Isha Arora

Boston, MA 02120 | (617) 652 6728 | arora.isha4128@gmail.com
Portfolio | LinkedIn | GitHub

EXPERIENCE

Research Trainee | Northeastern University, Boston — Feb 2023 – Dec 2023

- Revised **state-of-the-art** deep learning models for grading **Prostate Cancer** using **10616 WSIs** (PANDA) and **192 WSIs** (DiagSet)
- Acquired **0.66 QWK** with **0.81 weighted accuracy** for a new proposed configuration of EfficientNet-B1 model
- Engineered a use-case to extend to **Breast Cancer** using **9109 WSIs** (BreakHis) and RNA-sequencing data (TCGA) for **multi-modality**

Research Student | Massachusetts General Hospital, Dekel Laboratory, Boston — Jan 2023 – Aug 2023

- Investigated **traditional ML** and **deep learning** approaches using Electronic Medical Records to identify risk factors for CB-PTSD
- Explored PTSD reporting metrics using **statistical** modeling evaluated on **59 within 1-year postpartum patients**
- Authored a data-driven approach to reporting metrics, earning an impressive **0.94 AUC-ROC** and **0.82 correlation value**

Associate Engineer – Technology | Virtusa Consulting Services Pvt. Ltd., India — Aug 2020 – Aug 2021

- Coordinated with **Wolters Kluwer USA** to develop a system hosting regulations in banking and insurance in **Agile** using **PostgreSQL**
- Accelerated lookup time for laws, the system **increased client efficiency** by at least **60%** in **5 months**
- Modeled data sent in by client using **OpenRefine** to assemble according to relevant US states books

Data Analytics Intern | Financial Software and Systems Pvt. Ltd, India — Dec 2019 – May 2020

- Conceived a project on spam detection on **4480 reviews** for a banking application
- Initiated the integration of **Naive Bayes, Decision Tree, Apriori algorithm** for spam detection obtaining **61% accuracy**
- Executed **VADER algorithm** for sentiment analysis recording at least **50%** of reviews marked as non-spam **positive**

PUBLICATIONS

Establishing the validity of a diagnostic questionnaire for childbirth-related post-traumatic stress disorder — Nov 2023

- Validated use of self-reporting PCL-5 checklist to assess CB-PTSD against Clinician CAPS-5 for **59 patients**
- Generated **cutoff value 28** with maximized **sensitivity (0.80), specificity (0.93)**, diagnosing **86%** women
- Observed **Youden J-index 0.71** with an **86% overall diagnostic efficiency** for the cutoff score

PROJECTS

Exploring User Accessibility and Human-Machine Interaction Using EMG **[**GitHub**]**

- Designed a **gesture** recognition and **user** identification model using Electromyogram data to help people with mobility issues
- Formulated a neural network model for gesture recognition accomplishing **91% accuracy** and **0.9 F-1 score**
- Achieved LSTM user classification **accuracy 94%** accompanying cross-day **rank-5 accuracy of 80.3%**

The Song Search [GitHub]

- Developed an **information retrieval system** for audio files referencing the **MT3 model** from **TensorFlow Magenta**
- **Created** a specific dataset for audio data, programming a model by finding efficient representation of songs
- Attained **74% accuracy** in top 5 candidate set alongside **MAP of 0.68**

Deep Clustering for Unsupervised Learning of Visual Features - A Reproduction [GitHub]

- Reproduced the paper introduced by Facebook AI Research creating DeepCluster network with **Power Iteration Clustering** and **AlexNet**
- Clustered with a subset of **ImageNet dataset** with **64 classes, 600 images each**, alongside an **external dataset** with **28000 images**
- Assessed **NMI** between each new and previous cluster, produced an approximate value of **0.8**

EDUCATION

Northeastern University, Boston, MA — Dec 2023
Master of Science in Data Science — **GPA: 3.83**

Vellore Institute of Technology, Vellore, India — Jul 2020
Bachelor of Technology in Computer Science and Engineering — **GPA: 8.67/10**

TECHNICAL KNOWLEDGE

Languages:	Python | R | RStudio | SQL | C++ | Java | MATLAB
Database:	MySQL | PostgreSQL | Oracle PL/SQL
Libraries and Frameworks:	AWS | Pandas | NumPy | Matplotlib | Scikit | seaborn | Keras | TensorFlow | PyTorch | OpenCV | ggplot | GitHub | NLTK | OpenRefine | PowerBI
Technical Applied Skills:	Statistics | Data Mining | Data Science | Machine Learning | Neural Networks | Computer Vision | NLP | Artificial Intelligence