Moving Knowledge from Research to Impact

SUCCESS STORY

NIJ and Syracuse University
Improving DNA Mixture Interpretation with the Help of Machine Learning

PACE™
Probabilistic Assessment Contributor Estimator

> "This is something that every DNA crime lab worldwide should investigate. We now have a tool that with the press of one button does the entire suite of services and provides you with the end result. Having the simple decisions made for you enables focus on addressing the issues of most importance."
>
> —*Vic Meles, Marketing & Business Director, NicheVision*

Problem and Solution Synopses

When DNA is recovered from a crime scene, victim, or suspect, it may be a mixture of genetic information from multiple individuals. To successfully interpret a mixture of DNA, analysts must first determine the number of contributors present in the sample. This is a critical but challenging first step towards "deconvolution" of the DNA data.

Standard computer programs designed to assist with DNA mixture analysis require analysts to determine the number of contributors beforehand. To do so, users must identify, categorize, and remove irrelevant data called artifacts. However, in cases when the DNA is in limited quantities, degraded, or both—individual contributors are difficult to separate, some even artifacts. Generally, as the number of contributors increases and complexity of the DNA profile increases, requiring greater amounts of user time and experiential judgment to determine the number of contributors.

To address these challenges, Dr. Michael Marciano and Jonathan Adelman of Syracuse University developed the **Probabilistic Assessment for Contributor Estimation (PACE™)** a machine learning method to improve mixture interpretation. Previously, machine learning was not considered as useful to forensic DNA interpretation because of the lack of large and diverse data sets to "train" an algorithm. However, the rise of forensic databases and expansion of data-sharing practices has enabled machine learning to become a tool to tackle complex issues like DNA mixture analysis.

PACE™ employs a hands-off approach for determining the number of contributors in traditional DNA data sets (those based on the highly variable regions in our DNA that allow for human identification).

It defines a local noise threshold for each region of the DNA profile analyzed and then applies machine learning algorithms to classify the likely number of contributors in the sample. This tool is 20% more accurate than current methods in estimating the number of contributors in up to four-person mixtures, currently the upper limit.

With an overall accuracy rate of 98%, PACE™ is able to improve the confidence in assessing the number of contributors. PACE™ is unique in the field of forensic DNA analyses and is leading the way toward a future filled with myriad new tools and interpretation methods that better utilize complex, challenging samples. The development of PACE™ (patent pending), which is exclusively licensed to NicheVision, was supported through the National Institute of Justice (NIJ) forensic science R&D program with the goal of eventually seeing operational use in casework.

Key Benefits

▶ Estimates the number of contributors in a matter of seconds, eliminating a critical, time-intensive data-interpretation step that is contributing to growing backlog issues.

▶ Enables analysts to arrive at more confident conclusions by generating number of contributor estimates based on tested machine learning models and serves as a quality control measure in the analysis of complex data.

▶ Accommodates different DNA profiling kits and runs on standard computers, facilitating broader adoption with limited impact on existing workflows.

Visit us at
www.ForensicCOE.org | ForensicCOE@rti.org | 866.252.8415
RTI International
3040 E. Cornwallis Road PO Box 12194, Research Triangle Park, NC 27709 USA

@ForensicCOE #FTCoE

Publications

Marciano M.A. and Adelman J.D. PACE: Probabilistic Assessment for Contributor Estimation— A machine learning-based assessment of the number of contributors in DNA mixtures. Forensic Science International: Genetics 27 (2017) 82–91.

Marciano M.A., Williamson V.R. and Adelman J.D. A Hybrid Approach to Increase the Informedness of CE-based Data Using Locus-Specific Thresholding and Machine Learning. Forensic Science International: Genetics. 35 (2018) 26-37.

NIJ Research

NIJ support contributed to the development of the PACE™ tool, which is now available as an end-user tailored software that can be seamlessly integrated into crime laboratory work flows. The software, PACE™, introduces machine learning tools to the forensic DNA community for fast and accurate analysis of complex DNA mixtures. With NIJ backing, the researchers at Syracuse University's Forensic and National Security Sciences Institute were able to reach numerous crime laboratories and stakeholders to garner valuable insight that accelerated the technology on the path to commercialization. NIJ support of this project enabled widespread data sharing by crime laboratories, dissemination of knowledge through workshops, and incorporation of end-user feedback into software design and development.

Bringing Research to Practice

Several factors promoted the transition of this technology from the research bench to the forensic community. First, the team coupled Marciano's previous experience as a casework analyst with Adelman's expertise in computer science and machine learning to create a product that both addressed a problem faced by practitioners and integrated seamlessly into existing crime laboratory workflows. Second, partnering with laboratories such as the New York City Office of the Chief Medical Examiner–Department of Forensic Biology, the Onondaga County Center for Forensic Sciences, and the Promega Corporation generated valuable insight from key stakeholders. The team leveraged feedback from these and other stakeholders to improve the user interface, develop the deconvolution method, and ensure that the software provides value to real casework. The researchers are supporting dissemination efforts by conducting training initiatives that facilitate adoption of the technology in public laboratories. The team is currently in the process of patenting PACE™ and has exclusively licensed it to NicheVision Forensics, LLC.

About NicheVision

NicheVision Forensics, LLC empowers over 150 forensic crime labs worldwide with industry standard software tools to maximize their DNA human identity capabilities. The NicheVision suite of software tools include powerful and easy to use case management tools and robust mixture interpretation technologies that enable forensic scientists to resolve even the most complex forensic mixtures. PACE™ employs machine learning to quickly and accurately determine the number of contributors in a forensic mixture, and to identify and document all artifacts in a forensic mixture without extensive user interpretation. While PACE™ is currently limited to use with traditional capillary electrophoresis sequencing platforms, the technology is being further extended in collaboration with Syracuse University to use next generation sequencing data.

The Future

The introduction of PACE™ represents a launching point for machine learning-based applications in forensic DNA analysis, and additional versions of PACE™ will be commercially available soon. Marciano and Adelman are actively engaged in additional forensic science machine learning projects, including an NIJ-supported project to apply PACE™ to DNA sequence data. Additionally, Adelman is investigating machine learning to enhance latent print analysis and drug identification. The continued used of these machine learning methods, as illustrated by PACE™, will provide the forensic community with faster and more accurate analyses that support justice.

Image Credits

Page 1—Dr. Michael Marciano, Syracuse University

Page 2 (top)—Dr. Michael Marciano, Syracuse University

Page 2 (bottom)— Vic Meles, NicheVision Forensics, LLC