

# Puru Soni

Buffalo, NY | (716) 295-3948 | [purusoni@buffalo.edu](mailto:purusoni@buffalo.edu) | [purusoni@acm.org](mailto:purusoni@acm.org) | [github.com/purusoni](https://github.com/purusoni)

## RESEARCH INTERESTS

---

Multimodal AI, computer vision, natural language processing, and human-centered AI for healthcare and accessibility.

## EDUCATION

---

**University at Buffalo (UB), SUNY**

**B.S. Computer Engineering**

*University Honors College Scholar*

*Dean's List: Fall 2022, Spr 2023, Fall 2024, Fall 2025, Spr 2026 (5 semesters)*

*\$12,000 Annual Merit Scholarship*

Aug 2022 – May 2026

**Cum. GPA: 3.708 / 4.0**

### Relevant Coursework:

- **Mathematics:** Probability Theory (**A**), Differential Equations (**A**), Linear Algebra (**A-**), Honors Calculus 3 (**B**)
- **Computing:** Data Structures (**A-**), Computer Architecture (**A-**), Real-Time Embedded Systems (**A**), Hardware/Software Integrated Systems (**A**), Computer Vision & Image Processing (**A-**), Software Security (**A**), AI, Bias & Justice (**A**), Discrete Structures (**A**), Systems Programming (**B**), Digital Systems (**A**)

## EXPERIENCE

---

**AI Engineer** | *Auspex Medix, Buffalo, NY*

June 2026 – Present

Auspex Medix is a University at Buffalo spinout company founded by Dr. Wenyao Xu, developing AI-driven medical technologies for healthcare professionals and patients. The company is incubated at UB's Incubator @ Baird and has received funding from the National Institute on Aging (\$252,550 grant) and UB's Center of Excellence in Bioinformatics and Life Sciences.

Continuing work on commercializing research developed over 4 years in the ESC Group, including AI-powered diagnostic tools targeting accessibility and early detection across medical specialties.

## RESEARCH EXPERIENCE

---

**Lead Undergraduate Researcher** | *Embedded Sensing and Computing (ESC) Group, UB* Sep 2022 – May 2026 (4 years)

*Advisor: Dr. Wenyao Xu*

Conducted research across the entirety of my undergraduate education, contributing to four projects spanning medical AI, multimodal deep learning, and human-centered computing.

- **OralScan – Dental Diagnostics AI (Jul 2024 – May 2026, 2 years):** Designed and implemented a hybrid self-supervised learning framework combining contrastive and restorative objectives to pre-train a Swin Transformer backbone on unlabeled dental imagery, achieving a 94% pre-training loss reduction. Integrated the backbone with DINO-DETR, reaching mAP@.50 of 70.0%. Contributed to the OralScan mobile application (React Native), dataset curation with dental faculty, and a Community Engagement Studio (n=13 older adults) assessing usability and acceptability.
- **NoteMD – Medical NLP (May 2024 – May 2026, 1 year):** Engineered an end-to-end application for generating medical SOAP notes from clinical appointment audio. Built the full pipeline: WhisperX for transcription and diarization, LoRA-based PEFT fine-tuning of Mistral 7B, and a Gradio interface. The fine-tuned model outperformed GPT-4 by approximately 8% on ROUGE-1, 10% on ROUGE-2, and 6% on BLEURT. Conducted a human evaluation study with 12 healthcare professionals.
- **NeuroSpeech – Multimodal Speech Recognition (May 2023 – Jul 2024, 1 year):** Conducted a large-scale hyperparameter design space exploration on the university HPC cluster, executing over 9,420 experiments comparing GRU, ConvLSTM, and TD-CNN EEG encoder architectures. The TD-CNN encoder achieved an F1-score of 0.986 in clean conditions and 0.92 under high noise (SNR 0.5), compared to Whisper dropping to 0.46 under the same conditions.
- **EEG Vision – Neural Decoding (Sep 2022 – May 2023, 1 year):** Designed and trained a VGG-based CNN in PyTorch to classify visual stimuli from raw EEG data, achieving 95% classification accuracy. Served as an introduction to EEG signal processing and neural decoding.
- **Infrastructure & Documentation:** Managed lab server including root-level Linux administration and CUDA maintenance. Authored the [ESC Group CCR Handbook](#), used to onboard PhD students to HPC, SLURM, and EasyBuild workflows.

- **Last Mile Delivery Drone Project:** Worked in a three-person team to build a Python/Kivy ground control application for FPV drone flight, integrating OpenCV for dual-camera feeds and Pymavlink for telemetry. Resolved a real-time video lag bottleneck by implementing a multiprocessing architecture separating video rendering from data acquisition. Conducted flight experiments at UB's SOAR outdoor facility.
- **Crazyflie Swarm Project:** Worked with nano-quadcopter swarm robotics using the Crazyflie platform. Learned ROS and used the VICON Motion Capture Lab and the Crazyswarm2 library to coordinate and experiment with multiple drones simultaneously. Recognized as a completed project by UB's Office for Undergraduate Research.

## PUBLICATIONS & MANUSCRIPTS

---

Summary: 3 peer-reviewed publications. Currently leading 3 additional first-author manuscripts.

1. Das, A., **Soni, P.**, Zhao, H., Huang, M.-C., & Xu, W. (2025). Optimizing Deep Neural Networks for EEG-Based Speech Recognition: A Multimodal Approach to Assistive Communication. *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, 29(12), 8735–8742. [doi:10.1109/JBHI.2025.3618998](https://doi.org/10.1109/JBHI.2025.3618998)  
**Q1 journal, IF 6.8 (JCR 2025).** Contribution: Conducted a large-scale HPC design space exploration of 9,420+ experiments to identify the optimal EEG encoder. The resulting multimodal model maintained a 0.92 F1-score under extreme noise (SNR 0.5), where the audio-only Whisper baseline dropped to 0.46.
2. Das, A., **Soni, P.**, Huang, M.-C., Lin, F., & Xu, W. (2024). Multimodal speech recognition using EEG and audio signals: A novel approach for enhancing ASR systems *Elsevier Smart Health*, 32, 100477. [doi:10.1016/j.smhl.2024.100477](https://doi.org/10.1016/j.smhl.2024.100477)  
**Q2 journal (SJR 2024).** Contribution: Led the end-to-end software implementation of the initial multimodal deep learning framework. Established the baseline demonstrating that EEG-audio fusion outperforms single-modality models in degraded conditions.
3. Xiao, H., KrishhnaKumar, P., Pothuri, J.P.K.V., **Soni, P.**, Butcher, E., & Chowdhury, S. (2024). An Open-Source Hardware/Software Architecture and Supporting Simulation Environment to Perform Human FPV Flight Demonstrations for Unmanned Aerial Vehicle Autonomy. *AIAA Aviation Forum*. [doi:10.2514/6.2024-4458](https://doi.org/10.2514/6.2024-4458)  
**Flagship annual conference of AIAA, the largest aerospace professional society in the world.** Contribution: Built collaboratively the ground control station GUI and resolved a real-time video lag issue through a multiprocessing architecture.
1. **Soni, P.**, Dandapat, K., Gherardi, A., Bo, W., Li, R., & Xu, W. OralScan, an AI-Powered Mobile Tool for Geriatric Oral Healthcare: A Formative Usability and Acceptability Study. (In Preparation).  
*First Author.* Led the project end-to-end: mobile app development (React Native), dental pathology dataset curation, on-device YOLO deployment for image quality assessment, and a Community Engagement Studio (n=13 older adults).
2. **Soni, P.**, Dandapat, K., Gherardi, A., Bo, W., & Xu, W. A Hybrid Self-Supervised Learning Framework for AI-Powered Dental Diagnostics. (In Preparation).  
*First Author.* Designed a hybrid SSL framework combining contrastive and restorative objectives to pre-train a Swin Transformer backbone, achieving a 94% pre-training loss reduction. Integrated with DINO-DETR, reaching 70.0% mAP@.50.
3. **Soni, P.**, Das, A., & Xu, W. NoteMD: An End-to-End Application for Automated Medical Note Generation using Fine-Tuned LLMs. (In Preparation).  
*First Author.* Fine-tuned Mistral 7B with LoRA/PEFT, outperforming GPT-4 by approximately 8% on ROUGE-1, 10% on ROUGE-2, and 6% on BLEURT. Validated with a human evaluation study of 12 healthcare professionals.

## PATENTS

---

Das, A., **Soni, P.**, & Xu, W. *Optimizing Deep Neural Networks for EEG-Based Speech Recognition: A Multimodal Approach to Assistive Communication*. U.S. Provisional Patent Application No. 63/984,828, filed February 17, 2026 (University at Buffalo / SUNY).

## SELECTED PROJECTS

---

### AutoTA: Socratic LLM Teaching Assistant

Lead Software Engineer

Nov 2023

2nd Place (\$1,000) – UBHacking 2023

- Built a RAG-based chatbot for Socratic tutoring using an open-source Mixtral backend, with system-level constraints to prevent direct answer generation. Deployed with Docker and Nginx.

## SELECTED PRESENTATIONS

---

- **Poster:** “OralScan: AI-enabled Accessible Dental Care for Older Adults” – UB AI & Data Science Symposium (Mar 2025); UB Summer Research Symposium (Jul 2025)
- **Oral:** “NoteMD: Automated Medical Note Generation” – CTSI Clinical Research Group, UB (Apr 2024)
- **Oral:** “Multimodal Speech Recognition using EEG and Audio Signals” – SUNY Undergraduate Research Conference (Apr 2024)

- **Poster:** “Multimodal Speech Recognition using EEG and Audio Signals” – UB SEAS Summer Research Symposium (Jul 2024)
- **Poster:** “Human Vision Inference from Brainwave Processing” – UB Celebration of Academic Excellence (Apr 2023)

## TECHNICAL SKILLS

---

- **AI & Machine Learning:** PyTorch, TensorFlow, Hugging Face Transformers, PEFT/LoRA, OpenCV, Scikit-learn
- **Computer Vision:** YOLO, SAM, object detection, image segmentation
- **NLP & LLMs:** LLM fine-tuning, RAG, WhisperX, prompt engineering
- **Signal Processing:** EEG analysis, FFT, cross-covariance matrices, time-frequency domain analysis, MFCCs
- **HPC & Infrastructure:** SLURM, Linux administration (root), Docker, DevContainers, Git
- **Hardware:** Arduino, Raspberry Pi, TI Microcontrollers, Pixhawk, ROS2, Crazyflie
- **Languages:** Python, C, JavaScript, Assembly, SQL

## AWARDS, SERVICE & LANGUAGES

---

- **UB Health Futures Challenge, 1st Place (\$1,000)** – OralScan (May 2026)
- **UBHacking 2023, 2nd Place (\$1,000)** – AutoTA (Nov 2023)
- **AP Scholar with Honor** (June 2022)
- **ICAS Distinction**, Mathematics and Science (2018–2020)
- **Peer Reviewer:** *ACM Transactions on Computing for Healthcare (TOCH)*, 2025
- **Member:** Association for Computing Machinery (ACM)
- **Certification:** Deep Learning Specialization, DeepLearning.AI – avg. grade 99% across 5 courses (2022). [View certificate.](#)
- **Certification:** Machine Learning, Stanford University – 99.16% (2021). [View certificate.](#)