Tarun Gupta

B.Tech - Computer Science and Engineering
Indian Institute Of Technology Indore

tarungupta360@gmail.com
tarun360.github.io
linkedin.com/in/tarun-gupta-116860178

EDUCATION

Degree/Certificate	Institute/Board	CGPA/Aggregate	Year
B.Tech.	Indian Institute of Technology Indore	9.81 / 10	2018-2022
Senior Secondary	CBSE Board	94.2%	2018
Secondary	CBSE Board	10 / 10	2016

PUBLICATIONS

Rajaa, S., Anandan, K., Dalmia, S., **Gupta, T**. and Chng, E.S., 2023, June. Improving Spoken Language Identification with Map-Mix. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

Gupta, T., Truong, T.D., Anh, T.T., Chng, E.S. (2022) Estimation of speaker age and height from speech signal using bi-encoder transformer mixture model. Proc. Interspeech 2022, 1978-1982, doi: 10.21437/Interspeech.2022-567

Gupta, T., He, X., Uddin, M.R., Zeng, X., Zhou, A., Zhang, J., Freyberg, Z. and Xu, M., 2022. Self-supervised learning for macromolecular structure classification based on cryo-electron tomograms. Frontiers in Physiology, p.1757.

Tanveer, M., **Gupta, T.**, Shah, M. and Richhariya, B., 2021. Sparse twin support vector clustering using pinball loss. IEEE Journal of Biomedical and Health Informatics, 25(10), pp.3776-3783.

Tanveer, M., **Gupta, T.**, Shah, M. and Alzheimer's Disease Neuroimaging Initiative, 2021. Pinball loss twin support vector clustering. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(2s), pp.1-23.

RESEARCH EXPERIENCE

Research Intern at NTU Singapore   *Aug 2021 - Dec 2021*
Multilingual speech processing, Self-supervised learning, Transformers   Paper 1, Paper 2
- Built bi-encoder transformer mixture model for estimation of speaker age and height from speech signal, leading to relative improvement of 18.5% in male age estimation and 8.6% in female age estimation over current state-of-the-art on TIMIT dataset (refer Paper 1).
- Developed a new data augmentation method, Map-Mix, that leverages model training dynamics of individual data points to improve sampling for latent mixup. Map-Mix improves weighted F1 scores by 2% compared to the random mixup baseline and results in a significantly well-calibrated model (refer Paper 2).

Research Intern at Carnegie Mellon University   *Jan 2021 - May 2021*
Computer vision, Self-supervised learning, Computational biology   Paper
- Built contrastive self-supervised learning workflow for macromolecular structure classification based on Electron cryotomography (CryoET) data.
- Used techniques such as SwAV, MoCo and SimCLR to achieve state-of-the-art classification results for CryoET data.

Research Assistant at IIT Indore   *Jan 2020 - Mar 2020*
SVM, Clustering, Convex Programming.   Paper 1, Paper 2
- Designed a novel plane-based clustering algorithm based upon principles of SVM. Utilised pinball loss to improve generalization performance on noise-corrupted datasets. Used Concave-Convex Procedure (CCCP) for optimization.
- The proposed algorithm improved clustering accuracy by 1% on noise corrupted UCI datasets as compared to existing plane-based clustering algorithms.

WORK EXPERIENCE

Member Technical, D. E. Shaw India Pvt. Ltd   *July 2022 - Present*
Software development, Systems engineering, High-performance computing, Containerization, Python
- Developing high-performance cluster (HPC) scheduling software for internal cluster and public cloud platforms.
- Implementing ML workload pipeline with focus on security-first principles for public cloud.

Software Development Intern, D. E. Shaw India Pvt. Ltd.   *May 2021 - July 2021*
Software development, React, Kafka, Java
- Designed and implemented a powerful graph visualization tool using the GoJS web framework to enhance the workflow efficiency for traders. The tool provides an intuitive and interactive representation of the complex workflows followed by traders.
- Utilized Protocol Buffers to create contracts and Apache Kafka for inter-communication between microservices. Implemented multiple unit and integration tests using Mockito testing framework.

Projects

- **Parallelising Red Deer Algorithm (RDA) – A Nature Inspired Meta-heuristic Algorithm** *April 2021*
  Parallel Programming, MPI [Github](#)
  - Implemented Red Deer Algorithm (RDA) for solving Travelling Salesman Problem (TSP). RDA is a meta-heuristic algorithm inspired by the unique mating process of Scottish red deer.
  - Used Message Passing Interface (MPI) for parallelising the algorithm, achieving speed-up factor upto 4.

- **Automated Headline and Sentiment Generator** *March 2021*
  Silver medal at Inter IIT Tech Meet 9.0. [Github](#)
  - Employed DistillBERT transformer model for binary classification of tweets and articles and obtained 0.89 F1 score.
  - Used Aspect Based Sentiment Analysis driven Ada-BERT model to identify brands and their sentiments.
  - Generated headlines for articles using Pegasus, T5, BART models achieving 37% average similarity score

- **Adversarial Attack on Brain Tumour Segmentation** *April 2020-June 2020*
  AI for Medicine, Deep Learning. [GitHub](#)
  - Implemented 3D-UNet architecture using TensorFlow library and trained it on BraTS brain MRI data.
  - Successfully implemented Fast-Gradient Sign Method (FGSM), Iterative Fast-Gradient Sign Method (iFGSM), and Carlini & Wagner (CW) adversarial attacks from scratch using TensorFlow library.

- **Sentiment Analysis of Movie Reviews in Hindi Language** *May 2020 - July 2020*
  NLP, Flask framework, Web Scraping. [GitHub](#)
  - Created largest Hindi movie review dataset, containing 1714 movie reviews using web-scrapping techniques.
  - Used Universal Language Fine-tuning (ULMFiT) transfer learning method to achieve test-accuracy of 75.70%.
  - Created back-end APIs using FLASK framework.

- **Cache Oblivious Algorithms** *May 2020 - July 2020*
  Design and Analysis of Algorithms. [GitHub](#)
  - Implemented various cache oblivious algorithms including Van-Emde-Boas search tree, Funnel-sort and Median of medians algorithm.
  - Analysed memory-transfer complexity of above algorithms. Used Valgrind to calculate cache hit-miss ratio of cache oblivious algorithms and compared it with their cache ignorant counterparts.

Achievements

- **CSE Department Rank**: **1/66** and **Institute Rank: 2/273** at IIT Indore (based on CGPA).
- **Awarded AP grades**: for exceptional performance in **5** courses.
- **Best UG Researcher Award 2021**: For support vector clustering and its applications to biomedical domain.
- **Summer Research Fellowship**: Awarded research fellowship at Centre for Neuroscience, IISc Bangalore.
- **Inter IIT Tech Meet 9.0**: Awarded Silver medal in Bridgei2i's NLP competition.
- **Inter IIT Tech Meet 8.0**: Awarded Bronze medal in BITGRIT's data-science competition.
- **JEE Advanced 2018**: Secured All India Rank 1055 among 150,000 candidates (top 0.7%).

Relevant coursework

- **Computer Science**: Computer Vision, Cryptography & Network Security, Advanced Algorithms, Machine Learning, Compiler Techniques, Computational Intelligence, Computer Networks, Computer Graphics & Visualization, Optimization Algorithms & Techniques, Operating Systems, Computer Architecture, Parallel Computing, Design & Analysis of Algorithms, Software Engineering, Automata Theory & Logic, Logic Design, Data Structures & Algorithms, Database & Information Systems[†], Discrete Mathematical Structures.
- **Mathematics**: Numerical Methods[†], Complex Analysis & Differential Equations II, Linear Algebra & Differential Equations I, Real Analysis.
- **Other**: Psychology[†], Physics[†], Chemistry[†]
- **AI for Medicine [(Coursera)](#)**: AI For Medical Diagnosis, AI For Medical Prognosis & AI For Medical Treatment.

  ([†] indicates an exceptional grade (AP) achieved in the course.)

Technical Skills

- **Programming**: Python, C++, C, Java, SQL, MATLAB, VHDL, HTML, CSS.
- **Tools/Libraries**: Pytorch, Tensorflow, Keras, Xilinx, Docker, LaTeX.

Positions of Responsibility

- **President of Cynaptics Club**, AI and ML student club of IIT Indore. *Aug 2020 - Jun 2021*
- **Department Undergraduate Representative**, Discipline of Mathematics at IIT Indore. *Feb 2020 - Oct 2020*