

## EDUCATION

**Bachelor of Technology**, Manipal Institute of Technology

2018-2022

- BTech in Electronics and Communication Engineering (Minor Specialisation: Data Science)
- CGPA: 9.24/10.0 (Minor: 10.0/10.0)
- Ranked 1<sup>st</sup> in the 7<sup>th</sup> semester with a GPA of 9.8/10
- Awarded a [Certificate of Merit](#) for being ranked 11<sup>th</sup> out of 228 finishing in the top 5%

## UPCOMING PUBLICATIONS

- [In preparation] **A. Borthakur**, A. Zimmermann, *Goalkeeper distribution is a reliable indicator of squad possession*, **Journal of Quantitative Analytics in Sports**

## ACCEPTED PUBLICATIONS AND CONFERENCE ABSTRACTS

- [Published - 05 January 2024] G. Bhatta, S. Gharat, **A. Borthakur** and A. Kumar, *Gamma-ray Blazar Classification using Machine Learning with Advanced Weight Initialization and Self-Supervised Learning Techniques*, **Monthly Notices of The Royal Astronomical Society**, ([Link](#)) ([Code](#))
- [Published - 23 November 2023] S. Gharat, **A. Borthakur** and G. Bhatta, *Estimation of redshift and associated uncertainty of Fermi/LAT extra-galactic sources with Deep Learning*, **Monthly Notices of The Royal Astronomical Society**, ([Link](#)) ([Code](#))
- [Accepted - 31 October 2023] S.Gharat, G.Bhatta and **A. Borthakur**, *Gamma Ray AGNs: Estimating Redshifts and Blazar Classification using Neural Networks with smart initialization and self-supervised learning*, **37<sup>th</sup> Conference on Neural Information Processing Systems (NuerIPS) @ ML4PS workshop**, ([Paper #116](#)) ([Paper pdf](#)) ([Poster png](#))
- [Accepted - 14 October 2023] S.Gharat, G.Bhatta and **A. Borthakur**, *Gamma Ray AGNs: Estimating Redshifts and Blazar Classification using Neural Networks with smart initialization and self-supervised learning*, **243<sup>rd</sup> Meeting of the American Astronomical Society**
- [Published - 12 July 2023] S.Gharat, B.Bose, **A. Borthakur** and R. Mazumder, *An Image Processing approach to identify solar plages observed at 393.37 nm by the Kodaikanal Solar Observatory*, **Royal Astronomical Society Techniques and Instruments**, ([Link](#)) ([Code](#))
- [Accepted - 10 January 2023] S.Gharat, **A. Borthakur**, *A Bayesian approach to classify SDSS images*, **41<sup>st</sup> Meeting of the Astronomical Society of India**, ([Conference ID #ASI2023\\_751](#)) ([Code](#))
- [Accepted - 10 January 2023] S.Gharat, B.Bose, **A. Borthakur** and R. Mazumder, *An Image Processing approach to identify solar plages observed at 393.37 nm by the Kodaikanal Solar Observatory*, **41<sup>st</sup> Meeting of the Astronomical Society of India**, ([Conference ID #ASI2023\\_221](#)) ([Code](#))

## RESEARCH EXPERIENCE

**AI4Astro group @ CMINDS, IIT Bombay**

September 2022-Present

**Gamma Ray AGNs: Estimating Redshifts and Blazar Classification using Neural Networks with smart initialization and self-supervised learning**

Guide: Dr. Gopal Bhatta, University of Zielona Góra, Poland

- Designed and implemented a novel neural network architecture to regress redshifts present in the 4LAC-DR3 catalog and achieved better results when compared to the existing literature with correlation coefficient of 0.78 and an RMSE of 0.415.
- Extended the aforementioned network using variational inference to account for the uncertainty in making predictions for samples with unknown redshift values/ground truths.
- Proposed and implemented a lightweight neural network to classify BCU samples as either BLLacs or FSRQs with an accuracy of 93% and an F1-score of 0.914, while experimenting with techniques such as class weights and bias initialisation to combat class imbalance and deployed a webapp for the same on [AWS EC2](#)
- Identified research gaps in the existing literature and thus, explored and implemented various advanced techniques such as supervised and unsupervised greedy pre-training, and self-supervised learning with an accuracy and F1 score as high as 91.5% and 0.901, respectively.

**An Image Processing approach to identify solar plages observed at 393.37 nm by the Kodaikanal Solar Observatory**

Guide: Dr. Rakesh Mazumder, ARIES Nainital, India

- Designed a custom algorithm for semantic segmentation of solar plages and area calculation achieving a positive correlation of up to 97% between calculated and ground truth areas
- Integrated the algorithm with a Python webapp deployed on Streamlit with the BigQuery GCP service as a storage backend

**Efficient phase recovery with the help of a CNN****Guide: Dr. Sujit Kumar Sahoo, IIT Goa**

- Implemented [this paper](#) in TensorFlow 2.x
- Achieved impressive results (0.052 MAE) leading to sub-millisecond image retrieval on modifying the original architecture to create a lightweight Convolutional Neural Network (CNN)

**Department of ECE @ Manipal Institute of Technology, Manipal****Spring 2021****Enhancing small object detection in the RefineDet architecture on a custom dataset****Guide: Mrs Akshatha KR, MIT Manipal**

- Participated in a lengthy data annotation process where I helped annotate around 4000 frames with the aid of the Computer Vision Annotation Tool (CVAT) and wrote an efficient data ingestion pipeline in order to ingest the annotated data and train the model
- Implemented the model from scratch in the TensorFlow 2.0 framework and proposed changes to the anchor-box initialisation technique in order to improve the performance of the base model.
- Improved performance using k-means clustering on the annotations in order to improve the initialisation of the anchor box aspect ratios and scales led to an improvement in the mean average precision by 10.6%, while maintaining the speed of the original model.
- Final grade: A+

**WORK EXPERIENCE**

---

**Searce Inc @ Pune, Maharashtra, India****January 2022 - September 2022****Machine Learning Engineer****Manager: Dr. Muthukumaraswamy B, Associate Director - Applied AI**

- Prepared for and achieved the Google Cloud Professional Machine Learning Engineer Certificate, even though it's recommended only for industry professionals with 3 or more years of experience.
- Developed a JSON file parsing algorithm for a US-based automation client and **co-lead** the database design for the same on GCP Bigquery
- Built a custom OCR project for extracting fields from Income Tax Return Acknowledgement forms using YOLOv5 architecture, achieving an average extraction time of 3.43 s and mean average precision of 96.9% on the validation set and 92.1% on the test set
- Performed a literature review of various techniques for legal document summarization and developed Jupyter notebooks for the same
- Developed an object detection model for document localization on wooden surfaces, obtaining a mean average precision of 97.1% and used Google Cloud Vision API to extract barcodes from said documents

**SKILLS**

---

- **Languages:** C, C++, Python, Java, MATLAB, LabVIEW, R, RStudio, SQL
- **Frameworks:** OpenCV, TensorFlow, Scikit-Learn, PyTorch
- **Cloud:** GCP, AWS
- **Experiment tracking:** Weights and Biases, mlflow
- **Containerization:** Docker, Kubernetes
- **Version Control:** Git
- **Projects:** [Click here](#)

**MISCELLANEOUS**

---

- **Certifications:** [Click here for a complete list of certifications](#)
- **Courses and grades:** [Click here for the scanned transcript \(BTech\)](#)
- **TOEFL:** [Click here for the score report](#)

**EXTRACURRICULAR**

---

**The FootyPod****July 2023-Present**

- A football (soccer) podcast delivering unique insights through an analytical and data-driven approach.
- Around 300 plays on Spotify with 10+ five-star ratings.
- [Link to Spotify page](#)