2026 8th International Conference on Image, Video and Signal Processing (IVSP 2026)

March 17-19, 2026

Meiji University Surugadai Campus, Tokyo, Japan

Supported by

明治大学 MEIJI UNIVERSITY

Sensors and Systems Society of Singapore

Technically Supported by

Saitama University

DSU Dongseo University

RIDER UNIVERSITY

澳門理工大學 Universidade Politécnica de Macau Macao Polytechnic University

UTAR

UNIMAS UNIVERSITI MALAYSIA SARAWAK

UTHM Universiti Tun Hussein Onn Malaysia

Northumbria University NEWCASTLE

UNG UNIVERSITY of NORTH GEORGIA

SACRAMENTO STATE

Indexed by

Compendex on Engineering Village

Scopus

Table of Contents

Welcome Message

On behalf of the organizing committee, it is our great honor and pleasure to welcome you to the 2026 8th International Conference on Image, Video, and Signal Processing (IVSP 2026), taking place from 17th to 19th March 2026 at the Meiji University Surugadai Campus, Tokyo, Japan.

IVSP 2026 brings together leading experts, researchers, and practitioners from around the globe to share cutting-edge advancements, innovative ideas, and collaborative insights in the fields of image, video, and signal processing. This conference serves as a premier platform for fostering interdisciplinary discussions, exploring emerging technologies, and addressing the challenges and opportunities in these rapidly evolving domains.

We are delighted to host this event at the prestigious Meiji University, a hub of academic excellence and innovation. The vibrant city of Tokyo, with its rich cultural heritage and technological advancements, provides an inspiring backdrop for this gathering of minds.

Over the course of three days, you will have the opportunity to engage in keynote speeches, technical sessions, workshops, and networking events, all designed to facilitate knowledge exchange and collaboration. We are confident that the diverse perspectives and expertise represented here will lead to meaningful discussions and groundbreaking outcomes.

On behalf of all the conference committees, we would like to thank all the authors as well as the technical program committee members and reviewers. Their high competence, their enthusiasm, their time and expertise knowledge, enabled us to prepare the high-quality final program and helped to make the conference become a successful event.

We wish you a productive, inspiring, and enjoyable conference experience. Welcome to IVSP 2026!

IVSP 2026 Organizing Committee
March 2026

Conference Committees

Conference Chairs
Kiyoshi Hoshino, Meiji University, Professor Emeritus of University of Tsukuba, Japan
Jun Sakakibara, Meiji University, Japan

Conference Co-chair
Hideki Nakayama, The University of Tokyo, Japan

Program Co-Chairs
Vijayan Asari, University of Dayton, USA (SPIE Fellow)
Jae-Ho Han, Korea University, Korea

Organizing Committee Chair
Hum Yan Chai, Universiti Tunku Abdul Rahman, Malaysia

Steering Chairs
Ke-Lin Du, Concordia University, Canada
Caifeng Shan, Nanjing University, China

Chapter Chair
Tetsuya Shimamura, Saitama University, Japan

Technical Committee Chair
Zhiyu Jiang, Northwestern Polytechnical University, China

Publicity Chairs
Hideaki Kimata, Kogakuin University, Japan
Seong Tae Kim, Kyung Hee University, Korea

Technical Program Committees
Yuan-Kai Wang, Fu Jen Catholic University
Xiaochen Yuan, Macao Polytechnic University, China
Ming-Han Tsai, National Yang Ming Chiao Tung University
Md Liakat Ali, Rider University, USA
Waleed Abdulla, The University of Auckland, New Zealand
Rebeka Sultana, Tokyo University of Agriculture and Technology, Japan
Qiang Li, Northwestern Polytechnical University, China

Tengku Mohd Afendi Zulcaffle, Universiti Malaysia Sarawak, Malaysia
Roziana Binti Ramli, Northumbria University, UK
Gianluca Zaza, University of Bari Aldo Moro, Italy
Ahmad Al Badawi, Applied Cryptography, USA
Fangli Ying, East China University of Science and Technology, China
Hamimah binti Ujir, Universiti Malaysia Sarawak, Malaysia
Suk-Ho Lee, Dongseo University, Korea
Sukanya Phongsuphap, Mahidol University, Thailand
Francesco Zirilli, Universita di Roma La Sapienza, Italy
Xinwei Luo, Southeast University, China
Nabilah Ibrahim, Universiti Tun Hussein Onn Malaysia, Malaysia
Patrice Jean Delmas, The University of Auckland, New Zealand
Evgin Goceri, Akdeniz University, Turkey
Mariofanna Milanova, University of Arkansas at Little Rock, USA
Jianjun Yang, University of North Georgia, USA
Manuela Chessa, University of Genoa, Italy
Yi Chang, Huazhong University of Science and Technology, China
Kin Chung Kwan, California State University Sacramento, USA
Danilo Avola, Sapienza University, Italy
António J. R. Neves, University of Aveiro, Portugal
Shinnosuke Takamichi, Keio University, Japan
Ku Chin Soon, University of Tunku Abdul Rahman, Malaysia
Baoli Yao, XIOPM of Chinese Academy of Sciences, China
Dandan Ma, Northwestern Polytechnical University, China
Haonan Su, Xi'an University of Technology, China
Younhyun Jung, Gachon University, Republic of Korea
Cheng-Ta Huang, Yuan Ze University
Pei-Jung Lin, Feng Chia University
Irmawati, Universitas Multimedia Nusantara, Indonesia
Shyi-chyi Cheng, National Taiwan Ocean University
Guo-siang Lin, National Chin-Yi University of Technology
Feng-Cheng Lin, Feng Chia University
Yu-Ting Tsai, Yuan Ze University
Ali Reza Alaei, Southern Cross University, Australia
Sirikan Chucherd, Mae Fah Luang University, Thailand
Yi-Jheng Huang, Yuan Ze University
Rohan Loveland, New College of Florida, USA
Ran-Zan Wang, Yuan Ze University

Guideline for Onsite Attendance

Important Notes

- Please enter the meeting room at least 15 minutes before your session. Your punctual arrival and active involvement will be highly appreciated.
- Please wear your name tag for all the conference activities. Lending it to others is not allowed. If you have any companying person, please do inform our staff in advance.
- Please keep all your belongings (laptop and camera etc.) at any time. The conference organizer does not assume any responsibility for the loss of personal belongings.
- Please show name tag and meal coupons when dining.
- Due to force majeure including but not limited to earthquake, natural disaster, war and country policy, the organizer reserves the rights to change the conference dates or venue with immediate effect and takes no responsibility.

Oral & Poster Presentation

- Regular oral presentation: 13 minutes (including Q&A).
- Get your presentation PPT or PDF files prepared. Presentations MUST be uploaded at the session room at least 15 minutes before the session starts.
- Laptop (with MS-Office & Adobe Reader), projector & screen, laser pointer will be provided in all oral session rooms.
- Poster Presenters should bring your poster to the conference venue and put it on designated place.

Guideline for Online Attendance

Platform: Zoom

- Step 1: Download Zoom from the link: https://zoom.us/download

How to use Zoom

* A Zoom account is not required if you join a meeting as a participant, but you cannot change the virtual background or edit the profile picture.

- Rename: Before you enter the conference room, please change your name to Paper ID + Name
- Chat and raise your hand: During the session, if you have any questions, please let us know by clicking “raise your hands” and use “chat” to communicate with conference secretary.
- When you deliver your online speech, please open your camera.
- During the Question section, if you have any questions about keynote speakers or authors, you can also click “raise your hands” or “chat”
- Share Screen: Please open your power point first, and then click “share screen” when it’s your turn to do the presentation.

How to join the conference online

- Find your paper ID and suitable meeting ID on the conference program.
- Open the Zoom, click the join, paste the meeting ID, then you can join the conference.
- Click the stop share after you finish your presentation

Time Zone

- **Japan Standard Time (GMT+9)**

Device

- A computer with an internet connection (wired connection recommended)
- USB plug-in headset with a microphone (recommended for optimal audio quality)
- Webcam: built-in or USB plug-in

Online Room Information

Online Room Information

Zoom ID: **841 3471 6436**

Zoom Link: **https://us02web.zoom.us/j/84134716436**

* Please rename your Zoom Screen Name in below format before entering meeting room.

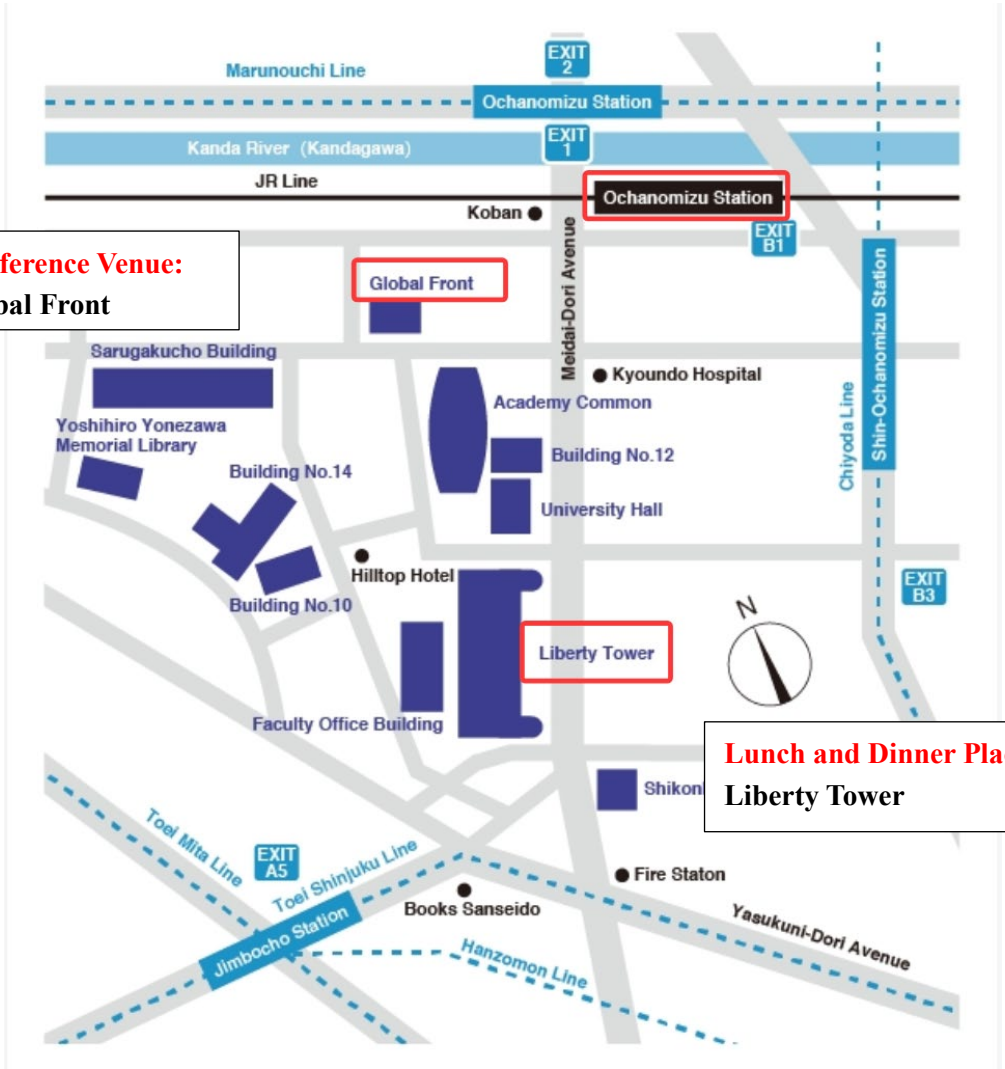
Role	Format	Example
Conference Committee	Position-Name	Conference Chair-Prof.
Keynote/ Invited Speaker	Position-Name	Keynote/Invited Speaker-Prof.
Author	Session Number-Paper ID-Name	S1-BH0001-Name
Delegate	Delegate-Name	Delegate-Name

Conference Venue

Meiji University Surugadai Campus

https://www.meiji.ac.jp/cip/english/about/campus/su_campus.html

Address: 1-1 Kanda-Surugadai, Chiyoda-ku, Tokyo



The organizer doesn't provide accommodation, and we suggest you make an early reservation.

Conference Building

Lunch and Dinner Place:
Liberty Tower

Conference Venue:
Global Front

1. Kishimoto Tatsuo HALL [23F]
2. Dining Hall (Sky Lounge Akatsuki) [17F]
3. Classroom
4. Hidamari Plaza
5. Central Library [1F—B3F]
6. Liberty Academy
7. Portrait relief of the founders of Meiji University
8. Meiji University Museum / AKU YOU Memorial Museum [B1F]
9. Shikon Hall [1F]
10. MEIJI UNIVERSITY Mental Health Clinic for Children and Adolescents
11. International Student Office [2F]
12. International Lounge [2F]

- A Lounge Marronnier
- B Café Pensée
- C Meidai Mart

Simple Program

March 17, 2026 (Tuesday)

Onsite Registration

Registration Time: 10:00-17:00
Venue: Global Front Building, Meiji University Surugadai Campus **Conference Room: 3rd Floor, Room 403A**
Address: 1-1 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8301
1. Arrive at Conference Room (403A), Global Front, Meiji University Surugadai Campus; **2. Inform the conference staff of your paper ID;** **3. Sign your name on the Participants list;** **4. Sign your name on Lunch & Dinner requirement list;** **5. Check your conference kits;** **6. Finish registration.**

Online Test
Time Zone: GMT+9

Online Test **Zoom ID: 841 3471 6436** **Link: https://us02web.zoom.us/j/84134716436**	**Duration**
Prof. Xudong Jiang	10:00-10:10
Session 12 (BH5093, BH5107, BH5105, BH0066, BH0011, BH5033)	10:10-10:40
Prof. Gonzalo Arce	10:40-10:50
Assoc. Prof. Ir. Dr. Kim Seng Chia	10:50-11:00
Prof. Maxim Bakaev	11:00-11:10
Session 13 (BH5078, BH5014, BH5015, BH5089, BH5091, BH5055, BH5054, BH0065, BH5006)	11:10-11:55

Note: If you want to do online zoom test after 11:55, please contact your conference secretary.

March 18, 2026 (Wednesday)

Morning Sessions Venue: Global Front, Room: 3rd Floor, 4031		Duration
Opening Remark	Conference Chair Prof. Jun Sakakibara, Meiji University, Japan	9:00-9:05
Keynote Speech 1	Prof. Tae-Kyun Kim, Korea Advanced Institute of Science and Technology, Korea Speech Title: Spatial Physical Generative AI	9:05-9:50
Invited Speech 1	Prof. Kiyoshi Hoshino, Meiji University, Professor Emeritus of University of Tsukuba, Japan Speech Title: Physical AI: From Collaboration to Combat	9:50-10:15
Group Photo & Coffee Break		**10:15-10:35**
Invited Speech 2	Prof. Hiromasa Oku, Gunma University, Japan Speech Title: Dynamic Image Control based on low-latency visual feedback and its applications	10:35-11:00
Invited Speech 3	Prof. Yoosoo Oh, Daegu University, Korea Speech Title: AI for 500 Inclusive Care: Understanding and Monitoring Challenging Behaviors Beyond the Classroom	11:00-11:25
Lunch Time, venue: 17th floor of Liberty Tower		**11:25-13:00**

Parallel Sessions		
13:00-14:35	**Session 1:** Image Detection, Pattern Recognition and Image Security **Session Chair:** Prof. Tohru Kamiya, Kyushu Institute of Technology, Japan **Presentations:** BH0037, BH0049-A, BH0016, BH0001, BH0023, BH0041, BH0019	**3F: 4031**
13:00-14:35	**Session 2:** Multimodal Natural Language Processing and Intelligent Conversation Generation Systems **Session Chair**: Prof. Shin Ando, Tokyo University of Science, Japan **Presentations:** BH5016, BH5080, BH5102, BH6001-A, BH5037, BH5017	**2F: 4021**
13:00-14:35	**Session 3:** Advanced Signal Processing and Multimodal Sensing in Biomedical and Industrial Applications **Session Chair:** Assistant Prof. Rebeka Sultana, Tokyo University of Agriculture and Technology, Japan **Presentations:** BH0012, BH0048, BH0056, BH0014, BH0034, BH0050, BH0057	**3F: 403D**
13:00-14:35	**Session 4:** Software Security and Human-Computer Interaction Design **Session Chair:** Assoc. Prof. Daniel Fernández Lanvin, University of Oviedo, Spain **Presentations:** BH5007, BH5019, BH5028, BH5046, BH5050, BH5090-A, BH5024-A	**3F: 403N**
14:35-14:50	**Coffee Break**	

14:50-16:25	**Session 5:** 3D Image Modeling and Virtual Technology Applications **Session Chair:** Prof. Shyi-chyi Cheng, National Taiwan Ocean University, Taiwan **Presentations:** BH0009-A, BH0015, BH0044, BH0022, BH0054, BH0010-A, BH0024	**3F: 4031**
14:50-16:25	**Session 6:** Artificial Intelligence and Applications in Digital Information Systems **Session Chair:** Assoc. Prof. Fuad Abujarad, Yale University, USA **Presentations:** BH5059, BH5062, BH5081, BH5042, BH5061, BH5063	**2F: 4021**
14:50-16:25	**Session 7:** Intelligent Text and Image Generation Systems Based on Cross-Modal Information Fusion **Session Chair:** Prof. Ran-Zan Wang, Yuan Ze University, Taiwan **Presentations:** BH0018, BH0029, BH0039-A, BH0045, BH0031, BH0052, BH0038	**3F: 403D**
14:50-16:55	**Session 8:** Intelligent Image Analysis and Processing Methods **Session Chair:** Prof. Tetsuya Shimamura, Saitama University, Japan **Presentations:** BH0032, BH0033, BH0040, BH0025, BH0053, BH0027, BH0004-A, BH0042-A, BH0028	**3F: 403N**
16:25-16:30	**Coffee Break**	
16:30-17:50	**Session 9:** Adaptive Machine Learning Models and Intelligent Computing Frameworks for Multimodal Scenarios **Session Chair:** Asst. Prof. Siew Mooi Lim, Tunku Abdul Rahman University of Management and Technology, Malaysia **Presentations:** BH5031, BH5040, BH5047-A, BH5066-A, BH5010-A, BH5051, BH5068-A	**3F: 4031**

16:30-17:50	**Session 10:** Key Technologies in Computer Vision and Image Processing **Session Chair:** Prof. Weichao Wang, University of North Carolina at Charlotte, USA **Presentations:** BH5003, BH5056, BH5086, BH5097, BH5073, BH5001	**2F: 4021**
16:30-17:50	**Session 11**: Human-Computer Interaction Design and Digital Multimedia Applications **Session Chair:** Prof. Chang-Hwan Im, Hanyang University, Republic of Korea **Presentations:** BH5099, BH5101, BH5045, BH5075, BH5027	**3F: 403D**
18:00-20:00	**Dinner Time, venue: 23rd floor of Liberty Tower**	

Note:

(1) One Best Presentation will be selected from each presentation session, and the Certificate for Best Presentation will be awarded at the end of each session by Session Chairs.

(2) Regular each Presentation: about 13 Minutes including 2-3 Minutes for Question and Answer.

March 19, 2026 Online (GMT+9)

Morning Sessions		Duration
Online Speeches and online sessions **Zoom ID: 841 3471 6436** **Zoom link: https://us02web.zoom.us/j/84134716436**		
Keynote Speech 2	Prof. Xudong Jiang, Nanyang Technological University, Singapore (IEEE Fellow) Speech Title: Critical Features of CNN Leads ML to DL and Transformer makes Model Intelligent	10:30-11:15
Session 12	**Topic:** Multimodal Sensing and Understanding for Agricultural and Biomedical Applications **Session Chair:** Assoc. Prof. Md Liakat Ali, Rider University, USA **Presentations:** BH5093, BH5107, BH5105, BH0066, BH0011, BH5033	11:15-12:50
Break Time		**12:50-14:00**
Keynote Speech 3	Prof. Gonzalo Arce, University of Delaware, USA (IEEE Life Fellow / SPIE fellow/ AAIA fellow) Speech Title: Computational Satellite Lidar Imaging	14:00-14:45
Invited Speech 4	Assoc. Prof. Ir. Dr. Kim Seng Chia Universiti Tun Hussein Onn Malaysia, Malaysia Speech Title: Adaptive Artificial Neural Network in near infrared spectroscopy for standard-free calibration transfer	14:45-15:10
Invited Speech 5	Prof. Maxim Bakaev Novosibirsk State Technical University, Russia Speech Title: AI-Based Automation in Design of Graphical User Interfaces: iterative generation and subjective evaluation	15:10-15:35

Session 13	**Topic:** Generative AI and Multimodal Sensing **Session Chair:** **Presentations:** BH5078, BH5014, BH5015, BH5089, BH5091, BH5055, BH5054, BH0065, BH5006	15:35-17:35

Detailed Program

March 18, 2026 (Wednesday)

Opening Remark

9:00-9:05	**Opening Remark**
Venue	**Global Front, Room: 3rd Floor, 4031**

Prof. Jun Sakakibara

Meiji University, Japan

Dr. Jun Sakakibara received his Ph.D. from the Department of Mechanical Engineering at Keio University in 1996. He was a visiting scholar at the University of Illinois at Urbana-Champaign from 1996 to 1997 before joining the Department of Engineering Mechanics at the University of Tsukuba in 1997. He later returned to the University of Illinois for a second visit from 2000 to 2001. In 2013, he joined the Department of Mechanical Engineering at Meiji University. Dr. Sakakibara is an associate editor of the Journal of Visualization and a member of the editorial board of Experiments in Fluids. He previously chaired the 46th Annual Meeting of the Visualization Society of Japan and is scheduled to chair the 16th International Symposium on Particle Image Velocimetry in 2025. He is currently serving as the President of the Visualization Society of Japan. His current research interests include experimental studies of complex turbulent flows, with particular emphasis on flows associated with wall-bounded systems, separated flows, and flows around wings or obstacles. He is also dedicated to the advancement of optical flow measurement techniques, including particle image velocimetry and laser-induced fluorescence.

Keynote Speech 1

Host	Prof. Jun Sakakibara	**Time**	9:05-9:50, March 18, 2026
	Meiji University, Japan	**Venue**	Global Front, Room: 3rd Floor, 4031

Prof. Tae-Kyun Kim

Korea Advanced Institute of Science and Technology, Korea

Tae-Kyun (T-K) Kim is Professor and the director of Computer Vision and Learning Lab at School of Computing, KAIST since 2020, and has been an adjunct reader of Imperial College London (ICL), UK for 2020-2024. He led Computer Vision and Learning Lab at Imperial College during 2010-2020. He obtained his PhD from Univ. of Cambridge in 2008 and Junior Research Fellowship (governing body) of Sidney Sussex College, Univ. of Cambridge during 2007-2010. His BSc and MSc are from KAIST in 1998 and 2000, he worked at Samsung AIT for 2000-2004 (military duty). His research interests primarily lie in machine (deep) learning for 3D computer vision, generative AI and Physics-based AI, including: articulated 3D hand/body reconstruction, face analysis and recognition, 6D object pose estimation, activity recognition, object detection/tracking, active robot vision, which lead to novel active and interactive visual sensing. He has co-authored over 100 academic papers in top-tier conferences and journals in the field, and has co-organised series of HANDS workshops and 6D Object Pose workshops (in conjunction with CVPR/ICCV/ECCV) since 2015 to 2020. He was the general chair of BMVC17 in London, the program co-chair of BMVC23, and is Associate Editor of IEEE Trans on PAMI, Pattern Recognition Journal, Image and Vision Computing Journal. He regularly serves as an Area Chair for top-tier vision/ML conferences. He received KUKA best service robotics paper award at ICRA 2014, and 2016 best paper award by the ASCE Journal of Computing in Civil Engineering, and the best paper finalist at CVPR 2020, and his co-authored algorithm for face image representation is an international standard of MPEG-7 ISO/IEC.

Speech Contents

Speech Title: Spatial Physical Generative AI

Abstract: Recent breakthroughs in large language models (LLMs) and generative AI have demonstrated remarkable capabilities in text, image, and video synthesis. However, despite their scale and fluency, these models remain fundamentally limited in their understanding of 3D space, physical interaction, and embodied reasoning. This keynote explores the next frontier beyond LLMs: Spatial Physical Generative AI — systems that not only generate content but understand and reason about the physical world.

We begin by examining the evolution from Transformers and large-scale pretraining toward vision-language and world models that attempt grounded intelligence. While current generative video and multimodal models achieve impressive visual realism, they often lack true spatial consistency, physical plausibility, and compositional generalization. Addressing these limitations requires integrating 3D scene representations, physics-based simulation, and generative diffusion frameworks.

The talk presents recent advances in 3D spatial AI, motion diffusion, physics-guided video generation, and category-level 6D pose estimation. We introduce methods that combine regression and diffusion modeling with score scaling sampling to capture multi-hypothesis pose distributions efficiently. Furthermore, we highlight MPMAvatar, a hybrid mesh–3D Gaussian Splatting framework that enables physically accurate cloth simulation and photorealistic avatar rendering, demonstrating realistic deformation, collision handling, and zero-shot scene interaction.

By unifying generative models with spatial representation and physical simulation, we move toward AI systems capable of embodied reasoning, real-world interaction, and scalable physical intelligence. Spatial Physical Gen AI represents a critical step toward grounded artificial general intelligence (AGI), bridging perception, generation, and physical understanding.

Invited Speech 1

Host	Prof. Jun Sakakibara	**Time**	9:50-10:15, March 17, 2026
	Meiji University, Japan	**Venue**	Global Front, Room: 3rd Floor, 4031

Prof. Kiyoshi Hoshino

Meiji University, (Professor Emeritus) University of Tsukuba, Japan

He received two doctor's degrees; one in Medical Science in 1993, and the other in Engineering in 1996, from the University of Tokyo respectively. From 1993 to 1995, he was an Assistant Professor at Tokyo Medical and Dental University School of Medicine. From 1995 to 2002, he was an Associate Professor at University of the Ryukyus. From 2002 to 2023, he served at the Biological Cybernetics Lab of University of Tsukuba as an Associate Professor and a Full Professor. He is now a Full Professor at Meiji University. He is awarded Professor Emeritus of University of Tsukuba in 2023. From 1998 to 2001, he was jointly appointed as a senior researcher of the PRESTO "Information and Human Activity" project of the Japan Science and Technology Agency (JST). From 2002 to 2005, he was a project leader of a SORST project of JST. He served as a member of the “cultivation of human resources in the information science field” WG, Special Coordination Funds for the Promotion of Science and Technology, MEXT, a member of “Committee for Comfort 3D Fundamental Technology Promotion”, JEITA, the General Conference Chair of the 43rd Annual Meeting of Japanese Society of Biofeedback Research, and a councilor and director of the Ibaraki Sports Association. He received IJCAI-09 AI Video Award, iFAN 2010 Best Paper Award, Laval Virtual Awards in 2009, 2013 and 2014, ISER 2015 Best Paper Award, and several domestic and international awards.

Speech Contents

Speech Title: Physical AI: From Collaboration to Combat

Abstract: The ideal movement in most martial arts and combat sports is to enter the opponent's interpersonal distance in the shortest time possible and strike him in the shortest time possible. Based on early cues obtained from the opponent's movements, it is necessary to predict and execute optimal defensive and offensive actions. My speech today focuses on close-quarters combat, where the ideal action is to defeat the opponent by covering the shortest distance in the shortest time, using a short bamboo sword, roughly equivalent to a bayonet or combat knife. The author introduces research examples on how to make a robot act as the defender acquires the ability to predict the type of attack motions produced by a human attacker early and accurately. The author also discusses methodologies for measuring an attacker's physiological signals.

Invited Speech 2

Host	Prof. Kiyoshi Hoshino	**Time**	10:35-11:00, March 18, 2026
	Meiji University, Professor Emeritus of University of Tsukuba, Japan	**Venue**	Global Front, Room: 3rd Floor, 4031

Prof. Hiromasa Oku

Gunma University, Japan

Hiromasa Oku received the Dr. (Eng.) degree in mathematical engineering and information physics from The University of Tokyo, Japan, in 2003. He was a Researcher with PRESTO, Japan Science and Technology Agency, from 2003 to 2005. He was a Research Associate (2005–2007), an Assistant Professor (2007–2011), and a Lecturer/an Assistant Professor (2011–2014) with The University of Tokyo. He has been an Associate Professor with the School of Science and Technology, Gunma University, since 2014, where he is currently a Professor with the Faculty of Informatics. He has received numerous awards in robotics, virtual reality, and human-computer interaction, including the Best Paper Award at VRST 2017, the Advanced Robotics Best Paper Award (2016), and multiple honors from JSME, SICE, and METI. His research interests include high-speed image processing, high-speed optical devices, and dynamic image control.

Speech Contents

Speech Title: Dynamic Image Control based on low-latency visual feedback and its applications

Abstract: While image recognition technology has advanced to a practical level with the recent breakthroughs in machine learning, there are still many unexplored areas in image utilization systems. Many optical systems and optical devices used to acquire images do not assume image recognition methods, leaving room for improvement, especially in response time. The author's group have been pointing out that high-speed imaging optics are also required, especially in systems based on high-speed image recognition. In particular, we have proposed a technology called "Dynamic Image Control" that enables dynamic image capture and projection by combining high-speed image processing and high-speed optics. This talk will introduce the basic concepts of Dynamic Image Control and its applications in combination with high-speed gaze controller using rotational mirrors, high-speed variable focus liquid lenses, and edible optical devices. Additionally, application examples of Dynamic Image Control, incorporating recent progress, will be also introduced.

Invited Speech 3

Host	Prof. Kiyoshi Hoshino	**Time**	11:00-11:25, March 18, 2026
	Meiji University, Professor Emeritus of University of Tsukuba, Japan	**Venue**	Global Front, Room: 3rd Floor, 4031

Prof. Yoosoo Oh

Daegu University, Korea

Yoosoo Oh received his B.S. degree in Electronics Engineering from Kyungpook National University, Daegu, Korea, in 2002. He received his M.S. degree in Information and Communications in 2003 and his Ph.D. degree in Information and Mechatronics in 2010 from Gwangju Institute of Science and Technology (GIST), Korea.

From 2005 to 2006, he was a Visiting Researcher at the Embedded Interaction Lab, University of Munich, Germany. From 2010 to 2012, he worked at the Culture Technology Institute (CTI), GIST, as a Technical Team Leader and General Team Leader. He was also a Visiting Scholar at Lancaster University, UK, and a Researcher at KAIST Graduate School of Culture Technology (GSCT) in 2012. In 2019–2020, he served as a Visiting Professor at Georgia Institute of Technology, USA.

Since joining Daegu University in 2012, he has held multiple academic leadership positions. He is currently a Full Professor in the School of Computer Software Engineering, School of Computer & Information Engineering. He previously served as Director of the Basic Education Center under the SW-Centered University Project (2025) and as Dean/Head of the School of Computer & Information Engineering (2025–2026). He has also served as Department Chair, Head Professor, and Director of several major research centers, including the Mixed Reality Convergence Research Center, DU Smart Drone Center, AZIT MakerSpace Center, and the Gyeongbuk Technopark–Daegu University Center. In addition, he led industry–university cooperation initiatives as Director of the Enterprise Supporting Office of the LINC Project Group.

Prof. Oh is the Founder and CEO/CTO of NWI Co., Ltd., an AI-based Edutech company established in 2021. He also serves as an Auditor of the Daegu-Gyeongbuk Metaverse

Industry Association. He has been a Board Member of the Korea Society of Industrial Information Systems (since 2019) and the HCI Korea Society (since 2014), and is a member of the ISMAR 2025 Organizing Committee.

His research focuses on machine learning, deep learning for Edutech, industrial AI applications, intelligent middleware, human–AI interaction, and AI convergence systems. He has received numerous awards and honors, including the Minister of Education Award (2025), selection among the Top 50 Outstanding Research Achievements supported by the National Research Foundation of Korea, the Albert Nelson Marquis Lifetime Achievement Award (2017), a Commendation from the Korea Industry-Academia Cooperation Association (2018), and the Daegu University Excellent Research Award in Academic Publications (2016).

Speech Contents

Speech Title: AI for Inclusive Care: Understanding and Monitoring Challenging Behaviors Beyond the Classroom

Abstract: Effective support for individuals with intellectual and developmental disabilities (IDD) necessitates systematic observation and consistent intervention across both educational and care environments. Traditional manual recording methods often yield fragmented data and subjective interpretations, hindering the development of effective support plans. This presentation introduces a human-centered artificial intelligence (AI) framework that integrates principles from both AI and human–computer interaction. The proposed framework operates within a "24-hour All-in-Care" ecosystem to provide inclusive behavioral support.

We introduce an AI-assisted behavioral analysis pipeline in which computer vision serves as an assistive sensing layer in complex, multi-user environments. Using the DACARE 2.0 system as a case study, we detail the technical architecture, including human detection, multi-object tracking, and behavior recognition models for identifying challenging behaviors. The system ensures continuous observation, even during visual occlusion or environmental changes, by combining face recognition embeddings with clothing-based person re-identification. This enables reliable identity association over time and across settings.

From a human–computer interaction (HCI) perspective, the system functions as both an automated recognition tool and an interactive decision-support interface for educators and caregivers. Behavioral data are rendered interpretable through visual analytics dashboards, which present metrics such as behavioral frequency and duration, as well as intervention phases including baseline, intervention, maintenance, and generalization. These interfaces facilitate sensemaking, reduce the cognitive demands of manual documentation, and promote data-driven collaboration among care stakeholders.

Field deployments in special education schools indicate that AI technologies can augment, rather than replace, professional expertise when integrated into well-designed, human-centered interfaces. By positioning AI as a collaborative partner within a broader sociotechnical care system, this work envisions an interconnected care ecosystem that enhances the understanding of challenging behaviors and extends inclusive support beyond the classroom.

Online Keynote Speakers
Keynote Speech 2 (Online)

Time	**10:30-11:15, March 19, 2026 (GMT+9)**
Zoom ID	**841 3471 6436**
Zoom Link:	**https://us02web.zoom.us/j/84134716436**

Prof. Xudong Jiang (IEEE Fellow)

Nanyang Technological University, Singapore

Dr. Xudong Jiang received the B.Eng. and M.Eng. from the University of Electronic Science and Technology of China (UESTC), and the Ph.D. degree from Helmut Schmidt University, Hamburg, Germany. From 1986 to 1993, he was a Lecturer with UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1998 to 2004, he was with the Institute for Infocomm Research, A-Star, Singapore, as a Lead Scientist and the Head of the Biometrics Laboratory, where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition in 2000. He joined Nanyang Technological University (NTU), Singapore, as a Faculty Member, in 2004, and served as the Director of the Centre for Information Security from 2005 to 2011. Currently, he is a professor in NTU. Dr Jiang holds 7 patents and has authored over 150 papers with over 40 papers in the IEEE journals, including 11 papers in IEEE T-IP and 6 papers in IEEE T-PAMI. Three of his papers have been listed as the top 1% highly cited papers in the academic field of Engineering by Essential Science Indicators. He served as IFS TC Member of the IEEE Signal Processing Society from 2015 to 2017, Associate Editor for IEEE SPL from 2014 to 2018, Associate Editor for IEEE T-IP from 2016 to 2020 and the founding editorial board member for IET Biometrics form 2012 to 2019. Dr Jiang is currently an IEEE Fellow and serves as Senior Area Editor for IEEE T-IP and Editor-in-Chief for IET Biometrics. His current research interests include image processing, pattern recognition, computer vision, machine learning, and biometrics. Additionally, he serves

as a Senior Editorial Board member of APSIPA Transactions on Signal and Information Processing and an Associate Editor of EURASIP International Journal on Image and Video Processing. His current research interests include image and video processing, computer vision, and human face analysis and recognition.

Speech Contents

Speech Title: Critical Features of CNN Leads ML to DL and Transformer makes Model Intelligent

Abstract: The powerfulness of machine learning was already proven more than 30 years ago in the boom of neural networks but its successful application to the real world is just in recent 10 years after the deep convolutional neural networks (CNN) have been developed. This is because the machine learning alone can only solve problems in the training data but the system is designed for the unknown data outside of the training set. This gap can be bridged by regularization: human knowledge guidance or interference to the machine learning. This speech will analyze these concepts and ideas from traditional neural networks such as MLP to the deep convolutional neural networks (CNN) and Transformer. It will answer the questions why the traditional neural networks fail to solve real world problems even after more than 30 years' intensive research and development and how the deep CNN and Transformer solve the problems of the traditional neural networks and now are very successful in solving various real world AI problems.

Keynote Speech 3 (Online)

Time	**14:00-14:45, March 19, 2026 (GMT+9)**
Zoom ID	**841 3471 6436**
Zoom Link:	**https://us02web.zoom.us/j/84134716436**

Prof. Gonzalo Arce (IEEE Life Fellow / SPIE fellow/ AAIA fellow)

University of Delaware, USA

Dr. Gonzalo R. Arce is the Charles Black Evans Professor in the Electrical and Computer Engineering Department at the University of Delaware. He is a JPMorgan-Chase Senior Faculty Fellow with the Institute of Financial Services Analytics at University of Delaware. He held twice the Nokia-Fulbright Distinguished Chair in Information and Communications Technologies at Aalto University in Helsinki, Finland. His research interests lie in computational imaging, signal processing, and machine learning. Dr. Arce is a Fellow of the IEEE, OPTICA, the SPIE and was elected to the National Academy of Inventors. He is Editor-in-Chief of the IEEE Transactions on Computational Imaging.

Speech Contents

Speech Title: Computational Satellite Lidar Imaging

Abstract: Spaceborne lidars play a critical role in observing Earth's urban, forest, and glacial ecosystems. However, existing satellite lidar systems are hindered by low spatial resolution and photon density, limiting their ability to produce detailed 3D surface topography and vegetation imagery. While airborne lidars provide higher resolution, they cannot achieve global coverage. This talk describes compressive satellite lidars (CS-Lidars), a novel approach utilizing coded laser illumination and dynamic wavelength scanning for wide-field 3D imaging. We propose a new framework based on hyperheight data cubes (HHDCs), a unique representation of waveform altimetry profiles that capture comprehensive 3D scene structures. HHDCs enable straightforward extraction of canopy height models (CHMs), digital terrain models (DTMs), and other scene features using simple statistical quantiles. Additionally, generative diffusion models are used for super-resolution imaging by learning conditional probabilities, employing forward imaging guidance, and incorporating high-resolution auxiliary data. These techniques were validated across multiple regions in the United States.

Online Invited Speakers
Invited Speech 4 (Online)

Time	14:45-15:10, March 19, 2026 (GMT+9)
Zoom ID	841 3471 6436
Zoom Link:	https://us02web.zoom.us/j/84134716436

Assoc. Prof. Ir. Dr. Kim Seng Chia

Universiti Tun Hussein Onn Malaysia, Malaysia

Ir. Dr. Kim Seng Chia is an Associate Professor at Universiti Tun Hussein Onn Malaysia (UTHM). He is a Chartered Engineer (CEng, UK) and Professional Engineer (BEM, Malaysia) with expertise in machine learning and embedded system applications. He holds a PhD in Electrical Engineering from Universiti Teknologi Malaysia (UTM), where his research focused on non-destructive fruit quality prediction using near-infrared spectroscopy. Dr. Chia has led multiple research grants on intelligent quality assessment and environmental monitoring, including honey, air, and wastewater. His applied research employs ESP32 and FreeRTOS for real-time data processing and automation, and he delivers HRD Corp claimable industrial training on embedded and IoT systems. He currently serves as Editor of the International Journal of Integrated Engineering (IJIE) and contributes actively to global academic communities as a forum speaker, session chair, technical committee member, technical reviewer, and IET mentor for Chartered Engineer registration.

Speech Contents

Speech Title: Adaptive Artificial Neural Network in near infrared spectroscopy for standard-free calibration transfer

Abstract: Near infrared spectroscopy (NIRS) is a secondary analytical method that could use machine learning to predict components of interest for various applications. Calibration transfer (CT) algorithms have been used to transfer the existing NIRS calibration model to another NIRS instruments to reduce the involvement of expensive and tedious conventional chemical analytical methods. However, most of CT algorithms involve both primary and secondary instruments to acquire NIRS signals from standard samples for spectral standardization. On the other hand, artificial neural network (ANN) that can adapt to new environmental conditions may be an alternative standard-free CT algorithm in transferring a trained ANN to secondary instruments without the involvement of the primary instrument. Although ANN has been widely implemented in NIRS research as a calibration model, its feasibility as an alternative CT method has not been evaluated. Thus, this study evaluates the feasibility of an adaptive ANN (AANN) in transferring ANN model from primary to secondary instruments using two traceable NIRS datasets. First, ANN and ANN with principal components (PCs-ANN) were developed and optimized using Bayesian learning algorithm. After that, these ANNs were adapted to secondary instruments using respective transfer data, in which, only the weights and biases of the ANNs were updated. Findings show that the lowest averaged RMSEP was achieved by the proposed PCs-AANN and AANN in the corn and wheat datasets, respectively. Particularly, the computation cost of having a calibration model in secondary instruments has been substantially reduced by means of the proposed AANN algorithm.

Invited Speech 5 (Online)

Time	**15:10-15:35, March 19, 2026 (GMT+9)**
Zoom ID	**841 3471 6436**
Zoom Link:	**https://us02web.zoom.us/j/84134716436**

Prof. Maxim Bakaev

Novosibirsk State Technical University, Russia

Maxim Bakaev got his PhD degree in Software Engineering in 2012. He currently works as Associate Professor of the Automated Control Systems department of Novosibirsk State Technical University (NSTU), Russia. He is also the Acting Head of the Data Collection and Processing Systems department. Previously, he received his Master Degree in Digital Design from Kyungsung University, South Korea. His research interests include Human-Computer Interaction, Universal Design, Web User Interfaces, User Behavior Models, Knowledge Engineering, Machine Learning, etc. (https://www.researchgate.net/profile/Maxim-Bakaev) His recent research results are related to perception of visual complexity in graphical user interfaces (UIs) and its relation to Gestalt principles and compression algorithms. So, he has proposed the Index of Difficulty for tasks that involve visual-spatial working memory. He oversees the development of the Web UI Measurement Platform (http://va.wuikb.info/) that integrates online services for collecting ML data for UI assessment. He has served as a committee member for several international conferences, particularly as PC Co-Chair for ICMSC 2018 and ICWE 2019, as Demo & Posters Chair for ICWE 2020, and as Workshops Co-Chair for ICWE 2021. He also served as a reviewer for several international conferences and journals, including CHI, UIST, International Journal of Human-Computer Studies, Applied Ontology, Symmetry, etc. He is the Guest Editor for "Complexity in Human-Computer Interfaces: Information-Theoretic Approaches and Beyond", a Special Issue in Mathematics journal. He is also a Section Editor for the Journal of Web Engineering. He has acted as PI or participant in several research grants, domestic and international. In 2016, he received Novosibirsk City Hall award in science and innovations as a "Best young researcher in higher education institutions". Under his supervision, more than 20 Master and Bachelor students graduated.

Speech Contents

Speech Title: Evoking Personas and Specialists from Large Language Models: going beyond the optimistic common-sense assistants

Abstract: Output of Large Language Models (LLM) based services tends to be common-sense and homogeneous, unless prompted otherwise. This trait has profound implications both for human users of the content created with the assistance of generative AI models in for the AI/ML field itself. As for the former, I expect the users to develop "generative blindness" (similar to the "banner blindness" at the onset of WWW) to filter out generic and low information density content. On the other hand, the default low diversity of synthetic texts hinders data augmentation for LLM, who are about to run out of the "natural" training data.

In my talk, I overview recent publications and techniques related to contextual prompting aimed at increasing the diversity and information density of LLM-produced texts. For instance, it has been demonstrated that prompting the models into being more seasoned specialists (e.g., senior software engineers) did increase the quality of the output with respect to solving the tasks in the professional domain. On the other hand, there are large-scale projects to create repositories of synthetic "authors" (e.g., with a whole a billion of them by Tencent AI), but relatively few evaluations of the effectiveness of different techniques in increasing the diversity of their "writings". Correspondingly, I report the results of our own research in evoking "authors" with different associative thesauri from LLMs through contextual prompts based on Cooper's personas descriptions, well-established in HCI. The conclusion is that so far the models fail to achieve the same level of diversity in the associations as the real humans.

Session 1 (13:00-14:35)

Topic: Image Detection, Pattern Recognition and Image Security
Session Chair: Prof. Tohru Kamiya, Kyushu Institute of Technology, Japan
Time: 13:00-14:35, March 18, 2026
Venue: Global Front, Room: 3rd Floor, 4031

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH0037

Image steganography using XOR transformation and Huffman coding

Lee Shu-Teng Chen[1] and **Yong-Wei Chen**[2]
1: Yuan Ze University, Taiwan
2: National Taiwan University, Taiwan

Abstract- In image steganography, achieving high stego-image quality while maintaining high embedding capacity remains a challenge. This paper proposes an image steganography method to embed a secret image in a cover image based on a row-wise exclusive-OR (XOR) transformation and Huffman coding. The secret image is first transformed using the row-wise XOR operation to generate a transformed image. The transformed image is divided into n-bit strings ($2 \leq n \leq 8$), converted into an integer sequence, and then compressed using Huffman coding to produce a variable-length bitstream. The resulting Huffman bitstream, together with the corresponding Huffman table, is embedded in the cover image using the least significant bit (LSB) substitution scheme to generate a stego-image. The embedded secret image can be losslessly extracted from the stego-image. Experimental results show that the proposed method achieves higher embedding capacity and peak signal-to-noise ratio (PSNR) improvements of 0.3–1.1 dB compared with an existing Huffman-based scheme, while maintaining structural similarity index (SSIM) values above 0.9994.

BH0049-A

A Deep Learning–Based Robust Image Watermarking System for Source Verification

Kai-Yi Chen, Huei-Ci Pan, Min-Han Weng, Zong-Min Zheng, Chen-Yu Lin, Po-Han Wu, Yi-Hui Chen and Chien-Lung Hsu

Chang Gung University, Taiwan

Abstract- This project presents a deep learning–based robust image watermarking system for source verification and content authenticity assurance. In response to the increasing challenges posed by large-scale image redistribution and the rapid development of generative AI, this study focuses on the design and evaluation of an end-to-end neural watermarking framework that preserves provenance information under common image manipulations.

The proposed method employs an encoder–decoder neural network architecture to embed an invisible and self-recoverable watermark directly into digital images. The watermarking model is trained to maintain stable decoding performance under a wide range of real-world distortions, including lossy compression, brightness variation, noise injection, and repeated transmission. By binding source-related information directly to image content, the system enables reliable provenance tracing without reliance on external metadata.

To enhance integrity and non-repudiation, the system integrates blockchain technology to store and verify watermark-related records. The decentralized ledger ensures that watermark identifiers and verification data cannot be altered or removed, providing a tamper-resistant reference for authenticity validation. The combination of AI-based invisible watermarking and decentralized verification establishes a trustworthy image authentication pipeline.

Experimental results demonstrate that the proposed watermarking model preserves high visual quality, achieving peak signal-to-noise ratio (PSNR) values between 37 and 40 dB, while maintaining a low bit error rate (BER) of 0.03–0.04 under multiple attack scenarios. These results indicate that the proposed approach effectively balances imperceptibility, robustness, and practical deployability.

Overall, this study shows that robust deep watermarking provides a viable and scalable solution for image provenance verification. The proposed system contributes a practical foundation for improving visual content trustworthiness and offers meaningful value for media professionals, enterprises, and AI content governance applications.

BH0016

Reversible Data Hiding in Encrypted Images Based on Shifting Encryption and Huffman Coding

Cheng-Hsing Yang[1], **Hao-Yu Weng**[2], Bo-Han Huang[3], Wei-Jen Wang[2] and Shiuh-Jeng Wang[4]

1: National Pingtung University, Taiwan

2: National Central University, Taiwan

3: National Changhua University of Education, Taiwan

4: Central Police University, Taiwan

Abstract- This paper presents a new reversible data hiding in encrypted images (RDHEI) scheme that performs data embedding in the compressed encrypted domain. The original image is first encrypted using a uniform shift-based block encryption, which preserves intra–block pixel dependency, while block scrambling is further applied to enhance security. Prediction is then conducted on encrypted pixels to generate prediction errors, which are Huffman–encoded to form compressed bitstreams. This compression step releases additional room for payload embedding, and the auxiliary information required for full image recovery is also stored within the encrypted image. Experimental results demonstrate that the proposed method achieves a higher embedding capacity compared with existing state-of-the-art RDHEI approaches, while preserving perfect reversibility.

BH0001

Plastic Bottle Recognition Using an Improved DeepLabv3+

Daiki Ideta and **Tohru Kamiya**

Kyushu Institute of Technology, Japan

Abstract- In recent years, Japan has experienced a worker shortage due to its declining birthrate and aging population, which has become a social problem. The manufacturing industry is particularly vulnerable to labor shortages because of the shrinking working-age population, the shortage of educators, and the industry's negative image. One solution to this problem is factory automation. Factory automation involves using machines, robots, and software to manufacture and assemble products without human intervention. Using AI (artificial intelligence) to perform simple tasks that were previously done by humans has many advantages in addition to reducing labor shortages. This paper focuses on the automatic sorting of plastic bottles at a waste disposal plant. As a preliminary study, this paper recognizes an image containing only one plastic bottle. We propose an image processing method based on semantic segmentation by applying a deep learning model. Specifically, we use DeepLabv3+ as the CNN (convolutional neural network) base model to perform recognition through semantic segmentation. To improve the recognition rate, we constructed an improved DeepLabv3+ model by introducing a Spatial and Channel Squeeze and Channel Excitation (scSE) block, a skip connection, and a DropBlock into the base model and conducted recognition experiments.

BH0023

Real-Time Fraudulent License Plate Detection through High-Angle and Front-Area Vehicle Images

Yu-Fang Cheng and Jen-Chun Chang

National Taipei University, Taiwan

Abstract- Fraudulent license plate usage poses increasing challenges to traffic monitoring and automated enforcement. Conventional license plate recognition systems rely on front-view imagery or OCR-based pipelines, which often become unstable under high-angle surveillance such as ETC gantries, where plates appear distorted and background interference is significant. This paper presents an improved new system, a detection and automatic reporting framework designed specifically for overhead monitoring scenarios. A front-area vehicle detection model, trained on a self-collected high-angle dataset, is introduced to reduce background noise and improve the stability of downstream recognition tasks. YOLOv8-based vehicle, plate, and character detection modules achieve high precision, recall, and mAP across diverse real-world conditions. To enhance vehicle identity verification beyond plate-text matching, a SIFT-based appearance comparison module is integrated, showing clearly distinguishable similarity differences between same-vehicle and different-vehicle comparisons. Latency measurements indicate an average processing time of 259 ms per image on an RTX 4070 GPU, confirming near real-time capability. Experimental results demonstrate that the improved system provides a robust, accurate, and practical solution for automated detection of fraudulent plate usage in intelligent transportation and roadside enforcement applications.

BH0041

Methodological Workflow for Creating Photorealistic Neoclassical Architecture Datasets

Neil Patrick Del Gallego, Gene Madison Franco, Lance Giljamin Vida and Jonathan Michael Gonzalez
De La Salle University, Philippines

Abstract- Computer vision models for architectural recognition and digital heritage require high-quality, labelled, real-world data, which can be obtained through various 3D reconstruction techniques. Alternatively, synthetic digital versions can be artist-driven, utilizing computer graphics and 3D modeling techniques. In this paper, we propose a methodological workflow for creating the ArcherVerse dataset, tailored for modelling neoclassical architecture. We specifically focus on re-creating De La Salle University (DLSU), which is a university that contains neoclassical structures. Our proposed approach integrates high-poly manual modeling, physically based rendering (PBR), block-out, and scene compositing to reconstruct the ornate facades and intricate motifs characteristic of the neoclassical style. A user evaluation involving 16 participants possessing expertise in 3D rendering yielded mean photorealism and artistic quality scores exceeding 4.0 out of 5.0. The results demonstrate that artist-driven manual modelling can provide “clean” geometric ground truth data as an alternative to traditional 3D scanning

methods for reconstructing architectural buildings. The ArcherVerse dataset can serve as a benchmark for sim-to-real research in CV applied to neoclassical architecture scene understanding.

BH0019

A Novel Reversible Data Hiding in Encrypted Technique for Medical Image Using Image Interpolation and a Two-Stage Difference Preservation Encryption

Cheng-Ta Huang[1], **Zhi-Qi Wang[1]**, Hao-Yu Weng[2], Chih-Ying Wu[3] and Shiuh-Jeng Wang[4]
1: Yuan Ze University, Taiwan.
2: National Central University, Taiwan
3: Radiological Diagnosis, Far Eastern Memorial Hospital, Taiwan
4: Central Police University, Taiwan

Abstract- With the rapid development of information technology and network transmission, the generation, storage, and sharing of images are widely applied in daily life. Reversible Data Hiding in Encrypted Images (RDHEI) is a key technology addressing security and privacy challenges. However, existing Vacating Room After Encryption (VRAE) methods for achieving high capacity often rely on MSB prediction, a technique whose performance is inherently tied to image texture. To overcome this, we propose a novel VRAE framework based on interpolation-based difference, creating a stable embedding space via a specialized encryption design. Our methodology uses a unique two-stage encryption process, applying Difference-Preserving Encryption (DPE) to downsampled reference pixels, followed by applying a pseudorandom constant mask (PRF Mask) to non-reference pixels. The core innovation is that the same constant mask is applied to both the original and interpolated images, precisely preserving the difference structure. This enables a "capacity-adaptive" embedding strategy that yields bits from zero-difference blocks to secret data, maximizing efficiency. Experimental results confirm that the proposed method achieves a significantly higher embedding capacity on multiple test images compared to existing state-of-the-art schemes. Security analysis confirms the method's robustness against statistical attacks, while complete reversibility is guaranteed.

Session 2 (13:00-14:35)

Topic: Multimodal Natural Language Processing and Intelligent Conversation Generation Systems

Session Chair: Prof. Shin Ando, Tokyo University of Science, Japan

Time: 13:00-14:35, March 18, 2026

Venue: Global Front, Room: 2rd Floor, 4021

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH5016

Real-Time Speech Emotion Recognition Using Wav2Vec2 and Bidirectional GRU for Multi-Dataset Emotion Classification

Chen Zhun Lee[1], **Siew Mooi Lim[1]**, Kevin Lee[1] and Kuan Yew Leong[2]

1: Tunku Abdul Rahman University of Management and Technology, Malaysia

2: AI System Research Co. Ltd., Japan

Abstract——Conversational agents often fail to detect affective cues in speech, limiting empathetic interaction. We present a real-time, end-to-end speech emotion recognition (SER) system that infers emotions across eight categories: neutral, calm, happy, sad, angry, fear, disgust, and surprise. Raw 16 kHz audio is denoised, normalized, and processed with Wav2Vec2 to extract contextual speech embeddings. Emotion classification is performed using a Bidirectional GRU (BiGRU), trained and evaluated on three publicly available datasets: TESS, RAVDESS, and SAVEE. Comparative evaluation shows that the BiGRU achieves superior performance over baseline models, with an accuracy of 76.35%, macro F1 of 75.75%, and weighted F1 of 76.23%, outperforming 1D CNN, LSTM-Attention, and CNN-LSTM architectures. The BiGRU demonstrates particularly strong recognition of high-arousal emotions such as angry, surprised, and disgusted, while confusions remain among acoustically similar classes such as happy, sad, and calm. These results highlight the effectiveness of BiGRU-based SER for real-time applications and its potential utility in empathetic conversational agents. Limitations such as class imbalance and subtle-emotion confusions are discussed, with future work aimed at larger, multimodal corpora and advanced neural architectures for further improving emotion recognition accuracy.

BH5080

AEMF: Text-Guided Cross-Attention and Multi-Loss Fusion for Noise-Robust Text-Audio Sentiment Analysis

Sizhe Zhao, Xin Kang, Kazuyuki Matsumoto and Minoru Yoshida
Tokushima University, Japan

Abstract—Multimodal Sentiment Analysis (MSA) aims to infer sentiment by combining different data streams, such as text and audio. Recent progress has been driven by powerful pre-trained models and multi-loss training strategies. A persistent challenge, however, is managing the inherent differences and frequent noise in non-verbal modalities like audio, which can disrupt learning if over-relied upon. This work introduces the Attention-Enhanced Multi-Loss Fusion (AEMF) framework. AEMF employs a textguided cross-attention mechanism to highlight the most relevant acoustic frames, coupled with a hierarchical late fusion strategy to combine utterance-level information. Evaluations on the CMUMOSI and CMU-MOSEI benchmarks show that AEMF delivers consistent, though incremental, gains over a strong late-fusion baseline. For MOSI, the improvement in binary accuracy (Acc- 2) is statistically significant at the item level, accompanied by a slight reduction in Mean Absolute Error (MAE) compared to the late-fusion baseline. Furthermore, experiments with artificially added audio noise demonstrate the model's robustness. These findings support the idea that directing cross-modal interaction with text is an effective and efficient approach for building resilience against acoustic noise in real-world applications.

BH5102

SemiFedER: Semi-Supervised Federated Averaging for Multimodal Emotion Recognition

Nhut Minh Nguyen, Thu Thuy Le, Thanh Trung Nguyen and Duc Ngoc Minh Dang
FPT University, Vietnam

Abstract—Multimodal Emotion Recognition (MER) is a powerful approach for human–computer interaction, leveraging complementary cues across multiple modalities to infer human affect. MER has broad real-world potential in applications such as healthcare monitoring and affect-aware virtual assistants. However, practical deployments face two major challenges: large-scale data is often unlabeled due to the high cost of emotion annotation, and speech signals and transcripts are privacy-sensitive, limiting centralized data collection and training. To address these limitations, we propose SemiFedER, a semisupervised federated learning framework for MER. SemiFedER performs client-side training in two stages, including supervised pre-training on labeled samples and semi-supervised learning that exploits unlabeled data via confidence-based pseudo-labeling and

weak-strong consistency regularization. The server aggregates client updates using Federated Averaging (FedAvg) to learn a global model without sharing raw data. We deploy the representative centralized MER backbones within SemiFedER to assess their effectiveness in this practical setting. Extensive experiments on the MELD dataset under speaker-disjoint non-IID federated splits demonstrate that SemiFedER provides stable performance across labeled ratios and client counts, and achieves competitive improvements in class-balanced evaluation compared to centralized baselines.

BH6001-A

Dimensionality Reduction of the Document–Term Matrix via SVD-Based Feature Orthogonalization

Ben Kim

Seattle University, USA

Abstract-In this paper, we present a method for dimensionality reduction through feature orthogonalization. This technique transforms correlated features into an orthogonal set within the same feature subspace, thereby eliminating multicollinearity and redundancy in large datasets. Unlike Principal Component Analysis (PCA), which projects data onto new principal components formed by linear combinations of the original variables, feature orthogonalization transforms correlated features into orthogonal ones within the same feature subspace, thereby preserving the span (information content) of the original features. Moreover, this approach does not require data centering.

While full orthogonalization retains all features, we select only the top kkorthogonal features associated with the largest singular values to enhance computational efficiency. Using Singular Value Decomposition (SVD), the orthogonalized feature matrix is obtained as $A_o=U_k\Sigma_k$, $A_o\in\mathbb{R}^{m\times k}$.

We implement this procedure in two ways: (1) by manually decomposing the matrix using Singular Value Decomposition (SVD), and (2) by utilizing the *TruncatedSVD* function from *sklearn.decomposition*. The optimal value of $kk$is selected visually using the elbow method. Model evaluation is conducted on a document–term matrix derived from a vectorized wine review dataset, employing cross-validation with Random Forest and Gradient Boosting Regression models.

BH5037

Significant Other AI: Toward Long-Term Relational Intelligence with Identity, Memory, and Emotional Regulation

Sung Park
Taejae University, Republic of Korea

Abstract—Significant Others (SOs) provide identity stability, emotional regulation, and meaning-making. However, many individuals today lack such long-term relational anchors. While empathic AI has advanced in emotional responsiveness, it lacks autobiographical memory, identity understanding, narrative coherence, and proactive emotional support. This paper introduces Significant Other Artificial Intelligence (SO-AI), outlining conceptual requirements and a multilayer architecture that combines identity modeling, long-term memory, narrative reasoning, and ethical boundaries. It formalizes SO-AI as a distinct class of relational intelligence, differentiating it from existing empathic or companion systems, and specifies how multi-timescale memory, identity-aware modeling, and narrative co-construction can be integrated into a coherent framework. The paper also proposes a research agenda that examines SO-AI's psychological, relational, and sociocultural implications, including risks of dependency and autonomy erosion. Overall, SO-AI explores whether AI can responsibly supplement missing relational functions in modern society without displacing human intimacy.

BH5017

A Framework for Contextual Prompt Generation and Predictive Text Modeling for Enhanced Scriptwriting Assistance

Wei Jian Gan[1], **Siew Mooi Lim[1]**, Kerr Chii Ooi[1] and Kuan Yew Leong[2]
1: Tunku Abdul Rahman University of Management and Technology, Malaysia
2: AI System Research Co. Ltd., Japan

Abstract—A modular framework that couples contextual prompt generation with predictive text modeling is presented to support screenplay authoring. From 11 movie scripts, 1,850 scenes were segmented, and five scene-aware attributes including Scene, Time, Characters, Action, Emotions, were automatically derived to produce 1,157 structured prompt–completion pairs (JSON/JSONL). These pairs were used to train and evaluate two complementary strategies: (A) domain adaptation via fine-tuning GPT-2 Medium and (B) zero-/few-shot prompting of larger locally served LLMs (e.g., Llama-2, Gemma) through Ollama. A lightweight, model-agnostic GUI was implemented to enable interactive comparison and real-time generation. Quantitatively, perplexity was reported for the fine-tuned model and ROUGE for prompted LLMs; qualitatively, human

inspection indicated that injecting structured scene attributes yielded continuations with improved narrative coherence, creative detail, and contextual relevance over unstructured baselines. In practice, reliable extraction and responsive inference were observed for Gemma in a local setup. Overall, structuring scene context into compact prompts is demonstrated as an effective, reusable bridge between raw screenplays and predictive engines, offering a practical path for creative tools in film and related narrative domains. Methodology and dataset construction details are provided to facilitate extension to larger corpora and alternative backbones.

Session 3 (13:00-14:35)

Topic: Advanced Signal Processing and Multimodal Sensing in Biomedical and Industrial Applications

Session Chair: Assistant Prof. Rebeka Sultana, Tokyo University of Agriculture and Technology, Japan

Time: 13:00-14:35, March 18, 2026

Venue: Global Front, Room: 3rd Floor, 403D

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH0012

Transformer-Augmented EfficientNetV2B3 for Robust Plant Disease Identification

Mohammad Farukh Hashmi, Pinagadi Meghana and Thota Bhagath

National Institute of Technology Warangal, India

Abstract- Plant-leaf diseases cause a significant damage to the agriculture yield if they are not diagnosed and treated early. It is also crucial in preserving global food security and promoting sustainable farming practices. These diseases can be detected through manual inspection but it is laborious to do by hand and the outcome is entirely dependent on the examiner. It has been observed that manual assessment prone to errors, particularly when there are irregularities in the illumination, abnormalities in the leaves, and small variations in disease symptoms. So, there is a need for a model that can successfully classify data by extracting features using computer vision and deep learning. This paper introduces a hybrid image classification deep-learning model using Convolutional Neural Network EfficientNetV2B3 combined with Transformer block made of Multi-head Attention and Multilayer Perceptron (Feedforward layers). EfficientNetV2B3 known for its scaling efficiency is used as a backbone for initial feature extraction, while the multi-head attention lets the model to learn relationships between distant regions by focusing on multiple areas of the image and the feedforward layers help model to figure complex features and then classified through a softmax output layer. The study tells that this model with less parameters, speed and high accuracy than existing image classification models like Resnet, VGG, Inception etc. generalizes better and performs well in detecting plant diseases with a validation accuracy of 99.70%.

BH0048

Style-Aware Data Augmentation for Deep Learning on Symbolic Music

Yung-Chi Tseng, Yu-Chia Wang, **Yung-An Chen**, Chin-Yun Yang and Yu-Cheng Lin
Yuan Ze University, Taiwan

Abstract- This study proposes a style-aware data augmentation framework that combines rule-based design with statistical constraints, applied to small-scale, highly constrained Jiangnan symbolic music generation tasks. By leveraging YNote representation, fixed rhythmic frameworks, and Markov-style local transition statistics, we systematically expand the training data while maintaining musical structural plausibility. Using the augmented data, we fine-tune a GPT-2 model to analyze how different training data scales affect generation behavior. Experimental results show that BLEU-based reference-overlap metrics exhibit only minor fluctuations across different training scales and are insufficient to directly reflect style improvement. In contrast, KL divergence and bigram statistics effectively characterize the style consistency of the generated set in terms of overall distribution proximity and local transition plausibility. Further analysis indicates that a medium-scale training set (approximately 3,000–12,000 samples) achieves the best balance between distribution consistency and transition coverage, whereas excessive augmentation may lead to distribution calibration drift. Overall, the study demonstrates that data augmentation has a positive but non-monotonic effect on style consistency, highlighting the need for carefully designed augmentation strategies in highly constrained symbolic music generation tasks.

BH0056

Non-Contact Heart Rate Estimation Using Automotive Millimeter-Wave Radar with Range-Azimuth Heatmap and Temporal Filtering

Chi-Hung Wang, **Xiang-Shun Yang** and Yu-Hsiang Hsiang
Feng Chia University, Taiwan

Abstract- In the post-COVID-19 era, the demand for continuous monitoring of vital signs, such as heart rate, has increased in clinical settings. Although electrocardiography (ECG) provides accurate heart rate measurements, the need for electrodes attached to the body surface introduces risks of infection. Recent studies have therefore explored non-contact methods for estimating heart rate. However, optical and thermal imaging techniques are sensitive to lighting conditions and occlusion. In contrast, millimeter-wave (mmWave) radar enables the measurement of subtle chest wall displacements without external illumination, making it suitable for long-term monitoring. Nevertheless, radar echo signals are still affected by respiration, body motion, and environmental reflections, which necessitate additional signal processing. To address these issues, this study proposes a heart rate measurement framework based on automotive-grade high-resolution mmWave radar. The framework integrates spatial-level Range-Azimuth heatmaps with temporal background suppression to extract stable one-dimensional human echo signals. The effects of three time-domain filtering methods, namely Amplifier (AMP)、Band-Pass (BP), and Median, on heart rate estimation performance are then evaluated. Synchronously measured ECG signals are used as the reference standard. Experimental results show that, under identical spatial localization conditions, BP filtering achieves the lowest estimation error (MAE = 1.34 bpm, MSE = 3.02 bpm). Compared with results obtained without this approach, the error distribution becomes more concentrated, with MAE reduced from 12.21 bpm to 5.18 bpm. No proportional bias is observed within the heart rate range of 60-120 bpm. These findings provide a practical reference for signal processing design in clinical non-contact heart rate monitoring using mmWave radar.

BH0014

CNN-Based Acoustic Spectrogram Analysis for Quality Assessment of Laser-Welded Screws

Ilinca-Laura Burdulea, Xiaoxu Zhang, Jörg Brünnhäußer and Jascha Seleenmeyer
Fraunhofer IPK, Germany

Abstract- Laser welding offers high precision and efficiency but remains challenging to inspect for internal defects using conventional Non-Destructive Testing (NDT) methods. This study presents a contact-free, post-process inspection framework combining laser-

induced acoustic excitation with Deep Learning (DL)-based spectrogram analysis for automated weld quality assessment. Acoustic responses from welded screw elements were recorded using a membrane-free optical microphone and converted into time-frequency spectrograms. A Convolutional Neural Network (CNN) was trained to classify weld integrity based on these spectrograms, outperforming classical Machine Learning (ML) baselines that relied on manually engineered features. The optimized CNN achieved 93.5% accuracy at the segment level and 92.9% at the specimen level, demonstrating reliable discrimination between defect-free and defective welds. Hyperparameter optimization revealed that spectrogram resolution and network depth were the primary performance drivers. The proposed method establishes a foundation for scalable, AI-driven, and non-contact Quality Assurance (QA) in laser-based manufacturing. Future work will extend the system toward multi-class defect characterization and real-time deployment.

BH0034

Shallow vs. Deep Features for Anomalous/Relevant Event Detection in High-Dimensional Streaming Imagery

Rohan C. Loveland, **Nathaniel A. Clark** and **Christopher M. Ritter**
New College of Florida, USA

Abstract- We compare the performance of the Anomalous/Relevant Event Detection (A/RED) algorithm on high-dimensional streaming imagery comparing shallow and deep features. A multi-class highly imbalanced dataset based on imagery of a parking lot is used with three different feature sets, one shallow and two deep, for inputs. The shallow dataset is the direct raw data, consisting of ≈ 4, 400 128x128 tiles, leading to 16,384 features. The deep feature sets are based on pre-processing the raw data with: i) Deep Autoencoder Gaussian Mixture Models (DAGMM), and ii) vision transformer DINOv2. Performance is compared using anomaly detection system appropriate equivalents to precision and recall over a number of scenarios with different values of κ, the “paranoia parameter”, and multiple relevant class deffnitions comprising 1% and 5% of the total data. The results indicate that DINOv2 has the best performance, followed by the shallow raw input features, and then DAGMM. The performance of the shallow features are sufffciently good to indicate that A/RED can still work well with high-dimensional features.

BH0050

Automatic Structural Tone Extraction from Monophonic Music

Yun-Chi Cheng, Ssu-Yu Yeh, Tung-En Chang, **Yung-An Chen**, Yu-Cheng Lin
Yuan Ze University, Taiwan

Abstract- This study proposes an AI-driven method for the automatic extraction of structural tones from monophonic music signals, in which the task is formulated as a weighted multi-objective optimization problem. The proposed framework jointly considers melodic similarity preservation, note selection sparsity, and rhythmic salience with respect to metrical strong beats, to balance perceptual fidelity and structural compactness in melodic representation. To address the resulting combinatorial optimization problem, Simulated Annealing and Genetic Algorithm are respectively employed as metaheuristic search strategies, and their optimization behaviors and performance characteristics are systematically compared. Experimental results demonstrate that the proposed method can significantly reduce redundant notes while effectively preserving key melodic contours and rhythmic features of the original melody, thereby confirming its feasibility and stability for structural tone extraction. The proposed approach shows strong potential for applications in music signal analysis, multimedia content indexing, music summarization, and AI-assisted music understanding.

BH0057

Synchronized RGB and NIR Imaging for Heart-Rate Estimation on the Thenar Eminence

Chi-Hung Wang, **Hong-Chen Chen**, Zhi-Jia Huang, Xiang-Shun Yang
Feng Chia University, Taiwan

Abstract- The demand for long-term, low-burden HR monitoring in both clinical and home settings is increasing, while electrode-based ECG remains limited by issues of comfort and contact stability. Image-based rPPG enables non-contact HR estimation from periodic changes in skin-reflected light. Still, it relies strongly on ROI geometric stability and is susceptible to illumination variation and camera auto-exposure. This study establishes a hand rPPG HR estimation pipeline using synchronized RGB and NIR imaging from AMT-PVS-50. Under consistent windowing and time-alignment conditions, window-level HR estimates are evaluated against a synchronized ECG reference. The experiments sequentially compare three candidate ROI, namely the thenar eminence, fingertip, and palm, and confirm that the thenar eminence provides superior stability and is therefore selected as the baseline ROI. We further compare three-channel and four-channel ICA, as well as spectral estimation using FFT and Welch PSD. The results show that ICA (RGB+NIR) combined with band-pass filtering and Welch PSD achieves the lowest ECG-referenced error, with MAE = 5.226 bpm, RMSE = 7.432 bpm, and dHR_{p95} = 10.225 bpm. This configuration is adopted as the baseline for the proposed pipeline.

Session 4 (13:00-14:35)

Topic: Software Security and Human-Computer Interaction Design

Session Chair: Associate Prof. Daniel Fernández Lanvin, University of Oviedo, Spain

Time: 13:00-14:35, March 18, 2026

Venue: Global Front, Room: 3rd Floor, 403N

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH5007

Integrated Sense of Agency: User Profiling for Safe Decision-Making in Human-Machine Interactions

Candelaria Gomez de la Calzada[1,2], Juan M. Nadales[3], Luis F. García[1] and Yumie Ono[2]

1: Meiji University, Japan

2: Autonomous University of Madrid, Spain

3: Institute of Science Tokyo, Japan

Abstract—The increasing integration of artificial agents into human decision-making processes raises important questions about autonomy and responsibility. Although many studies have examined the outcomes of human–machine collaboration, less attention has been paid to how individuals experience agency when interacting with artificial systems. In this perspective, we introduce the concept of the Integrated Sense of Agency (ISoA), which captures not only the feeling of ownership over actions but also the alignment of those actions with personal intentions and values. Within this framework, we define illustrative user profiles characterized by different configurations of sense of agency and voluntary endorsement in decision-making processes involving artificial systems. These profiles serve as conceptual tools for exploring potential patterns of human experience and identifying psychological vulnerabilities in human–machine interaction. Building on these insights, we conclude by proposing a series of design measures aimed at preserving the ISoA and promoting conscious, autonomous, and psychologically safe collaboration between humans and artificial agents.

BH5019

Performance Evaluation of Cloud–Edge IoT Architectures in the Philippine Context

Janice A. Abellana
MAPUA University, Philippines

Abstract—Recent advancements in the Internet of Things (IoT) have heightened the demand for real-time data processing, revealing the limitations of traditional cloud-centric architectures in latency, bandwidth, and scalability. This study presents a quantitative performance evaluation of cloud–edge collaborative computing architectures designed to support real-time IoT workloads. A Python-based simulation model was developed to compare three configurations: cloud-only, edge-only, and hybrid cloud–edge setups. Performance was analyzed using metrics such as latency, throughput, and energy consumption. Results indicate that edge computing offers superior latency and energy efficiency compared to cloud-only systems, while the cloud provides greater scalability and centralized resource control. The hybrid cloud–edge architecture achieves a balanced trade-off, improving response times and optimizing resource utilization. Statistical analysis using ANOVA confirmed significant performance differences ($p < 0.001$) across all configurations, demonstrating that architectural choices directly influence system efficiency.

BH5028

A Survey of NLP-Based Software Defect Prediction Methods

Wenkai Sun[1], **Dongcheng Li**[2] and Man Zhao[1]
1: China Univ. of Geosciences (Wuhan), China
2: Cal Poly, USA

Abstract—Software defect prediction (SDP) is a key research topic in software engineering. It seeks to identify potential faults during software development through data analytics and machine learning, thereby improving software quality and reliability. As software systems grow in complexity, natural language processing (NLP) has provided new perspectives and tools for SDP, covering data preprocessing, feature extraction, and model training. This paper presents a systematic survey that synthesizes NLP-based software defect prediction techniques in SDP based on NLP and their model variants. We analyze the strengths and limitations of representative models, the use of datasets and preprocessing pipelines, and differences among evaluation metrics. The survey aims to help researchers understand the state of the art in applying NLP techniques to SDP.

BH5046

An Eye-Gaze Based Human–Machine Interaction Framework for Assistive Wheelchair Navigation

Aivin Thankachan K, Hanock Issac, **Parthiv Keloth** and Stanley Raphel
Albertian Institute of Science and Technology, India

Abstract— Individuals with severe motor impairments often lose the ability to interact with conventional assistive mobility devices such as joystick, head, or voice-controlled wheelchairs. For users whose only reliable voluntary movement is through eye motion, existing solutions remain limited, costly, or impractical for everyday use. This paper presents a low cost, real-time eye-gaze based human–computer interaction system for assistive wheelchair navigation, designed to restore autonomy through an accessible and user-centered interface. The proposed system employs a compact embedded vision setup in which a standard webcam continuously captures ocular movements, while a lightweight perception-driven computer vision pipeline processes the video stream on embedded hardware. Using image preprocessing, facial landmark detection, and gaze-ratio analysis, the system interprets user intent and maps eye movements to differential motor commands in real time. A blink-based safety mechanism provides robust and intentional control, enabling immediate stopping and safe operation in dynamic indoor environments. Built entirely on affordable, off-the-shelf components, the system emphasizes low computational overhead, fast responsiveness, and ease of deployment in home and clinical settings. Beyond functional mobility, this work demonstrates how human-centered interaction design and interactive machine-learning principles can enhance accessibility, promote dignity, and support more inclusive eHealth and rehabilitation technologies.

BH5050

Tactual Tides: Embodying Environmental Crises through Tangible Interaction with Generative AI

Jian Wei Lin and Saiau Yue Tsao
National Taiwan Normal University, Taiwan

Abstract—Environmental crises often feel abstract and distant to the public. This paper presents Tactual Tides, an exploratory interactive installation that bridges Tangible User Interfaces (TUIs) and Generative AI systems. Unlike real-time text-based prompting, our system adopts a "Pre-rendered Latent Space" strategy, utilizing a physical sandbox as a proxemic controller to map user distance to pre-computed ecological narratives. A field study with 20 participants demonstrates how this approach fosters embodied agency and emotional resonance. While the visual assets are pre-rendered to ensure interaction fluidity,

our findings suggest that grounding generative media in physical interaction enhances the perceived weight of environmental narratives, though further quantitative validation is required to generalize these qualitative insights.

BH5090-A

Biosignal-based Human-Computer Interfaces for Communication and Control

Chang-Hwan Im
Hanyang University, Republic of Korea

Abstract-Biomedical signals, including electroencephalogram (EEG), electromyogram (EMG), and electrooculogram (EOG), can be employed to implement diverse human-computer interfaces (HCIs). These interfaces range from silent speech interfaces (SSIs) and control interfaces for XR devices to EEG-based brain-computer interfaces (BCIs). Computational intelligence and neural engineering (CoNE) laboratory at Hanyang University, South Korea has developed several practical HCI applications based on artificial intelligence (AI) technology. This presentation will showcase their notable projects, such as three-axis accelerometer-based SSI, EMG-based facial expression recognition, EEG-based BCIs for controlling home appliances, and biosignal-based HCIs for metaverse applications. These examples are anticipated to provide the researchers attending MLHMI 2026 with new ideas on the innovative application of HCI technologies.

BH5024-A

Advancing the Informed Consent Process for BioBank Participants Using a Multi-faceted AI and Digital Coaching Solution

Fuad Abujarad, Chelsea Edwards and Kien Pham
Yale University, USA

Abstract-The informed consent (IC) process is essential to research but faces persistent challenges. Traditional IC methods are often lengthy and difficult to comprehend, leading participants to consent without fully understanding risks and benefits. This limits shared decision making and may cause harm, especially in high-stakes clinical settings. We propose a multi-faceted solution using AI and digital coaching assets to improve comprehension, user engagement and personalization.
Our digital IC platform combines generative AI and interactive digital coaching to reshape creation and delivery of IC. In collaboration with the Yale Biobank, our pilot research will support this approach, including transformed displays of key content, multimedia (video animations, generative audio and text-to-speech, etc.), and features allowing deeper exploration into complex topics (Figure 1).

Preliminary findings demonstrate improved participants comprehension, acceptability and engagement. Participants enjoy the concise format and ability to dive deeper into complex topics and experience less cognitive burden while retaining essential IC content.

These results suggest that a multi-faceted approach using AI and digital coaching techniques can meaningfully improve the IC process to better support participants. Ongoing development will focus on refining the platform and establishing best practices for scalable, AI-supported IC to improve regulatory compliance, ethical standards, and participant trust.

Session 5 (14:50-16:25)

Topic: 3D Image Modeling and Virtual Technology Applications
Session Chair: Prof. Shyi-chyi Cheng, National Taiwan Ocean University, Taiwan
Time: 14:50-16:25, March 18, 2026
Venue: Global Front, Room: 3rd Floor, 4031

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH0009-A

Development and Evaluation of a Virtual Reality Learning System on Sandimen Indigenous Culture and Local Ecology: Implications for Learning Outcomes, Motivation, and Cognitive Load

Tzu-Yen Lin, Wernhuar Tarng and Tzu-Jen Ding
National Tsing Hua University, Taiwan

Abstract- This study integrates Virtual Reality (VR) and drone technology to create an immersive learning system combining Paiwan Indigenous culture with natural science education, focusing on the Sandimen tribe. Drone imagery and 3D modeling reconstructed key sites, including the Paiwan Village, Sandimen Cultural Center, Ailiao North Stream Valley, and Saija Forest Trail. The system integrates elementary-level science and social studies, embedding Paiwan cultural elements—such as traditional architecture, rituals, crafts, and ecology—into interactive experiences. Sixty fourth-grade students from Hsinchu City participated in a quasi-experiment: 30 used the VR system with worksheets, while 30 learned through traditional slides. Both studied the same topics on Paiwan culture and the environment across three 40-minute sessions. Research tools included satisfaction, achievement, motivation, and cognitive load assessments. Results showed high system usability, with the VR group achieving significantly better cultural and environmental understanding, higher motivation, and lower cognitive load than the control group. These findings suggest that VR and drone-based learning effectively integrate Indigenous knowledge into classrooms, enhance cultural and environmental education, and support the digital preservation of Paiwan heritage while fostering curiosity and engagement in science learning.

BH0015

SPAFormer: Super-Point Attention Based Point Cloud Transformer for 3D Object Segmentation

Yu-Jie Li, Hsun-Yu Lan, Shyi-Chyi Cheng and Yu-Jen Kao
National Taiwan Ocean University, Taiwan

Abstract- Recent advances in deep-learning assisted computer vision facilitate 3D scene modeling in real-time. In this paper, we propose a super-point attention based transformer, named SPAFormer, for segmenting objects in 3D point clouds. SPAFormer first uses the proposed super-point detector to quantize and order 3D points into a 1D spatial sequence; the sequence is then partitioned into fixed-size super-points with the same number of points (K = 128), while each super-point may have an irregular spatial shape. Next, a classifier annotates each super-point for 3D object segmentation. The key challenge in using point-cloud transformers is to reduce the quadratic and often prohibitive computational complexity of attention to encode the input point cloud. The proposed super-point attention applies standard multi-head self-attention (MHSA) within each fixed-size super-point, enabling GPU-friendly batched computation and reducing the quadratic cost of global attention. Furthermore, our super-point detector keeps super-point detection scalable for processing large-scale point clouds with linear complexity. Experimental results demonstrate that SPAFormer delivers competitive performance on 3D benchmarks, while significantly improving efficiency through super-point attention, enabling faster inference with a compact parameter budget.

BH0044

VMNet：Voxel-based Multimodal Fusion Networks for 3D Object Detection

Hao-Chun Chang, **Cheng-Hung Chen**, Hsing-Hua Lu, Qi-Yun Lu, Chia-Hung Lin and Pei-Zhen Lin
National Formosa University, Taiwan

Abstract- This study proposes a voxel-based multimodal fusion network (VMNet) that integrates LiDAR point clouds and image information to enhance the accuracy and stability of 3D object detection. Unlike conventional single-modality approaches, VMNet first voxelizes the point cloud data and employs a deep learning architecture to extract features from both images and point clouds. These features are fused three times at different stages of the multi-level backbone via the proposed LI-Fusion module, significantly improving the model's ability to recognize fine object details. Furthermore, the original 2D bird’s-eye view (BEV) backbone is replaced with a residual structure to alleviate the gradient vanishing problem and preserve more feature information.

Experimental results on the KITTI benchmark dataset demonstrate that the proposed method achieves improvements of 1.21% and 1.44% in accuracy under moderate and hard detection settings, respectively, compared to the original Focals Conv-F model. These results verify the effectiveness and potential of the proposed approach in 3D object detection tasks.

BH0022

A Study on Estimating Vertebral Positions Using Three-Dimensional Analysis of Back Surface Topography

Su-Jin Ha[1], Ju-An Kimb[1] and Chang-Suk Cho[2]
1: Hanshin University, Korea
2: ShapeSpaceXLab Co., Korea

Abstract- This study proposes a method for indirectly estimating vertebral alignment and scoliosis deformation by first estimating the spinal sulcus point from three-dimensional back surface geometry measured using a non-contact 3D sensor. The purpose of this study is to enable primary screening measurement of spinal deformities in environments where radiation-based imaging devices such as X-ray are difficult to use, thereby serving as a screening tool that can refer suspected cases to precise radiological diagnosis when abnormal findings are detected. To this end, a non-contact back surface measurement system based on an active stereo infrared (IR) 3D depth sensor was constructed, and a multi-channel one-dimensional convolutional neural network (1D-CNN)-based classification model was designed to detect spinal sulcus points from the depth topology of the back surface. In addition, to compensate for discontinuities and outliers in slice-wise estimation results, an autoencoder-based standardization and interpolation technique for spinal sulcus point vectors was applied. To validate the proposed method, a comparative analysis was conducted between the ground-truth spinal curvature values extracted from X-ray images and the back surface-based estimation results for 20 subjects. The results demonstrated a mean absolute error of 1.364% relative to the ground truth, confirming the high precision of the proposed approach. This study demonstrates the potential of the proposed method as a non-invasive spinal screening and early diagnostic support technology that enables quantitative evaluation of spinal deformities without radiation exposure.

BH0054

An Improved Beautification Method for Handwritten Character Using Generative Adversarial Networks

Kota Fukushima, Akihiro Ito and Qiu Chen

Kogakuin University, Japan

Abstract- Handwriting conveys rich personal characteristics and affective information that are difficult to express using standardized digital fonts; however, the widespread use of word processing software has reduced opportunities for handwriting, leading to uniform textual expression and declining writing skills. To address this issue, this paper proposes a handwritten character beautification method using the DG-Font model. We formulate this task as an Image-to-Image translation problem, where the model takes a single scanned handwritten character image as input and outputs a beautified character image that preserves individual stroke characteristics while improving structural clarity. The proposed method employs a two-stage training strategy: pre-training on 19 types of printed fonts and 2,141 Joyo Kanji (regular-use Chinese characters), followed by fine-tuning on handwritten character images. This approach successfully achieved character generation that retains the distinct stroke characteristics of the original handwriting while adopting the well-structured skeleton of the target font (KGothic). Furthermore, dilation and centering are introduced as pre-processing techniques to enhance thin stroke representation and stabilize generation quality. We compared our approach with the zi2zi baseline. Experimental results, evaluated using LPIPS against both the source printed font and the original handwriting, along with a user study on readability, demonstrate that the proposed method achieves a better balance of handwritten naturalness and structural correctness.

BH0010-A

An Immersive VR Learning System Combining Ground and Aerial Perspectives for Cultural Heritage Education: The Case of the Lin Family Mansion, Wufeng

Xiang-Ting Wang, Wernhuar Tarng and Tzu-Jen Ding
National Tsing Hua University, Taiwan

Abstract- This study integrates three-dimensional (3D) modeling and Virtual Reality (VR) technologies to develop an immersive learning system that merges first-person ground views with aerial perspectives, exploring the educational applications of digitized cultural heritage. While traditional digital preservation focuses on restoration and archiving, its pedagogical value remains underexplored. Addressing this gap, the study transforms 3D models into interactive VR learning materials, enabling learners to explore spatial structures, historical contexts, and cultural meanings from multiple viewpoints. The Lin Family Mansion and Garden in Wufeng, Taiwan's largest and best-preserved Qing dynasty residence, serves as the case study. Featuring ceremonial, residential, and landscape spaces arranged along a symmetrical axis, the site exemplifies Qing-era Minnan architectural hierarchy and aesthetics. Employing the Structure from Motion (SfM)

algorithm and Unity-based VR360 drone-guided exploration, learners can freely switch between perspectives for both detailed and holistic understanding. The system mitigates the cost and preservation limits of on-site teaching while maintaining spatial authenticity and presence. By integrating digital heritage, immersive VR, and exploratory learning, this study establishes a cross-disciplinary framework for cultural heritage education and technological innovation.

BH0024

ACANE: Asymmetric CLS Attention Nullification for Enhanced classification

Rinka Kiriyama[1], Akioa Sashima [1, 2] and Ikuko Shimizu[1]
1: Tokyo University of Agriculture and Technology, Japan
2: National Institute of Advanced Industrial Science and Technology, Japan

Abstract- Vision Transformers have revolutionized computer vision through their self-attention mechanisms, yet they exhibit an important inefficiency in few-shot learning: forced attention distribution from the [CLS] token to image tokens in early layers. Due to SoftMax normalization constraints, the [CLS] token must distribute weak but non-zero attention across all image tokens, creating attention noise that accumulates across layers and degrades classification performance. This problem is particularly acute in few-shot scenarios where limited training data prevents models from learning optimal attention suppression patterns. We propose ACANE (Asymmetric CLS Attention Nullification for Enhanced classification), a parameter-free approach that architecturally enforces zero attention from [CLS] to image tokens in early layers while preserving essential information flows in later layers. Our method is grounded in systematic analysis showing that [CLS] attention patterns naturally transition from potentially harmful (layers 1-8) to beneficial (layers 9-12) as representations mature. Through comprehensive experiments on 11 diverse datasets, ACANE demonstrates consistent improvements: 1.53% gain over linear baselines in 4-shot learning and competitive performance with LoRA while requiring zero additional parameters. When combined with LoRA, ACANE provides further 0.99% improvement, demonstrating the complementary nature of architectural and parameter-based adaptations.

Session 6 (14:50-16:25)

Topic: Artificial Intelligence and Applications in Digital Information Systems

Session Chair: Assoc. Prof. Fuad Abujarad, Yale University, USA

Time: 14:50-16:25, March 18, 2026

Venue: Global Front, Room: 2rd Floor, 4021

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH5059

Navigating the Control Paradox: Trust and Transparency in AI-Powered Educational Analytics

Ronel F. Ramos, Von Erick Mortel Magbitang, Roman M. De Angel and **Juan Paulo H. Magcuyao**

FEU Insitute of Technology, Philippines

Abstract—Artificial intelligence (AI) is increasingly integrated into learning analytics, offering predictive analytics, automated evaluations, and customized learning experiences. Although such systems are anticipated to be more efficient and fairer, the Control Paradox arises. This paradox combines human control with the autonomy of AI. Teachers aspire to control and be held accountable. However, AI tends to be a “black box,” generating values that are not easily understandable and disputable. This study will examine a mixed-methods approach to enhance the transparency of trust perceived within the context of the control paradox between human control and the autonomy of AI tools. In the quantitative method of this study, literature will be reviewed using the thematic method, while the quantitative method will survey educators about control, trust, and transparency views of AI tools. Results will be interpreted to inform a conceptual model that combines the tools of explainable AI (XAI), emotionally intelligent designs, and human-in-the-loop approaches.

BH5062

Design of a Human-Centered Digital Workflow System Supporting Learning-Driven Laboratory Operations

Josh Marco Goc-ong, Karyll Grace Bontuyan, Arabella Ortega and **Rhodessa Cascaro**

Mapúa Malayan Colleges Mindanao, Philippines

Abstract—This study presents the design of a human-centered digital workflow system that supports learning-driven laboratory operations in public agricultural laboratories. While workflow redesign initiatives are commonly adopted to address inefficiencies in public services, their effectiveness is often limited when redesigned processes are not adequately supported by interactive systems. To address this gap, this research applies a design science approach to develop a role-based Laboratory Information Management System (LIMS) that operationalizes redesigned laboratory procedures through guided task execution, workflow state visibility, and assisted automation features. The system emphasizes human–computer interaction principles by aligning interfaces with user roles, reducing cognitive load, and enabling users to learn workflow progress through real-time status feedback and audit trails. Observed design walkthroughs indicate that the system supports role clarity, reduces coordination ambiguity, and strengthens user trust during workflow execution, even prior to full automation of laboratory operations. The proposed design demonstrates how humancentered digital systems can support learning in process automation and strengthen accountability in public laboratory services.

BH5081

Design and Evaluation of a Generative AI– and AR-Enhanced Interactive Learning System in Elementary EFL Education

Kevin C. Tseng[1,2] and **Li-yu Tseng[1,3]**

1: College of Design, National Taipei University of Technology, Taiwan

2: Department of Industrial Design, Taiwan

3: China University of Science and Technology, Taiwan

Abstract—Generative AI and mobile technologies enable scalable multimodal instructional design, yet field-based HCI evidence in rural elementary settings remains limited. This study presents the design and evaluation of a generative AI–assisted, QR-triggered context-aware EFL learning system that links physical campus scenes to cloud-hosted multimedia cues. Learning materials were produced through a standardized AI-assisted image-to-image workflow grounded in on-site campus documentation. A two-session field deployment with rural second-grade students (N = 11) was conducted in 2023. Session 1 (May 18) included a vocabulary pre-test and guided exploration; Session 2 (June 1) implemented full QR-triggered instruction followed by an 8-item post-test. Significant learning gains were observed: Group 1 improved from 2.00 to 7.83 ($Z = -3.022$, $p = .003$), and Group 2 reached ceiling performance (1.80 to 8.00). Learner satisfaction was high (course content: M = 4.91/5). The findings suggest that lightweight contextual micro-interactions combined with controlled AI-assisted content authoring provide a deployable and pedagogically aligned HCI design strategy for rural EFL education, supporting SDG

4 (Quality Education) under resource constraints.

BH5042

How Many Models to Combine for Ensemble Learning?

Elio Zhang[1] and Craig A. Rolling[2]
1: Seattle University, USA
2: St. Louis University, USA

Abstract—Because high dimensionality makes model identification harder, a well-constructed ensemble—formed through weighted model combination—can often deliver substantial gains in predictive accuracy over selecting a single model when the number of covariates is large. Yet model combination in highdimensional regression remains under-studied, largely because it is unclear which models to combine and how many to include. A related question is whether model combination truly helps when a single strong model actually exists. To address these issues, we introduce Selected Model Averaging (SMA), a resampling-based procedure that adaptively determines which models to combine and how many to include. Unlike many existing model-averaging approaches, SMA collapses to model selection when appropriate, thereby bridging the gap between selection and combination. Both theoretical results and numerical experiments show that SMA performs well across a range of high-dimensional linear regression settings.

BH5061

Empowering Virtual Classrooms with Agentic AI: A Humanized Approach to Cognitive Load Reduction

Angelo C. Arguson, Shaneth C. Ambat, Elisa V. Malasaga and Ronel F. Ramos
FEU Institute of Technology, Philippines

Abstract—This study examined Agentic Artificial Intelligence (AI) in virtual classrooms to solve cognitive load issues and improve engagement in higher education. Using an explanatory sequential mixed-methods approach, this study surveyed 120 students and divided them into two groups, one using Agentic AI in virtual classrooms and other using traditional virtual classrooms. Quantitatively, this study found Agentic AI significantly reduces extraneous cognitive load and videoconferencing fatigue and increases germane processing, engagement, and retention. Chunking and switching interventions and breaks significantly reduce extraneous cognitive load without changing intrinsic task complexity. Qualitatively, this study found incorporating transparency, such as AI system dashboards and override options, into Agentic AI systems promotes trust and enhances user experience/functionality to comply with ethical considerations and user autonomy. The

system architecture, as presented, also relies on real-time telemetry and decision-making processes that ensure personalized interventions while adhering to privacy protections. The results affirm that human-centered, ethically driven agentic workflows play a crucial role in maximizing cognitive ergonomics and ensuring persistence in digital learning processes. The results of this research contribute significantly to strategies for higher educational institutions in the Philippines; thus, it contributes significantly to meeting global concerns and advocacy in utilizing technological innovations and responsible AI governance. It also touches on both cognitive and ethical aspects, emphasizing how Agentic AI has visions of transforming virtual learning processes and creating trusting environments.

BH5063

Quantifying and Redesigning Public Laboratory Workflows Using a Unified Swimlane–KPI Assessment Protocol

Arabella Ortega, Josh Marco Goc-ong, Karyll Grace Bontuyan and Rhodessa Cascaro
Mapúa Malayan Colleges Mindanao, Philippines

Abstract—This study presents a unified workflow assessment protocol that integrates Swimlane Process Mapping with logderived key performance indicators to learn operational performance patterns and guide data-informed redesign decisions in public laboratory workflows. Applied to a regional agricultural laboratory, the protocol established an AS-IS baseline of turnaround time, structural complexity, and service reliability, revealing prolonged delays associated with multiple handoffs, redundant approval layers, and manual data handling. Guided by Business Process Reengineering principles, a TO-BE workflow was designed to eliminate non-value-adding steps, consolidate approvals, and standardize data capture. The results demonstrate substantial reductions in workflow steps, handoffs, and approval levels, indicating that learning from historical interaction traces embedded in operational data can effectively inform human-centered process redesign. The proposed protocol offers a replicable approach for improving efficiency, transparency, and accountability in public laboratory services.

Session 7 (14:50-16:25)

Topic: Intelligent Text and Image Generation Systems Based on Cross-Modal Information Fusion
Session Chair: Prof. Ran-Zan Wang, Yuan Ze University, Taiwan
Time: 14:50-16:25, March 18, 2026
Venue: Global Front, Room: 3rd Floor, 403D

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH0018

A Dify-Integrated InternVL-Based Multimodal System for Breast Cancer Report Classification and Medical Diagnostic Report Question Answering

Cheng-Ta Huang[1], **Mei-Fang Zhuang[1]**, Hao-Yu Weng[2], Ruei-Chi Lin[2], Wei-Jen Wang[2] and Chia-Ying Lee[3]
1: Yuan Ze University, Taiwan.
2: National Central University, Taiwan
3: Radiological Diagnosis, Far Eastern Memorial Hospital, Taiwan

Abstract- This study presents an integrated multimodal diagnostic report analysis system that combines Dify, MinerU, and InternVL to support both breast cancer classification and medical question answering. The system employs Dify as the workflow orchestration framework to coordinate model execution, manage data flow, and unify the interaction between text based and vision-based components. MinerU is used for document parsing through optical character recognition, table extraction, and layout analysis, converting PDF based diagnostic reports into structured machine interpretable representations. InternVL provides multimodal semantic reasoning for two complementary tasks. First, it functions as a binary classifier to distinguish benign from malignant breast cancer reports. Second, it performs clinical question answering by integrating text and visual cues derived from structured document content. Experiments conducted on publicly available breast cancer and medical laboratory report datasets show that the system achieves perfect performance in benign malignant classification and produces fully correct responses in a ten-question arterial blood gas evaluation. These results demonstrate that the combined use of Dify, MinerU, and InternVL provides an effective and extensible framework for automated medical report interpretation. The findings highlight the potential of the proposed architecture as a scalable solution for intelligent clinical document processing

and multimodal reasoning in healthcare applications.

BH0029

VQA: improving ambiguous question answering through a multi-round clarification mechanism

Yang Wang and Katunobu ITOU
Hosei university, Japan

Abstract- This study investigates ambiguous question answering in multimodal visual dialogue, where users interact with a system through multi-round questions over images. In visually complex and semantically uncertain scenarios, users often pose ambiguous questions, leading to misunderstandings by the system. To address this, we propose a framework for automatically detecting question ambiguity types and generating targeted clarification questions. The framework first detects the ambiguity of a question using answer distribution entropy and visual-textual semantic similarity as signals, and identifies four types of ambiguity: referential, attribute, intent, and usage. To generate clarification questions corresponding to these ambiguities, the system constructs a belief state containing object attribute information to represent the visual context. Based on this belief state, an LLM generates contextually appropriate clarification questions over multiple dialogue turns, guiding the user to iteratively disambiguate their query. Experiments on the SIMMC 2.0 dataset demonstrate the effectiveness of the framework in ambiguity detection, accurately identifying both ambiguous questions and type of ambiguity. Evaluation on a self-collected supermarket shopping dataset shows that the belief state–based generation significantly improves the relevance and effectiveness of the clarification questions. The proposed framework provides a complete multi-turn pipeline that detects question ambiguity, tracks and utilizes belief state information, and leverages a large language model to generate natural, human-like clarification dialogue in multimodal visual scenarios.

BH0039-A

Medical Decision Support System with Knowledge Graphs for Alzheimer's Disease

Chi-Wei Fu[1], Tzu-Liang Hsu[1], Chieh-Ni Chen[2], Kuang-Chung Hsu[3] and **Chien-Lung Hsu[1]**
1: Chang Gung University, Taoyuan, Taiwan
2: National Chengchi University (NCCU), Taiwan
3: National Taipei University of Technology, Taiwan

Abstract- This study presents the design and implementation of a knowledge-graph-driven

medical decision support system (MDSS) for Alzheimer's disease (AD), aiming to bridge precision medicine with evidence-based clinical decision-making. Alzheimer's disease is characterized by complex and heterogeneous pathophysiology, slow and costly drug development, and substantial discrepancies among international clinical guidelines. These challenges limit the effectiveness of conventional clinical decision support systems, which often rely on isolated datasets and lack the capability to integrate heterogeneous biomedical knowledge across molecular, clinical, and guideline levels. To address these limitations, this research adopts PrimeKG, a large-scale precision medicine knowledge graph developed by Harvard University, as the core biomedical knowledge backbone. The system further incorporates international clinical guidelines, systematic reviews, and clinical trial data, enabling comprehensive integration of both structured and unstructured medical information. An automated workflow management platform is employed to support data preprocessing, transformation, and continuous updates, while the Neo4j graph database is used to model, store, and visualize complex biomedical entities and their relationships. The resulting platform provides an interactive, queryable, and extensible clinical decision support environment tailored to Alzheimer's disease. The key contributions of this study are summarized as follows: (i) A multi-source biomedical knowledge graph for clinical decision support is constructed and visualized, enabling clinicians and researchers to intuitively explore complex relationships among diseases, drugs, genes, and clinical evidence through an interactive graph-based interface. (ii) Disease-centric knowledge graph refinement for Alzheimer's disease is proposed, filtering and restructuring large-scale precision medicine knowledge to emphasize Alzheimer's-specific drug–gene–disease associations, thereby supporting in-depth mechanistic and therapeutic analysis. (iii) An AI-assisted knowledge extraction pipeline is introduced to transform unstructured textual sources—such as clinical guidelines, systematic reviews, and trial reports—into structured knowledge graph representations using automated relation extraction techniques. (iv) Evidence-based integration of clinical guidelines, systematic reviews, and clinical trials is achieved, enriching the knowledge graph with up-to-date therapeutic evidence, comparative treatment effectiveness, and evolving Alzheimer's drug development data. (v) A knowledge-graph-based clinical decision support system is developed to assist physicians in treatment planning by systematically referencing molecular evidence, clinical guidelines, and trial outcomes, and by supporting therapy recommendations grounded in aggregated evidence rather than isolated data sources. (vi) In summary, this study demonstrates that integrating knowledge graphs with precision medicine significantly enhances the capability of medical decision support systems to unify heterogeneous biomedical knowledge. By incorporating authoritative clinical guidelines, systematic reviews, and clinical trial evidence, the proposed system advances the practical applicability of knowledge-graph-based decision support in real-world Alzheimer's disease management. The framework is designed to be scalable and

transferable, offering a reference architecture for future clinical decision support systems targeting other complex diseases.

BH0045

Candlestick Chart Representations of Musical Scores: A Cross-Domain Deep Learning Study on Genre Classification and Provenance Tracing

Yung-An Chen, Yun-Chi Cheng, Ssu-Yu Yeh, Yu-Cheng Lin
Yuan Ze University, Taiwan

Abstract- Music genre recognition has significant practical value in the field of music information retrieval. However, traditional approaches mostly rely on audio features or symbolic analysis, and they still face challenges in capturing subtle differences between genres. To explore whether genres can be effectively identified through visual representations, this study builds on the YNote notation system proposed by Lu, S. C., et al., converts sequences of musical events into candlestick (K-line) charts, and conducts classification experiments using deep learning models. In terms of methodology design, this study compares two models: ResNet-50 and the Vision Transformer (ViT). The dataset is split into training, validation, and test sets according to a predefined ratio, and performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and the confusion matrix. The experimental results show that deep learning models can achieve stable classification performance under the proposed representation and can, to a certain extent, distinguish the Jiangnan style from other styles. By analyzing the training process and the confusion matrix, this study further investigates the models' classification behavior and limitations. The findings confirm that candlestick charts, as a visual representation converted from symbolic music, have the potential to preserve genre-related information, and they offer a feasible direction for integrating music notation systems with deep learning methods for image-based classification tasks.

BH0031

Integrating Image Information into Text2Loc: An Extension for Natural Language-Based 3D Point Cloud Localization

Tomoki Azuma and Seiichi Tani,
Nihon University, Japan

Abstract- With recent advancements in digital twin technology, high-precision localization within large-scale urban 3D point cloud data has attracted research attention. In particular, localization methods utilizing natural language inputs enable intuitive operations, contributing to the realization of flexible, user-friendly interfaces. While

conventional methods successfully match natural language descriptions with 3D point clouds, approaches integrating auxiliary image information remain underexplored. In our preliminary experiments, we integrated a ResNet-based image feature extraction model into the existing Text2Loc method. Although this yielded improvements on validation data, performance on test data was inconsistent, demonstrating that simply adding global image features has limited effectiveness. Addressing this limitation, we developed a novel auxiliary model that leverages direct object information. This model explicitly incorporates feature extraction based on object categories detected by YOLO, the addition of image positional information, and a feature integration process that accounts for relationships between multiple images. Experimental evaluations indicated that fusing this auxiliary model with the baseline achieved the highest performance on both validation and test data, although the difference was marginal. Furthermore, we conducted experiments by removing improvement components individually to verify the specific contribution of each element to the performance improvement. These experiments revealed that the YOLO-based object detection features and the relationship-aware feature integration were the primary contributors to accuracy gains. Finally, to verify generalizability, we applied our proposed integration method to another existing localization architecture. The results showed that while improvements on the validation data were limited, the test data exhibited a marginal increase in accuracy. These findings suggest that our method can serve as an effective enhancement for localization systems that utilize point clouds and natural language, leading to improved accuracy.

BH0052

Automatic Seed Optimization for Personalized Text-to-Image Generation with Diffusion Models

Yu Yamamoto and Qiu Chen
Kogakuin University, Japan

Abstract- Text-to-Image diffusion models are highly effective but remain heavily sensitive to the initial noise. This sensitivity causes significant instability in personalization tasks, where maintaining a specific subject's identity is crucial. While inferencebased methods like the W+ Adapter offer efficient alternatives to costly fine-tuning, they suffer from structural conflicts between identity preservation and prompt consistency depending on this noise. In this study, we address this issue by proposing an automated discrete "Latent Space Exploration" framework utilizing random search to optimize seed selection. We compare our discrete seed optimization approach against "Initial Noise Selection," a continuous optimization method that modifies the noise tensor directly via gradient descent. We define a multi-objective scoring function integrating text consistency (CLIP), identity preservation (ArcFace), and structural validity (MTCNN). Quantitative

experiments reveal a critical trade-off: while continuous optimization preserves identity competitively, it frequently degrades text consistency by ignoring prompt contexts like clothing or backgrounds. In contrast, our discrete exploration achieves a superior balance, ensuring a 100% face detection rate while maximizing both identity fidelity and text alignment. Furthermore, a subjective evaluation with 151 participants confirms that our method yields significantly higher overall visual quality and prompt fidelity. We conclude that discrete seed optimization offers a robust and practical solution for personalized generation.

BH0038

Artistic Dual-Message QR Code Generation by Text-Guided Diffusion Models

Chi-Han Lin, Chih-Chieh Wu and Ran-Zan Wang
Yuan Ze University, Taiwan

Abstract- This paper presents a novel method for generating artistic dual-message QR codes using Stable Diffusion models. The proposed approach first creates an aesthetic QR code by employing a text prompt to generate image content, guided by ControlNet to embed the first-layer message into the visual structure. A subsequent module stamping process is then applied to embed a second-layer message. The resulting QR code conveys visual information to human viewers through its artistic appearance, while simultaneously embedding two layers of QR code messages. One message can be decoded from a close distance, while the other message from a farther distance—both using standard QR code readers. Experiments are conducted to evaluate the robustness of decoding for both message layers, as well as to assess the visual quality of the generated QR codes. The ability to encode dual-layer messages within a single QR code offers new possibilities for designing enriched information interfaces. For instance, one layer may provide static content for immediate access, while the other offers a link to multimedia data retrievable via the Internet.

Session 8 (14:50-16:40)

Topic: Intelligent Image Analysis and Processing Methods
Session Chair: Prof. Tetsuya Shimamura , Saitama University, Japan
Time: 14:50-16:40, March 18, 2026
Venue: Global Front, Room: 3rd Floor, 403N

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH0032

A Reachability Distance Detector and a Nonlocal Means Framework for Mixed Noise Reduction in Color Images

Bogdan Smolka[1], Damian Kusnik[1] and Milena Smolka[2]
1: Silesian University of Technology, Poland
2: AGH University of Science and Technology, Poland

Abstract- In this paper, we introduce a fast method for reducing mixed Gaussian and impulsive noise in color images. The proposed approach integrates a reachability distance based impulse detector, which effectively identifies and localizes clusters of impulsive pixels that degrade the visual content. To further distinguish outliers from Gaussian-distributed fluctuations, a Cox–Box transformation is applied to the histogram of the estimated noise intensity map, enabling reliable separation of impulsive noise from the underlying Gaussian component.

Pixels identified as having been corrupted by impulsive noise are restored by replacing them with the average intensity of the undistorted pixels within their local neighborhood. After this initial correction stage, the remaining pixels, primarily affected by Gaussian noise, are denoised using a Non-Local Means (NLM) filter. The NLM parameters are adaptively tuned according to the estimated level of residual noise, ensuring effective smoothing while preserving essential image structures.

The experimental results indicate that the method is computationally efffcient, requiring only a small processing block size, and performs on par with the state-of-the-art filters. Moreover, it remains robust even when the noise level is high, a situation in which many alternative approaches suffer significant performance losses. The enhanced denoising accuracy broadens its potential use as a preprocessing stage for downstream image analysis tasks such as segmentation, edge extraction, and object recognition.

BH0033

Image enhancement with recursively separated fuzzy histogram equalization

Hafijur Rahman and Tetsuya Shimamura
Saitama University, Japan

Abstract- Enhancing image visibility through contrast enhancement (CE) is widely important in fields such as medical and satellite reconnaissance, as it enhances the distinctness of features. However, many CE methods still alter mean brightness, generate visual imperfections, and cause unnatural shifts toward overly bright or dark outputs. To reduce these problems and gain superior enhanced outputs, we propose the recursively separated fuzzy histogram equalization (RSFHE) method. Our method begins by calculating a fuzzy image histogram (FIH). Following this, a specific partitioning level (PL) is computed from the FIH. This PL is then used to divide the FIH into two sub-FIHs. In a similar way, these two sub-FIHs are divided, which produces four sub-FIHs in total. This FIH dividing is a crucial step for preserving mean brightness. The final enhanced image is produced by applying equalization to every sub-FIH. We compare the RSFHE method against several conventional and cutting-edge CE methods on a standard collection of slightly low-contrast images. Visual assessment and numerical measures—including AMBE, contrast, MSSIM, and PSNR—are used for comparison. Our findings show that the proposed RSFHE method reduces the described problems and provides superior performance relative to the other methods examined in this test

BH0040

Robust Supervision for NeRD-Rain: Stable Single-Image Deraining under a Limited Training Budget

Cian-Yi Yang and **Shang-Kuan Chen**
Yuan Ze University, Taiwan

Abstract- Single image deraining is a critical preprocessing step for outdoor vision systems, as rain streaks and veiling effects can degrade structural details and hinder downstream recognition tasks. This study proposes a robust training framework to enhance the NeRD-Rain backbone without increasing inference-time complexity by imposing explicit structural and spectral constraints during training. The framework introduces two lightweight, training-only auxiliary heads: a self-reconstruction head at the finest scale to enforce feature consistency, and a channel-consistency head at the mid-scale to regularize RGB relationships and reduce color distortion. A dynamic loss-weight scheduling strategy is adopted to stabilize early optimization, gradually shifting the learning focus from auxiliary objectives to fine-grained pixel-level reconstruction.

Experimental results show that the proposed framework achieves PSNR and SSIM comparable to NeRD-Rain, while delivering smoother convergence, fewer residual rain artifacts, and more stable cross-dataset performance, all without additional inference cost. These results suggest that multi-scale, frequency-aware supervision with carefully scheduled auxiliary heads effectively improves training stability and robustness in image deraining backbones.

BH0025

Fringe detection from optical coherence tomographic images: a hybrid approach

Md Emran Ali[1], Yume Hashimoto[2], Satoe Murazawa[2], Shuto Onodera[2], Rebeka Sultana[1], Ikuko Shimizu[1] and Tatsutoshi Shioda[2]

1: Tokyo University of Agriculture and Technology, Japan

2: Saitama University, Japan

Abstract- In this study, an automatic tomographic fringes (signal) detecting technique is proposed for single shot, time-domain, optical coherence tomographic images to study the morphological structure of an object. The images were captured by two-dimensional charge-couple devices in laboratory setup of time domain optical coherence tomography. A broadband light source in the visible spectrum was used for obtaining tomographic interference fringes of reflected lights. Light reflections are measured by a Michelson interferometer. The proposed automated approach for fringe detection relied on analyzing statistical variance which was coupled with the Hough transformation. This combined methodology was specifically implemented to accurately identify the fringe patterns, which are indicative of the object's internal structure. The performance of the automated fringe detection was rigorously validated by comparing the computed results against manually prepared ground truth images. The study demonstrated significant agreement between the automated detection and the ground truth, evidenced by strong quantitative scores across key similarity metrics, including the structure similarity index, Jaccard index, and dice coefficient.

BH0053

A Two-Stage Transformer-Based Framework for View Synthesis from a Single Image

Tamochi Fukuda and Qiu Chen
Kogakuin University, Japan

Abstract- View synthesis aims to generate images from novel viewpoints and is a key technology for applications such as free-viewpoint video and immersive VR/AR/MR content. However, existing view synthesis methods typically rely on multi-view input images and suffer from blurred or degraded results when synthesizing regions unseen in the input views, especially under limited input conditions. This paper proposes a method for high-degree-of-freedom view synthesis from a single input image. The proposed method incorporates a DiT-based generative model and consists of two stages: coarse novel-view generation and subsequent reconstruction. A conventional view-transformation method first generates a rough novel view, after which the generative model compensates for missing information and enhances image quality. By leveraging image generation to address insufficient interpolation in unseen regions, this two-stage process enables high-quality view synthesis without explicitly constructing a 3D model, even when only a single input image is available.

BH0027

Deep Learning for Cell-Level Segmentation of Canine Mast Cell Tumors in Histopathology: Dataset and Baselines Toward Automated Grading

Mark Horpynych[1], **Rebeka Sultana**[2], Momoka Kozawa[2], Tomoaki Murakami[2] and Ikuko Shimizu[2]
1: Czech Technical University, Czech Republic
2: Tokyo University of Agriculture and Technology, Japan

Abstract- Canine mast cell tumors (MCTs) exhibit variable malignancy, often leading to diagnostic ambiguity and interobserver variability in histopathological grading. To facilitate objective analysis, we present a new dataset of 49 H&E-stained microscopy images from five canine MCT cases, featuring dense manual segmentation masks for mast cells, their nuclei, and other cells. We benchmark U-Net-based architectures against a fine-tuned Cellpose foundation model and investigate the impact of backbone complexity, pretraining strategies, stain normalization, and data augmentation. Experimental results show that while a fine-tuned Cellpose baseline reaches a Dice score of 0.797 on the test set, the best-performing specialist model—a U-Net++ initialized with public histopathology weights improves performance, achieving a Dice score of 0.8605 for mast cells. Furthermore, it provides multi-channel outputs and predicts other cellular classes

with adequate accuracy. We observe that while stain normalization improves discriminability, it exacerbates overfitting in this low-data regime, whereas robust augmentation proved crucial for generalization. This work provides a baseline for dense cellular segmentation in canine MCTs, demonstrating that specialized architectures with transfer learning can outperform generalist foundation models in specific histopathology tasks.

BH0004-A

Land Use Semantic Segmentation Using Multilayered Raster Images

Aaron Bramson
GA technologies Inc. Waseda University, Japan

Abstract-Here we expand on earlier research to extract "pleasant vegetation" such as trees, bushes, gardens, and grass from readily and freely available aerial image data. Land use segmentation is an active and important area of study for many reasons, but typically utilizes specialized satellite or drone imagery datasets containing non-visible data bands. Because these datasets are limited in coverage and are too expensive for most researchers and planners to obtain, we attempt to extract land-use categories from three sets of freely available basemap image tiles. Our novel contribution is the use of multiple image sources in layers that act as additional data bands to provide additional context. In preliminary results these additional data layers failed to significantly improve results when used without processing; however, preprocessing and encoding features within the images fosters greater segmentation accuracy. Although we achieve usable results for detecting pleasant vegetation, they are not yet superior to alternative methods across all land use categories. We discuss features of the data likely lead to the relatively poor model performance and steps to include specialized transformers to further improve the results.

BH0042-A

Nonlinearity-Aware ACO-Based Routing for High-Quality Image Transmission

M Jyothi Kiran, Raja Datta and Goutam Das
Indian Institute of Technology Kharagpur, India

Abstract- The rapid advancement in imaging and video technologies has led to a significant increase in the demand for ultra-high-quality visual data across a wide range of applications. This demand is particularly critical in highly specialized domains such as medical diagnostics, telemedicine, and surgical imaging, where image and video quality directly influence decision-making accuracy and patient outcomes. Ensuring the delivery of high-resolution images and videos to end users therefore requires reliable data

transmission with extremely low bit error rates and high data throughput. Optical networks play a vital role in meeting these stringent requirements due to their high bandwidth and low latency characteristics. However, the performance of optical transmission is severely affected by various physical layer impairments, especially nonlinear effects, which degrade the overall quality of transmission (QoT). In this work, we propose a novel Ant Colony Optimization (ACO)–based routing and resource allocation framework that explicitly accounts for nonlinear impairments in optical networks. The proposed approach incorporates application-dependent assumptions to effectively evaluate and guarantee QoT while maintaining high transmission efficiency. By intelligently selecting optimal paths that balance performance and physical constraints, the proposed method enhances transmission reliability and supports high-speed data delivery. The results demonstrate the potential of the proposed ACO-based solution in enabling robust, high-quality image and video transmission over modern optical networks.

BH0028

An Android Facial Expression Control System Trainable with a Small Dataset

Kojiro Sato, Takeru Hiruta, Rebeka Sultana and Ikuko Shimizu
Tokyo University of Agriculture and Technology, Japan

Abstract- Controlling the facial expressions of humanoid robots (androids) requires coordinated operation of numerous actuators. Conventionally, actuator parameters are manually adjusted, which is extremely labor-intensive. Recently, deep learning-based methods have been proposed that automatically estimate actuator control values from input images of target facial expressions. However, these methods require large-scale datasets.

In this study, we propose a system that can learn and reproduce individual facial expressions from small training datasets. The key idea is to focus on expression variations relative to a neutral face baseline in a facial image-to-android expression control system.

Session 9 (16:30-17:50)

Topic: Adaptive Machine Learning Models and Intelligent Computing Frameworks for Multimodal Scenarios
Session Chair: Asst. Prof. Siew Mooi Lim, Tunku Abdul Rahman University of Management and Technology, Malaysia
Time: 16:30-17:50, March 18, 2026
Venue: Global Front, Room: 3rd Floor, 4031

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH5031

Toward Efficient Domain-adaptive Anomaly Detection
Tomoya Ueno and **Shin Ando**
Tokyo University of Science, Japan

Abstract—This paper addresses the task of Domain-adaptive Anomaly Detection (DA2D), aimed at detecting anomalies among diverse domains. The techniques to adapt pretrained foundation models to target application domains, such as few-shot learning, are less effective for anomaly detection where the example of the target class, i.e., anomalies, are scarce or inaccessible. Adapted models can be significantly outperformed by task-specific AD models in some target problems. The central challenge of DA2D is to determine the difficulty of a target problem with regard to the pretrained model and adapted features, based on which it chooses to either exploit the adapted model or train a taskspecific model, to achieve high detection performance without expending computational effort for relatively easier targets. DA2D implements an adaptation module using a normalizing flow model, to compare the distributions of normal samples and synthetic anomalies in the adapted feature space and assess its discriminative capacity. Our empirical study shows the advantage of DA2D compared to state-of-the-art few-shot AD and taskspecific AD models in terms of the ROC performances against the training time and data collection cost.

BH5040

Predicting the Green Purchase Intention of Generation Z Café Consumers in Manila using Machine Learning Techniques
Chelsee L. Bustos, Kirk Nathaniel A. Gamorot, Cassey S. Panganiban, Alexander A.

Hernandez and **Roman M. De Angel**
FEU Institute of Technology, Philippines

Abstract—This study analyzes the green purchase intention (GPI) of Generation Z and their factors influencing decisions to patronize green cafés in Manila. Although concerns over the environment are rising, there is much less known about the characteristics of green cafés that determine generation Z consumers to select a green café. To fill this, utilizing six machine learning techniques (Decision Trees, Random Forest, Gradient Boosting, K-Nearest Neighbors, XGBoost, and Support Vector Machines) on a dataset of 444 Gen Z consumers. The results showed that subjective norms, green perceived quality, and social influence were significant predictors, while environmental awareness and eco-social benefits were weak factors in this process. In the combined model, random forest and support vector machine achieved the highest overall model accuracy of 81% when predicting consumer purchasing behavior.

BH5047-A

Machine Learning Feature Selection for Early Autism Spectrum Disorder Screening

Bevina Desjwiandra Handari[1], Adriana Soekandar Ginanjar[1], Rosa Berliana Hasri[1], Siti Nur Salamah[1], Achmad Miftakhul Rachman Ardiva[1], Muhamad Hifzhudin Noor Aziz[2] and Gatot Fatwanto Hertono[1]
1: Universitas Indonesia, Indonesia
2: Universiti Malaya, Malaysia

Abstract-Autism Spectrum Disorder (ASD) is marked by social communication problems and limited, repetitive sensory and motor behavior. It is diagnosed in childhood, or in adulthood. This research yielded an early ASD screening summary based on the dataset ”Autism Screening for Toddlers” (from ages 1-3), and AQ-50 instrumentation (for adults). Machine Learning feature selection picks the most relevant features in early ASD screening. Feature selection was applied to: 1) ensemble filter-based feature selection on Support Vector Machine and Naïve Bayes, 2) hybrid feature selection on Random Forest and Logistic Regression, and 3) embedded feature selection on XGBoost and LightGBM. Filtering techniques such as Chi-Square Test, Mutual Information, and Phi Coeffiecient filtering methods were used in the process. Using an 80:20 training/testing data proportion, the best results for early ASD screening in toddlers were as follows: Logistic regression from 29 to 10 features with 100% values each for recall, f1-score, accuracy, and precision, with an AUC-ROC of 1.000 and a running time of 0.0243 seconds. Results for adults with logical regression from 50 to 28 features were 94.64%, 86.89%, 92.95%, and 80.30% for recall, f1-score, accuracy, and precision, respectively, with an AUC-ROC of 0.9844 and a running time of 0.0612 seconds.

BH5066-A

Comparative Analysis of Data Balancing Methods in Machine Learning Models for Predicting Local Recurrence After Breast Cancer Surgery

Chiu-Feng Chiu[1], Yi-An Wu[1], Chih-Yu Cheng[1], Chen-Cheng Wang[1], Pei-Ying Yang[1], Pei-Ju Chao[1,2,3] Jun-Ping Shiau[1] and Tsair-Fwu Lee[1]
1: National Kaohsiung University of Science and Technology, Taiwan
2: Kaohsiung Chang Gung Memorial Hospital, Taiwan
3: Chang Gung University, Taiwan

Abstract-Purpose: This study evaluates data balancing strategies to enhance machine learning models for predicting local recurrence after breast cancer surgery, proposing an interpretable framework for clinical support.
Materials and Methods: A retrospective cohort of 361 patients was analyzed using 15 perioperative variables. [cite_start]To address data imbalance, eleven techniques—including SMOTE, ADASYN, and Conditional Tabular GAN (CTGAN)—were tested across five classifiers (Random Forest, Logistic Regression, SVM, XGBoost, and Stacking) following LASSO feature selection. Performance was assessed via AUC, F1-score, and recall, utilizing SHAP analysis for interpretability.
Results: Results demonstrated that CTGAN-based augmentation outperformed traditional oversampling methods, achieving the optimal trade-off between sensitivity and generalization. Specifically, combining CTGAN with Logistic Regression yielded the highest predictive performance. Global SHAP analysis identified Oncoplasty, breast operation type, and lymph node operation as the most influential prognostic factors.
Conclusion: Among contemporary resampling approaches for imbalanced tabular clinical data, CTGAN most effectively enhanced minority-class detection in predicting postoperative local recurrence after breast cancer surgery. When combined with Logistic Regression, CTGAN produced an interpretable, high-performing model whose SHAP-based explanations emphasized the prognostic relevance of Oncoplasty, breast operation type, and lymph node surgery, supporting its potential integration into clinical decision workflows.

BH5010-A

Computational intelligence for multi-objective decision making in knapsack problems

Ramesh Srinivasan
University of California, Irvine, USA

Abstract-Artificial intelligence is usually designed to optimize a single objective. Real-

world problems often involve multiple competing objectives; there are a large number of candidate decisions (Pareto-optimal solutions) that reflect compromises on the importance of the objectives. Multi-objective decision making involves the fusion of human strategic judgements with computational optimization. Muliti-objective knapsack problems, a combinatorial optimization problem with constraints, are solved by heuristic strategies by human decision makers leading to suboptimal decisions. A novel computational solution to multi-objective knapsack problems using a Dynamic Bayes Net (DBN) is developed to provide probabilistic estimates of optimal decisions. The DBN can solve sequential and online knapsack problems and provides a robust decision support framework for hybrid thinking to integrate human strategy with artificial intelligence. This approach dynamically updates probabilistic estimates of best next actions based on ongoing human decisions. We demonstrate that this method can be used to allow the user to express their strategic preferences while also finding a Pareto-optimal solution in a simulated card game, and a real-world problem of selecting players for a sports team. We further advance our algorithms by dynamically integrating explicit and implicit human judgments during the computational modeling in order to scale to high-dimensional objective spaces.

BH5051

Application of Deep Reinforcement Learning to Cooperative Traffic Signal Control Considering Pedestrian Push-Button Signals

Taisei Motegi, Naoki Kodama and Taku Harada
Tokyo University of Science, Japan

Abstract— Deep reinforcement learning (DRL) is widely used for traffic signal control. Both vehicle and pedestrian waiting times must be considered in traffic signal control. Furthermore, in real-world environments, in addition to the vehicle signal, a pedestrian push-button signal exists. Therefore, this study proposes a method for applying DRL to traffic signal control that considers the pedestrian push-button signal in addition to the vehicle signal.

BH5068-A

Prediction of Post-Radiotherapy Survival in Lung Cancer Using Dual-Omics Features and Machine Learning with Platform Development

Yi-An Wu[1], Chen-Cheng Wang[1], **Chih-Yu Cheng**[1], Chiu-Feng Chiu[1], Hsi-Yu Niu[1], Pei-Ju Chao[1,2], Tsair-Fwu Lee[1]
1: National Kaohsiung University of Science and Technology, Taiwan
2: Kaohsiung Chang Gung Memorial Hospital, Taiwan

Abstract—

Purpose:

This study aimed to improve survival prediction for patients with non-small cell lung cancer (NSCLC) after radiotherapy by integrating dual-omics features with machine learning models and developing a clinically applicable prediction framework.

Materials and Methods:

Clinical data, radiotherapy dose information, and DICOM images were retrospectively collected from 93 NSCLC patients treated with volumetric modulated arc therapy. Radiomics and dosiomics features were extracted from CT images and dose distributions. A two-stage feature selection approach using Cox proportional hazards analysis and least absolute shrinkage and selection operator (LASSO) was applied to generate multiple feature combinations. Survival prediction models were developed using LASSO-Cox proportional hazards regression and the DeepSurv deep learning model. Model performance was assessed using the concordance index (C-index).

Results:

Models based on single-domain features demonstrated limited predictive performance. In contrast, dual-omics feature combinations significantly improved survival prediction. The LASSO-Cox model achieved a C-index of 0.91, while the DeepSurv model reached a C-index of 0.89.

Conclusion:

Integrating radiomics and dosiomics features substantially enhances survival prediction accuracy in NSCLC patients following radiotherapy, supporting the potential of dual-omics–based models for personalized treatment planning and clinical decision-making.

Session 10 (16:30-17:50)

Topic: Key Technologies in Computer Vision and Image Processing
Session Chair: Prof. Weichao Wang, University of North Carolina at Charlotte, USA
Time: 16:30-17:50, March 18, 2026
Venue: Global Front, Room: 2rd Floor, 4021

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH5003

Investigation of Driver Habituation in Virtual Enviroment with Functional Near-Infrared Spectroscopy

Radosław Piątek and Paweł Tomiło
Lublin University of Technology, Poland

Abstract—The study aimed to identify the phenomenon of habituation in drivers under simulated driving conditions. Functional Near-Infrared Spectroscopy (fNIRS) was used to measure prefrontal cortex activity during three consecutive drives along the same route in a driving simulator, performed by a group of students holding category B driving licenses. The analysis focuses on evaluating changes in neural responses across successive trials, testing the hypothesis that reaction stabilization occurs as a result of habituation.

BH5056

Research on the assessment of executive function based on recognition of hand movements during work

Hiraku Matsuda and Mutsuo Sano
Osaka Institute of Technology, Japan

Abstract—Assessing the executive function of individuals with higher brain dysfunction requires specialized knowledge. While some evaluation methods have attempted simplified assessments, evaluating executive function remains challenging. Therefore, this study proposes a system that evaluates executive function using segmentation and image recognition applied to wiping cleaning videos. This system takes videos as input and presents users with a numerical score reflecting their executive function evaluation,

aiming to facilitate improvement in executive function. Experiments confirmed the system's ability to perform accurate evaluations by running it on wiping cleaning videos with varied patterns.

BH5086

An Explainable Deep Learning for Rice Leaf Disease Classification Using Xception with Adaptive ArcMargin and Grad-CAM

Jiraporn Thomkaew, Pornprasert Thipsawet, Apichai Chanudom and Podjana Homhual
Rajamangala University of Technology Srivijaya, Thailand

Abstract—Rice is a major economic crop in Thailand, with a high export rate within the agricultural product group. Rice leaf disease is a significant factor affecting agrarian productivity. This paper proposes a model for classifying rice leaf diseases to support accurate, timely disease diagnosis. It uses the Xception model as its core and replaces the classification layer with ArcMargin, enabling adaptive margin adjustment for each class. And use Class Activation Mapping guided regularization to identify important disease lesion areas on rice leaves, helping the model classify each disease class more accurately. To increase both the accuracy and the ability to explain the decision results of the model. Experiments were conducted using the rice leaf dataset from www.kaggle.com, comprising four classes and 5,932 images. The evaluation methods were stratified, and the Group 5-Fold methods were used to reduce information leakage and to verify the ability to explain the results. Gradient-weighted Class Activation Mapping experimental results show that the proposed model achieves accuracy of 99.29%, macro-F1 of 99.29%, Cohen's kappa of 99.05%, and Matthews correlation coefficient of 99.06%, which are higher than the baseline Xception model that achieves accuracy of 98.16%, macro-F1 of 98.21%, kappa of 97.63%, and MCC of 97.67%. In addition, the proposed method is applied to other core models, including DenseNet121, MobileNetV2, VGG16, and ResNet50, achieving accuracies of 100%, 96.57%, 99.84%, and 99.54%, respectively. The results demonstrate that the proposed approach effectively enhances both the classification performance and the visual explainability of deep learning models for rice leaf disease diagnosis.

BH5097

Modified DenseNet121 with Entry-Flow and Swish Activation for Grape Leaf Disease Classification

Podjana Homhual, Jiraporn Thomkaew and **Apichai Chanudom**
Rajamangala University of Technology Srivijaya, Thailand

Abstract—Grapevine leaf diseases pose a persistent challenge to agricultural productivity, affecting both yield quality and crop management efficiency. Traditional diagnosis relies heavily on expert assessment and often lacks consistency and accuracy. This study introduces a Modified DenseNet121 architecture that incorporates an Xception-based Entry-Flow module and the Swish activation function to improve feature extraction and gradient propagation. The Entry-Flow module replaces Dense Block 1 to reduce early-stage computational complexity via depthwise separable convolutions. At the same time, the Swish activation substitutes ReLU across all layers to enhance nonlinear representation learning. The model was evaluated using 5-fold cross-validation on a Kaggle dataset containing 9,027 images across four classes: Black Rot, ESCA, Leaf Blight, and Healthy. The proposed model achieved an average accuracy of 99.74%, with precision, recall, and F1-score comparable to or exceeding those of established baselines, including ResNet50, MobileNetV2, Xception, and the original DenseNet121. Notably, the model contains only 6.7M parameters, substantially fewer than ResNet50 and Xception. These findings suggest that the modified architecture offers a compact yet highly accurate solution, with strong potential for deployment in automated plant disease diagnosis systems and advanced agricultural applications.

BH5073

HQM-Net: A Hybrid Quantum-Classical Medical Imaging Architecture for Multi-Disease Classification across Heterogeneous Modalities

Quang Nhan Hoang, Minh Tri Pham and Duc Ngoc Minh Dang
FPT University, Viet Nam

Abstract—This paper proposes HQM-Net, a lightweight quantum-classical hybrid architecture, for the classification of multi-disease medical images across diverse imaging modalities, particularly under limited data and computational resource constraints. Specifically, this architecture combines the IRSE backbone, a lightweight convolutional neural network optimized for extracting and compressing biomedical image features, with a Quantum Neural Network that functions as a nonlinear enhancement layer. As a result, the Quantum Neural Network expands the representation space, improves class separation, and minimizes the number of parameters. In this process, classical features are encoded into a parameterized quantum circuit, measured, and then mapped back to the classical domain for final classification. Notably, experimental evaluations on six publicly available medical image datasets spanning various imaging modalities indicate that HQM-Net consistently outperforms standard lightweight convolutional neural network models, despite using only 0.35 million parameters. Taken together, these results highlight the potential of quantum-classical hybrid models to enhance generalization and computational efficiency in biomedical image analysis.

BH5001

RINKA: Efficient Artificial Neural Network Model for Drone-Based Object Detection

Paweł Tomiło

Lublin University of Technology, Poland

Abstract— With the rapid development of unmanned aerial vehicle (UAV) technology, drone-based aerial imaging and remote sensing have gained importance as effective tools in civil and military applications. Despite their many advantages, object detection in images acquired from UAVs remains a challenge due to scale variability, complex backgrounds, low object resolution, and varying environmental conditions. In response to these difficulties, this paper proposes a new object detection model, RINKA (Repeated efficient layer aggregation network, INvolution, Kolmogorov-Arnold), which combines the Involution mechanism with an architecture based on the Kolmogorov-Arnold network. This model was designed with adaptability to local image features and high computational efficiency in mind. As part of the experiments, the RINKA model was compared with modern detection architectures from the YOLO family (YOLOv8. YOLOv9, YOLOv10, YOLOv11). The results showed that RINKA achieved the highest evaluation metrics, confirming its effectiveness in diverse conditions.

Session 11 (16:30-17:50)

Topic: Human-Computer Interaction Design and Digital Multimedia Applications

Session Chair: Prof. Chang-Hwan Im, Hanyang University, Republic of Korea

Time: 16:30-17:50, March 18, 2026

Venue: Global Front, Room: 3rd Floor, 403D

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH5099

Predicting Gender and Understanding Cross-Cultural Gesture Differences When Using Touchscreens: Preliminary Study

Haseeb Tariq, Martin Gonzalez-Rodriguez, **Daniel Fernandez-Lanvin** and Javier de Andres

University of Oviedo, Spain

Abstract—This study provides a machine learning-based mechanism to determine gender by analyzing cultural differences through swipe gestures collected from touchscreen devices. A dataset including multiple gesture types was subjected to multi-view learning for identifying the optimum gesture combinations for gender classification. According to our knowledge, this is one of the earliest attempts for gender recognition through mobile device gestures while also accounting for cultural differences. With 76% accuracy on swipe patterns, Support Vector Classifier performed better than the other classifiers, whereas Extreme Gradient Boosting only managed 57%. By representing touchscreen interactions as swipe gestures, the study obtained reliable and validated behavioral aspects related to gender. Based on our analysis of swipe interactions and cultural context, the experimental analysis revealed that neither gender nor cultural background significantly effects on how users interact with touchscreen devices. However, the swipe gesture models were trained to make precise predictions of gender from touch-based user behavior, emphasizing the ability of gesture-based behavioral modeling may capture valuable insights from users, regardless of the lack of cultural effects. These types of models have a wide range of practical uses, including secure authentication, customized services, and targeted advertising.

BH5101

SleePSyn : Synthetic Data Strategies for Sleep Prediction from Smartphone and Wearable Signals

Yonchanok khaokaew
King Mongkut's University of Technology North Bangkok, Thailand

Abstract—Sleep quality and duration are critical indicators of both physical and mental health, yet accurate prediction is hindered by the scarcity of data on poor or short sleep patterns. This research aims to systematically compare and evaluate multiple synthetic data generation strategies for predicting daily sleep quality using data collected from wearable devices and smartphones. We propose SleePSyn, a framework that benchmarks traditional oversampling (SMOTE/ADASYN) against modern generative methods, including Generative Adversarial Networks (GANs), Diffusion Models, and large language models (LLMs). Using the GLOBEM dataset, we simulate varying degrees of data imbalance and evaluate the downstream performance of predictive models. Our results reveal a task-dependent tradeoff: while LLMs significantly improve the prediction of complex metrics such as sleep efficiency, traditional methods remain superior for sleep duration, providing a nuanced foundation for robust digital health systems.

BH5045

Advancing Quality Education through Text-to-Image AI: Enhancing Visual Learning and Cognitive Engagement

Ronel Francisco Ramos, Alfredo Lucion Calimbo, Juan Paulo Hilario Magcuyao and Roman Madrid De Angel
FEU Institute of Technology, Philippines

Abstract— Artificial Intelligence (AI) is reshaping educational practices by introducing innovative tools that enhance teaching and learning. Among these, text-to-image AI systems offer dynamic visualizations that transform abstract concepts into accessible, engaging learning materials. This study investigates the pedagogical potential of text-to-image AI in fostering cognitive engagement, inclusivity, and equitable access to quality education, in alignment with Sustainable Development Goal 4 (SDG 4). Through experimental classroom integration, the research examines how AI-generated visuals influence comprehension, retention, and learner motivation compared to traditional static resources. The study also explores challenges such as accuracy, cultural sensitivity, and ethical considerations in deploying generative AI for education. Findings aim to provide a framework for educators and policymakers to responsibly integrate text-to-image AI tools, thereby advancing inclusive, equitable, and high-quality learning experiences for

diverse student populations.

BH5075

Data-Driven Optimization of Multimedia Learning Interaction Using the Taguchi Method

Li-yu Tseng[1,2]

1: College of Design, National Taipei University of Technology, Taiwan

2: China University of Science and Technology, Taiwan

Abstract—Optimizing multimedia learning materials requires balancing cognitive efficiency and user engagement in human–computer interaction (HCI) contexts. Existing approaches such as heuristic adjustment, A/B testing, and machine-learning-based personalization offer partial solutions but often require extensive data, high development cost, or provide limited interpretability for design decisions. This study proposes a data-informed usability optimization framework based on the Taguchi Method to systematically evaluate multimedia learning interaction configurations. Using an L18 mixed-level orthogonal experimental design, seven multimedia factors—sound/narration, font type, image presentation, background music, animation resolution, presentation speed, and font size—are systematically varied to construct 18 interaction configurations. A total of 35 participants (N = 35) were recruited to rate overall perceived effectiveness of multimedia learning interaction for each configuration based on structured descriptions and visual representations of design elements. Signal-to-noise (S/N) ratio analysis is applied to identify a robust multimedia interaction configuration. Results indicate that the optimized configuration achieves the highest mean perceived effectiveness with reduced variability. The findings provide interpretable, design-level guidance for early-stage usability-oriented optimization of multimedia learning interaction. An independent validation study (N = 84) further supported the stability of the optimized configuration.

BH5027

Affective Computing for Student Motivation: Sentiment Analysis of Classroom Feedback

Ronel F. Ramos, **Alfredo L. Calimbo**, Roman M. De Angel and Juan Paulo H. Magcuyao

FEU Insitute of Technology, Philippines

Abstract—This study investigates the role of affective computing in enhancing student motivation through sentiment analysis of classroom feedback. Traditional evaluation methods often fail to capture emotional dimensions of learning, limiting educators' ability to adapt strategies in real time. By applying lightweight machine learning models implemented in TensorFlow.js, this research classifies student comments into positive, neutral, and negative sentiments, providing actionable insights into motivational trends.

Validation confirmed strong performance (accuracy 87%, precision 85%, recall 83%, F1-score 84), supported by expert review and cross-validation. Findings demonstrate that lightweight, browser-based sentiment analysis can serve as a practical tool for fostering positive learning environments, improving student retention, and integrating affective computing into human-computer interaction frameworks for education. This work contributes to the discourse on machine learning and HCI by showcasing how emotional data can inform adaptive pedagogy and support student-centered learning.

Session 12 (Online)

March 19, 2026, Thursday (11:15-12:35, GMT+9)

Topic: Multimodal Sensing and Understanding for Agricultural and Biomedical Applications

Session Chair: Assoc. Prof. Md Liakat Ali, Rider University, USA

Time: 11:15-12:35, March 19, 2026

Zoom ID: 841 3471 6436

Zoom Link: https://us02web.zoom.us/j/84134716436

***Presenters are recommended to enter the meeting room 10 mins in advance.**

***Presenters are recommended to stay for the whole session in case of any absence.**

***After the session, there will be a group photo for all presenters.**

BH5093

A Comparative Study of Machine Learning Algorithms for Peach Leaf Segmentation and Physical Characteristic Estimation

Haixin Wang[1] and Chunxian Chen[2]

1: Fort Valley State University, USA

2: USDA Agricultural Research Service, USA

Abstract—Peach leaf image analysis provides an efficient tool for modern orchard management. This study aims to estimate physical characteristics, specifically length, width, area, perimeter, and color ratios. To achieve accurate trait extraction, we evaluated and compared machine learning algorithms for image segmentation, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Logistic Regression, Naive Bayes, Discriminant Analysis (DA), Ensemble Bagged Trees, Neural Networks, and K-means Clustering. Using a custom metric function, Discriminant Analysis was identified as the optimal segmentation method with a mean performance score of 0.957. The DA-based model was integrated into a two-branch analysis pipeline. For healthy leaves, geometric characteristics were computed. For unhealthy leaves, a complementary color analysis partitioned the segmented leaf into green, yellow, and other regions and their area ratios were used as quantitative indicators of disease severity. Experimental examples demonstrated effective background removal and trait extraction, supporting low-cost and interpretable monitoring of peach leaf status.

BH5107

CropViT: Climate-Aware Vision Transformer Model for Crop Yield Prediction

Srihith Chennareddy[1], Andrew Ma[2] and Yen-Chu Weng[3]
1: Bellevue High School, USA
2: Tesla STEM High School, USA
3: University of Washington, USA

Abstract—Early-stage yield prediction remains a challenge in agricultural forecasting. Although deep learning has enhanced accuracy, most existing methods rely on late-season observations or handcrafted vegetation indices. We propose CropViT, a Vision Transformer-based model that integrates climate awareness with Sentinel-2 imagery to improve the accuracy of early-season crop yield predictions. Trained on the CropNet dataset, our model extracts spatially detailed visual features directly from raw observations. Our analysis demonstrates that CropViT outperforms standard CNNs and index-based models across winter wheat, soybean, and corn. For corn, the model achieved RMSE of 11.5 bu/ac and R2 of 0.81. Our results show that CropViT generalizes effectively across different crop types and climatic conditions, enabling reliable yield estimates months before harvest and offering a foundation for scalable and reliable agricultural monitoring systems.

BH5105

A PRISMA-ScR Scoping Review of Software Sustainability Criteria for Non-Functional Requirements

Bahaa Jouda, **Issa Atoum** and Noor Owies
Philadelphia University, Jordan

Abstract—Software sustainability is increasingly regarded as a software quality issue, however, its implementation at the requirements level remains disjointed and inconsistently documented. This paper provides a PRISMA-ScR scoping review that compiles sustainability criteria from contemporary software engineering literature and categorizes them into a requirementsfocused taxonomy. We implemented the scoping in studies published between 2020 and 2025. We found 19 primary papers that covered conceptual frameworks, empirical studies, and technical specifications. We identified 22 sustainability criteria and organized them into nine eco-technical categories: energy and lifecycle impact, resource optimization, sustainable architecture, sustainability metrics, standards compliance, contextual deployment, system boundaries and transparency, user independence, and adaptive sustainability strategy. The results indicate that barriers to adoption remain, including inconsistent terminology, poor traceability from criteria to

requirements artifacts, and evaluation methods that are predominantly late-stage and ad hoc. The suggested taxonomy helps studies report their findings more consistently and provides a practical foundation for tool-assisted requirements analysis and reusable sustainability NFR patterns.

BH0066

Single Channel EEG Based Sleep Stage Classification Using Machine

Md Liakat Ali[1], Shrivishnu Mukundhan[1], John Perrott[1], Logan Carroll[1], Kutub Thakur[2]
1: Rider University, United States
2: College of Professional Studies, St. John's University, United States

Abstract- Sleep deprivation is an ongoing and serious problem that has a significant impact on both physical and mental health and can lead to severe sleep disorders. Obtaining high-quality sleep is critical for proper brain function, mood regulation, bone and muscle health, and immune system performance. While sleep disorders cannot be completely prevented, accurately classifying different sleep stages is an effective way to diagnose sleep disorders and identify early signs of their occurrence. Manual sleep stage classification by sleep specialists is a tedious process, as it requires continuous monitoring and is prone to human error. Automating this process not only reduces the workload for clinicians but also improves consistency and accuracy. The purpose of this project is to evaluate different machine learning algorithms to improve sleep stage classification using EEG signals. For our analysis, we used the ISRUC dataset, which contains sleep data from both healthy and unhealthy subjects. The following machine learning algorithms were applied: Random Forest, Gradient Boosting, Extreme Gradient Boosting, Support Vector Machine, and Quadratic Discriminant Analysis. We evaluated these models using data from five healthy subjects and five subjects with sleep disorders. Based on accuracy, precision, recall, and F1-score metrics, Quadratic Discriminant Analysis achieved the best performance, with an accuracy of 99.8% for both healthy and unhealthy subjects. While further research is needed, the results of this study demonstrate that machine learning models can accurately classify sleep stages and have strong potential for supporting automated sleep analysis and clinical diagnosis.

BH0011

Windowed Point Cloud Manifold Harmonic Transform

Jiamian Li and Bing-Zhao Li
Beijing Institute of Technology, China

Abstract- This paper proposes a novel windowed point cloud manifold harmonic

transform (WPMHT), providing a theoretical framework for local spectral analysis of signals on point cloud manifolds. This method extends translation and modulation operators from classical time-frequency analysis to point cloud manifolds, achieving local spectral analysis on irregular point cloud structures by convolving a local window function with the modulated manifold harmonic basis function. WPMHT effectively reveals the joint point-frequency distribution characteristics of point cloud signals and captures their multi-scale geometric features. Experimental results show that this method can effectively perform pointfrequency analysis on various point clouds, ranging from simple spheres and tori to complex real-world objects, providing a powerful new tool for geometric signal processing.

BH5033

Machine Learning Model for Early Detection of Autism Spectrum Disorder (ASD) during Childhood

Paolo A. Loro Ramírez, Angelo Chipulina Meza, Pedro Castañeda, Juan Mansilla-Lopez and Alejandra Oñate-Andino

Abstract—Early detection of Autism Spectrum Disorder (ASD) continues to be a challenge due to the different types of manifestations and the limited availability of specialized clinical evaluations. This paper presents the development of a Machine Learning model aimed at predicting the presence and level of ASD in children aged 6 to 9 years, based on information from behavioral assessment instruments completed by parents. The model was trained with a clinically validated dataset and evaluated using multiclass metrics, seeking to maximize metrics such as accuracy, sensitivity, and specificity. The results show an optimal performance, reaching an accuracy of 86.65%, weighted accuracy of 86.34%, recall of 86.65%, F1-score of 86.00%, and a weighted AUC-ROC of 97.30%, which demonstrates a great ability to identify patterns associated with ASD, offering a support resource for professional evaluation. This proposal does not replace medical diagnosis, but it provides a scientific tool to strengthen early detection processes.

Session 13 (Online)

March 19, 2026, Thursday (15:35-17:35, GMT+9)

Topic: Generative AI and Multimodal Sensing
Time: 15:35-17:35, March 19, 2026
Zoom ID: 841 3471 6436
Zoom Link: https://us02web.zoom.us/j/84134716436

***Presenters are recommended to enter the meeting room 10 mins in advance.**
***Presenters are recommended to stay for the whole session in case of any absence.**
***After the session, there will be a group photo for all presenters.**

BH5078

NutriGenRec: A Generative AI Framework for Personalized Meal Recommendation System

Zheng-Bin Ter, Su-Cheng Haw, Palanichamy Naveen and Kok-Why Ng
Multimedia University, Malaysia

Abstract—Personalized meal recommendation systems support healthy dietary choices by considering individual nutritional needs and preferences. However, the current approaches of content-based filtering and collaborative filtering models do not work efficiently to solve the problem of multiple nutritional constraints. This paper proposes NutriGenRec, an approach that combine content-based recommendation with Variational AutoEncoder (VAE) and similarity-based ranking. VAE provides learning of latent nutritional patterns and generating realistic meal representations whereas similarity ranking ensures alignment of recommendations with existing meal plans. Besides this, a Large Language Model (LLM) has been implemented for managing multiple nutritional constraints. Experimental results suggest that compared to the baseline approaches, NutriGenRec improves the correctness of recommendations as well as constraint satisfaction.

BH5014

Food Loss Forecasting with Recurrent Neural Networks Using Gated Recurrent Units

Ru Poh Tan[1], Siew Mooi Lim[1], Kuan Yew Leong[2], Wei Zheng Goh[1], Jun Kit Chaw[3], Siaw Hong Liew[4] and Shee Chia Lee[1]
1: Tunku Abdul Rahman University of Management and Technology, Malaysia
2: A.I. System Research Co., Ltd., Japan
3: Universiti Kebangsaan Malaysia, Bangi, Malaysia
4: Universiti Malaysia Sarawak, Malaysia

Abstract—Food loss forecasting is critical for global food security and sustainability. This study compares Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) architectures for predicting post-harvest food loss using Malaysian agricultural time-series data. Results demonstrate that GRU outperforms LSTM across all metrics, achieving 10.7% lower RMSE, 5.0% lower MAE, and 61.8% reduction in MAPE. GRU also exhibits superior computational efficiency with 23.9% fewer parameters and 19.5% faster inference. We release a reproducible setup that can be directly extended with Attention-GRU and seasonal-trend decomposition GRU (STL-GRU) to better capture long-range and seasonal dependencies. This work provides a practical foundation for next-generation models aimed at reducing losses and strengthening resilience in agri-food systems.

BH5015

Robust Real-Time Facial Emotion Recognition via Adversarial Feature Learning and GAN-based Domain Adaptation

Kevin Lee[1], Siew Mooi Lim[1], Chen Zhun Lee[1] and Kuan Yew Leong[2]
1: Tunku Abdul Rahman University of Management and Technology, Malaysia
2: AI System Research Co. Ltd., Japan

Abstract—The promise of infinite synthetic data for Facial Emotion Recognition (FER) comes with a critical, under-examined trade-off. While Generative Adversarial Networks (GANs) can brilliantly bridge the synthetic-to-real domain gap, we show they can also amplify dataset biases and inadvertently destroy the very emotional cues they are meant to capture. This paper provides a crucial case study on this "plug-and-play" data augmentation strategy. We use a standard CycleGAN to adapt synthetic faces for training a lightweight mini-XCEPTION classifier. Our quantitative evaluation confirms a stark performance paradox: the F1-score for the rare 'Disgust' class doubles from 0.44 to 0.78, yet performance on 'Angry' and 'Sad' plummets as the model misclassifies them as 'Neutral' over 40% of the time. We attribute this to the GAN amplifying the target dataset's majority-class bias. These findings serve as a data-driven cautionary tale, highlighting the urgent need for bias-mitigation techniques to ensure that the quest for photorealism does not come at the cost of emotional sensitivity and fairness in human-computer interaction.

BH5089

A Multidimensional Study of Factors Influencing Behavioral Intention to Adopt VR-Based Cognitive Rehabilitation Therapy (VR-CRT)

Joren Dela Cruz, Ezekiel Bernardo and Jazmin Tangsoc
De La Salle University, Philippines

Abstract—Virtual reality (VR) has emerged as a promising approach for cognitive rehabilitation due to its ability to deliver immersive, engaging, and personalized therapeutic experiences. While prior studies have demonstrated the clinical effectiveness of VR-based cognitive rehabilitation therapy (VR-CRT), limited research has examined the factors influencing users' behavioral intention to adopt such systems from a multidimensional perspective. This study proposes and validates the Three-Dimensional VR Rehabilitation Adoption Model (3D-VRAM), which conceptualizes adoption intention across technological, personal, and utility dimensions. Data was collected through an online survey of 216 adults and analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM). Results indicate that perceived realism, immersion, cognitive support, physical risk, curiosity, and enjoyment significantly influence behavioral intention, with perceived realism emerging as the strongest predictor. Second order analysis further reveals that the technological dimension exerts the greatest influence on adoption intention, followed by the utility dimension. The personal dimension, although significant, plays a comparatively weaker role. These findings highlight adoption as an evaluative process driven primarily by system credibility and perceived cognitive value rather than intrinsic motivation alone. System developers and healthcare providers should also prioritize the design of immersive and realistic environments that convincingly reflect real-life contexts to strengthen adoption intention.

BH5091

Iterative Self-Correction for Accessible Text Generation: An Agentic Approach for Supporting Inclusive Human-Computer Interaction

Eduard Berzeminskij, Markus Lange-Hegermann and Carsten Röcker
TH OWL University of Applied Sciences and Arts, Germany

Abstract—Accessible information is a legal and economic prerequisite for social inclusion. Leichte Sprache (German Easy Language) addresses this by defining strict linguistic constraints, such as the prohibition of passive voice and rigid limits on sentence complexity, to ensure comprehensibility for individuals with cognitive impairments. However, standard Large Language Models (LLMs) struggle to adhere to these rules, often prioritizing fluency over formal compliance. This paper proposes an Agentic Framework for rule-compliant text simplification. Drawing on Agentic Context Engineering (ACE), we implement a three-node architecture (Generator-Reflector-Curator) that utilizes iterative self-correction to enforce accessibility standards without the need for resource-intensive fine-tuning. Using a corpus of German cooking recipes as a proxy for complex instructional text, we demonstrate that this agentic approach noticeably simplifies text. Our results show that explicit reflection enables the system to reduce

structural violations inherent in the source material. These findings suggest that dynamic, step-wise reasoning is a viable and efficient method for highly constrained text generation tasks.

BH5055

Inferring Cognitive Strain from Eye-Tracking Data for Interactive Information Access using Augmented Reality

Ayesha Amin and Haiming Liu
University of Southampton, United Kingdom

Abstract—Traditional human-computer interaction methods, despite their effective design, often fall short in dynamic, realtime environments where users interact with intelligent systems. Many interactive systems remain largely static and do not adapt to users. They offer the same layout, speed, and tools to all users, regardless of who is using them. These limitations are specifically evident in immersive environments such as augmented realitybased search engines or interfaces where users must process complex information under pressure. This paper presents an exploratory, gaze-driven framework for characterizing cognitive strain during interaction using eye-tracking data. Raw binocular gaze logs were transformed into structured oculomotor features (fixation, saccade, pupil, and blink metrics) via a reproducible preprocessing pipeline. These features were integrated into a composite Gaze Fatigue Index (GFI) to capture temporal variations in cognitive strain. Four unsupervised clustering methods (K-Means, Gaussian Mixture Models, hierarchical clustering, and DBSCAN) were explored using internal validation metrics, with K-Means selected for primary analysis due to its stability and interpretability. The analysis identified four recurring gaze-based behavioral profiles: focused attention, high visual load, reduced arousal, and exploratory scanning. These profiles were interpreted using psychophysiological literature and supported by participant self-reports. Overall, this research can inform adaptive interfaces that respond intelligently to each user by reducing content density, adjusting pacing, or offering timely support before performance declines.

BH5054

The Effectiveness of Large Language Model Embeddings in Enhancing Document Clustering

Imed Keraghel and Mohamed Nadif
Université Paris Cité, France

Abstract—Document clustering is the unsupervised process of categorizing texts so that

those in the same group are more similar to each other than to texts in other groups. This task is highly dependent on learning to represent documents effectively. From the Bag-of-Words (BoW) model, there has been a shift towards more sophisticated contextual representations, particularly with the emergence of transformers. It is natural to question which of these embeddings is most suitable for unsupervised tasks such as clustering. Our research evaluates various representations in an unsupervised setting, including BoW, TF-IDF, Word2Vec, GloVe, BERT, JoSE, E5, MiniLM, INSTRUCTOR, and GPT. We use a variety of clustering methods, including the latest deep learningbased approaches, to ensure an equitable comparison. Through numerous experiments, our aim was to provide insight into the performance of these models in a clustering context.

BH0065

Thermal-Based Gait Anomaly Scoring for Concealed Firearm Detection using Unsupervised Deep Learning

Caldwell Wachira and Henry Muchiri
Strathmore University, Kenya

Abstract- Escalating firearm-related incidents exposes the limitations of contemporary security measures such as manual CCTV monitoring, pat-downs, and metal detectors, which are intrusive, labor-intensive, and susceptible to human oversight. Prior studies indicate that concealed firearm carry subtly alters an individual's natural gait, as the body unconsciously compensates for restricted movement and added weight, leading to measurable changes in motion biomechanics such as stride length and hip motion. While these disruptions are often imperceptible to the human eye, they can be identified through deep learning–based gait analysis. This study proposes a non-invasive deep learning framework that complements existing CCTV surveillance by detecting concealed firearm carry through gait anomalies captured in thermal video. The proposed pipeline enhances thermal imagery using Contrast Limited Adaptive Histogram Equalization (CLAHE), extracts human silhouettes, and aggregates gait cycles into Gait Energy Images (GEIs), which are processed by an autoencoder trained exclusively on normal gait patterns. Anomalous behavior is identified by jointly analyzing reconstruction error and latent-space Mahalanobis distance, enabling unsupervised detection without the need for extensive firearm-specific labeling. Experimental evaluation on 60 UCLM thermal gait sequences, comprising 32 concealed and 28 normal scenarios, yields an area under the ROC curve (AUC) of 0.881 at a 95% confidence interval, achieving 100% recall while maintaining operationally viable precision. These results demonstrate the potential of thermal gait analysis can deliver an appearance-agnostic, early-warning surveillance solution that reduces cognitive load on security personnel.

BH5006

Auto Vision: A YOLO – Based External Sedan Car Damage Detection and Cost Estimation System

Kent Gabrielle C. Cabiten, Jesse Benjamin C. Cerbo, Hosea James R. Zacarias and Patrick D. Cerna

Mapua Malayan Colleges Mindanao, Philippines

Abstract-This study presents the development and evaluation of Auto Vision: A YOLO-based External Sedan Damage Detection and Cost Estimation System, a smartphone application designed to automate the process of vehicle inspection. The application integrates the YOLOv8 convolutional neural network (CNN) to detect and localize exterior damages on sedan vehicles, while a Decision Tree model handles repair cost estimates based on the damaged components identified. This combination of advanced computer vision and machine learning techniques provides a practical method for estimating financial implications. To evaluate system performance, the model was tested using standard metrics including F1 confidence, precision, recall, mean average precision (mAP), and accuracy. Results indicated that the model achieved good performance under optimal conditions, reliably identifying and localizing damaged parts despite challenges in viewing angles and environmental variations. These findings demonstrate the potential of Auto Vision as a reliable and cost-effective tool for external sedan damage detection, offering significant benefits in efficiency, and accessibility for both users and automotive service providers.

Note

Call for Papers

ICCRT 2026

September 18-20, 2026 | Hiratsuka Japan Registration deadline: April 20, 2026

2026 4th International Conference on Control and Robot Technology (ICCRT). ICCRT 2026

Conference Website: https://www.iccrt.org/ e-mail: **iccrtconference@yeah.net**

Accepted and registered papers will be published in ***ICCRT 2026 international Conference Proceedings.***

Submission System: https://www.zmeeting.org/submission/iccrt2026

Keynote Speakers: Prof. Maria Pia Fanti, Polytechnic University of Bari, Italy (IEEE Fellow), Prof. Hisato KOBAYASHI, Emeritus Professor in Hosei University, Japan (IEEE Life Fellow)

Conference will be held in **Tokai University(Shonan Campus), Hiratsuka, Japan.**


SPML 2026

July 10-12, 2026 | Hangzhou, China Submission deadline: March 30, 2026

2026 9th International Conference on Signal Processing and Machine Learning (SPML 2026)

Conference Website: https://spml.net e-mail: **spml.contact@gmail.com**

The accepted and registered papers can be published in the ***conference proceedings to be archived into IEEE Xplore® Digital Library, indexed by Ei Compendex, Scopus, INSPEC, Thomson Reuters' Web of Science, and others.***

Submission System: **https://www.zmeeting.org/submission/SPML2026**

Keynote Speaker: Prof. Xudong Jiang, Nanyang Technological University, Singapore (IEEE Fellow), Prof. Dan Zeng , Shanghai University, China, Prof. Honggang ZHANG, Zhejiang University, China (IEEE Fellow, AAIA Fellow)

Conference will be held in **Hangzhou Dianzi University, Hangzhou, China.**


ICIVP 2026

November 18-20, 2026 | Kyoto, Japan **Submission deadline: May 30, 2026**

2026 2nd International Conference on Image and Video Processing (ICIVP 2026)

Conference Website: https://www.icivp.org/ e-mail: **icivpconference@163.com**

Accepted and registered papers will be published in ***ICIVP 2026 Conference Proceedings.***

Submission System: **https://www.zmeeting.org/submission/ICIVP2026**

Keynote Speakers: Prof. Jie Yang, Shanghai Jiao Tong University, China, Prof. Hiroshi Fujita, Gifu University, Japan

Conference will be held in **Kyoto Institute of Technology, Kyoto, Japan.**


AICCC 2026

December 18-21, 2026 | Tokyo, Japan **Submission deadline: July 5, 2026**

2026 9th Artificial Intelligence and Cloud Computing Conference (AICCC 2026)

Conference Website: https://www.aiccc.net/ e-mail: **aiccc.contact@gmail.com**

The accepted and registered papers can be published in the ***AICCC 2026 Conference Proceedings, which will be sent to be indexed by Ei Compendex and Scopus.***

Submission System: **https://www.zmeeting.org/submission/AICCC2026**

Keynote Speaker: Prof. Kenji Suzuki Institute of Integrated Research at Institute of Science Tokyo, Japan (the merger between Tokyo Medical and Dental University and Tokyo Institute of Technology), Prof. Gaurav Sharma (IEEE Fellow, SPIE fellow) University of Rochester, USA, Prof. Runhe Huang, Hosei University, Japan

Conference will be held in **The University of Electro-Communications, Tokyo, Japan.**