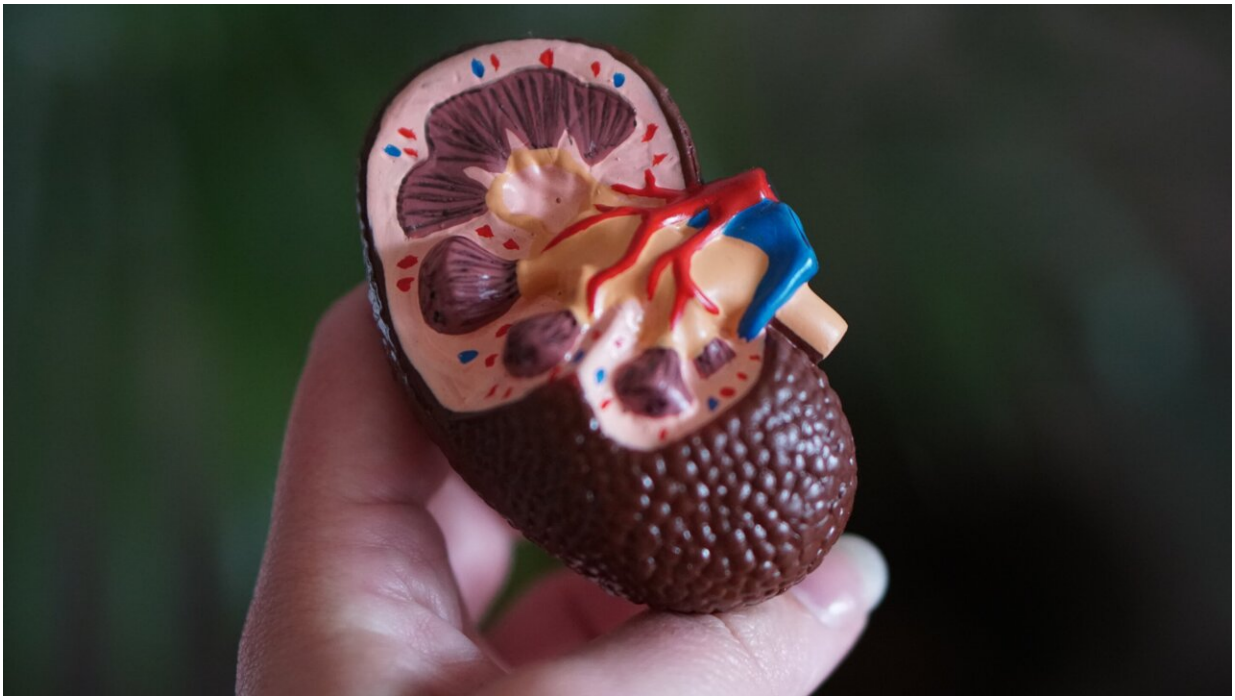
How machine learning could aid compatibility in kidney transplantation

August 9 2023, by Erica Moser



Credit: Unsplash/CC0 Public Domain

The United States saw a record 25,487 kidney transplants in 2021, according to the latest annual data report from the Organ Procurement and Transplantation Network and Scientific Registry of Transplant Recipients. Five years after transplantation, successful organ function—called graft survival—of kidneys from deceased donors was 81% among patients ages 18 to 34 and 68% among people older than 65.

Malek Kamoun of the Perelman School of Medicine and Ryan Urbanowicz of Cedars-Sinai Medical Center are developing machine learning strategies to improve kidney matching and decrease the risk of graft failure—with help from Penn students.

Three undergraduates worked remotely this summer on the project: Sphia Sadek, Antonios Kriezis, and Aryan Roy. Sadek and Roy completed the work through the Penn Undergraduate Research Mentoring Program. Each PURM student receives a $5,000 award for the 10-week program, supported by the Center for Undergraduate Research and Fellowships.

Sadek, a rising third-year student in computer science and cognitive science, says she has always been interested in a career in artificial intelligence but has never done anything with machine learning. She heard about PURM during orientation events for transfer students.

"For my project, I'm working on the threshold for how many amino acid mismatches would be considered low-risk or high-risk, for each bin," says Sadek, who is from Jersey City, New Jersey. "A bin is basically a list of these amino acid positions, and we use that bin to stratify the two groups."

Kamoun and Urbanowicz helped develop an algorithm called FIBERS (Feature Inclusion Bin Evolver for Risk Stratification), which assumed a threshold of zero, meaning that bins, a group of amino acid positions, with at least one mismatch are considered high-risk.

Sadek is exploring the possibility that the threshold could be higher, a beneficial exercise because it could lead to better risk stratification by race and ethnicity, for example, due to significant amino acid variations among various ethnic groups.

She is doing this using Python programming languages and data from the Scientific Registry of Transplant Recipients (SRTR), and Urbanowicz says her method "seems to be working really well so far."

Along with testing, coding, and programming, she went to meetings to learn about the human leukocyte antigen (HLA) system and the HLA biology behind kidney allograft rejection in transplant recipients.

She says the PURM experience helped solidify her interest in machine learning and AI and helped her gain research experience by "looking at different articles, thinking of new ideas to help test and debug, thinking of new ideas that can help improve an original existing project." And it helped with communication and networking skills, she says.

An evolutionary algorithm

Kamoun was inspired to do his current work by his former mentor Jean Dausset, the late Nobel Prize recipient who in the 1960s discovered the HLA system, a complex of genes that encode proteins responsible for immune system regulation. HLAs control the body's immune response to transplants and are a major cause of organ rejection.

"When a patient receives a transplant, they usually receive a transplant from somebody whose HLA are very different from their own," Kamoun says. "That leads to T lymphocyte activation and expansion because these cells see the recipient as carrying foreign antigen material; consequently, kidney allograft rejection could occur in transplant recipients who are not adequately immunosuppressed."

The traditional method of analyzing HLA differences is at the antigen level, whereas defining differences at the molecular level requires looking at amino acid sequence variation. But Kamoun says there is so much variation across populations at the amino acid level that it's

difficult to use traditional statistical methods to understand genetic variation in kidney graft survival.

So he turned to Urbanowicz, an expert in machine learning tools who helps groups study various biomedical problems. After a year or two of discussion, they submitted a National Institutes of Health grant and received funding last year.

Kamoun and Urbanowicz co-authored a paper assessing the performance of the FIBERS algorithm, published this spring in the *Journal of Biomedical Informatics*.

FIBERS is an evolutionary algorithm—and an approach without a hypothesis—that automatically discovers "bins" of amino acid-level mismatches, based on their ability to categorize donor-recipient pairs into low-risk and high-risk groups for graft survival.

Urbanowicz says FIBERS discovers which amino acid mismatch positions are most important to look at when determining the risk of graft failure, and this could work into a kidney allocation policy downstream, or at least assist in determining which transplant recipients might be at higher risk of graft failure.

Using SRTR data, the researchers applied FIBERS on 166,574 deceased-donor kidney transplants that occurred between 2000 and 2017. They found that compared to traditional methods, FIBERS categorized more than twice as many patients as low risk, indicating that traditional methods shroud relevant amino acid variability that can impact risk of graft failure.

Adjusting for variables

Rising second-year students Kriezis and Roy are exploring a different

avenue of the project than Sadek: adjusting for covariates—meaning variables such as age, sex, and race—using two different approaches.

Kriezis, a systems science and engineering major, was interested in the project because he wanted to do something with coding and machine learning. He was assigned to test the Akaike Information Criteria (AIC) model, a statistic that penalizes models that perform well through using larger numbers of covariates.

"Before I joined this project, I didn't really know what machine learning was," says Kriezis, who is from Athens. "I was basically able to learn what it is, how it works and how useful it can be. The team was very helpful. At every step, they were very easy to talk to, and they were willing to explain whatever questions you had."

As Kriezis spent his time debugging code and sorting through spreadsheet columns, he says he wasn't thinking about kidney transplantation much. But he says Kamoun, director of the Clinical Immunology and HLA Immunogenetics Laboratory at the Hospital of the University of Pennsylvania, talked to students about the larger application of their work.

Urbanowicz says the approach Kriezis worked on is expensive because it takes about 16 hours to run, "but it is a more established, recognized way in statistics to adjust for covariates," whereas "there's a lot more uncertainty" with the method Roy used.

Urbanowicz says Roy worked on an approach—one first proposed by statistician Keith McCullough, another co-author on the *Journal of Biomedical Informatics* paper—that involves adjusting for residuals in the FIBERS bins.

After running preliminary results on the original FIBERS approach and

both covariate approaches, Urbanowicz says the covariate approaches perform slightly better than the original "and that the residuals approach we implemented seems very promising and is working just as well as AIC," but faster.

More information: Satvik Dasariraju et al, HLA amino acid Mismatch-Based risk stratification of kidney allograft failure using a novel Machine learning algorithm, *Journal of Biomedical Informatics* (2023). DOI: 10.1016/j.jbi.2023.104374

Provided by University of Pennsylvania