Nihar Ranjan **Sahoo**

PhD Research Scholar, CFILT IITB

Indian Institute of Technology, Bombay, India

niharsahooigit@gmail.com | sahoonihar.github.io/ | sahoonihar | nihar-nrs26 | @nihar_runny

Interests

My interest lies in detection and mitigation of social biases/toxicity in text, ethics and fairness in ML, NLP. I am also interested in areas like Reinforcement Learning, Computational Cognitive Neuroscience, Robotics.

Education

Indian Institute of Technology, Bombay

Mumbai, India

PH.D, COMPUTER SCIENCE AND ENGINEERING, **CGPA: 9.36/10**

Ongoing

- **Advisor:** Prof. Pushpak Bhattacharyyaa
- **Seminar:** Social Bias in NLP: Detection and Mitigation
- Selected Coursework: Deep Learning for NLP, Automatic Speech Recognition, Organisation of web information, Probability and Random Process

Indian Institute of Science

Bengaluru, India

MASTER IN TECHNOLOGY, COMPUTER SCIENCE AND AUTOMATION

Aug. 2017 - June. 2019

- **Advisor:** Sridharan Devarajan
- Selected Coursework: Pattern Recognition and Neural Networks, Machine Learning for Signal Processing, Natural Language Processing, Probability and Statistics, Linear Algebra, Design and Analysis of Algorithms

Indira Gandhi Institute of Technology

Sarang, Odisha

BACHELOR IN TECHNOLOGY, COMPUTER SCIENCE AND ENGINEERING

July. 2012 - June. 2016

Publication

Hollywood Identity Bias Dataset: A Context Oriented Bias Analysis of Movie Dialogues

Poster Presentation

SANDHYA SINGH*, PRAPTI ROY*, **NIHAR RANJAN SAHOO***, NITEESH MALLELA*, HIMANSHU GUPTA*, PUSHPAK BHATTACHARYYA

LREC, 2022

Few-Shot Domain Adaptation for Low Light RAW Image Enhancement

Best Student Paper (Runner up) Award [BMVC Site]

K RAM PRABHAKAR, VISHAL VINOD*, **NIHAR RANJAN SAHOO***, AND R VENKATESH BABU

BMVC, 2021 (Oral)

Work Experience

Indian Institute of Technology, Bombay

Mumbai, India

TEACHING ASSISTANT, DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING

Jan. 2022 - May. 2022

- Received Excellence in TA award.

Indian Institute of Technology, Bombay

Mumbai, India

TEACHING ASSISTANT, ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Aug. 2021 - Dec. 2021

- Received Excellence in TA award.

Video Analytics Lab

Bangalore, India

RESEARCH ASSISTANT, CDS, IISc

Feb. 2020 - Dec. 2020

- Image enhancement in RAW domain.
- We have paired(low exposure and high exposure) abundant data in source domain(eg : Sony) and few (1-5) paired data in target domain(eg : Nikon).
- Using CNN we need to learn the way to enhance low exposure image to get high exposure image. As it is hard to collect paired abundant data for every camera, we are using few shot domain adaptation technique to achieve our goal.

Citrix R&D India Pvt. Ltd *Bangalore, India*

Software Engineer 2 *Jul. 2019 - Feb. 2020*

M.Tech Thesis

Alzheimer's Disease: Understanding Mechanisms for Early diagnosis and treatment using functional MRI of brain

Advisor: Sridharan Devarajan

Indian Institute of Science, Bengaluru *Jul. 2019*

- Longitudinal study on the Indian population to determine the early bio-markers of Alzheimer's disease using multimodal approaches.
- Psychophysical assessments has been done for behavioral task using 2-ADC model.
- The aim is to predict the psychophysical parameters (human behaviour) using neuroimaging (rest fMRI, task fMRI and DTI) data of the population.

Projects

Automatic Story Generation

Advisor: Pushpak Bhattacharyyaa

Indian Institute of Technology, Bombay *Dec. 2021*

- The aim of the project is to generate an automatic story from a given prompt with the help of neural techniques.
- It is a lot challenging to generate long, coherent stories using state of the art NLP techniques. Story generation, in general, suffers from fluency, topic drift problems.
- In the project, we used the GPT2 model to train the story generation task using *WRITINGPROMPTS* dataset.

ParConvE: Parrallel Convolutional 2D Knowledge Graph Embeddings for Link Prediction at Scale

Advisor: Partha Pratim Talukdar

Indian Institute of Science, Bengaluru *Dec. 2018*

- The aim of knowledge graphs is to gather knowledge about the world and provide a structured representation of this knowledge.
- Current knowledge graphs are far from complete. To address the incompleteness of the knowledge graphs, link prediction approaches have been developed which make probabilistic predictions about new links in a knowledge graph given the existing links.
- Link prediction is a key research direction within this area. In this project, we will focus on link prediction at large scale using ParConvE. In this work we introduce ParConvE, a multi-layer convolutional network model for link prediction.
- WN18 and FB1M data were used for model evaluations.

Decoding Music from the Dynamic Processing of Musical Features in the Brain

Advisor: Sriram Ganapathy

Indian Institute of Science, Bengaluru *Dec. 2018*

- Many studies are able to decode the visual stimulus by measuring the brain activity but none seems to decode the auditory stimuli from the brain activity. Here we propose to decode the music from the activity of the brain measured using fMRI.
- Six musical features (three musical stimuli), representing low-level (timbre) and high-level (rhythm and tonality) aspects of music perception, were computed from the acoustic signals, and classification into one of the three musical stimuli will be done on the parcellated fMRI time series data.

Sentiment Analysis of Twitter Data

Advisor: Partha Pratim Talukdar

Indian Institute of Science, Bengaluru *Apr. 2018*

- Given a message also know as "Tweet",classify whether the tweet is of positive or negative sentiment. For tweets conveying both a positive and negative sentiment, the sentiment that is stronger should be chosen.
- Different data preprocessing techniques was used for good representation of data.
- Different features and machine learning models were used to determine the best combination for sentiment analysis.
- SemEval-2016 and SAIL Codemixed-2017 data were used for model evaluations.

Flow Visualization using Line Integral Convolution Technique

Advisor: Vijay Natarajan

Indian Institute of Science, Bengaluru *Nov. 2017*

- The aim is to develop a model to visualize ocean current flow around the globe using Line Integral Convolution.
- First Coastline data was extracted in the form of line strips in (latitude, longitude) pair and each pair was draw on earth's surface using OpenGL.
- Finally, ocean current velocity data was extracted along with white noise texture to find out the pixel values at different ocean coordinates using LIC module and render pixel values.
- Technologies Used: OpenGL / C++

Skills

- Tools: Numpy, Scipy, sklearn, matplotlib, pandas, nltk, tensorflow, pytorch, huggingface, keras, fslview
- Experties: Machine Learning, NLP, Deep Learning, Cognitive Neuroscience
- Programming Languages: C, C++, Core Java, Python

Roles & Achievements

Achievements

2022 **Fellowship**, Received Prime Minister's Research Fellowship

2022 **Silver Medal**, 10^{th} Inter IIT Tech Meet

2018 **Winner**, Intel Campus Day, IISc Bengaluru

2017 **All India Rank - 9**, Graduate Aptitude Test in Engineering(GATE), Computer Science

2017 **Winner**, Codechef Certification Examination(CCDSAP)

2015 **First**, Hour of Code, Horizon, IGIT

2007 **Rank 40**, Mathematics Olympiad by Shikhya Vikash Samiti, Odisha

Roles

2018-19 **Placement coordinator**, Indian Institute of Science, Bengaluru

2019 **Event Coordinator**, Data Science Hackathon, IISc Open Day'19

2018 **Event Coordinator**, Hackathon, CSA Undergraduate Summer School'18

2016 **Event Coordinator**, Hour of Code,Horizon'16, IGIT