Ihsan ULLAH

ihsan2131@gmail.com — linkedin.com/in/ihsaan-ullah — github.com/ihsaan-ullah — ihsaan-ullah.github.io

Interests — Artificial Intelligence, Data Science, Datasets, Competitions/Benchmarks, Software Engineering

Skills

Languages Python, Swift, Javascript, PHP
Machine Learning Data processing and analysis, Datasets preparation, Challenge Organization
ML Tools Pytorch, Tensorflow, Keras, Scikit-Learn, Pandas, Numpy
Cloud GCP(Compute Engine, Cloud Storage)
Web React, Django, HTML, CSS, Bootstrap
Databases MySQL, PostgreSQL, NoSQL, Firebase
Technical Writing Research/Project proposals, Technical reports, Conference papers
Other Docker, CircleCI, Git/GitHub, OpenCV, LaTeX

Experience

ChaLearn U.S.A **Jan 2023 – Present**
Research Software Engineer
- **Data:** Data analysis, preprocessing, formatting, visualization, and datasets curation.
- **Competitions and Benchmarks:** Competitions/benchmarks design, metrics, leaderboards, tasks, evaluation, GitHub repository management, setting up competitions/benchmarks on open-sourced platforms (CodaLab and Codabench), and setting up compute workers on Google Cloud Platform. Creating competition bundles, starting kit Jupyter notebooks, and Google Colab notebooks. Creating Docker images for specific competition/benchmark needs.
- **Project Multitasking:** Managing concurrent and cross-disciplinary projects, managing deliverables, and meeting deadlines. Developing project websites for project visibility.
- **Solution Development:** Designing solutions using popular machine learning frameworks such as PyTorch, TensorFlow, and Scikit-learn. Writing Python code tailored to project needs, ensuring scalability and efficiency.
- **Platform Collaboration:** Collaborating with CodaLab to maintain their open-sourced platform by fixing reported issues, adding new features, and improving functionality based on user feedback.
- **Writing and Documentation:** Writing proposals, reports, and research papers. Preparing documentation for projects, competitions, and benchmarks to ensure clarity and accessibility.
- **LLMs for Research:** Utilizing Large Language Models (LLMs) for research tasks such as text and image generation, prompt engineering, and automating creative processes for better efficiency and scalability.

Crédit Agricole SA, Paris France **May 2022 – Oct 2022**
Data Science Intern (Data Analytics Team)
- Fairness in Machine Learning and AutoML
- Bias detection and mitigation

LISN, Université Paris-Saclay, France **May 2021 – Aug 2021**
Artificial Intelligence Intern
- Datasets preparation and ML Challenge Organization
- Transfer Learning, Meta-Learning, Computer Vision

Université Paris-Saclay, France **Jan 2021 – Feb 2021**
Travail d'etude et de recherche
- Data preprocessing
- Challenge Design and Organization

Codematics Inc, Abbottabad Pakistan **Jan 2020 – Aug 2020**
Software Project Manager

Codematics Inc, Abbottabad Pakistan **Jan 2019 – Dec 2019**
Software Engineer

Ocheng/Chinaccelerator, Shanghai China **May 2018 – Dec 2018**
Software Engineer

Education

Master's in Artificial Intelligence **Sep 2020 – Nov 2022**
Université Paris-Saclay, France
Scholarship: Labex Digicosme Excellence Scholarship

Bachelor of Science in Software Engineering **Sep 2013 —- Aug 2017**
University of Engineering and Technology Peshawar, Pakistan
CGPA: 3.70/4.0
Award: University Gold Medal – Graduated with Distinction

Publications

Stylized Meta-Album: Muti-domain computer vision meta-dataset **Submitted**
Journal of Data-centric Machine Learning Research
https://stylized-meta-album.github.io/

RelevAI-Reviewer: A Benchmark on AI Reviewers for Survey Paper Relevance
Conf´erence sur l'Apprentissage automatique (CAp) 2024
https://hal.science/LISN/hal-04608255v1

Meta-Album: Multi-domain Meta-Dataset for Few-Shot Image Classification
Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (2022)
https://meta-album.github.io/

Cross-Domain MetaDL: Any-Way Any-Shot Learning Competition with Novel Datasets from Practical Domains
NeurIPS 2022 Competition Track (2022)
https://metalearning.chalearn.org/#h.1u3w2mh9pbjf

Lessons learned from the NeurIPS 2021 MetaDL challenge
PMLR - Proceedings of the NeurIPS 2021 Competition and Demonstration Track (2022)
https://proceedings.mlr.press/v176/el-baz22a.html

MetaDL: Few Shot Learning Competition with Novel Datasets from Practical Domains
NeurIPS 2021 Competition Track (2021)
https://metalearning.chalearn.org/metadlneurips2021

Projects

NeurIPS Checklist Assistant
An LLMs based checklist assistant for NeurIPS submissions
https://www.codabench.org/competitions/2338/
https://blog.neurips.cc/2024/05/07/soliciting-participants-for-the-neurips-2024-checklist-assistant-study/
- This study explores the use of Large Language Models (LLMs) as an assistant to help authors verify their submissions against the NeurIPS Paper Checklist. The goal is to assess whether LLMs can improve submission quality at NeurIPS. Participants receive feedback from an experimental LLM assistant to check compliance with NeurIPS submission standards. The LLM provides detailed feedback on the checklist responses to help authors refine their papers before submission. While the tool offers valuable guidance, it is meant to complement, not replace, the author's judgment and expertise.

FAIR Universe
Unbiased Data Benchmark Ecosystem for Physics
https://fair-universe.lbl.gov/
- The FAIR Universe project, funded by the US Department of Energy, is a collaboration between Lawrence Berkeley National Laboratory, Université Paris-Saclay, University of Washington, and ChaLearn. The initiative aims to create a large-scale AI platform for hosting scientific datasets, models, and machine learning competitions to advance discoveries in high energy physics and cosmology. The project focuses on reducing systematic uncertainties in High Energy Physics through a series of challenges. Key events include a toy challenge (October 2023), a Particle Physics hackathon (November 2023), and the HiggsML Uncertainty Pilot Competition (March 2024). A major challenge on uncertainties in fundamental science is set to launch at NeurIPS 2024. The project is ongoing and will conclude in 2025.

Stylized Meta-Album
Muti-domain computer vision meta-dataset
https://stylized-meta-album.github.io/
- The Stylized Meta-Album (SMA) is a new image classification meta-dataset featuring 24 datasets (12 content and 12 stylized) to support research in out-of-distribution (OOD) generalization and related areas. SMA combines diverse subjects and styles, creating 4800 groups that offer extensive variability for rigorous studies. It introduces benchmarks for OOD generalization and group fairness, as well as unsupervised domain adaptation (UDA), showing the importance of group diversity in fairness and algorithm rankings, while also reducing error bars in benchmarking scenarios.

Meta-Album
A meta-dataset for few-shot image classification
https://meta-album.github.io/
- The Meta-Album is an image classification meta-dataset designed to support tasks such as few-shot learning, transfer learning, and meta-learning. It consists of 40 diverse open datasets from domains like ecology, manufacturing, human actions, and optical character recognition, each containing at least 20 classes with 40 examples per class. The datasets are uniformly preprocessed and available in three versions (Micro, Mini, Extended) to accommodate different computational needs. Meta-Album, already larger and more diverse than similar efforts, continues to expand through a series of competitions, creating a rolling benchmark for research.