

Program (ver. 2)

The 10th International Conference on Big Data Analytics Conference
(BDA) Dec. 19-22, 2022, Hyderabad, India.

(<https://www.bda2022.org/>)

(Two workshops, Four keynotes, 14 research papers,
Three invited talks, Four tutorials, One Panel)

I. December 19, 2022 (Monday): Workshop Program (Venue: IIIT Hyderabad, India)

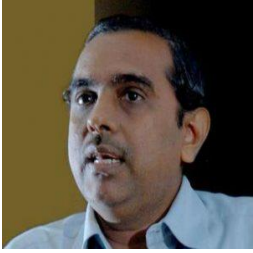

December 19, 2022 (Monday) All timings are as per Indian Standard Time (IST) (UTC + 05:30) For more details, please visit respective workshop website available at the following page. https://www.bda2022.org/Programme/workshops			
Time Duration	Hall 1	Hall 2	Hall 3
09.00-13.00	Data Challenges in Assessing (Urban & Regional) Air Quality (DACAAQ 2022) (https://dacaag.github.io/dacaag2022/) Organizers: Prof. Girish Agrawal (O.P. Jindal Global University, Sonapat, India), Prof. Anirban Mondal (Ashoka University, Sonapat, India), Prof. P. Krishna Reddy (IIIT Hyderabad, Hyderabad, India)	Workshop on Big Data Analytics using HPCC Systems (https://sites.google.com/rvce.edu.in/bda-workshop-iiith/home) Organizers: Dr G. Shobha (RV College of Engineering, Bengaluru, India), Jyoti Shetty (RV College of Engineering, Bengaluru, India)	Workshop on Data Science for Justice Delivery in India (DSJDI-2022) Organizers: P. Krishna Reddy, IIIT Hyderabad, Telangana State, India, K.V.K. Santhy, NALSAR University of Law, Telangana State, India
13.00-14.00	Lunch Break		
14.00-18.00	DACAAQ 2022 (Afternoon session) (https://dacaag.github.io/dacaag2022/)	Workshop on Universal Acceptance and Email Address Internationalization (http://dns.int.in/bda) Organizers: Shri Satish Babu (Technology Working Group Chair, UASG, www.uasg.tech), Harish Chowdhary (UA Ambassador, www.uasg.tech), Mr. K Mohan Raidu (President, ISoc India Hyderabad Chapter)	


II. December 20-22, 2022: Conference Program (Venue: IIIT Hyderabad, India) (<https://www.bda2022.org/>)


(Four keynotes, 14 research papers, three invited talks, four tutorials, One Panel)

December 20, 2022 (Tuesday)	
08.30-09.00	Registration and networking
09.00-09.30	Inauguration
09.30-10.30	Keynote 1 by Y. Narahari (Indian Institute of Science, Bangalore)
10.30-11.00	TEA BREAK
11.00-11.30	Invited talk 1 by Sridhar Viswanathan, Bank of America
11.30-12.30	Paper 1, Paper 2, Paper 3
12.30-13.30	LUNCH BREAK
13.30-14.30	Keynote Talk 2 by Sanjay Madria (Missouri University of Science and Technology, USA)
14.30-15.45	Paper 4, Paper 5, Paper 6, Paper 7
15.45-16.00	TEA BREAK
16.00-17.30	Tutorial 1
December 21, 2022 (Wednesday)	
08.30-09.30	Registration and networking
09.30 -10.30	Keynote 3 by Raj Sharman (University at Buffalo)
10.30 -11.00	TEA BREAK
11.00 -11.30	Invited Talk 3 by S. Bapi Raju (IIIT Hyderabad)
11.30-12.30	Paper 8, Paper 9, Paper 10
12.30-13.30	LUNCH BREAK
13.30-15.00	Tutorial 2
15.00-15.30	TEA BREAK
15.30-17.00	Panel
17.00-18.00	Networking break/ Steering committee meeting (invited members only)
18.00-21.00	Cultural program and Banquet Dinner
December 22, 2022 (Thursday)	
08.30-09.30	Registration and networking
09.30 - 10.30	Keynote 4 by Philippe Fournier-Viger (Shenzhen University, China)
10.30 -10.45	TEA BREAK
10.45-11.15	Invited talk 4 by Arvind Agarwal (IBM Research, India)
11.15-12.30	Paper 11, Paper 12, Paper 13, Paper 14
12.30-13.30	LUNCH
13.30- 15.00	Tutorial 3
15.00 -15.30	TEA BREAK
15.30 -17.00	Tutorial 4

I. KEYNOTES	
Keynote 1	<p>Title: Data Driven Crop Portfolio Recommendation for Agricultural Farmers</p> <p>by <u>Y. Narahari</u>, Indian Institute of Science, Bangalore.</p> <p>Abstract: Agriculture has a significant role to play in any emerging economy and provides the source of income and employment for a significant fraction of the population. A key challenge faced by small and marginal farmers is to determine which crops to grow to maximize their utilities. With a wrong choice of crops, farmers could end up with sub-optimal yields and low, and possibly significant loss of revenue. There is abundant literature in this area. In this talk, we describe a data driven system - ACRE (Agricultural Crop Recommendation Engine) - a novel tool designed us, that provides a scientific method to choose a crop or a portfolio of crops, to maximize the utility to the farmer. ACRE uses available data such as soil characteristics, weather conditions, and historical yield data, and uses state-of-the-art machine learning/deep learning models to compute an estimated utility to the farmer. A technical novelty in ACRE is to harness the use of Sharpe Ratio, a popular risk metric in financial investments. Using the Sharpe ratio, we generate a ranking on candidate recommendations</p>

	<p>of portfolios of crops. We use publicly available data from the agmarknet portal in India to present several promising data driven thought experiments with ACRE.</p> <p>Biography:</p>  <p>Narahari got his B.E. from Department of Electrical Communication Engineering in 1982, M.E. and Phd from Department of Computer Science and Automation in 1984 and 1987 respectively. In February 1988, he joined as the faculty of the Department of Computer Science and Automation and was Chair of the department during January 2010 – July 2014. He was the Dean of the Division of EECS (Electrical, Electronics, and Computer Sciences) from August 2014 to July 2021. He was also chairing the Office of DIGITS (Digital Campus and Informational Technology Services) from January 2016 to August 2020. During 1992, he was a Post-Doctoral Researcher at the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, USA and during 1997, he was a Visiting Scientist on sabbatical at the National Institute of Standards and Technology, Gaithersberg, Maryland, USA.</p> <p>The focus of Narahari's current research is to apply game theory, mechanism design, and machine learning to research problems at the interface of computer science and economics. In particular, he is interested in algorithmic game theory, design of auctions and electronic markets, dynamic mechanisms with learning, crowdsourcing , online education, social network analysis, and blockchains.</p>
Keynote 2	<p>Title: <TBD> by Raj Sharman, University at Buffalo.</p> <p>Abstract: <TBD></p> <p>Biography:</p>  <p>Sharman's research is focused on extreme events from a decision-support system perspective and on health information technology-related issues. This includes factors influencing online health information search, meaningful use of ambulatory EMR, resilience in hospital information systems, health information exchanges, health care social networks as well as a simulation based study for managing the hospital's emergency room capacity in extreme events, active shooter incidents and mass casualty event management.</p> <p>His expertise also includes information systems infrastructure management as it relates to information assurance, internet performance and distributed computing. Sharman's papers have been published in a number of national and international journals, and he is the</p>

	<p>recipient of several grants from the university as well as external agencies, including the National Science Foundation.</p> <p>He serves as an associate editor for the following journals: Journal of Information Systems Security, Journal of Information Privacy and Security, and Springer Security Informatics Journal.</p>
Keynote 3	<p>Title: Advances and challenges for the discovery of interesting patterns in data by Philippe Fournier-Viger, Distinguished Professor, Shenzhen University, China</p> <p>Abstract:</p> <p>Intelligent systems and tools play an important role in various domains such as for factory automation, e-business, and software engineering. To build intelligent systems and tools, high-quality data is generally required. Moreover, these systems need to process complex data and can yield large amounts of temporal data such usage logs, and data collected from sensors. Managing the data to gain insights and improve these systems is thus a key challenge. It is also desirable to be able to extract information or models from data that are easily understandable by humans. Based on these objectives, this talk will discuss the use of data mining algorithms for discovering interesting and useful patterns in temporal data generated from intelligent systems or from other applications.</p> <p>The talk will first briefly review early study on designing algorithms for identifying frequent temporal patterns in discrete sequences and time-interval data. Then, an overview of recent challenges and advances will be presented to identify other types of interesting patterns in complex data. Topics that will be discussed include high utility patterns, locally interesting patterns, trending patterns, time-interval patterns and periodic patterns. Lastly, the SPMF open-source software will be mentioned and opportunities related to the combination of pattern mining algorithms with traditional artificial intelligence techniques for intelligent systems will be discussed.</p> <p>Biography:</p>  <p>Philippe Fournier-Viger (Ph.D) is a Canadian researcher, distinguished professor at Shenzhen University (China). Five years after completing his Ph.D., he came to China and became the youngest full professor at the Harbin Institute of Technology (Shenzhen), after obtaining a title of national talent from the National Science Foundation of China. He has published more than 350 research papers related to data mining, intelligent systems and applications, which have received more than 10,800 citations (H-Index 52). He is editor-in-chief of Data Science and Pattern Recognition and former associate editor-in-chief of the Applied Intelligence journal. He is the founder of the popular SPMF data mining library, offering more than 240 algorithms, cited in more than 1,000 research papers. He is a co-founder of the UDML, PMDB and MLiSE series of workshops held at the ICDM, PKDD, DASFAA and KDD conferences.</p>
Keynote 4	<p>Title: EMOCOV: Machine Learning for Emotion Detection, Analysis and Visualization using COVID-19 Tweets by Sanjay Madria, Missouri University of Science and Technology, USA.</p>

	<p>Abstract:</p> <p>The adversarial impact of the Covid-19 pandemic has created a health crisis globally all over the world. This unprecedented crisis forced people to lockdown and changed almost every aspect of the regular activities of the people. Thus, the pandemic is also impacting everyone physically, mentally, and economically, and it, therefore, is paramount to analyze and understand emotional responses during the crisis affecting mental health. Negative emotional responses at fine-grained labels like anger and fear during the crisis might also lead to irreversible socio-economic damages. In this talk, I will discuss a neural network model trained using manually labeled data to detect various emotions at fine-grained labels in the Covid-19 tweets automatically. I will discuss about a manually labeled tweets dataset on COVID-19 emotional responses along with regular tweets data. A custom Q&A roBERTa model to extract phrases from the tweets that are primarily responsible for the corresponding emotions has been designed. None of the existing datasets and work currently provide the selected words or phrases denoting the reason for the corresponding emotions. The classification model outperforms other systems and achieves a Jaccard score of 0.6475 with an accuracy of 0.8951. The custom RoBERTa Q&A model outperforms other models by achieving a Jaccard score of 0.7865. Further, I will present a historical emotion analysis using COVID-19 tweets over the USA including each state level analysis.</p> <p>Biography:</p>  <p>Sanjay K Madria is a Curators' Distinguished Professor in the Department of Computer Science at the Missouri University of Science and Technology (formerly, University of Missouri-Rolla, USA). He has published over 290 Journal and conference papers in the areas of mobile and sensor computing, Big data and cloud computing, data analytics and cyber security. He won five IEEE best papers awards in conferences such as IEEE MDM and IEEE SRDS. He is a co-author of a book (published with his two PhD graduates) on Secure Sensor Cloud published by Morgan and Claypool in Dec. 2018. He has graduated 20 PhDs and 33 MS thesis students, with 9 PhDs currently progressing. NSF, NIST, ARL, ARO, AFRL, DOE, Boeing, CDC-NIOSH, ORNL, Honeywell, etc. have funded his research projects of over \$18M. He has been awarded JSPS (Japanese Society for Promotion of Science) invitational visiting scientist fellowship, and ASEE (American Society of Engineering Education) fellowship. In 2012 and in 2019, he was awarded NRC Fellowship by National Academies, US. He is ACM Distinguished Scientist, and served/serving as an ACM and IEEE Distinguished Speaker, and is an IEEE Senior Member as well as IEEE Golden Core Awardee.</p>
II. INVITED TALKS	
Invited Talk 1	<p>Title: “Big Data in Cognitive Neuroscience: Opportunities and Challenges”, by S. Bapi Raju, IIT Hyderabad</p> <p>Abstract: <TBD></p> <p>Biography:</p>



Dr. S. Bapi Raju is a professor and head of the Cognitive Science Lab, IIIT Hyderabad. He was a former professor of School of Computer and Information Sciences, University of Hyderabad, Hyderabad, India during 1999-2019. He worked as a Researcher at ATR Research Labs, Kyoto, Japan and as an EPSRC Research Fellow at University of Plymouth, UK before returning to India. He has over 20 years of teaching and research experience in AI, Machine Learning, Neural Networks and Cognitive Science. He has worked on a variety of inter-governmental collaborative projects such as Indo-French, Indo-Trento, in the areas of computational and cognitive neuroscience with multidisciplinary teams comprising computer scientists, linguists, neuroscientists, psychologists and clinicians. He is also currently heading the Healthcare vertical in the DST-funded National Mission for Cyberphysical Systems (NM-CPS) Technology Innovation Hub under at IIIT Hyderabad called IHub-Data.

He has degrees in BE (Electrical Engineering) from Osmania University, MS (Biomedical Engineering) and PhD (Computer Science) from University of Texas, Arlington, USA. He is a senior member of IEEE, a member of ACM, Society for Neuroscience, and Cognitive Science Society.

Invited
Talk 2

Title: <TBD> Sridhar Viswanathan, Bank of America

Abstract: <TBD>


Biography: Sridhar Viswanathan, working as architect at BA Continuum India Pvt Ltd, Hyderabad India. I have over 17 years of experience in Big Data, Statistical and business analytics. Skilled in Java, Big Data systems, Hadoop, Spark, Kafka, Tableau and HBase. I have been with Bank for close to 11 years and involved in flagship projects by leading teams to achieve high performance and deliver complex technical solutions. I have also trained professionals on Data visualization. I have interests on engineering problems related to real time streaming and enjoy solving them and learn lessons from failures. I have completed executive masters in data science from IIT Hyderabad in 2017. Prior to that i have worked in Deloitte and Accenture in healthcare, health insurance and banking domains. I have bachelor degree from Coimbatore institute of technology in information technology.

I have filed a patent (Patent Reference Number: P12952US01) on "Generating and providing enhanced user interfaces by implementing data, ai , intents and personalization (DAIP) technology.

Invited
Talk 3

Title: “Advances in NLP Research for Automated Business Intelligence” by Arvind Agarwal, IBM India,


Abstract: Automated business intelligence derives insights from data to help businesses make right decisions for their business processes. These business processes can range from back end IT operations, to designing and executing a marketing campaign, to creating a business strategy, among many others. Automated business intelligence attempts to automate these processes by removing dependency on human, by providing them new ways to interact with data. Some of these interactions, which not so long ago seemed almost impossible, have now become possible due to the recent advances in NLP, and particularly, in deep learning and large language models. Specifying a SQL query in natural language, let the data speak for itself in human understandable text, being able to converse with data and get insights are few examples of such interactions. In this talk, we will cover some of these recent advances in NLP research, and how they are influencing the area of automated business intelligence. The talk shall cover both, an industrial view of the automated business intelligence in the form of available tools; and an academic view in the form of technical problems.

	<p>We will cover a range of technical problems including data search and exploration through semantic technologies, data insights via natural language querying and free form interaction, and use of NLP for exploratory data analysis including for data insights and data stories. We will conclude the talk with some food for thought by discussing open research problems in this space.</p> <p>Biography:</p>  <p>Arvind Agarwal is a Senior Technical Staff Member and Manager at IBM Research, India (Gurgaon) where he leads a team of research scientists and software developers to develop solutions in the space of AI-driven data processing and data analytics. Prior to joining IBM, he was a research scientist at Palo Alto Research Centre (PARC), Webster, New York. His research interests are in the areas of machine learning, natural language processing, deep learning, and text analytics. He is especially interested in conversational data analytics, and in machine learning sub-areas that deal with the problem of limited supervised data, such as self-learning, semi(un)-supervised learning, zero shot learning, domain adaptation, multitask learning etc. Arvind completed his PhD in Computer Science from University of Maryland, his M.S. in Computer Science from University of Utah and Bachelor's from Birla Institute of Technology & Science, Pilani. He has about 20 patents, and more than 35 publications in top ML and NLP conference such as EMNLP, AAAI, KDD, NIPS, IJCAI, ATSTATS. He is also a recipient of Heidelberg Laureate Forum Young Researchers award, and ECML 2010 best student paper award.</p>
III. RESEARCH PAPERS	
Paper 1	“Learning enhancement using Question-Answer generation for e-book using contrastive fine-tuned T5” by Shobhan Kumar (IIIT Dharwad)*; Arun Chauhan (Indian Institute of Information Technology Dharwad); Pavan Kumar (IIIT Dharwad)
Paper 2	“A Deep Learning based Approach to Automate Clinical Coding of Electronic Health Records” by Ashutosh Kumar (ABV-IIITM Gwalior); Santosh Singh Rathore (ABV-IIITM Gwalior)*
Paper 3	“Drugomics: Knowledge Graph & AI to Construct Physicians' Brain Digital Twin to Prevent Drug Side-effects and Patient Harm” by Asoke K Talukder (SRIT india)*; Erwin Selg (SRH Fernhochschule GmbH.); Ryan Fernandez (SJRI); Tony Raj (SJRI); Abijeet Waghmare (SJRI); Roland Haas (IIITB)
Paper 4	“A Novel Feature Selection Based Text Classification using Multi-layer ELM” by Rajendra Kumar Roul (Thapar Institute of Engineering and Technology, Patiala, Punjab)*; Gaurav Satyanath (Department of Electrical Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania)
Paper 5	“Determining the severity of Dementia using ensemble learning” by Shruti Srivatsan (SVCE)*; Sumneet Kaur Bamrah (SVCE); Gayathri KS (SSN)
Paper 6	“A Machine and Deep Learning Framework to Retain Customers based on their Lifetime Value” by Kannan Kumaran (NCI)*
Paper 7	“A distributed ensemble machine learning technique for emotion classification from vocal cues” by Bineetha Vijayan (Cochin University of Science And Technology)*; Gayathri Soman (Cochin University of Science And Technology); Vivek M.V. (Cochin University of Science And Technology); M.V. Judy (Cochin University of Science And Technology)
Paper 8	“Hui2Vec: Learning Transaction Embedding Through High Utility Itemsets” by KHALED BELGHITH (NARD Intelligence)*; Philippe Fournier-Viger (Shenzhen University); Jassem Jawadi (ISTIC)

Paper 9	“Discovering Top-K Periodic Patterns in Temporal Databases” by Likhitha Palla (University of Aizu)*; Uday Kiran RAGE (University of Tokyo); Penugonda Ravikumar (The University of Aizu); Yutaka Watanobe (The University of Aizu)
Paper 10	“Extremely Randomized Tree based Sentiment Polarity Classification on Online Product Reviews” by Saranya R B R B (Narayana guru College of Engineering)*; K Ramesh (Anna University Regional Campus); NISHA K DEVI (Bannari Amman Institute of Technology)
Paper 11	“Analyze the Impact of Weather Parameters for Crop Yield Prediction using Deep Learning” by Pragnesh Patel (Ahmedabad University)*; Sanjay Chaudhary (Ahmedabad University); Hasit Parmar (L. D. College of Engineering)
Paper 12	“Analysis of Weather Condition based Reuse among Agromet Advisory: A Validation Study” by Mamatha Alugubelly (IIIT Hyderabad)*; Krishna Reddy P (International Institute of Information Technology, Hyderabad); Anirban Mondal (Ashoka University); Mahadevappa SG (P.J.T.S Agricultural University); Balaji Naik Banothu (P.J.T.S.Agricultural University); Sreenivas Gade (P.J.T.S. Agricultural University)
Paper 13	“Community Detection in Large Directed Graphs” by Siqi Chen (University of Cincinnati); Raj K Bhatnagar (University of Cincinnati)*
Paper 14	“ARCORE: Software Requirements Dataset for Service Identification” by Vijaya Peketi (Independant Consultant)*; Surekha satti (Independant Consultant)

IV. TUTORIALS

Tutorial 1	<p>Title: Malware Analysis and Detection by Mohit Sewak (Microsoft, India) and Hemant Rathore (BITS Pilani, Goa)</p> <p>Abstract: Today computing devices like laptops, mobile phones, smart devices, etc., have penetrated very deep into our modern society and have become an integral part of our daily lives. Currently, more than half of the world's population uses computers/mobile devices for their professional/ personal needs. However, these computing devices are targeted by malware designers encouraged by profits/gains associated with the attack. According to a recent report, monetary losses due to cybercrime are expected to reach 10 trillion dollars annually by 2025. The primary role in providing defense against malware attacks is designed and developed by the anti-malware community (researchers and the anti-virus industry). Traditionally anti-viruses are based on the signature, heuristic, and behavior based detection engines. However, these engines are unable to detect next-generation polymorphic and metamorphic malware. Thus researchers have started developing malware detection engines based on machine learning to complement the existing anti-virus engines. However, there are many open research challenges in these models like adversarial robustness, explainability, fairness, etc., which we are going to discuss in detail during the tutorial.</p> <p>Biography:</p> <p>Mohit Sewak is an Artificial Intelligence and Cybersecurity researcher with over 15 years of experience in designing innovative AI software and solutions. Mohit holds more than a dozen patents across the US, India, and worldwide for innovative AI solutions that empower many international products. Mohit is the author of multiple AI book titles on topics including technologies like Deep Reinforcement Learning and Convolutional Neural Networks. Mohit's research is focused on designing AI-based malware and other advanced threat detection and protection systems. Currently, Mohit serves as a Principal Data Scientist for Security & Compliance Research at Microsoft R&D.</p> <p>Hemant Rathore is a cyber security expert with more than ten years of experience in industry and academia. His current work focuses on the topic of Adversarial Robustness and Explainability in Malware Detection Models. His research interests are in the area of Malware Analysis, Network Security, Machine Learning, and Operating Systems. He has guided several undergraduate and postgraduate students in their independent research projects and published many research papers in reputed journals/ conferences.</p>
------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tutorial 2	<p>Neuro-Symbolic Techniques for XAI and Logical Reasoning by Raghava Mutharaju (IIIT Delhi, Delhi)</p> <p>Abstract: Neuro-Symbolic AI brings together the neural and symbolic aspects of AI. Symbolic techniques are transparent with provable guarantees for correctness. On the other hand, neural techniques are robust to noise and can easily pick up the patterns from the data. By combining the complementary strengths of these two approaches, it is possible to build AI systems that are robust and transparent. In this tutorial, we will discuss the use of neuro-symbolic techniques for explainable AI (XAI) and logical reasoning over Knowledge Graphs and ontologies.</p> <p>Biography:</p>  <p>Bio: Raghava Mutharaju is an Assistant Professor in the Computer Science and Engineering department of IIIT-Delhi, India and leads the Knowledgeable Computing and Reasoning (KRACR; pronounced as cracker) Lab. He got his PhD in Computer Science and Engineering from Wright State University, USA, in 2016. He has worked in Industry research labs such as GE Research, IBM Research, Bell Labs, and Xerox Research. His research interest is in Semantic Web and in general in Knowledge Representation and Reasoning. This includes knowledge graphs, ontology modelling, reasoning, querying, and its applications. He has published at several venues such as ISWC, ESWC, ECAI, and WISE. He has co-organized workshops at ISWC 2020, WWW 2019, WebSci 2017, ISWC 2015 and tutorials at ISWC 2019, IJCAI 2016, AAAI 2015 and ISWC 2014. He is/has been on the Program Committee of several (Semantic) Web conferences such as AAAI, WWW, ISWC, ESWC, CIKM, K-CAP and SEMANTiCS. More information is available on his lab's homepage at https://kracr.iiitd.edu.in/.</p>
Tutorial 3	<p>Federated Learning in the Real-World: From Theory to Practice By Tushar Semwal (Microsoft, India), Madhusudhanan Krishnamoorthy (Microsoft, India), Rajeev Gupta (Microsoft, India)</p> <p>Abstract: With the advent of the Internet of Things (IoT), there has been a huge surge in the volume of data collected by the devices at the edge of a network. This data is often collected and stored in the remote cloud servers to gain useful insights by training a model on this data. As an alternative, Federated Learning has been proposed where instead of learning a single global model centrally at the cloud server, each participating client device trains a model on its own local data and only share the weight gradients with the shared global model. Thus, in contrast to the sharing of raw data, the weights of the model are shared and distributed across the federation of client devices. One important use case could be an IoT in medical and health environments, Federated Learning (FL) can enable other organizations, which have similar data and have similar modelling requirements, to train in a single better global model which can then be distributed to each participating institute. In this tutorial, we</p>

will begin by providing a formal definition of FL, basic terminologies, architectures, and overview of challenges associated with centralized machine learning paradigms. We will then describe the federated learning framework through its various flavors such as horizontal federated learning, vertical federated learning, and federated transfer learning. In addition, the tutorial will also cover our recent published work on Federated Transfer learning. In this full day tutorial, after the first half of discussing the theoretical aspects of FL, the second half will begin with a hands-on introduction to python-based programming of a simple FL algorithm and testing on a benchmark dataset. In the tutorial, we will also introduce a fresh domain on Federated Graph Learning where the different components of FL are adapted for graph datasets.

Biography:

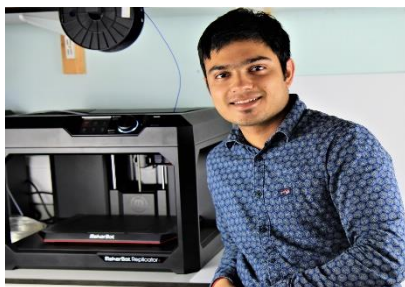
Biography

Tushar Semwal



Tushar Semwal is an Applied Scientist at Microsoft Search Assistant & Intelligence (MSAI), India. He got both master's and PhD degrees in Computer Science and Engineering from IIT Guwahati. Before joining Microsoft, Tushar served as a Research Associate at the Soft Computing Labs in the University of Edinburgh, Scotland. He is a recipient of the prestigious 4-year research fellowship award from TCS India, for his industry applicable research. He has won several travel grants from Microsoft, SERB India, SIAM, and TCS. His research interests include privacy-aware ML, Graph Representation Learning, and large-scale distributed systems.

Madhusudhan Krishnamoorthy



Madhusudhan Krishnamoorthy is a Senior Data Scientist at Microsoft Search Assistant & Intelligence (MSAI), India. He got his master's in Data science and Engineering from BITS Pilani. Before joining Microsoft, Madhu served as a Chief Data Scientist in Bank of America. He has more than 7 publications and 72 patents in the areas of cybersecurity, Mixed Reality, LiFi, Information extraction and cellular automata. His current work focuses on Graph representation learning and serving of embeddings at a larger scale.

Rajeev Gupta



Rajeev Gupta is a Principal Applied Scientist at Microsoft Search Assistant & Intelligence (MSAI), India. He got his PhD from Indian Institute of Technology (IIT) Mumbai (Bombay) in the area of distributed data management. He has more than 30 publications and 20 patents in the areas of data management, information extraction, and distributed computing in reputed conferences and journals.

Tutorial 4

Title: Self-Supervised Learning to Process Labeled and Unlabeled Medical Image Data

By Mayuri Mehta, Sarvajani College of Engineering and Technology, Sarvajani University, Surat - 395001, Gujarat, India.

Abstract: Medical imaging plays a significant role in developing automated clinical applications for early detection, monitoring, diagnosis, and treatment evaluation of various medical conditions. Deep learning is essential for in-depth and accurate analysis of medical images. Specifically, deep convolutional networks are most appropriate for extracting meaningful features from medical images. A huge amount of labeled data is required to train these deep convolutional networks. However, manually labeling medical images is time-consuming and expensive for medical experts. In addition, the major issue with the manual labeling of the huge dataset is the bias among human annotators. Therefore, applied deep learning is essential to process the dataset having a few labeled and largely unlabeled data.

Applied deep learning includes semi-supervised learning, Self-Supervised Learning (SSL) and reinforcement learning. Among them, SSL has been widely used in recent years to process medical data to reduce the data labeling cost and leverage the unlabeled data pool. SSL attempts to learn the visual representations of the data using proxy tasks perceived as pretext tasks. Pretext tasks are responsible for learning the prominent visual representations of data to use the learned representations or model weights obtained in the process for the downstream task.

The first half of this tutorial will comprise the emergence of AI in Healthcare, the significance of applied deep learning to process big healthcare data, various self-supervised learning frameworks, different types of pretext tasks, and how to design or select a suitable pretext task for processing medical images. In addition, medical image processing with labeled and unlabeled datasets will be discussed. In the second half of the tutorial, various SSL-based healthcare solutions (use cases) will be discussed. The discussion of each use case will include motivation, precise problem statement, the proposed solution, dataset, experimental results and challenges faced. Subsequently, the functioning of these healthcare solutions will be demonstrated. Finally, challenges and enormous future research opportunities will be discussed.

Biography:



Dr. Mayuri is a passionate learner, teacher and researcher. She received a doctorate in Computer Engineering from the National Institute of Technology, Surat, India. Her areas of teaching and research include Data Science, Healthcare Informatics, Machine Learning/Deep Learning, Computer Algorithms and Python Programming. Her 22 years of professional experience includes several academic and research achievements along with administrative and organizational capabilities. She is awarded the "Researcher of the Year Award (Engineering, Female)" by the 3rd International Business and Academic Excellence Award (IBAE-2021) committee for her Exceptional Calibre and Outstanding Performance as an Academician, Researcher, Mentor, Advisor, and a Thought Leader. She has 11 patents and 1 copyright to her credit. She has published two books: (1) Tracking and Preventing Diseases with Artificial Intelligence and (2) Knowledge Modelling and Big Data Analytics in Healthcare with Springer and CRC Press, respectively. Her books on "Explainable AI: Foundations, Methodologies and Applications" and "Recent Advances in Data and Algorithms for e-Government" with Springer will be published this year. She is the author of 33 research papers and 3 book chapters. She has worked on several academic assignments in collaboration with professors of universities across the globe. She has visited Germany, France, Switzerland, Oman, Dubai, Hongkong, Macau and Thailand for professional and personal purposes. She is an adjunct professor at Gujarat's largest private university- Parul University. Her AI-powered Healthcare project was approved for funding by the Multidisciplinary Research Unit of Surat Municipal Institute of Medical Education and Research (SMIMER). She has also received funds several times from Gujarat Council on Science and Technology (GUJCOST). She has received funds from Student Start-Up & Innovation Policy (SSIP), Government of Gujarat, India, for filing 2 patents. She has served in several International Conferences in different positions. She has conducted 80+ sessions in International Conferences, Short Term Training Programs (STTPs), Faculty Development Programs (FDPs), etc. With the noble intention of applying her technical knowledge for societal impact, she is working on several AI-powered research projects in Healthcare in association with doctors doing private practice and doctors of Medical Colleges. She is a member of professional bodies such as IEEE, ISTE, CSI.

V. PANEL

Panel	<p>Title: Data Science for sustainable development goals (SDS)</p> <p>Overview: Seventeen sustainable development goals (SDGs) are listed by United Nations in the year 2015 (https://sdgs.un.org/goals). The scope of the panel is as follows:</p> <p>"If data science researchers select the problems related to SDGs, the societal growth can be accelerated."</p> <p>The panel of eminent members share the perspectives on the above statement.</p> <ul style="list-style-type: none">• What are the typical research issues addressed by data science researchers?• List the potential research projects for SDGs, if any, in which data science has played a major
-------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>role.</p> <ul style="list-style-type: none"> • What are the challenges in carrying out research on Data Science for SDGs? • What is the framework for data science researchers to encourage research on Data Science for SDGs? <p>Moderators:</p> <p>Philippe Fournier-Viger, Shenzhen University and P.Krishna Reddy, IIIT Hyderabad</p> <p>Panellists: <TBD></p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------