Forum

Emergence of New Disease: How Can Artificial Intelligence Help?

Yurim Park,[1,*] Daniel Casey,[2] Indra Joshi,[3] Jiming Zhu,[4,5] and Feng Cheng[4]

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, the recent applications across disease prediction and drug development in relation to the COVID-19 pandemic are reviewed.

Principles of AI and Machine Learning

Evolution is inevitable and unpredictable. Infectious, rapidly-evolving viruses can inflict a significant burden of morbidity and mortality. Despite great advances in diagnosis, treatment, and prevention of infectious diseases, emergence of new disease poses an urgent challenge to the global population, stressing the need for better countermeasures. A current example is the novel coronavirus, COVID-19, which was first reported in Wuhan, China, and rapidly became a global pandemic [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 in relation to COVID-19, has been very difficult due to the enormity and complexity of the task.

However, the growing power of 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 (ML), 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 ML 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[i]. The great advantage of AI 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 analyzing real-time data for disease detection, ML-based disease risk models, and increasing the efficiency of health systems as a whole by modeling human behavior [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 disease 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 generation acting as outputs. An algorithm was developed to predict RNA sequences of successive generations, and the results were compared with 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 a calculated structure [7]. Such structure predictions could lead to a much wider availability of structural information compared with 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 on 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 pandemic[ii].

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 the reproduction of COVID-19. Over 4 days, they came up with new compounds that could block the function of the target by mimicking its' structure[iii].

Excitingly, the third strand of applications concern viral mutation prediction and pre-emption of the next generation of viral disease. Researchers sequenced the genome of several members of the same

family infected with COVID-19, and found viral mutation during person-to-person transmission [12]. Mutations can imply an increase in virulence, evasion of the host immune system, and evolution of resistance to antiviral treatments. Human coronavirus was first identified in the 1960s and has undergone several mutations. Using evolved viral sequences from the past decades, and recently sequenced COVID-19 samples with newly emerging 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 an ML-based system with known viral mutations in certain sequence regions, and their effects on viral behavior, it can foresee a pathogenic aspect of the predicted mutation as well.
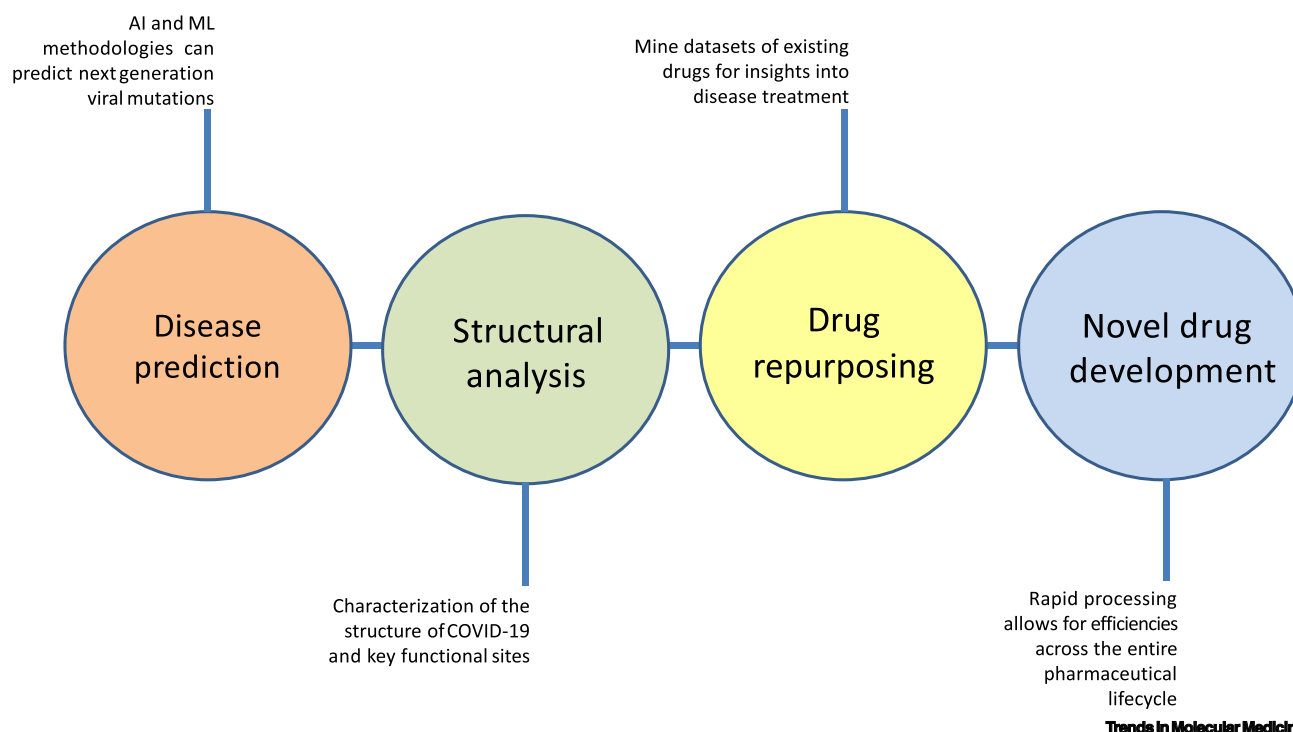

Figure 1. Applications of Artificial Intelligence (AI) in Emerging Disease. AI and machine learning (ML) methodologies allow for rapid processing of the vast amounts of medical data produced daily, and extraction of key insights across several applications. For disease prediction, viral mutations can be predicted before new strains emerge. These techniques also enable 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 redesigned, from pre-clinical stages to Phase IV studies.

Concluding Remarks

Putting aside practical applications, the ethical and societal implications of these new technologies must also be considered, and will require systematic examination, (e.g., issues around security, privacy, and confidentiality) [2]. Overall, AI is still in its' infancy with regard to widespread applications across the healthcare industry.

However, the prediction accuracy of AI methodology will be improved with increasing use, and will therefore be of more relevance in 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].

Resources

[i]www.innoplexus.com/blog/how-artificial-intelligence-works/

[ii]www.vox.com/recode/2020/2/7/21125959/artificial-intelligence-coronavirus-benevolent-ai-treatment

[iii]http://fortune.com/2020/02/06/ai-identifies-possible-coronavirus-treatment/

[1]Radcliffe Department of Medicine, Medical Sciences Division, University of Oxford, Oxford, UK
[2]Innovation Faculty, Celltrion Healthcare, Slough, UK
[3]Digital Health and Artificial Intelligence Laboratory, National Health Service, London, UK
[4]Vanke School of Public Health and School of Medicine, Tsinghua University, Beijing, China
[5]Institute for Hospital Management, Tsinghua Shenzhen International Graduate School, Shenzhen, China

*Correspondence:
yurim.park@st-hugh.ox.ac.uk (Y. Park).

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, 4
2. Davenport, T. and Kalakota, R. (2019) The potential for artificial intelligence in healthcare. *Future Healthc. J.* 6, 94–98
3. Panch, T. *et al.* (2018) Artificial intelligence, machine learning and health systems. *J. Glob. Health* 8, 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, 10
6. Zhang, Q. *et al.* (2016) A survey on rough set theory and its applications. *CAAI Trans. Intell. Technol.* 1, 323–333
7. Senior, A.W. *et al.* (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710
8. Kryshtafovych, A. *et al.* (2019) Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins* 87, 1011–1020
9. Schaarschmidt, J. *et al.* (2018) Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. *Proteins* 86, 51–66
10. Mamoshina, P. *et al.* (2016) Applications of deep learning in biomedicine. *Mol. Pharm.* 13, 1445–1454
11. Popova, M. *et al.* (2018) Deep reinforcement learning for *de novo* drug design. *Sci. Adv.* 4, eaap7885
12. Wei, X. *et al.* (2020) Evolutionary perspectives on novel coronaviruses identified in pneumonia cases in China. *Natl. Sci. Rev.* 7, 239–242