USING BIG DATA TO MAP FORESTS, TREE BY TREE

PROFESSOR TIAN ZHENG, PH.D.

USING BIG DATA TO MAP FORESTS, TREE BY TREE

TROPICAL FORESTS ARE MADE UP OF HUNDREDS OF TREE SPECIES. KNOWING WHAT FORESTS ARE MADE OF AT THE INDIVIDUAL TREE LEVEL CAN UNLOCK MANY SECRETS, NOT LEAST HOW FORESTS RESPOND TO THE EFFECTS OF CLIMATE CHANGE. **PROFESSOR TIAN ZHENG**, A STATISTICIAN AT COLUMBIA UNIVERSITY, USA, HAS TEAMED UP WITH **PROFESSOR MARIA URIARTE**, AN ECOLOGIST, TO COMBINE THEIR SKILLS WITH BIG DATA AND ECOLOGICAL KNOW-HOW TO MAP TREE SPECIES IN TROPICAL FORESTS

GLOSSARY

ARTIFICIAL INTELLIGENCE – the science of developing computer systems that can perform tasks usually reserved for humans, such as visual perception and decision-making

BIG DATA – extremely large datasets, within which computer programs may be able to recognise patterns or trends

LIDAR (LIGHT DETECTION AND RANGING) – a remote sensing method that uses a pulsed laser to measure distances between things

MACHINE LEARNING – a field of artificial intelligence, specifically concerning the development of computer programs that can learn and adapt without human intervention

REMOTE SENSING – using satellites or aircraft to capture data about the Earth's surface

AERIAL IMAGING – a remote sensing method whereby drones, or manned aircraft, equipped with a high-resolution camera and LiDAR take images of the forest, which can then be processed and analysed

Climate change is already having a profound effect on the world, not least through the rising frequency and intensity of cyclonic storms. Scientists know that powerful storms such as hurricanes can cause a lot of damage to forests, but how forests recover from these impacts is less well understood. There are winners and losers, with some species benefiting from storms at the expense of others. This has big implications for mitigating climate change, too: forests absorb a fifth of the carbon that we emit, but storms can release stored carbon and change the dominance of different species.

Understanding how hurricanes affect forests is a task that cannot be tackled by one scientific discipline alone. In Columbia University in New York City, this recognition has led to a unique collaboration. Professor Tian Zheng works in the Department of Statistics at Columbia and is also a member of Columbia's Data Science Institute, while her colleague Professor Maria Uriarte is a tropical ecologist within the Department of Ecology, Evolution and Environmental Biology. By combining their skillsets and working with others, they have taken strides in understanding what happens to forests after storms.

SEEING THE TREES FROM THE FOREST

Mapping a forest by traditional methods is no easy task. Documenting the trees that make up a forest from the ground level involves time-consuming treks over difficult terrain and usually only results in a small proportion of a forest being sampled. Given that forests are highly complex environments, and the diversity of trees within any one forested area, there is no guarantee that the sample area even represents the rest of the forest. Nevertheless, many intrepid ecologists have undertaken extensive mapping surveys, but now they have a helping hand – in the form of modern technology.

"We are interested in using big data produced by remote sensing technology to study the spatial distribution of species," says Tian. Aerial imaging is one kind of remote sensing technology that gives researchers access to images of massive areas of forest, but there is another method that can be even more revealing. LiDAR (Light Detection and Ranging) offers something different: this remote sensing technology uses lasers, emitted by apparatus on an aeroplane, for instance, which 'bounce' when they hit something below them and are then picked up again by the apparatus. By emitting thousands upon thousands of these lasers, a three-dimensional image of the world beneath is produced as a cloud of points. This is the same technology used by self-driving cars to detect the distance between themselves and

objects around them, so they do not bump into things.

BIG DATA, BIG POSSIBILITIES

These days, scientists can generate massive datasets using new technology. The trick to scientific discovery, however, is drawing insights from these data – which, when you have masses of the stuff, is not so straightforward. This is where machine learning comes in. Once a computer program has been 'taught' how to interpret a certain sort of data, it can then do the same process for massive datasets.

Take the case of remote sensing data. After data collection, the researchers will have a massive two-dimensional image and three-dimensional 'map' of a forest's canopy, which can be detailed enough to be able to pick out the individual leaves of some tree species. Researchers know that certain patterns in the map indicate that the lasers have bounced off a certain sort of tree but picking these out by hand would take years. Instead, they teach a computer program which patterns correspond to which species, and then let the machine run through the rest of the dataset. "The process mimics how human infants learn, through cognitive development," says Tian. The end product is a highly detailed map of the forest, right down to the individual species of the trees within it.

BEFORE (AND AFTER) THE STORM

In 2017, Hurricane Maria devastated the island of Puerto Rico, including its forests. Ecologist Maria Uriarte had been mapping trees in Puerto Rico's forests for fifteen years prior, which, together with remote sensing data, meant that a solid map existed for some of the island's forests. Using LiDAR and aerial imaging from the weeks and months following the hurricane, Tian and Maria are planning on building a second map – together, creating 'before' and 'after' datasets that can then be compared.

Initially, Tian and Maria focused on two species: *Prestoea acuminata*, a species of palm, and the distinctive tree *Cecropia schreberiana*. "The distributions of these species can act as signatures of hurricane damage," says Tian. For instance, *Cecropia schreberiana* is a pioneer species – this means that after a hurricane or some other event has caused devastation, it is one of the first species to 'colonise' the area due to its rapid growth conditions. Both species have visual features that are easy to identify by humans and the algorithms analysing the datasets.

The difference is that a well-trained computer algorithm will be able to identify these species 50 times faster than a human and with the same accuracy. This means that scientists are freed up from what could be a long and laborious task, giving them more time to focus on important questions such as how climate change is impacting forests.

NEXT STEPS

Indeed, now that Maria and Tian are equipped with an initial set of machine learning and imaging tools, they will be looking to answer specific questions. For instance, they are interested in discovering how landscape characteristics, such as a forest's underlying topography and geology, influence storm damage. They will be able to compare datasets to find out how hurricane damage influences species composition, and even whether human activity has an effect on forests' recovery. Beyond this study, there are potential applications for this fusion of technologies in all sorts of other fields. "We are evaluating how this sort of machine learning could be applicable to other areas, such as medical imaging," says Tian. As artificial intelligence becomes increasingly powerful, there will be increasing potential for discovering things about our world that have not been in our grasp before.

PROFESSOR TIAN ZHENG

Department of Statistics, Data Science Institute, Columbia University, New York, USA

PROFESSOR MARIA URIARTE

Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, USA

FIELD OF RESEARCH

Big Data and Machine Learning

RESEARCH PROJECT

Using artificial intelligence to map forests' species composition to understand the effects of powerful storms on rainforests.

FUNDER

Microsoft

ABOUT BIG DATA AND AI

Tian explains more about her career and opportunities within her discipline:

HOW MUCH MULTIDISCIPLINARY COLLABORATION DOES YOUR ROLE INVOLVE?

I started my career collaborating with computational biologists and geneticists. Over the years, I have collaborated with sociologists, political scientists, psychologists, climate scientists and, now, ecologists. Like John Tukey (a renowned mathematician) once said, "The best thing about being a statistician is that you get to play in everyone's backyard." I have played in many backyards!

WHAT EXCITES YOU ABOUT BIG DATA?

I am excited to be able to unleash the insights hidden within all the data we are collecting today, and how that will lead to new solutions. Machine learning offers new ways of extracting valuable information from large amounts of data.

COULD ADVANCES IN MACHINE LEARNING LEAD TO FEWER JOBS IN THE FUTURE?

I think it is unlikely. Some repetitive tasks (e.g. labelling trees from images of a forest) might be replaced by machines, but this opens the door for new creative careers to emerge. Within the development of machine learning, numerous career opportunities have already been created because every algorithm requires humans to design and optimise them. Data are not perfect, and algorithms need continuous fine-tuning to make sure they give valid conclusions. This calls for a lot of people who are talented in statistics, computer science, optimisation and data ethics.

ARE THERE MANY OPPORTUNITIES FOR STUDYING AND WORKING WITH BIG DATA, ARTIFICIAL INTELLIGENCE AND STATISTICS?

Yes, there is a great need for talent in these areas across all sectors in the USA. Finding real-world, data-driven solutions to problems often requires mathematical and statistical modelling, computational technologies and expert knowledge. In particular, statistical thinking has become ever more important to understanding the potential for bias or uncertainty. These are critical factors that need to be considered if big data is to realise its potential.

HOW TO BECOME A STATISTICIAN

- Some universities offer undergraduate degrees specifically in statistics, whereas others will incorporate it within mathematics, often with the option to specialise further during your degree.
- According to QS World University Rankings, the best universities for statistics are MIT, ETH Zurich, Stanford, Harvard, Oxford and Cambridge.
- Columbia University is among the top 15 in the QS World University Rankings and offers undergraduate and graduate programmes in statistics, AI, machine learning and data science.
- According to PayScale, the average annual salary for a statistician in the US is around $74k.

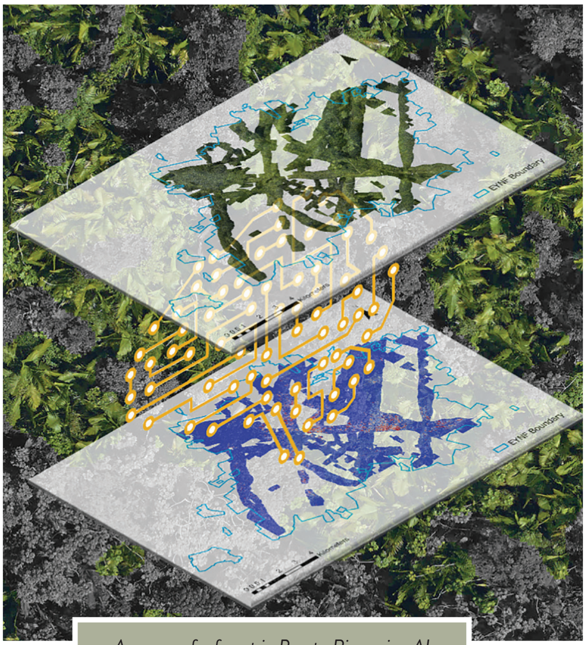
PATHWAY FROM SCHOOL TO STATISTICIAN

Tian recommends taking courses in mathematics, statistics and coding at school/college. Other subjects, especially the sciences, can also be useful.

TIAN'S TOP TIPS

01 If you enjoyed or continue to enjoy playing with LEGO or reading detective stories and solving crimes, then you are already prepared for a career in data science (and science, overall). The same curiosity and problem-solving skills apply to careers in data science.

02 There are a lot of similarities between a data scientist and a detective. You are using all the available tools to piece together clues. You may not succeed at first, some experiments may fail, but the joy in finally making something work and finding an answer to a problem is one of the most rewarding aspects of being a scientist.



A survey of a forest in Puerto Rico using AI

HOW DID TIAN BECOME A STATISTICIAN?

WHAT DID YOU WANT TO BE WHEN YOU WERE YOUNGER?

Originally, I wanted to be an architect, but this didn't work out. Instead, I discovered a fantastic career in mathematics!

WHAT CAREER OPTIONS WERE OPEN TO YOU AFTER COMPLETING YOUR UNDERGRADUATE DEGREE?

Some of my classmates went to work for banks and research labs directly after graduation. Others, like me, continued on to graduate school.

WHAT INSPIRED YOU TO TAKE A MASTER'S AND PHD IN STATISTICS?

I have always liked solving problems. Statistics is a field that uses data to solve problems, which was a very strong draw for me.

WHAT DO YOU LOVE MOST ABOUT YOUR WORK?

I love creating statistical models and machine learning algorithms that reveal interesting patterns in data.

YOUR TWITTER PROFILE SAYS YOU ARE A 'UNICORN TRAINER'! WHAT DOES THIS INVOLVE?

When data science first became established as a discipline, certain news articles referred to data scientists as 'unicorns'. Since I designed courses to teach students data science skills, I dubbed myself a 'unicorn trainer'!

MEET CHENGLIANG TANG

Chengliang Tang is a PhD student at Columbia University who works with Tian and Maria on the tree species mapping project.

WHY DID YOU CHOOSE TO STUDY MATHEMATICS AS AN UNDERGRADUATE?

I have been interested in mathematics since high school. People say that mathematics makes up the foundation of all sciences, which I believe without a doubt.

WHAT DREW YOU TO STUDY STATISTICS?

I took courses in many different fields besides pure maths. I found statistics to be a field that revealed the beauty of mathematical theories and led to them having an impact in the real world. As we enter the era of big data, there are more and more opportunities emerging within statistics.

WHAT DO YOU FIND ENJOYABLE ABOUT THE TREE SPECIES MAPPING PROJECT?

Tian and Maria are experts in their domains, so during our weekly meetings I learn a lot from them. The project is motivated by important questions about climate change and forests, involving data collected by ecologists and meaningful outputs that help answer these questions. The significance of this project really motivates me.

WHERE DO YOU SEE YOURSELF GOING NEXT?

The world is changing rapidly, so it is difficult to predict, but I do hope to continue working in statistics. It is an exciting subject with new ideas and challenges emerging every day. I am eager to learn more, do more and contribute to our growing knowledge base.

WOULD YOU ENCOURAGE OTHERS TO STUDY MATHEMATICS?

Yes, absolutely! As well as everything I mention above, mathematics has also fundamentally shaped my way of thinking. Everyday life is full of meaningless noise, but statistics teaches you how to extract the valuable signals from within it. These changes are so gradual and imperceptible that I didn't realise they were happening until recently.

WHAT ADVICE WOULD YOU GIVE TO YOUR YOUNGER SELF?

Take more exercise! I used to undervalue physical activity, but I now realise how crucial it is for physical and mental health. Research projects are long journeys full of ups and downs. Exercise helps us keep a positive attitude and enjoy what we are doing.

AI AND FORESTS WITH PROFESSOR TIAN ZHENG

TALKING POINTS

KNOWLEDGE

1. What is LiDAR?
2. What is a pioneer species?

COMPREHENSION

3. How can tropical storms affect a forest's ability to store carbon? Explain the process.
4. Why did Tian and Maria initially focus on two species of trees for their research?

APPLICATION

5. Scientists are worried that climate-change induced storms and subsequent forest damage may lead to a positive feedback loop, or 'snowball effect'. Describe what you think they mean by this.

ANALYSIS

6. Think about the two methods of remote sensing mentioned in the article: aerial imaging and LiDAR. What do you think are the benefits and drawbacks of each method?
7. The article says, "These days, it is often pretty easy for scientists to generate massive datasets". Why do you think this is the case?

SYNTHESIS

8. How do you think data from ecological surveys and remote sensing can be combined into one dataset? What challenges might this process face?
9. What sort of factors do you think Tian and Maria are hoping to compare in the 'before' and 'after' maps for Puerto Rico's forests? What conclusions might they draw from these comparisons?

EVALUATION

10. Tian mentions that the group's methodologies may have applications for medical imaging. How useful do you think these applications would be and why?
11. Some people are concerned that machine learning could become an existential threat for humanity. To what extent do you agree and why?

ACTIVITIES YOU CAN DO AT HOME OR IN THE CLASSROOM

Create a poster that acts as a flowchart of how Tian and Maria's project came about, the development of the methodology and why this research is important. It should cover the following steps:

1. Maria's ecological surveys
2. Remote sensing (aerial imaging and LiDAR)
3. Hurricane Maria and Puerto Rico
4. Tian's work with big data
5. Machine learning
6. Applications and implications

Undertake further research on the internet to flesh out each of these steps. For example, NASA is a good place to start:

https://earthobservatory.nasa.gov/images/144441/a-haircut-for-puerto-ricos-forests

At each step, think about how you can succinctly summarise the information whilst keeping it accurate and engaging. Feel free to illustratyour a poster or design it on a computer. Present your poster to your group or class in an engaging way. Think about the points you want to draw particular attention to and anticipate any questions you might get asked.

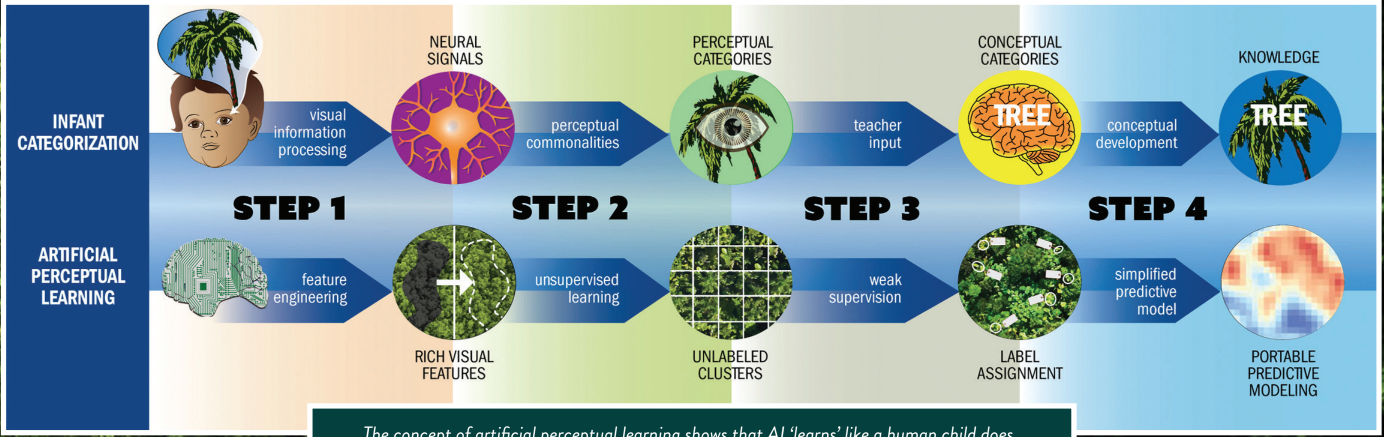
MORE RESOURCES

This grantee profile for Tian and Maria's project explains their research in more detail: **https://ai4edatasetspublicassets.blob.core.windows.net/grantee-profiles/Columbia%20University_US_LATAM_Climate_AI4E%20Grantee%20Profile.pdf**

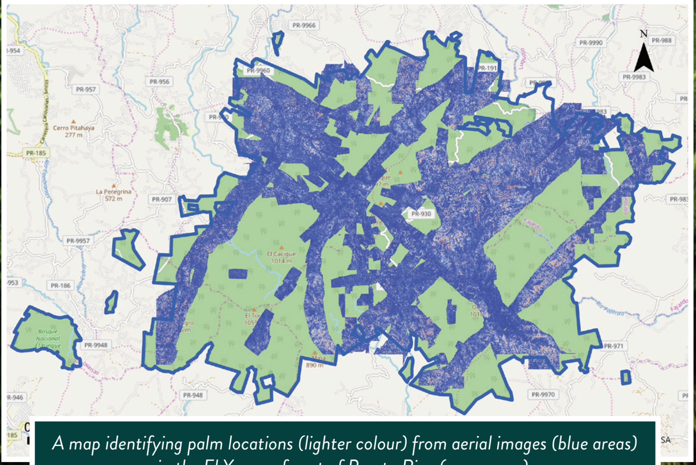
You can also find out more about the research on NASA's website: **https://gliht.gsfc.nasa.gov/index.php?section=49.** And explore the aerial photography taken of Puerto Rico: **https://glihtdata.gsfc.nasa.gov/puertorico/index.html**

This video from NEON Science uses clever graphics to explain how LiDAR remote sensing works: **https://www.youtube.com/watch?v=EYbhNSUnIdU**

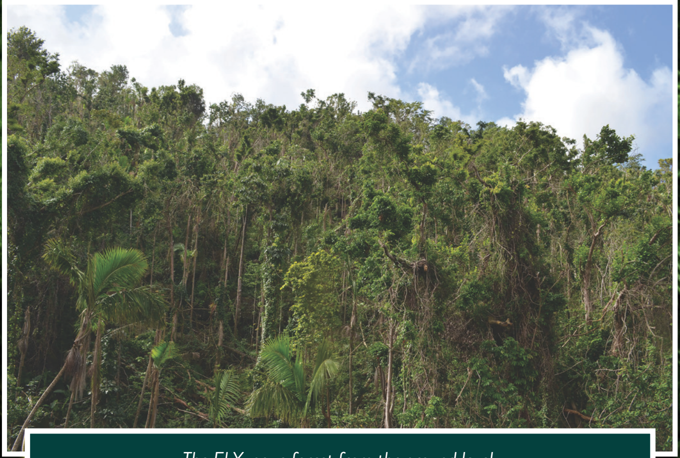
This Science article explains how scientists are studying the post-hurricane recovery of Puerto Rico's forests and the broader implications for climate change: **https://www.sciencemag.org/news/2018/09/puerto-rico-s-catastrophic-hurricane-gave-scientists-rare-chance-study-how-tropical**



The concept of artificial perceptual learning shows that AI 'learns' like a human child does.



A map identifying palm locations (lighter colour) from aerial images (blue areas) in the El Yunque forest of Puerto Rico (green area).



The El Yunque forest from the ground level.

Tian and Maria discuss their work. As Tian says, there are a lot of similarities between a data scientist and a detective. You are using all the available tools to piece together clues.

Chengliang stands below the Alma Mater statue at Columbia University in New York, USA.

futurum
Inspiring the next generation

WWW.FUTURUMCAREERS.COM

TEL: +44 117 909 9150

EMAIL: INFO@FUTURUMCAREERS.COM

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK