Smita Mohanty

Director DSAI, AI Innovation and Execution
Novartis Healthcare Pvt. Ltd.
Hyderabad, Telengana

+91 709 308 7638
smita.o.mohanty@gmail.com

Profile summary

10+ years providing analytics and R&D in pre-clinical research (drug repurposing, pharmacogenetics), pharma commercial analytics (portfolio optimization, patient health risk modeling, personalized recommendation engines) and software product development (Bodylogical, a mathematical simulation based product for modeling patient health).

Pharma Commercial Analytics

- ➔ NLP, Knowledge graphs and AI-based recommender engines for Healthcare professionals, scaled to billions of data points on cloud infrastructure
- ➔ Statistical models for predicting disease progression and prognostic markers for multiple therapeutic areas based on longitudinal patient data and claims data

Mathematical Modeling

- ➔ Simulations using ODEs to predict early onset of diseases from claims data
- ➔ Generating automated personalized lifestyle management recommendations based on patient wellness data
- ➔ Cloud based software development to scale to millions of users

Pre-clinical and Life Science Research

- ➔ Pharmacogenetics analysis for predicting impact of SNPs on the mechanism of action of drug and efficacy for an antidepressant drug and gastrointestinal disease drug.
- ➔ Developed software platform for drug repurposing, deployed for multiple clients (press releases with Takeda and Astellas).
  - ◆ Graph theory and NLP based algorithms for predicting new disease indications from drug-target-disease networks.
- ➔ Developed ML algorithm for improving protein-protein interaction detection accuracy that shows good correlation with experimental data.
- ➔ Developed computational methods for understanding protein dynamics and allostery further used in predicting impact of drug resistance mutations in oncology (published two papers)

Education

Indian Institute of Science / PhD in Computational Biology

July 2006 - October 2012, Bangalore, India

- ➔ Thesis title: Functionally interacting proteins: analyses and prediction
  - ◆ Development of approaches towards improving remote homology detection based methods for function annotation of proteins, this led to enhanced coverage of function prediction for proteins from genomes with high sequence divergence.
  - ◆ Leveraging large scale high throughput experimental data along with bioinformatics algorithms to predict functionally interacting proteins. Predicted associations showed high correlation with experimentally known interactions and further led to identification of novel function annotations for proteins.

Indian Institute of Science / Masters in Life Sciences

July 2003 - July 2006, Bangalore, India

Delhi University / Bachelor's in Biochemistry

July 2000- July 2003, Delhi, India

Experience

Director, Data Science AI Innovation and Execution, Novartis

March 2020 - Present

- ➔ Led the building of an enterprise wide knowledge graph by linking vast arrays of disparate data sources.
  - ◆ Instrumental in creating the next generation platform for engaging health care professionals. Platform leverages connected data from the knowledge graph, along with deep learning techniques like Graph convolutional networks to learn the graph representation and further use this in generating personalized recommendation.
  - ◆ Using knowledge graphs to build an Amazon like marketplace within Novartis by connecting products and their properties with user buying behavior. Built a product recommendation engine based on buying preferences and product attributes, useful towards nudging users to products with greater value proposition.
  - ◆ Deployed a scalable underlying graph architecture combining both OLAP and OLTP features to enhance the performance of the graph database with increasing amounts of data.
- ➔ Using deep learning to orchestrate multi-channel marketing for oncology products
  - ◆ Using recurrent neural networks like LSTMs to learn the underlying impact of multiple channels on the overall sales of drugs.
  - ◆ Combining the above mentioned channel attribution model with evolutionary algorithms to obtain optimal promotion recommendations
- ➔ Developed statistical predictive model as a virtual endpoint for clinical trial
  - ◆ Using longitudinal patient data to build statistical models for early diagnosis and potentially help in reducing the duration of clinical trials.
- ➔ Leading the development of a data driven platform leveraging advance data science approaches for evaluating pharma drug portfolio.
  - ◆ Developed a simulation platform to evaluate the entire multi-dimensional drug portfolio by integrating financial evaluation with advanced analytics and visualizations.
  - ◆ Time series analysis of the portfolio, evaluate impact of timeline perturbations.
  - ◆ Multi-objective optimization of conflicting trade-offs to select optimal projects to achieve business goals using several optimization algorithms such as genetic algorithms

Senior Manager, PricewaterhouseCoopers

February 2018- January 2020

- ➔ Managed a team of 6 data scientists and engineers to deliver solutions using mathematical modeling for multiple clients including Pharma and Payers.
  - ◆ Utilize mathematical models depicting human metabolic processes to evaluate health outcomes and impact of lifestyle on disease progression and drug effectiveness.
- ➔ Established simulation based algorithms to transform healthcare data ('big data') to high accuracy mechanistic predictive models.
- ➔ Integrated simulation modeling with real world data obtained from medical claims and electronic health records for US insurance clients.
- ➔ Developed a statistical model for identifying COPD burden using disease comorbidity network from medical claims data.

GVKBio Pvt Ltd/ Scientist, Manager of Algorithm Development Team

February 2015 - September 2016

- ➔ Applied chem-informatics and network biology to predict drug off-targets using adverse events data.
- ➔ Developed algorithm that shows enrichment of approved drug combinations with high confidence (50% of approved combinations show up in the top 10% of data (manuscript under preparation).
- ➔ Managed a team of 4 scientists and delivered 5 algorithms for drug repurposing.
- ➔ Developed Big Data visualization/analysis engine for integrated analysis of 'omics' data.
- ➔ Generation of *in silico* platform for pharmacogenomics analysis of SNPs, with special emphasis on non-coding SNPs. Developed approach led to accurate annotation of function of 30 non-coding SNPs in two different therapeutic areas.
- ➔ Developed a drug and disease phenotype map for aiding new target identification and drug repurposing

Department of Computational Medicine and Bioinformatics, University of Michigan / Postdoc

December 2016- January 2018

- ➔ Re-designed protein-protein interaction interfaces using computational biophysical methods,
- ➔ Specifically applied on oncogenic protein Ras and its interacting partners.

Experimental validation using high-throughput assays to measure stability and binding affinities..

Department of Biochemistry, University of Georgia / Postdoc

November 2012- December 2014

- ➔ Performed computational and experimental analysis of structural mechanisms of allostery in protein kinases.
- ➔ Found a unique tyrosine kinase sequence signature that allosterically controls the dynamics of ATP binding pocket and catalytic efficiency.
- ➔ Performed MD simulation analysis to predict three residues affecting ATP site dynamics and three rescue mutations.
- ➔ Using experimental thermal stability and enzyme kinetics validated the effect of MD-predicted changes in protein dynamics on enzyme regulation and function, especially drug and ATP binding kinetics.
- ➔ Identified regulatory regions in Cyclic nucleotide binding proteins of eukaryotic pathogens using Bayesian evolutionary analysis.

Skills

Computational

- Graph databases and knowledge graphs, Graph based deep learning
- Network algorithms
- Time series modeling using deep learning
- Statistical modeling using machine learning and deep learning, especially with patient longitudinal data.
- Expert in Perl and Python programming.
- Proficiency in statistical analysis using Scipy, Matlab and R
- Well versed with databases and tools in computational biology (high dimensional data)
- Expertise in BigQuery, SQL and SPARK databases
- Monte carlo simulation and other simulation techniques in biophysics
- Generating data visualization using d3 and other dashboarding tools

Experimental

- Standard molecular biology work, PCR, Gene construct generation, site-directed mutagenesis etc.

- Heterologous protein expression and purification.
- Enzyme kinetics using fluorescence and coupled enzyme assays.
- HIgh throughput thermal stability assays for protein stability measurements and ligand binding assays.

Publications

1. **Mohanty S,** Oruganty K, Annie K, Byrne DP, Ferries S, Zheng R, Hanold LE, Katiyar S, Kennedy EJ, Eyers PA, Kannan N (2016) Hydrophobic core variations provide a structural framework for tyrosine kinase evolution and functional specialization. PLoS Genet. 12(2), e1005885.
2. **Mohanty S**, Kennedy EJ, Herberg FW, Hui R, Taylor SS, Langsley G, Kannan N. (2015) Structural and evolutionary divergence of cyclic nucleotide binding domains in eukaryotic pathogens: Implications for drug design. Biochim Biophys Acta. S1570-9639(15).
3. **Mohanty S**, Purwar M, Srinivasan N, & Rekha N. (2013) Tethering preferences of domain families co-occurring in multi-domain proteins. Mol Biosyst. 9(7), 1708-25.
4. Amarnath S, Kawli T, **Mohanty S**, Srinivasan N & Nanjundiah V. (2012) Pleiotropic Roles of a Ribosomal Protein in Dictyostelium discoideum. PLoS ONE 7(2), e30644.
5. Srinivasan N, Agarwal G, Bhaskara RM, Gadkari R, Krishnadev O, Lakshmi B, Mahajan S, **Mohanty S**, Mudgal R, Rakshambikai R, Sankaran S, Sudha G, Swapna LS& Tyagi N (2011) Influence of genomic and other biological data sets in the understanding of protein structures, functions and interactions. Int.J. Knowledge Discovery Bioinf., 2, 24-44.
6. Nidhi T, Swapna LS, **Mohanty S**, Agarwal G, Gowri VS, Anamika K, Priya ML, Krishnadev O & Srinivasan N (2009) Evolutionary divergence of Plasmodium falciparum: Sequences, protein-protein interactions, pathways and processes. Infectious Disorders-Drug Targets, 9, 257-271.
7. **Mohanty S** & Srinivasan, N (2009) How effective Is The Data On Co-Occurrence Of Domains In Multi-domain Proteins In Prediction Of Protein-Protein Interactions? 7th IEEE International Workshop on Genomic Signal Processing and Statistics, GENSIPS 2009, pp 1-4
8. **Mohanty S** & Srinivasan N (2009) Identification of missing metabolic proteins of Plasmodium falciparum: A bioinformatics approach. Prot. Pept. Lett. 16, 961-968.
9. Krishnadev O, Rekha N, Pandit SB, Abhiman S, **Mohanty S**, Swapna LS, Gore S & Srinivasan N (2005). PRODOC resource for comparison of tethered protein domain architectures with in-built information on distantly related domain families. Nucl. Acids Res. 33, W126-129

Book chapters

1. **Mohanty S**, Ramakrishnan G, Dave P & Srinivasan N. (2014) Analysis of Sequence Divergence in Metabolic Proteins of Plasmodium falciparum: Implications for Remote Homology Detection. Accepted as a chapter for Frontiers in Protein and Peptide Science, Vol 1, Bentham eBooks series.
2. **Mohanty S**, Pandit SB & Srinivasan N (2009) Dynamics of protein-protein interaction network in Plasmodium falciparum in Biological Data Mining in Protein Interaction Networks (Editors: Xiao-Li Li & See-Kiong Ng). IGI Global Press. pp. 257-285.

Conferences

1. **Mohanty S**., Chelini C., Alessadro P., Dwivedi G. (2019) Apparent Plateau in Diabetes Diagnosis Explained by Misuse of Updated Diagnosis Criteria in Simulation Modeling Study of Macronutrient Consumption Trends in U.S. Diabetes 2019 Jun; 68(Supplement 1)
2. Pattnaik A.,Kanodia S., Chowdhary R. and **Mohanty S**. (2019) Predicting Tuberculosis Related Lung Deformities from CT Scan Images Using 3D CNN. CEUR Workshop Proceedings.