Fake honey, Machine Learning and Microscopy

By Alexis Gkantiragas

At a conference I bumped into an eccentric fellow student who was toying with the idea of developing a new method for honey authentication. His background was in machine learning, but he had little microscopy or bee research experience and so I offered to help. Honey is one of the world's most faked products, and one can theoretically identify the origin of the honey from the morphology and size of the pollen in the honey.

Since current methods are either ineffective or prohibitively expensive, and faked honey harms both beekeepers and bees, an authentication tool that was affordable but effective would have tangible benefits. We also came to the realisation that it could have practical benefits in bee research and environmental monitoring. Indeed, when I reached out to a former supervisor who works on both honeybees and bumblebees, it transpired that all pollen-based research in bee research was being done manually.

Unfortunately, my supervisor at the time did not want me using the lab's microscopes for the proposed project as this was not covered by the grants. So, we began by using my microscope that I had left over from my school days and mounting a smartphone camera to its eyepiece. Sadly, this promptly broke and we had to get a second one. We grabbed honey from various shops and supermarkets. The use of affordable equipment was crucial since current methods for honey authentication are prohibitively expensive for all but the largest scale beekeepers. We wanted to make a tool that could be used by any beekeeper with basic competency and around £100 spare to spend on a microscope.

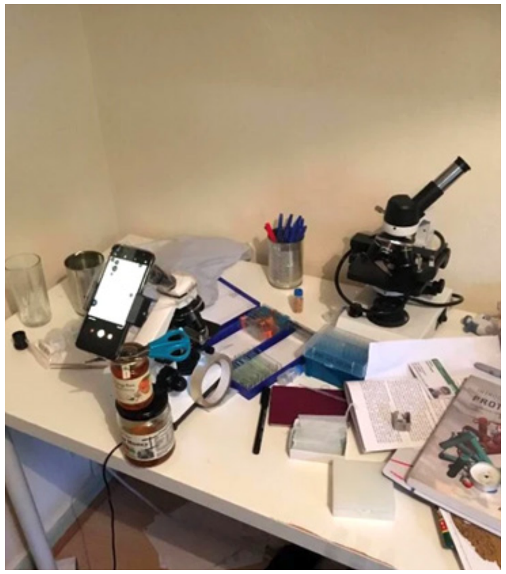

Figure 1. The makeshift rig.

I managed to set up a very makeshift microscope rig in my bedroom and began scanning slides of honey (figure 1). By appropriately compressing the cover slip onto the honey I was able to view the pollen in roughly a single focal plane. Amusingly, some of the samples tested had no pollen visible whatsoever due to ultrafiltration of the honey prior to its sale.

The initial work, which was later presented at NeurIPS[1] (Conference on Neural Information Processing Systems - formerly NIPS), the largest machine learning conference, was done using a 'supervised' approach. In practice, this consisted of my spending hours in front of a laptop painstakingly segmenting (drawing boxes around) and labelling the pollen. I attempted (unsuccessfully) to convince my university to fund my travel to the conference, since only one person would be funded by the conference itself. They expressed confusion as to how and why a biochemistry student had managed to get his work into a machine learning conference. After the work was presented, we were baffled to receive media coverage from *FastCompany* and *Techxplore*.

For the next set of work, we set up two microscopes (we bought another so we could work simultaneously) in my family's caravan, since my new room lacked a desk or space to put one. The algorithm used for this segment of work was unsupervised - meaning that it classified the pollen into groups without human input (figure 2). From the initial data we were able to deduce that it achieved roughly family level classification. After we presented our research again at the *International Conference on Machine Learning (ICML)*[2], the *World Bee Project CIC* kindly gave us a modest grant to continue our work.

Wanting to refine our work further and benchmark to human classification we came upon the realisation that we needed to be able to measure the dimensions of the pollen accurately. Unfortunately, the cheap microscopes we had bought did not have integrated eyepiece graticules and we realised that to require users to have an eyepiece graticule would make the technology prohibitively expensive for many beekeepers. Thus, we developed an open-source programme which calibrated the microscope from a stage micrometre alone[3]. Surprisingly this was novel, and we hope that it will increase access to microscopic techniques more broadly. Next, we plan to apply this sizing technology alongside our pollen identification and classification tools.

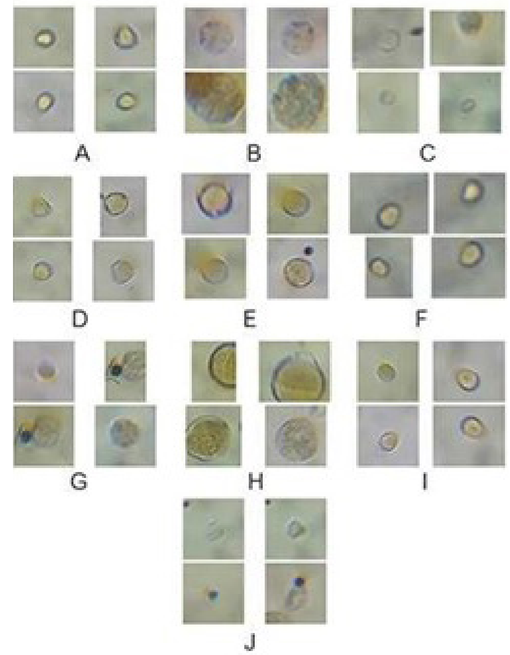

Figure 2. Unsupervised clustering of pollen types. From He et al., 2019.

It has surprised me just how accessible the intersection between light microscopy and machine-learning is and how viable it is to do actual novel and impactful research with little-to-no funding. I think it is quite plausible there are similar gaps and opportunities for relatively cheap, proof-of-concept work in other areas waiting to be addressed.

References

1. He, P., Gkantiragas, A. & Glowacki, G. Honey authentication with machine learning augmented bright-field microscopy. *arXiv* (2018).
2. He, P., Glowacki, G. & Gkantiragas, A. Unsupervised Representations of Pollen in Bright-Field Microscopy. in *international conference on machine learning* (2019). doi:1901.00516.
3. Glowacki, G., Gkantiragas, A., Brooke, B.-H., Mihalik, D. & He, P. Development of an open source program to identify the size of objects using bright-field microscopy - an automated calibration tool. *bioRxiv* (2021).