M. Alfi Hasan, PhD

Current Position: Staff Data Scientist at Walmart Global Tech

Email	malfihasan19@outlook.com	**Website**	www.malfihasan.com
Phone	+1-347-265-5476	**GitHub**	github.com/malfihasan
Address	Dallas, Texas [open to relocate]	**LinkedIn**	malfihasangeo

Data scientist and technical leader specialized in Deep Learning on Edge Devices, Multi-spectral Remote Sensing, and Large Vision Models across diverse domains, with over 12 years of experience (5+ years in academia, 7+ years in industry).

- As a project **Tech Lead**, led the deployment of iOS/Android efficient and React-supported deep learning models on IoT/edge devices for the recognition of over a million items at Walmart.
- Doctoral research focused on predicting disease epidemic cycles using real-time satellite images, such as Planet, LandSat, MODIS, and mathematical models like TabNet, LASSO (causal inference), and LSTMs.
- Have deep domain knowledge and hands-on experience with advanced ML/DL architectures, such as Generative Adversarial Networks (GANs), Segment Anything series, YOLO series, CLIP, Grounding DINO, Transformers, etc., in the Retail Industry.
- Advanced proficiency in Python, C, and R; designed end-to-end processes from device-level image capture to display on servers as dashboards (Tableau, Portal) while maintaining CI/CD pipelines.
- Have a unique strength in adopting multidisciplinary techniques in cross-functional corporate and academic environments.
- Supported large-scale projects with modeling expertise in deploying satellite- and drone-based projects for agricultural operations, mentoring junior team members, and teaching programs at universities like Purdue.
- Looking forward to senior scientist or leadership roles.

EXPERIENCE

Staff Data Scientist — **May, 2024 - Present**

Project Constellation, Walmart Global Tech, Bentonville, AR, USA

- Leading the data science technical effort to **provide a cost-saving management solution for store associates** by creating an visual chat intelligence from various cameras utilizing the latest large vision models (LVMs). This project leverages various LLM backbones such as **GPT-4o, LLaMA 3.1, LLaMA-LLava and nano-LLava, combined with LangChain and llama.cpp framerwork**, to provide inventory solution for walmart.
- **Coordinating and designing an in-house LLMs based framework** to reduce costs by avoiding commercial LLMs, using models like **LLava-LLaMA3 and prompt techniques like CoT, ReACT, VCoT, etc.**
- Guiding and coding a multi-barcode reader solution that **enhances the shelf-stocking process** in over 1,000 Walmart stores, utilizing deep learning architectures like Grounding DINO and SAM (Segment Anything).
- Designing and leading an auto-annotation framework that **saves time and reduces the cost of the annotation team**, utilizing the latest models like **Grounding SAM and multi-modal models**.
- Supporting and contributing to popular open-source models such as **YOLOX** and inspired to utilized within the company.
- Assisted the company in recruiting data scientists by participating in the technical interview panel.

Senior Data Scientist — **May 2022 - April 2024**

Project Diamond, Walmart Global Tech, Bentonville, AR, USA

- Led the development of an annotation tool for handheld devices (for both Android and iOS) as a tech lead, significantly **reducing the cost and time of manual human annotation** using **transfer-learned object detection (YOLOv8) and classification (embedding-based EfficientNet) models**. Managed cross-functional engineering and product teams.
- Conducted technical orchestration and coded a deep learning-based barcode detector, **replacing Walmart's current barcode detection system**, estimated to save about $4 million a year. The framework consists of a real-time object detection pipeline (for both Android and iOS, inspired by EdgeYOLO) with the **Zebra Crossing library**, providing storage efficiency and high performance.
- Proposed, orchestrated, and directed an **automated end-to-end model training and inference pipeline** for the data science team. This helped scale dynamic product detection for several Walmart stores using 100+ custom-made cameras. The tools used were **Looper, Docker, Concord, and GCP instance**.
- Developed and implemented a **novel architecture for dimension detection**, creating the potential to save thousands of dollars using an in-house deep learning architecture similar to **MiDAS (depth detection) and EfficientNet (classification)**.
- Orchestrated the deployment of a framework **handling over 1,000 image processing requests per second**, utilized for machine learning model (both object detection & classification) inference in conjunction with **embedding vector databases (Milvus)**.

- Organized and led innovative research by **hosting bi-weekly Data Science meetups** and maintaining a mono-repo to facilitate knowledge sharing and team objectives for the Data Science team.
- Developed a recommendation system that provides a **dynamic product recognition solution** for Walmart's inventory management process, utilizing **an embedding-based Siamese model with LightGBM**.
- Developed a modularized ML deployment pipeline in Go-Lang that **enabled the inference of core deep learning models** across multiple platforms, heavily utilizing tools and relational databases such as **SQL, PostgreSQL, BigQuery, and the company portal.**

Senior Data Scientist **Oct 2021 - May 2022**

Research and Development (Breeding), Bayer Crop Science, St. Louis, MO, USA

- Drove roadmaps for quantitative solutions and **reduced operational costs by over 50%** for North American and European canola operations by proposing and developing an automated UAV-based canola flowering detection model. The development process utilized image segmentation models like DeepLab3, image classification models such as the ResNet backbone, conventional computer vision (CV) algorithms using OpenCV, and cloud computing platforms like AWS-EC2 instances.
- Automated the cotton UAV pipeline operation for thousands of fields across North America by co-developing deep learning-based scalable solutions that detect the maturity of cotton from images. The project utilized AWS-Sagemaker in conjunction with parallelized Docker containers and YOLOv3 object detection architecture, with multi-year transfer learning inferences, and involved collaboration with stakeholders from diverse domains using version control systems such as Git.
- Created **novel and creative embeddings for temporal models**, which resulted in a patent. The development involved modeling with architecture families like **GANs (Pix2Pix, StyleGAN) and Transformers** on multi-band geospatial images using PyTorch (Python), alongside querying languages like SQL.
- Provided visibility of model-derived results that supported several operational teams by developing geospatial visualizations through a **Tableau dashboard**. This project required geospatial expertise using packages like GDAL and Shapely, skills that conventional software architects lacked.
- Explored new satellite acquisition opportunities, leading to the procurement of new satellite products by providing market analysis on the latest satellite vendors as part of the R&D team's 5+ year vision.
- Assisted the company in recruiting data scientists for three teams by participating in the technical interview panel during the HR interview process.

Spatial Data Scientist **Mar 2020 - Oct 2021**

Research and Development Breeding, Bayer Crop Science, St. Louis, MO, USA

- *Increased breeding pipeline efficiency* and introduced bias-free automated information for global cotton stakeholders with a cost-saving potential of almost half a million by leading and developing a UAV-based multi-flight cotton maturity detection framework to eliminate ground notes. To develop the framework in this multi-year project utilized semantic segmentation models like UNet and RCNN in conjunction with Decision tree models, CV algorithms like Canny Edge detector in Unix OS like Ubuntu using Tensorflow. The project is being implemented in production for 500+ flights each season using computer vision models.
- Developed several automation modules for the first scalable UAV processing pipeline for the company where the project *enabled the elimination of field notes for many crops and saved millions of dollars* on the field operations. In the development process used geospatial python libraries like geopandas, shapely, and Kerus in the Postgres server and conducted feasibility studies using IoT devices. The project won the breeding **excellence awards for 2021**.
- Orchestrated components of the global UAV-based soy maturity automation pipeline that *created several thousand acres of new soy maturity operations* by developing tools for color calibration, GCP detections, and maturity modeling. The work resulted in another patent for the company.

Research Data Scientist (Geo-Spatial) **Sep 2018 - Mar 2020**

Bayer Crop Science (Contract: Colaberry Inc), St. Louis, MO, USA

- Used traditional machine learning models like **Random Forest, TabNet, and XGBoost as well as utilized Causal inference & Bayesian approaches**.
- *Eliminated field monitoring operation by more than 50%* by proposing and developing a Satellite-based crop health quality recommendation system. The project conducted time series forecasting and anomaly detection using model families like GANs and also utilized high-resolution multi-band structures using model families like LSTM.
- Provided operational placement benefit for a global testing network by *generating satellite-based clustered embedding* using the word2vec model framework.
- Increased understanding and helped to generate new seeds for corn stakeholders by developing lodging & yield prediction models using Satellite in the company.

Remote Sensing Data Scientist (Summer Intern) **Jun 2018 - Aug 2018**

Monsanto (Contract: Colaberry Inc), St. Louis, MO, USA

- Automated open-source, cost-free stitching solution for hundreds of UAV flights by *developing a scalable stitching pipeline* of Open Drone Map (ODM).

- Developed code for the AWS pipeline, specifically in the automated streamlining of data in Google Big Query and AWS - S3.

Graduate Research and Teaching Assistant **Jan 2015 - May 2018**

University of Rhode Island, Kingston, RI, USA

- Developed a model that *predicts the epidemics of rotavirus* in the developing world 1-month prior using satellite under the project named "Control of Endemic Cholera in Bangladesh," funded by the Bill & Melinda Gates Foundation using Causal inference.
- Developed an irrigation app that *recommends irrigation to farmers* under a summer project with The International Maize and Wheat Improvement Center (CIMMYT).

Research Associate **Jul 2011 - Dec 2014**

Institute of Water and Flood Management (IWFM), Dhaka, Bangladesh

- Collaborated and worked on International Research Projects like: High End Climate Impact and Extremes (HELIX), High-resolution Regional Climate Change Information for Bangladesh to inform Impacts assessments.

Visiting Scientist **Jul 2012 - Sep 2012**

Met Office, Exeter, United Kingdom

Lecturer **Jun 2011 - Apr 2012**

World University of Bangladesh, Dhaka, Bangladesh

EDUCATION

Ph.D. in Civil and Environmental Engineering Aug 2018
University of Rhode Island, Kingston, RI, USA

Master of Science in Water Resources Development Dec 2014
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

Bachelor of Science in Water Resources Engineering Feb 2011
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

LEADERSHIP, AWARDS AND EXTRACURRICULAR ACTIVITIES

- Led the team and won 1st place in the Gen AI category and secured a top 20 position overall in the Techathon 2023, hosted by Walmart Global Tech.
- Received the 'Exceptional Performer' award within a large team of 50+ employees at Walmart Global Tech in 2023.
- Won the Digital Innovation Award 2021 from Bayer Crop Science for developing an automated UAV pipeline, which is an end-to-end process from capturing images by drones to backend image processing, deploying deep learning models, and visual inference on a dashboard.
- Facilitated novel solutions in remote sensing (RS) in agriculture, leading to the development of patents for the organization by connecting individuals and scientists through meetups on satellite and RS disciplines, and creating cultural awareness.
- Collaborated with universities such as Purdue University, the University of Illinois at Urbana-Champaign, and the University of Missouri as a corporate mentor, providing mentorship and emotional intelligence to graduate students in RS and agriculture.
- Managed 3,000+ student blood donors as the nominated President of the Blood Donation organization, Badhon-Suhrawardi Hall Unit, in 2008 and 2009.

PUBLICATIONS

Have published 9 journal papers and 20 conference papers and reviewed many more. Have more than 800+ citations as of now. Presented talks and posters in various conferences. List: **more details in google scholar.**