



## **IEEE BigData 2021 Program Schedule Virtual Event**

Conference time zone: USA Eastern Standard Time

The emails with the login information to access the BigData 2021 virtual platform were sent out to all the registered participants on Dec 10, 2021. Pls check your spam folder if you didn't receive the email. If you need the email login information to access the virtual platform pls email [bigdata2021@underline.io](mailto:bigdata2021@underline.io), don't forget to attach the registration receipt if you ask for the login information. For any questions about the BigData 2021 virtual platform, pls email [bigdata2021@underline.io](mailto:bigdata2021@underline.io)

*Sponsored by*



## Table of Contents

<b>IEEE Big Data 2021 Program Schedule</b>	<b>6</b>
<b><i>Day 1: Wednesday, December 15, 2021</i></b>	<b>6</b>
Tutorial 2: Peer Review at Scale: Challenges and Opportunities	6
Tutorial 4: Web Stream Processing with RSP4J: A Big Data Perspective	6
Tutorial 8: Community Detection: From Statistical Modeling to Deep Representation	6
Workshop 7th International Workshop on Methods to Improve Big Data Science Projects (MIDP-2021)	6
Workshop: 10th Workshop on Scalable Cloud Data Management	6
Workshop: The 4th International Workshop on “Big Data Engineering and Analytics in Cyber-Physical Systems” (BigEACPS)	6
Workshop: Discriminative Pattern Mining Workshop	6
Workshop: The Fifth Annual Workshop on Applications of Artificial Intelligence in the Legal Industry	6
Workshop: 6th Workshop on Real-time Stream Analytics, Stream Mining, CER/CEP & Stream Data Management in Big Data	6
Workshop: Applications of Big Data Technology in the Transport Industry	7
Workshop: the 4th International Workshop on Big Media Dataset Construction, Management and Applications	7
Workshop: The Fourth Workshop on Big Data for Economic and Business Forecasting	7
Workshop: The 8th Annual Workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI)	7
Workshop: The 4th International Workshop on Cross-disciplinary Data Exchange and Collaboration (CDEC2021)	7
Workshop: Fourth International Workshop on the Internet of Things Data Analytics (IoTDA)	7
Workshop: 9th International Workshop on Distributed Storage and Blockchain Technologies for Big Data	7
Workshop: The 3rd International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD)	7
Workshop: The 6th IEEE International Workshop on Big Spatial Data (BSD 2021)	7
Workshop: First International Workshop on Data science for equality, inclusion and well-being challenges (DS4EIW 2021)	7
Workshop: Multi-Modal Medical Data Analysis	7
Workshop: Advances in High Dimensional (AdHD) Big Data (6th Workshop)	7
Workshop: The 5th IEEE International Workshop on Big Data for Financial News and Data	7
Workshop: Workshop on Knowledge Graphs and Big Data	7
Special Session: Information Granulation on Data Science and Scalable Computing	7
Special Session: 2021 Special Session on machine Learning on Big Data (MLBD 2021)	7
Bigdata Cup Challenges: IARAI’s Weather4cast Competition	7
Workshop: The 5th Workshop on Graph Techniques for Adversarial Activity Analytics (GTA3 2021)	7
Workshop: 2nd International Workshop on User Understanding from Big Data &	7
Workshop: International Workshop on Data Science Systems (DSS)	7
Workshop: The 5th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in BigData (HMDData 2021)	8
Workshop: 1st International Workshop on Serverless Machine Learning for Intelligent and Scalable AI Workflow	8
Workshop: The 5th International Workshop on Big Data Analytic for Cybercrime Investigation and Prevention	8
Workshop: The 5th Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	8
Workshop: Applications of Topological Data Analysis to ‘Big Data’	8

Workshop: 1st International Workshop on Big Data Analytics for Sustainability (BDS4S)	8
&	8
5th International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2021)	8
Workshop: Big Data Analytics for COVID-19	8
Symposium: 2021 National Symposium for NSF REU Research in Data Science, Systems, and Security (REU 2021 Symposium)	8
Special Session: 4th Special session on Healthcare Data	8
Special Session: 7th Special Session on Intelligent Data Mining	8
<b>Keynote Speech 1: Towards Trustworthy Data Science</b>	9
<b>Day 2: Thursday, December 16, 2021</b>	9
<b>Keynote Speech 2: Database Systems and Information Management – Trends and a Vision</b>	9
Poster Session 01: Big Data Science and Foundation	9
Poster Session 02: Big Data Infrastructure/Management	9
BigData Cup Challenges: Soft Sensing at Scale	9
Workshop: Workshop on Big Data for CyberSecurity (BigCyber-2021)	9
Workshop on Smart Farming, Precision Agriculture, and Supply Chain (SmartFarm-2021)	9
Workshop: Eighth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2021)	9
Special Session: 2021 Special Session on machine Learning on Big Data (MLBD 2021)	9
Workshop: Workshop on Deviant Activities on Social Media	9
I&G S1: Machine Learning Applications 1	9
S6: Big Data Visualization & Analytics	9
S7: Big Data Systems and Search	10
S8: Big Data Model Attack, Bias, and Reliability	10
S12: Theoretical Models for Big Data	10
S13: Online, Temporal, and Time Series Systems	10
Tutorial 3: Network Science: Network analysis and network optimization to solve complex problems in the big data era	10
Tutorial 6: Interactive Demonstrations and Research Use of the net.science Cyberinfrastructure for Network Science	10
Poster Session 03: Big Data as Service for Health Management	10
Poster Session 04: Big Data Search and Mining	10
I&G S2: CRM and Recommender Systems	10
S3: Big Data Interpretability & Trust	10
S11: Spatio-Temporal Data Applications	10
S14: Theoretical and Computational Models for Big Data	10
S16: Efficient and Scalable Data Management and Mining	10
Poster Session 05: Big Data Applications I	10
Poster Session 06: Big Data Applications II	10
Workshop: 1st Workshop on Multimodal Big Data	10
I&G S3: Platforms and Systems	10
S18: Complex Big Data Applications	10
S22: Recommendation and Social Networks	10
S24: Big Data Services and Applications	10
S28: Graph Neural Network Learning	10
Tutorial 9: Making Sense of The Internet of Things: The Big Data Perspective	11
Reinforcement Learning based Recommender System Challenge	11
BigData Cup Challenges: Reinforcement Learning based Recommender Systems Challenges	11
<b>Day 3: Friday, December 17, 2021</b>	11
<b>Keynote Speech 3: Machine Learning for Robot Locomotion: Grounded Simulation Learning and Adaptive Planner Parameter Learning</b>	11
Special Session: 2021 Special session on Privacy and Security of Big Data (PSBD 2021)	11

Workshop: Workshop on Cyber Threat Intelligence and Hunting with AI (CyberHunt 2021)	11
Workshop: Computational Archival Science: digital records in the age of big data	11
S29: Contrastive Learning	11
S25: Big Data Analytics and Modeling	11
S23: Web Search and Recommendation	11
S17: Multi-structured Data - Big Variety Data	11
S5: Privacy, Interpretability, and Fairness	11
I&G S4: IoT and Sensor Systems	11
Tutorial 7: Healthcare Big Data Analytics for COVID-19 Challenges: Digital Contact Tracing and Infodemic Management	11
Predicting Victories in Video Games @KnowledgePit	12
BigData Cup Challenges: IARAI's Science4cast Competition: A Machine Learning Challenges in the Science of Science	12
S30: Big Data Learning and Analytics	12
S19: Social, Web Search, and Mining	12
S10: Big Transportation and Spatial Data	12
S4: Health and Medical Data Applications	12
I&G S5: Machine Learning Applications 2	12
NSF Information Session (Six NSF program officers)	12
Workshop: The 2nd International Workshop on Big Data Reduction (IWBD-2)	12
S31: Big Data Computing & Analytics (I)	12
S21: Transfer and Domain Adaptation Learning	12
S15: Big Data Uncertainty and Bias Reduction	12
S9: Big Human-Derived Data	12
S1: Mobility and Big Data	12
I&G S6: Data and Text Analytics	12
Tutorial 1: Optimization Meets Reinforcement Learning	12
<b>Day 4: Saturday, December 18, 2021</b>	<b>13</b>
<b>Keynote Speech 4: Building towards a Responsible Data Economy</b>	<b>13</b>
I&G S7: Deep Learning and Neural Networks	13
S2: Opinions, Contradiction, and Threat Detection	13
S20: Social, Web, Recommendation and Mining	13
S26: Link and Graph Mining	13
S27: Cloud/Grid/Stream Data Computing	13
S32: Big Data Computing & Analytics (II)	13
Tutorial 5: Big Data Cleaning: Method Overview and Challenges	13
<b>Paper Sessions</b>	<b>14</b>
<b>Industry and Government Program Sessions</b>	<b>24</b>
<b>Special Sessions</b>	<b>26</b>
<b>BigData Cup Challenges</b>	<b>34</b>
<b>Workshop Detailed Schedule</b>	<b>36</b>
<b>Posters</b>	<b>63</b>
<b>IEEE BigData 2022</b>	<b>66</b>

# IEEE Big Data 2021 Program Schedule

Virtual Event

December 15 - December 18, 2021

Keynote Lecture: **60 minutes** (about 45 minutes for talk and 15 minutes for Q and A)

Main conference regular paper: **25 minutes** (about 20 minutes for talk and 5 minutes for Q and A)

Main conference short paper: **15 minutes** (about 11 minutes for talk and 4 minutes for Q and A)

All conference activities take place virtually and are based on USA Eastern Standard time.

Day 1: Wednesday, December 15, 2021		
Workshops , Special Session, Tutorials, BigData Cup Challenges, Symposium		
Time	Sessions/Workshops/Tutorial	Session Chair
9am-11am	Tutorial 2: Peer Review at Scale: Challenges and Opportunities	Nihar Shah, nihars@cs.cmu.edu
9am-11am	Tutorial 4: Web Stream Processing with RSP4J: A Big Data Perspective	Riccardo Tommasini and Pieter Bonte pieter.bonte@ugent.be, riccardo.tommasini@ut. ee
9am-11am	Tutorial 8: Community Detection: From Statistical Modeling to Deep Representation	Di Jin, Shirui Pan, Pengfei Jiao, and Weixiong Zhang jindi@tju.edu.cn, shirui.pan@monash.edu , pjiao@tju.edu.cn, weixiong.zhang@wustl. edu
9am-1pm	Workshop 7th International Workshop on Methods to Improve Big Data Science Projects (MIDP-2021)	Jeffrey Saltz, jsaltz@syr.edu
9am-1pm	Workshop: 10th Workshop on Scalable Cloud Data Management	Felix Gessert, felix.gessert@baqend.com
9am-1pm	Workshop: The 4th International Workshop on “Big Data Engineering and Analytics in Cyber-Physical Systems” (BigEACPS)	Akbar Siami Namin, akbar.namin@ttu.edu
9am-1pm	Workshop: Discriminative Pattern Mining Workshop	rbenton@southalabama.edu
10am-2:10pm	Workshop: The Fifth Annual Workshop on Applications of Artificial Intelligence in the Legal Industry	Haozhen Zhao, haozhen.zhao@ankura.com Jianping Zhang <jianping.zhang@ankura.co m>
9am-1pm	Workshop: 6th Workshop on Real-time Stream Analytics, Stream Mining, CER/CEP & Stream Data Management in Big Data	Sabri Skhiri <sabri.skhiri@euranova.eu>

9am-1pm	Workshop: Applications of Big Data Technology in the Transport Industry	John Easton, <a href="mailto:j.m.easton@bham.ac.uk">j.m.easton@bham.ac.uk</a>
9am-1pm	Workshop: the 4th International Workshop on Big Media Dataset Construction, Management and Applications	Chaoqun Hong, <a href="mailto:cqhong@txmut.edu.cn">cqhong@txmut.edu.cn</a> , <a href="mailto:myyou@tongji.edu.cn">myyou@tongji.edu.cn</a> , <a href="mailto:xqzhang@wzu.edu.cn">xqzhang@wzu.edu.cn</a>
9am-1pm	Workshop: The Fourth Workshop on Big Data for Economic and Business Forecasting	<a href="mailto:shangwei@amss.ac.cn">shangwei@amss.ac.cn</a> , <a href="mailto:weixu@ruc.edu.cn">weixu@ruc.edu.cn</a>
9am-1pm	Workshop: The 8th Annual Workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI)	Donghui Wu < <a href="mailto:donghui.wu@ieee.org">donghui.wu@ieee.org</a> >
9am-1pm	Workshop: The 4th International Workshop on Cross-disciplinary Data Exchange and Collaboration (CDEC2021)	Teruaki Hayashi, <a href="mailto:hayashi@sys.t.u-tokyo.ac.jp">hayashi@sys.t.u-tokyo.ac.jp</a>
9am-1pm	Workshop: Fourth International Workshop on the Internet of Things Data Analytics (IoTDA)	Eyhab Al-Masri, <a href="mailto:ealmasri@uw.edu">ealmasri@uw.edu</a>
9am-1pm	Workshop: 9th International Workshop on Distributed Storage and Blockchain Technologies for Big Data	Hui Li ( <a href="mailto:huilihuge@163.com">huilihuge@163.com</a> , <a href="mailto:lih64@pkusz.edu.cn">lih64@pkusz.edu.cn</a> ) Han Wang, <a href="mailto:wanghan2017@pku.edu.cn">wanghan2017@pku.edu.cn</a>
9am-1pm	Workshop: The 3rd International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD)	Pravallika Devineni, <a href="mailto:devinenip@ornl.gov">devinenip@ornl.gov</a>
9am-1pm	Workshop: The 6th IEEE International Workshop on Big Spatial Data (BSD 2021)	Ashwin Shashidharan, <a href="mailto:AShashidharan@esri.com">AShashidharan@esri.com</a> ,
9am-1pm	Workshop: First International Workshop on Data science for equality, inclusion and well-being challenges (DS4EIW 2021)	Tania CERQUITELLI, <a href="mailto:tania.cerquitelli@polito.it">tania.cerquitelli@polito.it</a>
9am-1pm	Workshop: Multi-Modal Medical Data Analysis	Bo Zhang, <a href="mailto:zbo@bupt.edu.cn">zbo@bupt.edu.cn</a>
9am-1pm	Workshop: Advances in High Dimensional (AdHD) Big Data (6th Workshop)	Sotiris Tasoulis, <a href="mailto:stasoulis@uth.gr">stasoulis@uth.gr</a> Aris Vrahatis < <a href="mailto:aris.vrahatis@gmail.com">aris.vrahatis@gmail.com</a> >
9am-1pm	Workshop: The 5th IEEE International Workshop on Big Data for Financial News and Data	Quanzhi Li, <a href="mailto:quanzhi.li@gmail.com">quanzhi.li@gmail.com</a>
9am-1pm	Workshop: Workshop on Knowledge Graphs and Big Data	Yuan An, <a href="mailto:ya45@drexel.edu">ya45@drexel.edu</a>
9am-1pm	Special Session: Information Granulation on Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, <a href="mailto:tsumoto@med.shimane-u.ac.jp">tsumoto@med.shimane-u.ac.jp</a>
8:55am-2pm	Special Session: 2021 Special Session on machine Learning on Big Data (MLBD 2021)	Alfredo Cuzzocrea < <a href="mailto:alfredo.cuzzocrea@unical.it">alfredo.cuzzocrea@unical.it</a> >
11am-3PM	Bigdata Cup Challenges: IARAI's Weather4cast Competition	Aleksandra Gruca <a href="mailto:weather4cast@iarai.ac.at">weather4cast@iarai.ac.at</a> , D Kreil (IARAI) <a href="mailto:david.kreil@iarai.ac.at">david.kreil@iarai.ac.at</a>
1-6pm	Workshop: The 5th Workshop on Graph Techniques for Adversarial Activity Analytics (GTA3 2021)	Jiejun Xu ( <a href="mailto:jxu@hrl.com">jxu@hrl.com</a> ) Hanghang Tong ( <a href="mailto:htong@illinois.edu">htong@illinois.edu</a> )
1-6pm	Workshop: 2nd International Workshop on User Understanding from Big Data & Workshop: International Workshop on Data Science Systems (DSS)	Claire Ding, <a href="mailto:dingqi@amazon.com">dingqi@amazon.com</a> Wutao Wei, <a href="mailto:weiwutao@gmail.com">weiwutao@gmail.com</a> Carlos Ordóñez, <a href="mailto:carlos@central.uh.edu">carlos@central.uh.edu</a>

1-6pm	Workshop: The 5th IEEE Workshop on Human-in-the-Loop Methods and Future of Work in BigData (HMDData 2021)	Senjuti Basu Roy, <a href="mailto:senjuti.basuroy@njit.edu">senjuti.basuroy@njit.edu</a> Alex Quinn <a href="mailto:aq@purdue.edu">aq@purdue.edu</a> Atsuyuki Morihsima <a href="mailto:morishima-office@ml.cc.tsukuba.ac.jp">morishima-office@ml.cc.tsukuba.ac.jp</a>
1-6pm	Workshop: 1st International Workshop on Serverless Machine Learning for Intelligent and Scalable AI Workflow	Dhaval Patel ( <a href="mailto:patehdha@us.ibm.com">patehdha@us.ibm.com</a> ), sh uxin Lin, <a href="mailto:shuxin.lin@us.ibm.com">shuxin.lin@us.ibm.com</a>
9am-6pm	Workshop: The 5th International Workshop on Big Data Analytic for Cybercrime Investigation and Prevention	Andrii Shalaginov, <a href="mailto:andrii.shalaginov@kristiania.no">andrii.shalaginov@kristiania.no</a>
9am-6pm	Workshop: The 5th Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen, <a href="mailto:zhchen@umbc.edu">zhchen@umbc.edu</a>
9am-6pm	Workshop: Applications of Topological Data Analysis to 'Big Data'	Sarah Tymochko, <a href="mailto:tymochko@egr.msu.edu">tymochko@egr.msu.edu</a> u, Gregory Henselman- Petrusek <a href="mailto:greg.henselman@gmail.com">greg.henselman@gmail.com</a> Tegan Emerson <a href="mailto:tegan.emerson@pnnl.gov">tegan.emerson@pnnl.gov</a>
9am-6pm	Workshop: 1st International Workshop on Big Data Analytics for Sustainability (BDS4S) & 5th International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2021)	Stephen McGough, <a href="mailto:Stephen.mcough@newcastle.ac.uk">Stephen.mcough@newcastle.ac.uk</a>
9am-6pm	Workshop: Big Data Analytics for COVID-19	Duoduo (Lindi) Liao, <a href="mailto:dliao2@gmu.edu">dliao2@gmu.edu</a>
9am-6pm	Symposium: 2021 National Symposium for NSF REU Research in Data Science, Systems, and Security (REU 2021 Symposium)	Jianwu Wang, <a href="mailto:jianwu@umbc.edu">jianwu@umbc.edu</a>
9am-6pm	Special Session: 4th Special session on Healthcare Data	Ozgun Pinarer <a href="mailto:opinarer@gsu.edu.tr">opinarer@gsu.edu.tr</a> Sultan Turhan <a href="mailto:sturhan@gsu.edu.tr">sturhan@gsu.edu.tr</a> ,
9am-6pm	Special Session: 7th Special Session on Intelligent Data Mining	Uraz Yavanoglu < <a href="mailto:urazyavanoglu@gmail.com">urazyavanoglu@gmail.com</a> >
8:25 - 8: 35 PM	<b>Open Ceremony</b> <b>Conference Co-Chair:</b> Usama Fayyad, Xingquan Zhu <a href="mailto:u.fayyad@northeastern.edu">u.fayyad@northeastern.edu</a> ; <a href="mailto:xzhu3@fau.edu">xzhu3@fau.edu</a>	
8: 35 – 8: 50 PM	<b>Conference Program</b> <b>Program Committee Co-chairs:</b> Yixin Chen, Heiko Ludwig, Yicheng Tu ' <a href="mailto:chen@cse.wustl.edu">chen@cse.wustl.edu</a> '; ' <a href="mailto:tuy@mail.usf.edu">tuy@mail.usf.edu</a> '; ' <a href="mailto:hludwig@us.ibm.com">hludwig@us.ibm.com</a>	
8:50 – 9:00 PM	<b>Industry and Government Program</b> <b>Industry and Government Co-Chairs:</b> Suren Byna, Liu Xiong; Jianping Zhang <a href="mailto:sbyna@lbl.gov">sbyna@lbl.gov</a> , <a href="mailto:xiong.liu@novartis.com">xiong.liu@novartis.com</a> , <a href="mailto:jianping.zhang@ankura.com">jianping.zhang@ankura.com</a>	



## Keynote Speech 1: Towards Trustworthy Data Science

9:00 – 10:00 PM

Dr. Jian Pei  
[jpei@cs.sfu.ca](mailto:jpei@cs.sfu.ca)

Session Chair: Xingquan Zhu ([xzhu3@fau.edu](mailto:xzhu3@fau.edu))

## Day 2: Thursday, December 16, 2021

### Keynote Speech 2: Database Systems and Information Management – Trends and a Vision

9:00 - 10: 00 am

Dr. Volker Markl  
[Volker.Markl@dfki.de](mailto:Volker.Markl@dfki.de)

Session Chair: Yicheng Tu ([tuy@mail.usf.edu](mailto:tuy@mail.usf.edu))

10 - 10: 15 am

Coffee Break

Time	Sessions/Workshops	
9-11am	Poster Session 01: Big Data Science and Foundation	Ladjel Bellatreche, <a href="mailto:bellatreche@ensma.fr">bellatreche@ensma.fr</a>
9-11am	Poster Session 02: Big Data Infrastructure/Management	Yufei Tang <a href="mailto:tangyf@fau.edu">tangyf@fau.edu</a>
9-11am	BigData Cup Challenges: Soft Sensing at Scale	<i>Sthitie Bom &amp; Chao Zhang</i> <a href="mailto:sthitie.e.bom@seagate.com">sthitie.e.bom@seagate.com</a> , <a href="mailto:chaozhang@uchicago.edu">chaozhang@uchicago.edu</a>
9am-1pm	Workshop: Workshop on Big Data for CyberSecurity (BigCyber-2021)	Sudip Mittal, <a href="mailto:mittal@cse.msstate.edu">mittal@cse.msstate.edu</a>
9am-1pm	Workshop on Smart Farming, Precision Agriculture, and Supply Chain (SmartFarm-2021)	Maanak Gupta, <a href="mailto:mgupta@tntech.edu">mgupta@tntech.edu</a>
9am-1pm	Workshop: Eighth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2021)	Nesreen Ahmed, <a href="mailto:nesreen.k.ahmed@intel.com">nesreen.k.ahmed@intel.com</a>
9am-2:15pm	Special Session: 2021 Special Session on machine Learning on Big Data (MLBD 2021)	Alfredo Cuzzocrea < <a href="mailto:alfredo.cuzzocrea@unical.it">alfredo.cuzzocrea@unical.it</a> >
9am-6pm	Workshop: Workshop on Deviant Activities on Social Media	Muhammad Nihal Hussain, <a href="mailto:mnhussain@ualr.edu">mnhussain@ualr.edu</a> , u, <a href="mailto:elmead@ualr.edu">elmead@ualr.edu</a> , Samer Al- Khateeb1 <a href="mailto:skhateeb1@creighton.edu">skhateeb1@creighton.edu</a> , KiranKumar.Bandeli <a href="mailto:kiran.kumar.bandeli@walmart.com">kiran.kumar.bandeli@walmart.com</a> , <a href="mailto:nxagarwal@ualr.edu">nxagarwal@ualr.edu</a>
10: 15AM - 12: 20 pm	I&G S1: Machine Learning Applications 1	Xiong Liu, <a href="mailto:xiong.liu@novartis.com">xiong.liu@novartis.com</a>
	S6: Big Data Visualization & Analytics	Nancy Roberts <a href="mailto:nancy.roberts.2@us.af.mil">nancy.roberts.2@us.af.mil</a>

	S7: Big Data Systems and Search	Lauritz Thamsen lauritz.thamsen@tu-berlin.de
	S8: Big Data Model Attack, Bias, and Reliability	Yaoqing Wang yqyang@berkeley.edu
	S12: Theoretical Models for Big Data	Peng Zhao pzhao@tamu.edu
	S13: Online, Temporal, and Time Series Systems	Hoda Eldardiry hdardiry@vt.edu
	Tutorial 3: Network Science: Network analysis and network optimization to solve complex problems in the big data era	Carlos Pinheiro, carlos.pinheiro@sas.com
	Tutorial 6: Interactive Demonstrations and Research Use of the net.science Cyberinfrastructure for Network Science	Chris Kuhlman, Dustin Machi, and SS Ravi cjk8gx@virginia.edu, dm8qs@virginia.edu, ssr6nh@virginia.edu
11:00am-1:00pm	Poster Session 03: Big Data as Service for Health Management	Ladjel Bellatreche, bellatreche@ensma.fr
11:00am-1:00pm	Poster Session 04: Big Data Search and Mining	Yufei Tang tangy@fau.edu
12: 20 - 1: 30 pm	<b>Lunch Break</b>	
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>
1:30pm-3:35pm	I&G S2: CRM and Recommender Systems	Jianping Zhang, jianping.zhang@ankura.com
	S3: Big Data Interpretability & Trust	Xiao Han xiao.han@usu.edu
	S11: Spatio-Temporal Data Applications	Shubhi Asthana sasthan@us.ibm.com
	S14: Theoretical and Computational Models for Big Data	Radwa El Shawi radwa.elshaw@ut.ee
	S16: Efficient and Scalable Data Management and Mining	Noah Daniels noah_daniels@uri.edu
1:00pm-3:00pm	Poster Session 05: Big Data Applications I	Ladjel Bellatreche, bellatreche@ensma.fr
1:00pm-3:00pm	Poster Session 06: Big Data Applications II	Yufei Tang tangy@fau.edu
1-6pm	Workshop: 1st Workshop on Multimodal Big Data	Duoduo Liao, dliao2@gmu.edu
3: 35 - 3: 55 pm	<b>Coffee Break</b>	
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>
3: 55 - 6: 00 pm	I&G S3: Platforms and Systems	Suren Byna, sbyna@lbl.gov
	S18: Complex Big Data Applications	Li Li lili20111739@bistu.edu.cn
	S22: Recommendation and Social Networks	Yixin Chen chen@cse.wustl.edu
	S24: Big Data Services and Applications	Ajay Dholakia ajay.dholakia@gmail.com
	S28: Graph Neural Network Learning	Qiao Kang qkang@lbl.gov

4-6pm	Tutorial 9: Making Sense of The Internet of Things: The Big Data Perspective	Michael Sheng, Amin Beheshti, Wei Zhang, and Adnan Mahmood michael.sheng@mq.edu.au, amin.beheshti@mq.edu.au, wei.e.zhang@adelaide.edu.a u, adnan.mahmood@mq.edu.au
8:00-10:00pm	<u>Reinforcement Learning based Recommender System Challenge</u> BigData Cup Challenges: Reinforcement Learning based Recommender Systems Challenges	Runze Wu wurunze1@corp.netease.com

Day 3: Friday, December 17, 2021		
9:00 - 10: 00 am	<b>Keynote Speech 3: Machine Learning for Robot Locomotion: Grounded Simulation Learning and Adaptive Planner Parameter Learning</b> Dr. Peter Stone <a href="mailto:pstone@cs.utexas.edu">pstone@cs.utexas.edu</a> Session Chair: Yixin Chen ( <a href="mailto:chen@cse.wustl.edu">chen@cse.wustl.edu</a> )	
10: 00 - 10: 15 am	Coffee Break	
Time	Sessions/Workshops	
8:55am-1:15pm	Special Session: 2021 Special session on Privacy and Security of Big Data (PSBD 2021)	Alfredo Cuzzocrea < <a href="mailto:alfredo.cuzzocrea@unical.it">alfredo.cuzzocrea@unical.it</a> >
9am-6pm	Workshop: Workshop on Cyber Threat Intelligence and Hunting with AI (CyberHunt 2021)	Vasileios Mavroeidis, <a href="mailto:vasileim@ifi.uio.no">vasileim@ifi.uio.no</a>
9am-6pm	Workshop: Computational Archival Science: digital records in the age of big data	Mark Hedges, <a href="mailto:mark.hedges@kcl.ac.uk">mark.hedges@kcl.ac.uk</a> , <a href="mailto:marciano@umd.edu">marciano@umd.edu</a> , <a href="mailto:v.lemieux@ubc.ca">v.lemieux@ubc.ca</a>
10: 15 - 12: 20 pm	S29: Contrastive Learning	Xingquan Zhu <a href="mailto:xzhu3@fau.edu">xzhu3@fau.edu</a>
	S25: Big Data Analytics and Modeling	Yaoqing Yang <a href="mailto:yqyang@berkeley.edu">yqyang@berkeley.edu</a>
	S23: Web Search and Recommendation	Saed Rezayi <a href="mailto:saedr@uga.edu">saedr@uga.edu</a>
	S17: Multi-structured Data - Big Variety Data	Rameshwar Pratap <a href="mailto:rameshwar.pratap@gmail.com">rameshwar.pratap@gmail.com</a>
	S5: Privacy, Interpretability, and Fairness	Luyi Ma <a href="mailto:LuyiMa1993@gmail.com">LuyiMa1993@gmail.com</a>
	I&G S4: IoT and Sensor Systems	Cheng Shi, <a href="mailto:shicheng720@gmail.com">shicheng720@gmail.com</a>
	Tutorial 7: Healthcare Big Data Analytics for COVID-19 Challenges: Digital Contact Tracing and Infodemic Management	Chee Wei Tan <a href="mailto:cheewtan@cityu.edu.hk">cheewtan@cityu.edu.hk</a>

10am-12:30pm	Predicting Victories in Video Games @KnowledgePit	Dominik Ślęzak, slezak@mimuw.edu.pl, Andrzej Janusz, andrzej.janusz@qed.pl
12: 20 - 1: 30 pm	<b>Lunch break</b>	
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>
11:00am-3:00pm	BigData Cup Challenges: IARAI's Science4cast Competition: A Machine Learning Challenges in the Science of Science	Mario Krenn mario.krenn@univie.ac.at
1: 30 - 3: 35 pm	S30: Big Data Learning and Analytics	Alessio Bernardo alessio.bernardo@polimi.it
	S19: Social, Web Search, and Mining	Shuhan Yuan shuhan.yuan@usu.edu
	S10: Big Transportation and Spatial Data	Rage Uday Kiran uday.rage@gmail.com
	S4: Health and Medical Data Applications	Hongfeng Yu yu@cse.unl.edu
	I&G S5: Machine Learning Applications 2	Haozhen Zhao, Haozhen.zhao@ankura.com
	NSF Information Session (Six NSF program officers)	Xingquan Zhu xzhu3@fau.edu
3: 35 - 3: 55 pm	<b>Coffee Break</b>	
1:00-6pm	Workshop: The 2nd International Workshop on Big Data Reduction (IWBDR-2)	Xin Liang, xliang@mst.edu
<b>Time</b>	<b>Sessions/Workshops</b>	<b>Session Chair</b>
3: 55 - 6: 00 pm	S31: Big Data Computing & Analytics (I)	Hiroto Akatsuka hiroto.akatsuka.fb@nttdocomo.com
	S21: Transfer and Domain Adaptation Learning	Kyumin Lee kmlee@wpi.edu
	S15: Big Data Uncertainty and Bias Reduction	Raghav Kulkarni kulraghav@gmail.com
	S9: Big Human-Derived Data	Heiko Ludwig hludwig@us.ibm.com
	S1: Mobility and Big Data (x full + x short)	Mahima Agumbe Suresh mahima.agumbesuresh@jsu.edu
	I&G S6: Data and Text Analytics	Qiwei Han, <a href="mailto:qiwei.han@novasbe.pt">qiwei.han@novasbe.pt</a>
	Tutorial 1: Optimization Meets Reinforcement Learning	Shaofeng Zou, Yi Zhou, and Yingbin Liang szou3@buffalo.edu, yi.zhou@utah.edu, yingbinliang@gmail.com

Day 4: Saturday, December 18, 2021		
9: 00 - 10: 00 am	<b>Keynote Speech 4: Building towards a Responsible Data Economy</b> Dr. Dawn Song <a href="mailto:dawnsong@cs.berkeley.edu">dawnsong@cs.berkeley.edu</a> Session Chair: Heiko Ludwig ( <a href="mailto:hludwig@us.ibm.com">hludwig@us.ibm.com</a> )	
10:00 – 10:15 am	Coffee Break	
Time	Sessions/Workshops	Session Chair
10:15 - 12: 20 pm	I&G S7: Deep Learning and Neural Networks	Pranjul Yadav, <a href="mailto:pranjulyadav@google.com">pranjulyadav@google.com</a>
	S2: Opinions, Contradiction, and Threat Detection	Wei Shang <a href="mailto:shangwei@amss.ac.cn">shangwei@amss.ac.cn</a>
	S20: Social, Web, Recommendation and Mining	David Mosallanezhad <a href="mailto:amosalla@asu.edu">amosalla@asu.edu</a>
	S26: Link and Graph Mining	Yicheng Tu <a href="mailto:tuy@mail.usf.edu">tuy@mail.usf.edu</a>
	S27: Cloud/Grid/Stream Data Computing	Tanu Malik <a href="mailto:tanu.malik@depaul.edu">tanu.malik@depaul.edu</a>
	S32: Big Data Computing & Analytics (II)	Zhaohui Wang <a href="mailto:zhwang7@gmail.com">zhwang7@gmail.com</a>
	Tutorial 5: Big Data Cleaning: Method Overview and Challenges	Hongzhi Wang and Xiaou Ding <a href="mailto:wangzh@hit.edu.cn">wangzh@hit.edu.cn</a> , <a href="mailto:dingxiao_hit@163.com">dingxiao_hit@163.com</a>
12: 20 - 12:45 pm Location:	Closing Remarks	

## Paper Sessions

S1: Mobility and Big Data (x full + x short)	
BigD690 "A Human Mobility Data Driven Hybrid GNN+RNN Based Model For Epidemic Prediction" (Regular)	Shohaib Mahmud, Haiying Shen, Ying Natasha Zhang Foutz, and Joshua Anton
BigD571 "Placement of DNN Models on Mobile Edge Devices for Effective Video Analysis" (Regular)	George Constantinou, Cyrus Shahabi, and Seon Ho Kim
BigD421 "Federated Meta-Location Learning for Fine-Grained Location Prediction" (Regular)	Xiaopeng Jiang, Shuai Zhao, Guy Jacobson, Rittwik Jana, Wen-Ling Hsu, Manoop Talasila, Syed Anwar Aftab, Yi Chen, and Cristian Borcea
BigD497 "FedTriNet: A Pseudo Labeling Method with Three Players for Federated Semi-supervised Learning" (Regular)	Liwei Che, Zewei Long, Jiaqi Wang, Yaqing Wang, Houping Xiao, and Fenglong Ma
BigD583 "City-Scale System-Optimal Route Planning with Route Replanning" (Short)	Robert Fitzgerald and Farnoush Banaei-Kashani
S2: Opinions, Contradiction, and Threat Detection	
BigD346 "WikiContradiction: Detecting Self-Contradiction Articles on Wikipedia" (Regular)	Cheng Hsu, Cheng-Te Li, Diego Saez-Trumper, and Yi-Zhan Hsu
BigD275 "A Multi-Platform Analysis of Political News Discussion and Sharing on Web Communities" (Regular)	Yuping Wang, Savvas Zannettou, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, and Gianluca Stringhini
BigD373 "Target-dependent Event Detection: A New Task to Event Extraction from News" (Regular)	Tiantian Zhang and Man Lan
BigD398 "Identification of Driver Phone Usage Violations via State-of-the-Art Object Detection with Tracking" (Short)	Steven Carrell and Amir Atapour-Abarghouei
BigD566 "Zero-day Malware Detection using Threshold-free Autoencoding Architecture" (Short)	Chiho Kim, Sang-Yoon Chang, Jonghyun Kim, Dongeun Lee, and Jinoh Kim
BigD460 "How Inclusive Are Wikipedia's Hyperlinks in Articles Covering Polarizing Topics?" (Short)	Cristina Menghini, Aris Anagnostopoulos, and Eli Upfal
S3: Big Data Interpretability & Trust	
BigD394 "Learning Interpretation Representations with Explainable Knowledge Distillation" (Regular)	Raed Alharbi, Minh Vu, and My T. Thai
BigD369 "InterpretableSAD: Interpretable Anomaly Detection in Sequential Log Data" (Regular)	Xiao Han, He Cheng, Depeng Xu, and Shuhan Yuan
BigD325 "On Exploring Attention-based Explanation for Transformer Models in Text Classification" (Regular)	Shengzhong Liu, Franck Le, Supriyo Chakraborty, and Tarek Abdelzaher
BigD466 "Interpretable Explanations for Probabilistic Inference in Markov Logic" (Short)	Khan Mohammad Al Farabi, Somdeb Sarkhel, Sanorita Dey, and Deepak Venugopal

BigD510 "Fed-EINI: An Efficient and Interpretable Inference Framework for Decision Tree Ensembles in Vertical Federated Learning" (Short)	Xiaolin Chen, Shuai Zhou, Kai Yang, Hao Fan, Yongji Wang, and Hu Wang
BigD464 "Towards Explainable Visual Emotion Understanding" (Short)	Yue Zhang, Wanying Ding, Ran Xu, and Xiaohua Hu
<b>S4: Health and Medical Data Applications</b>	
BigD264 "Improving Brain Dysfunction Prediction by GAN: A Functional-Connectivity Generator Approach" (Regular)	Da Yan, Shengbin Wu, Mirza Sami, Abdullateef Almudaifer, Zhe Jiang, Haiquan Chen, Rangaprakash Deshpande, Gopikrishna Deshpande, and Yuen Ma
BigD390 "Using Model Class Reliance to Measure Group Effects on Non-Adherence to Asthma Medication" (Short)	Vanja Ljevar, James Goulding, Gavin Smith, and Alexa Spence
BigD598 "To Reduce Healthcare Workload: Identify Critical Sepsis Progression Moments through Deep Reinforcement Learning" (Short)	Song Ju, Yeo Jin Kim, Markel Sanz Ausin, Maria Mayorga, and Min Chi
BigD702 "DeepTrack: An ML-based Approach to Health Disparity Identification and Determinant Tracking for Improving Pandemic Health Care" (Short)	Jinwei Liu, Long Cheng, Ankur Sarker, and Li Yan
BigD388 "MedTextSeg: A Deep Dual Sequential Model for Section Segmentation in Medical Reports" (Short)	Shaika Chowdhury, Halid Yerebakan, Yoshihisa Shinagawa, and Philip S. Yu
BigD521 "Multi-Temporal Abstraction with Time-Aware Deep Q-Learning for Septic Shock Prevention" (Short)	Yeo Jin Kim, Markel Sanz Ausin, and Min Chi
BigD645 "Spike2Vec: An Efficient and Scalable Embedding Approach for COVID-19 Spike Sequences" (Short)	Sarwan Ali and Murray Patterson
<b>S5: Privacy, Interpretability, and Fairness</b>	
BigD467 "Improving Semi-supervised Federated Learning by Reducing the Gradient Diversity of Models" (Regular)	Zhengming Zhang, Yaoqing Yang, Zhewei Yao, Yujun Yan, Joseph E. Gonzalez, Kannan Ramchandran, and Michael W. Mahoney
BigD321 "Fairness-aware Bandit-based Recommendation" (Short)	Wen Huang, Kevin Labille, Xintao Wu, Dongwon Lee, and Neil Heffernan
BigD356 "Proposing an Interactive Audit Pipeline for Visual Privacy Research" (Short)	Jasmine DeHart, Chenguang Xu, Lisa Egede, and Christan Grant
BigD597 "ExgFair: A Crowdsourcing Data Exchange Approach To Fair Human Face Datasets Augmentation" (Short)	Ziyi Kou, Lanyu Shang, Huimin Zeng, Yang Zhang, and Dong Wang
BigD223 "Adversarial representation learning for synthetic replacement of private attributes" (Short)	John Martinsson, Edvin Listo Zec, Daniel Gillblad, and Olof Mogren
BigD382 "Privacy Preserving Text Data Encoding and Topic Modelling" (Short)	Dinusha Vatsalan, Raghav Bhaskar, Aris Gkoulalas-Divanis, and Dimitrios Karapiperis
BigD410 "Exploring privacy-enhancing technologies in the automotive value chain" (Short)	Gonzalo Munilla Garrido, Kaja Schmidt, Christopher Harth-Kitzerow, Andre Luckow, and Florian Matthes
<b>S6: Big Data Visualization &amp; Analytics</b>	
BigD411 "NeuralCubes: Deep Representations for Visual Data Exploration" (Regular)	Zhe Wang, Dylan Cashman, Mingwei Li, Jixian Li, Matthew Berger, Joshua Levine, Remco Chang, and Carlos Scheidegger
BigD397 "A Visual Analytics Framework for Distributed Data Analysis Systems" (Regular)	Abdullah-Al-Raihan Nayeem, Mohammed Elshamakey, Todd Dobbs, Huikyo Lee, Daniel Crichton, Yimin Zhu, Chanachok Chokwitthaya, William J. Tolone, and Isaac Cho
BigD354 "Glance: A Generative Approach to Interactive Visualization of Voluminous Satellite Imagery" (Short)	Saptashwa Mitra, Daniel Rammer, Shrideep Pallickara, and Sangmi Pallickara

BigD527 "In Situ Adaptive Spatio-Temporal Data Summarization" (Short)	Soumya Dutta, Humayra Tasnim, Terece Turton, and James Ahrens
BigD409 "Predicting Financial Suspicious Activity Reports with Online Learning Methods" (Short)	Roxane Desrousseaux, Gilles Bernard, and Jean-Jacques Mariage
BigD694 "Augmentation of Body-in-White Dimensional Systems through Artificial Intelligence" (Short)	Carlos Escobar, Debejyo Chakraborty, Jorge Arinez, and Ruben Morales-Menedez
<b>S7: Big Data Systems and Search</b>	
BigD416 "A Systematic Approach to Efficiently Find All Cyclic Quorum Sets with All-pairs Property" (Regular)	Yiming Bian and Arun Somani
BigD519 "Streaming Sparse Graphs using Efficient Dynamic Sets" (Regular)	Brian Wheatman and Randal Burns
BigD569 "Efficient Scalable Temporal Web Graph Store" (Regular)	Duy-Khoi Vo, Sergej Zerr, Xiaofei Zhu, and Wolfgang Nejdl
BigD517 "A Parallel Chain Mail Approach for Scalable Spatial Data Interpolation" (Short)	Albert Asratyan, Sina Sheikholeslami, and Vladimir Vlassov
BigD634 "Multi-Objective Submodular Optimization with Approximate Oracles and Influence Maximization" (Short)	Xiaoyun Fu, Rishabh Rajendra Bhatt, Samik Basu, and A. Pavan
BigD489 "P-QALSH: Parallelizing Query Aware Locality-Sensitive Hashing for Big Data" (Short)	Yikai Huang, Zhili Yao, and Jianlin Feng
<b>S8: Big Data Model Attack, Bias, and Reliability</b>	
BigD436 "A Synergetic Attack against Neural Network Classifiers combining Backdoor and Adversarial Examples" (Regular)	Guanxiong Liu, Issa Khalil, Abdallah Khreishah, and NhatHai Phan
BigD417 "Resisting Distributed Backdoor Attacks in Federated Learning: A Dynamic Norm Clipping Approach" (Regular)	Yifan Guo, Qianlong Wang, Tianxi Ji, Xufei Wang, and Pan Li
BigD632 "A Multimodal Misinformation Detector for COVID-19 Short Videos on TikTok" (Regular)	Lanyu Shang, Ziyi Kou, Yang Zhang, and Dong Wang
BigD601 "gen2Out: Detecting and Ranking Generalized Anomalies" (Regular)	Meng-Chieh Lee, Shubhranshu Shekhar, Christos Faloutsos, T. Noah Hutson, and Leon Iasemidis
BigD615 "Evaluating Gender-Neutral Training Data for Automated Image Captioning" (Short)	Jack J Amend, Albatool Wazzan, and Richard Souvenir
<b>S9: Big Human-Derived Data</b>	
BigD620 "Human-like Explanation for Text Classification with Limited Attention Supervision" (Regular)	Dongyu Zhang, Cansu Sen, Jidapa Thadajarassiri, Thomas Hartvigsen, Xiangnan Kong, and Elke Rundensteiner
BigD484 "Reconstructing High-resolution Turbulent Flows Using Physics-Guided Neural Networks" (Regular)	Shengyu Chen, Shervin Sammak, Peyman Givi, Joseph P.Yurko, and Xiaowei Jia
BigD272 "c-Eval: A Unified Metric to Evaluate Feature-based Explanations via Perturbation" (Regular)	Minh Vu, Truc Nguyen, NhatHai Phan, Ralucca Gera, and My Thai
BigD434 "A text-based mining approach for real estate policy impact monitoring and analysis" (Short)	Lei Cao, Peng Xu, and Wei* Shang*
BigD516 "Cash flow prediction of a bank deposit using scalable graph analysis and machine learning" (Short)	Ryo Kawahara and Mikio Takeuchi
BigD662 "VarSKD: A Variational Student Knowledge Diagnosis for Efficient Representing Student Latent Knowledge Space" (Short)	Huan Dai, Yue Yun, Yupei Zhang, and Xuequn Shang
<b>S10: Big Transportation and Spatial Data</b>	
BigD376 "Co-Prediction of Station-based Multimodal Transportation Demands With Self-learned Spatiotemporal Dependence" (Regular)	Mingzhe Liu, Bowen Du, and Leilei Sun



BigD481 "Traffic Dispersion by Predicting Traffic Conditions based on Population Distribution" (Regular)	Hiroto Akatsuka, Yasunori Kamata, Tomohiro Nagata, Naoko Komiya, Makoto Goto, and Masayuki Terada
BigD532 "Citywide reconstruction of cross-sectional traffic flow from moving camera videos" (Short)	Ashutosh Kumar, Takehiro Kashiya, Hiroya Maeda, and Yoshihide Sekimoto
BigD529 "HastGCN: A Hierarchical Attention-Based Spatiotemporal Graph Convolutional Network for Traffic Incident Impact Forecasting" (Short)	Kaiqun Fu, Taoran Ji, Nathan Self, Zhiqian Chen, and Chang-Tien Lu
BigD588 "Traffic Flow Prediction using Recurrent Neural Network and Optimized Spatial Temporal Dependencies" (Short)	Zinat Ara and Mahdi Hashemi
BigD206 "A Hybrid Graph Convolutional Network For Dynamic Origin-Destination Prediction using a Large-Scale Public Transportation Data" (Short)	Juan Nathaniel and Baihua Zheng
<b>S11: Spatio-Temporal Data Applications</b>	
BigD266 "Context Integrated Relational Spatio-Temporal Resource Forecasting" (Regular)	Hongjie Chen, Ryan Rossi, Kanak Mahadik, and Hoda Eldardiry
BigD461 "Attention-augmented Spatio-Temporal Segmentation for Land Cover Mapping" (Regular)	Rahul Ghosh, Praveen Ravirathinam, Xiaowei Jia, Chenxi Lin, Zhenong Jin, and Vipin Kumar
BigD692 "Optimizing Bike-Share Repositioning: Networked Inventory Management with Spatiotemporal Modeling" (Regular)	Chunyi Liu, Hao Feng, Jiang Xu, Zhiwei Qin, and Hongtu Zhu
BigD401 "Influence of Land Use information over performance when predicting spatiotemporal electricity load demand" (Short)	Guillaume Habault, Shinya Wada, and Chihiro Ono
BigD433 "DDGNet: A Dual-Stage Dynamic Spatio-Temporal Graph Network for PM2.5 Forecasting" (Short)	Dong Li, Haomin Yu, Yangli-ao Geng, Xiaobao Li, and Qingyong Li
BigD647 "A Label Correction Algorithm Using Prior Information for Automatic and Accurate Geospatial Object Recognition" (Short)	Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H. Uhl, and Craig A. Knoblock
<b>S12: Theoretical Models for Big Data</b>	
BigD347 "Analyzing Data Selection Techniques with Tools from the Theory of Information Losses" (Regular)	Brandon Foggo and Nanpeng Yu
BigD593 "Detecting Gradual Structure Changes of Non-parametric Distributions via Kernel Complexity" (Regular)	So Hirai and Kenji Yamanishi
BigD685 "A Scalable Linear-Time Algorithm for Horizontal Visibility Graph Construction Over Long Sequences" (Regular)	Colin Stephen
BigD614 "YOLO-Green: A Real-Time Classification and Object Detection Model Optimized for Waste Management" (Short)	Wesley Lin
BigD677 "REAPS: Quasi-active Fault Tolerance for Big Data Publish-Subscribe Systems" (Short)	Hang Nguyen, MD YUSUF SARWAR UDDIN, and Nalini Venkatasubramanian
BigD562 "Approximate Bayesian Estimation with Subsampled Logistic Regression in Big Data Settings" (Short)	Peng Zhao and Guanyu Hu
<b>S13: Online, Temporal, and Time Series Systems</b>	
BigD414 "Concept Drift and Covariate Shift Detection Ensemble with Lagged Labels" (Regular)	Yiming Xu and Diego Klabjan
BigD589 "Discovering Maximal Partial Periodic Patterns in Very Large Temporal Databases" (Regular)	Uday Kiran RAGE, Likitha P, Veena P., Yukata Watanobe, and Koji Zettsu
BigD666 "Multi-feature Urban Traffic Prediction based on Unconstrained Graph Attention Network" (Regular)	Hangtao He, Kejiang Ye, and Cheng-Zhong Xu
BigD561 "Forecast-based Multi-aspect Framework for Multivariate Time-series Anomaly Detection" (Regular)	Lan Wang, Yusan Lin, Yuhang Wu, Huiyuan Chen, Fei Wang, and Hao Yang

BigD299 "End-to-End Performance Optimization for Training Streaming Convolutional Neural Networks using Billion-Pixel Whole-Slide Images" (Short)	Liang-Wei Tao, An-Fong Hwu, Yu-Jen Huang, Chi-Chung Chen, Chao-Yuan Yeh, and Shih-Hao Hung
BigD320 "Continual Neural Network Model Retraining" (Short)	Xiaofeng Zhu and Diego Klabjan
<b>S14: Theoretical and Computational Models for Big Data</b>	
BigD329 "MixNN: Combating Noisy Labels in Deep Learning by Mixing with Nearest Neighbours" (Regular)	Yangdi Lu and Wenbo He
BigD650 "Guided Generative Models using Weak Supervision for Detecting Object Spatial Arrangement in Overhead Images" (Regular)	Weiwei Duan, Yao-Yi Chiang, Stefan Leyk, Johannes H. Uhl, and Craig A. Knoblock
BigD358 "cSmartML: A Meta Learning-Based Framework for Automated Selection and Hyperparameter Tuning for Clustering" (Short)	Radwa El Shawi
BigD284 "Is Systematic Data Sharding able to Stabilize Asynchronous Parameter Server Training?" (Short)	Nikodimos Provatas, Ioannis Konstantinou, and Nectarios Koziris
BigD478 "Graph Structure-Aware Bi-directional Graph Convolution Model for Semantic Role Labeling" (Short)	Wenke Ding, Congcong Zhang, Gaofei Xie, Xiaojie Hu, Xiajiong Shen, and Yatian Shen
BigD617 "ELR-Net: An Enhancement and Location-Refinement Network for Co-Saliency Detection" (Short)	Haiyang He, Xiaolin Li, Zhihong Wang, and Shiguo Huang
BigD591 "Deep Sentence Denoising beyond Grammatical Error Correction" (Short)	Zhantong Liang and Abdou Youssef
<b>S15: Big Data Uncertainty and Bias Reduction</b>	
BigD631 "Cross-Concatenation: Tackling Uncertainty in Imbalanced Big Data Classification" (Regular)	Hadi Mansourifar and Weidong Shi
BigD296 "A Simple Approach to Balance Task Loss in Multi-Task Learning" (Regular)	Sicong Liang, Chang Deng, and Yu Zhang
BigD425 "HAR: Hardness Aware Reweighting for Imbalanced Datasets" (Regular)	Rahul Duggal, Scott Freitas, Sunny Dhamnani, Duen Chau, and Jimeng Sun
BigD366 "Improving Algorithmic Decision-Making in the Presence of Untrustworthy Training Data" (Short)	Wenting Qi and Charalampos Chelmis
BigD500 "Boosting the Training Time of Weakly Coordinated Distributed Machine Learning" (Short)	Erika Duriakova, Elias Tragos, Aonghus Lawlor, Barry Smyth, and Neil Hurley
BigD549 "How Shrinking Gradient Noise Helps the Performance of Neural Networks" (Short)	Zhen Deng, Jiaoyang Huang, and Kenji Kawaguchi
BigD599 "Uncertainty-Aware Multiple Instance Learning from Large-Scale Long Time Series Data" (Short)	Yuansheng Zhu, Weishi Shi, Deep Pandey, Yang Liu, Xiaofan Que, Daniel Krutz, and Qi Yu
<b>S16: Efficient and Scalable Data Management and Mining</b>	
BigD371 "Optimizing Performance of Parallel I/O Accesses to Non-contiguous Blocks in Multiple Array Variables" (Regular)	Qiao Kang, Scot Breitenfeld, Kaiyuan Hou, Wei-keng Liao, Robert Ross, and Suren Byna
BigD625 "CM-GCN: A Distributed Framework for Graph Convolutional Networks using Cohesive Mini-batches" (Regular)	Guoyi Zhao, Tian Zhou, and Lixin Gao
BigD555 "Building an Accessible, Usable, Scalable, and Sustainable Service for Scholarly Big Data" (Regular)	Jian Wu, Shaurya Rohatgi, Sai Raghav Reddy Keesara, Jason Chhay, Kevin Kuo, Arjun Manoj Menon, Sean Parsons, Bhuvan Uргаonkar, and C. Lee Giles

BigD270 "Tarema: Adaptive Resource Allocation for Scalable Scientific Workflows in Heterogeneous Clusters" (Regular)	Jonathan Bader, Lauritz Thamsen, Svetlana Kulagina, Jonathan Will, Henning Meyerhenke, and Odej Kao
BigD319 "FedLess: Secure and Scalable Federated Learning Using Serverless Computing" (Short)	Andreas Grafberger, Mohak Chadha, Anshul Jindal, Jianfeng Gu, and Michael Gerndt
BigD406 "Multi-step LRU: SIMD-based Cache Replacement for Lower Overhead and Higher Precision" (Short)	Hiroshi Inoue
S17: Multi-structured Data - Big Variety Data	
BigD459 "Feature Hashing with Insertion and Deletion of Features" (Regular)	Rameshwar Pratap, Hrushikesh Sudam Sarode, Suryakant Bhardwaj, and Raghav Kulkarni
BigD444 "Efficient Huge Page Management with Xpage" (Regular)	Wenqi Cao, Ling Liu, Gong Su, and Arun Iyengar
BigD526 "Scientific Formula Retrieval via Tree Embeddings" (Regular)	Zichao Wang, Mengxue Zhang, Richard Baraniuk, and Andrew Lan
BigD623 "Scalable Sequence Clustering for Large-Scale Immune Repertoire Analysis" (Regular)	Prem Bhusal, AKM Mubashwir Alam, Keke Chen, Ning Jiang, and Jun Xiao
BigD415 "Compressing large sample data for discriminant analysis" (Short)	Alexander Lapanowski and Irina Gaynanova
BigD353 "Scalable Approaches to Selecting Key Entities in Large Networked Infrastructure Systems" (Short)	Arun Sathanur and Arif Khan
S18: Complex Big Data Applications	
BigD389 "Language-Agnostic and Language-Aware Multilingual Natural Language Understanding for Large-Scale Intelligent Voice Assistant Application" (Regular)	Daniel Zhang, Jonathan Hueser, Yao Li, and Sarah Campbell
BigD596 "InferNet for Delayed Reinforcement Tasks: Addressing the Temporal Credit Assignment Problem" (Regular)	Markel Sanz Ausin, Hamoon Azizsoltani, Song Ju, Yeo Jin Kim, and Min Chi
BigD604 "Click-Based Student Performance Prediction: A Clustering Guided Meta-Learning Approach" (Regular)	Yun-Wei Chu, Elizabeth Tenorio, Laura Melissa Cruz Castro, Kerrie Douglas, Andrew Lan, and Christopher Brinton
BigD446 "A Data-Driven Approach to Nation-Scale Building Energy Modeling" (Short)	Andy Berres, Brett Bass, Mark Adams, Eric Garrison, and Joshua New
BigD463 "CalCROP21: A Georeferenced multi-spectral dataset of Satellite Imagery and Crop Labels" (Short)	Rahul Ghosh, Praveen Ravirathinam, Xiaowei Jia, Ankush Khandelwal, David Mulla, and Vipin Kumar
BigD498 "A Variational Bayesian Approach for Fast Adaptive Air Pollution Prediction" (Short)	Zhiyuan Wu, Ning Liu, Guodong Li, Xinyu Liu, Yue Wang, and Lin Zhang
S19: Social, Web Search, and Mining	
BigD268 "HAMLET: Hierarchical Attention-based Model with muLti-task sELf-Training for user profiling" (Regular)	Fuxin Ren, Zhongbao Zhang, Yang Yan, Zhi Wang, Sen Su, and Philip S. Yu
BigD323 "Hidden Buyer Identification in Darknet Markets via Dirichlet Hawkes Process" (Regular)	Panpan Zheng, Shuhan Yuan, Xintao Wu, and Yubao Wu
BigD370 "Understanding User Topic Preferences across Multiple Social Networks" (Regular)	Ziqing Zhu, Jiuxin Cao, Tao Zhou, Huiyu Min, and Liu Bo
BigD418 "From depression to suicidal discourse on Reddit" (Regular)	Benedetta Iavarone and Anna Monreale
BigD439 "A Latent Customer Flow Model for Interpretable Predictions of Check-In Counts" (Regular)	Daniel Tschernutter and Stefan Feuerriegel
S20: Social, Web, Recommendation and Mining	
BigD606 "Optimizing Iterative Algorithms for Social Network Sharding" (Regular)	Zishi Deng and Torsten Suel

BigD607 "Generating Topic-Preserving Synthetic News" (Regular)	Ahmadreza Mosallanezhad, Kai Shu, and Huan Liu
<b>BigD315 "Automatic Identification of Teachers in Social Media using Positive Unlabeled Learning" (Short)</b>	Hamid Karimi, Jiliang Tang, Xochitl Weiss, and Jiangtao Huang
BigD332 "On Minimal and Maximal High Utility Itemsets Mining using Propositional Satisfiability" (Short)	Amel Hidouri, Said Jabbour, Imen Ouled Dlala, and Badran Raddaoui
<b>BigD258 "An Effective, Robust and Fairness-aware Hate Speech Detection Framework" (Short)</b>	Guanyi Mou and Kyumin Lee
BigD664 "UGCLink: User Identity Linkage by Modeling User Generated Contents with Knowledge Distillation" (Short)	Hao Gao, Yongqing Wang, Jiangli Shao, Huawei Shen, and Xueqi Cheng
<b>S21: Transfer and Domain Adaptation Learning</b>	
BigD303 "Non-Parallel Text Style Transfer using Self-Attentional Discriminator as Supervisor" (Regular)	Kuan Feng, Yanmin Zhu, and Jiadi Yu
BigD568 "A Few Shot Transfer Learning Approach Identifying Private Images With Fast User Personalization" (Regular)	Edoardo Serra, Anna Squicciarini, Sujeet Ayyapureddi, and Qudrat Ratul
BigD263 "Multi-Source Domain Adaptation with Weak Supervision for Early Fake News Detection" (Short)	Yichuan Li, Kyumin Lee, Nima Kordzadeh, Brenton Faber, Cameron Fiddes, Elaine Chen, and Kai Shu
BigD380 "TransJury: Towards Explainable Transfer Learning through Selection of Layers from Deep Neural Networks" (Short)	Md Adnan Arefeen, Sumaiya Tabassum Nimi, MD YUSUF SARWAR UDDIN, and Yugyung Lee
BigD542 "Variational Open Set Recognition" (Short)	Luke Buquicchio, Walter Gerych, Kavin Chandrasekaran, Abdulaziz Alajaji, Hamid Mansoor, Thomas Hartvigsen, Elke Rundensteiner, and Emmanuel Agu
BigD636 "A Dropout Style Model Augmentation for Cross Domain Few-Shot Learning Given Naive Experience" (Short)	Pei-Cheng Tu and Hsing-Kuo Pao
BigD327 "Towards building a robust large-scale Bangla text recognition solution using a unique multiple-domain character-based document recognition approach" (Short)	AKM Shahariar Azad Rabby, Md. Majedul Islam, Nazmul Hasan, and Fuad Rahman
<b>S22: Recommendation and Social Networks</b>	
BigD680 "NEAT: A Label Noise-resistant Complementary Item Recommender System with Trustworthy Evaluation" (Regular)	Luyi Ma, Jianpeng Xu, Jason H.D. Cho, Evren Korpeoglu, Sushant Kumar, and Kannan Achan
BigD475 "Hidden Workforce: Analysis on Recognizing Unpaid Domestic Work Through Social Determinants" (Regular)	James Hicks and Thilanka Munasinghe
BigD324 "REFORM: Fast and Adaptive Solution for Subteam Replacement" (Short)	Zhaoheng Li, Xinyu Pi, Mingyuan Wu, and Hanghang Tong
BigD343 "Node-Polysemy Aware Recommendation by Matrix Completion with Side Information" (Short)	Bo Hui, Da Yan, and Wei-Shinn Ku
BigD480 "Two-stage Clustering Method for Discovering People's Perceptions: A Case Study of the Covid-19 Vaccine from Twitter" (Short)	Takako Hashimoto, Takeaki Uno, Yuka Takedomi, David Shepard, Toyoda Masashi, Naoki Yoshinaga, Masaru Kitsuregawa, and Ryota Kobayashi
BigD660 "How Useful is Meta-Recommendation? An Empirical Investigation" (Short)	Nassim Bouarour, Idir Benouaret, and Sihem Amer-Yahia
<b>S23: Web Search and Recommendation</b>	
BigD393 "GLOW : Global Weighted Self-Attention Network for Web Search" (Regular)	Xuan Shan, Chuanjie Liu, Yiqian Xia, Qi Chen, Yusi Zhang, Kaize Ding, Yaobo Liang, Angen Luo, and Yuxiang Luo
BigD602 "A Framework for Knowledge-Derived Query Suggestions" (Regular)	Saed Rezayi, Nedim Lipka, Vishwa Vinay, Ryan A. Rossi, Franck Dernoncourt, Tracy H. King, and Sheng Li

BigD331 "Temporal-aware QoS Prediction for Service Recommendation via Deep Neural Network and Feature Integration" (Regular)	Guobing Zou, Tengfei Li, Ming Jiang, Chenhong Cao, Bofeng Zhang, Yanglan Gan, and Yixin Chen
BigD657 "Tensor-based Complementary Product Recommendation" (Regular)	Negin Entezari, Evangelos Papalexakis, Haixun Wang, Sharath Rao, and Shishir Kumar Prasad
BigD584 "Distant-Supervised Slot-Filling for E-Commerce Queries" (Short)	Saurav Manchanda, Mohit Sharma, and George Karypis
<b>S24: Big Data Services and Applications</b>	
BigD652 "CoMe-KE: A New Transformers Based Approach for Knowledge Extraction on Conflict and Mediation Domain" (Regular)	Erick Skorupa Parolin, Yibo Hu, Latifur Khan, Javier Osorio, Patrick Brandt, and Vito D'Orazio
BigD273 "Scalable Data Management on Hybrid Memory System for Deep Neural Network Applications" (Regular)	Wei Rang, Donglin Yang, Zhimin Li, and Dazhao Cheng
BigD452 "Modelling of Destinations for Data-driven Pedestrian Trajectory Prediction in Public Buildings" (Short)	Andrew Kwok-fai Lui, Yin-hei Chan, and Man-fai Leung
BigD491 "Evaluating Actions in Sports Analytics with Deep Learning" (Short)	Dimitrios Klagkos and Vana Kalogeraki
BigD642 "AI-Driven Agent-Based Models to Study the Role of Vaccine Acceptance in Controlling COVID-19 Spread in the US" (Short)	Parantapa Bhattacharya, Dustin Machi, Jiangzhuo Chen, Stefan Hoops, Bryan Lewis, Henning Mortveit, Srinivasan Venkatramanan, Mandy L. Wilson, Achla Marathe, Przemyslaw Porebski, Brian Klahn, Joseph Outten, Anil Vullikanti, Dawen Xie, Abhijin Adiga, Shawn Brown, Christopher Barrett, and Madhav Marathe
BigD518 "TSLib: A unified Traffic Signal Controls framework using Deep Reinforcement Learning and Benchmarking" (Short)	Toan V. Tran, Thanh-Nam Doan, and Mina Sartipi
BigD592 "Solar Flare Forecasting with Deep Neural Networks using Compressed Full-disk HMI Magnetograms" (Short)	Chetraj Pandey, Rafal Angryk, and Berkay Aydin
<b>S25: Big Data Analytics and Modeling</b>	
BigD676 "Parallel Decoders Guided Lexically Constrained Response Generation" (Regular)	Daixuan Cheng, Haifeng Sun, Qi Qi, Jingyu Wang, and Yan Qi
BigD298 "Logit-based Uncertainty Measure in Classification" (Regular)	Huiyu Wu and Diego Klabjan
BigD687 "CLARA: A Constrained Reinforcement Learning Based Resource Allocation Framework for Network Slicing" (Regular)	Yongshuai Liu, Jiabin Ding, and Xin Liu
BigD238 "Keyword-based Topic Modeling and Keyword Selection" (Short)	Xingyu Wang, Lida Zhang, and Diego Klabjan
BigD557 "RLAC: Random Line Approximation Clustering" (Short)	Petros Barmpas, Aristidis Vrahatis, and Sotiris Tasoulis
BigD580 "Assessing Deep Neural Networks as Probability Estimators" (Short)	Yu Pan, Kwo-Sen Kuo, Michael Rilee, and Hongfeng Yu
<b>S26: Link and Graph Mining</b>	
BigD243 "Knowledge Graph Representation via Hierarchical Hyperbolic Neural Graph Embedding" (Regular)	Shen Wang, Xiaokai Wei, Cicero Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, Andrew Arnold, and Philip S. Yu
BigD582 "Schema-Aware Deep Graph Convolutional Networks for Heterogeneous Graphs" (Regular)	Saurav Manchanda, Da Zheng, and George Karypis

BigD653 "Graph Convolution Neural Network For Weakly Supervised Abnormality Localization In Long Capsule Endoscopy Videos" (Regular)	Sodiq Adewole, Philip Fernandes, James Jablonski, Andrew Copland, Michael Porter, Sana Syed, and Donald Brown
BigD426 "GraSSNet: Graph Soft Sensing Neural Networks" (Regular)	Yu Huang, Chao Zhang, Jaswanth Yella, Sergei Petrov, Xiaoye Qian, Yufei Tang, Xingquan Zhu, and Sthitie Bom
BigD531 "Anomaly Detection in Edge Streams Using Term Frequency-Inverse Graph Frequency (TF-IGF) Concept" (Short)	Prabin Lamichhane and William Eberle
<b>S27: Cloud/Grid/Stream Data Computing</b>	
BigD442 "Translation of Array-Based Graph Programs to Spark SQL on Block Arrays" (Regular)	Md Hasanuzzaman Noor and Leonidas Fegaras
BigD537 "Transparent Network Memory Storage for Efficient Container Execution in Big Data Clouds" (Regular)	Juhyun Bae, Ling Liu, KaHo Chow, Yanzhao Wu, Gong Su, and Arun Iyengar
BigD337 "Point cloud indexing using Big Data technologies" (Regular)	Kevin Kocon and Pascal Bormann
BigD469 "Cherry: A Distributed Task-Aware Shuffle Service for Serverless Analytics" (Regular)	Nikolaos Nikitas, Ioannis Konstantinou, Vana Kalogeraki, and Nectarios Koziris
BigD248 "Scaling the monitoring of approximate top-k queries in streaming windows" (Short)	Víctor Rampérez, Shima Zahmatkesh, and Emanuele Della Valle
BigD651 "A Framework for Decentralized Parallel Complex Event Processing on Heterogeneous Infrastructures" (Short)	Henriette Röger, Sukanya Bhowmik, and Tobias Linn
<b>S28: Graph Neural Network Learning</b>	
BigD267 "Object Interaction Recommendation with Multi-Modal Attention-based Hierarchical Graph Neural Network" (Regular)	Huijuan Zhang, Lipeng Liang, and Dongqing Wang
BigD361 "Pre-training Recommender Systems via Reinforced Attentive Multi-relational Graph Neural Network" (Regular)	Xiaohan Li, Zhiwei Liu, Stephen Guo, Zheng Liu, Hao Peng, Philip Yu, and Kannan Achan
BigD540 "Position-based Hash Embeddings For Scaling Graph Neural Networks" (Regular)	Maria Kalantzi and George Karypis
BigD586 "Improving Taxonomy-based Categorization with Categorical Graph Neural Networks" (Short)	Tianchuan Du, Keng-hao Chang, Paul Liu, and Ruofei Zhang
BigD413 "Graph Neural Networks to Predict Sports Outcomes" (Short)	Peter Xenopoulos and Claudio Silva
BigD293 "Graph Compression Networks" (Short)	Ting Guo, Xingquan Zhu, Yang Wang, and Fang Chen
<b>S29: Contrastive Learning</b>	
BigD628 "Contrastive Learning in Neural Tensor Networks using Asymmetric Examples" (Regular)	Maminur Islam, Somdeb Sarkhel, and Deepak Venugopal
BigD453 "Reducing and Exploiting Data Augmentation Noise through Meta Reweighting Contrastive Learning for Text Classification" (Regular)	Guanyi Mou, Yichuan Li, and Kyumin Lee
BigD441 "TRAJCROSS: Trajectory Cross-Modal Retrieval with Multi-level Contrastive Learning" (Short)	Quanliang Jing, Yao Di, Chang Gong, Xinxin Fan, Baoli Wang, Haining Tan, and Jingping Bi
BigD374 "CONAN: Contrastive Fusion Networks for Multi-view Clustering" (Short)	Guanzhou Ke, Zhiyong Hong, Zhiqiang Zeng, Zeyi Liu, Yangjie Sun, and Yunnan Xie
BigD682 "Session-based Recommendation via Contrastive Learning on Heterogeneous Graph" (Short)	Hangyue Li, Xucheng Luo, Qinze Yu, and Haoran Wang



BigD407 "Enhanced Seq2Seq Autoencoder via Contrastive Learning for Abstractive Text Summarization" (Short)	Rachel Zheng, Kunpeng Zhang, Harry Wang, Ling Fan, and Zhe Wang
S30: Big Data Learning and Analytics	
BigD474 "Model Selection in Local Approximation Gaussian Processes: A Markov Random Fields Approach" (Regular)	Hamed Jalali, Martin Pawelczyk, and Gjergji Kasneci
BigD603 "FlowDyn: A Dynamic Web-based Streamflow Prediction Pipeline" (Regular)	Nushrat Humaira, Sadegh Sadeghi Tabas, Vidya Samadi, and Nina Hubig
BigD226 "Weather radar echo prediction method based on recurrent convolutional neural network" (Regular)	Xiajiong Shen, Kunying Meng, Daojun Han, Kai Zhai, and Lei Zhang
BigD630 "Chronic Respiratory Disease: Risk Modeling Potential and Limitations" (Short)	Alexander He and Thilanka Munasinghe
BigD644 "Stock Price Prediction Leveraging Reddit: The Role of Trust Filter and Sliding Window" (Short)	Dennis Huynh, Garrett Audet, Nikolay Alabi, and Yuan Tian
BigD689 "A Predictive Framework for Multi-Horizon Financial Crises Forecasting using Macro-Economic Data" (Short)	Taruna Seth and Vipin Chaudhary
S31: Big Data Computing & Analytics (I)	
BigD242 "Supporting Data Compression in PnetCDF" (Regular)	KAI-YUAN HOU, Qiao Kang, Sunwoo Lee, Ankit Agrawal, Alok Choudhary, and Wei-keng Liao
BigD701 "Discovering Relative High Utility Itemsets in Very Large Transactional Databases Using Null-Invariant Measure" (Regular)	Uday Kiran Rage, Pradeep Pallikila, Luna J.M, Fournier-Viger Philippe, Toyoda Masashi, and Krishna Reddy P
BigD514 "Bench-Ranking: A First Step Towards Prescriptive Performance Analyses For Big Data Frameworks" (Regular)	Mohamed Ragab, Feras M. Awaysheh, and Riccardo Tommasini
BigD383 "MultiBlock: A Scalable Iterative Approach for Progressive Entity Resolution" (Regular)	Dimitrios Karapiperis, Aris Gkoulalas-Divanis, and Vassilios Verykios
BigD554 "Efficient and Distributed Temporal Pattern Mining" (Short)	Nguyen Ho, Van Long Ho, Torben Bach Pedersen, and Mai Vu
BigD696 "LDI: Learned Distribution Index for Column Stores" (Short)	Dai-Hai Ton That, Alexander Rasin, and Tanu Malik
S32: Big Data Computing & Analytics (II)	
BigD404 "Collectively Learned Multi-level Spatial Embeddings for Residential Rental Price Prediction" (Regular)	Wenlu Wang
BigD283 "An Analysis of Relation Extraction from Wet Lab Protocols" (Regular)	Xianjun Yang, Xinlu Zhang, Julia Zuo, Stephen Wilson, and Linda Petzold
BigD430 "Faster Depth Estimation for Situational Awareness on Urban Streets" (Regular)	Sanjana Srinivas and Mahima Agumbe Suresh
BigD443 "Multivariate Hawkes Processes for Incomplete Biased Data" (Regular)	Zihan Zhou and Mingxuan Sun
BigD512 "HyperASPO: Fusion of Model and Hyper Parameter Optimization for Multi-objective Machine Learning" (Regular)	Aswin Kannan, Anamitra Roy Choudhury, Vaibhav Saxena, Saurabh Manish Raje, Parikshit Ram, Ashish Verma, and Yogish Sabharwal
BigD671 "A study on the calibration of fingerprint classifiers" (Short)	Daniel Peralta, Lin Tang, Maxim Lippeveld, and Yvan Saey
BigD665 "Identifying Salient Entities of News Articles Using Binary Salient Classifier" (Short)	Nirupama Appikatala, SansWord Huang, Balachandar Sankar, Shweta Tripathi, and Eyan Goldman

## Industry and Government Program Sessions

I&G S1: Machine Learning Applications 1	
N216	Science-Guided Machine Learning for Wall-Modeled Large Eddy Simulation
N217	Chemicals Informatics: Explore Structural Factors and Potential Chemicals based on
N219	Public Literatures
N233	TripTracker: Unsupervised Learning of Fishing Vessel Routine Activity Patterns
	Qasar: Self-Supervised Learning Framework for Extractive Question Answering
I&G S2: CRM and Recommender Systems	
N203	Mitigating Divergence of Latent Factors via Dual Ascent for Low Latency Event
N212	Prediction Models
N225	A Real-World Implementation of UnbiasedLift-based Bidding System
N229	Dynamic Length Factorization Machines for CTR Prediction
N248	Multi-Task and Multi-Scene Unified Ranking Model for Online Advertising
	Predicting Loss Risks for B2B Tendering Processes
I&G S3: Platforms and Systems	
N209	Pathfinder: Building the Enterprise Data Map
N246	Efficient Contingency Analysis in Power Systems via Network Trigger Nodes
N221	An Advanced Open Data Platform for Integrated Support of Data Management,
N223	Distribution, and Analysis
N228	Online Distributed Maritime Event Detection & Forecasting over Big Vessel
N230	Tracking Data
	Move Real-Time Data Analytics to the Cloud: A Case Study on Heron to Dataflow
	Migration
	An AI-powered Smart Routing Solution for Payment Systems
I&G S4: IoT and Sensor Systems	
N210	Soft Sensing Transformer: Hundreds of Sensors are Worth a Single Word
N211	Soft Sensing Model Visualization: Fine-tuning Neural Network from What Model Learned
N213	Soft-Sensing ConFormer: A Curriculum Learning-based Convolutional Transformer
N252	Optimized One-vs-One Algorithm for Online Learning on Low-cost IoT Hardware
N241	CrowdQuake+: Data-driven Earthquake Early Warning via IoT and Deep Learning
I&G S5: Machine Learning Applications 2	
N236	Multi-Task Deep Learning Based Spatiotemporal Arctic Sea Ice Forecasting
N240	AutoGeoLabel: Automated Label Generation for Geospatial Machine Learning
N242	Virtual-SRE For Monitoring Large Scale Time-series Data
N251	FASCA: Framework for Automatic Scalable Acceleration of ML Pipeline
N238	A Systematic Literature Review on Machine Learning Approaches for Quality Monitoring
	and Control Systems for Welding Processes
I&G S6: Data and Text Analytics	
N201	Twitter Sparrow : Reduce Event Pipeline latency from hours to seconds
N218	Can we Estimate Truck Accident Risk from Telemetric Data using Machine Learning?
N222	Markdown Pricing For a Large Scale Retailer
N234	Efficient Multi-Lingual Sentence Classification Framework with Sentence Meta
N227	Encoders



	Mufin: Enriching Semantic Understanding of Sentence Embedding using Dual Tune Framework
<b>I&amp;G S7: Deep Learning and Neural Networks</b>	
N206	A Deep Graph Reinforcement Learning Model for Improving User Experience in Live
N226	Video Streaming
N243	Addressing Stability in Classifier Explanations
N250	Learning Code Representations Using Multifractal-based Graph Networks
N247	Online Advertising Revenue Forecasting: An Interpretable Deep Learning Approach
	Recent Efficiency Gains in Deep Learning: Performance and Sustainability

## Special Sessions

<b>4<sup>th</sup> Special Session on HealthCare Data</b> <i>Workshop Chairs: Sultan Turhan, Ozgun Pinarer</i>		
Time	Title	Presenter/Author
09:00 – 09:15	<b>Welcome</b>	
09:15 – 09:35	<i>SP01201 - GraphQL Patient Case Presentation using the Problem Oriented Medical Record Schema</i>	Sabah Mohammed
09:35 – 09:55	<i>SP01202 - Thick Data Analytics for Rating Ulcerative Colitis Severity Using Small Endoscopy Image Sample</i>	Jinan Fiaidhi
09:55 – 10:15	<i>SP01214 - A Lakehouse Architecture for the Management and Analysis of Heterogeneous Data for Biomedical Research and Mega-biobanks</i>	Edmon Begoli
10:15 – 10:35	<i>BigD287 - Multimodality Data Fusion for COVID-19 Diagnosis</i>	Jagadish D N
10:35 – 10:55	<i>SP01217 - Detecting Genetic Disposition of Ethnicity to Autoimmune Diseases via Clustering</i>	Damla Şentürk
10:55 – 11:25	<b>Coffee Break</b>	
11:25 – 11:45	<i>SP01209 - GAN For Generating User-Specific Human Activity Data From An Incomplete Training Corpus</i>	Walter Gerych
11:45 – 12:05	<i>SP01211 - Identifying Precursors to Long-Term Crisis in Veterans Using Associative Classifier</i>	Priyanka Annapureddy
12:05 – 12:25	<i>SP01220 - Detecting Transistor Defects in Medical Systems Using a Multi Model Ensemble of Convolutional Neural Networks</i>	Teoman Naskali
12:25 – 12:45	<i>SP01213 - Explainable image analysis for decision support in medical healthcare</i>	Roberto Corizzo
12:45 – 13:05	<i>BigD686 - COVID-19 Infection Detection Using Machine Learning</i>	Leo Wang
13:05 – 14:00	<b>Lunch Break</b>	
14:00 – 14:20	<i>SP01218 - Respiratory Rate Prediction Algorithm based on Pulse Oximeter</i>	Sultan Turhan
14:20 – 14:40	<i>SP01208 - MURAL: An unsupervised random forest-based embedding for electronic health record data</i>	Smita Krishnaswamy
14:40 – 15:00	<i>BigD262 - Study on prediction model of stroke risk based on decision tree and regression model</i>	Wang Yan
15:00 – 15:20	<i>SP01212 - Passive Smartphone Contact Tracing and Continuous COVID-19 Infection Risk Assessment</i>	Emmanuel Agu
15:20 – 15:40	<i>SP01204 - TBI2Vec: Traumatic Brain Injury Smartphone Sensing using AutoEncoder Embeddings</i>	Sayali Shelke
15:40 – 16:10	<b>Coffee Break</b>	
16:10 – 16:30	<i>SP01219 - 3D Mesh Model Generation from CT and MRI data</i>	Ozgun Pinarer
16:30 – 16:50	<i>SP01205 - Analysis of Gene Expression Cancer Data Set: Classification of TCGA Pan-cancer HiSeq Data</i>	Yusaku Nitta
16:50 – 17:10	<i>BigD326 - Modeling Influenza with a Forest Deep Neural Network Utilizing a Virtualized Clinical Semantic Network</i>	AKM Shahariar Azad Rabby
17:10 – 17:30	<i>BigD556 - Trusted Compliance Enforcement Framework for Sharing Health Big Data</i>	Dae-young Leroy Kim
17:30 – 17:50	<i>SP01216 - A Machine Learning Approach to Predict Length of Stay for Opioid Overdose Admitted Patients</i>	Praveen Madiraju
17:50 – 18:00	<b>Closing Remarks</b>	

<b>7<sup>th</sup> Special Session on Intelligent Data Mining</b> <i>Workshop Chair: Uraz Yavanoglu</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
09:00-09:15	SP02201- Classifying human rights violations using deepmulti-label co-training	<b>Ragini Kihlman</b> and Maria Fasli
09:15-09:30	SP02202 - On the combined effect of class imbalance and concept complexity in deep learning	<b>Kushankur Ghosh</b> , Colin Bellinger, Roberto Corizzo, Bartosz Krawczyk, and Nathalie Japkowicz
09:30-09:45	SP02203 - From Static to Dynamic Prediction: Wildfire Risk Assessment Based on Multiple Environmental Factors	<b>Tanqiu Jiang</b> , Sidhant Bendre, Hanjia Lyu, and Jiebo Luo
09:45-10:00	SP02204 - Fed xData: A Federated Learning Framework for Enabling Contextual Health Monitoring in a Cloud-Edge Network	<b>Khoa Anh Tran</b> , Do-Van Nguyen, Minh-Son Dao, and Koji Zettsu
10:00-10:15	SP02205 - Remaining Useful Life Estimation of Hard Disk Drives using Bidirectional LSTM Networks	<b>Austin Coursey</b> , Gopal Nath, Srikanth Prabhu, and Saptarshi Sengupta
10:15-10:30	SP02206 - Personalized Fashion Recommendation from Personal Social Media Data: An Item-to-Set Metric Learning Approach	<b>Haitian Zheng</b> , Kefei Wu, Jong-Hwi Park, Wei Zhu, and Jiebo Luo
10:30-10:45	SP02207 - Prediction of higher-order links using global vectors and Hasse diagrams	Kalpnil Anjan, <b>Willam Andreopoulos</b> , and Katerina Potika
10:45-11:00	SP02208 - Online Machine Learning from Imbalanced Data Streams in Real-world IoT	<b>Bharath Sudharsan</b> , John G. Breslin, and Muhammad Intizar Ali,
11:00-11:15	Coffee Break	
11:15-11:30	SP02209 - MM-trafficEvent: An Interactive Incident Retrieval System for First-view Travel-log Data	<b>Minh-Son Dao</b> , Dinh-Duy Pham, Phu Nguyen, Thanh-Binh Nguyen, and Koji Zettsu
11:30-11:45	SP02210 - Mining Learner-friendly Security Patterns from Huge Published Histories of Software Applications for an Intelligent Tutoring System in Secure Coding	<b>Sayem Mohammad Imtiaz</b> , Kazi Zakia Sultana, and Aparna Varde
11:45-12:00	SP02211 - Estimating Forest Canopy Height With Multi-Spectral and Multi-Temporal Imagery Using Deep Learning	<b>Stefan Oehmcke</b> , Thomas Nyegaard-Signori, Kenneth Grogan, and Fabian Gieseke

12:00-12:15	SP02212 - Towards Multi-class Sentiment Analysis With Limited Labeled Data	Md. Riyadh and <b>Omair Shafiq</b>
12:15-12:30	SP02213 - Can a machine win a Grammy? An evaluation of AI-generated song lyrics	<b>Junlan Lu</b> and Magdalini Eirinaki
12:30-12:45	BigD225 - Mi YouTube es Su YouTube? Analyzing the Cultures using YouTube Thumbnails of Popular Videos	<b>Songyang Zhang</b> , Tolga Aktas, and Jiebo Luo
12:45-1:00	BigD230 - cgSpan: Closed Graph-Based Substructure Pattern Mining	<b>Shaul Zevin</b> and Naaz Sheikh
1:00-1:15	BigD255 - Unsupervised Learning for Identifying High Eigenvector Centrality Nodes: A Graph Neural Network Approach	<b>Appan Rakaraddi</b> and Mahardhika Pratama
1:15-1:30	<b>Coffee Break</b>	
1:30-1:45	BigD289 - TTERCL: An onSite Real-time Alarm Root-Cause Location Algorithm	<b>Jianbing Ding</b> , Xidong Wang, Xiaozhou Ye, Ye Ouyang, and Yuanyuan Chai
1:45-2:00	BigD313 - Discovering Top-k Spatial High Utility Itemsets in Very Large Quantitative Spatiotemporal databases	<b>Pradeep Pallikila</b> , Veena P, Uday Kiran Rage, Ram Avatar, Sadanori Ito, Koji Zettsu, and Krishna Reddy P
2:00-2:15	BigD322 - Asynchronous Federated Learning for Sensor Data with Concept Drift	<b>Yujing Chen</b> , Zheng Chai, Yue Cheng, and Huzefa Rangwala
2:15-2:30	BigD501 - Understanding the Hoarding Behaviors during the COVID-19 Pandemic using Large Scale Social Media Data	<b>Xupin Zhang</b> , Hanjia Lyu, and Jiebo Luo
2:30-2:45	BigD541 - Embedding propagation over heterogeneous event networks for link prediction	<b>Paulo Carmo</b> and Ricardo Marcacini
2:45-3:00	<b>Coffee Break</b>	
3:00-3:15	BigD565 - Modeling Longitudinal Behavior Dynamics Among Extremist Users in Twitter Data	Priyadarshini Murugan, Younes Karimi, <b>Anna Squicciarini</b> , and Christopher Griffin
3:15-3:30	BigD574 - Learning Tabular Embeddings at Web Scale	<b>Sophie Pavia</b> , Montasir Shams, Rituparna Khan, Anna Pyayt, and Michael Gubanov
3:30-3:45	BigD654 - Analyzing Scientific Publications using Domain-Specific Word Embedding and Topic Modelling	Trisha Singhal, Junhua Liu, Lucienne T.M. Blessing, and <b>Kwan Hui Lim</b>
3:45-4:00	<b>Closing Remarks</b>	

Information Granulation in Data Science and Scalable Computing		
<i>Special Session Chairs: Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong and Shyue-Liang Wang</i>		
Time	Title	Presenter/Author
0900-0925	SP05210 Detection of Trajectory Outliers in Intelligent Transportation Systems	Usman Ahmed, Gautam Srivastava, Youcef Djenouri, Jimmy Ming-Tai Wu, and Jerry Chun-Wei Lin
0925-0950	SP05211 Analysis of Students' Concentration Levels for Online Learning Using Webcam Feeds	Linh Le, Ying Xie, Sumit Chakravarty, Michale Hales, John Johnson, and Tu Nguyen
0950-1015	SP05212 Toward automatic assessment of a risk of women's health disorders based on ontology decision models and menstrual cycle analysis	Łukasz Sosnowski and Jakub Wróblewski,
1015-1040	SP05209 Deep-learning-based Extraction of Electronic Component Parameters from Datasheets	Tzung-Pei Hong, Hsiu-Wei Chiu, Shih-Feng Huang, and Yi-Ting Chen
1040-1105	<b>Coffee Break</b>	
1105-1130	SP05205 NSPIS: Mining Negative Sequential Patterns with Individual Support	Gengsen Huang, Wensheng Gan, Shan Huang, Jiahui Chen, and Chien-Ming Chen
1130-1155	SP05203 Joint Utility and Frequency for Pattern Classification	Qi Lin, Wensheng Gan, Yongdong Wu, Jiahui Chen, and Chien-Ming Chen
1155-1230	SP05204 Targeted High-Utility Itemset Querying	Jinbao Miao, Shicheng Wan, Wensheng Gan, Jiayi Sun, and Jiahui Chen
1230-1400	<b>Lunch Break</b>	
1400-1425	SP05202 Mining On-shelf High-utility Quantitative Itemsets	Lili Chen, Wensheng Gan, Qi Lin, Jinbao Miao, and Chien-Ming Chen
1425-1450	SP5208 Empirical Rule Induction Methods Selection	Shusaku Tsumoto
1450-1525	SP5207 Granular Computing based Comparison of Agglomerative Clustering	Shusaku Tsumoto
1525-1530	<b>Closing Remarks</b>	

2021 Special Session on Machine Learning on Big Data (MLBD 2020) Program – 15-16 December 2021

MLBD 2021 Detailed Program Schedule		
Chair: Alfredo Cuzzocrea		
Wednesday 15 December, 2021		
Time	Title	Presenter/Author
08:55am – 09:00am	Session MLBD21_1: Session Opening Chair: Alfredo Cuzzocrea	
09:00am – 10:15am	Session MLBD21_2 Chair: TBA	
09:00am – 09:15am	SP03201 SMOTE-OB: Combining SMOTE and Online Bagging for Continuous Rebalancing of Evolving Data Streams	Alessio Bernardo and Emanuele Della Valle
09:15am – 09:30am	SP03202 Shallow Node Representation Learning using Centrality Indices	Masoud Malek, Mostafa Haghiri Chehrehgani, Ehsan Nazerfard, and Morteza Haghiri Chehrehgani

09:30am – 09:45am	<b>SP03203</b> ConvTab: A Context-Preserving, Convolutional Model for Ad-Hoc Table Retrieval	Akansha Bhardwaj, Vibhav Agarwal, Paolo Rosso, and Philippe Cudré-Mauroux
09:45am – 10:00am	<b>SP03204</b> Kalman Filtering for Learning with Evolving Data Streams	Giacomo Ziffer, Alessio Bernardo, Emanuele Della Valle, and Albert Bifet
10:00am – 10:15am	<b>SP03205</b> CUP: Cluster Pruning for Compressing Deep Neural Networks	Rahul Duggal, Cao Xiao, Richard Vuduc, Duen Chau, and Jimeng Sun
<b>10:15am – 11:30am</b>	<b>Session MLBD21_3</b> <b>Chair: TBA</b>	
10:15am – 10:30am	<b>SP03206</b> Constraining Weighted Word Co-occurrence Frequencies in Word Embeddings	Paula Lauren
10:30am – 10:45am	<b>SP03207</b> Implementing Efficient and Scalable In-Database Linear Regression in SQL	Patrick Giesser, Gabriel Stechschulte, Anna da Costa Vaz, and Michael Kaufmann
10:45am – 11:00am	<b>SP03208</b> Open Data Lake to Support Machine Learning on Arctic Big Data	Anifat Olawoyin, Carson Leung, and Alfredo Cuzzocrea
11:00am – 11:15am	<b>SP03210</b> DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning	Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu
11:15am – 11:30am	<b>SP03211</b> Identifying Malicious Users in the Offshore Leaks Networks via Structural Node Representation Learning	Brian Daley, Edoardo Serra, and Alfredo Cuzzocrea
<b>11:30am – 12:45pm</b>	<b>Session MLBD21_4</b> <b>Chair: TBA</b>	
11:30am – 11:45am	<b>SP03212</b> Evaluating Attribution Methods in Machine Learning Interpretability	Qudrat E Alahy Ratul, Edoardo Serra, and Alfredo Cuzzocrea
11:45am – 12:00pm	<b>SP03214</b> Privacy-Preserving Big Data Exchange: Models, Issues, Future Research Directions	Alfredo Cuzzocrea and Ernesto Damiani
12:00pm – 12:15pm	<b>BigD213</b> Does Dataset Complexity Matters for Model Explainers?	José Ribeiro, Raissa Silva, Lucas Cardoso, and Ronnie Alves
12:15pm – 12:30pm	<b>BigD214</b> Active Learning with Numerical Feature Annotation	Zhiye Fu, Hsing-Kuo Pao, and Jiabin He
12:30pm – 12:45pm	<b>BigD228</b> Graph Classification with Minimum DFS Code: Improving Graph Neural Network Expressivity	Jhalak Gupta and Arijit Khan
<b>12:45am – 02:00pm</b>	<b>Session MLBD21_5</b> <b>Chair: TBA</b>	
12:45am – 01:00pm	<b>BigD246</b> An Efficient and Flexible Inference System for Serving Heterogeneous Ensembles of Deep Neural Networks	Pierrick Pochelu, Serge G. Petiton, and Bruno Conche
01:00pm – 01:15pm	<b>BigD265</b> A Big Data Experiment to Evaluate the Effectiveness of Traditional Machine Learning Techniques Against LSTM Neural Networks in the Hotels Clients Opinion Mining	Thiago de Oliveira Lima, Methanias Colaço Júnior, Kleber H. de J. Prado, Israel Dias de J. Junior, and Frederico Resende de Oliveira
01:15pm – 01:30pm	<b>BigD274</b> Drivable Area Detection Using Deep Learning Models for Autonomous Driving	Donghao Qiao and Farhana Zulkernine
01:30pm – 01:45pm	<b>BigD288</b> Model-Centric and Data-Centric Aspects of Active Learning for Deep Neural Networks	John Daniel Bossér, Erik Sörstadius, and Morteza Haghir Chehreghani
01:45pm – 02:00pm	<b>BigD290</b> Overcoming Rare-Language Discrimination in Multi-Lingual Sentiment Analysis	Jasmin Lampert and Christoph H. Lampert
<b>Thursday 16 December, 2021</b>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>

<b>09:00am – 10:15am</b>	<b>Session MLBD21_6</b> <b>Chair: TBA</b>	
09:00am – 09:15am	<b>BigD309</b> Deep Neural Network to tradeoff between Accuracy and Diversity in a News Recommender System	Shaina Raza and Chen Ding
09:15am – 09:30am	<b>BigD314</b> Clustered Hierarchical Anomaly and Outlier Detection Algorithms	Najib Ishaq, Thomas Howard, and Noah Daniels
09:30am – 09:45am	<b>BigD362</b> Machine Learning Assisted Troubleshooting Flows	Heng-Ru May Tan and Lisa Pham
09:45am – 10:00am	<b>BigD364</b> AI Tool with Active Learning for Detection of Rural Roadside Safety Features	Hong Yi, Chris Bizon, David Borland, Matthew Watson, Matthew Satusky, Robert Rittmuller, Randa Radwan, Raghavan Srinivasan, and Ashok Krishnamurthy
10:00am – 10:15am	<b>BigD365</b> Hierarchical MultiClass AdaBoost	Charalampos Chelmis and Wenting Qi
<b>10:15am – 11:30am</b>	<b>Session MLBD21_7</b> <b>Chair: TBA</b>	
10:15am – 10:30am	<b>BigD378</b> Optimal Branch Location Finding for Cost effective Inference on Branchynet	Chang-Han Chiang, Pangfeng Liu, Da-Wei Wang, Ding-Yong Hong, and Jan-Jan Wu
10:30am – 10:45am	<b>BigD392</b> Scaling Deep Learning Data Management with Cassandra DB	Francesco Versaci and Giovanni Busonera
10:45am – 11:00am	<b>BigD405</b> Training Graph Neural Networks by GraphonEstimation	Siby Charley Pulikottil and Manish Gupta
11:00am – 11:15am	<b>BigD419</b> MapReduce Clustering for Big Data	Badih Ghattas, Antoine Pinto, and Sambou Diao
<b>11:30am – 12:45pm</b>	<b>Session MLBD21_8</b> <b>Chair: TBA</b>	
11:30am – 11:45am	<b>BigD427</b> Data Regression Framework for Time Series Data with Extreme Events	Yifan Zhang, Jiahao Li, Ablan Carlo, Alex Manda, Scott Hamshaw, Sergiu Dascalu, Frederick Harris, Jr., and Rui Wu
11:45am – 12:00pm	<b>BigD431</b> Bounding Means of Discrete Distributions	Eric Bax and Frederic Ouimet
12:00pm – 12:15pm	<b>BigD458</b> Margin-based Greedy Shapelet Search for Robust Time Series Classification of Imbalanced Data	Thomas Schlegl, Stefan Schlegl, Amélie Sciberras, Nikolai West, and Jochen Deuse
12:15pm – 12:30pm	<b>BigD502</b> Residual Connection based TPA-LSTM Networks for Cluster Node CPU Load Prediction	Zhaojun Yang, Lan Chen, He Zhang, and Zhenjie Yao
12:30pm – 12:45pm	<b>BigD520</b> Towards Unveiling Dark Web Structured Data	Montasir Shams, Sophie Pavia, Rituparna Khan, Anna Pyayt, and Michael Gubanov
<b>12:45am – 02:00pm</b>	<b>Session MLBD21_9</b> <b>Chair: TBA</b>	
12:45am – 01:00pm	<b>BigD559</b> Calibration of the GRACE-FO Satellite Platform Magnetometers and Co-Estimation of Intrinsic Time Shift in Data	Kevin Styp-Rekowski, Claudia Stolle, Ingo Michaelis, and Odej Kao
01:00pm – 01:15pm	<b>BigD560</b> Tracking Property Ownership Variance and Forecasting Housing Price with Machine Learning and Deep Learning	Junfeng Jiao, Seung Jun Choi, and Weijia Xu
01:15pm – 01:30pm	<b>BigD684</b> Benchmarking Robustness of Deep Learning Classifiers Using Two-Factor Perturbation	Wei Dai and Daniel Berleant
<b>02:00pm – 02:15pm</b>	<b>Session MLBD21_10: Session Closing</b> <b>Chair: Alfredo Cuzzocrea</b>	

2021 Special Session on Privacy and Security of Big Data  
(PSBD 2021) Program – 17 December 2021

<b>PSBD 2021 Detailed Program Schedule</b>		
<b>Chair: Alfredo Cuzzocrea</b>		
<b>Friday 17 December, 2021</b>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
<b>08:55am – 09:00am</b>	<b>Session PSBD21_1: Session Opening</b> <b>Chair: Alfredo Cuzzocrea</b>	
<b>09:00am – 10:15am</b>	<b>Session PSBD21_2</b> <b>Chair: TBA</b>	
09:00am – 09:15am	<b>SP04201</b> Preventing Attribute and Entity Disclosures: Combining k-Anonymity and Anatomy over RDF Graphs	Maxime Thouvenot, Olivier Curé, and Philippe Calvez
09:15am – 09:30am	<b>SP04202</b> DECADE - Deep Learning Based Content-hiding Application Detection System for Android	Mingming Peng, Max Khanov, Saikethi Reddy Madireddy, Hongmei Chi, Esra Akbas, and Gokila Dorai
09:30am – 09:45am	<b>SP04203</b> Bayesian Hyperparameter Optimization for Deep Neural Network-Based Network Intrusion Detection	Mohammad Masum, Hossain Shahriar, Hisham Haddad, Md Jobair Hossain Faruk, Mohammad Rahman, Muhaiminul Adnan, Alfredo Cuzzocrea, Md Abdullah Khan, Maria Valero, and Fan Wu
09:45am – 10:00am	<b>SP04204</b> Securing Kafka with Encryption-at-Rest	Chris Giblin, Sean Rooney, Pascal Vetsch, and Adrian Preston
10:00am – 10:15am	<b>SP04206</b> Deep Fraud Detection on Non-Attributed Graph	Chen Wang, Yingdong Dou, Min Chen, Jia Chen, Zhiwei Liu, and Philip Yu
<b>10:15am – 11:30am</b>	<b>Session PSBD21_3</b> <b>Chair: TBA</b>	
10:15am – 10:30am	<b>SP04207</b> Privacy-Preserving Publishing and Visualization of Spatial-Temporal Information	Anifat Olawoyin, Carson Leung, and Alfredo Cuzzocrea
10:30am – 10:45am	<b>SP04208</b> Malware Detection and Prevention using Artificial Intelligence Techniques	Md Jobair Hossain Faruk, Hossain Shahriar, Maria Valero, Farhat Lamia Borsha, Shahriar Sobhan, Md Abdullah Hafiz Khan, Michael Whitman, Alfredo Cuzzocrea, Dan Lo, Akond Rahman, and Fan Wu
10:45am – 11:00am	<b>SP04209</b> Unsupervised Risk for Privacy	Christophe Cerisara and Alfredo Cuzzocrea
11:00am – 11:15am	<b>SP04210</b> Android Malware Identification and Polymorphic Evolution via Graph Representation Learning	Miguel Quebrado, Edoardo Serra, and Alfredo Cuzzocrea
11:15am – 11:30am	<b>SP04211</b> Detecting Botnet Nodes via Structural Node Representation Learning	Justin Carpenter, Janet Layne, Edoardo Serra, and Alfredo Cuzzocrea
<b>11:30am – 12:45pm</b>	<b>Session PSBD21_4</b> <b>Chair: TBA</b>	
11:30am – 11:45am	<b>SP04212</b> Ride-Hailing for Autonomous Vehicles: Hyperledger Fabric-Based Secure and Decentralized Blockchain Platform	Ryan Shivers, Mohammad Rahman, Md Jobair Hossain Faruk, Hossain Shahriar, Alfredo Cuzzocrea, and Victor Clincy
11:45am – 12:00pm	<b>N255</b> A Methodology of Trusted Data Sharing across Telecom and Finance Sector under China's Data Security Policy	Yuxin Liang, Zhiyong Liu, Yong Song, Aidong Yang,



		Xiaozhou Ye, and Ye Ouyang
12:00pm – 12:15pm	<b>BigD318</b> Achieving Differential Privacy in Vertically Partitioned Multiparty Learning	Depeng Xu, Shuhan Yuan, and Xintao Wu
12:15pm – 12:30pm	<b>BigD445</b> Big Data Analytics Framework for Predictive Analytics using Public Data with Privacy Preserving	Duy Ho and Yugyung Lee
12:30pm – 12:45pm	<b>BigD496</b> Detecting Fake Points of Interest from Location Data	Syed Raza Bashir and Vojislav Misic
<b>12:45am – 01:00pm</b>	<b>Session PSBD21_5</b> <b>Chair: TBA</b>	
01:00pm – 01:15pm	<b>BigD667</b> Classifying Android Applications Via System Stats	Joshua Hightower, William Bradley Glisson, Ryan Benton, and J. Todd McDonald
<b>01:00pm – 01:15pm</b>	<b>Session PSBD21_6: Session Closing</b> <b>Chair: Alfredo Cuzzocrea</b>	

## BigData Cup Challenges

<b>Reinforcement Learning based Recommender System Challenge</b> <i>WorkshopChairs: Runze Wu, wurunze1@corp.netlease.com</i>		
Time	Title	Presenter/Author
20:00-20:15	SC02207 :Compressive Features in Offline Reinforcement Learning for Recommender Systems	Minh Pham, Hung Nguyen, Long Dang, and Jennifer Adorno Nieves
20:15-20:30	SC02204 :Deep Reinforcement Learning based Recommender System with State Representation	Peng Jiang, Jiafeng Ma, and Jianming Zhang
20:30-20:45	SC02206 :A Combinatorial Recommendation System Framework Based on Deep Reinforcement Learning	Fei Zhou, Biao Luo, Tianmeng Hu, Zihan Chen, and Yilin Wen
20:45-21:00	SC02202 :A DQN-based Recommender System for Item-list Recommendation	Haicheng Chen
<b>Coffee Break</b>		
21:10-21:25	SC02201 :Session-aware Item-combination Recommendation with Transformer Network	Tzu-Heng Lin and Chen Gao
21:25-21:40	SC02205 : A multi-step decision prediction model based on LightGBM	Yuhao Luo, Qianfang Xu, Wenliang Li, Feng Jiang, and Bo Xiao
21:40-21:55	SC02203 :Interest-aware Item Combination Prediction	Shiwen Wu and Jian Zhao
<b>Closing Remarks</b>		

<b>Predicting Victories in Video Games @KnowledgePit</b> <i>WorkshopChairs: Dominik Ślęzak, slezak@mimuw.edu.pl</i>		
Time	Title	Presenter
10:00 - 10:05	Session start – introduction	Dominik Ślęzak
10:05 – 10:40	SC01205 : Predicting Victories in Video Games - IEEE BigData 2021 Cup Report	Andrzej Janusz
10:40 – 11:05	SC01201 :WP-GBDT: An Approach for Winner Prediction using Gradient Boosting Decision Tree	Haitao Xiao
11:05 – 11:30	SC01202 :A combination of general and specific models to predict victories in video games	Quang Hieu Vu
11:30 – 11:55	SC01203: Predicting Victories in Video Games: Using Single XGBoost with Feature Engineering	Kai-Shen Tseng
11:55 – 12:20	SC01204 :Automated feature engineering for prediction of victories in online computer games	Ming Liu
12:20 – 12:30	<b>Closing Remarks</b>	

<b>Soft Sensing at Scale -Seagate</b> <i>WorkshopChairs: Sthitie Bom &amp; Chao Zhang, chaozhang@seagate.com, sthitie.e.bom@seagate.com</i>		
--	--	--

Time	Title	Presenter/Author
8:00-8:15	Keynote	Dr. Matthew Johnson
8:15-8:35	Overview	Chao Zhang
8:35-8:55	Presentation 1	Bodhi Satwa Mishra
8:55-9:15	Presentation 2	Ali Abdin
9:15-9:35	Presentation 3	Neeta Shinde
9:35-10:00	Summary and discussion	all

<b>Weather4cast Competition</b> <i>Workshop Chairs: David Kreil, Michael Kopp</i> <i>Aleksandra Gruca <a href="mailto:aleksandra.gruca@iarai.ac.at">aleksandra.gruca@iarai.ac.at</a>, <a href="mailto:weather4cast@iarai.ac.at">weather4cast@iarai.ac.at</a></i>		
Time	Title	Presenter/Author
11:00	Keynote talk: Insights on end-to-end weather forecasting	Nal Kalchbrenner, Google Brain, Netherlands
11:40	First insights on transfer learning from the Weather4cast Competitions 2021	Pedro Herruzo, Institute of Advanced Research in Artificial Intelligence, Austria
12:00	Improvements to short-term weather prediction with recurrent-convolutional networks	Jussi Leinonen, MeteoSwiss, Switzerland
12:20	UNet based future weather prediction on Weather4cast 2021 stage 2	Sungbin Choi, South Korea
<b>12:40</b>	<b>Coffee break (IARAI topia world)</b>	
1:00	Invited talk: The NWC SAF satellite derived weather nowcasting & data source for ML	Llorenç Lliso, AEMET
1:20	3rd prize: Enhanced variational U-Net for weather forecasting	Pak Hay Kwok and Qi Qi, UK
1:40	Submission: Spatiotemporal Vision Transformer for short time weather forecasting	Alabi Bojesomo, Khalifa University, UAE
2:00	Closing session keynote talk: Skilful precipitation nowcasting using deep generative models of radar	Piotr Mirowski, DeepMind, UK
<b>2:40</b>	<b>Get together (IARAI topia world)</b>	

## Workshop Detailed Schedule

Workshop on Cyber Threat Intelligence and Hunting with AI (CyberHunt 2021)		
<b>Workshop Chair:</b> Vasileios Mavroeidis <b>PC Chair:</b> Andre Büttner		
Time (EST)	Title	Presenter/Author
09:00-09:05	Opening Remarks	
09:05-09:45	Keynote Presentation: What is (cyber)secure enough?	Simen Bakke (Senior information security advisor, Norwegian Police ICT-services, Focus area: Risk management)
09:45-10:05	Vulnerabilities Manager, a platform for linking vulnerability data sources (S01207)	Konstantinos Touloumis
10:05-10:25	Effects of Demographics and Photometric Normalization on Image Translation GANs for Cross-Spectral Face Recognition (S01211)	Thirimachos Bourlai
10:25-10:45	Words of Suicide: Identifying Suicidal Risk in Written Communications (S01213)	Amendra Shrestha
10:45-11:05	PCAT: PLC Command Analysis Tool for automatic incidence response in Water Treatment Plants (S01222)	Siddhant Shrivastava
11:05-11:30	Coffee Break	
11:30-11:50	A Machine Learning Approach to Detection of Critical Alerts from Imbalanced Multi-Appliance Threat Alert Logs (S01206)	Samuel Ndichu
11:50-12:10	A Survey of Using Process Data and Features of Industrial Control Systems in Intrusion Detection (S01212)	Jon-Martin Storm
12:10-12:30	Detection of Advanced Persistent Threats using Artificial Intelligence for Deep Packet Inspection (S01218)	Allard Dijk
12:30-12:50	Comparing Ransomware using TLSh and @DisCo Analysis Frameworks (S01214)	Rafer Cooley
12:50-13:10	Malware Generation with Specific Behaviors to Improve Machine Learning-based Detection (S01219)	Michael Smith
13:10-13:30	Coffee Break	
13:30-13:50	On the Integration of Course of Action Playbooks into Shareable Cyber Threat Intelligence (S01217)	Vasileios Mavroeidis
13:50-14:10	Inside the Matrix: CTI Frameworks as Partial Abstractions of Complex Threats (S01208)	Kris Oosthoek
14:10-14:30	Threat Hunting as a Similarity Search Problem on Multi-positive and Unlabeled Data (S01209)	Tomáš Komárek
14:30-14:50	Polymer: An Adaptive Kill Chain Expanding Cyber Threat Hunting to Multi-Platform Environments (S01204)	Antonio Jose Horta Neto
14:50-15:00	Closing Remarks	

Computational Archival Science		
<i>Workshop Chairs: Mark Hedges, Victoria Lemieux, Richard Marciano</i> <b>Friday 17 December</b>		
Time	Title	Presenter/Author

9:00 – 9:15	Welcome	Mark Hedges (King's College London), Victoria Lemieux (U. British Columbia), Richard Marciano (U. Maryland)
9:15 – 10:00	Keynote	Michael Kurtz
<b>10:00 – 10:40</b>	<b>SESSION 1: Infrastructures and services</b>	
10:00 – 10:20	#1: Computational Archival Science is a Two-Way Street (S02208)	B. Ambacher, M. Conrad
10:20 – 10:40	#2: Managing Records in Enterprise Resource Planning Systems (S02201)	S. Katuu
<b>10:40 – 11:00</b>	<b>Coffee Break (coffee not supplied!)</b>	
<b>11:00 – 12:40</b>	<b>SESSION 2: Applying analytics to archival materials</b>	
11:00 – 11:20	#3: Using Transfer Learning to contextually Optimize Optical Character Recognition (OCR) output and perform new Feature Extraction on a digitized cultural and historical dataset (S02212)	A. Inbasekaran, R. Kumar Gnanasekaran, R. Marciano
11:20 – 11:40	#4: Using AI/Machine Learning to Extract Data from Japanese American Confinement records (S02204)	M. Friedman, C. Ford, M. Elings, V. Singh, T. Tan
11:40 – 12:00	#5: A Framework for Unlocking and Linking WWII Japanese American Incarceration Biographical Data (S02203)	L. Beltran, E. Ping O'Brien, G. Jansen, R. Marciano
12:00 – 12:20	#6: Computational Curation and the Application of Large-Scale Vocabularies (S02214)	S. Grabus, J. Greenberg
12:20 – 12:40	#7: Inference of Absolute Time Value from Temporal Expressions (S02207)	J. Sung, S. Mori, H. Kameko, A. Kubo, T. Sekino
<b>12:40 – 1:40</b>	<b>Lunch Break (lunch not supplied!)</b>	
<b>1:40 – 2:00</b>	<b>Discussion</b>	
<b>2:00 – 3:20</b>	<b>SESSION 3: Analytics to support archival processing</b>	
2:00 – 2:20	#8: EMCODIST: A Context-based Search Tool for Email Archives (S02205)	S. Kuppili Venkata, S. Decker, D. Kirsch, A. Nix
2:20 – 2:40	#9: An AI-Assisted Framework for Rapid Conversion of Descriptive Photo Metadata into Linked Data (S02202)	J. Proctor, R. Marciano
2:40 – 3:00	#10: Organizing a Content Profile for a Large, Heterogeneous Collection of Interactive Projects (S02211)	E. Kaltman, R. Lorelli, A. Larson, E. Wolfe
3:00 – 3:20	#11: Constructing Archives of Population Dynamics and Migration Network as a Gateway to Get Deep Insights into Hard-to-reach Populations: Research on Taiwan Indigenous Peoples (S02206)	Ji-Ping Lin
<b>3:20 – 3:40</b>	<b>Coffee Break (coffee not supplied!)</b>	
<b>3:40 – 4:20</b>	<b>SESSION 4: New forms of archives</b>	
3:40 – 4:00	#12: NFTs: Tulip Mania or Digital Renaissance? (S02210)	D. Ross, E. Cretu, V. Lemieux
4:00 – 4:20	#13: Making Case for Using RAFT in Healthcare Through Hyperledger Fabric (S02209)	A. Alexandridis, G. Al-Sumaidace, R. Alkhudary, Z. Zilic
<b>4:20 – 5:00</b>	<b>SESSION 5: Discussion</b>	
<b>5:00 – 5:20</b>	<b>Closing Remarks (Organizers)</b>	

<b>7th International Workshop on Methods to Improve Big Data Science Projects</b>		
<i>Workshop Chair: Jeff Saltz</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00am	<i>A survey study of success factors in data science projects (S03204)</i>	Iñigo Martinez, Elisabeth Viles, and Igor G Olaizola
9:20am	<i>The CRISP-ML Approach to Handling Causality and Interpretability Issues in Machine Learning (S03205)</i>	Inna Kolyshkina and Simeon Simoff
9:40am	<i>Managing and Composing Teams in Data Science: An Empirical Study (S03206)</i>	Timo Aho, Terhi Kilamo, Lucy Lwakatare, Tommi Mikkonen, Outi Sievi-Korte, and Sezin Yaman
10:00am	<i>Identifying and Addressing 6 Key Questions when Using Data Driven Scrum (S03209)</i>	Jeffrey Saltz, Alex Sutherland, and Thibaut Jombart

10:20am – 10:40am	<b>Coffee Break</b>	
10:40am	<i>Big Data Testing Framework for Recommendation Systems in e-Science and e-Commerce Domains (S03201)</i>	Meryem Uzun-Per, Ali Burak Can, Ahmet Volkan Gürel, and Mehmet S. Aktaş
11:00am	<i>An Approach to Support Human-in-the-Loop Big Data Software Development Projects (S03208)</i>	Nathalia Nascimento, Paulo Alencar, and Donald Cowan
11:20am	<i>A Socio-Technical Framework for Measuring Organizational Capacity During Cyberinfrastructure Diffusion (S03203)</i>	Kerk Kee, Alex Olshansky, and Shan Xu
11:40am	<i>Unsupervised Anomaly Detection in Data Quality Control (BigD511)</i>	Lex Poon, Siamak Farshidi, Na Li, and Zhiming Zhao
<b>Noon</b>	<b>Closing Remarks</b>	

<b>BigData Engineering and Analytics in Cyber-Physical Systems</b> <i>Session/Workshop Chair: Dr. Akbar Namin (Texas Tech University)</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 – 9:30	S05205 – A Lifecycle for Engineering IoT Neural Network-based Systems	Nathalia Nascimento, Paulo Alencar, Donald Cowan
9:30 – 10:00	S05203 – A Comparison of TCN and LSTM Models in Detecting Anomalies in Time Series Data	Saroj Gopali, Faranak Abri, Akbar Siami Namin
10:00 – 10:30	S05202 – Detecting Anomalies using Overlapping Electrical Measurements in Smart Power Grids	Sina Sontowski, Nigel Lawrence, Deepiyoti Deka, and Maanak Gupta
10:30 – 11:00	S05204 – The Applications of BlockChain in Addressing the Integration and Security of IoT Systems: A Survey	Zulfigar Khan and Akbar Siami Namin
11:00 – 11:30	S05201 – Identifying Social Media Content Supporting Proud Boys	MD Fahim and Swapna Gokhale
	<b>Closing Remarks</b>	

<b>Deviant Activities on Social Media (DEVIANCE 2021)</b> <i>Workshop Chairs: Drs. Nitin Agarwal, Esther Mead, Nihal Hussain, Samer Al-Khateeb, Kiran Bandeli</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00am-9:15am	Introduction to the Workshop	Workshop Chairs
	<b>Track 1</b>	
9:15am - 9:40am	BigD402 - Am I Being Bullied on Social Media? An Ensemble Approach to Categorize Cyberbullying	Mohsinul Kabir
9:40am - 10:05am	BigD396 - Hate or Non-hate: Translation based hate speech identification in Code-Mixed Hinglish data set	Shankar Biradar
10:05am - 10:30am	S06205 - An Implementation of Fake News Prevention by Blockchain and Entropy-based Incentive Mechanism	Chien-Chih Chen
10:30am - 10:55am	S06204 - Identifying the Popularity and Persuasiveness of Right- and Left-Leaning Group Videos on Social Media	Julia Hirschberg
10:55am - 11:20am	S06207 - MetaHate: A Meta-Model for Hate Speech Detection	Daniel G. Kyrollos
11:20am - 12:30pm	<b>Lunch Break</b>	
	<b>Track 2</b>	
12:30pm - 12:55pm	S06210 - If it retweets like a troll, follows like a troll and looks like a troll; is it a troll?: Targeting the 2021 Ecuadorian Presidential Elections Trolls	Sariah López-Fierro
12:55pm - 1:20pm	S06201 - Beyond a binary of (non)racist tweets: A four-dimensional categorical detection and analysis of racist and xenophobic opinions on Twitter in early Covid-19	Deval Mehta
1:20pm - 2:25pm	<b>Keynote</b>	Dr. Anatoliy Gruzd
2:25pm - 2:40pm	<b>Short Break</b>	
	<b>Track 2 (continued)</b>	
2:40pm - 3:05pm	S06206 - Identifying Social Media Content Supporting Proud Boys	Swapna Gokhale

3:05pm - 3:30pm	S06208 - Internet Trolls against Russian Opposition: A Case Study Analysis of Twitter Disinformation Campaigns against Alexei Navalny	Iuliia Alieva
3:30pm - 3:45pm	<b>Closing Remarks</b>	

*The 5<sup>th</sup> International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention*

Thursday, December 15<sup>th</sup>, 2021; Location: online (USA Eastern Standard time zone)

Workshop Chairs: **Andrii Shalaginov, Asif Iqbal, Igor Kotsiuba, Mamoun Alazab**

Kristiania University College, Norway

KTH Royal Institute of Technology in Stockholm, Sweden

Pukhov Institute for Modelling in Energy Engineering, Ukraine

Charles Darwin University, Australia

Time (EST)	Time (CET)	Title of the talk	Presenter/Author
<b>09:00-09:05</b>	<b>15:00-15:05</b>	<b>Opening Remarks and Welcome</b>	Andrii Shalaginov
09:05-09:30	15:05-15:30	<i>S08201: Template Clustering for the Foundational Analysis of the Dark Web</i>	Viswajit Vinod Nair, Dion Oosterman, and Mark van Staalduijn
09:30-09:55	15:30-15:55	<i>S08202: A Convolutional Autoencoder Based Method with SMOTE for Cyber Intrusion Detection</i>	Xinyi She and Yuji Sekiya
09:55-10:20	15:55-16:20	<i>S08203: Privacy-Preserving Framework to Facilitate Shared Data Access for Wearable Devices</i>	Dane Troyer, Justin Henry, Hoda Maleki, Gokila Dorai, Bethany Sumner, Gagan Agrawal, and Jon Ingram
10:20-10:45	16:20-16:45	<i>S08204: Gun Violence News Information Retrieval using BERT as Sequence Tagging Task</i>	Hung-Yeh Lin, Teng-Sheng Moh, and Bryce Westlake
<b>10:45-11:00</b>	<b>16:45-17:00</b>	<b>Coffee Break</b>	
11:00-11:25	17:00-17:25	<i>S08211: Securing Smart Future: Cyber Threats and Intelligent Means to Respond</i>	Andrii Shalaginov and Tor-Morten Grønli
11:25-11:50	17:25-17:50	<i>S08208: Pattern Recognition and Reconstruction: Detecting Malicious Deletions in Textual Communications</i>	Abiodun Abdullahi SOLANKE, Xihui Chen, and Yuniior Ramirez-Cruz
11:50-12:15	17:50-18:15	<i>S08210: A Semi Supervised Multi-relational Graph Convolutional Network for Uncovering Key Actors on Hackforums</i>	Nikita Saxena and Vinti Agarwal
12:15-12:40	18:15-18:40	<i>S08207: AMoC: A Multifaceted Machine Learning-based Toolkit for Analysing Cybercriminal Communities on the Darknet</i>	Chao Chen, Claudia Peersman, Matthew Edwards, Ziauddin Ursani, and Awais Rashid
<b>12:40-13:00</b>	<b>18:40-19:00</b>	<b>Coffee Break</b>	
13:00-13:30	19:00-19:30	<b><i>Tales from the Cloud</i></b>	Keynote speaker: Ashwin Ram, Check Point Software Technologies, Australia
13:30-13:55	19:30-19:55	<i>N202: Digital Forensics Process of an Attack Vector in ICS environment</i>	Todd Mason and Bing Zhou
13:55-14:00	19:55-20:00	<b>Closing Remarks</b>	Andrii Shalaginov

**Discriminative Pattern Mining Workshop**

Workshop Chairs: Ryan Benton, Tom Johnsten and Suresh Choubey

Ryan Benton <[rbenton@southalabama.edu](mailto:rbenton@southalabama.edu)>

Time	Title	Presenter/Author
9:00-9:10	Welcome	Ryan Benton
9:10 – 9:40	BigD659: “Scotty: Fast a priori Structure-based Extraction from Time Series”	Athanasios Tsitsipas, Pascal Schiessle, and Lutz Schubert

9:40 – 10:10	BigD655: “RARE: Rare Action Rule Exploration”	Blake Johns, Ryan Benton, Tom Johnsten, and David Bourrie
10:10 – 10:40	BigD533: “A Parallelized DynTARM Algorithm for the Discovery of Predictive Co-Occurrences Within Streaming Time Stamped Data”	Matthew Higginbotham, Ethan Franks, Pratchi Roy, and Jennifer Lavergne
10:40-10:55	<b>Coffee Break</b>	
10:55 – 11:25	N235: “A Scalable Multi-factor Fault Analysis Framework for Information Systems”	Hong Hai Phan Vu, Brigitte Jaumard, Tristan Glatard, Justin Whatley, and Sylvain Nadeau
11:25 – 11:35	Closing Remarks	Tom Johnsten
	<b>Closing Remarks</b>	

<b>GTA<sup>3</sup> 5.0: The 5<sup>th</sup> workshop on Graph Techniques for Adversarial Activity Analytics</b> <i>Workshop Chairs: Jiejun Xu, Hanghang Tong, Andrea Bertozzi, Vince Lyzinski, George Chin, Hang Liu</i>		
<b>Time (US EST)</b>	<b>Title</b>	<b>Presenter/Author</b>
1:00 pm	<b>Opening Remarks</b>	
1:05 - 1:50 pm	Keynote 1: Interplay of Linear Algebra, Machine Learning, and HPC	Dr. Sherry Li
1:50 - 2:35 pm	Keynote 2: Consensus, Discord, and Adversarial Intervention in Models of Opinion Formation	Prof. Jon Kleinberg
2:35 - 2:50 pm	Transactional Knowledge Graph Generation to Model Adversarial Activities (S10208)	Sumit Purohit, Patrick Mackey, William Smith, Madelyn Dunning, Miquette J Orren, Trevor M Langlie-Miletich, Rahul D Deshmukh, Ankur Bohra, Tonya J Martin, Dan Aimone, and George Chin
2:50 - 3:05pm	Higher-order Structure Based Anomaly Detection on Attributed Networks (S10209)	Xu Yuan, Na Zhou, Shuo Yu, Huafei Huang, Zhikui Chen, and Feng Xia
3:05 – 3:15	Influence in Transient Populations (S10204)	Dana Warmesley and Samuel Johnson
3:15 – 3:25	Evaluating the subgraph matching problem in the presence of categorical constraints (S10206)	Christopher Ebsch, Joseph Cottam, and George Chin
3:25 – 3:45pm	<b>Coffee Break</b>	
3:45 – 4:30pm	Keynote 3:	Prof. Marinka Zitnik
4:30 – 4:45pm	Active Learning for the Subgraph Matching Problem (S10205)	Yurun Ge and Andrea Bertozzi
4:45 – 5:00pm	An Analysis of COVID-19 Knowledge Graph Construction and Applications (S10202)	Dominic Flocco, Bryce Palmer-Toy, Ruixiao Wang, Hongyu Zhu, Rishi Sonthalia, Junyuan Lin, Andrea Bertozzi, and Jeffrey Brantingham
5:00 – 5:10pm	A Constraint Propagation Approach for Identifying Biological Pathways in COVID-19 Knowledge Graphs (S10203)	Thomas Tu
5:10 – 5:20pm	Characterizing Disease Spreading via Visibility Graph Embedding (S10207)	Kangyu Ni, Jiejun Xu, Shane Roach, Tsai-Ching Lu, and Alexei Kopylov
5:20 – 5:30pm	A Graph Embedding Approach to User Behavior Anomaly Detection (S10201)	Alexander Modell, Jonathan Larson, Melissa Turcotte, and Anna Bertiger
5:30 – 5:40pm	Semantic Property Graph for Scalable Knowledge Graph Analytics (S10210)	Sumit Purohit, Nhuy Van, and George Chin
5:40 pm	<b>Closing Remarks</b>	



<b>Fourth International Workshop on the Internet of Things Data Analytics (IoTDA)</b> <i>Location: Online (virtual)</i> <i>Workshop Chairs: Eyhab Al-Masri (University of Washington Tacoma), Chi-Hua Chen (Fuzhou University, China), Alireza Souri (Islamic Azad University, Iran)</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 am – 9:05 am	Opening Remarks and Welcome	
9:10 am – 9:25 am	(S27204) Spatially-distributed Federated Learning of Convolutional Recurrent Neural Networks for Air Pollution Prediction	Do-Van Nguyen and Koji Zettsu
9:30 am – 9:45 am	(S27205) Improving the Awareness of Sustainable Smart Cities by Analyzing Lifelog Images and IoT Air Pollution Data	Tuan-Vinh La, Minh-Son Dao, Kazuki Tejima, Rage Uday Kiran, and Koji Zettsu
9:50 am – 10:05 am	(S27206) Automated Kernel Search for Gaussian Processes on Data Streams	Jan David Hüwel, Fabian Berns, and Christian Beecks
10:10 am – 10:25 am	(S27207) Lightweight Video Analytics for Cycle Time Detection in Manufacturing	Andrew Walker, Daisuke Maeda, and Joydeep Acharya
10:30 am – 10:45 am	(S27210) Context-Aware Data Analytics Variability in IoT Neural Network-Based Systems	Nathalia Nascimento, Paulo Alencar, and Donald Cowan
10:50 am – 11:05 pm	(S27211) Deep Learning Framework for Internet of Things for People with Disabilities	Syed Jawad Hussain Shah, Ahmed Awad Albishri, and Yugyung Lee
<b>11:10 pm – 11:15 pm</b>	<b>Closing Remarks</b>	

<b>The Fifth Annual Workshop on Applications of Artificial Intelligence in the Legal Industry</b> <i>Workshop Chairs: Jianping Zhang, Nathaniel Huber-Fliflet and Haozhen Zhao</i>		
<b>Time (EST)</b>	<b>Title</b>	<b>Speakers/Authors</b>
10:00-10:10	<b>Opening Remarks</b>	
10:10-10:30	Invited Talk by Rishi Chhatwal – AT&T	
10:30-10:50	S11205 “Comparison of Deep Learning Technologies in Legal Document Classification”	Qian Han and Derek Snidauf
10:50-11:10	S11201 “PRESIDE: A Judge Entity Recognition and Disambiguation Model for US District Court Records”	Chris Rozolis and Adam Pah
11:10-11:40	Invited Talk by Peter Henderson and Diego A. Zambrano – Stanford University Presentation on: Vulnerabilities in Discovery Tech	
11:40-12:00	<b>Break</b>	
12:00-12:20	S11208 “Use Image Clustering to Facilitate Technology Assisted Review”	Fusheng Wei and Hilary Quatnetz
12:20-12:40	S11202 “The need for a numeric measure of explainability”	Wim De Mulder and Peggy Valcke
12:40-13:00	Invited Talk by Nikki Stern – Sony	
13:00-13:20	S11206 “Constraint-based Hierarchical Clustering for Time Sequences”	Yufeng Kou and Chris Knackstedt
13:20-13:40	S11207 “An Empirical Study on Transfer Learning for Privilege Review”	Haozhen Zhao
13:40-14:00	Invited Talk by Heather Heindel – Perkins Coie LLP	
14:00-14:10	<b>Closing Remarks</b>	

<b>6th Workshop on Real-time Stream Analytics, Stream Mining, CER/CEP &amp; Stream Data Management in Big Data</b> <i>Workshop Chairs: Sabri SKHIRI, Albert BIFET, Alessandro MARGARA</i> <i>Sabri Skhiri &lt;sabri.skhiri@euranova.eu&gt;</i>		
Time	Title	Presenter/Author
9:00 - 9:45	Keynote 1: A Year in Flink: The most important changes of the last versions and what's coming next	Till Rohrmann - Ververica
9:45 - 10:30	Keynote 2: Managing State in ksqlDB and Kafka Streams	Bill Bejeck- Confluent
	<b>Coffee Break</b>	
10:45 - 11:00	<b><i>A GPU Algorithm for Detecting Contextual Outliers in Multiple Concurrent Data Streams (S12203)</i></b>	Abinash Borah, Le Gruenwald, Eleazar Leal, and Egawati Panjei
11:00 - 11:15	<b><i>Reducing numerical precision preserves classification accuracy in Mondrian Forests (S12201)</i></b>	Marc Vicuna, Martin Khannouz, Gregory Kiar, Yohan Chatelain, and Tristan Glatard
11:15 - 11:40	<b><i>PASCAL-G: a Probabilistic Stream Clustering Analysis on Graphs (BigD674)</i></b>	Nidia Yadira Vaquera Chavez and Trilce Estrad
11:40 - 12:05	<b><i>Matrix Profile Index Approximation for Streaming Time Series (BigD550)</i></b>	Maryam Shahcheraghi, Trevor Cappon, Samet Oymak, Evangelos Papalexakis, Eamonn Keogh, Zachary Zimmerman, and Philip Brisk
12:05 - 12:15	<b>Break</b>	
12:15 - 12:40	<b><i>RStream: Simple and Efficient Batch and Stream Processing at Scale (BigD291)</i></b>	Alessio Fino, Alessandro Margara, Gianpaolo Cugola, Marco Donadoni, and Edoardo Morassutt
12:40 - 13:05	<b><i>Temporal Pattern Recognition in Graph Data Structures (BifgD250)</i></b>	Pietro Daverio, Hassan Nazeer Chaudhry, Alessandro Margara, and Matteo Rossi
13:05 - 13:15	<b>Closing Remarks</b>	

<b>The 8th Annual Workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDDBHI)</b> <i>Workshop Chairs: Donghui Wu, PhD, Xin Deng, PhD</i>		
Time	Title	Presenter/Author
9:00-9:15	Chairs' Opening Remarks	Donghui Wu
9:15-9:40	BigD450 : IDEA: Integrating Divisive and Ensemble-Agglomerate hierarchical clustering framework for arbitrary shape data	Hongryul Ahn, Inuk Jung, Heejoon Chae, Minsik Oh, Inyoung Kim, and Sun Kim
9:40-10:05	S13202 : <i>Empirical analysis of the residual information needed for image classification</i>	Rebecca Leygonie and Guillaume Vimont
10:05-10:30	S13204: <i>Generalization in Cardiac Image Segmentation</i>	Zhengjie Xu, Zixiang Zhou, Garrison W. Cottrell, and Mai H. Nguyen
10:30-10:55	N208 : Programmatic Database Language Generation for Big Data Applications	Hangu Yeo, Elahe Khorasani, Vadim Sheinin, Ngoc Phuoc An Vo, Octavian Popescu, and Petros
10:55-11:00	<b>Coffee Break</b>	

11:00-11:25	S13203 : <i>Integrating Heterogeneous Data for a Multi-disease Outbreak Detection Framework</i>	Ismael Villanueva-Miranda and Monika Akbar
11:25-11:50	BigD249 :The Influence of Social Factors on Mental Health and Wellbeing during the COVID-19 Pandemic	Alex Sartain and Mahmoud El-Haj
11:50-12:15	BigD447 :Attention based Sub-Network Search for Molecular Property Prediction	Xia Xiao and Sanguthevar Rajasekaran
12:15-12:40	S13205 : <i>A Secure and Reusable Software Architecture for Supporting Online Data Harmonization</i>	Zlatan Feric, Nicolas Bohm Agostini, Daniel Beene, Antonio J. Signes-Pastor, Yuliya Halchenko, Deborah Watkins, Debra MacKenzie, Margaret Karagas, Justin Manjourides, Akram Alshawabkeh, and David Kaeli
12:40-1:00	<b>Closing Remarks</b>	

<b>Applications of Big Data Technology in the Transport Industry</b> <i>Workshop Chair: John Easton, University of Birmingham, UK</i>		
<b>Time (US EST)</b>	<b>Title</b>	<b>Presenter/Author</b>
09:00 – 09:20	<i>S14206 - Predicting Air Ticket Demand Using Deep Neural Networks</i>	<i>Kodai Imanaka and Chiaki Sakama</i>
09:20 – 09:40	<i>S14205 - Clustering of Human Movement Trajectories based on Distributional Representations Derived from Bi-directional LSTM Network with Geographical Coordinates</i>	<i>Hiroki Tanaka, Takeshi Saga, and Satoshi Nakamura</i>
09:40 – 10:00	<i>S14209 - A Video Game-Inspired Approach for a Pedestrian Guidance System Within a Railway Station</i>	<i>Joseph Preece, Mohamed Samra, Richard Thomas, and John Easton</i>
10:00 – 10:10	<b>Q&amp;A Papers 1 - 3</b>	
10:10 – 10:15	<b>Comfort Break</b>	
10:15 – 10:35	<i>BigD473 - BIGcoldTRUCKS: a BIG data dashboard for the management of COLD chain logistics in refrigerated TRUCKS</i>	<i>Aurora González-Vidal, Paula Gómez-Bernal, José Mendoza-Bernal, and Antonio F. Skarmeta</i>
10:35 – 10:55	<i>BigD522 - A Systematic Literature Review of Machine Learning Tools for Supporting Supply Chain Management in the Manufacturing Environment</i>	<i>Johannes Breitenbach, Sara Haileselassie, Christoph Schuerger, Jonas Werner, and Ricardo Buettner</i>
10:55 – 11:15	<i>S14202 - A processing architecture for real-time predictive smart city digital twins</i>	<i>Sven van den Berghe</i>
11:15 – 11:25	<b>Q&amp;A Papers 4 - 6</b>	
11:25 – 11:50	<b>Coffee Break</b>	
11:50 – 12:10	<i>S14203 - A year into the pandemic: a passenger perspective on its impact at Paris-Charles de Gaulle airport</i>	<i>Geoffrey Scozzaro, Clara Buire, Aude Marzuoli, Eric Feron, and Daniel Delahaye</i>
12:10 – 12:30	<i>S14204 - Computational Analysis to Capture Cluster Lifetime Dynamics</i>	<i>Ivens Portugal, Paulo Alencar, and Donald Cowan</i>
12:30 – 12:50	<i>BigD581 - NodeSense2Vec: Spatiotemporal Context-Aware Network Embedding for Heterogeneous Urban Mobility Data</i>	<i>Dakshak Keerthi Chandra, Jennifer Leopold, and Yanjie Fu</i>
12:50 – 13:10	<i>S14201 - OnlineDC: Leveraging Temporal Driving Behavior to Facilitate Driver Classification</i>	<i>Hashim Abu-gellban, Yu Zhuang, Long Nguyen, Fang Jin, and Zhenkai Zhang</i>
13:10 – 13:25	<b>Q&amp;A Papers 7 - 10</b>	
13:25 – 13:30	<b>Closing Remarks</b>	

<p align="center"><b>The Fifth IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD), Wednesday, December 15</b></p> <p align="center"><i>Workshop Chairs: Zhiyuan Chen, Jianwu Wang, Feng Chen, Liqiang Wang</i></p>
---

Workshop website: <a href="https://workshops.minds.edu/workshop-BP19/">https://workshops.minds.edu/workshop-BP19/</a>		
Time (US Eastern Time)	Title	Presenter/Author
	Session 1 (Analytics & Optimization)	
9-9:05 am	Opening Remark	Workshop Chairs
9:05-9:20 am	S17203: Performance Optimization of Allreduce Operation for Multi-GPU Systems	Akira Nukada
9:20-9:40 am	S17210: Anomaly detection of high-dimensional sparse data based on Ensemble Generative Adversarial Networks	Wanghu Chen, Meilin Zhou, Chenhan Zhai, Mengyang Shen, Pengbo Lv, and Ali Arshad
9:40-10:00 am	BigD308: Impact of Attention on Adversarial Robustness of Image Classification Models	Prachi Agrawal, Narinder Singh Punna, Sanjay Kumar Sonbhadra, and Sonali Agarwal
10:00-10:15am	S17204: Benchmarking of Quantization Libraries in Popular Frameworks	Tejas Dubhir, Mayank Mishra, and Rekha Singhal
	Break	
	Session 2 (Benchmarks & Optimization)	
10:30-10:50	BigD494: A Performance Evaluation of Distributed Deep Learning Frameworks on CPU Clusters Using Image Classification Workloads	Andreas Krisilias, Nikodimos Provatias, Ioannis Konstantinou, and Nectarios Koziris
10:50-11:05	S17207: On the Potential of Execution Traces for Batch Processing Workload Optimization in Public Clouds	Dominik Scheinert, Alireza Alamgiralem, Jonathan Bader, Jonathan Will, Thorsten Wittkopp, and Lauritz Thamsen
11:05-11:20	S17208: Training Data Reduction for Performance Models of Data Analytics Jobs in the Cloud	Jonathan Will, Onur Arslan, Jonathan Bader, Dominik Scheinert, and Lauritz Thamsen
11:20-11:40	S17209: Energy Efficiency vs. Performance of Analytical Queries: The case of Bitmap Join Indexes	Issam Ghabri, Ladjel Bellatreche, and Sadok Ben Yahia
	Lunch Break	
1-2 pm	S17216: <b>Keynote</b>	Liqiang Wang, Central Florida University
	Session 3 (Analytics & Optimization)	
2:00-2:15 pm	BigD524: Deep-Learning-as-a-Workflow (DLaaW): an Innovative Approach to Enabling Deep Learning in Scientific Workflows	Junwen Liu, Ziyun Xiao, Shiyong Lu, and Dunren Che
2:15-2:35 pm	S17213: Efficient heap monitoring tool for memory leak detection and root-cause analysis	Vahid Azhari, Simar Bhamra, Naser Ezzati-Jivan, and Francois Tetreault
2:35-2:50 pm	S17205: Performance Profile of Transformer Fine-Tuning in Multi-GPU Cloud Environments	Seung-Hwan Lim, Edmon Begoli, and Sudarshan Srinivasan,
2:50-3:10 pm	S17201: A Decoupled Data Pipeline and its Reliability Assessment: Case Study in Extreme Climatic Humans Studies	Catherine Inibhunu, Jennifer Yeung, Aaron Gates, Bradley Chicoine, and Carolyn McGregor
3:10-3:25 pm	Break	
	Session 4 (Analytics)	
3:25-3:45 pm	S17214: Programming Languages in Data Science: a Comparison from a Database Angle	Xiantian Zhou and Carlos Ordonez
3:45-4:05 pm	S17202: WinnowML: Stable feature selection for maximizing prediction accuracy of time-based system modeling	Oceane Bel, Sinjoni Mukhopadhyay, Nathan Tallent, Faisal Nawab, and Darrell Long
4:05-4:25 pm	S17206: Exploratory Data Analysis with Database-backed Dataframes: A Case Study on Airbnb Data	Phanwadee Sinthong and Michael Carey
4:25-4:45 pm	S17211: Large-Scale Causality Discovery Analytics as a Service	Xin Wang, Pei Guo, and Jianwu Wang
4:45-5:05 pm	BigD548: Understanding the state of the Art in Animal detection and classification using computer vision technologies	Gabriel Ferrante, Felipe Rodrigues, Fernando Andrade, Rudinei Goularte, and Rodolfo Meneguette
	Closing Remarks	

<b>The Fourth Workshop on Big Data for Economic and Business Forecasting</b> <i>Workshop Chairs: Wei Shang, Wei Xu</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 – 9:20	Application of Machine Learning and Government Finance Statistics for macroeconomic signal mining to analyze recessionary trends and score policy effectiveness (S19202)	Chandrasekar Vuppalapati
9:20 – 9:40	Risk identification and default time analysis of corporate bonds (S19204)	Yang-guang LI
9:40 – 10:00	Financial Forecasting with Clustering Guided Modeling (BigD357)	Urjitkumar Patel
10:00-10:20	Temporal and Spatial Attention Network Model Based Evolution Model for Bulk Commodity Price Fluctuation Risk (S19210)	Wenjing Zhang
<b>Coffee Break</b>		
10:30-10:50	Prediction of Adult Chronic Kidney Disease With Class-Imbalanced Datasets (S19203)	ZHU Cuiling
10:50-11:10	Predicting the next stage of China's household sustainable energy consumption research based on a comparative knowledge graph analysis between China and foreign countries (S19205)	Yiming Yuan
11:10-11:30	Research on Quality Prediction of Typical Workpieces Based on Feature Recombination and XGBoost Algorithm (S19206)	Yaoyao Fan
11:30-11:50	How Does Movie Production Determine Box Office? Tests of Mediation and Moderation Effects (S19207)	Lin Huo
<b>Closing Remarks</b>		

<b>BigCyber 2021</b>		
<i>Workshop Chairs: Dr. Karuna Pande Joshi, Rajeev Agrawal, Sudip Mittal</i>		
<b>Time in EST</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 – 10:00 AM	Keynote	
10:00 – 10:20 AM	Evaluating Model Robustness to Adversarial Samples in Network Intrusion Detection (S20201)	Madeleine Schneider, David Aspinall, and Nathaniel Bastian
10:20 – 10:40 AM	Efficient and Privacy-Preserving Collaborative Intrusion Detection Using Additive Secret Sharing and Differential Privacy (S20202)	Laylon Mokry, Paul Slife, Patrick Bishop, Jose Quiroz, Cooper Guzzi, Zhiyuan Chen, Adina Crainiceanu, and Don Needham
10:40 – 11 AM	Cybersecurity Knowledge Graph Improvement with Graph Neural Networks (S20206)	Soham Dasgupta, Aritran Piplai, Priyanka Ranade, and Anupam Joshi
11 – 11:20 AM	Machine Learning Approaches for Authorship Attribution using Source Code Stylometry (S20204)	Sophia Frankel and Krishnendu Ghosh
11:20 – 11:40 AM	A Systematic Literature Review of Data Privacy Solutions for Smart Meter Technologies (S20208)	Jan Gross, Johannes Breitenbach, Waldemar Granson, Daniel Japs, Ardi Reci, Aaron Koengeter, and Ricardo Buettner
11:40 – 12 Noon	CyBERT: Contextualized Embeddings for the Cybersecurity Domain (S20207)	Priyanka Ranade, Aritran Piplai, Anupam Joshi, and Tim Finin
12 – 12:20 PM	Colab Cloud Based Portable and Shareable Hands-on Labware for Machine Learning to Cybersecurity (S20203)	Dan Lo, Hossain Shahriar, Kai Qian, Michael Whitman, and Fan Wu
12:20 – 12:40 PM	Combating Fake Cyber Threat Intelligence using Provenance in Cybersecurity Knowledge Graphs (S20205)	Shaswata Mitra, Aritran Piplai, Sudip Mittal, and Anupam Joshi
12:40 – 1 PM	Identifying Protection Motivation Theory Factors that Influence Smartphone Security Measures (BigD222)	Marvin Schneider and Shawon Rahman
1 – 1:20 PM	Detection of Newly Registered Malicious Domains through Passive DNS (BigD339)	Marcos Rogrio Silveira, Leandro Marcos da Silva, Adriano Mauro Cansian, and Hugo Koji Kobayashi
1:20 – 1:40 PM	Detection of Malicious Webpages Using Deep Learning (BigD395)	Amit Kumar Singh and Navneet Goyal
1:40 – 2 PM	Tolerating Adversarial Attacks and Byzantine Faults in Distributed Machine Learning (BigD455)	Yusen Wu, Hao Chen, Xin Wang, Chao Liu, Phuong Nguyen, and Yelena Yesha
	<b>Closing Remarks</b>	

<b>Workshop on Smart Farming, Precision Agriculture and Supply Chain (SmartFarm-2021)</b>		
<i>Maanak Gupta, Sudip Mittal and Mahmoud Abdelsalam</i>		
<b>Time (Eastern)</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 – 9:05	Welcome and Introduction	Maanak Gupta
9:05 – 9:25	(S21207) <b><i>Towards Data-and-Innovation Driven Sustainable and Productive Agriculture: BIO-AGRI-WATCH as a Use Case Study</i></b>	Asanee Kawtrakul, Kasetsart University, Thailand Hutchatai Chanlekha, Kasetsart University, Thailand Kitsana Waiyamai, Kasetsart University, Thailand Thanapat Kangkachit, Dhurakij Pundit University, Thailand * Laurent d'Orazio, Univ Rennes, CNRS, IRISA, France Dimitris Kotzinos, Cergy-Pontoise, ENSEA, CNRS, France Dominique Laurent, Cergy-Pontoise, ENSEA, CNRS, France Nicolas Spyrtatos, University of Paris-South, France
9:30 – 9:50	(S21206) <b><i>"An Ontological Knowledge Representation for Smart Agriculture"</i></b>	Bikram Pratim Bhuyan, University of Petroleum and Energy Studies, India * Ravi Tomar, University of Petroleum and Energy Studies, India Maanak Gupta, Tennessee Tech University, United States

		Amar Ramdane Cherif, University of Versailles Versailles, France
9:50 – 10:10	(S21204) <b>"RDNet: Deep Learning Model for Predicting pH H2O and pH KCl from Soil Vis-NIR Spectra"</b>	* Vung Pham, Sam Houston State University, United States David Weindorf, Central Michigan University, United States Tommy Dang, Texas Tech University, United States
10:10 – 10:15	Break	
10:15 – 11	Keynote	Michael Natrass, Tennessee Tech University
11- 11:20	(S21203) <b>"A systematic review of traceability issues in beef supply chain management"</b>	* <b>Oratile Leteane</b> , University of Botswana, Botswana Yirsaw Ayalew, University of Botswana, Botswana Tshiamo Motshegwa, University of Botswana, Botswana
11:20 – 11:40	(S21205) <b>"Autoencoder-based Anomaly Detection in Smart Farming Ecosystem"</b>	<b>Mary Adkisson</b> , Tennessee Technological University, United States <b>Jeffrey C Kimmell</b> , Tennessee Technological University, United States * Maanak Gupta, Tennessee Tech University, United States Mahmoud Abdelsalam, North Carolina A&T State University, United States
11:40 – 12:00	(S21202) <b>"Blockchain Based Farm-to-Fork Supply Chain Tracking"</b>	Kendall Land, Tennessee Technological University, United States Ambareen Siraj, Tennessee Technological University, United States
12:00 – 12:20	(S21201) <b>"Plant Growth and LAI Estimation using quantized Embedded Regression models for high throughput phenotyping"</b>	<b>Dhruv Sheth</b> , Edgelpulse Inc., India
	Closing Remarks	

The 4th International Workshop on Cross-disciplinary Data Exchange and Collaboration (CDEC2021)		
Workshop Chairs: Teruaki Hayashi, Yoshiaki Fukami, Yukio Ohsawa		
Time	Title	Presenter/Author
9:00-9:05	Opening Remarks	Teruaki Hayashi
9:05-9:25	Growing Process of Communities on Data Platforms: Case Analysis of a COVID-19 Dataset (S22202)	Teruaki Hayashi
9:25-9:45	Data Distribution and Competition Law Issues (S22201)	Koki Arai
9:45-10:05	The Impact of Decentralized Identity Architecture on Data Exchange (S22204)	Yoshiaki Fukami
10:05-10:25	A Numerical Study with Experimental Data on Risk-Averse Subcontractors in Procurement Auctions with Subcontract Bids (S22209)	Naoki Watanabe
10:25-10:35	Coffee Break	
10:35-11:15	Invited Talk: How to use data & video analysis in sports	Hotaka Yonezawa
11:15-11:35	Market Trend Analysis Using Polarity Index Generated from Analyst Reports (S22203)	Rei Taguchi
11:35-11:55	How can we collect satisfactory answers in art appreciation experiment? (S22205)	Akinori Abe
11:55-12:15	Retrieving of Data Similarity using Metadata on a Data Analysis Competition Platform (S22206)	Hiroki Sakaji
12:15-12:35	Supporting Test Case Design on Reasoning Scheme with Natural Language Processing Technique (S22210)	Noriyuki Kushiro



12:35-12:40	Closing Remarks
-------------	-----------------

<b>The Fifth IEEE Workshop on Human-in-the-loop Methods and Future of Work in BigData (HMDData 2021)</b> <i>Workshop Chair: Senjuti Basu Roy (New Jersey Institute of Technology), Alex Quinn (Purdue University), Atsuyuki Morishima (University of Tsukuba)</i> <i>The latest program is available at <a href="https://humanmachinedata.org">https://humanmachinedata.org</a></i>		
Time	Title	Author(s)
1:00PM	<b>Opening</b>	
1:05	<b>Keynote:</b> Lora Aroyo (Google)	
1:55	<b>Short Break</b>	
2:00	[S25209] Adaptive Budget Allocation for Cooperative Task Solving in Crowdsourcing	Yuya Itoh and Shigeo Matsubara
2:20	[S25214] Visualization for Enabling Human-in-the-Loop in Trace Clustering-based Process Mining Tasks	Thais Rodrigues Neubauer, Glaucia Pamponet Sobrinho, Marcelo Fantinato, and Sarajane Marques Peres
2:32	[S25205] A Skill-based Worksharing Approach for Microtask Assignment	Kanta Negishi, Hiroyoshi Ito, Masaki Matsubara, and Atsuyuki Morishima
2:44	[S25202] Editors-in-the-loop News Article Summarization Framework with Sentence Selection and Compression	Shotaro Ishihara, Yuta Matsuda, and Norihiko Sawa
2:56	<b>Break</b>	
3:20	[S25206] BUBBLE : A Quality-Aware Human-in-the-loop Entity Matching Framework	Naofumi Osawa, Hiroyoshi Ito, Yukihiro Fukusima, Takashi Harada, and Atsuyuki Morishima
3:40	[BigD375] An Explainable Person-Job Fit Model Incorporating Structured Information	Yunchong Zhang, Baisong Liu, Jiangbo Qian, Jiangcheng Qin, Xueyuan Zhang, and Xueyong Jiang
3:52	[S25208] A Practical and Universal Framework for Generating Publicly Available Medical Notes of Authentic Quality via the Power of Crowds	Rina Kagawa, Yukino Baba, and Hideo Tsurushima
4:04	[S25203] Does Multi-Hop Crowdsourcing Work? A Case Study on Collecting COVID19 Local Information	Ying Zhong, Masaki Kobayashi, Masaki Matsubara, and Atsuyuki Morishima
4:16	[S25218] Spammer Detection Based on Task Completion Time Variation in Crowdsourcing	Ayato Watanabe and Keishi Tajima
4:30	<b>Break</b>	
4:40	[S25213] BIRD-QA: A BERT-based Information Retrieval Approach to Domain Specific Question Answering	Yuhao Chen and Farhana Zulkernine
5:00	[S25201] An Alert Notification Subsystem for AI Based Clinical Decision Support: A Prototype in NICU	Catherine Inibhunu, Carolyn McGregor, and Edward Pugh
5:12	[S25212] Development of Prototype System to Generate Chronological Response Scenario Dataset by Assembling Multi-responders' Action Logs at Past Disaster	Munenari Inoguchi
5:24	[S25210] A Rival Recommendation Approach for Acoustic AR Running Support System Considering the Athletic Ability of Users	Ryusei Arisawa, Panote Siriaraya, Da Li, Kazutoshi Sumiya, Yukiko Kawai, and Shinsuke Nakajima
5:36	[S25207] A Method to Correct Perspective Distortion of Ground Area without Camera Parameters	Hisatoshi Toriya, Ashraf Dewan, and Itaru Kitahara
5:50	<b>Closing Remarks</b>	

<b>9th International Workshop on Distributed Storage and Blockchain Technologies for Big Data</b> <i>Workshop Chair: Hui Li</i>		
Time	Title	Presenter/Author
9:00 – 9:20 am	S29201, “Improving Transaction Speed and Scalability in Blockchain Systems”	Joshua Aaron DeNio and Simone Ludwig



9:20 – 9:40 am	S29202, “A Data Lightweight Scheme for Parallel Proof of Vote Consensus”	Zixian Wang, Hui Li, Han Wang, Zhenwei Xiao, Ping Lu, Zhenyuan Yang, Minglong Zhang, and Peter Han Joo Chong
9:40 – 10:00 am	S29203, “Optimal Copyset in Distributed Object Storage”	Yaoguang Huo, Junfeng Ma, Hui Li, Xin Yang, Han Wang, and Xiangzhen Meng
10:00 – 10:20 am	S29204, “A Complete Privacy-Preserving Credit Score System Using Blockchain and Zero Knowledge Proof”	Yuye Han, Hong Chen, Zhijie Qiu, Lei Luo, and Gongbin Qian
10:20 – 10:40 am	BigD440(->S29), “Distributed file system for rewriting Big Data files using a local-write protocol”	Erico Silva, Liria Sato, and Edson Midorikawa
<b>Closing Remarks</b>		

<b>The 3rd International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD) 2021</b> <i>Workshop Chairs: Sangkeun (Matt) Lee, Pravallika Devineni, Jong Youl Choi</i>		
Time	Title	Presenter/Author
8:50-9:00	Opening Remarks	
9:00-9:25	Kara Schatz, Cleber Melo-Filho, Alexander Tropsha, and Rada Chirkova, <i>Explaining Drug-Discovery Hypotheses Using Knowledge-Graph Patterns</i> (S30201)	Kara Schatz <kmschat2@ncsu.edu>
9:25-9:50	Jalal Khalil, Da Yan, Guimu Guo, Mirza Tanzim Sami, Joy Bhadhan Roy, and Virginia Sisiopiku, <i>Traffic Study of Shared Micromobility Services by Transportation Simulation</i> (S30202)	Da Yan <yanda@uab.edu>
9:50-10:15	Sheila Alemany, Jason Nucciarone, and Niki Pissinou, <i>Jespipe: A Plugin-Based, Open MPI Framework for Adversarial Machine Learning Analysis</i> (S30203)	Sheila Alemany <saalem010@fiu.edu>
10:15-10:40	Jeremy Logan, Greeshma Agasthya, Heidi Hanson, Matthew Wolf, Anuj Kapadia, Heechan Lee, and Hong-Jun Yoon, <i>Creating a Tools Ecosystem for Cross-Discipline Environmental Data Reuse</i> (S30204)	Jeremy Logan <loganjs@ornl.gov>
10:40-10:50	<b>Coffee Break</b>	
10:50-11:15	Xintong Zhao, Jane Greenberg, Yuan An, and Xiaohua Tony Hu, <i>Fine-Tuning BERT Model for Materials Named Entity Recognition</i> (S30205)	Xintong Zhao <xz485@drexel.edu>
11:15-11:40	Alina Lazar, Ling Jin, Caitlin Brown, C. Anna Spurlock, Alexander Sim, and Kesheng Wu, <i>Performance of the Gold Standard and Machine Learning in Predicting Vehicle Transactions</i> (S30206)	Alina Lazar <alazar@ysu.edu>
11:40-12:05	António Correia, Benjamim Fonseca, Hugo Paredes, Ramon Chaves, Daniel Schneider, and Shoaib Jameel, <i>Determinants and Predictors of Intentionality and Perceived Reliability in Human-AI Interaction as a Means for Innovative Scientific Discovery</i> (S30207)	António Correia <antonio.g.correia@inesctec.pt>
12:05-12:20	Bill Kay, Hao Lu, Pravallika Devineni, Anika Tabassum, Supriya Chintavali, and Sangkeun Lee, <i>Identification of Critical Infrastructure via PageRank</i> (S30208)	Bill Kay <kaybw@ornl.gov>
12:20-12:45	Katerina Antypas, Debbie Bard, Johannes Blaschke, Shane Canon, Bjoern Enders, Arjun Shankar, Suhas Somnath, Dale Stansberry, Thomas Uram, and Sean Wilkinsin, <i>Enabling discovery data science through cross-facility workflows</i> (S30209)	Katerina Antypas <kantypas@lbl.gov>
12:45	<b>Closing Remarks</b>	

<b>The 6th IEEE International Workshop on Big Spatial Data (BSD 2021)</b> <i>Workshop Chairs: Ashwin Shashidharan, Abdeltawab Hendawi, Chengyang Zhang, Farnoush Banaei-Kashani</i>		
Time (EST)	Title	Presenter/Author

9:00 – 9:05	<b>Opening Remarks</b>	
9:05 – 10:10	Keynote	Dr. Robert Stewart, ORNL
10:10 – 10:30	S31208: Towards a Model-Driven Datacube Analytics Language	Peter Baumann
10:30 – 10:40	<b>Coffee Break</b>	
10:40 – 11:00	S31203: Fast Gaussian Process Estimation for Large-Scale In Situ Inference using Convolutional Neural Networks	Divya Banesh
11:00 – 11:15	S31202: Prioritized Sampling on Knowledge Distillation for Nowcasting Pluvial Flood Prediction	Ryoma Kondo
11:15- 11:30	S31210: Development of a Reinforcement Learning based Agent Model and People Flow Data to Mega Metropolitan Area	Yanbo Pang
11:30 – 11:50	S31209: A Holistic Spatial Platform For Managing Infectious Diseases, Case Study on COVID-19 Pandemic	Louai Alarabi
11:50 – 12:00	<b>Coffee Break</b>	
12:00 – 12:20	S31205: Containerization of Model Fitting Workloads over Spatial Datasets	Menuka Warushavithana
12:20 – 12:35	S31211: A Scalable System for Searching Large-scale Multi-sensor Remote Sensing Image Collections	Yifan Zhao
12:35 – 12:50	S31204: Map4OLAP: A web-based tool for interactive map visualization of OLAP queries	Sandro Bimonte
12:50 – 1:10	BigD420: Decentralized Storage for Scientific Data	Shirish Patel
1:10 – 1:15	<b>Closing Remarks</b>	

<b>Applications of Topological Data Analysis to Big Data</b> Workshop Chairs: Tegan Emerson, Mason A. Porter, Sarah Tymochko, Wlodek Zadrozny		
<b>Time (EST)</b>	<b>Title</b>	<b>Presenter/Author</b>
10:00 – 10:50	Keynote	Gregory Henselman-Petrusek
11:00 – 11:25	S32202 Combining Geometric and Topological Information for Boundary Estimation <i>(regular paper)</i>	Hengrui Luo and Justin Strait
11:30 – 11:55	S32205 The Impact of Changes in Resolution on the Persistent Homology of Images <i>(regular paper)</i>	Teresa Heiss, Sarah Tymochko, Brittany Story, Adélie Garin, Hoa Bui, Bea Bleile, and Vanessa Robins
12:00 – 12:15	S32203 Topological Learning for Motion Data via Mixed Coordinates <i>(short paper)</i>	Hengrui Luo, Jisu Kim, Alice Patania, and Mikael Vejdemo-Johansson
12:20 – 12:35	S32204 Activation Landscapes as a Topological Summary of Neural Network Performance <i>(short paper)</i>	Matthew Wheeler, Jose Bouza, and Peter Bubenik
12:40 – 12:55	S32208 Topological Segmentation and Tracking for Space Weather Modeling <i>(short paper)</i>	Divya Banesh, Natalie Klein, Steven Morley, Ayan Biswas, Vania Jordanova, Michael Henderson, and Earl Lawrence
1:00 – 2:00	Lunch Break	
2:00 – 2:25	S32206 Go with the Flow? A Large-Scale Analysis of Health Care Delivery Networks in the United States Using Hodge Theory <i>(regular paper)</i>	Thomas Gebhart, Xiaojun Fu, and Russell Funk
2:30 – 2:55	S32212 Adaptive Covers for Mapper Graphs Using Information Criteria <i>(regular paper)</i>	Nithin Chalapathi, Youjia Zhou, and Bei Wang
3:00 – 3:25	S32213 Topological Data Analysis in an Economic Context: Property Tax Maps <i>(regular paper)</i>	David Damiano and Anbo Wu
3:30 – 3:45	S32214 Diagnosis of Pediatric Obstructive Sleep Apnea via Face Shape Analysis <i>(short paper)</i>	Milad Kiaee, Adam Kashalak, Jisu Kim, Giseon Heo, and Prachi Loliencar
3:50 – 4:05	S32215 Data Reduction and Feature Isolation for Computing Persistent Homology on High Dimensional Data <i>(short paper)</i>	Rishi Verma, Nicholas Malott, and Philip Wilsey
4:10	Closing Remarks	

<b>Multi-Modal Medical Data Analysis</b> <i>Workshop Chairs: Wendong Wang</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
8:55-9:00	Opening	Workshop Chairs
9:00-9:15	S34201: Measles Rash Identification Using Transfer Learning and Deep Convolutional Neural Networks	Kimberly Glock
9:15-9:30	S34203: DRA U-Net: An Attention based U-Net Framework for 2D Medical Image Segmentation	Sicheng Shen
9:30-9:45	S34204: Cyst segmentation on kidney tubules by means of U-Net deep-learning models	Simone Monaco
9:45-10:00	S34205; R2Net: Recurrent Recalibration Network for Medical Image Segmentation	Wenbin Wu
10:00-10:15	<b>Coffee Break</b>	
10:15-10:30	S34206: ISE-YOLO: Improved Squeeze-and-Excitation Attention Module based YOLO for Blood Cells Detection	Cong Liu
10:30-10:45	S34209: MFSL-Net: A Modality Fusion and Shape Learning based Cascaded Network for Prostate Tumor Segmentation	Fan Zhang
10:45-11:00	S34212: Early Detection of Alcohol Use Disorder Based on a Novel Machine Learning Approach Using EEG Data	Dennis Flathau
11:00-11:15	BigD513: FREIDA - Fracture Risk Evaluation using Highly Efficient Information Retrieval and Analysis of Large Healthcare Datasets	COROIU Adriana Mihaela
	<b>Closing Remarks</b>	

<b>1<sup>st</sup> International Workshop on Big Data Analytics for Sustainability &amp; 5<sup>th</sup> International Workshop on Big Data Analytics for Cyber Intelligence and Defense</b> <i>Workshop Chairs: Stephen McGough, Matthew Forshaw Alex Kell</i> <i>Workshop Chairs: Stephen McGough, Amir Atapour-Abarghouei, David Wall</i> <b>As we're running these two workshops together here are the two workshop schedules.</b>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00 – 9:10	Introduction to workshop	Stephen McGough, Matthew Forshaw Alex Kell
9:10 – 9:35	<b>Keynote:</b> Geospatial Big Data analytics to model the long-term sustainable transition of residential heating worldwide (S36207)	<b>Diego Moya</b> , Sara Giarola, and Adam Hawkes
9:35 – 10:00	Spatio-Temporal Clustering based on HHT and Its Applications in Thermal Boiler Controlling (S36203)	Wanghu Chen, <b>Yan Sun</b> , Jing Li, Chenhan Zhai, Pengbo Lv, and Shengfang Jin
10:00 – 10:15	<b>Coffee Break</b>	
10:15 – 10:40	Multi Scale Graph Wavenet for Wind Speed Forecasting (S36201)	Neetesh Rathore, <b>Pradeep Rathore</b> , Arghya Basak, Sri Harsha Nistala, and Venkataramana Runkana
10:40 – 11:05	Hard Disk Failure Prediction on Highly Imbalanced Data Using LSTM Network (S36205)	<b>Cahyadi</b> and Matthew Forshaw
11:05 – 11:30	Survival Analysis and Predictive Maintenance Models for non-sensored Assets in Facilities Management (S36206)	<b>Genevieve Moat</b> and Shirley Coleman
11:30 – 11:55	Application of deep learning to camera trap data for ecologists in planning / engineering – Can captivity imagery train a model which generalises to the wild? (S36208)	<b>Ryan Curry</b> , Cameron Trotter, and Stephen McGough

11:55 – 12:20	Forecasting Air Pollution using a Modified Compositional Learning Approach (S36202)	<b>Samuel A. Ajila</b> and Karthik Dilliraj
12:20 – 13:30	<b>Lunch</b>	
13:30 – 13:55	Top-k user-based Collaborative Recommendation System using MapReduce (S36204)	<b>Sheheeda Manakkadu</b> , Srijan Prasad Joshi, Tom Halverson and Sourav Dutta
13:55 – 14:20	<b>Round Table: What is the future of Big Data and Sustainability?</b>	
14:20 – 14:35	<b>Closing Remarks</b>	

Time	Title	Presenter/Author
14:35 – 14:45	Introduction to Workshop	<i>Stephen McGough, Amir Atapour-Abarghouei, David Wall</i>
14:45 – 15:10	A review of recent advances in machine learning approaches for cyber defense (BigD525)	<b>Ricardo Buettner</b> , Daniel Sauter, Jonas Klopfer, Johannes Breitenbach, and Hermann Baumgartl
15:10 – 15:35	Transforming Fake News: Robust Generalisable News Classification Using Transformers (BigD422)	<b>Ciara Blackledge</b> and Amir Atapour-Abarghouei
15:35 – 15:55	<b>Coffee Break</b>	
15:55 – 16:20	Rank over Class: The Untapped Potential of Ranking in Natural Language Processing (BigD278)	<b>Amir Atapour-Abarghouei</b> , Stephen Bonner, and Andrew Stephen McGough
16:20 – 16:45	“Just Drive”: Colour Bias Mitigation for Semantic Segmentation in the Context of Urban Driving (S35201)	<b>Jack Stelling</b> and Amir Atapour-Abarghouei
16:45 – 17:10	<b>Round Table: What is the future of Big Data and Cyber Intelligence and Defence?</b>	
17:10 – 17:30	<b>Closing Remarks</b>	

<b>1st International Workshop on Serverless Machine Learning for Intelligent and Scalable AI Workflow</b> <i>Workshop Chairs: Dhaval Patel, Shuxin Lin</i>		
Time	Title	Presenter/Author
1 PM : 1:30 PM	Key Note: Prof. Evgenia Smirni	
1 :30 PM : 1:45 PM	S37201 : Deployment and Management of Time Series Forecasts in Ocean Industry	Fearghal O'Donncha, Albert Akhriev, Bradley Eck, Meredith Burke, Ramon Filgueira, and Jon Grant
1:45 PM : 2:00 PM	S37204 : <i>FedSmarteum: Secure FederatedMatrix Factorization Using Smart Contracts for Multi-Cloud Supply Chain</i>	Srini Bhagavan, Mohamed Gharibi, and Praveen Rao,
2:00 PM : 2:15 PM	S37203: TableNN: Deep Learning Framework for Learning Domain Specific Tabular Data	Pranav Sankhe, Bhavna Agrawal, Elham Khabiri, and Yingjie Li
2:15 PM : 2:30 PM	S37202 : Efficient Supervision Net: Polyp Segmentation Using EfficientNet and Attention Unit	sabari nathan, Suganya R, and Rampriya RS
	<b>Coffee Break</b>	
3:00 PM : 3:20 PM	S37205: Predicting family physicians based on their practice using machine learning	Arunim Garg, David W. Savag, Salimur Choudhury, and Vijay Mago
3:20 PM : 3:40 PM	S37206 : FederatedTree: A Secure Serverless Algorithm for Federated Learning to Reduce Data Leakage	Mohamed Gharibi, Srini Bhagavan, and Praveen Rao
3:40 PM : 4:00 PM	S37207: Asset Modeling using Serverless Computing	Srideepika Jayaraman, Chandra Reddy, Elham Khabiri, Dhaval Patel, Anuradha Bhamidipaty, and Jayant Kalagnanam
4:00 PM : 4:20 PM	N249 : Automated Counterfactual Generation in Financial Model Risk Management	Jingwei Gan, Shinan Zhang, Chi Zhang, and Andy Li
	<b>Closing Remarks</b>	

<b>2nd International Workshop on User Understanding from Big Data &amp; International Workshop on Data Science Systems (DSS)</b> <i>Workshop Chairs: Ladjel Bellatreche, Carlos Ordonez, Claire Ding, Wutao Wei</i>		
<b>Time (EST, Orlando)</b>	<b>Title</b>	<b>Presenter/Author</b>
<b>1:00pm</b>	DSS workshop: Welcome	Ladjel Bellatreche
1:10pm	BigD619: K-nearest Multi-agent Deep Reinforcement Learning for Collaborative Tasks with a Variable Number of Agents	Hamed Khorasgani
1:25pm	S33205: Towards Automated Variability-Aware Machine-Learning-Based Modeling Analysis	Cristina Tavares
1:40pm	S33201: Coxswain: Guided Data Analytics	Carlos Garcia-Alvarado
1:55pm	BigD304: Communication efficient distributed learning of neural networks in Big Data environments using Spark	Fouad Alkhoury
2:10pm	Open discussion about research trends and all papers	
<b>2:20-2:40pm</b>	<b>Social Time</b>	
2:40-3:10pm	S16205 "A Short Survey on Forest Based Heterogeneous Treatment Effect Estimation Methods: Meta-learners and Specific Models"	Hao Jiang, Peng Qi, Jingying Zhou, Jack Zhou, and Sharath Rao
3:20-3:50pm	S16203 "Predicting Depression Symptoms from Discord Chat Messaging Using AI Medical Chatbots"	Venkata Duvvuri, Qihan Guan, Swetha Daddala, Mitch Harris, and Sudhakar Kaushik
4:00-4:10pm	<b>Social Time</b>	
4:10-4:40pm	S16201 "Test Script Generation Based on Hidden Markov Models Learning From User Browsing Behaviors"	Izzettin Erdem, Ramazan Faruk Oguz, Erdi Olmezogullari, and Mehmet Siddik Aktas
<b>4: 40pm</b>	<b>Closing Remarks</b>	

<b>Eighth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining</b> <i>Workshop Chairs: Nesreen Ahmed, Shaikh Arifuzzaman, Mohammad Al Hasan, Kamesh Madduri</i>		
<b>Time</b>	<b>Title</b>	<b>Presenter/Author</b>
9:15am-9:20am	Opening Remarks	Shaikh Arifuzzaman, Mohammad Al Hasan
9:20am-9:40am	BigD231: Building Graphs at a Large Scale: Union Find Shuffle	Sai Kiran Reddy Malikireddy
9:40am-10am	BigD595: Event-based Product Carousel Recommendation with Query-Click Graph	Luyi Ma
10am-10:15am	<b>Coffee Break</b>	
10:15am-11am	Keynote	
11am-11:20am	S38202: A Pre-training Oracle for Predicting Distances in Social Networks	Gunjan Mahindre
11:20-11:40am	S38204: HPCGCN: A Predictive Framework on High Performance Computing Cluster Log Data Using Graph Convolutional Networks	Avishek Bose
11:40am-12pm	S38201: Correlation and pattern detection in event networks	Valerio Bellandi
12pm-12:20pm	BigD626: Drug Abuse Detection in Twitter-sphere: Graph-Based Approach	Khaled Mohammed Saifuddin
<b>12:20pm-12:25pm</b>	<b>Closing Remarks</b>	

## Data science for equality, inclusion and well-being challenges (DS4EIW)

Workshop Chairs: Genoveva Vargas-Solar, Tania Cerquitelli December 15 <sup>th</sup> , 2021 (Morning)		
Time	Title	Presenter/Author
8:30	Opening	Genoveva Vargas-Solar, Tania Cerquitelli
8:40 – 9:00	Invited Keynote: "Social Mining and Big Data Analytics"	Fosca Giannotti
9:00 – 9:20	S40201 : <b><i>hUGe2: An Interdisciplinary Research Program for Sustainability</i></b> by J. David Patón-Romero and Letizia Jaccheri	J. David Patón-Romero
9:20 – 9:40	S40202: <b><i>Covid-19 digital Contact-tracing: a doorway to well-being or a backdoor to security vulnerabilities?</i></b> by Nishit Patel, David Cancel, Moitrayee Chatterjee, and Md Shahinoor Rahman	Moitrayee Chatterjee
9:40 – 10:00	S40204 : <b><i>Star inventors and gender gaps in patented innovations</i></b> by Federico Caviggioli, Alessandra Colombelli, and Chiara Ravetti	Chiara Ravetti
10:00 – 10:20	S40211: <b><i>E-MIMIC: Empowering Multilingual Inclusive Communication</i></b> by Giuseppe Attanasio, Salvatore Greco, Moreno La Quatra, Luca Cagliero, Rachele Raus, Michela Tonti, and Tania Cerquitelli	Salvatore Greco
10:20 – 10:40	S40206 : <b><i>Detecting Discrimination Risk in Automated Decision-Making Systems with Balance Measures on Input Data</i></b> by Mariachiara Mecati, Antonio Vetrò, and Marco Torchiano,	Mariachiara Mecati
10:40 – 10:50	Coffee Break	
10:50- 11:10	S40209 : <b><i>Demographic Data-driven Deprivation Index for Predicting Chronic Diseases</i></b> By Olugbenga Iyiola and Monika Akbar	Monika Akbar
11:10- 11:30	S40207 : <b><i>Green innovative start-ups and gender shares</i></b> by Alessandra Colombelli, Anna D'Ambrosio, and Chiara Ravetti,	Alessandra Colombelli
11:30- 11:50	S40210 : <b><i>DS4ALL: All you need for democratizing data exploration and analysis</i></b> by Paolo Bethaz, Khalid Belhajjame, Genoveva Vargas-Solar, and Tania Cerquitelli,	Paolo Bethaz
11:50- 12:10	S40212 : <b><i>Estimating urban socioeconomic inequalities through airtime top-up transactions data</i></b> By Eduardo Cruz, Carmen Vaca, and Monica Villavicencio,	Eduardo Cruz
12:10- 12:25	S40208 : <b><i>ASSIST: Article eXtraction and statistical Analysis</i></b> by Justine Fouillé, Thi Lan Huong Nguyen, Baptiste Alix, Brett Becket, Matthieu Rochard, Hélène De Ribaupierre, and Laurent d'Orazio	Laurent d'Orazio
12:25- 12:40	S40205 : <b><i>Clinical gait analysis: a gender medicine perspective</i></b> by Valentina Agostini	Valentina Agostini
12:40- 1:00	S40203 : <b><i>Bayesian model for Fairness in sampling from clustered data and FP-FN error rates</i></b> Ishani Chakraborty	Ishani Chakraborty
1:00-1:10	Closing Remarks	

# IEEE Big Data 2021 Workshop on Big Data Analytics for COVID-19

[BDA-COVID 2021](#)

Dec. 15, 2021

## Workshop Chairs:

Duoduo (Lindi) Liao, [dliao2@gmu.edu](mailto:dliao2@gmu.edu)

Harry J. Foxwell, [hfoxwell@gmu.edu](mailto:hfoxwell@gmu.edu)

Time (EST)	ID	Paper Title	Author(s)	Session Chair
9:30am - 9:55am	S41202	Distinct Structural Flexibility within SARS-CoV-2 Spike Protein Reveals Potential Therapeutic Targets	Serena H. Chen, M. Todd Young, John Gounley, Christopher Stanley, and Debsindhu Bhowmik	Dr. Lindi Liao
9:55am - 10:20am	S41205	Forecasting High-risk Areas of COVID-19 Infection Through Socioeconomic and Static Spatial Analysis	Abdulaziz Alhamadani, Shailik Sarkar, Lei Zhang, Lulwah Alkulaib, and Chang-Tien Lu	
10:20am - 10:45am	S41210	Wastewater-Based Epidemiological Modeling for Continuous Surveillance of COVID-19 Outbreak	Mehrdad Fazli, Samuel Sklar, Michael D. Porter, Brent A. French, and Heman Shakeri	
10:45am - 11:10am	S41212	TemPredict: A Big Data Analytical Platform for Scalable Exploration and Monitoring of Personalized Multimodal Data for COVID-19	Shweta Purawat, Subhasis Dasgupta, Jining Song, Shakti Davis, Kajal T. Claypool, Sandeep Chandra, Ashley Mason, Varun Viswanath, Amit Klein, Patrick Kasl, YingJing Wen, Benjamin Smarr, Amarnath Gupta, and Ilkay Altintas	
11:10am - 11:20am	Coffee Break			
11:20am - 11:45am	S41217	Mobility-Based Individual POI Recommendation to Control the COVID-19 Spread	Abhirup Ghosh and Tong Xia	Dr. David Wong
11:45am - 12:10pm	S41220	Characterizing the Online Discourse in Twitter: Users Reaction to Misinformation around COVID-19 in Twitter	Niloofar Kalantari, Duoduo Liao, and Vivian Motti	



Time (EST)	ID	Paper Title	Author(s)	Session Chair
12:10pm - 12:35pm	S41221	Classification of COVID-19 using Deep Learning and Radiomic Texture Features Extracted from CT Scans of Patients Lungs	Jayalakshmi Mangalagiri, Jones Sam Sugumar, Sumeet Menon, David Chapman, Aryya Gangopadhyay, Yelena Yesha, and Phuong Nguyen	
12:35pm - 1:10pm	<b>Lunch Break</b>			
1:10pm - 1:30pm	S41203	Visual Understanding of COVID-19 Knowledge Graph for Predictive Analysis	Seung-Hwan Lim, Junghoon Chae, Guojing Cong, Dasha Herrmannova, Robert Patton, Ramakrishnan Kannan, and Thomas E. Potok	Dr. Harry Foxwell
1:30pm - 1:50pm	S41204	Course Scheduling to Minimize Student Wait Times For University Buildings During Epidemics	Arvin Hekmati, Bhaskar Krishnamachari, and Maja J Mataric	
2:10pm - 2:30pm	S41208	COVID-19++: A Citation-Aware COVID-19 Dataset for the Analysis of Research Dynamics	Lukas Galke, Eva Seidlmayer, Gavin Lüdemann, Lisa Langnickel, Tetyana Melnychuk, Konrad U. Förstner, Klaus Tochtermann, and Carsten Schultz	
2:30pm - 2:50pm	S41215	Unsupervised Data Mining on Spatial-Temporal Passenger Mobility and Survey Data during COVID-19	Philippe Budinger and Tor Morten Gronli	
3:10pm - 3:30pm	S41223	Learning Domain-Specific Word Embeddings from COVID-19 Tweets	Steve Aigbe and Christoph Eick	
3:30pm - 3:40pm	<b>Coffee Break</b>			
3:40pm - 3:55pm	S41216	Effects of Image Quality and Quantity on Building a Competitive COVID-19 Diagnosis Model	Philipp Skovelef Orfanoudakis, Paraskevi Tzouveli, and Stefanos Kollias	Dr. Chen Shen
3:55pm - 4:10pm	S41218	COVID Vaccine Risk Stratification	Lucas Standaert, Thilanka Munasinghe, Dongyoung	

Time (EST)	ID	Paper Title	Author(s)	Session Chair
			Jang, Catherine Tate, Sean Lossef, and Wilson Wong	
4:10pm - 4:25pm	S41206	Covid Vaccine Sentiment Analysis by Geographic Region	Ebelechukwu Nwafor, Ryan Vaughan, and Christopher Kolimago	
4:25pm - 4:40pm	BigD219	Compliance with Stay-at-Home Orders during COVID-19	Colin Berry	
4:40pm - 4:55pm	BigD621	Applying Machine Learning to Analyze Anti-Vaccination on Tweets	Maryam Taeb, Hongmei Chi, and Jie Yan	
4:55pm - 5:10pm	S41208	My Mind is a Prison: A Boosted Deep Learning approach to detect the rise in depression since COVID-19 using a stacked bi-LSTM Catboost model	Shahana Nandy and Vishrut Kumar	

## IEEE Big Data 2021 Workshop on Multimodal Big Data

[MMBD 2021](#)

Dec. 16, 2021

**Workshop Chair:** Duoduo (Lindi) Liao, [dliao2@gmu.edu](mailto:dliao2@gmu.edu)

Time	ID	Paper Title	Author(s)	Session Chair
1:00pm - 1:25pm	S43201	Predicting Mental Health Problems with Personality, Behavior, and Social Networks	Dongyu Zhang, Teng Guo, Shiyu Han, Sadaf Vahabli, Mehdi Naseriparsa, and Feng Xia	Dr. Lindi Liao
1:25pm - 1:50pm	S43208	A Re-thinking ASR Modeling Framework using Attention Mechanisms	Chih-Ying Yang and Kuan-Yu Chen	

Time	ID	Paper Title	Author(s)	Session Chair
1:50pm - 2:15pm	S43206	Improving Sign Video Modeling Using Graph Neural Network	Al Amin Hosain, Huzefa Rangwala, and Jana Kosecka	
2:15pm - 2:40pm	BigD276	Multimodal Product Matching and Category Mapping: Text+Image based Deep Neural Network	Ketki Gupte and Linsey Pang	
2:40pm - 2:50pm	<b>Coffee Break</b>			
2:50pm - 3:10pm	S43204	Fingerphoto Presentation Attack Detection: Generalization in Smartphones	Emanuela Marasco and Anudeep Vurity	Dr. Ge Jin
3:10pm - 3:30pm	S43203	ConOffense: Multi-Modal Multitask Contrastive Learning for Offensive Content Identification	Debaditya Shome and T Kar	
3:30pm - 3:50pm	BigD545	Deep Learning Approach Towards Squat Isolation in a Multi-Embedded Track Geometry Defects	Ibrahim Balogun, Mark Leadingham II, Dominique Gulliot, and Nii Attoh-Okine	
3:50pm - 4:05pm	S43209	COVID-19 Multi-Modal Data Analysis with Alexa Voice and Conversational AI Applications	Alex Kaplunovich	
4:05pm - 4:20pm	BigD635	Multi-Modal Hate Speech Detection Using Machine Learning	Fariha Tahosin Boishakhi, Ponkoj Chandra Shill, and Md. Golam Rabiul Alam	

<b>The 5th IEEE International Workshop on Big Data for Financial News and Data</b> <b>Date: Wednesday, Dec. 15, 2021</b> <b>Workshop Chairs: <i>Quanzhi Li, Xiaozhong Liu, Sameena Shah</i></b>			
ID	Time	Title	Presenter/Author
S45204	9:30am - 9:45am	S45204 : <i>Do Candlestick Patterns Work in Cryptocurrency Trading?</i>	Kin-Hon Ho, Tse-Tin Chan, Haoyuan Pan, and Chin Li

S45203	9:45am - 10:00am	S45203 : <i>Customer Segmentation Using Mobile Phone Usage Data to Reveal Finance Application User's Behavior</i>	Gomathi Sengodan
S45201	10:00 - 10:25am	S45201 : <i>Contextual Sentence Analysis for the Sentiment Prediction on Financial Data</i>	Elvys Linhares Pontes and Mohamed Benjannet
S45202	10:25 - 10:40am	S45202 : <i>Multilingual Financial Word Embeddings for Arabic, English and French</i>	Nadhem Zmandar, Mahmoud El-Haj, and Paul Rayson
S45206	10:40 - 11:05am	S45206 : <i>Dimension Estimation using Second Order data in Finance</i>	Nitish Bahadur and Randy Paffenroth
S45205	11:05 - 11:20am	S45205 : <i>Recognition for Creating a Skill Inventory of Today's Workplace</i>	Gjorgjina Cenikj, Bozhanka Vitanova, and Tome Eftimov
BigD360	11:20 - 11:35am	BigD360 : <i>Joint time-series learning framework for maximizing PO renewals</i>	Shubhi Asthana, Pawan Chowdhary, Indervir Singh Banipal, Shikhar Kwatra, and Taiga Nakamura
Closing Remarks			

<b>Workshop on Knowledge Graphs and Big Data</b> <b>In Conjunction with IEEE Big Data 2021</b> <b>Dec. 15, 2021</b> <i>Workshop Co-Chairs: Yuan An, Dejing Dou, Yuan Ling, Alex Kalinowski</i>		
<b>Time (EST)</b>	<b>Title</b>	<b>Presenter/Author</b>
9:00-9:05am	Welcome	Yuan An
9:05-9:30am	Knowledge Graph-based Data Transformation Recommendation Engine (Paper Id: S48201)	Garima Natani and Satoru Watanabe
9:30-9:55am	Trustworthy Knowledge Graph Population From Texts for Domain Query Answering (Paper Id: S48202)	Jing Ao, Swathi Dinakaran, Hongjian Yang, David Wright, and Rada Chirkova
9:55-10:20am	A Semantic Approach to Identifier Management in Engineering Systems (Paper Id: S48203)	Rustam Mehmandarov, Arild Waaler, David Cameron, Roar Fjellheim, and Thomas B. Pettersen
10:20-10:30am	<b>Coffee Break</b>	
10:30-10:55am	Privacy Preserving Survival Prediction (Paper Id: S48204)	Stefano Fedeli, Frida Schain, Sana Imtiaz, Zainab Abbas, and Vladimir Vlassov
10:55-11:20am	Knowledge Graph-Empowered Materials Discovery (Paper Id: S48206)	Xintong Zhao
11:20-11:45am	Knowledge-Oriented Graph-Based Approach to Capture the Evolution of Developers' Knowledge (Paper Id: S48207)	Edson Lucas, Toacy Oliveira, Paulo Alencar, and Donald Cowan
11:45-12:10am	Latent Learning with pyroMind.2020 (Paper Id: BigD282)	Vyacheslav Romanov
<b>12:10-12:30am</b>	<b>Closing Remarks and Future Suggestions</b>	

**2021 National Symposium for NSF REU Research in Data Science, Systems, and Security  
(REU 2021 Symposium)  
December 15, 2021**

*Symposium Website: <https://bigdatareu.umd.edu/reu-symposium>*

*Program Chairs:*

- [Dr. Jianwu Wang](#), University of Maryland, Baltimore County, Organizer of [the NSF REU Site on Online Interdisciplinary Big Data Analytics in Science and Engineering](#)
- [Dr. Matthias K. Gobbert](#), University of Maryland, Baltimore County, Co-organizer of [the NSF REU Site on Online Interdisciplinary Big Data Analytics in Science and Engineering](#)
- [Dr. Nirmalya Roy](#), University of Maryland, Baltimore County, Organizer of [the NSF REU Site on Smart Computing and Communications](#)
- [Dr. Fusheng Wang](#), Stony Brook University, Organizer of [the NSF REU on Big Data, AI and Medical Informatics](#)

Time (ET)	Title	Presenter/Author
<b>Session 1 (Data Science)</b>		
9:00-9:20	SP06209: Drug-Drug Interaction Prediction: a Purely SMILES Based Approach	Bri Bumgardner, Farhan Tanvir, Khaled Mohammed Saifuddin, and Esra Akbas
9:40-10:00	SP06214: Promising Hyperparameter Configurations for Deep Fully Connected Neural Networks to Improve Image Reconstruction in Proton Radiotherapy	Sokhna York, Alina Ali, David Lashbrooke, Rodrigo Yopez-Lopez, Carlos Barajas, Matthias Gobbert, and Jeremy Polf
10:00-10:20	SP06207: Convolutional Neural Network Optimization for Phishing Email Classification	Cameron McGinley and Sergio Salinas Monroy
10:20-10:40	SP06205: Overhead Projection Approach For Multi-Camera Vehicle Activity Recognition	George Strauch, JiaJian Lin, and Jelena Tesic
<b>Coffee Break</b>		
<b>Session 2 (Systems)</b>		
11:00-11:20	SP06213: 3D Modeling of Cities for Virtual Environments	Calvin Davis, Jaired Collins, Emily Lattanzio, Joshua Fraser, Shizeng Yao, Haoxiang Zhang, Bimal Balakrishnan, Ye Duan, Prasad Calyam, and Kannappan Palaniappan
11:20-11:40	SP06211: Woolery: Extending Frame Semantics to Structured Documents	Chloe Eggleston and Jeremy Abramson
11:40-12:00	SP06206: Creation of EMA-KN – A Knowledge Network for Ecological Momentary Assessment	Cade Winters, Justin Ebby Varghese, George Stafford, Fengxiang Zhao, Songxi Chen, and Yi Shang
12:00-12:20	SP06202: Visualizing Internet Measurements of Covid-19 Work-from-Home	Erica Stutz, Yuri Pradkin, Xiao Song, and John Heidemann
<b>Coffee Break</b>		
<b>Session 3 (Security)</b>		
12:40-13:00	SP06208: Investigating the Changes in Software Metrics after Vulnerability is Fixed	Andy Zhou, Kazi Zakia Sultana, and Bharath Samanthula
13:00-13:20	SP06203: Identifying ATT&CK Tactics in Android Malware Control Flow Graph Through Graph Representation Learning and Interpretability	Jeffrey Fairbanks, Andres Orbe, Christine Patterson, Janet Layne, Edoardo Serra, and Marion Scheepers
13:20-13:40	SP06212: Private Blockchain for Visitor Authentication and Access Control	Ka Ying Chan, Mason Lovett, and Yesem Kurt Peker

13:40-14:00	SP06204: A Scalable and Domain Adaptive Respiratory Symptoms Detection Framework using Earables	Nhan Nguyen, Avijoy Chakma, and Nirmalya Roy
14:00-14:20	SP06210: A Fine-grain Geospatial and Demographic Analysis of Breast Cancer Patterns in New York State	Fiona Murphy, Kayley Abell-Hart, and Fusheng Wang
<b>Closing Remarks</b>		

**Advances in High Dimensional (AdHD) Big Data (Advances in High Dimensional (AdHD) Workshop Chairs: Sotiris Tasoulis, Aristidis Vrahatis and Nicos Pavlidis**

<b>Time</b>	<b>Paper ID</b>	<b>Title</b>	<b>Presenter/Author</b>
09:00 – 09:10		Intro to AdHD	
09:10 – 09:40		Keynote Speech	Luca Martino
09:40 – 10:00	S42202	Approximate Top-k Inner Product Join with a Proximity Graph	Hayato Nakama
10:00 – 10:20	S42201	Approximate High Dimensional Graph Mining With Matrix Polar Factorization: A Twitter Application	Georgios Drakopoulos
10:20 – 10:40	S42205	WATCH: Wasserstein Change Point Detection for High-Dimensional Time Series Data	Kamil Faber
10:40 – 11:00		<b>Break</b>	
11:00 – 11:20	S42206	Graph-based clustering for time series data	Peiyu Li
11:20 – 11:40	BigD368	Scalable Outlier Detection Using Distance Projections	Jin Cao
11:40 – 12:00	BigD492	Graph-guided Bayesian SVM with Adaptive Structured Shrinkage Prior for High-dimensional Data	Wenli Sun
12:00 – 12:20	S42207	Recent Dimensionality Reduction Techniques for Visualizing High-Dimensional Parkinson's Disease Omics Data	Georgios Dimitrakopoulos
12:20 – 12:40	BigD610	Cascaded Dimension Reduction for Effective Anomaly Detection	Xiaoyan Zhuo
12:40 – 13:00		<b>Closing</b>	

## Posters

<b>Thursday, December 16<sup>th</sup>, 2021, 9:00am – 11:00am (Chair: Ladjel Bellatreche)</b> <b>Poster Session 01: Big Data Science and Foundations</b>	
Poster ID	Accept Posters
P222	Eric Bax, <i>When to Re-Draw a Sample, and Why</i>
P227	Keqian Li, Yifan Hu, Manisha Verma, Fei Tan, Changwei Hu, Tejaswi Kasturi, and Kevin Yen, <i>Hadoop-MTA: a system for Multi Data-center Trillion Concepts Auto-ML atop Hadoop</i>
P229	Lamine Diop and Cheikh Ba, <i>Parallelization of Sequential Pattern Sampling</i>
P237	Saba Syed and Vaibhav Anu, <i>Digital Evidence Data Collection: Cloud Challenges</i>
P239	Dhaval Patel, Shuxin Li, Srideepika Jayaraman, Giridhar Ganapavarapu, Anuradha Bhamidipaty, and Jayant Kalagnanam, <i>Scaling Anomaly Detection Service Using Serverless Technology</i>
P255	Orestis Rafail Nerantzis, Apostolos Tselios, and Alexandros Karakasidis, <i>MI-OPJ: A Microservices-based Online Programming Judge</i>
P257	Yin-Hei Chan and Andrew Kwok-Fai Lui, <i>Leveraging data sampling patterns in data-driven traffic characteristic modeling</i>
P262	Atsuki Kamo and Saneyasu Yamaguchi, <i>I/O Performance Improvement with Striping File Layout Considering Storage of Intermediate Data</i>
P263	Peiquan Jin, Xiangyu Zhuang, Yongping Luo, and Mingchen Lu, <i>Exploring Index Structures for Zoned Namespaces SSDs</i>
P266	Zexuan LYU, Yuchen WANG, Chao LI, Guigang ZHANG, Chunmei HUANG, Li ZHENG, Mengyuan DU, Yongbo WANG, Qin CHEN, and Weimin LI, <i>DaaS: Internet-perception big data systems based on AI</i>
P268	Gonçalo Carvalho, Jorge Bernardino, Vasco Pereira, and Bruno Cabral, <i>A holistic data modeling approach for multi-database systems</i>
P273	HAFSA EL HAFYANI, Mohammad Abboud, and Yehia Taher, <i>A Microservices Based Architecture for Implementing and Automating ETL Data Pipelines for Mobile Crowdsensing Applications</i>
P283	Mai Bui, Lin-Ching Chang, Hang Liu, Qi Zhao, and Genshe Chen, <i>Comparative Study of 3D Point Cloud Compression Methods</i>
P286	Sami Uddin, Bipasha Banerjee, Jian Wu, William A. Ingram, and Edward A. Fox, <i>Building A Large Collection of Multi-domain Electronic Theses and Dissertations</i>
P291	Lisa Singh, Colton Padden, Pam Davis-Kean, Rabin David, Virinche Marwadi, Yiqing Ren, and Rebecca Vanarsdall, <i>Text analytic research portals: supporting large-scale social science research</i>

<b>Thursday, December 16<sup>th</sup>, 2021, 9:00am – 11:00am (Chair: Yufei Tang)</b> <b>Poster Session 02: Big Data Infrastructure/Management</b>	
Poster ID	Accept Posters
P209	Aws Naser Jaber AL., <i>Efficient Visualization Framework for Real-Time Monitoring Network Traffic of High-Speed</i>
P213	Jalal Khalil, Da Yan, Guimu Guo, Mirza Tanzim Sami, Joy Bhadhan Roy, and Virginia Sisiopiku, <i>Realistic Transport Simulation for Studying the Impacts of Shared Micromobility Services</i>
P214	Gregory Smith, <i>Impact of Organizational Factors on Big Data Analytics Adoptions in U.S. Public Sector Organizations</i>
P216	Britnie Gonzalez-Moodie, Shane Daiek, Jorge Lorenzo-Trueba, and Aparna Varde, <i>Multispectral Drone Data Analysis on Coastal Dunes</i>
P221	Sangwhan Cha, Yunze Tian, Majid Shaalan, Jeongwoo Hyun, and Ki H. Lee, <i>Implementation of Prototype Environment of Virtual Reality Platform for Virtual Lab (Use case of ESL)</i>
P224	Mathieu Roche and Maguelonne Teisseire, <i>Integrating Textual Data into Heterogeneous Data Ingestion Processing</i>
P232	Jieling Wu and Mitsugu Saito, <i>Big Data Analysis for Predicting Post-Earthquake Municipal Road Recovery</i>
P234	Linli Qi, Peiquan Jin, and Shouhong Wan, <i>Efficient Selectivity Estimation for Relation-Tree Joins in Multi-Model Databases</i>
P238	Wali Mohammad Abdullah, David Awosoga, and Shahadat Hossain, <i>Intersection Representation of Big Data Networks and Triangle Counting</i>
P248	Yao Li, Zhenxiao Luo, Chunxu Tang, Mainak Ghosh, Huijun Wu, Lu Zhang, Anneliese Lu, Ruchin Kabra, Nikhil Kantibhai Navadiya, Prachi Mishra, Prateek Mukhedkar, and Vrushali Channapattan, <i>A Performance Evaluation of Spark GraphFrames for Fast and Scalable Graph Analytics at Twitter</i>
P259	Janani Priyanka, Daniel Cheriyan, and Jae-ho Choi, <i>ISSUES WITH CURRENT PM MONITORING TECHNIQUES AND CONTROL MEASURES AND THE WAY FORWARD USING BIG DATA</i>
P261	Naomichi Fukuda, Kenichi Kourai, and Saneyasu Yamaguchi, <i>Improvement of ARC Considering Negative Locality of Reference in Virtualized Environment</i>

P264	Alberto Berenguer, Jose-Norberto Mazón, and David Tomás, <i>Towards a tabular open data search engine for public sector information</i>
P276	Jose-Norberto Mazón, Rob Brennan, and Markus Helfert, <i>Overcoming misattribution to understand open data reuse in smart cities</i>
P280	Nariman Ammar, Parya Zareie, Marion E Hare, Lisa Rogers, Sandra Madubuonwu, Jason Yaun, and Arash Shaban-Nejad, <i>SPACES: Explainable Multimodal AI for Active Surveillance, Diagnosis, and Management of Adverse Childhood Experiences (ACEs)</i>
P293	Torsten Priebe, Sebastian Neumaier, and Stefan Markus, <i>Finding Your Way Through the Jungle of Big Data Architectures</i>

**Thursday, December 16<sup>th</sup>, 2021, 11:00am – 1:00pm (Chair: Ladjel Bellatreche)**  
**Poster Session 03: Big Data as Service for Health Management**

Poster ID	Accept Posters
P203	Soumya Kundu, <i>A Distributed Deep Learning Framework for Federated Big Medical Image Analysis</i>
P235	Vui Huang Tea and Karine Tonson La Tour, <i>Relationship Discovery Graph of EU COVID Unique Vaccination Certificate Identifier (UVCI)</i>
P242	Natasia Fernandez and Vaibhav Anu, <i>Text Mining Algorithms for Cancer Diagnostics</i>
P243	Antonio Iyda Paganelli, Pedro Velmovitsky, Pedro Miranda, Adriano Branco, Paulo Alencar, Donald Cowan, Markus Endler, and Plinio Morita, <i>IoT-Based COVID-19 Health Monitoring System: Context, Early Warning and Self-Adaptation</i>
P244	Admitos-Rafael Passadakos, Anastasios Vlahos, Paraskevi Tzouveli, and Stefanos Kollias, <i>EXTRACTING GEOGRAPHICAL CHARACTERISTICS ABOUT COVID-19 EVOLUTION WORLDWIDE USING MACHINE LEARNING</i>
P246	Thejas Prasad and Shakeel Rashed, <i>Real World Data to Real World Evidence using Machine Learning in Clinical Trial Designs</i>
P247	Ziwei Wang, Minzhu Zhao, Yu Chen, Yunya Song, and Liang Lan, <i>A Study of Cantonese Covid-19 Fake News Detection on Social Media</i>
P250	Tanya Graham and Sean Thorpe, <i>ADAPTIVE BIOMETRIC CLOUD FACIAL AGING TEMPLATE POLICY - A TIME POINT ANALYSIS</i>
P271	Shizuku Iida and Kazuo Hara, <i>Finding Experiential Stress in Tweets by Utilizing both Explicit and Implicit Stress Data</i>
P274	Whitney S Brakefield, Nariman Ammar, and Arash Shaban-Nejad, <i>UPHO: Leveraging an Explainable Multimodal Big Data Analytics Framework for COVID-19 Surveillance and Research</i>
P297	Pedro Velmovitsky, Paulo Alencar, Scott Leatherdale, Donald Cowan, and Plinio Morita, <i>Towards Real-Time Public Health: A Novel Mobile Health Monitoring System</i>

**Thursday, December 16<sup>th</sup>, 2021, 11:00am – 1:00pm (Chair: Yufei Tang)**  
**Poster Session 04: Big Data Search and Mining**

Poster ID	Accept Posters
P205	Chao Zhang, Jaswanth Yelle, Yu Huang, and Sthitie Bom, <i>Learning from Failures in Large Scale Soft Sensing</i>
P207	Mirza Sami, Da Yan, Huang Huang, Xinyu Liang, Guimu Guo, and Zhe Jiang, <i>Drone-Based Tower Survey by Multi-Task Learning</i>
P218	Qiyu Li and Reza Langari, <i>MYOELECTRIC HUMAN COMPUTER INTERACTION USING CNN-LSTM NEURAL NETWORK FOR DYNAMIC HAND GESTURES RECOGNITION</i>
P223	Keqian Li, Kevin Yen, Yifan Hu, Shaunak Misra, Changwei Hu, and Manisha Verma, <i>BAN: Large Scale Brand ANonymization for Creative Recommendation Via Label Light Adaptation</i>
P226	Xueqin DU, Yulin HE, and Joshua Zhexue HUANG, <i>Random Sample Partition-Based Clustering Ensemble Algorithm for Big Data</i>
P228	Jiayin YU, Yulin HE, and Joshua Zhexue HUANG, <i>A Two-Stage Missing Value Imputation Method Based on Autoencoder Neural Network</i>
P236	Jiamu Li, Ji Zhang, Jian Wang, Youwen Zhu, Mohamed Jaward Bah, Gaoming Yang, and Yuquan Gan, <i>VAGA: Towards Accurate and Interpretable Outlier Detection Based on Variational Auto-Encoder and Genetic Algorithm for High-Dimensional Data</i>
P245	Xinlan Tan, Zaid Shaikh, Alice Mello, and Fiona Creed, <i>The UN General Assembly Voting Coincidence and Clustering Analysis</i>
P251	Dev Dave, Vaibhav Anu, and Aparna Varde, <i>Automating the Classification of Requirements Data</i>



P252	Hani Ramadhan and Joonho Kwon, <i>Enhancing Learned Index for A Higher Recall Trajectory K-Nearest Neighbor Search</i>
P253	Hongxuan Huang, Qingyuan Linghu, Fan Zhang, Dian Ouyang, and Shiyu Yang, <i>Truss Decomposition on Multilayer Graphs</i>
P260	Zhenjie Yao, Yixin Chen, Jinwei Wang, Shouling Wu, Yanhui Tu, Minghui Zhao, and Luxia Zhang, <i>Trend Analysis Neural Networks for Interpretable Analysis of Longitudinal Data</i>
P272	Valsamis Ntouskos, Chrysa Iliopoulou, and Konstantinos Karantzas, <i>Production Machine Learning Frameworks for Geospatial Big Data</i>
P282	Nikolas Burzynski, Yuhao Yuan, Adiel Felsen, David Reitano, Zhibo Wang, Khalid Sethi, Fake Lu, and Kenneth Chiu, <i>Deep Learning Techniques for Unmixing of Hyperspectral Stimulated Raman Scattering Images</i>
P289	Terrence Fries, <i>Big Data Analytics Using Fuzzy Clustering for Network Security</i>
P290	Kaito Kawashima and Saneyasu Yamaguchi, <i>Understanding Language Model from Questions in Social Studies for Students</i>
P294	Mengmeng Zhu, Shouhong Wan, Peiquan Jin, and Qijun Tian, <i>A Feature Fusion Method Based on Multi-Classification Losses for Fine-Grained Visual Categorization</i>

**Thursday, December 16<sup>th</sup>, 2021, 1:00pm – 3:00pm (Chair: Ladjel Bellatreche)**  
**Poster Session 05: Big Data Applications I**

Poster ID	Accept Posters
P201	Arman Abdoli, MohammadReza Majma, and Parisa Rahmani, <i>Can Time Series Data Mining Facilitate Wider Adoption of Mobile Health?</i>
P204	Osvaldo Gogliano Sobrinho, Liedi Légi Bariani Bernucci, Pedro Luiz Pizzigatti Corrêa, Rosângela dos Santos Motta, Jeaneth Machicao, Angelo Samuel Junqueira, Robson Correia da Costa, Wellington Dias de Queiroz, Thales Cesar Giriboni de Mello e Silva, Pedro Lopes Ferraz, Luciano Oliveira, and Luciano Cassaro, <i>Big data analytics in support of the under-rail maintenance management at Vitória – Minas Railway</i>
P208	Jianxin Sui, Xiao Ling, Xing Xiang, Genwei Zhang, and Xiangchi Zhang, <i>Transformer Oil Temperature Prediction Based on Long and Short-term Memory Networks</i>
P217	Rui Li and Weitian Wang, <i>Augmenting the Communication Naturalness via an Audio-Visual Virtual Agent for Collaborative Robots</i>
P225	Tiago Coelho and Alvaro Figueira, <i>Analysis of Top-Ranked HEI Publications' Strategy on Twitter</i>
P230	Sean Thorpe, <i>TOWARDS A CYBER AWARE CHATBOT SERVICE</i>
P233	Amulya Kantharaj and Alireza Abdoli, <i>Improving Aviation and Flight Safety By Analyzing Surveillance Data</i>
P240	Ryuta Yamaguchi, Panote Siriaraya, Da Li, Tomoki Yoshihisa, Shinji Shimojo, and Yukiko Kawai, <i>A Proposal of Data Collection by Mobility Users on Real-time for e-Bike Navigation System</i>
P241	Thanapol Phungtua-eng, Yoshitaka Yamamoto, and Shigeyuki Sako, <i>Dynamic Binning for the Unknown Transient Patterns Analysis in Astronomical Time Series</i>
P249	Toshiki Yamaguchi, Kazuo Hara, and Ikumi Suzuki, <i>Robust Method to Convert HIRAGANA Sequences into Japanese Text</i>
P254	Seungjoo Lee and Minsoo Park, <i>Understanding EV Market Trend: Using Time Series Dynamic Topic Modeling with Youtube Data</i>
P256	Tomoya Sasaki, Arisa Nakamura, Jun-ichi Harasawa, Kazuo Hara, Ikumi Suzuki, and Tatsuhito Takahashi, <i>Bayesian Optimization with an Auxiliary Classifier for the Development of Polymer Materials</i>

**Thursday, December 16<sup>th</sup>, 2021, 1:00pm – 3:00pm (Chair: Yufei Tang)**  
**Poster Session 06: Big Data Applications II**

Poster ID	Accept Posters
P258	Jin young Choi, Eun-Ju Ha, and Jae-ho Choi, <i>Data Mining Approach for Characterizing Construction Accidents</i>
P265	Angela Geronazzo and Markus Ziegler, <i>QMLEx: Data Driven Digital Transformation in Marketing Analytics</i>
P270	Abhishek Saxena and Stefan A. Robila, <i>Analysis of the New York City's Vehicle Crash Open Data</i>
P275	Marius E. Penteliuc, <i>Processing Large Satellite Imagery to Estimate Solar Irradiance</i>
P277	Kambiz Vahedi and Khadijeh Afhamisizi, <i>Cloud Based Malware Detection Through Behavioral Entropy</i>
P284	Adiel Felsen, Yuhao Yuan, Nikolas Burzynski, David Reitano, Zhibo Wang, Khalid Sethi, Fake Lu, and Kenneth Chiu, <i>Cell Nuclei and Lipid Droplets Quantification in Stimulated Raman Images</i>
P287	Jose Delgado, Xingquan Zhu, Karin Scarpinato, Jason O. Hallstrom, and Terje Hill, <i>Understanding and Predicting Faculty Success in Winning Grant Awards</i>

P288	Jai Radhakrishnan, Daniel Uyematsu, Martin Collazo, Yunfei Hou, and Mariam Salloum, <i>Towards Predicting Bus On-Time Performance in the Inland Empire</i>
P292	Ethan Zhu and Dingding Wang, <i>Small Data for Cyberbullying Prediction</i>
P295	Yasuko Kawahata, <i>Motorization Utilizing Spatio-Temporal Information Case Study of Domestic and International Cross-Border Mobility Trends In Cities</i>
P296	Filip Markoski, Lasko Basnarkov, Biljana Stojkoska, Petre Lameski, and Eftim Zdravevski, <i>Are central-zone restaurants better for consumers? - An analytical approach</i>

# IEEE BigData 2022

## IEEE BIGDATA 2022

Middle December, 2022, Tokyo, Japan

The IEEE Big Data conference series is a leading forum for disseminating the latest advances in big data research, development and application. We solicit high-quality original research papers (including significant work-in-progress) in any aspect of Big Data with emphasis on 5Vs (Volume, Velocity, Variety, Value and Veracity): big data science and foundations, big data infrastructure, big data management, big data searching and mining, big data privacy/security, and big data applications. Relevant topics include but are not limited to:

### 1. Big Data Science and Foundations

- a. Novel Theoretical Models for Big Data
- b. New Computational Models for Big Data
- c. Data and Information Quality for Big Data
- d. New Data Standards

### 2. Big Data Infrastructure

- a. Cloud/Grid/Stream Computing for Big Data
- b. High Performance/Parallel Computing Platforms for Big Data
- c. Autonomic Computing and Cyber-infrastructure, System Architectures, Design and Deployment
- d. Energy-efficient Computing for Big Data
- e. Programming Models and Environments for Cluster, Cloud, and Grid Computing to Support Big Data
- f. Software Techniques and Architectures in Cloud/Grid/Stream Computing
- g. Big Data Open Platforms
- h. New Programming Models for Big Data beyond Hadoop/MapReduce, STORM
- i. Software Systems to Support Big Data Computing

### 3. Big Data Management

- a. Advanced database and Web Applications
- b. Novel Data Model and Databases for Emerging Hardware
- c. Data Preservation
- d. Data Provenance
- e. Interfaces to Database Systems and Analytics Software Systems
- f. Data Protection, Integrity and Privacy Standards and Policies
- g. Information Integration and Heterogeneous and Multi-structured Data Integration
- h. Data management for Mobile and Pervasive Computing
- i. Data Management in the Social Web
- j. Crowdsourcing
- k. Spatiotemporal and Stream Data Management
- l. Scientific Data Management
- m. Workflow Optimization
- n. Database Management Challenges: Architecture, Storage, User Interfaces

### 4. Big Data Search and Mining

- a. Social Web Search and Mining
- b. Web Search
- c. Algorithms and Systems for Big Data Search
- d. Distributed, and Peer-to-peer Search
- e. Big Data Search Architectures, Scalability and Efficiency
- f. Data Acquisition, Integration, Cleaning, and Best Practice
- g. Visualization Analytics for Big Data
- h. Computational Modeling and Data Integration
- i. Large-scale Recommendation Systems and Social Media Systems
- j. Cloud/Grid/Stream Data Mining- Big Velocity Data

- k. Link and Graph Mining
- l. Semantic-based Data Mining and Data Pre-processing
- m. Mobility and Big Data
- n. Multimedia and Multi-structured Data- Big Variety Data

### 5. Big Data Security, Privacy & Trust

- a. Intrusion Detection for Gigabit Networks
- b. Anomaly and APT Detection in Very Large Scale Systems
- c. High Performance Cryptography
- d. Visualizing Large Scale Security Data
- e. Threat Detection using Big Data Analytics
- f. Privacy Threats of Big Data
- g. Privacy Preserving Big Data Collection/Analytics
- h. HCI Challenges for Big Data Security & Privacy
- i. User Studies for any of the above
- j. Sociological Aspects of Big Data Privacy

### 6. Big Data Learning and Analytics

- a. Predictive analytics on Big Data
- b. Machine Learning algorithms for Big Data
- c. Deep Learning for Big Data
- d. Feature representation learning for Big Data
- e. Dimension reduction for Big Data
- f. Physical informed Big Data learning

### 7. Big Data Applications

- a. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication
- b. Big Data Analytics in Small Business Enterprises (SMEs)
- c. Big Data Analytics in Government, Public Sector and Society in General
- d. Real-life Case Studies of Value Creation through Big Data Analytics
- e. Big Data as a Service
- f. Big Data Industry Standards
- g. Experiences with Big Data Project Deployments

### INDUSTRIAL and GOVERNMENT Track

The Industrial and government Track solicits papers describing implementations of Big Data solutions relevant to industrial settings. The focus of industry track is on papers that address the practical, applied, or pragmatic or new research challenge issues related to the use of Big Data in industry. We accept full papers (up to 10 pages) and extended abstracts (2-4 pages).