Kopal Garg

kopal.garg@mail.utoronto.ca

github.io/kopalgarg

in/gargkopal/

Education

- **University of Toronto** — Toronto, ON
  Master of Science in Computer Science and Machine Learning (cGPA: 4.0) — *Sept. 2021 – Present*
- **University of Waterloo** — Waterloo, ON
  Bachelor of Biomedical Engineering with Honors (Co-op) (GPA: 3.9) — *Sept. 2016 – Apr. 2021*

Technical Skills

- **Languages/Programs**: Python, R, C++, MySQL scripting/querying, Unix, Bash scripting, LaTeX, Git, AWS, GCP, MATLAB/Octave, JavaScript, Terra, WDL/Cromwell, Jupyter Lab, Jupyter Lite, PyScript
- **Packages/APIs**: PyTorch, TensorFlow, Keras, tslearn, statsmodel, Scikit-Learn, OpenCV, spaCy, NLTK, NumPy, BeautifulSoup, Scrapy, Matplotlib, Seaborn, Pandas, dplyr, shiny, tidyverse, ggplot2, plotly
- **Relevant Coursework**: Machine Learning, Deep Learning, Data Mining and Warehousing, Algorithms and Optimization, Probability Theory and Statistics, Cloud Computing, Natural Language Computing, Mobile Computing
- **ML Skills**: Building statistical and algorithmic models with complex and large datasets, Data modeling and evaluation, Signal processing, NLP, Supervised statistical learning, Clustering, Computer vision, Time-series forecasting, Predictive modeling, Data visualization, Performing complex data analysis on large volumes of data
- **Certifications**: Neural Networks and Deep Learning (DeepLearning.AI), IBM Machine Learning Professional Certificate*, IBM Data Engineering Professional Certificate*

Professional Experience

Vector Institute for Artificial Intelligence — May 2021 - Present
Machine Learning Researcher — *Toronto, ON*

- **Non-parametric Change Point Detection**: Led the development of a statistical CPD method for multivariate time series data. Employed a non-parametric two-sample hypothesis test that maximizes MMD test power by aggregating multiple tests over a range of kernel bandwidths.
- **Graphical Lasso for Tracking Feature Interactions**: Demonstrated a novel use of Time-Varying Graphical Lasso for quantifying changes in feature interactions, presenting its capability in detecting change points that occur due to changes in the covariance of the joint distribution of features over time. Used pair-wise changes in feature interactions to explain and further sub-categorize change points. Conducted extensive benchmark evaluation to show the method's ability in outperforming existing state-of-the-art CPD methods in characterizing various CP types.
- **Model-Agnostic Explainability**: Contributed to a model-agnostic explainability method that assigns importance to observations in multivariate time-series based on their counterfactual influence on future predictions.
- **Multi-variate Time-series Forecasting**: Developing and tuning algorithms (in PyTorch, Keras, Tensorflow) for extracting insight from multi-variate sensor data and forecasting disease outcomes.
- ***Skills***: DL and ML Python Packages, R, Bash, Jupyter Notebooks, Multivariate Time-Series, Predictive Modeling

Broad Institute of MIT and Harvard — Jan. 2020 - May, 2021
Data Science Intern — *Cambridge, MA*

- **Exploratory and Statistical Analysis**: Led the exploratory and statistical analysis of large high-throughput and complex multi-omics datasets to test new hypothesis about complex diseases.
- **Cloud-based GWAS Framework**: Built a cloud-based pipeline to facilitate a scalable genome-wide association study (GWAS) workflow through parallelization and effective distribution of computational resources. Built an R package with workflows for common GWAS visualizations like regional association plots, linkage disequilibrium patterns, population stratification and admixture plots.
- **Web-applications for genomic analysis**: Developed an array of web-interfaces to facilitate analysis of single cell RNA sequencing data, and for visualization of differential gene expression in erythroid cells during hematopoiesis.
- **Single-cell Transcriptomics**: Built a command line tool in Python for generating heteroplasmic estimations from single-cell mitochondrial DNA data. Devised a data-driven technique of optimizing inputs to the tool, created and maintained a Terra pipeline and Docker image for the workflow.

* Ongoing

- **SRA Downloads, Pseudoalignment and Feature Quantification Cloud Pipeline**: Built a cloud-based pipeline for automation of SRA downloads through NCBI toolkits, pseudoalignment, and feature quantification using custom Docker images of various bionformatics tools and packages.
- ***Skills***: Python, Jupyter Notebooks, R, Bash, Terra, WDL, Cromwell, Docker, Bioinformatic CL Tools, Databases and annotation tools (GenBank, RefSeq, ENSEMBL, dbSNP, UCSC genome browser, BLAST, NCBI, OMIM)

Broad Institute of MIT and Harvard Apr. 2019 - Sept, 2019
Data Science Intern *Cambridge, MA*

- **Exome Sequencing Pipeline**: Built a pipeline for exome sequencing to identify high impact and rare genomic variants classified by Mendelian patterns of inheritance. Used JavaScript to define and facilitate operations on arbitrary family structures (trios, quartets, singletons, etc.), as inferred from pedigree files.
- **Interactive Web Application**: Developed a customizable web application for rare variant discovery using VEP annotations, in silico tools for novel variant interpretation (SIFT, PolyPhen and SpliceAI), as well as known genes and phenotype keywords from ClinVar.
- **QC, Phasing and Imputation Cloud Pipeline**: Wrote a pipeline for quality control, phasing and imputation of whole-genome sequencing data.
- ***Skills***: Python, R, Bash, Terra, WDL, Cromwell, Docker, Bioinformatic CL Tools, JavaScript, Jupyter Notebooks

University of Calgary, McCaig Institute Sept. 2018 - Dec. 2018
Data Science, Co-op *Calgary, AB*

- **Imaging Software Tool**: Developed an imaging software to assess anterior chest wall deformities before and after non-operative chest-wall correction.
- **Iterative Close Point Registration**: Implemented registration methods to refine alignment in 3D surface topography triangular mesh data to detect anatomical landmark locations, obtain cross-sectional depth profiles and calculate external chest wall indices of a patient's torso.
- **Assessment of Alignment**: Quantified differences in patients' torsos at subsequent time-points (weeks or months into thoracic bracing) by refining alignment using point-set registration methods.
- **Visualization**: Devised methods of quantifying treatment efficacy by using chest contour mapping to show elevation or depression of the central chest and adjacent areas.
- ***Skills***: MATLAB, Python, R, Bash, Jupyter Notebooks

University of Toronto, SickKids Research Institute Jan. 2018 - Apr. 2018
Data Science Intern *Toronto, ON*

- **Methylation Data Analysis**: Wrote scripts to analyze differential methylation data for identification of genomic alterations that lead to neurodevelopmental and growth disorders. These included quality control assessments, within-array normalizations, gender prediction, and bump hunting for Differentially Methylated Regions.
- **Data Visualizations**: Wrote scripts to visualize and further explore sample quality through density bean plots that show Beta value densities of samples, control probe plots that help assess sample preparation steps, and gender prediction plots that cluster samples based on predicted gender.
- **Interactive Web Application**: Built a screening tool and an underlying computational framework that uses ML models trained on significant CpG sites from a discovery cohort to correlate epigenome-based signatures with variant pathogenicity.
- ***Skills***: Python, R, Bash, Jupyter Notebooks

Journal Articles and Conference Publications

- **(In Submission) KDD, 2022**: **Garg, K.**, Tonekaboni S., Goldenberg, A. *Categorizing Change Points with Time Varying Networks.*
- **(In Submission) Nature, 2022**: Caleb A. Lareau, Sonia M. Dubois, **Kopal Garg**, Pauline Kautz, Lena Nitsch, Samantha Praktiknjo, Jeffrey M. Verboon, Wendy Lu, Eleni Mimitou, Christoph Muus, Rhea Malhotra, Patrick Maschmeyer, Sumit Parikh, Mark D. Fleming, Anshul Kundaje, Peter Smibert, Ansuman T. Satpathy, Aviv Regev, Vijay G. Sankaran, Suneet Agarwal, Leif S. Ludwig *Cellular states, purifying selection, and clonal dynamics in Pearson syndrome revealed via single-cell multi-omics.*
- **(In Submission) The 1st Single Cell Genomics Symposium, 2022**: Caleb A. Lareau, Sonia M. Dubois, **Kopal Garg**, Pauline Kautz, Lena Nitsch, Samantha Praktiknjo, Jeffrey M. Verboon, Wendy Lu, Eleni Mimitou, Christoph Muus, Rhea Malhotra, Patrick Maschmeyer, Sumit Parikh, Mark D. Fleming, Anshul Kundaje, Peter Smibert, Ansuman T. Satpathy, Aviv Regev, Vijay G. Sankaran, Suneet Agarwal, Leif S. Ludwig *Cellular states, purifying selection, and clonal dynamics in Pearson syndrome revealed via single-cell multi-omics.*

- **Journal of Experimental Medicine, 2021**: Verboon, J., Wahlster, L., Ludwig, L., Black, S., Luo, W., **Garg, K.**, Collins, R., Garimella, K., Costello, M., Chao, K., Ditroia, S., O'Donnell-Luria, A., Talkowski, M., Michelson, A., Cantor, A., Sankaran, V. *Familial thrombocytopenia due to a complex structural variant resulting in a WAC-ANKRD26 fusion transcript.*

- **The American Journal of Human Genetics, 2020**: Choufani, S., Gibson, W., Turinsky, A., Chung, B., Wang, T., **Garg, K.**, Vitriolo, A., Cohen, A., Cyrus, S., Goodman, S., Chater-Diehl, E., Brzezinski, J., Brudno, M., Ming, L., White, S., Lynch, S., Clericuzio, C., Temple, I., Flinter, F., McConnell, V., Cushing, T., Bird, L., Splitt, M., Kerr, B., Scherer, S., Machado, J., Imagawa, E., Okamoto, N., Matsumoto, N., Testa, G., Iascone, M., Tenconi, R., Caluseriu, O., Mendoza-Londono, R., Chitayat, D., Cytrynbaum, C., Tatton-Brown, K. and Weksberg, R. *DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes.*

- **International Society of Biomechanics, 2019**: Bugajski, T., Vafadar, B., Gray, E., **Garg, K.**, Schneider, M., Nettel-Aguirre, A., Brindle, M., Lam, J., Lopushinsky, S. and Ronsky, J.,2019. Reliability of a *Three-Dimensional Scanning Technique and Metrics Quantifying Pectus Deformities.*

- **American Society of Human Genetics – Reviewers' Choice Abstract, 2019**: S. Choufani, **K. Garg**, C. Cytrynbaum, S. Cyrus, B. H.Y. Chung, W.T. Gibson, R. Weksberg *Novel first tier DNA methylation-based diagnostic platform for neurodevelopmental disorders.*

- **The 5th Canadian Conference on Epigenetics, 2019**: S. Choufani, **K. Garg**, C. Cytrynbaum, S. Cyrus, B. H.Y. Chung, W.T. Gibson, R. Weksberg *Novel first tier DNA methylation-based diagnostic platform for neurodevelopmental disorders.*

Scholarships and Awards

- **Vector Institute for AI**: Research Grant
- **University of Toronto**: Graduate Scholarship
- **University of Waterloo, Department of Engineering**: Graduation with Distinction
- **University of Waterloo, Department of Engineering**: Dean's Honors List
- **University of Waterloo, Department of Engineering**: President's Scholarship of Distinction
- **University of Waterloo, Department of Engineering**: Engineering Entrance Scholarship
- **Alberta Education**: Alexander Rutherford Scholarship