Isha Arora

Boston, MA | (617) 652-6728 | arora.isha4128@gmail.com
isha-04.github.io | linkedin.com/in/isha-arora04 | github.com/isha-04

EXPERIENCE

Deep Learning Research Assistant **Boston, MA**
Northeastern University *Feb 2023 – Dec 2023*

- Formulated Python pipelines for managing **10616** whole slide images (PANDA dataset), testing on **192** whole slide images (DiagSet dataset) of Prostate Cancer, achieving efficient extraction and transformation
- Refined deep learning models, including EfficientNet, ResNet, PROMETEO, enhancing Prostate Cancer grading precision
- Elevated grading accuracy, attaining **0.66** quadratic weighted kappa (QWK) alongside **0.81** weighted accuracy (WA) with an innovative EfficientNet-B1 configuration, marking around 8% enhancement over previous models
- Initiated deployment of the EfficientNet-B1 model on a Flask web application, optimizing research analysis and aggregation, targeting reduction in pathologists' grading time
- Pioneered model generalizability with Breast Cancer analyzing **9109** images (BreakHis dataset), performing image segmentation on Prostate Cancer slides, alongside integrating TCGA RNA-sequencing data, expanding research scope with multimodality

Data Science Researcher **Boston, MA**
Massachusetts General Hospital *Jan 2023 – Aug 2023*

- Researched an extensive review of over **100** Machine Learning, Deep Learning publications, applying insights to Electronic Medical Records (EMR/EHR) for identifying risk factors in Childbirth-Posttraumatic Stress Disorder (CB-PTSD)
- Processed information from hospital records, aligning **8000** data points with IRB standards, setting the stage for model implementation
- Investigated and validated self-reporting metrics against clinician metrics for accessible PTSD detection through R statistical modeling, achieving **0.94** AUC-ROC, evaluating on **59** patients, with **66%** within first year postpartum
- Visualized patient statistics using R for NIH, achieving a **25%** boost in data clarity and research decision-making efficiency

Associate Engineer **India**
Virtusa Consulting Services Pvt. Ltd. *Aug 2020 – Aug 2021*

- Constructed a PostgreSQL system for hosting US federal and state banking and insurance regulations, improving lookup efficiency by **30%**
- Expedited client data retrieval by **45%** in **5** months, surpassing targets; spearheaded a modeling project using OpenRefine, leading an **8**-member cross-functional team to **40%** increase in analytical capacity.
- Applied Agile methodologies and JIRA for project management, focusing on data storage monitoring to develop visualizations in Python
- Cultivated collaboration with Wolters Kluwer USA, enhancing delivery and inter-organizational ties

Data Scientist Intern **India**
Financial Software and Systems Pvt. Ltd. *Dec 2019 – May 2020*

- Crafted a cost-effective spam detection system for banking app reviews, attaining **65%** accuracy, reducing manual review time by **26%**
- Advanced product development through review sentiment analysis employing VADER (Valence Aware Dictionary for sEntiment Reasoning), leading to **40%** increase in positive feedback with **15%** rise in user engagement
- Designed an AI algorithm for app review authenticity, boosting relevance by **50%**, presenting strategic impacts detailed in a comprehensive stakeholder report

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, MATLAB, Java, C/C++, HTML, CSS, JavaScript
Libraries and Frameworks: TensorFlow, Keras, PyTorch, OpenCV, scikit-learn, SciPy, Plotly, HuggingFace, ggplot2, NLTK, dlib
Databases: MySQL, PostgreSQL, MS SQL Server, Oracle PL/SQL
BI Tools and IDEs: Tableau, Power BI, JupyterLab, Visual Studio Code, RStudio, JIRA, Excel (Pivot, VBA), OpenRefine
Technical Skills: Data Analysis, Data Science, Machine Learning, Deep Learning, Artificial Intelligence, NLP, Computer Vision, Image Processing, git, AWS, Spark, Linux

EDUCATION

Northeastern University, Khoury College of Computer Sciences **Boston, MA**
Master of Science (M.S.) in Data Science *Sep 2021 – Dec 2023*

Vellore Institute of Technology **India**
B.Tech. in Computer Science and Engineering *Jun 2016 – Jul 2020*

PROJECTS

Exploring User Accessibility and Human-Machine Interaction Using EMG — Oct 2023 – Dec 2023

- Crafted a user identification and gesture recognition model using Electromyogram data, aiding those with mobility issues
- Built XGBoost, neural network models for gesture recognition, securing **91%** accuracy alongside an F-1 score of **0.9**
- Attained **94%** accuracy in LSTM user classification, with a notable cross-day rank-5 accuracy of **80.3%**

The Song Search — Oct 2022 – Dec 2022

- Implemented a music retrieval system, harnessing the GTZAN collection for a curated audio dataset, advancing music search functionality
- Secured top 5 candidate set accuracy of **74%** with **0.68** MAP, utilizing TensorFlow MAGENTA's MT3 model (based on T5 architecture), highlighting enhanced precision in music information retrieval

Deep Clustering for Unsupervised Learning of Visual Features - A Reproduction — Mar 2022 – May 2022

- Replicated Facebook AI Research's DeepCluster network, integrating Power Iteration Clustering (PIC) and AlexNet clustering with ImageNet dataset, encompassing **64** classes with **600** images, extending analysis to an external dataset of **28000** images
- Assessed Normalized Mutual Information (NMI) between clusters, reaching an approximate value of **0.8**, indicating robust clustering

A Literature Review on BERT, RoBERTa, and T5 — Mar 2022 – May 2022

- Analyzed distinctions between attention models BERT, RoBERTa, T5, delineating their characteristics and performance metrics
- Reviewed foundational papers of these attention models, detailing the evolutionary improvements from BERT to subsequent models

Cryptocurrency Price Prediction — Nov 2021 – Dec 2021

- Led a comprehensive project aimed at predicting daily price of cryptocurrency, leveraging **5** years of historical data
- Performed Data Analysis and Visualization on cryptocurrency pricing using Python and Tableau, emphasizing daily price fluctuations
- Optimized a Random Forest model, achieving an RMSE score of **0.222**, evidencing model accuracy

Facial Emotion Recognition — Oct 2021 – Dec 2021

- Orchestrated development of a model classifying **28273** facial images into 6 emotions: anger, sadness, happiness, fear, surprise, neutral
- Executed a range of classification algorithms including Decision Tree, Random Forest, Gaussian Naïve Bayes, k-NN, VGG-16, and incorporated the dlib library for image processing, demonstrating versatile modeling skills
- Realized **63%** accuracy benchmark with top-performing VGG-16 model, highlighting model optimization and application

PUBLICATIONS

Arora, Isha Hemant, et al. "Establishing the validity of a Diagnostic Questionnaire for Childbirth-Related Posttraumatic Stress Disorder." American Journal of Obstetrics & Gynecology, Nov. 2023.

From work at Massachusetts General Hospital as a Data Science Researcher

- Highlighted effectiveness of the self-reporting PTSD Checklist (PCL-5) as a preliminary diagnostic tool for CB-PTSD to significantly streamline and enhance efficiency of the diagnostic approach.
- Established cutoff value of **28**, optimizing sensitivity (**0.8**) and specificity (**0.93**), which led to accurate diagnosis in **86%** of women.