Khandokar Md. Nayem

Luddy Hall
700 N Woodlawn Ave, #3061P
Bloomington, IN 47408

+1 (812) 369-6358
knayem@iu.edu
nayem.github.io

RESEARCH SUMMARY

I am Khandokar Md. Nayem, currently pursuing a Ph.D. in Computer Science at Indiana University, specializing in Deep Learning, Speech/Audio Machine Learning (ML), Large Language Models (LLM), and Natural Language Processing (NLP). My academic journey has provided me with a strong foundation in the intersection of speech and NLP, and my contributions are reflected in top-tier publications, including TASLP, ICASSP, INTERSPEECH, and MLSP. In addition to my academic pursuits, I've gained valuable industry experience through projects at companies such as Amazon, Microsoft Research, and BOSE. Moreover, I've had the privilege of sharing my knowledge and mentoring students both at Indiana University and during my time as a lecturer in Bangladesh.

EDUCATION

Ph.D. in Computer Science December 2023
Indiana University, Bloomington, IN, USA
Committee: David J. Crandall (co-chair), Donald S. Willamson (co-chair), Minje Kim, Christopher S. Raphael, Zoran Tiganj
Dissertation: "Towards Realistic Speech Enhancement: A Deep Learning Framework Leveraging Speech Correlations, Spectral Language Models, and Perceptual Evaluation"

M.Sc. in Computer Science December 2019
Indiana University, Bloomington, IN, USA

B.Sc. in Computer Science & Engineering (CSE) July 2014
Bangladesh University of Engineering & Technology (BUET), Dhaka, Bangladesh

TECHNICAL EXPERTISE

• Deep Learning • Machine Learning • Large Language Model (LLM) • Natural Language Processing (NLP) • Speech Enhancement, Translation & Recognition • Data Science • Data Visualization • Classification & Regression • Knowledge Distillation • Machine Learning Models (Bayesian Network, HMM, GMM, Clustering, Decision Tree, Ensemble Methods) • Deep Neural Network (DNN) • Long-Short Term Memory (LSTM) • Convolutional Neural Network (CNN) • Transformers • Deep Learning Libraries (PyTorch, TensorFlow, Keras, Fairseq, HuggingFace) • NLP Libraries (NLTK, Scikit-Learn) • Human Survey Platform (Qualtrics, Amazon MTurk) • Computer Vision & Graphics Libraries (CImg, OpenCV, OpenGL) • Docker • Git • AWS • Sagemaker • Python • Matlab • C\C++ • Java • Python Libraries (Pandas, Numpy, Matplotlib, Seaborn) • HTML, CSS & Javascript • PHP • SQL • R • Shell

PUBLICATIONS

Peer-reviewed Journal:

- **Khandokar Md. Nayem**, Donald S. Williamson, "Attention-based Speech Enhancement Using Human Quality Perception Modelling", in Proceedings of journal IEEE/ACM Transactions on Audio Speech and Language Processing (TASLP) 2024. (paper)
  Journal Impact factor: 5.4, SCImago Journal Rank (SJR): 1.35

Peer-reviewed Conference:

- **Khandokar Md. Nayem**, Ran Xue, Ching-Yun Chang, Akshaya Vishnu Kudlu Shanbhogue, "Knowledge Distillation on Joint Task End-to-End Speech Translation", in Proceedings of the Annual Conference of the International Speech Communication Association (ISCA), INTERSPEECH 2023. (paper, poster)
  Conference Impact factor: 3.47, SCImago Journal Rank (SJR): 0.689
- **Khandokar Md. Nayem**, Donald S. Williamson, "Incorporating Embedding Vectors from a Human Mean-Opinion Score Prediction Model for Monaural Speech Enhancement", in Proceedings of the Annual Conference of the International Speech Communication Association (ISCA), INTERSPEECH 2021. (paper, slides, video)
  Conference Impact factor: 3.47, SCImago Journal Rank (SJR): 0.689

- **Khandokar Md. Nayem**, Donald S. Williamson, "Towards an ASR approach using Acoustic and Language Models for Speech Enhancement", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2021. (paper, poster, slides, video)
  Conference Impact factor: 3.59, SCImago Journal Rank (SJR): 0.997

- **Khandokar Md. Nayem**, Donald S. Williamson, "Monaural speech enhancement using intra-spectral recurrent layers in the magnitude and phase responses", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020. (paper, slides, video)
  Conference Impact factor: 3.59, SCImago Journal Rank (SJR): 0.997

- **Khandokar Md. Nayem**, Donald S. Williamson, "Incorporating Intra-Spectral Dependencies with a Recurrent Output Layer for Improved Speech Enhancement", in Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP) 2019. (paper, poster)
  Conference Impact factor: 1.81, SCImago Journal Rank (SJR): 0.401

Poster:

- Shujon Naha, **Khandokar Md. Nayem**, Md. Lisul Islam, "RSGAN: Recurrent Stacked Generative Adversarial Network for Conditional Video Generation", presented at IU computer vision project showcase, 2017. (paper, poster)
- **Khandokar Md. Nayem**, Mir Toornaw Islam, Md. Monirul Islam, "Handwritten Writer Independent Bangla Character Recognition", presented at undergrad thesis at BUET, 2014. (paper, poster)

Archive:

- **Khandokar Md. Nayem**, Donald S. Williamson, "Attention-based Speech Enhancement Using Human Quality Perception Modelling", in arXiv, 2023. (paper)

RESEARCH EXPERIENCE

Indiana University, Bloomington, IN, USA Summer 2018 - Now
Research Assistant, ASPIRE research lab

- Developed a novel speech enhancement model that aims to maximize human perceptual satisfaction of speech after cleaning real-world noise from the captured audio. We incorporate perceptually important features learned from a separate perception rating prediction model in achieving this task. (TASLP-2024, arXiv, INTERSPEECH-2021)
- Proposed & implemented a quantized speech prediction model that classifies speech spectra into a corresponding quantized class and applies a language-style model to generate more realistic speech. Acceptable quantization level was determined by listener subjective study conducted on Amazon MTurk, designed using Qualtrics. (ICASSP-2021)
- Formulated a new type of recurrent output layer in the context of deep learning architecture that shows significant improvement in speech enhancement. This new design utilizes the internal relations along the frequency axis between speech frequency points. (ICASSP-2020, MLSP-2019)
- Engineered a deep architecture named Recurrent Stacked Generative Adversarial Network (RSGAN) which generates video clips given different types of signal like human's brain fMRI signal or sentence description of the video clips. (IU-VISION-2017)
- Created data visualization techniques for monitoring daily physical and sleep activity data from pregnant women using wearable devices, contributing to the development of a smart system aimed at identifying potential risks related to gestational complications such as gestational diabetes and pre-eclampsia. This work was conducted as part of the NSF Proactive Health Informatics (PHI) project.
- Researched on speech emotion detection systems that analyze speech to monitor human emotions, a valuable cue for tackling sensitive emotional situations and maintaining a healthy conversation.
- Developed an automatic classification of rhetorical questions with stress detection on our own collected dataset using Recurrent Neural Network (RNN) and Convolutional Recurrent Neural Network (CRNN) models.

Amazon, Seattle, WA, USA Summer 2023
Applied Scientist II Intern, Seller Partner Services

- Conducting research on the application of the Large Language Model (LLM) for class labeling on closed taxonomy utilizing product descriptions, while also generating chain-of-reasoning explanations for improved overall comprehension.

Amazon, Cambridge, MA, USA Fall 2022
Applied Scientist II Intern, Alexa AI

- Researched the development of a real-time, end-to-end compressed multi-lingual speech translation system. Investigated the use of Large Language Models (LLMs) and applied knowledge distillation approach to transfer their performance to smaller models with 50% and 75% fewer parameters. (INTERSPEECH 2023)

Microsoft Research, Redmond, WA, USA Summer 2022
Research Intern, Audio and Acoustics Research Group

- Focused on analyzing and improving the performance of speech enhancement algorithms to generate high-fidelity (Hi-Fi) speech by removing distortions and extending speech bandwidth. Applied real-time deep learning models with various data augmentation techniques to recover codec and clipping distortions, and performed deep noise suppression.

BOSE Corporation, Boston, MA, USA Summer 2020
Machine Learning/Neural Signal Processing Intern

- Researched on enhancing speech in remote microphone applications by removing self-speech in order to provide better quality sound with low latency to hearing aids and voice-assistive wearable devices. Utilized a deep learning architecture with speaker-dependent features for speaker identification, to ensure real-time operation.

REVE Systems, Dhaka, Bangladesh July 2014 - January 2015
Jr. Software Engineer

- Programmed controller of 'media-gateway' server to facilitate both phone calls and faxes between the telephone network and VoIP network or another telephone network. Also designed a front-end panel of controllers for VoIP administrators and customers.

TEACHING EXPERIENCE

Indiana University, Bloomington, IN, USA Fall 2016 - Spring 2023
Teaching Assistant, Research Assistant

- Mentored 2 undergrad students Muhammad Asghar (currently works as Software Engineer at L3Harris) and Daniel Quintans (currently works as Software Development Engineer Intern at Amazon) as a part of the **NSF Research Experience for Undergraduates (REU) program**.
- Mentored **summer visitor undergrad** student Chitrank Gupta (currently works at the University of Texas, Austin) in audio focused machine learning research project.
- Designed an online course curriculum on the "Introduction of Probability" for the Data Science Onramp program.
- Taught Computer Science course: Introductory Python Programming (CSCI-A-201) to freshman-level undergrad class with a class size of 35 students.
- Taught Computer Science course: Discrete Mathematics (CSCI-C-241) to sophomore-level undergrad class with an average class size of 40 students.
- Taught Data Science course: Data Representation (DSCI-D-321) to junior-level undergrad class with a class size of 50 students, and coordinated with other TAs as Lead TA.
- Taught Computer Science course: Machine Learning (CSCI-B-455) to senior-level undergrad class with a class size of 25 students.

United International University (UIU), Dhaka, Bangladesh February 2015 - August 2016
Lecturer, Department of Computer Science & Engineering

- Evaluated and judged senior-level undergrad final research thesis projects.
- Coached undergrad students for **programming contest** at south-asia region of International Collegiate Programming Contest (ICPC), 2016.
- Designed two computer science system design courses: Digital Logic Design (CSE-1325) and Digital Logic Design Laboratory (CSE-1326) for ABET accreditation, being as a member of the committee.
- Introduced and designed curriculum for a senior-level undergrad Computer Science course: Pattern Recognition (CSE-415).

- Taught over 600 freshman-level undergrads with an average class size of 50 students in Computer Science courses: Structured Programming Language (CSE-1111), Structured Programming Language Laboratory (CSE-1112), Object Oriented Programming (CSE-1115), Object Oriented Programming Laboratory (CSE-1116), Digital Logic Design (CSE-1325), and Digital Logic Design Laboratory (CSE-1326).
- Taught over 400 sophomore-level undergrads with an average class size of 38 students in Computer Science course: Data Structure and Algorithms I (CSE-2215).
- Taught over 250 junior-level undergrads with an average class size of 35 students on Computer Science courses: Digital Logic Design (CSE-1325), Digital Logic Design Laboratory (CSE-1326), Artificial Intelligence (CSE-3811), Artificial Intelligence Laboratory (CSE-3812).
- Taught over 100 senior-level undergrads with an average class size of 50 students on Computer Science course: Pattern Recognition (CSE-415).

ARTICLE, TALKS & PRESENTATION

- Presented "Fine-tuning LLM for Product Classification and Explanation" in **Amazon**, 2023.
- Article published in **Amazon Science** on "Knowledge Distillation on Joint Task End-to-End Speech Translation", 2022. (link)
- Presented "Knowledge Distillation on Joint Task End-to-End Speech Translation" in **Amazon**, 2022.
- Presented "Unified Speech Enhancement Approach for Speech Degradations and Noise Suppression" in **Microsoft Research**, 2022.
- Presented "Incorporating Embedding Vectors from a Human Mean-Opinion Score Prediction Model for Monaural Speech Enhancement" in conference **INTERSPEECH 2021**. (video)
- Presented "Towards an ASR approach using Acoustic and Language Models for Speech Enhancement" in conference **ICASSP 2021**. (video)
- Presented "Speech Enhancement" in **BOSE Corporation**, 2020.
- Presented "Monaural speech enhancement using intra-spectral recurrent layers in the magnitude and phase responses" in conference **ICASSP 2020**. (video)
- Talk on "Speech Emotion Recognition" as Guest Lecturer in course **CSCI-B 659**, 2020.
- Talk on "Machine Learning & Deep Learning" as Guest Lecturer in **NSF Research Experience for Undergraduates (REU) program**, 2019.
- Presented "Improved Speech Enhancement by Incorporating Intra-Spectral Dependencies" in workshop **Midwest Music and Audio Day (MMAD)**, 2019.
- Presented "Incorporating Intra-Spectral Dependencies with a Recurrent Output Layer for Improved Speech Enhancement" in conference **MLSP 2019**.

AWARDS & SERVICES

- **Reviewer** for International Speech Communication Association (ISCA) at INTERSPEECH 2023 in Dublin, Ireland.
- **Student Volunteer** responsible for the managing virtual platform at ISCA INTERSPEECH 2021 in Brno, Czech Republic.
- Recipient of a **Travel grant** from Indiana University to attend IEEE ICASSP 2020 in Barcelona, Spain.
- **Panelist** at the career development session titled "Career Choice: Academia and Industry" organized by **NSF Research Experience for Undergraduates (REU) program** in 2019.
- Recipient of a **Travel grant** to attend at the Annual Conference of ISCA INTERSPEECH 2019 in Graz, Austria.
- **Student Volunteer** involved in organizing conference at the Annual Conference of ISCA INTERSPEECH 2019 in Graz, Austria.
- **President** of the Bangladesh Student Association (BDSA) at Indiana University, Bloomington, IN, overseeing a 50+ member student organization in 2020. Also, served as **Vice-President**, **General Secretary**, and **Treasurer** of BDSA consecutively from 2017 to 2019.
- Attained the rank of **Brown belt** in Taekwondo and served as a judge for **Yellow belt** students at the Indiana University Taekwondo club from 2017 to 2020.