Choudur K. Lakshminarayan
Teradata Labs
9390 Research Boulevard
Austin, TX 78759
512-993-3121 (o)
512-983-3879 (c)
choudur.lakshminarayan@gmail.com

Fields of Interest

Research Areas

Large-Scale Machine Learning, Mathematical Statistics, Applied Mathematics, Biostatistics, Statistical Signal Processing, Data Mining, Text Mining, Machine Learning, Parallel Computing, Industrial Mathematics, High Dimensional Statistics, Sparsity and Regularization, Numerical Methods (Gradient Descent and Function Approximations), Kernel Methods, Bayesian Learning, Time Series

Application Areas

Highly Scalable Algorithms, Sensor Analytics, Healthcare and Pharmaceuticals, Energy, Semiconductor Manufacturing, Database Technology, Large Scale Data Centers, Finance, Marketing Optimization, Web Analytics, Digital Marketing, Cyber-Security, Clinical Trials and Pharmaceuticals

Academic Degrees

Doctor of Philosophy, Mathematical Sciences, The University of Texas, 1990
Master of Science, Mathematical Sciences, The University of Texas, 1986

Professional Experience

Industrial

- *Chief Data Scientist (Advanced Analytics), Engineering Fellow, and Sr. Director*, Teradata Labs, 2016-Present
- *Principal Research Scientist*, Big Data Advanced R&D Center, Hewlett-Packard Laboratories, 2013-2016
- *Principal Research Scientist*, Hewlett-Packard Laboratories, 2007-2013
- *Senior Research Scientist*, Hewlett-Packard Laboratories, 2002-2007
- *Sr. Member, Technical Staff*, E-Business Research, Compaq Computer 2001-2002

- *Member, Technical Staff*, Advanced Product Development and Research Labs, Motorola, 2000-2001
- *Member, Technical Staff*, Texas Instruments Incorporated, 1992-2000
- *Analyst/Biostatistician*, Johnson and Johnson, 1988-1989

Startups

- *Co-Founder and Chief Scientific Officer*, Pattern Sciences LLC
  - Startup involved in Data Science and Machine Learning for Healthcare and Defense applications, Austin, TX
- *Mentor*, Capital Factory, 2016-Present
  - Incubator and accelerator for start-ups, Austin, TX
- *Scientific Advisor,* Imanage*,* 2016-Present
  - Document analytics company for clustering, classification, and legal fraud detection, Chicago, IL

Academic

- **Adjunct Assistant Professor**, Department of Statistics and Data Sciences, The University of Texas at Austin, 2018-Present
- Faculty Affiliate (SSI), Department of Statistics and Data Sciences, The University of Texas at Austin, 2012-Present
- Visiting Professor, Department of Electrical Engineering, Indian Institute of Technology, Hyderabad, 2012-2013
- Professor, Indian Institute of Information Technology, Bangalore, 2003-2004
- Adjunct Lecturer, Department of Mathematical Sciences, University of Texas at Dallas, 1992, 1998
- Adjunct Faculty Member and Lecturer, Department of Management Sciences and Information Systems, University of Texas at Arlington, 1989-1992
- Graduate Research Assistant, Departments of Mathematics, Finance, and Information Systems, The University of Texas at Austin, 1986-1989
- Graduate Teaching Assistant, Department of Mathematics, The University of Texas at Austin, 1983-1986
- Research Scholar, Production and Quantitative Methods Department, Indian Institute of Management, Ahmedabad, 1981-1983

Industrial Training

- Design and Analysis of Statistical Experiments, Texas Instruments Incorporated
- Statistical Process and Quality Control, Texas Instruments Incorporated

Consulting

1. LiveOak Inc., California
   o Machine Learning and Forecasting for luxury real estate price prediction
2. 7X.Energy, Texas
   o Battery Storage Optimization and Bidding for wind and solar substations
3. Rollick, Austin, TX
   o Campaign Optimization for forklifts by Toyota
4. BrilliantMD, Austin, TX, 2016-2017
   o Analytical pipelines for optimized care-giving from acute care to rehabilitation
5. Move Inc., A Newscorp Company, Santa Clara, CA
   o Machine Learning and Forecasting for real estate price prediction
6. Hologic Systems, Marboro, MA, 2017
   o Image Recognition for automated cervical cancer detection
7. Alberta Open Data Institute, Government of Canada
   o Public data analytics for Healthcare, Energy, and other public policy verticals
8. McKinsey & Company, Institute for Corporate Excellence, Washington, D.C., 2007
   o Text Analytics and Mining methods to analyze enduring corporate excellence
9. BHEL, Research and Development, Government of India, 2007
   o Machine Learning for gas turbine diagnostics and monitoring
10. B2K Corporation and Brickwork, Bangalore, India, 2003-2005
    o Quantitative methods, Analytics, and Training in finance, retail, business process outsourcing, and engineering
11. National Bank of Kuwait, Kuwait City, Kuwait, 2004
    o Customer recruitment and retention for project on "Analytics in Banking
12. Revenue Department, Government of Karnataka, Karnataka, India, 2003
    o Mining government land records data to detect trends in crop yields, mutation time for deed transfers, and other aspects of government-owned lands and tracts
13. World Omni Financial Corporation (subsidiary of Toyota), Deerfield Beach, FL, 2000
    o Business Intelligence and Customer Relationship Management analytics for customer segmentation to determine leasing and buy-out strategies for loss mitigation in the automobile industry
14. Office of Civil Rights, United States Government, 1992
    o Analysis of discriminatory practices in granting athletic scholarships in colleges and universities within the State of Texas

Short Courses and Workshops

1. 13th Summer Statistics Institute, The University of Texas at Austin, May 2020
   - *"Big Data Analytics: Structured, Semi-Structured, and Unstructured"*
2. Teradata Labs, Spring 2020
   - *"Large-Scale Machine Learning and Parallel Computation"*
3. American Statistical Association, American University, June 2019
   - *"Big Data Analytics"*
4. 12th Summer Statistics Institute, The University of Texas at Austin, May 2019
5. 11th Summer Statistics Institute, The University of Texas at Austin, May 2018
6. 10th Summer Statistics Institute, The University of Texas at Austin, May 2017
7. 9th Summer Statistics Institute, The University of Texas at Austin, May 2016
8. 8th Summer Statistics Institute, The University of Texas at Austin, May 2015
9. 2nd International Conference on Big Data Analytics, Mysore, India, 2013
   - *"Tutorial on High Dimensional Big Data"*
10. 7th Summer Statistics Institute, The University of Texas at Austin, May 2014
    - *"Multivariate Methods in Big Data: Theory and Methods"*
11. 6th Summer Statistics Institute, The University of Texas at Austin, May 2013
    - *"Predictive Analytics and Big Data: Theory and Methods"*
12. SAS Institute, Beijing, China, December 2009
    - *"Marketing Optimization, From Theory to Practice"*
13. SAS Institute, Shanghai, China, December 2008
    - *"Data Mining in Financial Applications"*
14. National Bank of Kuwait, Kuwait, 2008
    - *"Predictive Analytics and Statistical Modeling in Banking"*
15. SAS Institute, Hong Kong, December 2007
    - *"Data Mining, From Theory to Practice in Banking and Finance"*
16. SAS Institute, Bangalore, India
    - *"From Analytics to Profit"*
17. McKinsey & Company, Institute for Corporate Excellence, Washington, D.C., 2006
    - *"Text Mining and Unstructured Data Analysis: Mining annual reports, quarterly earnings reports, news reports, and analyst reports to extract insights about corporate performance indices"*
18. Semiconductor Manufacturing Tech Consortium (SEMATECH), Austin, TX, 1999
    - *"Intelligent Data Analysis in Semiconductor Manufacturing and Research"*

Publications

Patents Issued

1. Incrementally Updating a Database Statistic, 10,922,314
2. Singular Value Decompositions, 10,012,1527
3. Segments of Contacts, 10,565,603
4. Managing Database with Counting Bloom Filter, 10,452,676
5. Incrementally Updating Statistics, 10,430,411
6. Estimating Unique Entry Counts Using a Counting Bloom Filter, 9,465,826.
5. Generating a query plan for estimating a number of unique attributes in a database, 9,576,027
6. Energy Based Wavelet Thresholding, 9,405,731
7. Anomaly Detection in Streaming Data, 9,218,527
8. System and Method for Ranking Anomalies, 9,141,914
9. Estimating the number of unique values in a list, 9,158,815
10. Detecting Regime Change in Streaming Data, 9,078,629
11. Classifying Unclassified Samples, 9,037,518
12. Estimation of Unique Database Values, 8,549,004
13. Website Analysis using Quantitative and Qualitative Data, 8,396,737
14. Method and System for Site Path Evaluation Using Web Session Clustering, 8,572,223
15. Method and apparatus for identifying anomalies of a signal, 8,620,609
16. A system for accurately predicting revenues given a portfolio of investments, 8,180,693
17. A system and method for incorporating expert opinion in statistical models, 8,180,694
18. A system and method for optimizing financial performance generated by marketing investments under budget constraints, 8,027,897
19. Multi Regime Detection in Streaming Data, 8,620,987
20. Anomaly Detection in Data Centers, 8,668,620
21. Classification of a signal in a time domain, 8,706,203
22. Compression of Non-dyadic Sensor data Organized within a Non-dyadic Hierarchy, 8,898,209
23. System and Method for estimating unique attributes in a database, 7,428,550
24. Processing Capability Metrics in On-line Monitoring (Defensive Publication), 700210682
25. Autoregressive Model for Time Series Model (Allowed for Grant), 83209810
26. Methods and Systems for Processing Data Arrays Using Bloom Filters (Allowed for Grant), 82837112

Patents Pending and Filed

1. System and methods for acceleration of machine learning functions, Pending
2. A Graph Theoretic Learning Model for Generating Edges and Missing Value Imputation (700221953), Disclosed
3. Information Theoretic Models for Statistical Classification (84224822), Pending
4. Anomaly detection in utility clouds using parametric and non-parametric statistical approaches (201005668), Filed
5. Using Wavelet Transforms for Smoothing in The Estimation of Regression Parameters For Applications In Marketing (200905071), Filed
6. Wavelet Compression with Bootstrap Sampling (20110184934), Published
7. Clustering and Analysis of Electronic Medical Records (83166859), Filed
8. Dataset Compression (20130191309), Published

Invention Disclosures

1. A System and Method for Analytical Processing and Image Recognition, DN 17-1054
2. A System and Method for Predicting Customer Behavior, DN 17-1053
3. Scalable Solution for Dimensionality Reduction by Maximizing Mutual Information, DN 18-1055

Papers

1. "Social Determinants of Recidivism: A Machine Learning Solution", arXiv:2011.11483, 2020 (with Vikrant Shirvaikar)
2. "Topological Data Analysis in Digital Marketing", Applied Stochastic Models in Business and Industry, John Wiley & Sons, vol. 36(6), pp. 1014-1028, November 2018 (with Mingzhang Yin)
3. "Enterprise-wide Machine Learning Enterprise Solution using Teradata Vantage: An Integrated Analytics Platform," IEEE International Conference on Big Data, pp. 2043-2046, 2019 (with Khaled Bouaziz, Awny Al Omari, Thiagarajan Ramakrishnan, Faraz Ahmad, Srinivasan Raghavan, Prama Agarwal)
4. "Model Management and Handwritten Character Recognition", pp. 6110-6112 (with Khaled Bouaziz, Awny Al Omari, Thiagarajan Ramakrishnan, Faraz Ahmad, Srinivasan Raghavan)
5. "Character Recognition by Deep Learning: An Enterprise Solution", IEEE International Conference on Big Data, pp.1717-1727, 2018 (with Khaled Bouaziz, Thiagarajan Ramakrishnan, Srinivasan Raghavan, Kyle Grove, Awny Al Omari)
6. "Modeling Complex Clickstream Data by Stochastic Models: Theory and Practice", WWW (Companion Volume), pp. 879-884, 2016 (with Ram Kosuru, Meichun Hsu)

7. "Automatic Classification of Heartbeats", European Signal Processing Conference, pp. 1542-1546, 2014 (with Tony Basil)

8. "High Dimensional Big Data and Pattern Analysis: A Tutorial," Lecture Notes in Computer Science, vol. 8302, pp. 68-85, 2013

9. "Pattern Recognition in Large-Scale Data Sets: Application in Integrated Circuit Manufacturing", Lecture Notes in Computer Science, vol. 8302, pp. 185-196, 2013

10. "Adaptive Estimators of Process Capability Indices Using Preliminary Test," JSM Proceedings, Section on Statistical Quality and Productivity, American Statistical Association, 2013 (with Chien-Pai Han)

11. "Robust Versions of the Tukey Boxplot with Their Application to Detection of Outliers," Proceedings of IEEE, ICASSP, pp. 6506-6510, 2013 (with Georgy Shevlyakov, Kliton Andrea, Pavel Smirnov, Alexander Ulanov, Natalia Vassileva)

12. "Detection of Classes of Heart Arrhythmias based on Heartbeat Morphology Patterns," Proceedings of 2nd International Workshop on Analytics for Cyber-Physical Systems, 2013 (with Tony Basil, C. Krishnamohan)

13. "A comparison of Statistical Machine Learning Methods in Heartbeat Detection and Classification," Big Data Analytics, Lecture Notes in Computer Science, vol. 7678, pp. 16-25, 2012 (with Tony Basil, Bollepalli Chandra)

14. "Nearest Neighbor Distributions for Imbalanced Classification," Proceedings Of IEEE-INNS International Joint Conference on Neural Networks, pp. 1-5, 2012 (with Jose C. Principe, Evan Kriminger)

15. "Mixture of designer-experts for multi-regime detection in streaming data", Proceedings of 20th European Signal Processing Conference (EUSIPCO), pp. 410-414, 2012 (with Jose C. Principe, Evan Kriminger),

16. "Ultra low power automaton for heartbeat classification based on integrate and fire sampler," Machine Learning and Signal Processing Conference (MLSP), 2012 (with Gabriel Nallathambi, Jose C. Principe)

17. "Live Operational Intelligence-Characterizing Operational Processes," HP Business Intelligence Conference, 2012 (with Chetan Gupta et al.)

18. "Enabling Partial Data Cube Computations using the Bloom Filter," Proceedings of the 1st Intl. Workshop on Analytics for Cyber-Physical Systems, 2012 (with Ram Kosuru)

19. "Ranking Anomalies in Data Centers," Proceedings Of 13th IEEE Network Operations and Management Symposium (NOMS) pp. 79-87, 2012 (with W. Chengwei, K. Viswanathan, V. Talwar, S. Wade, G. Macdonald)

20. "Time-based Compression and Classification of Heartbeats", IEEE Transactions on Biomedical Engineering, vol. 59, pp. 1641-1648, (with Alexander Singh Alvarado, Jose C. Principe)

21. "Early Detection of Anomalous Patterns in Sensor Streams," Tech Con '12, 2012 (with Alan Benson, Ram Kosuru)

22. "Modified embedding for multi-regime detection in non-stationary streaming data", Proceedings of ICASSP, pp. 2256-2259, 2011 (with Jose C. Principe, Evan Kriminger)

23. "Time encoding using the integrate and fire sampler: Discriminative Representation for Neural Action Potentials", Proceedings of International Conference on Sampling Theory and Applications, 2011 (with Alexander Singh Alvarado, Jose C. Principe)

24. "Local Frequency-Based Estimators for Anomaly Detection in Oil and Gas Applications", JSM Proceedings, Section on Statistical Learning and Data Mining, American Statistical Association, 2011 (with Alexander Singh Alvarado, Evan Kriminger)

25. "Statistical Techniques for Online Anomaly Detection in Data Centers", Proceedings of 12th IFIP/IEEE International Symposium on Integrated Network Management, pp. 385-392, 2011(with W. Chengwei, K. Viswanathan, V. Talwar, S. Wade, S. Karsten)

26. "Better Drilling Through Sensor Analytics: A Case Study in Live Operations Management," Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, pp. 8-15 (with Umeshwar Dayal et al.)

27. "Finding Needles in Haystacks: Analytic Techniques for Anomaly Detection in Data Centers," Tech Con'11, 2011, (with Nigel Cook, Jeff Hilland, Partha Ranganathan, Karsten Schwan, Krishnamurthy, Vishwanatha, Vanish Talwar, Chengwei Wang)

28. "Applications of Sensor Analytics over Streaming Data in Oil Production," in Proceedings of 17th International Conference on Management of Data (COMAD), 2011 (with Umeshwar Dayal et al.)

29. "Statistical Techniques for Online Anomaly Detection in Data Centers", HP Labs Technical Report, 2011 (with Chengwei Wang, Krishnamurthy Viswanathan, Vanish Talwar, Wade Satterfield, Karsten Schwan)

30. "Parallelizing Statistical Operators in a Database Management System Framework," HP Labs Technical Report, 2011 (with Suresh Subbiah, Robert Wehrmeister)

31. "Online Detection of Utility Cloud Anomalies Using Statistical Metric Distributions", Tech Con '10, 2010 (with Nigel Cook, Jeff Hilland, Partha Ranganathan, Karsten Schwan, Ram, Swaminathan, Vanish Talwar, Chengwei Wang)

32. "Non-Dyadic Haar Wavelets for Streaming and Sensor Data," Proceedings of IEEE 26th International Conference on Data Engineering, pp. 569-580, 2010 (with Chetan Gupta, Song Wang, Abhay Mehta)

33. "On Wavelet Compression and Cardinality Estimation in Enterprise Data", HP Labs Technical Report, 2010 (with U. Dayal, G. Chetan, R. Swaminathan)

34. "Using page sequence analytics to understand complex web behavior", JSM Proceedings, Section on Statistical Learning and Data Mining, American Statistical Association, pp. 3926-3929, 2009 (with Alan Benson)

35. "Adaptive wavelet filtering in database applications", International Conference on Time-Frequency Analysis, 2009

36. "Marketing Optimization for Increased Revenues", 12th Annual Data Mining Conference, 2009

37. "A new composite estimator of distinct values that performs well in a wide range of skewness", HP Labs Technical Report, 2008 (with Vinay Deolalikar, Hernan Laffitte)

38. "Marketing Innovation: Budget Optimization for Increased Revenues," Proceedings of the 15th Annual Workshop of the HP Software University Association (HP-SUA), pp. 247-253, 2008

39. "A New UEC Estimator for Neoview: Enhancing Query Throughput", Tech Con' 08, 2008 (with Vinay Deolalikar et al.)

40. "Session-centric page sequence clustering for improved web experience on hp.com", Tech Con '05, 2005 (with Alan Benson)

41. "Improving Customer Comments via Text Mining", Lecture Notes in Computer Science, Springer Berlin, pp. 288-299, 2005 (with Alan Benson, Qingfeng Yu)

42. "Vector Space Based Text Mining Models", 7th Annual Data Mining Conference, 2005

43. "Automated Comment Analysis", Tech Con' 04, 2004 (with Alan Benson)

44. "Quantitative Web Analytics and Qualitative Usability Research-Strange Bedfellows", Tech Con '03, 2003 (with Alan Benson)

45. "Markov Random Fields in Pattern Recognition for Semiconductor Manufacturing", Institute of Mathematical Statistics/American Statistical Association, Spring Research Conference, 2001

46. "Markov Random Fields in Pattern Recognition for Semiconductor Manufacturing", Technometrics, vol. 43, pp. 66-72, 2001 (with Michael Baron, Zhenwu Chen)

47. "Pattern Recognition in IC Diagnostics Using the Linear Discriminant Classifier and Artificial Neural Networks", IEEE Transactions in Semiconductor Manufacturing, 2000 (with Michael Baron, Chen Zhenwu)

48. "Spatial Statistical Interpolation Methods for Image Enhancement," INRIA, Sophia Antopolis, 1999 (with Mukul Shirvaikar)

49. "A Preliminary Test Estimator of Process Capability Index", Texas Instruments Technical Journal, 1998 (with Chien-Pai Han)

50. "Knowledge Discovery and Data Mining for Improved Manufacturing Quality - An Overview", Texas Instruments Technical Journal, 1998

51. "A biased estimator of the process capability index", International Indian Statistical Association, 1998

52. "On Estimating the Mean in a Bivariate Normal Distribution", Journal of Statistical Simulation and Computation, vol. 58, pp. 155-170, 1997 (with Chien-Pai Han)

53. "Signature Analysis: Statistical Methods and Their Applications to Failure Analysis", Proceedings of the 22nd International Symposium for Testing and Failure Analysis, pp. 183-187, 1997. (with Seshu Pabbisetty and Chien-Pai Han)

54. "Introducing Dependency in Integrated Circuit Signature Analysis Models", The University of Texas at Dallas Technical Report, No. 238, 1997 (with Michael Baron)

55. "Signature Analysis-Based IC Diagnostics," Proceedings of the 6th International IEEE Symposium on the Physical and Failure Analysis of Integrated Circuits, pp. 167-171, 1997 (with Seshu Pabbisetty and Chien-Pai Han)
56. "An Auto-Logistic Model for the Joint Distribution of Failed Chips on a Regular Lattice", Proceedings of the American Statistical Association Section on Physical and Engineering Sciences, pp. 75-80, 1996 (with Michael Baron)
57. "An Adaptive Estimator of Location Based on the t-family", Communications in Statistics, Theory and Methods, vol. 23, pp. 747-761, 1996 (with D.L. Hawkins)
58. "Detecting Outliers in Semiconductor Wafer Yields and Manufacturing", Texas Instruments Technical Journal, pp. 52-57, 1993
59. "Estimation of Regression Coefficients Under Multicollinearity Condition", Non-Parametric Statistics and Related Topics, Saleh (Ed.), Elsevier Publishers, pp. 297-309, 1992 (with Chien-Pai Han)

Lecture Notes, Manuals, and Books

1. *Data Mining and Analytics for Decision Making*
2. *Data mining and Analysis: Theory to Practice*
3. *From Analytics to Profit using Statistical Modeling*
4. *Statistics in Data Mining: Exploration, Modeling, and Analysis* (unpublished manuscript)
5. *Pattern Analysis, Predictive Analytics, and Big Data: Theory and Methods* (manuscript in preparation), CRC Press, Taylor & Francis, Spring 2021, ISBN 978-1-4987-5492-7

Invited Talks

1. Data Science Colloquium, College of Natural Sciences, The University of Texas at Arlington, 2020
2. Capital Factory Innovation Forum, Austin, TX, 2019
3. Fall Seminar Series, Department of Statistics and Data Sciences, The University of Texas at Austin, 2019
4. Loyola Institute of Business Administration, Loyola College, Chennai, India, 2018
5. Distinguished Lecture, Department of Industrial Engineering, College of Business, Northwestern University, 2017
6. The University of Texas at Arlington, 2017
7. S. Rajagopalan Endowed Lecture, Department of Statistics, University of Madras, 2017
8. Department of Mathematics, American University, Washington D.C., 2016
9. Distinguished Lecture, Department of Operations Management and Management Sciences, The University of Texas at Arlington, 2016
10. The University of Texas at San Antonio, 2016
11. National Science Foundation Lecture, The University of Texas at Tyler, 2016
12. Department of Computer Science, The Hyderabad Central University, 2013
13. Numerical and Harmonic Analysis Group, University of Vienna, 2012

14. Department of Electrical Engineering, Department of Mathematics, Indian Institute of Technology, 2012
15. Indian Institute of Science, 2012
16. Department of Computer Science, University of North Texas, 2010
17. Department of Computer Science, Imperial College, London 2002
18. Department of Statistics, University of Madras, 2002
19. Department of Operations Research, McCombs School of Business, The University of Texas at Austin, 2001
20. INRIA, Sophia Antopolis, Nice, France, 2000
21. Department of Statistics, Osmania University, Hyderabad, India, 2000

Research Grants (as Principal Investigator or Co-PI)

1. "Borrowing Digital Data Across Studies in Clinical Trials Designs" (with Dr. Peter Mueller, The University of Texas, Dr. Robertino Mera, Gilead Sciences)
2. "Talent Management and Pilot Training Analytics," U.S. Air Force, *$100,000*, Pattern Sciences LLC
3. "Sensor data, sparse sampling, and signal reconstruction," Department of Electrical Engineering, The Indian Institute of Science, Bengaluru, *$10,000* (with Dr. Seelamankula Chandrasekhar)
4. "Comparative Effectiveness for Evidence-based Health Care Advances," Innovation Research Program, HP Laboratories, 2012-2013, *$75,000* (with Dr. Eva K. Lee, Georgia Institute of Technology)
5. "Real-Life Data Analysis: Distribution Model Fitting and Anomaly Detection," Innovation Research Program, HP Laboratories, 2012-2013 (with Georgy Shevlyakov, St. Petersburg State Polytechnic University)
6. "Anomaly Detection in Multivariate Data Streams using Kernel Methods and Information Theoretic Cost Functions," Innovation Research Program, HP Laboratories, 2010-2011, *$50,000* (with Dr. Jose C. Principe, University of Florida)
7. "Sparse Sampling, Stochastic Point Processes, and Classification of Signals," Innovation Research Program, HP Laboratories, 2011-2012, *$50,000* (with Dr. Jose C. Principe)
8. "Statistical Models for Signature Analysis," Texas Instruments Corporate Technical Council, 1996-1997, *$25,000* (with Dr. Chien-Pai Han)
9. "Statistical Pattern Recognition for Yield Enhancement and Cycle Time Reduction," Texas Instruments Corporate Technical Council, 1997-1998, *$25,000* (with Dr. Michael Baron, University of Texas at Dallas)
10. "Applying Markov Random Fields to Model Failed Chip Patterns in Semiconductor Manufacturing," Advanced Technology Program of Texas University System, 1998 (with Dr. Michael Baron)

Significant Industry Projects and Programs

1. Digital Marketing
Patented technologies for predicting online user behavior using page navigation sequences on enterprise websites, through stochastic models and computational topology.

2. Scientific Computation
Highly scalable and parallel algorithms for large-scale machine learning and efficient implementation and regularization of algorithms in big data analytics.

3. Randomized Clinical Trials
Statistical designs which combine cohorts who volunteered for an experimental therapy (treatment group) with a synthetically generated control group of patients found in real world data. Adjustment for lack of randomization to prevent misleading comparisons when patients exposed to the treatment group differ systematically from observed data.

4. Business Intelligence and Live Analytics
Lead scientist for creation of a unified data and analytics platform, called "Live Business Intelligence (BI)", that shifts BI from the traditional back-office, report-generation orientation, to an enabler for delivering data and computationally-intensive mathematical methods that transform operational business processes and customer interactions for near-real-time decision making. Massive parallel processing, supporting analytics over both streaming and historical data, for pattern detection by non-linear dynamical systems, detectors based on time-frequency procedures, auto-regressive processes, and anomalies.

5. Comparative Effectiveness Research
Assessment, analysis, and future directions of comparative effective research (CER) for improving healthcare quality and efficient delivery using the latest scientific and technological advances. Creation of a blueprint for personalized medical care of patients with chronic illnesses that controls escalating costs and increases patient satisfaction.

6. Sensors, Sensing, and Analytics
Investigation of embedding sampling algorithms for compression at the sensor level as well as analytics in the compressed domain. Implementation of event-based samplers in low power analog VLSI circuits for sampling and low-bandwidth transmission, treating data as realizations of an underlying stochastic point process for real-time signal classification, detection, and prediction.

7. Query Optimization in Relational Database Systems
Technology to optimize queries in large databases by providing an estimate of the number of distinct values in specified columns of the database, also known as unique entity count. Implementation of a product based on a family of generalized Jackknife estimators, resulting in a market differentiator now used by the New York Stock Exchange, Yahoo, Wal-Mart and other major companies

8. Multi-patent Technology for Marketing Optimization
Lead scientist for the effective management of a $500 million marketing budget to determine investments in marketing instruments and maximize revenue and profit. Application of regression analysis, signal processing, mathematical optimization, and qualitative analysis to build solutions and a framework to optimally allocate investments, resulting in a program implemented in Germany, Spain, Netherlands, Canada, Korea, Australia, and Mexico with incremental revenue across countries of ~$30 million.

9. Total Customer Experience and Satisfaction Models
Lead scientist for development of text mining to process and understand customer concerns collected in on-line surveys by the total customer experience (TCE) group on www.hp.com. Application of latent semantic indexing, the EM algorithm, numerical linear algebra procedures and Gaussian mixture models to read, collate, and cluster unstructured (textual) customer comments, and identify issues that were previously undetected. Use of non-parametric density estimation in statistical classification to track customer satisfaction by problem areas, increasing customer satisfaction by up to 5%.

10. On-line Customer Behavioral Models and Analysis
Lead statistician of a team that set up a ~10 TB database of on-line customer information, using Keylime and Omniture web analytics systems, to capture client-side on-line customer click-stream data in a central repository at for the first time at HP. Development of key technical ideas, web metrics, and mathematical methods to find predictors of purchase which helped to double revenues.

11. Integrated Circuit Fault Diagnostics
Lead scientist for the development of an automated system for detecting faults in integrated circuit (IC) manufacturing in a modern wafer fabrication facility. Application of artificial neural network systems and generalized Markov random fields to analyze manufacturing defects and find the root cause of defects, resulting in cost savings on the order of millions and reduction in failure analysis cycle times from days to hours.

12. Motorola Advanced Products and Research Development Labs (APRDL)
Lead scientist on the effort to set up a statistics laboratory for analysis of semiconductor processing and measurement data. Creation of a centralized database of chip manufacturing and failure analysis data, and an on-line system to identify process problems in real-time to minimize faulty product, minimize cycle-time, and improve yield.

13. Texas Instruments
Lead statistician to develop software tools for intelligent data analysis and pattern recognition. Development of algorithms for IC fault diagnostics, process control, quality control, and image enhancement in digital signal processing chips.

Service

Technical Positions (Elected)

1. Nominated for **Fellowship of the American Statistical Association**, 2019
2. **Engineering Fellow**, Teradata Labs, 2017-Present
3. **Master Technologist**, Technical Career Ladder, Hewlett-Packard Labs, 2003-2016
4. **Senior Member of Technical Staff**, Advanced Technical Career Ladder, The Compaq Computer Corporation, 2001-2003

Leadership and Committee Memberships

1. Quality and Productivity Research Conference (QPRC), American Society for Quality, Program Committee, 2019
2. American Statistical Association, Austin Chapter, Speaker Recruitment, Nominee for Vice President, 2018-Present
3. IEEE Computer Science and EMBS, Austin Chapter, Communications Director, 2013-2016
4. SIAM International Conference on Data Mining, Program Committee
5. International Conference on Big Data Analytics, Program Committee, 2012
6. HP Tech Con '11, Review Committee, 2011

Referee Activities

1. Technometrics (Journal of American Society of Quality and Technology)
2. Communications in Statistics, Theory and Methods
3. IEEE Transactions
4. VLDB (International Conference on Very Large Databases)
5. ICDE (International Conference on Data engineering)
6. ICDM (International Conference on Data Mining)
7. DEXA (Conference on Databases and Expert Systems)
8. DNIS (Databases and Networked Information Systems)
9. DKE (Databases and Knowledge Engineering)
10. SIAM (Society for Industrial and Applied Mathematics)
11. EDBT (Enterprise Database Technology Conference)
12. HP Tech Con (Annual Technical Conference of HP Technologists)
13. American Society of Quality
14. SIGMOD, Association of Computing Machinery
15. ACM SIGKDD

Advising

1. Erin Burgoon, Ph.D Thesis, Department of Psychology, The University of Texas at Austin, 2014

2. Tony Basil, M.Tech Thesis, Department of Computer Science and Engineering, The Indian Institute of Technology, Hyderabad, "Features for Detection of Heart Arrhythmias", 2013
3. Qingfeng Yu, Ph.D. Thesis, Department of Electrical and Communication Engineering, The University of Texas at Austin, "Human Extremity Detection and Its Applications in Action Detection and Recognition", 2009

Mentorship

1. Harshita Vemula, Ph.D. Student, Department of Industrial Engineering and Operations Research, The University of Texas at Austin, 2018-
2. Thiagarajan Ramakrishnan, Software Engineer, Teradata Labs, Austin, TX, 2018-
3. Nikhil Chakravarthula, Software Engineer, Teradata Labs, Hyderabad, India, 2019-
4. Dr. Faraz Ahmad, Sr. Architect, Teradata Labs, Santa Clara, 2017-
5. Vikrant Shirvaikar, Data Science Intern, Teradata Labs, Undergraduate Student, The University of Texas at Austin, 2017-
6. Mingzhang Yin, Ph.D Student, The University of Texas at Austin, Summer Intern, HP R&D, 2016
7. Erin Burgoon, Ph.D Student, The University of Texas at Austin, Summer Intern, HP R&D
8. Matthew Hagen, Ph.D Student, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2013
9. Alexander Singh Alvarado, Ph.D Student, Department of Electrical Engineering, University of Florida, 2011
10. Evan Kriminger, Ph.D Student, Department of Electrical Engineering, University of Florida, 2011
11. Qingfeng Yu, Ph.D. Student, Department of Electrical Engineering, The University of Texas at Austin, 2009
12. A. Bharat, Ph.D. Student,. Birla Institute of Technology and Sciences, Pilani, India
13. Kashif Siddique, M.S. Student, Department of Computer Science, The University of Texas at Austin, Non-stop Engineering Division, HP, 2005
14. Greg Nolder, M.S. Student, Department of Management Sciences, Southern Methodist University, Dallas, TX, 2003
15. Jay Gutzler, Ph.D. Student, Department of Mathematics, University of Texas at Arlington, Texas Instruments, 1997-1999
16. Zhenwu Chen, Ph.D. Student, Department of Mathematics, University of Texas at Dallas, Texas Instruments, 1999-2000

Awards and Honors

1. "Topological Analysis in Digital Marketing," Special Invited Paper, International Society of Business and Industrial Statistics (ISBIS) Conference, 2021
2. HP Innovation Award, 2013
3. HP Employee Award, 2008 (for "Update Statistics for Query Optimization in SQL/MX Databases")
4. HP Employee Award, 2007 (for "A standardized platform for optimal marketing")
5. HP Employee Award, 2007 (for "A robust estimator of number of unique attributes in a database")
6. HP Star Award, 2006 (for "Return on Marketing Investment")
7. Marketing Circle Award Finalist, Marketing Innovation, 2006 (for "Return on Marketing Investment")
8. HP Best Innovation Finalist, 2006 (one of 19 from a total of 433 international entries)
9. "Markov Random Fields in Pattern Recognition for Semiconductor Manufacturing", Special Invited Paper, Institute of Mathematical Statistics/American Statistical Association Spring Research Conference, 2001
10. Texas Instruments IDEA Program Award, 2000 (for "Image Enhancement using Gaussian Kernel Methods")
11. Texas Instruments IDEA Program Award, 1998 (for "Pattern Recognition for IC Failure Diagnostics")

Personal

Married with no children, United States Citizen