

Nikunj C. Oza, Ph.D.

NASA Ames Research Center
Mail Stop 269-2
Moffett Field, CA 94035-1000
E-mail: Nikunj.C.Oza@nasa.gov
<http://ti.arc.nasa.gov/people/oza>
Tel: (650)604-2978
Fax: (413)771-6476

RESEARCH INTERESTS

Data Mining, Machine Learning, Ensemble Machine Learning methods, Online Learning, Anomaly Detection, Integration of Data Mining with other areas of Artificial Intelligence such as Automated Planning, Applications of Data Mining and Machine Learning to problems in Aviation Safety, Satellite Image Understanding, Fault Detection, and Integrated System Health Monitoring

EDUCATION

- Ph.D. **University of California, Berkeley**
Computer Science, September 2001.
Thesis: **Online Ensemble Learning**
- M.S. **University of California, Berkeley**
Computer Science, May 1998.
Thesis: **Probabilistic Models of Driver Behavior**
- B.S. **Massachusetts Institute of Technology**
Mathematics with Computer Science, February 1994.

BOOK

Nikunj C. Oza, Robi Polikar, Josef Kittler, and Fabio Roli (editors), *Sixth International Workshop on Multiple Classifier Systems*, Springer-Verlag, Heidelberg, 2005.

BOOK CHAPTER (refereed)

Nikunj C. Oza, Ensemble Data Mining Methods, *Encyclopedia of Data Warehousing and Mining*, Idea Group Reference, Hershey, 2005.

JOURNAL SPECIAL ISSUE

Nikunj C. Oza and Kagan Tumer (editors), *Information Fusion, Special Issue on Applications of Ensemble Methods*, 9(1), Elsevier, Amsterdam. 2008.

INVITED JOURNAL ARTICLES

Nikunj C. Oza and Kagan Tumer, Key Real-World Applications of Classifier Ensembles, *Information Fusion, Special Issue on Applications of Ensemble Methods*, 9(1), 4-20. (2007).

REFEREED JOURNAL PUBLICATIONS

Ashok N. Srivastava, Nikunj C. Oza, and Julienne Stroeve, Virtual Sensors: Using Data Mining to Efficiently Estimate Spectra, *IEEE Transactions on Geosciences and Remote Sensing, Special Issue on Advances in Techniques for Analysis of Remotely Sensed Data*, 43(3), pp. 590-600, 2005.

Kagan Tumer and Nikunj C. Oza, Input Decimated Ensembles, *Pattern Analysis and Applications*, 6(1):65-77, 2003.

CONFERENCE SPECIAL SESSION

Nikunj Oza and Robert Morris, "Planning for Data Mining and Data Mining for Planning: Strategies and Methods," Special Session, IEEE Symposium on Computational Intelligence and Data Mining, 2007.

REFEREED CONFERENCE PUBLICATIONS

Elif Kurklu, Robert M. Morris, and Nikunj C. Oza, Learning Points of Interest for Observation Flight Planning Optimization: A Preliminary Report, *Workshop on AI Planning and Learning, International Conference on Automated Planning and Scheduling (ICAPS)*, 2007.

Nikunj C. Oza, Online Bagging and Boosting, *International Conference on Systems, Man, and Cybernetics, Special Session on Ensemble Methods for Extreme Environments*, 2005.

Nikunj C. Oza, Ashok N. Srivastava, and Julienne Stroeve, Improvements in Virtual Sensors: Using Spatial Information to Estimate Remote Sensing Spectra, *International Geoscience and Remote Sensing Symposium*, 2005.

Nikunj C. Oza, AveBoost2: Boosting for Noisy Data, *Fifth International Workshop on Multiple Classifier Systems*, Cagliari, Italy. 2004.

Ashok N. Srivastava, Nikunj C. Oza, and Julienne Stroeve, Virtual Sensors: Using Data Mining to Efficiently Estimate Spectra, *International Geoscience and Remote Sensing Symposium*, 2004.

Nikunj C. Oza, Boosting with Averaged Weight Vectors, *Fourth International Workshop on Multiple Classifier Systems*, Guildford, UK. 2003.

Nikunj C. Oza, Kagan Tumer, Irem Y. Tumer, and Edward M. Huff, Classification of Aircraft Maneuvers for Fault Detection, *Fourth International Workshop on Multiple Classifier Systems*, Guildford, UK. 2003.

Nikunj C. Oza and Stuart Russell, Experimental Comparisons of Online and Batch

Versions of Bagging and Boosting, *The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California. 2001.

Nikunj C. Oza and Kagan Tumer, Input Decimation Ensembles: Decorrelation through Dimensionality Reduction, *Second International Workshop on Multiple Classifier Systems*, Cambridge, UK. 2001.

Nikunj C. Oza and Stuart Russell, Online Bagging and Boosting, *Eighth International Workshop on Artificial Intelligence and Statistics*, Key West, Florida. 2001.

Kagan Tumer and Nikunj C. Oza, Decimated Input Ensembles for Improved Generalization, *International Joint Conference on Neural Networks*, Washington, D.C. 1999, **Recipient: Best Presentation Award.**

Alex Cuthbert, Christopher Stecker, Inna Aleksandrovsky, Sheryl Ehrlich, Nikunj Oza, and Paula Rogers, Instructional Effects on Spatial and Temporal Memory for Videotaped Events in a Large-scale Environment, *Nineteenth Annual Conference of the Cognitive Science Society*, Stanford, CA. 1997.

OTHER PUBLICATIONS

Mark Schwabacher, Nikunj C. Oza, and Bryan Matthews. Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring. *AIAA Infotech@Aerospace Conference*, 2007.

Rodney A. Martin, Mark Schwabacher, Nikunj C. Oza, and Ashok N. Srivastava, Comparison of Unsupervised Anomaly Detection Methods for Systems Health Management Using Space Shuttle Main Engine Data. *JANNAF Propulsion Meeting*, 2007.

Sandra Hayden, Nikunj Oza, Robert Mah, Ryan Mackey, Sriram Narasimhan, Gabor Karsai, Scott Poll, Somnath Deb, and Mark Shirley, "Diagnostic Technology Evaluation Report For On-Board Crew Launch Vehicle," NASA/TM-2006-214552.

Nikunj C. Oza and Upender K. Kaul, Machine Learning for Fault Detection in a Rotating Gear (oral presentation only), *SAE World Aerospace Congress*, 2005.

Upender K. Kaul and Nikunj C. Oza, Machine Learning for Detecting and Locating Damage in a Rotating Gear, *SAE World Aerospace Congress*, 2005. **Recipient: Arch T. Colwell Merit Award for originality of contribution (chosen from among 3300 papers published at SAE conferences in 2005).**

Ashok N. Srivastava and Nikunj C. Oza, Knowledge Driven Image Mining with Mixture Density Mercer Kernels, *ESA-EUSC 2004: Workshop on the Theory and Applications of Knowledge Driven Image Information Mining*, with focus on Earth Observation, 2004.

Ashok Srivastava, Julianne C. Stroeve, and Nikunj C. Oza, Using Kernel Methods to Detect Clouds, Snow, Ice and other Geophysical Processes, *Transactions of the American Geophysical Union*, 84(46), Fall Meeting Supplement, Abstract C12A08-65, 2003.

Nikunj C. Oza, Online Ensemble Learning, Ph.D. thesis, University of California,

Berkeley, 2001.

Nikunj C. Oza, Online Ensemble Learning, *17th National Conference on Artificial Intelligence, Doctoral Consortium*, Austin, TX. 2000.

Nikunj C. Oza and Kagan Tumer, Dimensionality Reduction through Classifier Ensembles, Technical Report NASA-ARC-IC-1999-124.

Nikunj C. Oza, Probabilistic Models of Driver Behavior, *Spatial Cognition Conference*, Berkeley, CA. 1999.

Nikunj C. Oza, Probabilistic Models of Driver Behavior, M.S. report, University of California, Berkeley, 1998.

Jeffrey Forbes, Nikunj Oza, Ronald Parr, Stuart Russell, Feasibility Study of Fully Automated Vehicles Using Decision-Theoretic Control, California PATH Research Report UCB-ITS-PRR-97-18.

FUNDED PROPOSALS

Principal Investigator in “Automated Data Assimilation and Flight Planning for Multi-Platform Observation Missions,” in *Advanced Information Systems Technology (AIST) Program, Earth Science Technology Office, NASA*, with Robert Morris and Anthony Strawa

Co-Investigator in “Background Interferent Measurements and Standards,” in *Joint Science and Technology Program, Defense Threat Reduction Agency, 2006-2007*

Co-Investigator in “Virtual Sensors,” in *Intelligent Systems Program (NASA)*, directed funding, with Ashok N. Srivastava and Julienne Stroeve, 2003-2005.

HONORS AND AWARDS

- Arch T. Colwell Merit Award (for originality of contribution, one of six papers chosen from among the 3300 papers published by SAE conferences in 2005), 2006. Papers are judged for their value as contributions to existing knowledge of mobility engineering, and primarily with respect to their value as an original contribution to the subject matter.
- Joint Army Navy NASA Air Force (JANNAF) "Outstanding Achievement in Liquid Propulsion Award" for paper with Rodney Martin, Mark Schwabacher, and Ashok Srivastava on the results of Space Shuttle Main Engine data analysis.
- NASA Incentive Award (for collaboration with Puresense Environmental, NASA Research Park), 2003.
- Best Presentation Award, International Joint Conference on Neural Networks, Washington, D.C., 1999
- Schlumberger Foundation Fellowship, 1999
- NSF Graduate Research Training Program in Cognitive Science (Fellowship), 1996-1998
- California Fellowship in Microelectronics, 1994-1995
- MIT Class of 1950 Scholarship 1990-1993
- Life Membership, California Scholarship Federation, 1990

INVITED PRESENTATIONS

Nikunj Oza and Robert Morris, "Automated Data Assimilation and Flight Planning for Multi-Platform Observation Missions," Invited talk, *Special Session on Planning for Data Mining and Data Mining for Planning: Strategies and Methods, IEEE Symposium on Computational Intelligence and Data Mining*, 2007.

"Virtual Sensors," SIAM International Conference on Data Mining, April 2005.

"Being a Civil Servant," AAAI National Conference on Artificial Intelligence Doctoral Consortium, July 2004.

"Artificial Intelligence Applied to Fault Detection," Information Systems Audit and Control Association (ISACA) San Jose State University chapter meeting, April 2004.

"Boosting with Averaged Weight Vectors," Seminar on Computational Learning and Adaptation, Computational Learning Laboratory, Stanford University, Stanford, CA. January 2003.

"Machine Learning and Data Mining in System Health and Safety," California Menay Institute, July 2002.

PROFESSIONAL ACTIVITIES

Member of Editorial Board, Information Fusion Journal.

Session Chair, "Planning for Data Mining and Data Mining for Planning: Strategies and Methods, " Computational Intelligence and Data Mining conference, April, 2007.

Conference Chair, Sixth International Workshop on Multiple Classifier Systems, June 2005.

Member of Program Committee

- SIAM International Conference on Data Mining, 2008.
- Sixth International Conference on Machine Learning and Cybernetics, Multiple Classifier System Theme, 2007.
- Seventh International Workshop on Multiple Classifier Systems, 2007.
- Special Session on Ensembles in Extreme Environments, IEEE Conference on Systems, Man, and Cybernetics, 2005.
- Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security; SPIE Defense and Security Symposium, 2005, 2006, 2007.

Reviewed articles for

- Journals/Books
 - Computational Statistics and Data Analysis
 - IEEE Transactions on Knowledge and Data Engineering
 - IEEE Circuits and Systems
 - IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)
 - Pattern Recognition Letters
 - Encyclopedia of Data Warehousing and Mining
 - Information Fusion (Journal)

- Journal of Intelligent and Fuzzy Systems
- Journal of Machine Learning Research
- IEEE Transactions on Neural Networks
- IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control
- International Journal for Pattern Recognition and Artificial Intelligence
- Conferences/Workshops
 - Computational Intelligence and Data Mining
 - International Workshop on Multiple Classifier Systems
 - IEEE International Conference on Data Mining
 - SIAM Data Mining Conference
 - SIAM Workshop on Data Mining and Machine Learning for Counter-Terrorism
 - SIAM Workshop on High Performance Data Mining
- Grant Agencies
 - Defense Advanced Research Projects Agency
 - Experimental Program to Stimulate Competitive Research

NASA Grant Writing Seminar, September 10-11, 2003.

NASA Project Management course, Jan 27, 2003—Feb 6, 2003.

RESEARCH EXPERIENCE

- Oct. 2005 to Sep. 2006 *Associate Principal Investigator, Integrated Vehicle Health Management Program, Aeronautics Research Mission Directorate, NASA Ames Research Center, Moffett Field, California.*
Managing the Architectures, Databases, and Data Mining element within the IVHM project.
- Oct. 2004 to Sep. 2005 *Program Manager, Knowledge Discovery Tools for System-Wide Security, NASA Ames Research Center, Moffett Field, California.*
Determined research plan for data mining applied to aviation security problems, including flight deviation and cargo screening. Currently manage a group with four other people to implement this agenda.
- Sep. 2001 to Present *Research Scientist, NASA Ames Research Center, Moffett Field, California.*
Worked on applications of machine learning and data mining to aircraft fault detection and diagnosis for aircraft; satellite image understanding; water distribution system monitoring; and aviation security. Currently working on text mining for aviation safety and the integration of planning and data mining for the control of mobile instruments.
- Jun. 1998—Aug. 1998 and Jun. 1999—Aug. 1999 *Summer Researcher, NASA Ames Research Center, Moffett Field, California.*
Worked on input decimation--a new method of correlation reduction in ensemble classifiers that presents different features to different individual classifiers within the ensemble. Achieved classification results several standard deviations above the level of standard neural networks on several well-known datasets within the UCI Machine Learning Repository and numerous synthetic datasets.
- Aug. 1996 to *Member, NSF Graduate Research Training Program in Spatial Cognition, Institute for Cognitive Studies, University of California, Berkeley.*

- May 1998 Worked with Professor Stephen Palmer and students from several departments on the issue of separability of spatial and sequential mental representations as revealed by having subjects view a video depicting movement through a large-scale space and having them draw maps and make ordered lists of events occurring in the video.
- Oct. 1995 *Graduate Student Researcher, Computer Science Division, University of California, Berkeley.*
to
Aug. 2001 Worked with Professor Stuart Russell on Bayesian Automated Taxi project. Worked on learning probabilistic models of automobile driver behavior to allow an automated vehicle to predict the actions of nearby vehicles, thereby improving its own driving. Designed online versions of the bagging and boosting ensemble learning algorithms.
- Jun. 1995 *Researcher, Information Technology Laboratory, General Electric Corporate Research and Development, Schenectady, New York.*
to
Aug. 1995 Worked with Dr. Tomek Strzalkowski on the Natural Language Toolkit, a C++ library of tools to analyze documents. Specifically, designed and implemented module to classify sentences as imperative, declarative, fragments, or questions. Used this module to find procedures---sequences of imperative sentences---in maintenance manuals.

TEACHING EXPERIENCE

- Jan. 1995 *Teaching Assistant, University of California, Berkeley*
to
Dec. 1999 Structure and Interpretation of Computer Programs (Fall 1999), Artificial Intelligence (Fall 1995, Fall 1998), Computer Architecture (Spring 1995).
- May 1999 *Member, Part-time Instructor Pool, City College of San Francisco, California.*
to
May 2000 Awarded membership in part-time pool, allowing members to choose each semester whether to teach and which courses to teach each semester.
- Jan. 1998 *Instructor, City College of San Francisco, California.*
to
May 1998 Designed and lectured intermediate C++ course.
- Jan. 1993 *Instructor, MIT Educational Studies Program, Cambridge, Massachusetts*
to
May 1993 Designed and taught a course in Artificial Intelligence at MIT.
- Sep. 1990 *Instructor, MIT Educational Studies Program, Cambridge, Massachusetts*
to
Nov. 1991 Taught courses on the Mathematics portion of the Scholastic Aptitude Test. (SAT).

OTHER PROFESSIONAL EXPERIENCE

- Nov. 1996 *Independent Item Writer, Educational Testing Service, Oakland, California*
to
Nov. 1998 Independent author of questions for the new Graduate Record Examination (GRE) Mathematical Reasoning (MR) test.

- Jan. 1994 to Aug. 1994 *Engineer, General Electric Nuclear Energy, San Jose, California*
 Evaluated several potential document management solutions for GE Nuclear Energy. Performed Database and Document Management System Administrator duties for Verity TOPIC databases (required C, UNIX, shell script, and MS-DOS batch programming). Established network remote access for Materials Services Site Representatives.
- Jun. 1993 to Aug. 1993 *Summer Intern, General Electric Nuclear Energy, San Jose, California*
 Tested and wrote operator's manual for General Electric's ultrasonic Remote Inspection System and Motion Control System (a four-axis motor control system).
- Jun. 1992 to Sep. 1992 *Summer Intern, General Electric Nuclear Energy, San Jose, California*
 Set up data acquisition and analysis system for fuel rod vibration data analysis using Viewdac software. Established it as part of standard operating procedure for vibration testing by performing required testing and writing appropriate documentation.
- Apr. 1992 to May 1993 *Chairman, MIT Educational Studies Program, Cambridge, Massachusetts*
 Made all major executive decisions and oversaw the work of the directors of each educational program.
- Jan. 1992 to Apr. 1992 *Director, SAT Preparation Program, MIT Educational Studies Program, Cambridge, Massachusetts*
 Directed program of 16 teachers and 150 students. Made all executive decisions specifically involving the SAT Preparation Program, presented reports on this program as a member of the Executive Board, and provided support to teachers.
- May 1991 to Sep. 1991 *Researcher, Undergraduate Research Opportunity Program, Massachusetts Institute of Technology, Cambridge, Massachusetts*
 Designed and wrote a Survival Analysis software package for Professor Michael Rappa of the Sloan School of Management at MIT. Wrote the package in the C programming language for the Apple Macintosh series computers. Designed and implemented the user interface, algorithms, and code to perform survival data analysis.

PROFESSIONAL MEMBERSHIPS

Society for Industrial and Applied Mathematics (SIAM)
 American Association for Artificial Intelligence (AAAI)
 Association for Computing Machinery (ACM)

OTHER

United States citizen (by birth)