Hiren Madhu

✉ hirenmadhu@iisc.ac.in
https://hirenmadhu.github.io/
LinkedIn: Hiren Madhu
Google Scholar: Hiren Madhu
GitHub: HirenMadhu

Education

Aug 2024	**Yale University**, *Computer Science*, Ph.D., Advised by Prof. Smita Krishnaswamy and Prof. Rex Ying.
Aug 2021–Jun 2023	**Indian Institute of Science, Bangalore**, *Artificial Intelligence*, Master of Technology, Advised by Prof. Sundeep Chepuri. GPA – 8.1/10
Aug 2017–Jun 2021	**LDRP-ITR, Gandhinagar**, *Computer Engineering*, Bachelor of Engineering. GPA – 8.4/10

Research Experiences

July 2023 – Jun 2024

ECE Predoc Fellow, INDIAN INSTITUTE OF SCIENCE, BENGALURU.
- My work mainly focused on weakly-supervised representation learning techniques for topological data structures, specifically for simplicial complexes.
- Currently, I am working on developing scattering transforms for simplicial complexes that extend geometric scattering techniques for simplicial complex data.

Feb 2021–Jun 2021

Research Intern, INDIAN INSTITUTE OF MANAGEMENT, AHMEDABAD.
- Performed deep, bottom-up fundamental analysis of Amul Hits by scrapping user interaction from social media platforms like Twitter, Facebook, LinkedIn, and Instagram to scrutinize their effects.
- Further, collaborated with senior analysts to build a customized independent web scarping agent to harvest accurate and complete daily/monthly data from the Weather Underground website.
- All work was done under the mentorship of Dr. Hyokjin Kwak, Prof. of Marketing at the Indian Institute of Management, Ahmedabad.

Conference and Journal Publications

* - Equal contribution

Unsupervised Parameter-free Simplicial Representation Learning with Scattering Transforms,
H. Madhu*, S. Gurugubelli*, SP Chepuri,
2024, ICML 2024.
https://icml.cc/virtual/2024/poster/32736

TopoSRL: **Topology Preserving Self-Supervised Simplicial Representation Learning**,
H. Madhu, SP Chepuri,
2023, NeurIPS 2023.
https://neurips.cc/virtual/2023/poster/72124

Detecting Offensive Speech in Conversational Code-Mixed Dialogue on Social Media: A Contextual Dataset and Benchmark Experiments [Site],
H. Madhu, S. Satapara, S. Modha, T. Mandl, P. Majumder,
Nov 2022, Accepted at Expert Systems with Application.
https://www.sciencedirect.com/science/article/pii/S0957417422023600

Shared Task Overview Papers

Overview of the HASOC Subtrack at FIRE 2022: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages,
Shrey Satapara, Prasenjit Majumder, Thomas Mandl, Sandip Modha, **Hiren Madhu**, *Tharindu Ranasinghe, Marcos Zampieri, Kai North, Damith Premasiri*,
2022, FIRE 2022: Forum for Information Retrieval Evaluation
`https://dl.acm.org/doi/abs/10.1145/3574318.3574326`.

Overview of the HASOC Subtrack at FIRE 2022: Identification of Conversational Hate-Speech in Hindi-English Code-Mixed and German Language,
Sandip Modha, Thomas Mandl, Prasenjit Majumder, Shrey Satapara, Tithi Patel, **Hiren Madhu**,
2022, FIRE 2022: Forum for Information Retrieval Evaluation
`https://ceur-ws.org/Vol-3395/T7-1.pdf`.

Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages and Conversational Hate Speech,
S. Modha, T. Mandl, G. Shahi, **H. Madhu**, *S. Satapara, T. Ranasinghe, M. Zampieri*,
2021, FIRE 2021: Forum for Information Retrieval Evaluation
`https://doi.org/10.1145/3503162.3503176`.

Overview of the HASOC Subtrack at FIRE 2021: Conversational Hate Speech Detection in Code-mixed language,
Shrey Satapara, Sandip Modha, Thomas Mandl, **Hiren Madhu**, *Prasenjit Majumder*,
2021, Working Notes of FIRE, 2021
`http://ceur-ws.org/Vol-3159/T1-2.pdf`.

Shared task and workshop papers

Analysis On Identification Of Hate Speech In Indo-European Languages With Fine-Tuned Transformers,
H. Madhu, *S. Satapara, H. Rathod*,
2020, FIRE '20, Forum for Information Retrieval Evaluation, December 16–20, 2020, Hyderabad, India
`http://ceur-ws.org/Vol-2826/T2-7.pdf`.

Projects

May 2022 – June 2023 **Weakly supervised topological representational learning for Simplicial Complexes**,
Master's thesis, Advised by: Prof. Sundeep Chepuri.
- There are publicly available large simplicial complexes that contain higher-order relations between a set of entities. These simplicial complexes can be used in many downstream applications, such as trajectory prediction, simplicial closure, and graph classification, to name a few.
- But the challenge is that learning representations and making inferences from such large simplicial complexes would require many costly quality annotations.
- As a part of my master's thesis, I have developed two methods for unsupervised and self-supervised simplicial representation learning.

Mar 2022 – April 2022 **Multi Modal Fake news detection**.
- This is a course project on Advanced Image Processing course I credited in my Master's.
- The project's main idea was to exploit the dissimilarity between the images in a news article and the news text. The main contribution of our work was to fuse the multimodal features. I trained a Siamese network that extracts images and text features such that the dissimilarity between cross-modal features is higher when the news is fake. [Slides]

Nov 2021 – Dec 2022 **Effects of segmentation on Detection of Casting Defects.** .
- This is a course project on Digital Image Processing course I credited to my Master's.
- The project's main idea was to analyze the effects of segmentation on the classification of a casting defect dataset. We first segmented the images using Otsu's binarization and watershed algorithm. Then we analyzed how segmentation affects the classification performance under conditions such as shorter training and smaller dataset.
- In the end, we concluded that segmentation helps the model learn the patterns even with a smaller dataset and shorter training. [Slides]

Feb 2021 – May 2021	**GAN**-**BERT**. ○ This is a major project I worked on during the final semester of my bachelor's. ○ The project's main idea was to explore semi-supervised techniques for text classification. We extended our previous work done for HASOC 2021 and used the same dataset for our analysis here. We replicated experiments done in a paper titled GAN-BERT, and we were able to achieve better performance compared to all the systems submitted to HASOC 2021. ○ `https://aclanthology.org/2020.acl-main.191.pdf`
Nov 2020 – Dec 2020	**CyGAN**. ○ Training semi-supervised GAN to generate real-life data packets resembling the features of the NSL-KDD data set to decrease the abnormality present in the data set, which leads to a decrease in the number of false positives for a network intrusion detection system.
Jan 2020 – Fev 2020	**HealthCard**. ○ The project's main idea was a problem statement from SSIP-SGH 2020. This project aimed to create a centralized system for health-related records. I was responsible for creating the REST API. ○ The whole project consists of an Android app for patients and a web app for doctors, labs, and medicals. ○ I was responsible for creating the REST API.

Test scores

GATE CS	Marks: 68.12, All India Rank: 87, Score: 860
TOEFL	Total: 102, L:27, R:26, W:27, S:22

Awards & Grants

SERB International Travel Support (ITS)

NeurIPS Student Travel Grant

Teaching Experiences

Spring Term, 2023	**Teaching Assistant**, OPTIMIZATION FOR ML, Indian Institute of Science, Bengaluru, Primary Instructor : Professor Sundeep Chepuri.

Skills

Languages	Python(A), C/C++(B), JavaScript(B)
Frameworks	Keras, PyTorch, Tensorflow
Utilities	Anaconda, Git, Sublime Text, Jupyter Notebook
Communication	Gujarati (Mother Tongue), English (SRW), Hindi (SRW)

Reviewing

COLING	2022
ACL	2023
FIRE	2021,2022

Relevant Courses

Online	DeepLearning.AI Specialization, Introduction to Machine Learning, Introduction to Statistics and Probability
Classroom	Computational Linear Algebra, Computational Methods of Optimization, Stochastic Modelling and Applications, Digital Image Processing, Pattern Recognition and Neural Networks, Reinforcement Learning, Advanced Image Processing, Data Structures and Algorithms, Advanced Deep Representational Learning, Data Analytics