PRIYANSHU SINHA ⬈

Seeking Full Time Data Scientist / ML Engineer Role

Data Science and Machine Learning Professional with 3+ Years' Experience

📱 +1-317-982-4563 ✉ priyanshu_sinha@outlook.com 🔗 linkedin.com/in/pri2si17 ⬡ github.com/pri2si17-1997

Experience

Labcorp Drug Development   **Durham, North Carolina**
Student Intern - Data Science   *September 2022 - Present*

- Utilized **AWS Textract** and **TF-IDF** to extract text from document-based images and then applying **spectral clustering** to group them into clusters, which can serve as pseudo labels to train a multi-class classifier.
- Developed a tool utilizing **Learning Interpretability Tool (LIT)** and **Shapely** analysis to evaluate the **explainability** of deep neural network black box models, with the goal of understanding the factors that impact predictions and meeting stakeholder requirements.
- Developed an **AWS Lambda** function to automate the extraction of John Hopkins COVID-19 data and store it in an AWS S3 bucket.

Labcorp Drug Development   **Princeton, New Jersey**
Data Science Intern   *May 2022 - August 2022*

- Utilized **TensorFlow** and **AWS Sagemaker** to conduct time-series data analysis, built a deep **LSTM** network for weekly clinical screening rate prediction, and investigated the impact of COVID-19 on screening rate.
- Designed a proof-of-concept for augmented annotation that employs **BERT** and **Hugging Face** framework to classify patient emails with minimal samples, reducing manual labeling efforts by 10x without compromising model accuracy.

Indiana University   **Indianapolis, Indiana**
Graduate Research Assistant   *August 2021 - Present*

- Working on early detection of Dementia/Alzheimer's disease by utilizing image and text data modalities and **language models (BERT/adapter)**, with the aim of developing a screening tool for neuro-degenerative diseases and enhancing patient healthcare.
- Collaborated with Emory University researchers and surgeons to develop a **Virtual Reality (VR)** teaching tool prototype for immersive learning during surgical procedures. The prototype features VR annotations via voice and hand controllers for real-time interaction between students and doctors, and employs **image segmentation** techniques to segment surgical scenes in VR.
- Optimized chest x-ray and brain MRI deep learning imaging models using **INT-8 quantization** with **TensorFlow Model Optimization Toolkit**. Achieved 75% smaller models, lower memory usage, and 4x faster inference on **edge devices**, potentially benefiting healthcare development in low-income countries.
- Created **MIMIC-IV** dataset-based tutorials with **PostgreSQL** and **TensorFlow** to teach healthcare professionals and students about applying machine learning and deep learning in healthcare, enhancing AI understanding in healthcare.

Siemens Digital Industries Software   **Noida, India**
Senior Software Engineer   *June 2018 - August 2021*

- Developed **Lidar point cloud clustering** and projection components in **Python** and **C++** for Mentor's low-latency framework (LoLa) to enable efficient classification and detection tasks in autonomous vehicles.
- Developed a semi-automated annotation tool using deep learning models for efficient labeling of autonomous car data, achieving up to 60% reduction in manual labeling efforts
- Performed comparative evaluation of object detection models **YOLO V3** and **RetinaNet**, with a proprietary MSC hierarchical classifier, on KITTI dataset and Simcenter Prescan simulated data.

Skills

Technologies : Machine Learning, Deep Learning, Data Structures and Algorithms, Computer Vision, Natural Language Processing, Time Series Analysis, Database Management System, REST APIs, Object Oriented Programming, AWS
Languages: Python, C++, C, R
Databases: SQL, MongoDB
Libraries & Tools: Tensorflow, PyTorch, Keras, Numpy, Pandas, Scikit-Learn, Scipy, spaCY, OpenCV, Matplotlib, NLTK, Plotly, Apache Spark, PyTest, SHAP, Linux, Git, Docker, MATLAB

Education

Indiana University   **Indianapolis, Indiana**
Master of Science in Health Informatics   *May 2023*

Jaypee Institute of Information Technology   **Noida, India**
Bachelor of Technology in Computer Science and Engineering   *May 2018*

Publications

1. **Priyanshu Sinha**, Sai Sreya Tummala, Saptarshi Purkayastha, and Judy Gichoya. Energy efficiency of quantized neural networks in medical imaging. In *Medical Imaging with Deep Learning*, 2022

2. **Sinha, Priyanshu**, Judy W. Gichoya, and Saptarshi Purkayastha. Leapfrogging medical ai in low-resource contexts using edge tensor processing unit. In *2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, pages 67–70, 2022

3. Judy W. Gichoya, **Priyanshu Sinha**, Melissa Davis, Jeffrey W. Dunkle, Scott A. Hamlin, Keith D. Herr, Carrie N. Hoff, Haley P. Letter, Christopher R. McAdams, Gregory D. Puthoff, Kevin L. Smith, Scott D. Steenburg, Imon Banerjee, and Hari Trivedi. Multireader evaluation of radiologist performance for COVID-19 detection on emergency department chest radiographs. *Clinical Imaging*, 82:77–82, 2021

4. Areeba Abid, **Sinha, Priyanshu**, Aishwarya Harpale, Judy Gichoya, and Saptarshi Purkayastha. Optimizing Medical Image Classification Models for Edge Devices. In Kenji Matsui, Sigeru Omatu, Tan Yigitcanlar, and Sara Rodríguez González, editors, *Distributed Computing and Artificial Intelligence, Volume 1: 18th International Conference*, pages 77–87, Cham, 2021. Springer International Publishing

5. Pradeeban Kathiravelu, Puneet Sharma, Ashish Sharma, Imon Banerjee, Hari Trivedi, Saptarshi Purkayastha, **Sinha, Priyanshu**, Alexandre Cadrin-Chenevert, Nabile Safdar, and Judy Wawira Gichoya. A DICOM Framework for Machine Learning and Processing Pipelines Against Real-time Radiology Images. *Journal of Digital Imaging*, 34(4):1005–1013, 2021

6. Ananth Bhimireddy, **Sinha, Priyanshu**, Bolu Oluwalade, Judy W Gichoya, and Saptarshi Purkayastha. Blood Glucose Level Prediction as Time-Series Modeling using Sequence-to-Sequence Neural Networks. In *CEUR workshop proceedings*, 2020

7. Saptarshi Purkayastha, Ananth R Bhimireddy, **Sinha, Priyanshu**, and Judy W Gichoya. Using ImageBERT to improve performance of multi-class Chest Xray classification. 2020

8. Imon Banerjee, **Sinha, Priyanshu**, Saptarshi Purkayastha, Nazanin Mashhaditafreshi, Amara Tariq, Jiwoong Jeong, Hari Trivedi, and Judy W Gichoya. Was there COVID-19 back in 2012? Challenge for AI in diagnosis with similar indications. *arXiv preprint arXiv:2006.13262*, 2020

9. **Sinha, Priyanshu**, Saptarshi Purkayastha, and Judy Gichoya. Full training versus fine tuning for radiology images concept detection task for the ImageCLEF 2019 challenge. In *CEUR workshop proceedings*, 2019

Open-Source Contributions

Google Summer of Code (GSoC) Mentor @ LibreHealth | *Python, ML, Unity, VR*   **May 2018 - September 2022**
- Mentored GSoC students to develop UI for Virtual-Reality application and integrate machine learning models for classification, detection, and segmentation of disease in a VR environment for immersive learning during medical procedures. This app includes features such as voice chat, live streaming and video playback.

Google Summer of Code (GSoC) Student @ LibreHealth | *Django, DBMS*   **May 2017 -August 2017**
- Refactored and optimized the LibreHealth EHR database by migrating code to Django, implementing ORM and object-oriented practices (including inheritance), and writing unit tests for various modules to enhance query performance.

Awards and Achievements

- Our team (of 3 members) won **1st** place in the **AT&T 5G Sports hackathon** in Fan Engagement category (held in Indianapolis, USA) where more than 25 teams participated.
- Our team (of 2 members) was selected in **top-50** in Hackerearth IndiaHacks hackathon (2016) out of 1500 teams globally.

Projects

Explainable M5 Accuracy Forecasting | *Python, Tensorflow, Time Series*   **December 2022**
- Created an **LSTM** model for predicting one-day and 28-day unit sales on the M5 Forecasting - Accuracy dataset, with explanations provided through counterfactual analysis using **what-if tool**.

Explainable AI in Radiology Image Analysis Pipeline | *Python, Tensorflow, Computer Vision*   **December 2022**
- Created an **explainable AI** prototype for radiology image analysis using **Densenet121** and **SHAP** values on NIH-14 chest x-ray dataset, to improve trust in black box AI models and assist radiologists in understanding predictions.