AMAN WAOO

(+1) 919-985-9088 | Raleigh, NC | awaoo@ncsu.edu | Portfolio | linkedin.com/in/aman-waoo | github.com/amanwaoo

EDUCATION

North Carolina State University, Raleigh **Aug 2022 – May 2024**

Master of Science, Electrical and Computer Engineering *GPA – 4.0/4.0*

Coursework: Advanced Machine Learning, Automated Learning and Data Analysis, Data Science, Neural Networks and Deep Learning, Object Oriented Design and Development, Computer Vision, Pattern Recognition, Digital Imaging Systems

National Institute of Technology Bhopal, India **Aug 2016 – May 2020**

Bachelor of Technology, Electrical Engineering *GPA – 7.9/10*

PROFESSIONAL AND RESEARCH EXPERIENCE

North Carolina State University (Sozzani Lab) **Sep 2023 – Present**

Research Assistant *Raleigh, North Carolina*

- Built protein family classification network for functional annotation of protein sequences for signaling network inference and achieved F1 score of **96%**, by developing a custom **CNN-Attention-BiLSTM** model and **SHAP** for explainability, leading to a publication in *Nature Journal.*
- Developed a Conditional **Variational AutoEncoder** model to analyze Soybean mass spectroscopy data, linking agronomic traits such as seed production. Collaborated in creating **Modified Shannon Entropy** method in R, to capture differential expression in temporal genomics data.
- Conducted Exploratory Data Analysis (**EDA**) on **50,000+** genomic data samples of Arabidopsis gene expression data using SQL and Python.

Schlumberger Limited (SLB) **May 2023 – Aug 2023**

Data Science Intern *Houston, Texas*

- Improved predictive maintenance workflows for determining Remaining Useful Life (**RUL**) of electronic sensor boards on oil drilling tools, by implementing **classical machine learning models** (xgboost, random forest, GBM) with recall scores of **92%.**
- Mitigated dataset imbalance using oversampling techniques and Generative Adversarial Network (**GAN**). Leveraged advanced statistical features (histograms, quantiles, probability mass function) to integrate **data-driven insights** with domain knowledge.
- Conducted **ETL** operations on various datasets of raw and unstructured sensor data, streamlining it for subsequent failure analysis.
- Applied **clustering and processing techniques**, such as K-means and anomaly detection, to analyze datasets from **1,500+** electronic channels.

Larsen & Toubro Limited (L&T) **Aug 2020 – May 2022**

System Engineer *Gujarat, India*

- Utilized data visualization techniques using **RStudio** and created dashboards using **PowerBI** for reporting, and displaying **KPIs** to business stakeholders for thermal power plant protection system.
- Researched and implemented relay coordination studies using **ETAP** and **PLCs**, leading to a **30%** reduction in operational downtime.
- Established seamless collaboration with **cross-functional teams** spanning various engineering departments, project management group and third-party suppliers. Proposed data-driven optimizations, yielding substantial cost savings across **6 projects** in India.
- Facilitated the creation of **automatic sizing and calculation sheets** using Macros, for electrical and instrumentation packages such as battery packs, switchgears, generators and transformers, lowering the potential for human error.

ACADEMIC PROJECTS

- **Large Language Models – ChatGPT Clone for Question-Answering Application** | OpenAI, MongoDB **Jul 2024**
  - Implemented a semantic search feature using **vector embeddings and Atlas** Vector Search to enhance natural language querying for movies.
  - Developed a question-answering system leveraging the **RAG** architecture and Large Language Models (LLMs) to retrieve and generate accurate responses from user-specific data, achieving **15%** improvement in response accuracy.
- **Deep Learning – Deepfake Images Detection Algorithm** (GitHub) | PyTorch, sklearn **Jun 2023**
  - Generated **120,000** fake images from CelebA real images dataset for different **GANs**, stored them using **AWS S3** and trained a combined **Siamese Network** (Common Fake Feature Network + Classification Network) for fake/real images annotations.
  - Utilized **AWS EC2** GPU instances to expedite model training and achieved impressive accuracy of 99.37% and F1 score of 98.97%.
- **Machine Learning – 2-D Object Detection for Autonomous Vehicle** (GitHub) | TensorFlow, Keras **Mar 2023**
  - Trained **YOLOv3** model with DarkNet-53 architecture as codebase and for MSCOCO dataset with **10,000+** car dashcam images.
  - Surpassed model performance over other algorithms in terms of Frames Per Second **(FPS)** and mean Average Precision **(mAP)** scores.

PUBLICATION

- *Morffy, N. et al. "*Identification of plant transcriptional activation domains.*" Nature (2024).*

TECHNICAL SKILLS

Programming Languages: Python, SQL, C++, MATLAB, R Programming, SAS, VBA (Excel Macros)

Frameworks & Libraries: PyTorch, TensorFlow, Keras, OpenCV, scikit-image, scikit-learn, Pandas, Seaborn, Matplotlib, NumPy

OS & Tools: Linux (Ubuntu), PowerBI (DAX), Databricks, Dataiku (AWS, Azure), Spark, Roboflow, Git, Docker, Kubernetes, Jira, Agile/Scrum

Miscellaneous: Experienced in statistical analysis and conveying narratives through data visualization. Knowledge of machine learning concepts, computer vision algorithms and strong mathematical skills in linear algebra, statistics, probability theory, and geometry.

CERTIFICATIONS AND EXTRACURRICULARS

- Winner in the Machine Learning track at the annual N.C. PSI **Hackathon** as a member of team of four students.
- Dataiku DSS – "Machine Learning Practitioner", "Core Designer", NVIDIA – "Deep Learning Fundamentals", "Image Segmentation Techniques", "Time Series Data Modeling with RNN", DeepLearning.AI – "Generative AI with LLMs".
- Vice Chairperson (Administration) – IEEE MANIT Student Branch – Won Darrel Chong Activity Award (Gold Category).