by the interplay of competing warming and cooling processes. Since 1850, there has been a net warming owing to greenhouse-gas emissions from human activity[4]. Over this time, the global mean temperature has increased[5] by 0.9 °C, although the warming caused by greenhouse-gas emissions has been partially compensated for by the cooling effect of aerosol emissions. Therefore, a precise quantification of this cooling could have profound implications for projections of future climate.

Clouds that form in the presence of high aerosol concentrations contain droplets that are smaller and more numerous than usual (Fig. 1). These droplets therefore have a large total surface area for sunlight to bounce off. Consequently, the reflectance of polluted clouds is greater than that of unpolluted ones[1]. As a result of the aerosol emissions of the early twenty-first century, as opposed to pre-industrial conditions, this enhanced reflectance generates a substantial cooling effect on Earth's climate[6].

Whether cloud reflectance is increased or decreased by changes in water content, and to what extent, has been highly uncertain. A greater water content in polluted clouds than in unpolluted ones could enhance the net cooling effect of aerosol emissions[2]. This possibility is suggested by many global climate models[7]. Some scientists have argued that the increase in cloud water content, and the associated cooling effect, caused by aerosols might be even larger than these models indicate[8]. By contrast, other evidence suggests that there could be considerably less water in polluted clouds than in unpolluted ones, which would reduce the net cooling effect substantially[9].

To address this uncertainty, Toll and colleagues looked at features of polluted clouds called pollution tracks (see Fig. 1 of the paper[3]). These features were produced downwind of sources of human-made aerosols such as coal-fired power plants, oil refineries, smelters, cities, ships and wildfires. Like the cloud trails that form behind aircraft at high altitudes, these pollution tracks in low-level clouds are visible from space. As a result, polluted and less polluted cloudy regions can be clearly distinguished. Observed changes in droplet size or cloud water content can, therefore, be unequivocally attributed to variations in aerosol concentrations.

Using 15 years of high-resolution satellite data of near-global coverage, the authors built an unprecedented database of thousands of such tracks across Earth's climate zones. Overall, they found that the average droplet size was at least 30% lower in polluted clouds than in unpolluted ones. Although differences in cloud water content varied, the mean water content was slightly lower in polluted clouds than in unpolluted ones (Fig. 1). This finding suggests that the effect of aerosols on cloud water content slightly reduces the overall aerosol-induced increase in cloud reflectance.

Toll *et al.* then extrapolated their findings to all low-level clouds across Earth, considering global changes in aerosol emissions from human activity. They estimate that the identified decrease in cloud water content offsets only 23% of the net cooling effect caused by the reduction in droplet size. However, the precise estimate remains uncertain. Although the authors sampled thousands of pollution tracks, these features are scarce. For example, it is extremely rare for a ship to leave a pollution track in its wake[10], and probabilities of track generation for the other sources of human-made aerosols are likely to be similarly low. This rarity raises the question of whether observations made using pollution tracks can be generalized to all other conditions in which pollution tracks are not seen.

The most common hypothetical situations in which pollution tracks are not identified are: when clouds are already bright, so that added aerosols have no impact on reflectance; and when cloud properties are rapidly varying because of changes in humidity, stability or horizontal winds. The aerosol-induced decrease in cloud water content might therefore be smaller or larger than is estimated from pollution tracks. However, there is no a priori reason for the clouds to respond in a fundamentally different manner in conditions in which pollution tracks are not observed. Toll and colleagues' work therefore strongly suggests that the sensitivity of cloud water content to changes in the concentration of human-made aerosols might not be accurate in many current global climate models, and that large cooling effects caused by variations in cloud water content are unlikely. ■

**Anna Possner** *is at the Institute for Atmospheric and Environmental Sciences, Goethe University, 60438 Frankfurt, Germany.*
*e-mail: apossner@iau.uni-frankfurt.de*

1. Twomey S. *Atmos. Environ.* **8**, 1251–1256 (1974).
2. Albrecht, B. A. *Science* **245**, 1227–1230 (1989).
3. Toll, V., Christensen, M., Quaas, J. & Bellouin, N. *Nature* **572**, 51–55 (2019).
4. Myhre, G. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 659–740 (Cambridge Univ. Press, 2013).
5. Hartmann, D. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 161 (Cambridge Univ. Press, 2013).
6. McCoy, D. T. *et al. J. Geophys. Res. Atmos.* **122**, 1779–1796 (2017).
7. Ghan, S. *et al. Proc. Natl Acad. Sci.* USA **113**, 5804–5811 (2016).
8. Rosenfeld, D. *et al. Science* **363**, eaav0566 (2019).
9. Gryspeerdt, E. *et al. Atmos. Chem. Phys.* **19**, 5331–5347 (2019).
10. Campmany, E., Grainger, R. G., Dean, S. M. & Sayer, A. M. *Atmos. Chem. Phys.* **9**, 1899–1905 (2009).

MEDICAL RESEARCH

# Deep learning detects impending organ injury

**Organ damage is often detected late, when treatment options are limited. The use of artificial intelligence to continuously monitor a patient's medical data can identify people at risk of imminent kidney injury.** **See Letter p.116**

ERIC J. TOPOL

Acute injury to the kidneys occurs in one in five patients in US hospitals[1]. It is a common condition in hospital patients because it can be caused by a number of factors, including abnormal blood pressure or blood volume. But the ability to predict whether or when acute kidney injury will happen is limited. For people who are at high risk of developing this condition, the standard clinical approach is daily assessment of their laboratory test results, including the concentration of creatinine in their blood, because high levels of this molecule are a hallmark of kidney problems.

Tomašev *et al.*[2] report on page 116 that an approach involving artificial intelligence makes it possible to identify impending acute kidney injury, for most patients, one or two days before the condition would be diagnosed using standard clinical tests. Kidney injury is usually spotted only at a late stage, when irreversible damage has occurred that could lead to death or the need for temporary or long-term dialysis. Being able to catch the condition early would be a major step forward in enabling effective treatment.

In the artificial-intelligence method known as deep learning, an algorithm is developed to identify patterns in the data that are associated with an outcome of interest — in this case, the development of acute kidney injury. The authors used this approach on data collected between 2011 and 2015 from more than 700,000 adults treated in 172 hospitals and 1,062 outpatient clinics run by the US Department of Veterans Affairs — a health-care provider for military workers and their families. The anonymized information

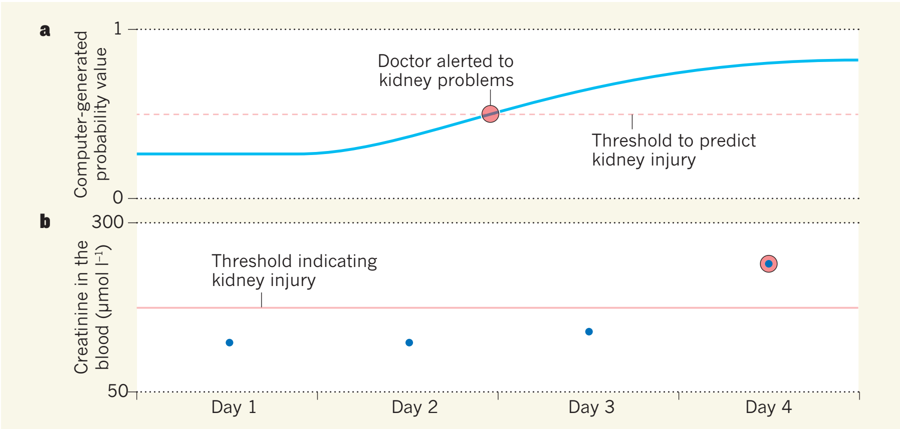


**Figure 1 | Predicting kidney malfunction.** **a**, Tomašev *et al.*[2] report a way of providing advance warning that patients will develop acute kidney injury. The authors used an artificial-intelligence approach called deep learning to train a computer to detect patterns associated with subsequent kidney injury. This computer-based approach was then used to analyse previously collected medical data that included a range of information such as electronic health records and laboratory test results. When such data were assessed for individual patients, as for the hypothetical patient shown here, the computer continuously generated a probability of kidney injury occurring within 48 hours. If this probability exceeded a threshold value, the prediction was considered positive, alerting the doctor (red circle). **b**, A standard way to monitor those at risk of developing kidney problems is to track a daily measurement of the level of creatinine in the blood (in micromoles per litre). For this hypothetical patient, Tomašev and colleagues' approach would provide a warning of imminent kidney problems earlier than is possible by tracking this molecule.

provided the authors with data for these individuals that included: demographics, electronic health records, laboratory test results, medications prescribed and records of procedures undergone. Tomašev and colleagues used these cumulative data to train their computer by running a time-series analysis of around 6 billion data points and more than 600,000 recorded features. They chose a method for deep learning called a recurrent neural network, which is ideal for assessing sequential data inputs that are obtained over time.

The authors tested the system using data for individual patients that had been set aside for this purpose. They obtained computer-generated probability values that continuously traced the likelihood over time that any individual would develop acute kidney injury within the next 48 hours. If the probability exceeded a threshold value, the prediction was considered positive (Fig. 1). Checking whether the patient was subsequently diagnosed with the condition revealed the accuracy of the prediction. The authors' model also provided an indication of the level of uncertainty for the probability value, enabling a doctor to assess the strength of the predictive signal.

Tomašev and colleagues' approach is more accurate than other statistical or machine-learning methods that have been proposed for identifying impending kidney damage[3,4]. As might be expected, the prediction accuracy of the authors' system was highest for people in hospital settings, where acute kidney injury is more frequent, has a more rapid onset and occurs within a shorter window of time than is typical in outpatient clinics. For all patients and any type of acute kidney injury, including less severe forms, the system was 56% accurate. Successful predictions for more serious forms of the condition were 84% and 90% for people who subsequently required dialysis treatment within 30 and 90 days, respectively. The model's accuracy was similar across the different health-care sites and throughout the time period studied.

The authors used a method called ablation analysis to determine the factors linked to the risk of developing acute kidney injury. They found many contributory factors, which might explain why trying to determine this risk has been a vexing task in the past.

Consideration of the case of a hypothetical patient (Fig. 1) underscores the potential usefulness of the system developed by Tomašev and colleagues. This patient's daily creatinine values gave no indication of acute kidney injury until their fourth day in hospital. By contrast, the authors' system predicted organ damage two days earlier, giving more time for treatment interventions such as increasing the patient's fluid intake, or avoiding the use of drugs that could cause kidney toxicity.

However, the authors' system generated many 'false positive' predictions — predictions of injury that did not occur. For each accurate prediction, there were two false positives. Most of these occurred in people who had chronic kidney disease, which would make superimposed acute kidney injury more difficult to predict.

A limitation of the authors' work is that it is a retrospective study. There are examples of the use of artificial intelligence in retrospective studies of medical data in which the model's accuracy declined when it was tested prospectively[5]. Such declines probably occur because dealing with data in a real-world clinical environment is more complicated than dealing with a 'cleaned', pre-existing data resource.

Prospective studies are essential for determining the true clinical value of a predictive system. Moreover, successful prediction is not the only factor that should be assessed. One way to determine whether these predictive warnings result in a reduction in acute kidney injury would be to carry out a clinical trial using a randomized design, in which only half of the predictions of impending injury are relayed to doctors. The authors' model should also be tested to determine how well it works in other groups of patients. Moreover, less than 7% of Tomašev and colleagues' study group were women. Whether the model's ability to predict acute kidney injury differs depending on gender thus needs further investigation.

Although the authors' system included a variety of data types, other data sources might also be valuable for inclusion. For example, it is possible that written notes in medical records, or continuous monitoring of vital signs, such as heart rate, from wearable sensors, might provide relevant information.

For patients who are not in an intensive-care unit, the standard monitoring approach is to take their vital signs once daily. However, all too often, the daily doctors' rounds can reveal a patient who has suddenly become critically ill. Tomašev and colleagues' study shows the benefit of being able to anticipate serious organ damage well before it occurs. Most predictive studies using artificial intelligence in a clinical context have previously focused on patient outcomes such as deaths, readmissions or the time spent in hospital[6]. The work by Tomašev *et al.* stands out by providing a prediction that might enable effective clinical intervention.

The use of deep learning has considerable promise as a way of alerting doctors to concerns about any organ. Its implementation will probably require a change in the medical mindset. But moving from infrequent, one-off tests to relying more on systems that allow continuous assessment might provide a better way of predicting what lies ahead for a patient. ■

**Eric J. Topol** *is in the Department of Molecular Medicine, Scripps Research Translational Institute, La Jolla, California 92037, USA.*
*e-mail: etopol@scripps.edu*

1. Wang, H. E., Muntner, P., Chertow, G. M. & Warnock, D. G. *Am. J. Nephrol.* **35**, 349–355 (2012).
2. Tomašev, N. *et al. Nature* **572**, 116–119 (2019).
3. Mohamadlou, H. *et al. Can. J. Kidney Health* **5**, https://doi.org/10.1177/2054358118776326 (2018).
4. Pan, Z. *et al.* Preprint at https://arxiv.org/abs/1901.04364 (2019)
5. Keane, P. A. & Topol, E. J. *NPJ Dig. Med.* **1**, 40 (2018).
6. Rajkomar, A. *et al. NPJ Dig. Med.* **1**, 18 (2018).

**The author declares competing financial interests. See go.nature.com/2jwac69 for details.**