

**2025 IEEE 11th International
Conference on Big Data
Computing Service and Machine
Learning Applications
(BigDataService 2025)**

**Tucson, Arizona, USA
21-24 July 2025**



**IEEE Catalog Number: CFP25A91-POD
ISBN: 979-8-3315-8533-4**

**Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP25A91-POD
ISBN (Print-On-Demand):	979-8-3315-8533-4
ISBN (Online):	979-8-3315-8532-7
ISSN:	2996-2269

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2025 IEEE 11th International Conference on Big Data Computing Service and Machine Learning Applications (BigDataService) **BigDataService 2025**

Table of Contents

Welcome Message from the General Chairs	xi
Committees	xii

Invited Papers

Scalable Framework for Behavior Execution of Mobility and Transport Digital Twins	1
<i>Pierfrancesco Bellini (University of Florence, DISIT lab, Italy), Luciano Alessandro Ipsaro Palesi (University of Florence, DISIT lab, Italy), Fabrizio Mereu (University of Florence, DISIT lab, Italy), and Paolo Nesi (University of Florence, DISIT lab, Italy)</i>	
An Approach to Transiting from Divergent Thinking to Convergent Thinking for Creativity	10
<i>Chong Zeng (University of Leicester, England) and Hongji Yang (University of Leicester, England)</i>	
An Automated Framework for Unsupervised Binary Class Distribution Estimation	17
<i>Mary Anne Walauskis (Florida Atlantic University) and Taghi M. Khoshgoftaar (Florida Atlantic University)</i>	
Finding Optimal Teams: An Analysis of NBA Statistics and Constraints	26
<i>Shaykh Siddique (Prairie View A&M University, USA), Lin Li (Prairie View A&M University, USA), and Yonghui Wang (Prairie View A&M University, USA)</i>	
Deep Learning-Based Anomaly Detection and Object Classification for Campus Surveillance	35
<i>Shakshi Richhariya (San Jose State University, USA), David Wang (San Jose State University, USA), and Jerry Gao (San Jose State University, USA)</i>	

Session 1 - Smart cities

Efficient and Scalable Semantic Data Ingestion for Smart City Digital Twin Platforms	44
<i>Pierfrancesco Bellini (UNIFI, DISIT lab, Snap4City, Italy), Paolo Nesi (UNIFI, DISIT lab, Snap4City, Italy), Enrico Collini (UNIFI, DISIT lab, Snap4City, Italy), Marco Fanfani (UNIFI, DISIT lab, Snap4City, Italy), and Christian Panconi (UNIFI, DISIT lab, Snap4City, Italy)</i>	

A Smart City Cloud Platform for Road Inspection and Analysis Using Machine Learning	52
<i>Sweekruthi Balivada (San Jose State University, USA), Yuting Sha (San Jose State University, USA), Manisha Lagisetty (San Jose State University, USA), Damini Vichare (San Jose State University, USA), and Jerry Gao (San Jose State University, USA)</i>	
Modeling Perceived Exertion with Deep Neural Networks and Wearable Sensors	60
<i>Aref Smiley (dept. Biomedical Informatics, The University of Utah School of Medicine, USA) and Joseph Finkelstein (The University of Utah School of Medicine, USA)</i>	

Session 2 - Health and medical informatics

A Lightweight and Interpretable Ensemble of CNN and Transformer Models for IDC Breast Cancer Classification	66
<i>Faith Tobore Edafetanure-Ibeh (Harrisburg University of Science and Technology, USA), Sonachi Bukola Mogbogu (Harrisburg University of Science and Technology, USA), and Ralchukwu Mogbogu (Harrisburg University of Science and Technology, USA)</i>	
Advanced Anomaly Detection Framework for Enhancing Prediction of Severe Health Outcomes in Cancer Patients Undergoing Radiotherapy	73
<i>Yang Yan (Southern Illinois University), Christopher Lominska (University of Kansas Medical Center), Gregory N. Gan (University of Kansas Medical Center), Hao Gao (University of Kansas Medical Center), and Zhong Chen (Southern Illinois University)</i>	
Automated Medical Coding Using a Hybrid Decision Tree with Deep Learning Nodes	81
<i>Spoorthi Bhat (University of Massachusetts Dartmouth, USA), Haiping Xu (University of Massachusetts Dartmouth, USA), and Joshua Carberry (University of Massachusetts Dartmouth, USA)</i>	
Hybrid Deep Learning for Early Detection of Symptom Escalation in Patients with Cancer	89
<i>Joseph Finkelstein (The University of Utah, USA), Aref Smiley (The University of Utah, USA), Xingyue Huo (The University of Utah, USA), Christina Echeverria (The University of Utah, USA), and Kathi Mooney (The University of Utah, USA)</i>	

Session 3 - Systems and architectures

AutoDrift: A Forecast-Aware Concept Drift Detection and Retraining Pipeline in MLOps with CMAPSS	94
<i>Raj Kumar Myakala (CVS Health, USA), Praveen Kumar Nagata (Trine University, USA), Sampath Lavudya (Campbellsville University, USA), Sanjeev Kumar Pedhapally (University of North Texas, USA), and Vinithya Reddy Podduturi (World Bank, USA)</i>	
Deep Learning Enhanced Load Balancing with NeuroPointSelect for Edge Computing	101
<i>Yeswanth Surya Srikar Nuchu (Software Engineer II, Texas) and Devi Navya Nuchu (Senior Software Engineer, Texas)</i>	

Collaborative Reasoning Framework for Edge-Deployable EEG Sleep Staging via Local LLM	108
<i>Cheng-Lin Cheng (National Central University, Taiwan), Ting-Chuan Lin (National Central University, Taiwan), and Chia-Kai Chang (National Central University, Taiwan)</i>	
Distilling Big Data Outlier Detection Model by Bayesian Spline-Based Score for Portable Service	113
<i>Angela Li (Stony Brook University, USA), Xinyan Cui (Yeshiva University, USA), and David Li (Yeshiva University, USA)</i>	

Session 4 - Innovative Applications of AI, Optimization, and Analytics

Unraveling the Impact of Curricular Complexity on Graduation Time: A Causal Analysis in Higher Education	118
<i>Ameer Slim (University of New Mexico, USA), Ahmad Slim (University of New Mexico, USA), and Charla Orozco (University of New Mexico, USA)</i>	
A Mixed-Integer Optimization Framework for Degree Audit and Course Allocation in Higher Education	126
<i>Ahmad Slim (University of New Mexico, USA), Chaouki Abdallah (Lebanese American University, Lebanon), Michael Hickman (University of New Mexico, USA), Elisha Allen (University of New Mexico, USA), Husain Al Yusuf (University of Arizona, USA), and Ameer Slim (University of New Mexico, USA)</i>	
Culturally-Aware Multiclass Bangla Text-to-Image Generation via Fine-Tuned Stable Diffusion Turbo	134
<i>Susmita Mondal Sristi (Chittagong University of Engineering and Technology, Bangladesh) and Moumita Sen Sharma (Kansas State University, USA)</i>	
Stock Market Forecasting with Pretrained Deep Learning Models	142
<i>Son Tung Do (Fordham University, USA), Anthony Chu (Stanford University, USA), Yijun Zhao (Fordham University, USA), and Yanjun Li (Fordham University, USA)</i>	

Session 5 - Graph learning

Multirelational Twitter Bot Detection using Graph Neural Networks	147
<i>Ketan Jadhav (San Jose State University, USA), Petros Potikas (National Technical University of Athens, Greece), Chris Pollett (San Jose State University, USA), and Katerina Potika (San Jose State University, USA)</i>	
An Exploration of Integrative Analysis of Simultaneously Profiled scHi-C and scRNA-seq Data Using Graph Convolutional Network	155
<i>Utsha Saha (North Dakota State University, USA) and Lu Liu (North Dakota State University, USA)</i>	
Graph-Based User and Content Embeddings for Improved Content Ranking	163
<i>Mariia Bulycheva (Zalando SE, Germany) and Eugene Alooef (R&D Department ATEK, Poland)</i>	

Poster Session

AI-Based Steam Estimation for Excess Air Ratio Optimization in Industrial Boilers	168
<i>Kiwoong Kwon (Korea Electronics Technology Institute, South Korea), Youngmin Joo (Korea Electronics Technology Institute, South Korea), Hyun Woo (Korea Electronics Technology Institute, South Korea), and Seunghyeon Park (Korea Electronics Technology Institute, South Korea)</i>	
Load Forecasting in Power Grids Using Spatio-Temporal Graph Neural Networks	170
<i>Seunghyeon Park (Korea Electronics Technology Institute, Korea), Kiwoong Kwon (Korea Electronics Technology Institute, Korea), and Yongho Kim (Korea Electronics Technology Institute, Korea)</i>	
Selection of Lightweight AI Models for CCTV-Based Occupancy Estimation in Buildings on Edge Devices	172
<i>Hyun Woo (Korea Electronics Technology Institute, South Korea), Seunghyeon Park (Korea Electronics Technology Institute, South Korea), Youngmin Joo (Korea Electronics Technology Institute, South Korea), and Kiwoong Kwon (Korea Electronics Technology Institute, South Korea)</i>	
Toward a Hybrid Ontology Framework for Semantic Data Understanding in AI-Augmented Data Warehousing	174
<i>Hayagriv Sridharan (Chicory AI, USA), Saurabh Sarkar (Chicory AI, USA), Parag Paul (RIG AI, USA), and Eugene Cheng (Chicory AI, USA)</i>	
A Scalable Orchestration Architecture for Managing AI Models at Scale on Kubernetes Platforms	176
<i>Franklin Jebadoss Mohandoss (Dell Technologies, USA)</i>	

Session 7 - Recommendation systems & personalization

Multimodal Benchmarking and Recommendation of Text-to-Image Generation Models	179
<i>Kapil Wanaskar (San Jose State University, USA), Gaytri Jena (Independent researcher, USA), and Magdalini Eirinaki (San Jose State University, USA)</i>	
Domain-Specific Graph RAG Pipelines: Optimized Approaches for Building Efficient Personal Knowledge Repositories	187
<i>Omkar Yadav (San José State University, USA) and William Andreopoulos (San José State University, USA)</i>	
Enhancing Recommender Systems Using Graph Neural Networks, Long Short-Term Memory and Textual Embeddings	195
<i>Tianxiang Chen (San Jose State University, USA) and William B. Andreopoulos (San Jose State University, USA)</i>	

Session 8 - Image processing & analysis applications

DREPADetect: An Optimized YOLOv8 Model for High-Precision Detection of Sickle Cell	203
<i>Abdourahmane Balde (Assane Seck University, Sénégal), Simon Schabat (Cell Environment, France), Awele Bassene (Cheikh Anta Diop University, Senegal), Youssou Faye (Assane Seck University, Senegal), Dieu-Donné Okalas Ossami (University of Science and Technology of Masuku, Gabon), Ousmane Sall (Cheikh Hamidou Kane Digital, Sénégal), Mamadou Soumboundou (Iba Der Thiam University, Sénégal), Bruno Colicchio (IRIMAS Laboratory, University of Haute Alsace, France), and Radhia M'kacher (Cell Environment, France)</i>	
X-ray Illicit Object Detection Using Hybrid CNN-Transformer Neural Network Architectures	210
<i>Jorgen Cani (Harokopio University of Athens, Greece), Christos Diou (Harokopio University of Athens, Greece), Spyridon Evangelatos (Netcompany-Intrasoft S.A., Luxembourg), Panagiotis Radoglou-Grammatikis (K3Y Ltd, Bulgaria; University of Western Macedonia, Greece), Vasileios Argyriou (Kingston University, United Kingdom), Panagiotis Sarigiannidis (University of Western Macedonia, Greece), Iraklis Varlamis (Harokopio University of Athens, Greece), and Georgios Th Papadopoulos (Harokopio University of Athens, Greece)</i>	
Optimized Face Recognition Using Reinforcement Learning and Deep Learning Feature Extraction	218
<i>Youssef Harrath (Dakota State University, USA), Muhammad Bhutta (Dakota State University, USA), Oswald Adohinzin (Dakota State University, USA), and Nirvik KC (Dakota State University, USA)</i>	
Laparoscopic Image Desmoking Using the U-Net with New Loss Function and Integrated Differentiable Wiener Filter	226
<i>Chengyu Yang (New Jersey Institute of Technology, USA) and Chengjun Liu (New Jersey Institute of Technology, USA)</i>	

Session 9 - Novel ML methods

Enhancing Machine Learning Models using a Hierarchical Two-Level Genetic Algorithm for Feature Engineering	231
<i>Armin Felahatpisheh (University of Winnipeg, Canada), Aditi Gulati (Indira Gandhi Delhi Technical University for Women, India), and Camilo E. Valderrama (University of Winnipeg, Canada)</i>	
A Scalable Binary CUR Low-Rank Approximation Algorithm	239
<i>Bowen Su (Michigan State University, USA)</i>	
A Federated Learning Study on Bias and Fairness in Small Data Medical Applications	244
<i>Pratik Korat (San Jose State University, USA), Kirthika Ashokkumar (San Jose State University, USA), Sakina Rahman (San Jose State University, USA), Palak Agarwal (San Jose State University, USA), and Mahima Agumbe Suresh (San Jose State University, USA)</i>	

Session 10 - Big Data analysis and security

Leveraging Big Data Analytics to Detect Backroom Inaccuracies at Target	249
<i>Swapnil Deshpande (Product Availability Target Data Sciences, USA), Xinbo Wang (Product Availability Target Data Sciences, USA), and Timothy Murray (Product Availability Target Data Sciences, USA)</i>	
Federated Analytics for Cyber Threats at Scale	256
<i>Denzel Prince (Howard University, USA), Kesi Chapman (Howard University, USA), Aprille Thomas (Howard University, USA), Dylan Canel (Howard University, USA), and Jeremy Blackstone (Howard University, USA)</i>	
Efficient Organization and Search for Growing Datasets: A Benchmarking Framework	264
<i>Andris Docaj (Texas Tech University, USA) and Yu Zhuang (Texas Tech University, USA)</i>	
Maxie: Distributed Threat Attribution Calculus	272
<i>Abishay Reddy (University of Maryland College Park, USA) and Nate Taylor (Princeton University, USA)</i>	
Author Index	277