# MOHIT PANDEY

1777 Kingsway, Unit 201, Vancouver, BC V5N 2S5 | +1-848-237-9939 | pandey.mohitk@gmail.com | Website | Google Scholar

## SUMMARY


- Interdisciplinary researcher with skills and experience in machine learning, natural language processing, and applied deep learning in life sciences and healthcare.
- Worked on and led collaborative projects, resulting in peer-reviewed publications in high impact journals.
- Abreast of state of the art Natural Language Processing algorithms and tools and experienced in applied machine learning.
- Self-motivated, problem-solving and collaborative worker with excellent communication skills.
- Looking to research further on multi-modal learning at the intersection of protein crystallographic structures and ligand sequences to enhance generative drug discovery.


## EDUCATION

- Ph.D., Bioinformatics (Machine Learning)
  University of British Columbia, 2020-
  Advisor: Dr. Artem Cherkasov

- M.S., Computer Science
  Rutgers University, 2017 3.65/4.0
  *Courses: Algorithms • Computer Animations and Game Design • Artificial Intelligence •Computer Vision • Machine Learning • Pattern Recognition • Probability Theory • Mathematical theories for Optimization • Text Mining in Big Data*

- B. Eng., Electronics and Telecommunication
  Shri G.S. Institute of Tech and Science, 2015 7.26/10.0 (absolute)
  *Thesis: Cost effective solutions for Human Computer Interfacing*

## EXPERIENCE

**Vancouver Prostate Centre** **Vancouver**
*Graduate Researcher* *Jan 2021 to -*
- Led collaborative survey study on advancements in GPU computing for drug discovery. The work is conditionally accepted for publication in Nature Machine Intelligence.
- Mentored a team of developers to build and release deep learning based web platform for democratizing docking. This solution was used to identify potential inhibitors of SARS-CoV-2 Main Protease by docking 1.3 Billion compounds.

**Signzy Technologies** **Bengaluru**
*Senior AI Research Engineer* *Nov 2020 to June 2021*
- Created an industry standard solution for detecting passive liveliness from single frame webcam feed using state of the art deep learning techniques.
- Interviewed and managed hiring process for senior level engineering positions within the AI team.

**Ipsos Public Affairs** **New York**
*Senior Machine Learning Scientist* *July 2019 to June 2020*
- Created and streamlined deep learning based NLP solutions for Aspect based Sentiment Analysis towards top political entities spanning multiple countries for deeper understanding of electoral standings through social media data.
- Developed English and Arabic emotion analysis model for text in large scale social media stream using transfer learning.
- Productionized several real-time machine learning solutions through cloud based deployment.
- Worked on semantic segmentation to distinguish buildings footprint in geospatial imagery.

**Weill Cornell Medicine** **New York**
*Machine Learning Scientist and Cognitive Developer* *Feb 2017 to June 2019*
- Designed deep learning based NLP solution for information retrieval from unstructured Electronic Health Records to create large scale datasets for effectively carrying out outcomes research. A large patient pool hence created showed excellent MACE forecast.
- Worked on context aware neural data augmentation policy for low resource task in specialized domains like medical sciences.
- Co-Authored several machine learning papers for prognostic assessment and prediction of congestive heart failure in top-tier journals and conferences.
- Deployed and studied various state of the art deep learning models and techniques for challenging problem of LV Segmentation across multi-modality radiological imaging.

**Cariclub Inc.** **New York**
*Software Developer, AI* *June 2016 to June 2019*
- Designed matching algorithm to suggest most suitable match of non-profits for the users on CariClub platform using supervised machine learning. Also made location based, interest based recommendations using both content and collaborative filtering.
- Deployed distributed collabrative filtering using Alternating Least Sqaures using Apache Spark on AWS.

## TEACHING AND MENTORING EXPERIENCE

- 2018 - Mentored 6 high-school interns in Cardiovascular anatomical annotations and basics of medical image segmentation.
- 2016 - Graduate Grading Assistant for Computer Animations and Game Design

## RESEARCH PROJECTS

**Computer Vision and Deep learning pipeline for Self Driving Car**

- Used Convolutional Neural Network to detect and identify traffic signals
- Using state of the art computer vision models like LeNet and AlexNet, trained models to drive a car on simulator. The data was collected using human behavior cloning.
- Found lane lines to constrain car's driving using deep learning and traditional computer vision algorithms like Hough Transform.

**Predicting a Terrorist Attack**

- Implemented and compared various classification algorithms on Global Terrorism Dataset to predict the group responsible for a terrorist attack. Achieved accuracy of over 70% when using Random Forests for classification. K-Nearest Neighbor, Naive Bayes were other classifiers used.
- Performed linear and polynomial (2nd order) regression to predict number of casualties and wounded people in a given terrorist attack.

**Probabilistic Topic Modelling**

- Used Latent Dirchlet Allocation and Hierarchical latent Dirichlet allocation to cluster news articles in 20 NewsGroup Dataset.
- Did topic modeling to study and understand underlying pattern in essays of students applying to Rutgers Grad School.

**Effect of Word Embeddings on Sentiment Analysis using Spark**

- Used Apache Spark to pre-process Amazon book reviews and calculate word frequency correlation to labels.
- Presented a survey on various word embeddings on classical machine learning models as well as deep learning based neural network architecture.

**Web Scale Recommendation System**

- Implemented Distributed Stochastic Gradient Descent in Apache Spark for low rank decomposition of similarity matrix for collaborative filtering on Netflix dataset.

**Low Cost Gesture Based Mouse**

- Developed Java Desktop Application to access mouse based on user's gesture. The Application does not expect the user to wear additional hardware in order to detect fingers. The application detects 9 static gestures including opened and closed palm, claw, OK and O signs, Gun Sign, Pinch, Pointing and Finger Tap and supports common mouse actions like Pinch to Zoom, Scroll, Swipe, Wave, Finger push to click.

## POSTERS AND PUBLICATIONS


[1] M. Pandey, M. Fernandez, F. Gentile, S. D. Huver, V. Mysore, A. C. Stern, and A. Cherkasov, "Citius altius fortius: The transforming role of gpu computing in drug discovery," *In Submission*.

[2] M. Pandey, Z. Xu, E. Sholle, G. Maliakal, G. Singh, Z. Fatima, D. Larine, B. C. Lee, J. Wang, A. R. van Rosendael, *et al.*, "Extraction of radiographic findings from unstructured thoracoabdominal computed tomography reports using convolutional neural network based natural language processing," *PloS one*, vol. 15, no. 7, p. e0236827, 2020.

[3] M. Pandey and S. Haymann, "Tracking immigration discussion in social media: A survey on deep learning based natural language processing for social media insights," *American Association for Public Opinion Research, 2020*, Accepted. Could not present due to COVID-19.

[4] Y. E. Yoon, L. Baskaran, B. C. Lee, M. K. Pandey, B. Goebel, *et al.*, "Differential progression of coronary atherosclerosis according to plaque composition: Cluster analysis of the paradigm registry," *Journal of the American College of Cardiology*, vol. 77, no. 18_Supplement_1, pp. 1274–1274, 2021.

[5] S. J. Al'Aref, K. Anchouche, G. Singh, P. J. Slomka, K. K. Kolli, A. Kumar, M. Pandey, G. Maliakal, A. R. Van Rosendael, A. N. Beecy, *et al.*, "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," *European heart journal*, vol. 40, no. 24, pp. 1975–1986, 2019.

[6] G. Singh, S. Alaref, G. Maliakal, M. Pandey, A. van Rosendael, B. Lee, J. Wang, Z. Xu, and J. Min, "Deep learning based automatic segmentation of cardiac computed tomography," *Journal of the American College of Cardiology*, vol. 73, no. 9 Supplement 1, p. 1643, 2019.

[7] G. Singh, S. J. Al'Aref, M. Van Assen, T. S. Kim, A. van Rosendael, K. K. Kolli, A. Dwivedi, G. Maliakal, M. Pandey, J. Wang, *et al.*, "Machine learning in cardiac ct: basic concepts and contemporary data," *Journal of Cardiovascular Computed Tomography*, vol. 12, no. 3, pp. 192–201, 2018.

[8] A. R. van Rosendael, G. Maliakal, K. K. Kolli, A. Beecy, S. J. Al'Aref, A. Dwivedi, G. Singh, M. Panday, A. Kumar, X. Ma, *et al.*, "Maximization of the usage of coronary cta derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the confirm registry," *Journal of cardiovascular computed tomography*, vol. 12, no. 3, pp. 204–209, 2018.

[9] S. J. Al'Aref, G. Maliakal, G. Singh, A. R. van Rosendael, X. Ma, Z. Xu, O. A. H. Alawamlh, B. Lee, M. Pandey, S. Achenbach, *et al.*, "Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the confirm registry," *European Heart Journal*, vol. 41, no. 3, pp. 359–367, 2020.

[10] S. J. Al'Aref, G. Singh, A. R. van Rosendael, K. K. Kolli, X. Ma, G. Maliakal, M. Pandey, B. C. Lee, J. Wang, Z. Xu, *et al.*, "Determinants of in-hospital mortality after percutaneous coronary intervention: a machine learning approach," *Journal of the American Heart Association*, vol. 8, no. 5, p. e011160, 2019.

[11] S. Al'Aref, G. Maliakal, M. Cheng, K. Kolli, G. Singh, M. Pandey, A. Kumar, A. Dwivedi, D. Andreini, F. Cademartiri, *et al.*, "A novel ensemble machine learning-based method versus clinical risk scoring for discrimination of individuals who will versus will not experience acute coronary syndrome after coronary computed tomographic angiography: Results from the iconic study," *Journal of the American College of Cardiology*, vol. 71, no. 11 Supplement, p. A1628, 2018.

[12] G. Singh, G. Maliakal, S. Al'Aref, A. Dwivedi, M. Pandey, A. Kumar, M. Gummalla, P. Dunham, M. Gomez, H.-J. Chang, *et al.*, "Automated diagnosis of echocardiographic views using deep learning: P2-39," *Journal of the American Society of Echocardiography*, vol. 31, no. 6, 2018.

[13] S. Al'Aref, A. van Rosendael, G. Maliakal, G. Singh, X. Ma, M. Pandey, J. Wang, V. Do, T. S. Kim, M. Gummalla, *et al.*, "Tct-55 clinical predictors of obstructive coronary artery disease in individuals with suspected coronary artery disease," *Journal of the American College of Cardiology*, vol. 72, no. 13 Supplement, pp. B24–B25, 2018.

[14] L. Baskaran, G. Maliakal, S. J. Al'Aref, G. Singh, Z. Xu, K. Michalak, K. Dolan, U. Gianni, A. van Rosendael, I. van den Hoogen, M. Pandey, *et al.*, "Identification and quantification of cardiovascular structures from ccta: an end-to-end, rapid, pixel-wise, deep-learning method," *JACC: Cardiovascular Imaging*, vol. 13, no. 5, pp. 1163–1171, 2020.

## TECHNICAL SKILLS

- **Programming Languages:** Python, Matlab, MySQL, LaTeX, Java, Shell, JavaScript (familiar), R (beginner)
- **Software:** Photoshop, Illustrator, OsiriX
- **Operating System:** MacOS, Windows, Linux

## SCHOLARSHIPS AND AWARDS

- President's Academic Excellence Initiative PhD Award, University of British Columbia, 2021
- International Tuition Award, University of British Columbia, 2021
- Computer Science MS Program Award, Indiana University, 2015 *[declined]*
- Tuition Assistance Award for Technical Studies, HEG India, 2011-2015
- Voted among best 5 grad level algorithm class project for "Feature representation of Amazon Book Recommendation System", USA 2015
- Winner Vijana Bharti State level technical poster presentation beating over 60 colleges, India 2012
- All India Rank 1887 in National Aptitude Test by NIIT, India, 2013 (Top 1 percentile.)