Achint Soni

📱 +1 5198976489   •   ✉ achint.s046@gmail.com   •   🌐 trickyjustice.github.io

Education

Research Masters (MMath Thesis), University of Waterloo — **Ontario, Canada**
Major: Computer Science — *Sep'23 - Jun'25 (Expected)*

B.Tech, Indian Institute of Technology Kanpur — **Kanpur, India**
Major: Electrical Engineering | Minor: Machine Learning, Theory of Computing | GPA: 3.57/4 — *Jul'19 - May'23*

Research Interests

Probabilistic ML (Generative models); Deep representation learning; Foundational models (text to multimedia generation); Time series forecasting

Publications

Energy efficient time table scheduling for metro systems using machine learning
Paper under review at EJIS

Research Experiences

Mobility data to address public health decision making post COVID-19
Mentor: Sahar Saeed, Queen's University, Canada — *Jun'22 – Dec'22*

- Evaluated the viability of monitoring mobility data to specific points of interest in Canada in order to continue informing post-COVID-19 pandemic public health decision making.
- Accessed the representativeness of SafeGraph mobility data in Canada by comparing the number of sampled devices with Census population counts at different geographic levels.
- Described the temporal patterns of visits to healthcare institutions at several epochs and contrasted these patterns across provinces and Material Deprivation Indices by utilising a quasi-Poisson hierarchical generalised additive model.
- Estimated the utilization patterns of each facility during the pandemic using geographic catchment analytical methods.
- The ratio of expected and observed healthcare utilization over time for each census block group was used to construct an inequity map.

Fast and accurate bayesian polygenic risk modelling using variational inference
McGill University, Canada — *May'22 – Jul'22*

- Implemented fast and efficient Bayesian polygenic risk score method that approximates posteriors for the effect sizes of genetic variants on the phenotype using variational inference techniques.
- Conducted comprehensive set of experiments using simulated and real traits to assess the predictive ability of the model in comparision with some of the most popular Bayesian and non-bayesian methods.
- Observed fast model convergence enabled by variational inference algorithm in contrast to stochastic MCMC approaches implemented by other methods.

Carbon footprint reduction using plastic bag price optimization
WRIHW Lab, Durham University — *Jan'22 – Apr'22*

- Formulated a multi-objective optimization model to determine the Pareto-optimal pricing strategy for plastic bags, effectively balancing supermarket profitability and the quantifiable carbon footprint associated with plastic production and consumption.
- Engineered an optimal reward policy, effectively curbing the carbon footprint of plastic bags while ensuring stable sales trajectories.

Forecasting COVID-19 Vaccination demand using Twitter
LMS, University of Hull — *Sep'21 – Jan'22*

- Proposed a novel method to predict the demand information for COVID-19 vaccinations in a tweet by utilizing semi-supervised learning with recurrent neural networks (RNN).
- Designed a mathematical model that calculates the likelihood of demand information of a tweet at a particular time.
- A new data dictionary is derived with words that show the feeling of support and request, which increases the robustness of the model to temporal losses and long term dependencies.
- The result is compared with other neural network based control to demonstrate the superiority of the proposed technique.

Student Research Associate
Department of Biotechnology, Government of India *Sep'21 – Nov'21*
- Deployed a classification model involving deep learning that can detect motor imagery brain signals using an open-source brain-computing interface electroencephalogram (OpenBCI-EEG) headset.
- Implemented intelligent EEG artefacts removal such as eye and tongue movements, electrode and sweat artefacts by Autoregressive-Deep Variational Autoencoder model.

Undergraduate Research Associate- SURGE'21, IIT Kanpur
Intelligent Systems and Controls Lab, IIT Kanpur *Jun'21 – Sep'21*
- Built a lightweight keyword spotting system that can recognize multiple different short speech commands.
- Calculated Mel-Frequency Cepstral Coefficients (MFCC) to extract spectral features and gain more perceptually-relevant representation of speech audio.
- Deployed a classification model involving deep CNN that can detect speech commands based on their mel spectrograms.
- Assembled the model on unmanned mine safety inspection vehicle and tested on English and Hindi language speakers.

Projects

Document Image Classification with Intra-Domain Transfer Learning
Mentor: Prof. Tushar Sandhan, IIT Kanpur *Jan'23 – Mar'23*
- Implemented a region-based DCNN framework for document structure learning on RVL-CDIP dataset.
- Separated the image into four different regions and applied transfer learning using Vgg16 architecture trained on ImageNet.
- Combined the predictions from individual base deep neural network using a stacking generalisation based ensembling.
- Achieved a maximum mean F1-Score of 79.5% on test and 92% on train dataset and secured an All India Rank 21.

Acoustic Event Detection
Mentor: Prof. Vipul Arora, IIT Kanpur *Apr'22 – May'22*
- Identified onset, offset times, and labels of acoustic events in audio clips using mel-spectrogram features at millisecond precision.
- Leveraged a diverse array of advanced models, including CNNs, RNNs, hybrid CRNNs, LSTM, and HMMs, attaining a noteworthy 0.91 F1 score for precise acoustic event detection.
- Applied rigorous statistical techniques, including Gibbs sampling, Importance/Rejection sampling, Monte-Carlo sampling, and Normalizing Flows, enhancing the robustness of the analysis.
- Constructed Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) from the ground up, employing neural networks to model Gaussian Mixture Models for a comprehensive exploration of acoustic event distributions.

Twitter Sentiment Analysis
Mentor: Prof. Faiz Hamid, IIT Kanpur *Aug'21 – Oct'21*
- Designed and implemented an information retrieval and classification system for sentiment analysis on Twitter.
- Cleaned, parsed and segmented tweets content; counted most frequent words, ngrams and hashtags.
- Used TF-IDF and GloVe pretrained Word Embeddings to obtain vector representations for words.
- Modeled Support Vector Machine and Naive Bayes algorithm to determine sentiment polarity of data set.
- Implemented a Bidirectional LSTM model using Tensorflow to classify the tweets into appropriate categories of sentiment.

Deep learning approaches for COVID-19 detection based on chest X-Ray images
Self project *Jul'20 – Sep'20*
- Designed a deep learning system to extract features and detect COVID-19 from chest X-ray images.
- Automatized the process of analyzing X-ray images with high accuracy using deep Convolotional Neural Networks (CNNs).
- Three powerful networks, namely ResNet50, InceptionV3, and VGG16, were fine-tuned on an enhanced dataset.
- Implemented transfer learning to train a ResNet50 model on the same data to achieve an accuracy of 96.32% on train data and 94.53% on validation data

Technical Skills

Programming Skills: C/C++, Python, LaTeX

Libraries: PyTorch, Tensorflow, HuggingFace, Keras, OpenCV, Sklearn

Tools & OS: Git, VS Code, MacOS, Linux(Ubuntu)

Achievements and Accolades

- Full academic ride by the Cheriton School of Computer Science, University of Waterloo
- Received MITACS Graduate Fellowship for aiding my graduate research
- Academic Excellence Award for exceptional performance in Mathematics and Science