

Tarun Gupta

B.Tech., Computer Science and Engineering
Indian Institute Of Technology (IIT), Indore

tarungupta360@gmail.com
tarun360.github.io (personal website)
linkedin.com/in/tarun-gupta-116860178

EDUCATION

Degree/Certificate	Institute/Board	CGPA/Aggregate	Year
B.Tech., Computer Science and Engineering	Indian Institute of Technology (IIT), Indore	9.81/10 (Department rank 1/64, Institute rank 2/273)	2018-2022
Senior Secondary	CBSE Board	94.2%	2018

PUBLICATIONS

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

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. Interspeech 2022.

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.

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 (JBHI).

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).

INTERNATIONAL RESEARCH EXPERIENCE

Research Intern at NTU Singapore and [Skit.ai](#)

Dec 2021 - June 2022

Multilingual Speech Processing, Self-Supervised Learning, Transformers

[Paper 1](#), [Paper 2](#)

- Built a bi-encoder transformer mixture model to estimate speaker age and height from speech signal, leading to a relative improvement of 18.5% in male age estimation and 8.6% in female age estimation over the state-of-the-art benchmarks on the TIMIT dataset (refer to 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, resulting in a significantly well-calibrated model (refer to Paper 2).

Research Intern at Carnegie Mellon University

Jan 2021 - May 2021

Computer Vision, Self-Supervised Learning, Computational Biology

[Paper](#)

- Built a contrastive self-supervised learning (CSSL) pipeline for macromolecular structure classification using cryo-electron tomography (cryo-ET) data.
- Extended techniques such as SimCLR, MoCo, and SwAV to the cryo-ET domain, achieving a significant improvement over non-self-supervised learning methods.

RESEARCH ASSISTANTSHIP

Research Assistant at IIT Indore

Jan 2020 - May 2020

SVM, Clustering, Convex Programming

[Paper 1](#), [Paper 2](#)

- Designed a novel plane-based clustering algorithm based on the principles of SVM, utilizing 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 compared to existing plane-based clustering algorithms.

WORK EXPERIENCE

Member Technical at D. E. Shaw India

July 2022 - Present

Software Development, High-Performance Computing, Security

- Developing an internal High-Performance Clustering (HPC) scheduling software.
- Implemented an integrated ML workload and data pipeline for scheduling workloads on public cloud platforms.
- Contributed to the open-source container technology tool Podman by adding a feature to encrypt container-images ([GitHub PR](#)).

ACHIEVEMENTS

- **CSE Department Rank 1/64** and **Institute Rank 2/273** at IIT Indore (based on CGPA).
- **Best Undergraduate Researcher Award, 2021.**
- **Awarded AP grades:** for exceptional performance in **5** courses.
- **Inter IIT Tech Meet 9.0 (2021):** Awarded **Silver medal** in Bridgei2i's NLP competition.
- **Inter IIT Tech Meet 8.0 (2020):** Awarded **Bronze medal** in BITGRIT's data-science competition.
- **Awarded Research Fellowship (2020)** at **Center for Neuroscience, IISc Bangalore.**
- **JEE Advanced 2018:** Secured All India Rank 1055 among 150,000 candidates (top 0.7%).

KEY PROJECTS

- **Satellite video prediction (in collaboration with [REint.co](#))** *August 2023 - September 2023*
Computer Vision, Spatio-Temporal Predictive Learning
 - Implemented and tested various spatio-temporal prediction models, such as ConvLSTM and SimVP, on real-world satellite data. The predicted frames can be fed to models such as MetNet-2 for weather prediction.
 - Explored and employed PyTorch DDP for multi-node and multi-GPU training.
- **Parallelizing Red Deer Algorithm (RDA) – A Nature Inspired Meta-heuristic Algorithm** *April 2021*
Parallel Programming, MPI [GitHub](#)
 - Implemented Red Deer Algorithm (RDA) for solving the Traveling 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 parallelizing the algorithm, achieving a speed-up factor up to 4.
- **Bridgei2i's NLP competition at Inter IIT Tech Meet 9.0 (2021)** *March 2021*
Awarded Silver Medal [GitHub](#)
 - *Theme identification task:* Applied DistilBERT, a distilled version of the BERT model, for fast and scalable binary classification of tweets and articles, achieving an F1 score of 0.89.
 - *Sentiment analysis task:* Utilized Ada-BERT transformer model fine-tuned for aspect-based sentiment analysis to identify the sentiments associated with a particular brand.
 - *Headline generation task:* Utilized Pegasus, T5, and BART models, achieving 37% average similarity score.
- **Adversarial Attacks on Brain Tumor Segmentation Models** *April 2020-June 2020*
AI for Medicine [GitHub](#)
 - Implemented a 3D-UNet model using the TensorFlow library and trained it on BraTS brain MRI data for brain tumor segmentation.
 - Tested the model's robustness by implementing adversarial attacks such as Fast-Gradient Sign Method (FGSM), Iterative Fast-Gradient Sign Method (iFGSM), and Carlini & Wagner (CW) attack.
- **Cache-Oblivious Algorithms** *May 2020 - July 2020*
Design and Analysis of Algorithms [GitHub](#)
 - Implemented various cache-oblivious algorithms such as Van-Emde-Boas search tree, Funnel-sort, and Median of medians algorithm.
 - Analyzed the memory-transfer complexity of the above algorithms. Used Valgrind to calculate the cache hit-miss ratio of cache oblivious algorithms and compared it with their cache ignorant counterparts.

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, Calculus.
- **Other:** Psychology[†], Physics[†], Chemistry[†].
- **AI for Medicine (Coursera):** AI For Medical Diagnosis, AI For Medical Prognosis & AI For Medical Treatment.
- **Quantum Computation (David Deutsch).**
([†] indicates an exceptional grade (AP) achieved in the course.)

TECHNICAL SKILLS

- **Programming:** Python, C++, C, POSIX C, Go, Java, SQL, VerilogHDL, HTML, CSS
- **Tools/Libraries:** PyTorch, TensorFlow, Container technology (Podman, Docker), LaTeX.

POSITIONS OF RESPONSIBILITY / EXTRA CURRICULAR ACTIVITIES

- **Head, Cynaptics Club (AI and ML student club of IIT Indore):** Led, mentored, and taught club members and students. Organized competitions and administered tutorials. *Aug 2020 - Jun 2021*
 - **Department Undergraduate Representative,** Department of Mathematics. *Feb 2020 - Oct 2020*
-