



# Soma S. Dhavala

---

#5, 15th Cross, Apt 201, Bhadra Landmarks  
Girinagar 2nd Phase, Banashankari 3rd Stage  
Bengaluru, 560085, Karnataka, India.

Phone: +91 990 209 6435

Email: [soma.dhavala@gmail.com](mailto:soma.dhavala@gmail.com)

LinkedIn: <http://www.linkedin.com/in/somasdhavala>

Github: <https://github.com/dhavala>

## Experience

### Summary

---

I have over twenty years of experience in

- Formulating problems with academic rigor
- Implementing industrial grade solutions
- Applying them to various business contexts, spanning verticals such as Agriculture, Health Care, Life Sciences and Education, among others
- Operating effortlessly at the interface of Engineering and Science

#### Academic Leadership:

- Trained in both Frequentist and Bayesian approaches, Applied Maths, Statistics
- Exposed to extensive coursework, both formal and self-learnt, in Communications, Signal Processing, Algorithms and Data Structures, Object Oriented Design, Applied Maths, Programming
- At ease with evaluating, improvising, and applying hundreds of classical statistical techniques, and state-of-the-art literature in Statistics, ML, AI, Deep Learning and Computer Science

#### Development:

- Built and deployed large scale scientific applications
- Gained first hand experience in architecting large scale, data-first, microservices platforms
- Applied platform thinking to developing tools and best practices to democratizing AI
- Hands-on with various computational frameworks
- Handled both Product and Consulting Services portfolios
- Contributed to open source software

#### Teaching and Mentoring:

- Taught undergraduate courses
- Mentored senior graduate, Ph.D students, and post-docs
- Designed and conducted bootcamps for fresh graduates in Industry
- Advised Data Science teams

## Skill

### Summary

---

ML Areas:

- Frameworks for Democratizing AI & Deep Learning, Machine Learning As Infrastructure, Deep Learning, NLP, IoT, Semantic Computing, Social Network Analysis, Dependence modeling, Copulas & Quantiles, Nonparametric regression, Bayesian Modeling, Markov-Chain Monte Carlo (MCMC), Bayesian Modeling, Approximate Bayesian Computing, Parallel & Distributed Computing

Languages

- Python, R, MATLAB, Julia, JavaScript, Scala

ML Frameworks

- Python: PyTorch, keras, scikit-learn, pandas, networkx, etc..
- R: tidy, dplyr, ggplot2, igraph, caret, mlr, etc..
- Bayesian: winBUGS, stan, Pyro

Distributed Computing:

- Apache Spark, TORQUE

Databases:

- mysql, cassandra, neo4j, druid, elasticsearch

Web, Networking and Cloud:

- Familiarity with MEAN Stack, Ember, Polymer, Web Components, OAuth, Sockets, Load Balancers, Proxying, Firewalls, DNS, SSL, Google Cloud Computing Platform

IoT:

- Hands-on with Raspberry Pi, Motor Controllers, Power Supplies, Arduino boards, messaging protocols

Architecture:

- Design Patterns, MicroServices, Serverless Computing

Publishing:

- MS Office Suite, LaTeX, G Suite

## Education

---

### University of Washington / M.S

2020 - 2022, Seattle, USA (online)

Masters in Applied & Computational Maths:

Dynamical Systems, Stochastic Processes, PDEs, Numerical Linear Algebra, Deep Learning in NLP

### Texas A & M University / Ph.D

2005 - 2010, College Station, TX, USA

Advisor: Prof. Bani K. Mallick

Topic: Bayesian Semiparametric Models For Heterogeneous Cross-Platform Differential Gene Expression [[thesis](#)]

Summary:

- Established connection between FDR control, and Knapsack problem
- Proposed a novel Bayesian non-parametric method to analyze over-dispersed count data and designed, and implemented efficient MCMC samplers for inference
- Proposed a novel Bayesian non-parametric meta analysis method to fuse data from heterogeneous sources, and designed and implemented efficient MCMC samplers for inference in a high dimensional setting
- Mentored senior graduate students, and postdocs in  
Bayesian paradigm, Designing MCMC samplers, Parallel Computing, High Dimensional Graphical Models
- Taught undergraduate courses in Statistics, with complete responsibility, including designing the teaching material, setting homework assignments and exam papers, and grading
- Consulted graduated students from various disciplines such as Construction Management, Civil Engineering, Electrical Engineering, Education Psychology, Food Sciences, to name a few. Some of them got best dissertation awards, in their respective departments
- Volunteered and played leadership roles in various community services

### Iowa State University / Ph.D (transferred to TAMU)

2003 - 2005, Ames, IA, USA

Advisor: Prof. Alexander Doganzic

Topic: Distributed Estimation Techniques

Summary:

- Took courses in Linear Algebra, Statistical Signal Processing, Bayesian Analysis
- Moved to TAMU to pursue PhD in Statistics

### Indian Institute of Technology, Madras / M.S

1997 - 2000, Chennai, TN, India

Advisor: Prof. K. M. M. Prabhu, Prof. S. Srinivasan

Topic: Time-frequency Representations: Analysis, Synthesis and Implementation [[thesis](#)]

Summary:

- Developed Chirplet decomposition, and its estimation using a variant of the EM algorithm
- Proposed time-recursive architectures for efficient implementation of time-frequency distributions
- Taught undergraduate labs in Digital Electronics, VLSI
- Consulted students from various Engineering disciplines in the application of Signal Processing Techniques

## S.R.K.R. Engg College, Andhra University / B.E (ECE)

1993 - 1997, S.R.K.R. Engg. College, Bhimavaram, AP, India

### Summary:

- Participated in National Level Symposiums. Authored technical papers on Autonomous Vehicles, Voice Cloning, in 1996, and 1997, respectively.

## Teaching Experience

---

### Adjunct Faculty, III-Jammu

March' 2025 - , Jammu, J & K, India

Sector Expert- Agri and AI, Central Project Management Unit, IIT Jammu

### Adjunct Faculty, IIIT-B

Aug' 2024 - Dec' 20'24, Bengaluru, KA, India

Teach [ML In Production](#) course for post-graduate students

### Adjunct Faculty, PES University

Aug' 2019 - Dec', 2019, Bengaluru, KA, India

Taught (classical) AI course for undergraduate students

## Industry Experience Wadhvani AI 2020-24

---

### Director, ML- [Wadhvani Institute for Artificial Intelligence](#)

Jan' 2023 - May' 2024 , Bengaluru, KA, India

### Principal ML Scientist- [Wadhvani Institute for Artificial Intelligence](#)

Dec' 2020 - , Dec'2022

### About:

- Wadhvani AI is an institute working on *AI for Social Impact*.

### Responsibilities:

- Mentor young Research Fellows, contribute to overall enterprise AI strategy,
- Work with Federal Government agencies to shape the AI policies, and solutions in the space of public health, agriculture, education, among others.
- Design, Develop, Deploy and Scale AI solutions.
- Develop concept notes/ whitepapers to drive partnerships and collaborations
- Write grant proposal to fund the activities,

### Selected Works:

- Built teams and contributed to the overall enterprise AI strategy.
- Worked with several State and Central Government agencies in shaping AI policies and solutions in Public Health, Agriculture, Education, and other sectors.
- Oversaw the design, development, deployment, and scaling of AI solutions.
- Developed concept notes and whitepapers to drive partnerships and collaborations.
- Wrote grant proposals to fund these activities.
- Helped the creation several thematic ML specialized areas, including the Tabular group, LLM group, and Document Intelligence group, to leverage synergies across solutions.
- Established collaborations with academic and industry partners to further research and adoption, overseeing the development of more than fifteen problems across various domains, and leading initiatives to bring the latest developments in the Generative AI space.
- Represented the Government of India in various national and international forums, such as the G20 and GPAI.
- Managed the entire Agriculture portfolio, expanded the ML team from two to ten members, and secured grants exceeding 4 million USD through funding and partnerships.
- Developed a comprehensive strategy to address the reliability challenges with LLMs and implemented multiple design patterns such as RAG, multi-genetic, and router patterns for developing chatbot applications.
- Expanded the Agriculture portfolio to grow from one problem to many including Pest Detection, Disease Identification in Cotton and Rice, Synthetic Data Generation, Crop Advisories using LLMs, Crop Yield Estimation, Ground Truth Correction using GIS data.
- Built trusted partnerships with industry, academia and other NGOs.

## Industry Experience

**Embibe**  
2019- 20

### Sr Principal Statistical Engineer - [Embibe](#)

Dec' 2019 - Nov'2020 , Bengaluru, KA, India

#### About:

- Embibe is an EdTech company on a mission to improve learning outcomes, across the spectrum.

#### Responsibilities:

- Mentor Data Science team, bring best practices to develop quality data products
- Develop rigorous scientific methods to implement learning products

#### Selected Works:

- Designed, Prototyped, and Managed the implementation of core user-facing APIs that allow users to
  - pick a career path based on their current competencies
  - optimize the learning process to achieve a chosen goal
- Implement scalable Item Response Theory models via Deep Learning frameworks
- Develop Causal Inference frameworks

## Industry Experience

**Consultant**  
2015- 2019

### Consulting ML Architect - [FrameWirk](#), NIIT [StackRoute](#)

2015 -2019 , Bengaluru, KA, India

#### Client:

- [EkStep](#) is a not-for-profit co-founded by Nandan Nilekani, and the core team is constituted by the Aadhar UIDAI team. Its mission is to provide access to learning opportunities to two hundred million children. The platform is adopted by the central Government and more than twenty States are onboarded.

#### Responsibilities:

- Mentor Data Science team
- Provide leadership in AI for Education
- Collaborate in defining and reviewing Platform Architecture and Solutions

#### Selected Works: ([deck](#))

##### Content Recommendation Engines

- Designed and helped team prototype multiple types of Recommendation Engines for Content Consumption based on Factorization Machines, and Item Similarity metrics
- PoC was done in R, and productionized in Spark Scala. Deployed in Production but rolled back as use cases have changed

##### VidyaVani Architecture: Every ML intent is just a Query

- Realized that there is uncertainty at both the source (data models) and sink (use cases) and it was not worthwhile to treat every ML problem as anew
- Proposed an architecture where data models are infinitely additive and extensible, and recommendation Intents are fulfilled via Queries -- making the application evolve with business
- Prototyped the architecture - using Noe4j as the Graph database, Cypher as the querying language, and an orchestration layer written in Python.
- Market-basket Analysis based recommendation algorithm for Content creation was productionized based on the architecture. Rolled back due to changing business dynamics, albeit with little cost to implement

##### Plug-n-Play Analytics Architecture: Enable Conversational Business Intelligence

- Proposed an architecture where database details are abstracted out, and user can focus on asking what is needed, rather than how to get it and from where to get it
- Prototyped the architecture: Apache Spark exposed via REST is used as the on-demand OLAP engine, and Druid, Neo4j, and ElasticSearch are chosen as representative heterogeneous databases. Isomorphism between Graphs and Tables are exploited to enable database hopping
- Developed JSON-based DSL for common ETL operations which can be translated to either SQL or open Cypher

##### An ML-As-a-Service Platform: Provide manufacturer backed guarantees to ML Products

- Designed [daggit](#), an ML-As-a-Service platform, to improve the productivity of a Data Scientist. It glues several frameworks each handling certain responsibility such as:
  - `airflow` for managing and scheduling computational DAGs,
  - `pandas` for data shaping
  - `dvc` for data versioning
  - `git` for versioning models
  - `flask` for exposing functionality via APIs etc.,

along with wrappers for ML libraries. It takes away the mundane parts such as a boiler plate code for ETL, and abstracts out where and how data is stored.

- Version 0.5 is released under MIT license
  - An NLP-based Content Auto Tagging Service was built using the platform
- DevCon'19: EkStep Developer Conference, 2019
- Co-Designed the theme of the Conference.
  - Chalked areas for hacking so that individual solutions fit into providing overall, holistic experience for every stakeholder in the ecosystem
  - Hacked and demo'd several new ideas such as
    - Auto generating questions from a piece of text using NLP
    - A reinforcement learning paradigm to curate and incorporate feedback data
  - Auto question generation was a precursor to one of EkStep's larger vertical solutions: EQB (Energized Question Banks), which CBSE is spearheading

## Open Source Experience

At ML Square  
2018-

---

### mlsquare.org / Founder

2018 -, Bengaluru, KA, India

#### About:

- [mlsquare.org](https://mlsquare.org) is a foundation for a) contributing open source tools to democratize AI and ML, b) drive the development of open source tools and community to create an open AI ecosystem and c) disseminate knowledge about the processes and tooling generated in the due course

#### Realize the vision:

- Make "Explainability" a first class citizen
- Make ML interoperable
- Make ML self-learnable
- Quantify the uncertainty of Deep Learning systems
- A framework for democratizing AI: [report](#)

#### Selected Works:

- Introduced new "layers" such as Decision Trees and LDA into Deep Learning ecosystem
- Mapped and transpiled several APIs in sklearn to their Deep Learning counterparts. An algorithm developed Python can now be scored in a Browser, making ML interoperable
- Designed a framework to turn any black box ML model into an explainable model via Deep Learning
- Mentoring developers in architecture, implementation and product experience design
- Collaborated with Academicians across the globe in developing new approaches in
  - Providing uncertainty quantification metrics in Deep Learning by introducing new losses
  - Likelihood-free inference and Computational sufficient statistics, and others
- [FedEm](#), a proof-of-concept for distributed, decentralized development of LLMs, which one hackathon competition

## Startup Experience

At VitalTicks  
2017 -

---

### VitalTicks Ltd / Founder and CTO

May' 2017 -, Bengaluru, KA, India

#### About:

- [VitalTicks](#)' mission is to make connected health care accessible to anyone, anywhere, and turn existing Dental Clinics into Smart Diagnostic Health Clinics.

#### Responsibilities:

- Architect the platform, Develop Product Roadmap and Project Management

#### Selected Works:

- Studied Product and User Experience design
- Studied DevOps, Deploying embedded applications on the Cloud
- Prototyped the IoT application and integrated it with the Progressive Web App
- Supervised overall product development, and is in piloting stage

**Industry  
Experience**  
At Ilimi  
2014-16

---

**Perceptron Learning Private Limited / Chief Data Scientist**

2014 - 2016, Bengaluru, KA, India

**About:**

- It was a boutique large-scale software architecture consulting company, with rich experience in developing software solutions across FinTech, Education, Oil and Gas Industries. Its learning platform, Ilimi, was seeded to develop EkStep's learning services and the association began in 2015. In 2016, Ilimi was acquired by NIIT, under new Brand *StackRoute Labs*.

**Responsibilities:**

- Design, Develop Learning Analytics Services
- Provide leadership in the Data Science area

**Selected Works:**

- Designed telemetry and data models for Learning Analytics using Ontologies and Dimensional Modeling techniques
- Prototyped a peer grading algorithm
- Developed indices to measure engagement and effectiveness of Learners using Graph and Social Network analysis algorithms
- Coordinated NLP projects to search, curate and tag relevant Content, and align it with taxonomies
- Consulted NIIT StackRoute in designing and running a bootcamp on Applied NLP and ML

**Industry  
Experience**  
At DAS  
2011-14

---

**Dow AgroSciences / Associate Research Scientist**

2011 - 2014, Indianapolis, IN, USA

**About:**

- It was a subsidiary of the Dow Chemical company, with businesses in Ag Chemicals and Seeds

**Responsibilities:**

- Provide Statistical Consulting Services in various domains
- Create, Develop, Apply Statistical methods in Genomics, and Systems Biology

**Selected Works:**

- Designed Quality Control dashboard and Integrated them into user workflows
- Introduced Dynamic DoEs and administered DoE spanning multiple years to creating valuable experimental data
- Established SOPs for cleaning, normalizing high-throughput Genomics data
- Developed and productioned a highly scalable distributed Scientific application for processing Genomics data
- Evaluated and implemented a large scale Gene Network estimation algorithm
- Hired and supervised PhD interns from UC Berkeley and other Universities
- Developed novel methods for constructing high dimensional nonparametric distributions based pair copula constructions

**Research  
Experience**  
At TAMU  
2010-11

---

**Texas A & M University, PostDoc**

2010 - 2011, College Station, TX, USA

**Advisors:**

- Prof. Bani K. Mallick and Prof. Raymond J. Carroll.

**Selected Works:**

- Developed a novel Bayesian method to quantify uncertainty in Dynamical systems
- Advised other graduate students and PostDocs, largely in Bayesian paradigm and efficient inferencing

**Research  
Experience**  
At IIT-M  
2003-04

---

**Center for Scientific and Industrial Research, IIT-M / Project Officer**

2002 - 2003, Chennai, TN, India

About:

- CSIR is the consulting arm of IIT-Madras.

Responsibilities:

- Develop signal processing methods to analyze SONAR data

Selected Works:

- Developed a method to estimate and detect SONAR signals using Chirplet decomposition

## Industry

### Experience

At GE

2000-02

---

#### GE John F. Welch Technology Center / Research Engineer

2000 - 2002, Bengaluru, KA, India

About:

- GE JFWTC is a global research center of GE, a global conglomerate

Selected Works:

- Invented Universal Position Coding method, an adaptive, variable bit-rate quantization method for compressing non-stationary data. It was discovered later that, the optimization criteria used in the algorithm is Genie criteria used Decision Trees, one of the classical Machine Learning techniques
- Optimized Monte Carlo simulation algorithms for simulating PET systems on supercomputers - achieving 3x acceleration in software and 30x on hardware

---

## Journal

### Publications

[16] Learning to detect PII: Tabular vs Document Classification models for network traffic analysis, Rishika Kohli, Shaifu Gupta, Manoj Gaur, Dhavala, SS; Journal of Information Security and Applications, 2025,[accepted], [[link](#)]

[15] Artificial intelligence for advancing eye care in resource-poor settings: Assessing the predictive accuracy of an AI-model for diabetic retinopathy screening in India, Rohan Chawla a, Prachi Karkhanis b, Malay Shah b, Aritra Das b, Rishabh Sharma b, Dhvani Almaula b, Pradeep Venkatesh a, Harsh Vardhan Singh b, Mukul Kumar b, Ramanuj Samanta d, Vinod Kumarl a, Amar Shah c, Bhavin Vadera c, Nakul Jain b, Akanksha Sen b, Shyamsundar Shreedhar b, Vipin Garg b, Soma Dhaval b, Kowshik Ganesh b, Srinivas Rana b, Radhika Tandon a, 2024. [accepted] [[link](#)]

[14] quantile-Long Short Term Memory (qLSTM): A Robust, Time Series Anomaly Detection Method, Jyotirmoy, S; Santanu, S; Dhavala, SS; Saha, S; *IEEE Trans. on AI*, 2024 [accepted]

[13] Estimation and Applications of Quantiles in Deep Binary classification tasks, Tambwekar, A; Maiya, A; Dhavala, SS; Saha, S; *IEEE Trans. on AI*, 2021 [arXiv]

[12] LipGene: Lipschitz continuity guided adaptive learning rates for fast convergence on Microarray Expression Data Sets, Tejas Prashanth, Snehanshu Saha, Sumedh Basarkod, Suraj Aralihalli, Soma S Dhavala, Sriparna Saha, Raviprasad Aduri, *IEEE Trans. on Computational Biology*, 2021

[11] AdaSwarm: Augmenting gradient-based optimizers in Deep Learning with Swarm Intelligence, Mohapatra, R; Saha, S; Coell, C.; Bhattacharya, A; Dhavala, S; Saha, S; , *IEEE Trans. on Emerging Topics in Computational Intelligence*, 2022

[10] A Methodology to Design Heuristics for Model Selection Based on Characteristics of Data: Application to Investigate When the Negative Binomial Lindley (NB-L) is Preferred Over the Negative Binomial (NB), Shirazi, M., S.S. Dhavala, D. Lord, and S.R. Geedipally, *Accident Analysis and Prevention*, Vol. 107, pp. 186-194, 2017. [[link](#)]

[09] A Semiparametric Negative Binomial Generalized Linear Model for Modeling Over Dispersed Count Data with a Heavy Tail: Characteristics and Applications to Crash Data" Mohammad A. Shirazi, Soma S. Dhavala, Sreenivas Geedipally, Lord Dominique, *Accident Analysis and Prevention*, Vol 91, pp. 10-18, 2016. [[link](#)]

- [08] Emulation of Numerical Models with Over-specified Basis Functions, A. Chakraborty, D. Bingham, S. S. Dhavala, C. C. Kuranz, R. P. Drake, M. J. Grosskopf, E. Rutter, B. R. Torralva, J. P. Holloway, R. G. McClarren and B. K. Mallick, *Technometrics*, Vol. 59, No. 2, pp. 153-163, 2016. [[link](#)]
- [07] A caution about using the Deviance Information Criterion while modeling traffic crashes, Sreenivas Geedipally, Lord Dominique, Soma S. Dhavala, *Safety Science*, Vol. 62, pp. 495–498 2014. [[link](#)]
- [06] Bayesian object classification of gold nanoparticles, Bledar A. Konomi, Soma S. Dhavala, Jianhua Huang, Subrata Kundu, David Huitink, Hong Liang, Yu Ding, Bani K. Mallick, *Annals of Applied Statistics*, April, 7(2) pp. 640-668, 2013. [[link](#)].
- [05] Evaluation of Corn Hybrids Expressing Cry1F, Cry1A.105, Cry2Ab2, Cry34Ab1/Cry35Ab1, and Cry3Bb1 Against Southern United States Insect Pests, Melissa W. Siebert, Steve Nolting, Soma S. Dhavala, William Henrix, Craig Chism, Scott D. Stewart, John N All, Fred Musser, David Buntin, Samuel Luke, *Journal of Economic Entomology*, Vol. 32, No.5, pp. 1825-35, 2012. [[link](#)]
- [04] The Negative-Binomial-Generalized-Lindley Model: Characteristics and Application using Crash Data, Sreenivas Geedipally, Lord Dominique, Soma S. Dhavala, *Accident Analysis and Prevention*, Vol. 45, No.2, pp. 258-265, 2011. [[link](#)]
- [03] Characterizing the Performance of the Conway-Maxwell Poisson Generalized Linear Model, Royce A. Francis, Srinivas Reddy Geediapply, Seth D. Guikema, Soma S. Dhavala, Dominique Lord, Sara LaRocca, *Risk Analysis*, Vol. 32, No.1, pp. 167-183, 2011. [[link](#)]
- [02] The Salmonella enterica serotype Typhi Vi capsular antigen is expressed after entering the ileal mucosa, Quynh T. Tran, Gabriel Gomez, Sangeeta Khare, Sara D. Lawhon, Manuela Raffatellu, Andreas J. Baumler, Dharani Ajithdoss, Soma S. Dhavala, and L. Garry Adams, *Infection & Immunity*, Vol. 78, No.1, pp. 527-535, 2010. [[link](#)]
- [01] Bayesian Modeling of MPSS data: Gene Expression Analysis of Bovine Salmonella infection, Soma S. Dhavala, Bani K. Mallick, Sangeeta Khare, Sara Lawhon, Raymond J. Carroll and, L. G. Adams, *Journal of the American Statistical Association*, Vol. 105, No. 491, pp. 956-967, 2010. [[link](#)]

## Book Chapters/

### Invited Discussions

- 
- [06] Deploying COVID-19 Case Forecasting Models in the Developing World, Sansiddh Jain, Avtansh Tiwari, Nayana Bannur, Ayush Deva, Siddhant Shingi, Vishwa Shah, Mihir Kulkarni, Namrata Deka, Keshav, Ramaswami, Vasudha Khare, Harsh Maheshwari, Soma Dhavala, Jithin, Sreedharan, Jerome White, Srujana Merugu, and Alpan Raval, *AI for Social Good, Part 2* [[link](#)]
- [05] Probabilistic Programming with Pyro, Prakash Bisht., Dhavala S.S. (work-in-progress) [[link](#)]
- [04] A New Approach for Momentum Particle Swarm Optimization, Mohapatra R., Talesara R.R., Govil S., Saha S., Dhavala S.S., Sudarshan T. In: Patnaik S., Yang X.S., Sethi I. (eds) *Advances in Machine Learning and Computational Intelligence. Algorithms for Intelligent Systems*. Springer, Singapore, 2021
- [03] AI for Education, *UNESCO Blue Dot Magazine*, pp. 47-49, 2019. [[link](#)]
- [02] Discussion of “Bayesian models for sparse regression analysis of high dimensional data by Sylvia Richardson, Leonordo Bottolo, Jeffrey Rosenthal”, Soma Dhavala, Rajesh Talluri, Bani K. Mallick, Faming Liang, Mingqi Wu, *Bayesian Statistics*, Vol 9, Proceedings of Ninth Valencia International Conference on Bayesian Statistics, Oxford University Press, pp. 539-568, 2011. ISBN 978-0-19-969458-7. [[link](#)]
- [01] Bayesian Model-based Principal Component Analysis, Bani K. Mallick, Shubhankar

Ray, and Soma Dhavala, *Frontiers of Statistical Decision Making and Bayesian Analysis*, in Honor of James. O. Berger, eds. Ming-Hui Chen, Dipak K. Dey, Peter Mueller, Dongchu Sun, and Keying Ye, Springer, New York, 2010, ISBN 978-1-4419-6943-9. [\[link\]](#)

## Consulting

### Acknowledged

---

[09] Maintaining user trust through uncertainty aware inference, Chandan Agarwal, Ashish Papanai, Jerome White, [\[link\]](#), workshop on [Deployable AI](#), AAAI, 2024.

[08] Book: Highway Safety Analytics and Modeling, D. Lord, X. Qin, and S. Geedipally, Elsevier, 2021. [\[link\]](#)

[07] Development of a Random Parameters Negative Binomial-Lindley Generalized Linear Model to analyze Over-Dispersed Data, Shaon, M.R.R., X. Qin, A. Shirazi, D. Lord, and S. Geedipally, *Analytic Methods in Accident Research*, Vol. 18, pp. 33-44, 2018. [\[link\]](#)

[06] Characteristics Based Heuristics to Select a Logical Distribution between the Poisson Gamma and the Poisson Lognormal for Crash Data Modeling, Shirazi, M., and D. Lord, 97th Annual Meeting of the TRB. Aug 2, 2017. Submission under revision. [\[pdf\]](#)

[05] Exploring the Application of the Negative Binomial-Generalized Exponential Model for Analyzing Traffic Crash Data with Excess Zeros, Prathyusha Vangala, Dominique Lord, Srinivas Reddy Geedipally, *Analytic Methods in Accident Research*, Vol 7, pp. 29-36, 2015. [\[link\]](#)

[04] Morphologic and Cytokine Profile Characterization of Salmonella enterica Serovar Typhimurium Infection in Calves With Bovine Leukocyte Adhesion Deficiency (J. S. Nunes, S. D. Lawhon, C. A. Rossetti, S. Khare, J. F. Figueiredo, T. Gull, R. C. Burghardt, A. J. Baumler, R. M. Tsois, H. L. Andrews-Polymeris, and L. G. Adams, *Vet. Pathol.*, 47(2), pp. 322-33, 2010. [\[link\]](#)

[03] The negative binomial Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros, Dominique Lord, Srinivas Reddy Geedipally, *Accident Analysis and Prevention*, 43, 1738-42, 2011. [\[link\]](#)

[02] Assessing circuit breaker performance using condition-based data and Bayesian approach, Satish Natti, Mladen Kezunovic, *Electric Power Systems Research*, 81, 1796-1804, 2011. [\[link\]](#)

[01] Estimation of two parameter multilevel item response models with predictors - simulations and substantiation for an Urban school district, Prathiba Natesan, Ph. D Thesis, TAMU, ,2007. [\[link\]](#)

## Conference

### Proceedings

---

[17] Lattice-Based Vector Quantization for Low-Bit Quantization-Aware Training Rishika Kohli, Soma S Dhavala, Shaifu Gupta, Manoj S. Gaur, Proceedings of the Machine Learning Research, Conference on Parsimony and Learning, 2026, [\[accepted\]](#)[\[peer reviewed\]](#)

[17] Health Sentinel: An AI Pipeline For Real-time Disease Outbreak Detection, Devesh Pant, Rishi Raj Grandhe, Jatin Agrawal, Jushaan Singh Kalra, Sudhir Kumar, Saransh Khanna, Vipin Samaria, Mukul Paul, Satish V Khalikar, Vipin Garg, Himanshu Chauhan, Pranay Verma, Akhil Vssg, Neha Khandelwal, Soma S Dhavala, Minesh Mathew, Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI), 2025, [\[accepted\]](#)[\[peer reviewed\]](#)[\[link\]](#)

- [16] QuantProb: Generalizing Probabilities along with Predictions for a Pre-trained Classifier , Aditya Challa, Soma S. Dhavla, Snehanshu Saha, 40th Conference on Uncertainty in Artificial Intelligence (UAI 2024), 2024. [accepted][peer reviewed]
- [15] Automatic Interpretation of Line Probe Assay Test for TB , Agrawal, J., Kumar, M., Tiwari, A., Danisetty, S., Dhavala, S., Jain, N., Balraj, P., Singh, N., Shingi, S., Kurada, J., Rao, R., Anand, S., & Kumar, N. (2024), AAAI, 2024. [accepted][peer reviewed]
- [14] HMC-PSO: A Hamiltonian Monte Carlo and Particle Swarm Optimization-based optimizer, Omatharv Bharat Vaidya, Rithvik Terence DSouza, Soma Dhavala, Snehanshu Saha, Swagatam Das ,The 29th International Conference on Neural Information Processing, New Delhi , ICONIP, 2022. [accepted][peer reviewed]
- [13] A Deep Learning Framework for Machine Learning Interoperability, Shakkeel Ahmed, Prakash Bisht, Ravi S Mula, Soma S Dhavala, 1st International Conference on AI-ML Systems, Bangalore, India, ACM, 2021. [accepted][peer reviewed][[paper](#)]
- [12] LALR: Theoretical and Experimental validation of Lipschitz Adaptive Learning Rate in Regression and Neural Networks, Snehanshu Saha, Tejas Prashanth, Suraj Aralihalli, Sumedh Basarkod, T.S.B Sudarshan and Soma S Dhavala, International Joint Conference on Neural Networks, Glasgow, UK,, 2020. [accepted][peer reviewed]
- [11] Auto-generation of diagnostic assessments and their quality evaluation, Soma S Dhavala, Chirag Bhatia, Joy Bose, Keyur Faldu, Aditi Avasthi, Educational Data Mining Conference, Virtual Conference, 2020. [accepted][peer reviewed]
- [10] MOOC2.0: A Case for India, Raviteja Garlapati, Soma S Dhavala, Nagaraju Pappu, Knowledge Discovery and Data Mining Conference, New York, NY, 2014. [[pdf](#)][peer reviewed]
- [09] A comparison between the Geometric Mean Estimator and the bias-corrected Geometric Mean Estimator for Lognormal Median, Zgenglei Gao, Iain Proctor, Soma Dhavala, Klauss Hammel, Pesticide Behaviour in Soils, Water and Air , University of York, York, September 2013. [[pdf](#)]
- [08] A Bayesian Semiparametric Hierarchical Model for Analyzing Differential Expression in Sequence-Based Gene Expression Data, presentation at the Joint Statistical Meeting, Washington DC, Aug 1-6, 2009.
- [07] A sequential Bayesian approach to distributed estimation in wireless sensor networks, poster presentation at the Joint Statistical Meeting, Denver, CO, Aug 3-7, 2008. [[pdf](#)]
- [06] Characterizing the Performance of a Bayesian Conway-Maxwell Poisson GLM , paper presented at the 2008 Joint Statistical Meetings, Denver, Co, August 3-7, 2008, with Geedipally S., S.D. Guikema and D. Lord.
- [05] Gene Expression Analysis of Bovine Salmonella infection, poster presented at ABCS Workshop on Bioinformatics, Computational Biology and Systems Biology, TAMU, College Station, Feb 22-23, 2008, with Sujay Datta, Bani K Mallick, Sangeeta Khare, Sara Lawhon, Raymond J Carroll, L G Adams
- [04] An empirical Bayesian kernel density estimator, poster presented at the AMSTAT Conference of Texas Statisticians, March 31- April 1, 2006, Austin, Texas. [[pdf](#)]
- [03] Importance sampling in PET collimators, IEEE Nuclear Science Symposium Conference Record, 2003, Volume: 4 , 19-25 Oct. 2003, Page(s): 2559 -2561, with

Harrison, R.L, Kumar, P.N, Yiping Shao, Lewellen, T.K. [[link](#)][peer reviewed]

[02] Acceleration of SimSET photon history generation, IEEE Nuclear Science Symposium Conference Record, 2002 , Volume: 3 , 10-16 Nov. 2002, Page(s): 1835 -1838, with Harrison, R.L, Kumar, P.N, Yiping Shao, Manjeshwar, R, Lewellen, T.K, Jansen, F.P. [[link](#)][peer reviewed]

[01] On the Use of Chirp Transform, International Conference on Robotics, Vision and Parallel Processing for automation, University Sains, Malaysia, July 14-16, 1999, Vol. 1, pp . 218-225, with K. M. M. Prabhu

## Patents

---

[08] System and method to generate sets of similar assessment papers, K Faldu, S Dhavala, A Avasthi, *filed*, 2021. [USA]

[07] System and method for generating diagnostic assessment question papers and evaluating their quality S Dhavala, K Faldu, A Avasthi, *filed*, 2021. [USA]

[06] Distributed system and method to monitor health parameters of a patient, Soma S Dhavala, Mehaboob Basha, *filed*, 2018. [India]

[05] System and Method to remotely operate a Dental Chair, Soma S Dhavala, Mehaboob Basha, *filed*, 2018. [India]

[04] Method and system for image compression and decompression using span of interest of an image sequence, Sudipta Mukhopadhyaya, Soma S Dhavala, E M Lalitha, *filed*, 2002. [[USA](#), [Japan](#)]

[03] System System and Method for coding Data, Soma S Dhavala and Fredirick Wheeler, *granted*, 2006. [[USA](#), [Japan](#)]

[02] Position Coding System and Method, Soma S Dhavala, Fredirick Wheeler, Sudipta Mukhopadhyaya, Nithin Nagaraj, *filed*, 2002. [[USA](#)]

[01] Progressive transmission and reception of image data using modified Hadamard transform, John E. Hershey and Soma S. Dhavala, *filed*, 2002. [[USA](#), [Japan](#)]

## Invited Talks

---

[09] Role of ML in Cybersecurity, First GE HealthCare Symposium on Cyber Security, GE HealthCare, Bengaluru, 2019.

[08] Building Enterprise-grade Data Products, GE John F Welch Technology Center, Bengaluru, 2018.

[07] Network Analysis: Lending insights into the connected world, IISc, Bengaluru, 2014. [[link](#)]

[06] Semiparametric inference with discrete expression data and FDR control, International Chinese Statistical Association's 2010 Applied Statistics symposium, Indianapolis, IN, June 20-23, 2010.

[05] A Bayesian Semiparametric Model to Analyze Discrete Gene-expression, MD Anderson Cancer Center, 2008

[04] Bayesian Adaptive Regression Splines (BARS), guest lecture, Advanced Bayesian Bioinformatics class, Fall 2008. [[pdf](#)]

[03] Introduction to programming in SAS, presented to INFORMS members, TAMU, 2006. [[pdf](#)]

[02] Distributed Source Localization in wireless sensor networks, presented to graduate students at National Institute of Advanced Studies, IISc campus, India, 2005. [[pdf](#)]

[01] An Introduction to Statistics, presented to employees of Veloz Systems, Bangalore, India, 2005. [[pdf](#)]

## Selected

---

[13] Let's dope it - ML interoperability via Deep Learning: [deck](#)

## Blogs/Videos

- [12] Learning as a location-based service: [blog](#)
- [11] VitalTicks: Connected Health Solutions: [deck](#)
- [10] Developing a Workbench: [deck](#)
- [09] SQUAD question - answering system: [video](#)
- [08] Word Embeddings: [video](#)
- [07] Text Similarity: [video](#)
- [06] An Analysis of Gated Recurrent Units: [video](#)
- [05] Optimization in DNNs: [video](#)
- [04] Deep Learning as a Machine Learning Infrastructure Framework: [part-1](#), [part-2](#)
- [03] A [baby] step towards Data Science Automation: [here](#)
- [02] 15 Commandments for Building Data Products: [blog part-1](#), [part-2](#)
- [01] DataFrame is the new JSON: [blog](#)

## Selected

### Reports

- 
- [17] A framework for predicting, interpreting, and improving Learning Outcomes, Donda, C; Dasgupta, S; Dhavala, SS; Faldu, K; Avasthi, A; 2020, [\[arXiv\]](#)
  - [16] A similarity metric for multivariate ordinal features: Soma S. Dhavala, 2016, [\[link\]](#)
  - [15] Machine Learning At EkStep, SS Dhavala, Feroz Sheik, Nagaraju Pappu, 2015: [\[link\]](#)
  - [14] A Dynamic Programming Approach to False Discovery Rate Control, Soma S Dhavala, Bank K. Mallick, 2010: [\[pdf\]](#)
  - [13] Towards multivariate p-values and FDR control, Soma S Dhavala, 2009: [\[pdf\]](#)
  - [12] Joint estimation in State-Space Models, Soma S Dhavala, 2008: [\[pdf\]](#)
  - [11] Slice Sampler, Soma S Dhavala, 2008: [\[pdf\]](#)
  - [10] An empirical Bayesian Kernel Density Estimator, SS Dhavala, S. Ghosh, 2005: [\[pdf\]](#)
  - [09] Algorithms for distributed source locations in wireless sensor networks, SS Dhavala, 2005: [\[pdf\]](#)
  - [08] A Bayesian approach to relative node location in wireless sensor networks, Soma S Dhavala, David Faden, Matthew A. Halverson, 2004: [\[pdf\]](#)
  - [07] Wavelet Packets and Adaptive Filterbanks, Soma S Dhavala, 2004: [\[pdf\]](#)
  - [06] Generalized Digital Modulation Techniques with emphasis on Instantaneous Frequency modulation, Soma S Dhavala, 2004: [\[pdf\]](#)
  - [05] Numerical Methods for Monte Carlo Integration, Soma S Dhavala, 2004: [\[pdf\]](#)
  - [04] Laplace Approximation, Soma S Dhavala, 2004: [\[pdf\]](#)
  - [03] Gaussian Mixture Modeling for Multi-Modal Posterior, Soma S Dhavala, 2004: [\[pdf\]](#)
  - [02] EM Algorithm for Mixture Modeling, Soma S Dhavala, 2003: [\[pdf\]](#)
  - [01] Chirplet Decomposition, Soma S Dhavala, 2003: [\[pdf\]](#)

## Awards

- 
- [07] Increase The Slope Award Finalist, Dow AgroSciences, 2014
  - [06] ACTG Award, Dow AgroSciences, 2014
  - [05] JSM Travel Award, 2008,2009
  - [04] General Electric Intellectual Property Award, 2002
  - [03] General Electric Global Research Center Management Award, 2001
  - [02] Six Sigma Green Belt Certification, 2001
  - [01] Andhra Pradesh State Merit Scholarship, 1991-97

## Professional

### Service

- 
- [04] Executive Committee Member: IEEE Computer Society, Bangalore Chapter, 2019-
  - [03] Reviewer: Bayesian Analysis, JASA, Integrative Biology, 2010-14
  - [02] Treasurer: ASA South East Texas Chapter, 2008-10
  - [01] Judge: Student Research Week, TAMU, 2008

## Community

### Service

- 
- [08] Regular Teacher - Purna Pramati, Bengaluru, 2018 -  
Have been teaching Maths to High School students since 2018.  
Designed a teaching method around NCERT curriculum to foster mathematical

- thinking, and prepare them for taking a career in Maths
- [07] Treasurer, Carnatic Music Association of Indianapolis, Indianapolis, 2013-14  
Coordinated fundraising events and music event management, ran newsletter campaigns, managed budget and balance sheets
  - [06] Student Advisor, CRY, TAMU chapter, 2007-08  
Mentored student chapter executive body in organizing fundraising, event management
  - [05] President, CRY, TAMU chapter, 2006-07  
Organized fundraising events, awareness campaigns
  - [04] Volunteer, SPICMACAY, TAMU chapter, 2005-09  
Organized fundraising events, managed Artists hospitality, ran video campaigns
  - [03] Volunteer, AID, TAMU chapter, 2006-07  
Helped in AID fundraising events, created awareness videos
  - [02] Volunteer, ISA, TAMU chapter, 2006-07  
Facilitated on-boarding and assimilation of new Indian graduate students
  - [01] Volunteer, SANKALP, ISU, 2003-05  
Participated in fundraising drives

## Hobbies

### Selected Works

- 
- [06] IoT
    - PiGanesha: A Raspberry Pi based Chariot controlled via REST APIs. [\[youtube\]](#)
    - VentroBot: A Raspberry Pi based Ventriloquist. [\[youtube\]](#)
    - PiGenie: A Raspberry Pi based augmented learning agent. As a child plays game, feedback manifests over the bot. [\[youtube\]](#)
    - HoverCroft: A Raspberry Pi based hover croft controlled via REST APIs. [\[youtube\]](#)
  - [05] Audio Mixing
    - Jana Gana Mana remix of A. R. Rehman's vocal and instrumental versions. Released on the occasion of India's 60th Independence Day. [\[youtube\]](#)
  - [04] Video Mixing
    - Shake a leg or two [\[youtube\]](#)
    - Shades of Bhangra [\[youtube\]](#)
  - [03] Script, Direction, Editing and Camera work
    - Desi Got Talent [\[youtube\]](#)
    - Proportion of Graduate Life [\[youtube\]](#)
    - Shiva Born Again - short movie [\[part-1, part-2\]](#)
    - The Signs of the Car-centric [\[youtube\]](#)
    - News for the Deaf and Dumb [\[youtube\]](#)
  - [02] Theatre and Skits
    - The Great Celestial Cow [in the [news](#)]
    - DFSS: Design for Securing Sita [\[youtube\]](#)
  - [01] Poetry (in Telugu)
    - embedding Chess moves into a poetic meter [\[knight, bishop\]](#)
    - 101 of composing poetry with constraints [\[blog\]](#)
    - [others](#)

References available on Request

Updated March, 2025