Artificial Intelligence and One Health: Knowledge Bases for Causal Modeling

Nitin Pandit[1*] and Abi T. Vanak[1,2,3]

Abstract | Scientists all over the world are moving toward building database systems based on the One Health concept to prevent and manage outbreaks of zoonotic diseases. An appreciation of the process of discovery with incomplete information and a recognition of the role of observations gathered painstakingly by scientists in the field shows that simple databases will not be sufficient to build causal models of the complex relationships between human health and ecosystems. Rather, it is important also to build knowledge bases which complement databases using non-monotonic logic based artificial intelligence techniques, so that causal models can be improved as new, and sometimes contradictory, information is found from field studies.

1 Background

The recently launched National Mission on Biodiversity and Human Well-Being (NMBH)[1] aims to conserve and restore the rich but rapidly degrading biodiversity of India. Launched by the Prime Minister's Science, Technology and Innovation Advisory Council in 2019, the NMBH is designed to bring together several disciplines which impact and are impacted by biodiversity. Driven by respected research institutions in India, the NMBH is the first step toward developing the science and for building the capacity needed for the integration of biodiversity in the areas of agriculture, disaster management, climate change, bioeconomy, ecosystem services and health. Post its launch, COVID-19 happened, providing a fillip to the component on biodiversity and health, as the role of zoonotic diseases came into limelight[2]. The world over, the scientific community is focused on the emerging trans-disciplinary approaches of One Health, a discipline that characterizes the relationships of biodiversity vis-a-vis human and public health[3].

The glue that binds the components of the ambitious mission is a geospatial database for cataloguing and mapping life (CML). The design of the CML is largely based on the experience of the India Biodiversity Portal (IBP[4]), which has been designed to support researchers and interested citizens in collection and collation of biodiversity related data sets. Concurrently, many other systems for biodiversity data have been created around the world, such as GBIF[5], with applications ranging from species identification[6] to reintroduction[7]. Modern algorithms using big data driven machine learning (ML)[8] and neural networks (NN)[9], coupled with sensors with new capabilities such as bioacoustics[10] and analytical approaches such as genomics[11], are used to complement traditional approaches of biodiversity conservation in situ and in vivo.

Meanwhile, data and models about human health are also becoming increasingly complex, as medical discoveries utilize new computation assisted approaches for health management from prevention to cure for the human body[12]. In fact, biomedical technologies for curing human health ailments are being projected as the next frontier of growth for the global economy toward an ageless generation[13].

2 Human and Public Health Meets Ecosystems

The COVID-19 pandemic has provided an impetus for establishing a closer relation between individual and public health. In

[1] Ashoka Trust for Research in Ecology and the Environment (ATREE), Bangalore 560064, India.
[2] Wellcome Trust/DBT India Alliance Program, Hyderabad 500034, India.
[3] School of Life Sciences, University of KwaZulu-Natal, Durban 4001, South Africa.
*nitin.pandit@atree.org

Springer

looking to quickly tide over this global emergency, the medical community has been spurred on to develop a vaccine to protect the public and reduce individual risk. Whereas a vaccine from the best minds in biomedical research will be welcomed by one and all, public health and biodiversity experts are now under pressure to speed up their work on preventive approaches which include early warning systems, delaying and hopefully even preventing such outbreaks, and if it occurs, better management of such outbreaks.

The existing surveillance apparatus rightly concentrates on early outbreak detection among people, and includes containment and response. While new standards for interoperability[14] are being adopted in India for clinical health of individuals, standards are silent about including causal information, such as wild and domestic animal surveillance for understanding the dynamics of the pathogen-host cycles between outbreaks. Such long-term longitudinal surveillance provides insight into disease burden and helps detect possible predictable patterns in outbreaks at a much lower economic cost than responding after the pathogens emerge[15].

In an attempt to create an integrated mechanism for surveillance, detection and treatment of such zoonoses, a multi-disciplinary engagement in the form of the Roadmap to Combat Zoonoses in India (RCZI) initiative was established in 2008[35]. The RCZI had identified key thrust areas and provided several strategies for research and action. Yet, large-scale and long-term integrated surveillance, involving human, veterinary and wildlife monitoring have failed to materialise[36]. As a consequence, we still lag in our understanding of the burden and dynamics of emerging and re-emerging infectious diseases (ERID).

The Indian government's Integrated Disease Surveillance Project (IDSP), launched in 2004, sought to establish a decentralised state-run India-wide surveillance programme. This programme began with the establishment of surveillance units at the district level, led by a district surveillance officer and a rapid response team to respond to outbreaks. The IDSP has generated clear information flow on outbreaks of 22 conditions and publishes periodic reports of outbreaks on their website[16].

While the outbreak detection and rapid response functions are taken care of by the IDSP, the programme is unable to integrate human and animal (livestock and wildlife) surveillance. This is not surprising given that the IDSP is structured within the department of health and thus, there is limited scope for convergence with other departments. Independent evaluations of the IDSP have pointed out the need for its strengthening and have identified key limitations in achievement of timely outbreak detection and proactive monitoring of ERIDs[17]. An integrated human and animal surveillance system that collects primary data on disease parameters from people, livestock and wildlife is needed as it will improve our understanding of the dynamics of ERIDs and as well as our response (both locally and also policies).

Globally, there are increasing demands for the establishment of responsive and scientifically sound surveillance systems to better understand the connections between deforestation, wildlife, and pandemic risk[18] and, possibly to predict outbreaks and the spread of ERIDs. Recent reviews of surveillance systems have recognized that these need to be strengthened in developing countries. There is also moderate evidence to suggest that most efforts in strengthening response to zoonoses have been focused on "laboratory capacity and technical training, with relatively little attention given to the collection of field data, particularly at the interface between human and livestock populations"[19].

3 Artifacts: In Silico Models of One Health

The biomedical profession is developing advanced algorithms using machine learning and neural networks to derive hypotheses with strong correlations to enable drug discovery for medicines and vaccines to address human health[20]. The health industry has been captivated by cost savings through efficient transactions and better diagnostic outcomes through the use of artificial intelligence (AI) techniques[21]. In fact, current systems of medical informatics focus on human biology only, with most of the research efforts evolving to solve health problems of the individual[22]. Even in the developed health care systems in the west, the vision of future medical systems does not include much about zoonotic diseases[23]. Some AI techniques are being used to further derive correlations using large data sets for individual human-centric medicine[24].

Meanwhile, there is much to be done to develop proactive, in silico models of One Health for public health related applications for prevention and management of outbreaks. When causal models of outbreaks are known,

718

⌾ Springer ⬚

J. Indian Inst. Sci.|VOL 100:4|717–723 October 2020|journal.iisc.ernet.in

e.g., free-ranging dogs causing zoonotic diseases, targeted management approaches can be designed using modern tools such as agent-based modeling[25]. However, the main difficulty with developing in silico causal models of One Health are founded on the lack of data which can help us characterize the ecosystem of pathogens in which the human is simply one actor, who we tend to focus on. Scientists are calling for the NMBH to create a decentralized, national system of surveillance of zoonotic disease outbreaks[26] which also will collate data about ecosystems and biodiversity, since it is their degradation due to human actions which leads to ERIDs. But is that enough?

In fact, modeling such complex ecosystems requires us to understand the myriad behavioral patterns of pathogens and other actors who possess different contextual mechanisms of problem solving intelligence best described in the "ants on a beach" parable in Herbert Simon's classic 1969 book, *Sciences of the Artificial*,[27]. It is, therefore, quite understandable that research in One Health calls for decades long, painstaking, and heroic efforts to discover causal linkages[28] which can provide sufficient data for deriving correlations with confidence[29], and which then can be used as predictive causal models. Surveillance databases need to be coupled with such causal models in the form of knowledge bases to create useful artifacts, i.e., in silico models of One Health.

4 Reasoning with Incomplete Information

The One Health system for data management is a necessary and immediate requirement to enhance our understanding and for rapid response to outbreaks. When such a data management system is available and continually updated, and if we know a well founded causal "law of nature", we can deduce conclusions from observations. For example:

Causal law: IF all < humans with *Ixodes* tick bites in the US > have < Lyme disease >.

Observation: < Arundhati > is a < human with *Ixodes* tick bite in the US >.

Deduction: THEN < Arundhati > has < Lyme disease >.

Deductive rules are represented by the famous syllogism that:

Causal law: IF all < men > are < mortal >.

Observation: < Socrates > is a < man >.

Deduction: THEN < Socrates > is < mortal >.

However, the complexity of ecosystems and zoonotic diseases rarely present such simple situations for the application of rules of deductive logic. Definitive causal laws of nature simply are not established or well founded. Therefore, the analytical approach will still be reactive in nature and largely dependent on correlations between observations and hypotheses generated by the integration of knowledge from the diverse disciplines such as public health, epidemiology, and biodiversity. The research question is whether knowledge from disparate sources can be captured and utilized to create causal models which, in turn, are capable of generating hypotheses for a proactive response to ERIDs.

Recent developments in ML and NN have proliferated in the data analytics community to solve many complex problems. Similar to traditional time series forecasting methods, ML and NN algorithms work well when there is no dearth of data[30]. Some slight variations in the applications of such algorithms also allow for "learning" and deriving models that fit reality to an acceptable degree[31]. In fact, all such more or less statistical methods allow for deriving causal models from large datasets for which virologists created the metaphor in Fig. 1 to represent problem solving for prediction of occurrence of the Kyasanur Forest Disease (KFD) in India.

That is:

Case $n = 1$:

Observation: IF < KFD Virus > is < Present >

Observation: IF < population > is < Susceptible to KFD >

Observation: IF < Climate and Environment > is < Conducive for KFD >

Observation: IF < Vector Population > is < Present for KFD >

Observation: IF < Susceptible Monkey > is < Present for KFD >

Observation: IF < Arundhati > is < a human in the population >

Observation: IF < Arundhati > has < KFD >

Case n = 2:

Observation: IF < KFD Virus > is < Present >

Observation: IF < population > is < Susceptible to KFD >

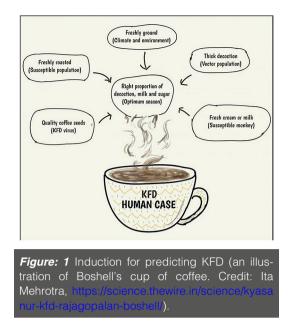Observation: IF < Climate and Environment > is < Conducive for KFD >

Observation: IF < Vector Population > is < Present for KFD >

Observation: IF < Susceptible Monkey > is < Present for KFD >

Observation: IF < Arnab > is < a human in the population >

Observation: IF < Arnab > has < KFD >

719

J. Indian Inst. Sci. | VOL 100:4 | 717–723 October 2020 | journal.iisc.ernet.in   Springer

Figure: 1 Induction for predicting KFD (an illustration of Boshell's cup of coffee. Credit: Ita Mehrotra, https://science.thewire.in/science/kyasanur-kfd-rajagopalan-boshell/).

… and so on for all known humans (or mathematically, as $n \rightarrow$ all members in the population…

Induction: THEN All < humans in the population > have < KFD >

The corresponding syllogism is:
Observation: < Socrates > is a < man >.
Observation: < Socrates > is < mortal >.
Observation: < Plato > is a < man >.
Observation: < Plato > is < mortal >.
Observation: < Aristotle > is a < man >.
Observation: < Aristotle > is < mortal >.
Induction: THEN all < men > are < mortal >.

The rules of inductive logic are not as automatically applicable as the rules of deductive logic. However, when one has statistically representative datasets of the population, inductive rules can enable low-risk reasoning with some predictive capabilities. History is replete with stories of poor, inductive reasoning leading to beliefs which were difficult to revise. Galileo would have agreed.

Perhaps the most interesting case of reasoning for problem solving arises when there is paucity of data. In such cases, problem solving requires that we make hypotheses and test them as we obtain more information. The painstaking gathering of information, leading to incrementally improving hypotheses leads scientists to causal models such as the one developed by scientists working on KFD. The causal models, often represented as directed graphs, show the current state of knowledge based on whatever information is available.

That is:
Causal law: IF all < migratory birds from Russia > have < encephalitis >.
Observation: < KFD > has same origins as < encephalitis >.
Abduction: THEN < KFD > will be in < migratory birds from Russia >.

But, < KFD > could be indigenous! And, in fact, this was the logic that was used in the quest to find KFD, and found to be an erroneous assumption.

Abductive rules are represented by the famous syllogism that:
Causal law: IF all < men > are < mortal >.
Observation: < Socrates > is < mortal >.
Abduction: THEN < Socrates > is a < man >.
But < Socrates > could be a dog!

Abductive reasoning carries significant risk, and can lead to dangerous assumptions which can have subsequent knock-on effects. Furthermore, such hypothetical models carry the inherent risk of being disproved when additional information conflicts with the information gathered to date.

The scientific method essentially incorporates such "abductive" reasoning based on hypothesis testing, and it was in full display in the mystery of the KFD outbreaks which re-emerged after half a century as an ERID in India. Abductive reasoning was applied to develop hypotheses that small mammals on the forest floor could be the reservoirs for KFD and yet again, was proven wrong. Through a process of hypothesis testing, causal chains such as 'small mammal-*Haemaphysalis*-small mammal' chain, the 'small mammal-*Ixodes*-small mammal' chain, and 'small mammal-*Haemaphysalis*-monkey' chain were all eliminated. Before the development of data intense techniques like ML and NN, the science of AI cultivated sophisticated methods[32] to enable building artifacts, i.e., in silico problem solving knowledge bases to emulate such reasoning and support incremental development of causal models.

5 Discussion

The current causal model (Fig. 2) for the re-emergence of KFD was traced to human interventions which reduce biodiversity and provide opportunities for the virus to infest species that they otherwise may not have. The important lesson from the KFD story is that for different types of reasoning to be applied, it is important to develop tools which go beyond simple databases to store and retrieve datasets. It will be important
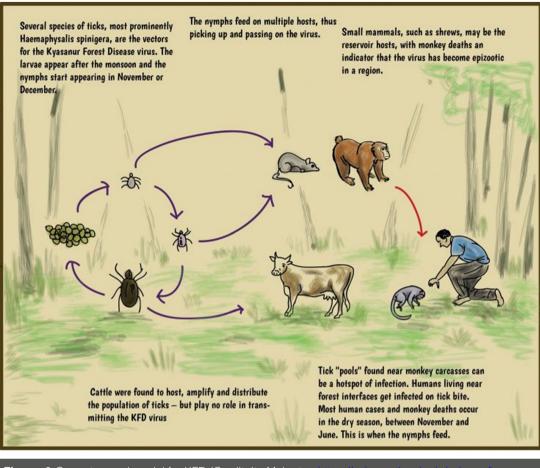
Several species of ticks, most prominently Haemaphysalis spinigera, are the vectors for the Kyasanur Forest Disease virus. The larvae appear after the monsoon and the nymphs start appearing in November or December.

The nymphs feed on multiple hosts, thus picking up and passing on the virus.

Small mammals, such as shrews, may be the reservoir hosts, with monkey deaths an indicator that the virus has become epizootic in a region.

Cattle were found to host, amplify and distribute the population of ticks — but play no role in transmitting the KFD virus

Tick "pools" found near monkey carcasses can be a hotspot of infection. Humans living near forest interfaces get infected on tick bite. Most human cases and monkey deaths occur in the dry season, between November and June. This is when the nymphs feed.

Figure: 2 Current causal model for KFD (Credit: Ita Mehrotra, https://science.thewire.in/science/kyasanur-kfd-rajagopalan-boshell/).

to develop statistical approaches to enable the use of large datasets. But more realistically, it will be important to assist the ecologists, field biologists, epidemiologists, and other scientists with systems which can represent the current state of knowledge, that can be changed as more information is obtained to consolidate and revise the best known models of the time.

Models based on incomplete information can be dangerous. They can set up societal trends that can influence societies in good and bad ways[33]. As the world responds to the COVID-19 crisis with emphasis on health financing[34], it would behoove us to invest in technologies that actually assist One Health scientists in building not only databases, but also their knowledge bases toward prevention and management of zoonotic diseases. Investment in developing such comprehensive artifacts for One Health is the need of the day.

Publisher's Note

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

References

1. https://psa.gov.in/pmstiac-mssions/national-biodiversity-mission. Accessed date 08 Aug 2020
2. https://www.theguardian.com/commentisfree/2020/jul/28/pandemic-era-rainforest-deforestation-exploitation-wildlife-disease?utm_term=677510011d2a929750445fde3e5db9f9&utm_campaign=BestOfGuardianOpinionUK&utm_source=esp&utm_medium=Email&CMP=opinionuk_email. Accessed date 10 Aug 2020
3. https://india.mongabay.com/2020/04/can-biodiversity-loss-lead-to-more-infectious-disease-spread/. Accessed date 29 Jul 2020

721

J. Indian Inst. Sci. |VOL 100:4|717–723 October 2020|journal.iisc.ernet.in

Springer

4. (https://indiabiodiversity.org/. Accessed date 01 Aug 2020

5. https://www.gbif.org/. Accessed date 01 Aug 2020

6. https://www.inaturalist.org/. Accessed date 01 Aug 2020

7. https://www.cbsg.org/integrated-data-management -reintroductions-and-translocations. Accessed 1 Oct 2020

8. https://www.amnh.org/research/center-for-biodiversi ty-conservation/capacity-development/biodiversi ty-informatics/machine-learning-for-conservation. Accessed date 01 Aug 2020

9. https://www.researchgate.net/publication/220704972_ Knowledge_Discovery_using_Artificial_Neural_Netwo rks_for_a_Conservation_Biology_Domain. Accessed date 01 Aug 2020

10. https://conservify.org/. Accessed date 01 Aug 2020

11. https://link.springer.com/article/10.1007%2Fs12041- 019-1159-1. Accessed date 01 Aug 2020

12. https://medium.com/@ideaxme.mail/systems-medic ine-with-dr-leroy-hood-c207e2052e8a. Accessed date 01 Aug 2020

13. https://www.amazon.com/Ageless-Generation-Advan ces-Biomedicine-Transform/dp/0230342205. Accessed date 01 Aug 2020

14. https://go.lyniate.com/blog/fhir-as-explained-by-a- physician . Accessed date 12 Aug 2020

15. https://science.sciencemag.org/content/369/6502/379. full Accessed date 20 Jul 2020

16. https://idsp.nic.in/. Accessed 10 May 2014

17. CDC Evaluation Report—https://idsp.nic.in/Write ReadData/l892s/CDC_Sept07.pdf. Accessed 1 Oct 2020

18. https://www.thegef.org/news/connecting-deforestation- wildlife-and-pandemic-risk. Accessed date 10 Aug 2020

19. https://doi.org/10.1098/rstb.2016.0163. Accessed date 10 Aug 2020

20. https://www.youtube.com/watch?v=G5IiEuXHvk8. Accessed date 01 Aug 2020

21. https://www.youtube.com/watch?v=jZg5QhL3Ckc. Accessed date 03 Aug 2020

22. Adapted from https://www.amazon.com/Systems-Biolo gy-Properties-Reconstructed-Networks/dp/0521859034. Accessed 1 Oct 2020

23. https://www.itl.nist.gov/div897/ctg/it_healthcare/JackC orley2_files/frame.html. Accessed date 12 Aug 2020

24. https://www.nature.com/articles/s41591-018-0300-7. Accessed date 20 Jul 2020

25. (https://www.authorea.com/users/343316/articles/46997 5-modelling-the-challenges-of-managing-free-ranging- dog-populations?commit=bee6875e868203128961adbc9 a2dbd5c277331cf. Accessed date 12 Aug 2020

26. https://www.thehindu.com/sci-tech/science/the-time- is-right-for-onehealth-science/article31069639.ece. Accessed date 15 Jul 2020

27. https://theconversation.com/weve-been-looking-at-ant- intelligence-the-wrong-way-17619#:~:text=In%20his %201969%20book%2C%20The,the%20complexity%20 in%20the%20ant. Accessed date 30 Jul 2020

28. https://science.thewire.in/health/kyasanur-kfd-rajagopala n-boshell/. Accessed on 08/08/2020

29. https://journals.plos.org/plosntds/article?id=10.1371/ journal.pntd.0008179. Accessed on 10 Aug 2020

30. https://towardsdatascience.com/3-facts-about-time-serie s-forecasting-that-surprise-experienced-machine-learn ing-practitioners-69c18ee89387. Accessed date 25 Jul 2020

31. https://towardsdatascience.com/a-short-introduction- to-model-selection-bb1bb9c73376. Accessed date 25 Jul 2020

32. https://www.amazon.com/Building-Problem-Solvers- Artificial-Intelligence/dp/0262061570. Accessed date 25 Jul 2020

33. https://www.epw.in/engage/article/man-machine-asoci al-construction-health-0. Accessed 30 Jul 2020

34. https://pib.gov.in/PressReleasePage.aspx?PRID=1637002. Accessed date 25 Jul 2020

35. Sekar N, Shah NK, Abbas SS, Kakkar M, Roop R (2011) Research Options for Controlling Zoonotic Disease in India, 2010–2015. PLoS ONE 6(2):e17120

36. Chatterjee P, Kakkar M, Chaturvedi S (2016) Integrating one health in national health policies of developing coun- tries: India's lost opportunities. Infect Dis Poverty 5(1):2

722

Springer

J. Indian Inst. Sci.|VOL 100:4|717–723 October 2020|journal.iisc.ernet.in

Nitin Pandit is the Director of the Ashoka Trust for Research in Ecology and Environment (ATREE) in Bangalore, India. Previously, Dr. Nitin Pandit was the Director of Priority Initiatives at the World Resources Institute (WRI) in Washington, DC, USA, focusing on restoration and energy efficiency. Prior to this assignment, he was the CEO of WRI India and led WRI's work in India. He was responsible for formulating and implementing WRI India's strategy, including a new program in the restoration of degraded lands. Before WRI, Nitin was President of International Institute for Energy Conservation (IIEC), with offices and programs in a dozen countries, for implementing novel sustainable energy approaches for developing countries, such as market transformation, energy-efficient buildings, and demand-side management using renewable energy hybrids and energy efficiency improvement. In the 90s, Nitin formed a boutique high-tech consultancy specializing in artificial intelligence (AI) applications in environmental and renewable energy systems. Using AI, he developed tools and solutions for integrated "closed-loop" systems of water, energy, and materials, synoptic climatology, air pollution, and non-point source pollution. In the 80s, Nitin worked with reputed consulting firms in the areas of pollution prevention and waste management, geohydrology and geo-technical construction, and water resources engineering. Nitin has a bachelors and couple of masters' degrees in engineering, and a doctorate in public policy.

Abi T. Vanak is a Senior Fellow (Associate Professor), and Convener of the Centre for Biodiversity and Conservation with the Ashoka Trust for Research in Ecology and the Environment (ATREE). He is also a Fellow of the DBT/Wellcome Trust India Alliance Clinical and Public Health Program. His research areas include animal movement ecology, disease ecology, One-Health, savanna ecosystems, invasive species and wildlife in human-dominated systems. Much of his research work focuses on the outcome of interactions between species at the interface of humans, domestic animals and wildlife in semi-arid savannas and agro-ecosystems. Under OneHealth systems and disease ecology, he studies dynamics of rabies transmission in multi-host systems and the role of small and medium mammals in the transmission dynamics of Kyasanur forest disease. Abi Vanak has a Master's in Wildlife Biology from the Wildlife Institute of India and a Ph. D. in Wildlife Science from the University of Missouri.

723

J. Indian Inst. Sci. |VOL 100:4|717–723 October 2020|journal.iisc.ernet.in