PUJA SAHA

Website | Google Scholar | LinkedIn | GitHub | Blog
psaha03@uoguelph.ca | Guelph, ON, Canada.

RESEARCH INTERESTS

Distributed Learning, Multimodal AI, Model Architecture, Machine Learning Systems and Optimization, Computer Aided Diagnosis, Computer Vision and Robotics.

EDUCATION

University of Guelph, Canada [Sept 2023 - Cont.]
M.A.Sc. in Computer Engineering CGPA: **3.93**/4.00
Courses: Deep Learning, Statistical Machine Learning, Big Data Analysis, Machine Learning for Sequences.
Academic Highlights:
- Average grade is 88%.
- Recipient of multiple prestigious scholarships (i.e., QE-II GSST)

Thesis Domain: Decentralized Machine Learning for Medical Image Analysis **Advisor:** Dr. Eranga Ukwatta

Bangladesh University of Professionals, Bangladesh [Feb 2016-Jan 2020]
Military Institute of Science and Technology (MIST)
B.Sc. in Biomedical Engineering (164 Credit) CGPA: **3.68**/4.00
Academic Highlights:
- CGPA of last four semesters is **3.86**/4.00.
- Enlisted on MIST Dean's Lists (CGPA >3.75) – 6 Semesters.
- Recipient of MIST Merit Scholarships (Academic Rank in Top 3) – 2 Semesters

Thesis: *Classification of Brain Diseases using Convolutional Neural Network.* **Advisor:** Dr. M J. Alam
The goal of this study was to develop a CNN from scratch that can learn effectively from nearly unprocessed and limited number of medical images (MRI and CT scan). The proposed CNN in this work, trained with 100 unique and unprocessed images for each class and identified hemorrhagic and tumorous scans against normal ones with 72% and 74% accuracy, respectively.

PROJECTS

Multiclass Semantic Segmentation | *Tech Stack: PyTorch, Torchvision, OpenCV, NumPy, matplotlib, tqdm, etc.*

A PyTorch-based solution for multiclass semantic segmentation on the KiTS19 dataset, specifically targeting kidney and tumor regions in abdominal CT scans. After thorough data analysis and preparation, advanced image preprocessing techniques and optimized training and testing workflows were applied, resulting in strong quantitative performance metrics: a Dice score of 91.03% for kidney segmentation and 62.82% for tumor segmentation. *GitHub Link*

Medical Chatbot | *Tech Stack: Pinecone, LangChain, Ragas, PyTorch, NumPy, Seaborn, tqdm etc.*

Employing language model with retrieval-augmented generation, trained on a reliable dataset from PubMed, our chatbot achieved 96.7% context precision, 95% context recall, 85% faithfulness, 73% answer relevancy, and 69.4% answer correctness in its responses. *GitHub Link*

Job Market Analysis and Salary Prediction | *Tech Stack: Apache Spark, Pandas, NumPy, Scikit-learn, Seaborn, etc.*

This big data-driven project involves exploratory data analysis and salary prediction for the global job market from 2021 to 2023, encompassing over 1.6 million records. Using PySpark for EDA, the project uncovered exciting insights and achieved salary predictions with an RMSE error of approximately $10K. *GitHub Link*

Human Pose Estimation | *Tech Stack: PyTorch, NumPy, Scikit-learn, TensorBoard, tqdm, Matplotlib, etc.*

Using a CNN-based regression technique to track human body posture by identifying 14 key joint coordinates, we achieved an impressive mean squared error (MSE) of 0.0878. GitHub Link

Predictive Modeling of Property Prices | *Tech Stack: Pandas, Scikit-learn, NumPy, Matplotlib etc.*

In this project, we performed regression using spectrum of machine learning models and ensemble techniques (e.g., Bagging and Boosting) to predict house prices. The Random Forest model achieved an impressive mean squared error (MSE) of 0.0012. GitHub Link

EMG-Biofeedback | *Tech Stack: Arduino IDE, MATLAB*

Built an interactive technology to observe muscles' progress to accomplish rehabilitation goals by exploiting muscles' electricity sensing capacity of EMG sensor-AD8232.

CONFERENCE PROCEEDINGS

1. *Saha P.,* Jhalak S.C., Mehrab A., Alam J. (2022) *Convolutional Neural Network to Classify Medical Images of Rare Brain Disorders*. International Conference on Healthcare Engineering (ICHE'22) in Malaysia, 23-25 Sept, 2022

In some cases, it is sometimes practically impossible to gather a large number of samples, especially for rare disorders (e.g., Multiple Sclerosis, Cerebral Palsy), limiting the use of traditional CNN requiring large number of images. The aim of this study was to develop a CNN that can learn from limited and nearly unprocessed samples to learn and classified unseen samples of the same condition with an accuracy of 85%.

2. *Saha P.,* Tasnim A., Omi O.A., Rahman T., Ashrafuzzaman M. (2020) *Photosweep: An Engineering Approach to Develop Cost Effective Sterilization System for Hospitals.* Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems, vol 3. Springer, Cham. International Conference on Emerging Trends in Engineering (ICETE 2019) in Hyderabad, India, 22-23 March, 2019.

This study represented an engineering approach for the development of a cost-effective sterilization device called "Photosweep": a dual-chamber sterilizing system. Whereas the first chamber uses solar heat to convert water into steam for sterilizing glassware, metallic tools, and accessories, the second chamber uses UV rays extracted from the sun to sterilize plastics (ventilation tubes, bio-bags, media solutions, etc.) that would usually melt at high pressure and temperature. The proposed model is applicable to hospitals having enough daylight and can reduce the operation costs for sterilization.

3. *Saha P*., Tuba M.A., Ahmed K.A., Ashrafuzzaman M. (2020) *Development of an Inexpensive Proficient Smart Walking Stick for Visually Impaired People*. Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems, vol 3. Springer, Cham. International Conference on Emerging Trends in Engineering (ICETE 2019) in Hyderabad, India, 22-23 March, 2019

This study represented the development of an assistive device for visually impaired people. It was designed and developed with the goal of keeping it simple and affordable for people of various economic backgrounds, which could also mitigate their disability and dependency and, hence, improve their quality of life.

TECHNICAL SKILLS

- **Frameworks and Tools**: PyTorch, Flower, PySpark, OpenCV, TensorFlow, LangChain, Ragas, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, NiBabel, Keras.
- **Programming Languages**: Python, MATLAB, C. **OS:** Linux, Mac, Windows.
- **Design and Simulation:** SOLIDWORKS, Ansys, OrCAD PSpice.
- **Technologies:** Cloud Computing (AWS, Compute Canada), Federated Learning, Git.

PROFESSIONAL DEVELOPMENT

Graduate Research Assistant | *AI-Enabled Medical Image Analysis Lab, CA* [Sept 23 - Cont.]

In this role, currently I am exploring kidney and tumor segmentation in CT scans using various versions of U-Net (such as Attention U-Net, Residual U-Net, and Recurrent Residual U-Net, etc.). However, my primary research goal is to focus on federated learning for medical image analysis.

Graduate Teaching Assistant | *School of Engineering, University of Guelph, CA* [Jan 24 - Cont.]
Courses: Medical Imaging Modalities (Fall-2024), Engineering Design (Winter-2024, Winter-2025).

Responsibilities included conducting laboratory sessions, providing mentorship and guidance to students, helping them grasp the course material effectively, supporting them in completing their lab tasks, invigilating exams, evaluating reports, exam papers and technical memos with detailed feedback, and working closely with faculty and fellow teaching assistants.

Intern Engineer | *Radiotherapy Department, Ahsania Mission Cancer and General Hospital, BD* [Dec 2018]

Responsibilities included collaborating with a multidisciplinary team of radiologists and radiotherapists to capture computed tomography scans of cancer patients, understanding their clinical conditions, analyzing patient data, conducting manual tumor segmentation, and assisting in the administration of radiotherapy sessions for patients with diverse cancer types and stages.

Certified Courses:

- 'MRI Fundamentals' offered by KAIST in Coursera.
- 'SOLIDWORKS Application in Biomedical Engineering' offered by MIST.
- 'Microcontroller [PIC-16F877A] and Robotics' offered by MIST.

STANDARDIZED TEST SCORES

Graduate Record Examination **(GRE)** [Sept 2022]
Overall: **308**; Quantitative: 159; Verbal: 149; AWA: 3.0

International English Language Testing System **(IELTS)** [Aug 2021]
Overall: **7.0** (CEFR C1), Reading: 7.5; Listening: 6.5; Writing: 6.0; Speaking: 7.5

CO-CURRICULAR ACTIVITIES

- Judge: Senior and Junior Project Competitions at the University of Guelph
- Oral Presentations: ICHE-2022 in Malaysia, ICETE-2019 in India.
- Poster Presentations: BME Festival-2018, ICMPROI-2018, Dhaka
- Project Presentations: Microcontroller-Based Project Fair, BME Festival-2018, Dhaka
- Club Memberships: Active Member of BME Einthoven Club, Robotics Club, MIST

AWARDS & ACHIEVEMENTS

- Queen Elizabeth-II Graduate Scholarship in Science and Technology ($15k). [May 2024- Apr 2025]
- Graduate Tuition Scholarships at the University of Guelph ($16k). [Sept 2023- Aug 2025]
- Research Fellowship at the University of Guelph ($12k). [Sept 2023- Aug 2025]
- Internal Growth Scholarship at the University of Guelph ($11k). [Sept 2023- Aug 2025]
- 5th Best Paper Award at IEEE International Conference on Healthcare Engineering-2022. [Sept 2022]
- MIST Silver Coin (Research). [Feb 2020]
- MIST Dean's List and Merit Scholarships. [2017-2019]
- Top Project Award and Best Informative Poster Award at BME Fest in MIST. [Oct 2018]
- 2nd Runners up: Inter-Department 'Microcontroller Based Project Fair' at MIST. [Jan 2018]
- 2nd Runners up: HULTPRIZE@MIST; Selected in Shanghai Regional Final. [Dec 2017]
- Bangladesh Education Board Merit Scholarships. [2010, 2013]