1st International Workshop on Sustainable AI for Edge (SAI4E) - 2023

in conjuntion with AIMLSys 2023 (25-28 October 2023)

The Chancery Pavilion, Bangalore

Workshop Date: 28th October 2023

Keynote talks:

- **Importance of AI Explainability for Hight Frequency Data** - Mr. Shyam Prabhakar (Director of Engineering, Reality AI/Renesas)
- **How can neuroscience aid in designing a mortal computer?** - Prof. Ayon Borthakur (Asst. Professor, IIT Hyderabad)

Invited talks:

- **Unlocking the Value of Vehicular Data: A Simple, Unsupervised Approach for On-board Driver Behavior Classification** - Prof. Punit Rathore (IISc Bangalore)
- **Machine Learning for Resource-efficient Industrial and Infrastructural Inspections** - Dr. Hrishikesh Sharma (Senior Scientist, TCS Research)

Technical Papers:

Paper ID	Title	Authors
1	Multi-Task Learning for massive MIMO CSI Feedback	Sharan Mourya Bathala (Indian Institute of Technology Hyderabad)
2	A Study on Tiny YOLO for Resource Constrained XRay Threat Detection	Raghav Venkata Ambati (IIT Hyderabad)
4	GPU Implementation: Accelerating 3D-Bin Packing Problem	Krishna Vamsi, Ashwin Krishnan, Manoj Nambiar (TCS Research)
7	Binarized-Distance Computation for performing low-power Face Recognition with optimized storage	Vivek K Parmar, Manan Suri (IIT Delhi)

Tutorial:

Tutorial on TinyML and Embedded NAS - **Generation and Deployment of Tiny Machine Learning Programs on TinyML platform** - Gitesh Kulkarni, Swarnava Dey (TCS Research)

Panel Discussion:

Topic: Motivation, challenges, state of the art and future directions of Sustainable AI for Edge
Host: Dr. Balamurali P (Research Advisor, TCS Research)
Panelists: Mr. Shyam Prabhakar, Prof. Ayon Borthakur, Prof. Punit Rathore, Dr. Hrishikesh Sharma

Talk Abstracts and Speaker Biographies:

Mr. Shyam Prabhakar

Shyam was most recently a Director of Software Engineering at Reality AI, acquired by Renesas Electronics Corp last year. Reality AI focuses on edge AI for high frequency data. Shyam leads the Software As a Service product that helps engineers build Machine Learning models in a no-code way and directly deploy them onto microcontrollers and other devices at the edge. In addition, Shyam led the team building a new product line called RealityCheck AD which is an anomaly detection edge AI solution for industrial purposes. Prior to this Shyam was at a clean tech company, Opower (acquired by Oracle) as a Sr. Director of Engineering for close to a decade where they created an entirely new industry vertical called Behavioral Energy Efficiency. This product was one of a kind that combined behavioral science with energy efficiency leading to households reducing energy between 1.4 to 3.3 % at each utility. Shyam is currently doing a 2 year expat stint in Bangalore, India. Shyam has a Masters degree from the University of Maryland at College Park and an MBA from Virginia Tech university.

Talk Title: Importance of AI explainability for high frequency data

Abstract: Data created by endpoint devices (that typically sit at the extreme edge of an IoT network) is expected to grow at a staggering 85% with total data generated by IoT devices at roughly 73ZB in 2025. Just to put things in perspective, Zetta has 21 zeroes! But data is data till we do something with it… and this is where AI/ ML becomes so powerful and meaningful. Now with all the data, follows the hundreds of millions of AI models being trained, tested and deployed all over the world. A lot of these are being deployed in mission critical systems. But how do we know the innerworkings of these AI models, is it all black-box magic? Are we able to peel the curtain even a little bit to understand why the model is inferring the way it is OR what feature it's using to make those inferences? In this talk, let's take a look at how one could, for high frequency data.

Prof. Ayon Borthakur

I am an Assistant Professor in the Department of Artificial Intelligence at IIT Hyderabad. Previously, I worked at Innatera Nanosystems in the Netherlands as a Senior Neuromorphic Engineer-Machine Learning. Before that, I completed my Ph.D. at Cornell University, USA. As part of my Ph.D., I worked on neuroscience-inspired Artificial Intelligence for learning in the wild, focusing on its implementation in neuromorphic chips such as Intel Loihi. I completed my Bachelor of Technology in Electrical Engineering from IIT Dhanbad.

Talk Title: How can neuroscience aid in designing a mortal computer?

Abstract: The demand for intelligent computing is on the rise. But for numerous practical applications that involve real-time interactions, current AI algorithms, and the associated computing architectures still cannot match the performance of the brain. For instance, unlike biological neural networks, artificial neural networks (ANNs) tend to forget previously learned information upon learning new information (catastrophic forgetting). Moreover, today's computers used for the implementation of ANNs are slow and energy-consuming due to factors such as the physical separation of computing units and memory, unlike biological neural networks where computation and memory are co-localized. Accordingly in this research, we take neuroscience inspiration and work towards designing a mortal computer. In this talk, I will describe the properties of such a mortal computer and showcase selected neuroscience-inspired AI design results. I will present EPLff - inspired by the external plexiform layer of mammalian olfaction, which learns from a few shots without catastrophic forgetting. Next, I will discuss Sapinet - a multilayer recurrent neural network for learning in the wild optimized for implementation in Intel Loihi.

Dr. Punit Rathore

Dr. Punit Rathore is currently an Assistant Professor at Indian Institute of Science, Bangalore in Robert Bosch Centre for Cyberphysical Systems, jointly with Centre for infrastructure, Sustainable Transportation, and Urban Planning. Before joining IISc, Dr. Rathore worked as a Postdoctoral Fellow in Senseable City Lab at Massachusetts Institute of Technology (MIT), Cambridge, USA during for two years, and in AI Lab at National University of Singapore for almost one year. Dr Rathore completed his Ph.D. from the University of Melbourne, Australia in 2019. Prior to PhD, Dr. Punit worked as a Researcher in Automation Division at Tata Steel Limited, Jamshedpur, where he developed several real-time systems based on machine learning and machine vision for manufacturing industries. His research interests are in big data analytics, unsupervised learning, explainable ML, and data-driven analytics for smart cities, including intelligent transportation systems.

Talk Title: Unlocking the Value of Vehicular Data: A Simple, Unsupervised Approach for On-board Driver Behavior Classification

Abstract: In the fast-evolving sphere of modern vehicles, the extensive data generated, running up to several Gigabytes per day, hold not only instrumental significance for vehicular functioning but also for amplifying our understanding of nuanced driving experiences. Dr. Punit Rathore will delve into their novel unsupervised methodology that harnesses the wealth of data collected from vehicles to classify driver behavior within specific geographical contexts. This simple, yet efficient approach, tested on datasets involving professional truck drivers and validated against human-labeled ground truth data, stands out as the first to scrutinize and validate the driving behaviors of professional truck drivers on such a significant scale, correlating them with tangible safety scores. Dr. Punit will also present challenges and limitation of existing driving behaviour analysis algorithms for Indian traffic conditions, followed by a brief introduction of their ongoing work on driving behaviour analysis in Indian traffic conditions.

Dr. Hrishikesh Sharma

Hrishikesh Sharma is a Senior Scientist at TCS Research. His core research areas are Computer Vision and Machine Learning, with specific long-term focus on anomaly detection as applied to inspection and safety scenarios. He holds a PhD in Digital Systems from IIT Bombay for which he recieved an award, and a Dual Degree in Signal Processing also from IIT Bombay. He has published 14 peer reviewed papers and has filed over 30+ unique patents across multiple geographies. He also has prior hands-on experience of complete embedded product life cycle, from conception to deployment.

Talk Title: Machine Learning for Resource-efficient Industrial and Infrastructural Inspections

Abstract: Visual inspections for safety and quality assessment are commonplace in utilities/infrastructure and manufacturing industries, respectively. Vast outdoor infrastructures such as bridges, dams, rail corridors, tunnels, pavements and roads, towers and buildings etc. require periodic health inspections, to maintain their safety and durability. Due to vastness, most of such inspections in practice are visual inspections, done aerially e.g. using drones. Such inspections are quite challenging for many reasons. Nascent damage/fault regions get imaged as small-sized regions quite often, making their segmentation a tough task. The other challenges include types and variations of faults themselves. Many intrinsic variations in various fault regions such as cracks are in fact continuous factors of variations, e.g. region's shape, region's texture etc. Extrinsic factors of variations mainly arise from outdoor conditions such as lighting condition, surface wetness, occlusion etc. Due to such variations, the fault regions generally display a lack of any obvious region-wide pattern. Similar challenges exist when manufacturing industry's output has to be visually inspected for quality assessment, though some of the outdoor factors of variations such as lighting condition do not hold. Especially the surface defects which more commonly happen in material production industry e.g. steel, are highly variable classes of defects. The defect region needs to be accurately detected so that the downstream decision about which way to dispense with faulty produce/lot be enabled. In this talk, I will discuss these challenges and give insights into an in-house multipurpose deep learning based model and its deployments.