

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 drug molecules to enhance drug discovery.

EDUCATION

- Ph.D., Bioinformatics (Machine Learning)
University of British Columbia, Canada, 2021-
Advisor: Dr. Artem Cherkasov
Prospective Thesis: Accelerating Drug Discovery through Structural Analysis of Proteins using Geometric Representation Learning
- M.S., Computer Science
Rutgers University, U.S.A, 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, India, 2015 7.26/10.0 (absolute)
Thesis: Cost effective solutions for Human Computer Interfacing

EXPERIENCE

Vancouver Prostate Centre **Vancouver**
Graduate Research Assistant *Jan 2021 to -*

- Led collaborative survey study on advancements in GPU computing for drug discovery.
- 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.

Ipsos Public Affairs **New York**
Senior Machine Learning Scientist *July 2019 to June 2020, Sep 2021 to - Dec 2022*

- 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.

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.

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 collaborative filtering using Alternating Least Squares using Apache Spark on AWS.

TEACHING AND MENTORING EXPERIENCE

- 2022* - Teaching Assistant for courses Supervised Learning II; Statistical Inference and Computation II; Advanced Machine Learning (NLP); Algorithms and Data Structures; and Python programming for Data Science at University of British Columbia's Masters in Data Science program.
- 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

Binding Affinity prediction from AlphaFold generated protein structures

- Performed regression on negative log of K_d for protein-ligand interaction pairs using Siamese-like network.
- Created protein structural graphs to utilize graph neural networks for succinct representation.
- Visualized active regions of protein, learned through graph attention network.
- The method achieved state-of-the-art performance across several standard benchmarking datasets (PDBBind, BindingDB, DAVIS, KIBA)

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 Dirichlet 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. Radaeva, H. Morin, M. **Pandey**, F. Ban, M. Guo, E. LeBlanc, N. Lallous, and A. Cherkasov, "Novel inhibitors of androgen receptor's dna binding domain identified using an ultra-large virtual screening," *Molecular Informatics*.
- [2] O. Garland, M. Radaeva, M. **Pandey**, A. Cherkasov, and N. Lallous, "Purificationdb: database of purification conditions for proteins," *Database*, vol. 2023, p. baad016, 2023.
- [3] O. Garland, A.-T. Ton, S. Moradi, J. R. Smith, S. Kovacic, K. Ng, M. **Pandey**, F. Ban, J. Lee, M. Vuckovic, *et al.*, "Large-scale virtual screening for the discovery of sars-cov-2 papain-like protease (plpro) non-covalent inhibitors," *Journal of Chemical Information and Modeling*, 2023.
- [4] M. Radaeva, M. **Pandey**, H. MsLati, and A. Cherkasov, "Deep learning model for prediction of compound activities over a panel of major toxicity-related proteins," in *Machine Learning and Deep Learning in Computational Toxicology*, pp. 579–593, Springer, 2023.
- [5] M. **Pandey**, M. Radaeva, H. Mslati, O. Garland, M. Fernandez, M. Ester, and A. Cherkasov, "Ligand binding prediction using protein structure graphs and residual graph attention networks," *Molecules*, vol. 27, no. 16, 2022.
- [6] M. **Pandey**, M. Fernandez, F. Gentile, O. Isayev, A. Tropsha, A. C. Stern, and A. Cherkasov, "The transformational role of gpu computing and deep learning in drug discovery," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 211–221, 2022.

- [7] A.-T. Ton, M. **Pandey**, J. R. Smith, F. Ban, M. Fernandez, and A. Cherkasov, "Targeting sars-cov-2 papain-like protease in the post-vaccine era," *Trends in Pharmacological Sciences*, 2022.
- [8] 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.
- [9] 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.
- [10] Y. E. Yoon, L. Baskaran, B. C. Lee, M. K. **Pandey**, B. Goebel, S.-E. Lee, J. M. Sung, D. Andreini, M. H. Al-Mallah, M. J. Budoff, *et al.*, "Differential progression of coronary atherosclerosis according to plaque composition: a cluster analysis of paradigm registry data," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] A. R. van Rosendael, G. Maliakal, K. K. Kolli, A. Beecy, S. J. Al'Aref, A. Dwivedi, G. Singh, M. **Pandey**, 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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)
- **Operating System:** MacOS, Windows, Linux

SCHOLARSHIPS AND AWARDS

- Canadian Institutes of Health Research (CIHR) Doctoral award, Govt. of Canada, \$105000, 2023-2025
- President's Academic Excellence Initiative PhD Award, University of British Columbia, 2021-2022
- Faculty of Science Graduate Award, University of British Columbia, 2021-2022
- International Tuition Award, University of British Columbia, 2021-2022
- Computer Science MS Program Award, Indiana University, 2015 [*declined*]
- Tuition Assistance Award for Technical Studies, HEG India, 2011-2015
- Winning solution for in-conference code accelerator challenge at NVIDIA's GPU hackathon 2020.

- 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.)
- Secured All India Rank 9411 in IIT-JEE 2011 among 530,000 students.

VOLUNTEER SERVICES

- 2023 - Reviewer Academic Radiology
- 2023 - Reviewer Springer Vegetos
- 2022 - Reviewer IEEE/ACM Transactions on Computational Biology and Bioinformatics
- 2022 - Reviewer Journal of Cheminformatics
- 2021-2022 - Senior Editor, Health Science Enquiry Journal, Canada.
- 2021-2022 - Adjudicator, Multidisciplinary Undergraduate Research Conference, University of British Columbia.
- 2016 - Student Volunteer, International Joint Conference of Artificial Intelligence, New York, USA.