Insilico, a Hong Kong-based startup, used ML to identify a 3C-like protease critical for COVID-19’s reproduction. Over 4 days, they came up with new compounds that can block the target’s function by mimicking its structure.

Excitingly, the third strand of applications can be on viral mutation prediction and pre-empting the next generation of viral disease. Researchers sequenced the genome of several members of the same family infected with the COVID-19 and found viral mutation during person-to-person transmission[12]. Mutations can imply increases in virulence, the evasion of the host immune system and the evolution of
resistance to antiviral treatments. Human coronavirus was first identified in 1960s and it has undergone several generations. Using the evolved viral sequences for the past decades and newly sequenced COVID-19 samples with ongoing mutations, researchers can use prediction systems to come up with potential sequences in the next generation. Given the sequence, systems such as AlphaFold can provide information about abnormally folded proteins as a result of viral mutation. By training a ML-based system with known viral mutations in certain sequence regions and their effects on viral behaviour, it can foresee a pathogenic aspect of the predicted mutation as well.

**Challenges and the future of AI**

Putting aside practical applications, the ethical and societal implications of these new technologies must also be considered and will require systematic examination – for example, issues around security, privacy and confidentiality[2]. Overall, AI is still in its infancy with regard to widespread applications across healthcare industry. However, the prediction accuracy of AI methodology will be improved with increasing use and will be therefore of more relevance to real-world settings. Recently, Patel and colleagues showed that a deep-learning AI model for diagnostic imaging could be augmented by having a radiologist stationed at key checkpoints where the algorithm has difficulty. So-called ‘human in the loop’ AI may represent medicine’s near future[4].

**Figure 1: Applications of AI in emerging disease.** AI and ML methodologies allow for rapid processing of the vast amount of medical data being produced daily and extraction of key insights across several applications. For disease prediction, viral
mutations can be predicted before new strains emerge. These techniques are additionally enabling new structural predictions and wider availability of structural information. Mining existing datasets will allow for efficient drug repurposing. The entire pharmaceutical lifecycle could potentially be disrupted, from pre-clinical stages to Phase IV studies.

Resources:

i)  https://www.innoplexus.com/blog/how-artificial-intelligence-works/

ii) https://www.vox.com/recode/2020/2/7/21125959/artificial-intelligence-coronavirus-benevolent-ai-treatment

iii) https://fortune.com/2020/02/06/ai-identifies-possible-coronavirus-treatment/

References

1. Jin, Y.H. et al. (2020) A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil Med Res 7 (1), 4.
2. Davenport, T. and Kalakota, R. (2019) The potential for artificial intelligence in healthcare. Future Healthc J 6 (2), 94-98.
3. Panch, T. et al. (2018) Artificial intelligence, machine learning and health systems. J Glob Health 8 (2), 020303.
4. Patel, B.N. et al. (2019) Human-machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2, 111.
5. Salama, M.A. et al. (2016) The prediction of virus mutation using neural networks and rough set techniques. EURASIP J Bioinform Syst Biol 2016 (1), 10.
6. Zhang, Q. et al. (2016) A survey on rough set theory and its applications. CAAI Transactions on Intelligence Technology 1 (4), 323-333.
7. Senior, A.W. et al. (2020) Improved protein structure prediction using potentials from deep learning. Nature 577 (7792), 706-710.
8. Kryshtafovych, A. et al. (2019) Critical assessment of methods of protein structure prediction (CASP)-Round XIII. Proteins 87 (12), 1011-1020.
9. Schaarschmidt, J. et al. (2018) Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. Proteins 86 Suppl 1, 51-66.
10. Mamoshina, P. et al. (2016) Applications of Deep Learning in Biomedicine. Mol Pharm 13 (5), 1445-54.
11. Popova, M. et al. (2018) Deep reinforcement learning for de novo drug design. Sci Adv 4 (7), eaap7885.
12. Wei, X. et al. (2020) Evolutionary perspectives on novel coronaviruses identified in pneumonia cases in China. National Science Review 7 (2), 239-242.