Machines that ape brain cells

THE CURRENT CAPABILITY IN MASSIVE COMPUTING CAN SPEED UP THE PROGRESS OF DRUG DESIGN, WRITES S ANANTHANARAYANAN

Knowing the biochemistry of the body is one way of designing therapy to deal with medical conditions. Most known drugs or herbal remedies, however, were not discovered in this way but after centuries of trial and error. Modern medicine has analysed traditional cures to arrive at the functional chemicals, which has helped drug synthesis and also to understand physiology. Such wisdom, however, does not help the development of new drug remedies, which must rely on intense and largely, blind assays of the great variety of preparations or substances and their effects on pathogens or body processes.

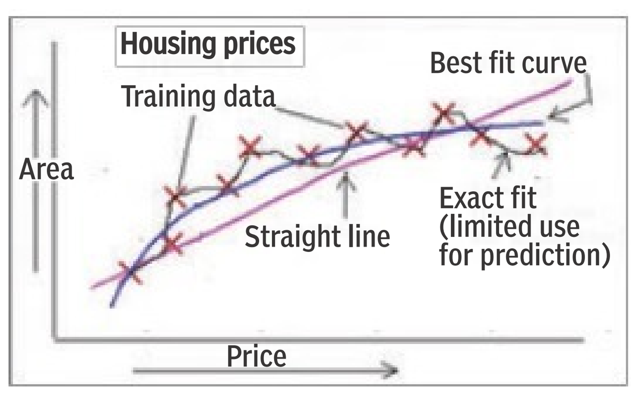
A way to accelerate the process has been the use of machine learning and computer-based artificial intelligence, which can find patterns in the features of known effective substances and screen out the less promising lines of inquiry. An improvement in this strategy, which was used to scan 72 million compounds and create a tractable list of promising candidates for anti-cancer application, has been reported by researchers Artur Kadurin, Alexander Aliper, Andrey Kazennov, Polina Mamoshina, Quentin Vanhaelen, Kuzma Khrabrov and Alex Zhavoronkov, from Moscow, Tatarstan, St Petersburg and Dolgoprudny in Russia, Truro, in Cornwall, the University of Oxford and John Hopkins University.

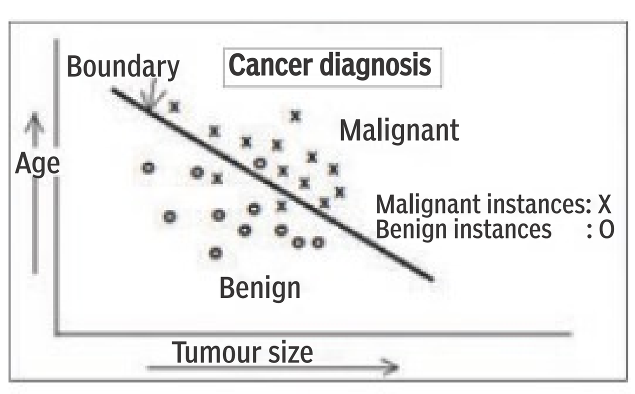
Their paper in the journal, *Oncotarget*, begins with the words, "despite the many advances in biomedical sciences, the productivity of research and development programmes in the pharmaceutical industry is on the decline." It explains that nearly 90 per cent of clinical trials for all disease categories end in failure, and almost 95 per cent fail in the case of cancer. The main reason, the paper says, is that trials have to start without a hint of where to look and the huge effort needed leads to high prices of cancer drugs.

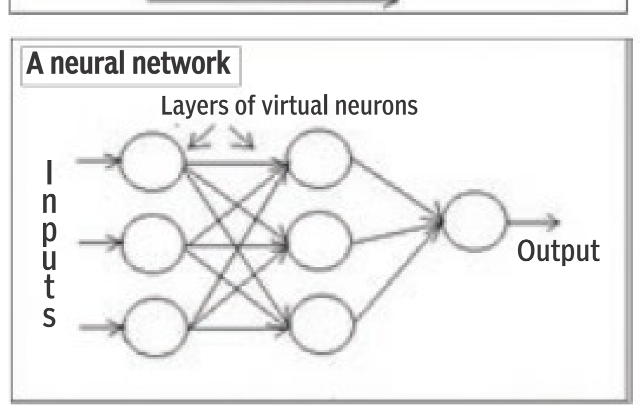
The methods of machine learning have been strikingly successful in trend analysis, based on just some known examples. Areas of application include analyses of customer trends, improving availability of products and services, traffic control, health administration and automated diagnosis. Machine learning is now capable of voice and image recognition, even surpassing human ability, and has been successful in running driverless motor vehicles in a busy street.

The basic method used is to make rapid calculations with a set of known data to find a mathematical formula that fits their distribution. The formula is then tested on some more known examples and if it qualifies, then it is likely to make correct predictions with unknown data too. An elementary example, from a popular course on machine learning that is offered online by Stanford University, is to consider the prices of apartments as a function of the covered area, the number of rooms, bathrooms, windows, location, et al. Computer-aided analysis could discover a relationship among many of these, which could predict the price of apartments, given its features, to help sellers or buyers.

Another application would be to connect the







tumour size, age of the patient and some other features, in the case of cancer, with malignancy. The formula, as developed from known cases, could be regulated to predict malignancy either in most malignant cases, or rarely to make a mistake, depending on the priority.

While predictions like this were first made purely by varying and refining the formulae used — a process which is possible with large data and the help of computers — it was soon realised that the animal brain seems to perform a lot better through a different angle of attack. In playing a game of chess, for instance, one

way a computer could go about it is by working out all possible moves and counter-moves, by both players, from a given board position, and then choosing what move to make. It was seen however, that human players were able to do better than even powerful computers that planned several moves ahead.

It was then realised that the human brain did not follow the brute force method of the computer, but seemed to take in some features of the chess board position, which may seem to be unrelated, and use these, and experience, to play in a more effective way. And the mechanism by which the brain did this was by charting the manner in which arrays of brain cells reacted to the different inputs of features of the chess board. Given a set of responses by brain cells, the same responses were either strengthened or weakened, depending on their outcome. Over a series of actual instances, the brain adapts to making more effective responses and continues to learn with experience. This is the mechanism that leads not only to good chess playing but also how a child internalises the nuances of a language faster than years of study by scholars.

Computers were now programmed to simulate this architecture by creating virtual neurons, or software that behaved like brain cells. In a simple instance of recognising a single feature, the feature could be presented to a single virtual neuron. The neuron responds at random, from a pair of choices. If the answer is correct, there is feedback that adds to the probability of that response, and if the answer is wrong, the feedback lowers the probability. We can see that this device would soon learn, through a random process, to consistently make the correct response.

We can also see that the same process could deal with a choice of more than only two responses and we could also have a brace of artificial neurons that would send their responses to another set of neurons, and so on. By creating this kind of network, and sending feedback from layer to layer, it is possible to develop a computer system that can identify an image as being that of a car or a pedestrian, for instance, and then if the pedestrian was a man or a woman.

In applying this process to drug design, a large number of features of many candidate substances, whose therapeutic value is known, is presented to neural networks, to generate responses of whether the substances are useful in cancer therapy or not. The feedback of whether the responses were correct is the process by which the network "learns" the importance of different combinations of features. Large computer facilities are able to simulate extensive neural networks and train the system with a range of features, to filter out substances of doubtful potential and reduce the load at the final stages of drug research.

While these applications of machine learning deal with classification, even discovering classes, or discrimination, other applications are of generating instances that should lie within a given or discovered class. These applications are useful in image generation, synthesising a texture or cleaning signals of noise. An application in drug research would be to go from characteristics of known curative agents to the specifications for new drugs.

The authors of the *Oncotarget* paper made use of available databases of drug characteristics and an advanced machine learning technique where one part of the programme tries to generate new instances that cannot be distinguished from a base set while another part (as an "adversary") works to uncover these generated instances. The result is an optimum of similarity with novelty, or new and effective drug templates.

The paper describes a trial where the system was trained with the features of 6,252 compounds that were effective against *MCF-7*, a recognised line of breast cancer cells. The system was then used to screen 72 million compounds from a freely accessible database of compounds and small molecules. The properties of the resulting set of 69 compounds, culled out of the 72 million, were then assessed with the help of another database of compounds that had anti-cancer capability. It was found that several of the 69 compounds were known as anti-cancer agents and most were related to a group of highly effective cancer drugs.

This is a significant achievement, that a computer trial has isolated a biologically relevant subset out or a very large collection, an exercise that would not have been possible through laboratory methods. While nature, after centuries of evolution, provides us with a nearly infinite variety of substances, we now have a machine learning procedure that helps us sift through the mass to find those that serve specific purposes.

THE WRITER CAN BE CONTACTED AT response@simplescience.in

CHROMOSOME NUMBERS

TAPAN KUMAR MAITRA EXPLAINS THE LIFE CYCLES OF SEXUAL ORGANISMS

Since gametes are haploid, they cannot be produced from diploid cells by mitosis because mitosis creates daughter cells that are genetically identical to the original parent cell. In other words, if gametes were formed by mitotic division of diploid cells, both sperm and egg would have a diploid chromosome number, just like the parent diploid cells. The hypothetical zygote created by the fusion of such diploid gametes would be tetraploid (possess four homologous sets of chromosomes). Moreover, the chromosome number would continue to double for each succeeding generation - an impossible scenario. Thus, for the chromosome number to remain constant from generation to generation, a different type of cell division must occur during the formation of gametes. That special type of division, called meiosis, reduces the chromosome number from diploid to haploid.

Meiosis involves one round of chromosomal DNA replication followed by two successive nuclear divisions. This results in the formation of four daughter nuclei (usually in separate daughter cells) containing one haploid set of chromosomes per nucleus. The principle of meiosis starts with a diploid cell containing four chromosomes. A single round of DNA replication is followed by two cell divisions, meiosis I and meiosis II, leading to the formation of four haploid cells.

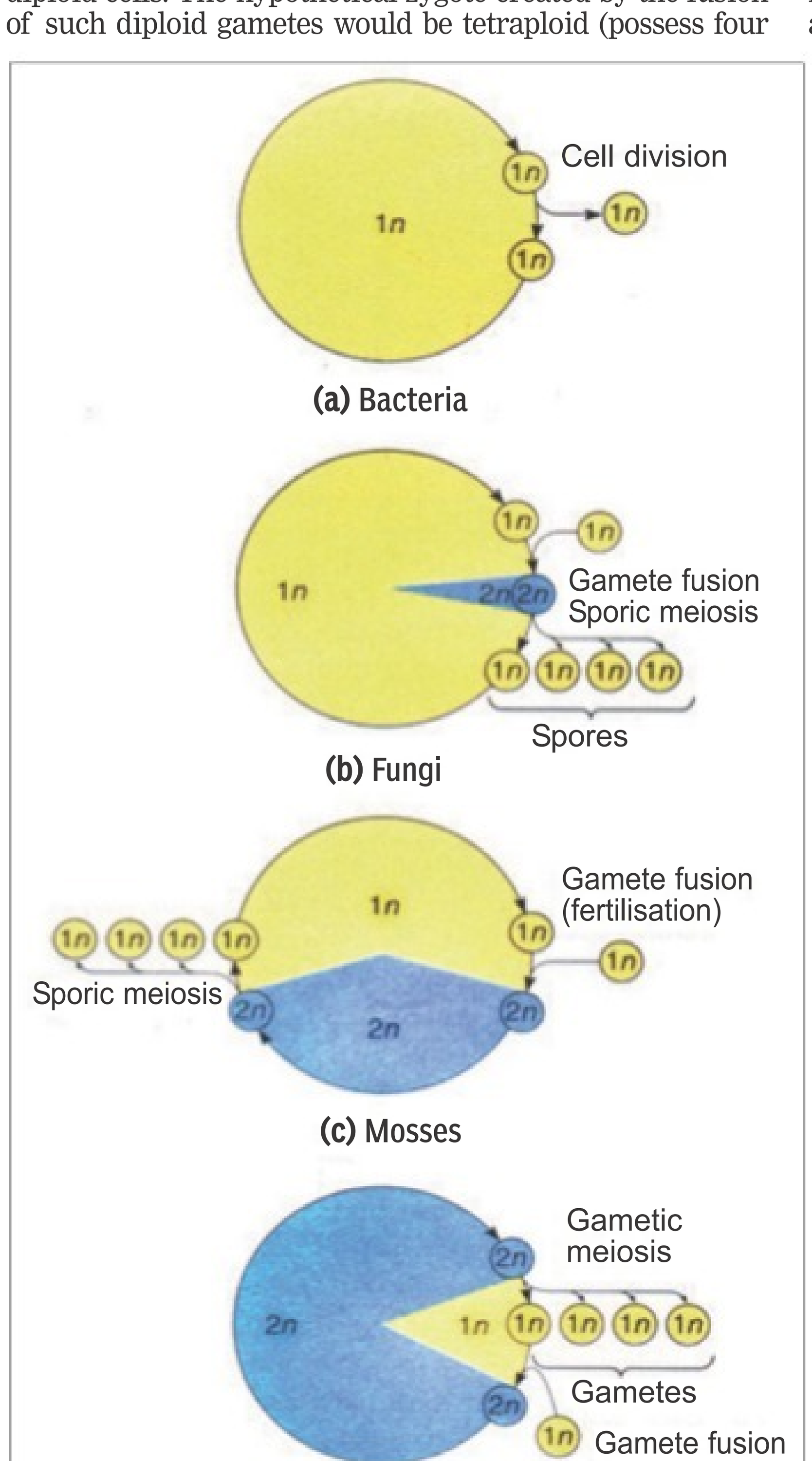
Meiosis and fertilisation are indispensable components of the life cycle of every sexually reproducing organism, because the doubling of chromosome number that takes place at fertilisation is balanced by the halving that occurs during meiosis. As a result, the life cycle of sexually reproducing organisms is divided into two phases — a diploid (2n) phase and a haploid (1n) phase. The diploid phase begins at fertilisation and extends until meiosis, whereas the haploid phase is initiated at meiosis and ends with fertilisation.

Organisms vary greatly in the relative prominence of the haploid and diploid phases of their life cycles. Fungi are examples of sexually reproducing organisms whose life cycles are primarily haploid but include a brief diploid phase that begins with gamete fusion (the fungal equivalent of fertilisation) and ends with meiosis. Meiosis usually takes place almost immediately after gamete fusion, so the diploid phase is very short, and, accordingly, only a very small fraction of fungal nuclei are diploid at any one time. Fungal gametes develop, without meiosis, from cells that are already haploid.

Mosses and ferns are probably the best examples of organisms in which both the haploid and diploid phases are prominent features of the life cycle. Every species of these plants has two alternative, morphologically distinct, multi-cellular forms — one haploid and the other diploid. For mosses, the haploid form of the organism is larger and more prominent, and the diploid form is smaller and more short-lived. For ferns, it is the other way around. In both cases, gametes develop from pre-existing haploid cells.

Organisms that alternate between haploid and diploid multi-cellular forms in this way are said to display an alternation of generations in their life cycles. In addition to mosses and ferns, eukaryotic algae and other plants exhibit an alternation of diploid and haploid generations. In all such organisms, the products of meiosis are haploid spores, which, after germination, give rise by mitotic cell division to the haploid form of the plant or alga. The haploid form in turn produces the gametes by specialisation of cells that are already haploid.

On the other hand, the best examples of life cycles dominated by the diploid phase are found in animals. In such organisms, including humans, meiosis gives rise not to spores but to gametes directly, so the haploid phase of the life cycle is represented only by the gametes. Meiosis in such species is called gametic meiosis to distinguish it from the sporic meiosis observed in spore-producing organisms exhibiting an alternation of generations. Meiosis is thus gametic in animals and sporic in plants.



The relative prominence of the haploid (1n) and diploid (2n) phases of the life cycle differ greatly, depending on the organism, (a) Bacteria exist exclusively in the haploid state, (b) Many fungi exemplify a life form that is predominantly haploid but has a brief diploid phase. (c) Mosses (and ferns as well) alternate between haploid and diploid forms, both of which are significant components in the life cycles of these organisms. (d) Higher animals are the best examples of organisms that are predominantly diploid, with only the gametes representing the haploid phase of the life cycle.

THE WRITER IS ASSOCIATE PROFESSOR, HEAD, DEPARTMENT OF BOTANY, ANANDA MOHAN COLLEGE, KOLKATA, AND ALSO FELLOW, BOTANICAL SOCIETY OF BENGAL, AND CAN BE CONTACTED AT tapanmaitra59@yahoo.co.in

Message in a starry bottle

SCIENTISTS ARE SET TO BEGIN SPEAKING TO ALIENS DESPITE WARNINGS THAT IT COULD TRIGGER EARTH'S DESTRUCTION. ANDREW GRIFFIN REPORTS

A major programme to try and speak to aliens has been launched — despite warnings that it could lead to the destruction of humanity.

Meti stands for Messaging Extra-terrestrial Intelligence and is a new project that will look to send out signals into space so that extra-terrestrials can pick them up. It will start doing that from 2018 — allowing humanity to get in touch with aliens, rather than waiting for them to talk to us.

But scientists have long warned that sending out such messages could be dangerous, and that alerting aliens to our existence might lead to our complete destruction. That has

A car moves along the Extraterrestrial Highway near Rachel, Nevada, US on the east side of Area 51.

included Stephen Hawking, who has warned that sending out signals could put us in danger.

Earlier this year he said that we should be "wary" of responding to any messages from aliens. Doing so would probably be like when the Native Americans first met Christopher Columbus, he said in September, and in that case things "didn't turn out so well" for the people being visited.

And Professor Hawking has suggested that any civilisation we did make contact with is likely to see us as no more developed than bacteria. As a consequence it might not actively look to kill us — but wipe us out just because it doesn't care about us.

There are no regulations governing whether or not messages can be sent out into space, or what those messages should say. That means that Meti and other groups can send out their signals despite any objections.

But those behind Meti say that it will instead be used as a way to "learn and share information", if it is successful. The team will now look to work out how best to do that, and put together a message that could be understood by other living things in the distant universe that we might be able to make contact with. The initial message is likely to use basic mathematical and scientific concepts.

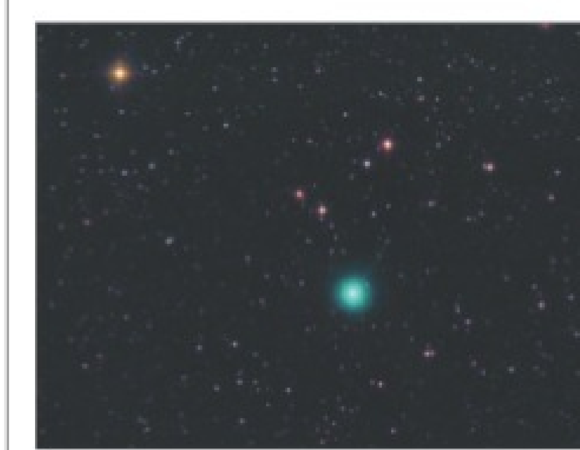
The group hopes to raise $1 million to start that work. Some of that money will go towards building or borrowing a powerful transmitter that can send a message out into the universe.

Scientists have tried in the past to send out messages into space. Those have included Nasa's pioneer 10 and 11 spacecraft that carried messages written onto plaques and a record, as well as radio messages that have been sent out into the universe. It isn't clear whether any of those have ever been received.

Some have suggested that we might have received similar messages ourselves, but most scientists believe that we have so far failed to hear or talk to any alien life.

THE INDEPENDENT

PLUS POINTS



Rare celestial sight

A comet is set to be visible from Earth for the first time in what could be millions of years, Nasa has announced.

The comet, named C/2016 U1 Neowise, will be close enough to Earth between now and 14 January to potentially be seen in the night sky with binoculars or even the naked eye. It was first detected by Nasa in October and is currently heading past Earth on its way towards the sun, where it will pass through the orbit of Mercury before heading back to the outer Solar System.

Paul Chodas, manager of Nasa's Centre for Near-Earth Object Studies, said, "C/2016 U1 has a good chance of becoming visible through a good pair of binoculars, although we can't be sure because a comet's brightness is notoriously unpredictable." The best chance of seeing the comet is to look in the south-eastern sky just before dawn. The comet will be seen further south every day until it disappears from sight sometime in mid-January.

C/2016 U1 has an orbit that could take millions of years, meaning this could be the first and last time it gets so close to Earth. Comets typically only have a lifespan of one or two million years. But while it is likely to be comparatively near to us, the comet is still many millions of miles away and poses no threat to the planet.

Scientists are not entirely sure whether to class C/2016 U1 as a comet or an asteroid but believe it is more likely to be a comet.

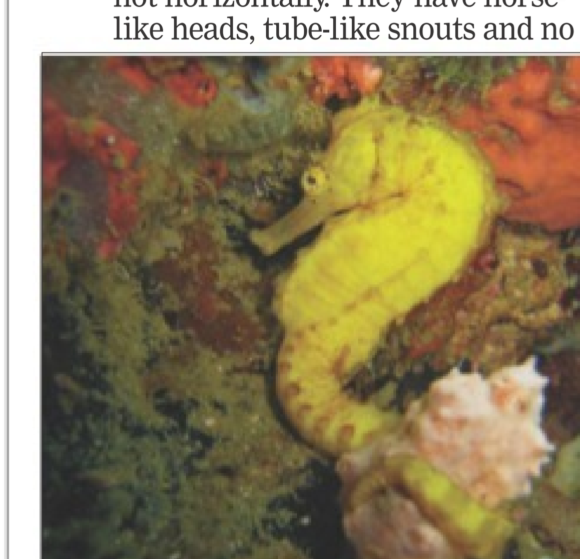
BEN KENTISH/THE INDEPENDENT

Male pregnancy

Scientists have unlocked some of the genetic secrets of the weird and wondrous seahorse, including its exotic eccentricity of male pregnancy.

Researchers said recently that they sequenced the genome of a seahorse species for the first time and identified the genetic underpinning for certain peculiarities in this equine-looking fish group that inhabits coastal waters around the world.

Seahorses boast a host of oddities. Males, not females, carry and give birth to babies. They swim upright, not horizontally. They have horse-like heads, tube-like snouts and no



teeth. They have grasping tails to grip sea grasses and corals to avoid being swept away by currents. Their bodies are covered in bony plates. Unlike most fish, they lack tail and pelvic fins. Their eyes work independently, letting them look forward and backward simultaneously. And they can change colours to camouflage themselves.

"They are such iconic animals, one of the examples of the exuberance of evolution," said evolutionary biologist and genome researcher Axel Meyer of Germany's University of Konstanz, one of the researchers in the study published in the journal *Nature*.

"Their numbers are declining due to habitat destruction and harvest by humans," added molecular biologist Byrappa Venkatesh of Singapore's Agency for Science, Technology and Research.

The researchers analysed the genome of the Southeast Asian tiger tail seahorse, which reaches 10 cm long and boasts a yellow-and-black banded tail. It had the fastest rate of molecular evolution among any fish whose genome has been studied.

Male seahorses possess a brood pouch. During mating, a female deposits eggs into the male's pouch. The male fertilises the eggs internally and carries them in the pouch until they hatch, releasing the fully formed offspring into the sea.

A gene present in other fish that plays a role in egg hatching underwent duplication in the seahorse and assumed a new role, helping the advent of the male pouch.

THE JAPAN NEWS/ANN