CONFERENCE PROGRAM

2025 10th International Conference on FRONTIERS OF SIGNAL PROCESSING

ICFSP IPAI 2025

The 2025 International Conference on IMAGE PROCESSING AND ARTIFICIAL INTELLIGENCE

September 10-12, 2025 | Paris, France

Table of Content

Welcome Message

We are pleased to welcome you to 2025 10th International Conference on Frontiers of Signal Processing (ICFSP 2025) and the 2025 International Conference on Image Processing and Artificial Intelligence (IPAI 2025), held in Paris, France, during September 10-12, 2025.

ICFSP 2025 and IPAI 2025 invite authors to submit papers on different aspects of signal processing, image processing and artificial intelligence. Key areas of interest include, but are not limited to, image analysis and computational models, digital image processing and multimedia applications, next generation communication systems and signal processing.

The conferences aim to provide an interactive platform for professionals to learn about cutting-edge academic and industrial trends, share the latest scientific research and technological achievements, and discuss innovative ideas and methods. The goal is to elevate academic research and industrial application levels, supporting global initiatives around technological transformation. The conference seeks to promote technology research, development, and application both locally and internationally.

We are profoundly grateful to everyone who has helped make this event possible, including the respected authors, the invited speakers, and the peer reviewers. Special thanks also go to the conference committees for their dedication throughout the planning and execution of the conference.

We hope all participants will benefit from this event. Your contributions are essential in advancing the frontiers of knowledge and technology.

Wishing you a successful and inspiring conference experience at ICFSP 2025 and IPAI 2025!

ICFSP 2025 & IPAI 2025
Conference Committee

Onsite Conference Information

Conference Venue

ECE - Paris Campus

Address: Eiffel 4, 43 Quai de Grenelle, 75015 Paris
LOCATE THE CAMPUS ON GOOGLE MAPS:
https://goo.gl/maps/1YU6QYgZSwvecUwm9

Auditorium Location
(ECE Building Eiffel 4):

OMNES Education
43 Quai de Grenelle,
75015 Paris

Metro Station:
Bir Hakeim

ECE Building Eiffel 1:
10 Rue Sextius
Michel, 75015 Paris
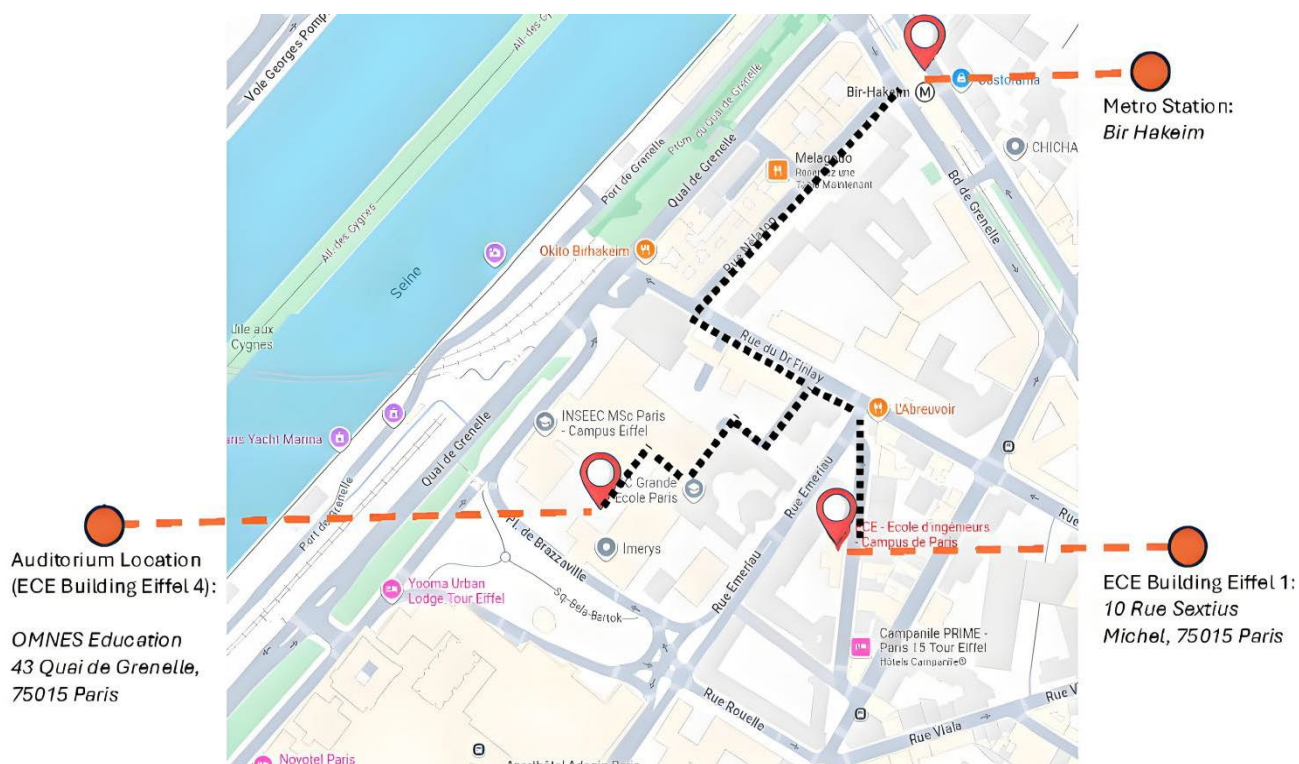
Oral Presentation Tips
✓ The duration of a presentation slot is 15 minutes. Please prepare your presentation for about12 minutes plus about 3 minutes for questions from the audience;
✓ An LCD projector & computer will be available in every session room for regular presentations;
✓ Presentations MUST be uploaded at the computer at least 15 minutes before the session start.

Dress Code
✓ All participants are kindly requested to dress formally, as casual wear is discouraged.
✓ National formal dress is welcome.

Attention Please
✓ Please ensure the safety of your belongings in public areas. For personal and property security, delegates are advised to wear their identification badges during the conference and refrain from lending them to unauthorized individuals. The conference cannot be held responsible for the loss of personal items.

Emergency Call: 112

Online Conference Information

Zoom information
Room A **ZOOM ID:88386795743** **ZOOM Link: https://us02web.zoom.us/j/88386795743** **Room B** **ZOOM ID: 85941126636** **ZOOM Link: https://us02web.zoom.us/j/85941126636** **Passcode: 091012**

Time Zone
✓ Paris Standard Time, UTC/GMT 2
✓ please make sure that both the clock and the time zone on your computer are set to the correct Paris standard time.

Sign in and Join
- ✓ Join a meeting without signing in: 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.
- ✓ Sign in with a Zoom account: All the functions are available.

Additional Suggestions
- ✓ A computer with an internet connection (wired connection recommended)
- ✓ USB plug-in headset with a microphone (recommended for optimal audio quality)
- ✓ Webcam (optional): built-in or USB plug-in
- ✓ Stable Internet Connection
- ✓ Quiet environment
- ✓ Proper lighting

Presentation Tips
Each presentation slot is 15 minutes. Please prepare to speak for around 12 minutes, allowing 3 minutes for audience questions.
Join the meeting room at least 15 minutes before the session begins.

September 10, 2025 | Wednesday

For Onsite Participants	
Venue: G007, Eiffel 4, 43 Quai de Grenelle, 75015 Paris	
14:00-17:00	Sign in and Collect Conference Materials

September 10, 2025 | Wednesday

For Online Participants	
ZOOM ID: 88386795743 **ZOOM Link: https://us02web.zoom.us/j/88386795743** **Passcode: 091012**	
10:00-12:00	ZOOM Test for Online (Presenters / Invited Speakers / Session Chairs)

Daily Schedule

September 11, 2025 | Thursday

Venue: Auditorium, Eiffel 4, 43 Quai de Grenelle, 75015 Paris

09:00-09:30	**Sign In & Morning Reception**
	Opening Address & Keynote Speech
09:30-09:35	**Conference Chair Opening Address** **Jacques Blanc-Talon**, Université Paris XI, France
09:35-10:15	*Keynote Speech I: Towards hyper-realistic and immersive visual communications* **Frederic Dufaux,** CNRS・Laboratoire des Signaux et Systèmes, France, IEEE Fellow
10:15-10:40	**Break Time & Group Photo**
10:40-11:20	*Keynote Speech II: Physics Informed Neural Network (PINN), New Bayesian Approach (BPINN), and the Applications* **Ali Mohammad-Djafari,** Former Research Director at CNRS, France Shaoxing University, China; IDT-EIT, China
11:20-11:40	*Invited Speech I: Bit Error Rate of MIMO with One Active Antenna Over Nakagami-m Fading* **Sergo Shavgulidze,** Georgian Technical University, Georgia
11:40-12:00	*Invited Speech II: Artificial Intelligence in Audio-Based Voice Pathology Diagnosis: Challenges and Advances* **Ioanna Miliaresi,** Ionian University, Greece
12:00-14:00	**Lunch	Cafeteria, Eiffel 1, 10 Rue Sextius Michel, 75015 Paris**

	Author Presentation Sessions	
	Venue: Eiffel 4, 43 Quai de Grenelle, 75015 Paris	
14:00-15:45	**Onsite Session 1** **Digital Image Analysis and Processing Technology** Session Chair: Jacques BlancTalon, Université Paris XI, France Ngoc Tam-Bui, Shibaura Institute of Technology, Japan FI105, FI2019, FI2017, FI2026, FI1012, FI2018, FI619-A	**G019**
14:00-15:45	**Onsite Session 2** **Multimodal Signal Detection, Estimation and Analysis Method** Session Chair: Sergo Shavgulidze, University of Georgia, Georgia FI106, FI2022, FI3036, FI3035, FI3031-A, FI4043, FI4042	**G024**
15:45-16:00	**Break Time (in front of the Auditorium)**	
16:00-17:30	**Onsite Session 3** **Intelligent Communication System and Key Technologies** Session Chair: José Antonio Díaz, PLOCAN, Spain FI103, FI2028, FI2014, FI2021, FI3033, FI2023	**G019**
16:00-17:15	**Onsite Session 4** **Applications of Artificial Intelligence in Information Systems** Session Chair: Ioanna Miliaresi, Ionian University, Greece FI2024, FI614, FI620, FI724, FI3030-A	**G024**
19:00-20:30	**Dinner	Cafeteria, Eiffel 1, 10 Rue Sextius Michel, 75015 Paris**	

Daily Schedule

September 12, 2025 | Friday

	Online Presentation Session	
10:00-11:45	**Online Session 1** **Image Analysis and Computing** Session Chair: Umesh Chandra Pati, National Institute of Technology, India Invited Speaker-Umesh C. Pati, Invited Speaker-Herminiño C. Lagunzad, Invited Speaker-San Gultekin, FI3029-A, FI506, FI4039	**Room A** **ZOOM ID:** **88386795743** **Passcode: 091012**
10:00-11:45	**Online Session 2** **Artificial Intelligence Theory and Innovative Applications** Session Chair: Ruksar Fatima, Khaja Bandanwaz University, India Invited Speaker-Ruksar Fatima, Invited Speaker-Warren B. Panizales, Invited Speaker-Bo Sun, FI617, FI826, FI616	**Room B** **ZOOM ID:** **85941126636** **Passcode: 091012**
12:00-13:00	**Break Time**	
13:00-14:40	**Online Session 3** **Advanced Information Technology and Data Management** Session Chair: Ronakkumar N Patel, Charotar University of Science and Technology, India Invited Speaker-Ronakkumar N Patel, Invited Speaker-Xinguo Yu, FI3037, FI4041, FI502, FI827	**Room A** **ZOOM ID:** **88386795743** **Passcode: 091012**

Keynote Speaker

Frederic Dufaux, CNRS · Laboratoire des Signaux et Systèmes, France
IEEE Fellow

Speech Time:09:35-10:15
Venue: Auditorium, Eiffel 4, 43 Quai de Grenelle, 75015 Paris

Speech Title: Towards hyper-realistic and immersive visual communications

Abstract: Nowadays, thanks to rapid technological progresses over the last decades, digital images and video sequences are ubiquitous, with many remarkable and successful applications and services. A key driver to research and development activities has been the objective to provide an ever-improving visual quality and user experience.

In this context, one of the next frontiers is to be able to faithfully represent the physical world and to deliver a perceptually hyperrealist and immersive visual experience. On the one hand, the human visual system is able to perceive a wide range of colors, luminous intensities, and depth, as present in a real scene. However, current traditional imaging technologies cannot capture nor reproduce such a rich visual information. On the other hand, immersive applications aim at giving to the user the sense of being present and immersed in one location or environment, without being physically there.

Recent research innovations have made it possible to address current bottlenecks in multimedia systems. As a result, new multimedia signal processing areas have emerged such as ultra-high definition, high dynamic range imaging, light fields, and point clouds. These technologies have the potential to bring a leap forward for upcoming multimedia systems. However, the effective deployment of hyper-realistic video technologies entails many technical and scientific challenges.

In this talk, I will discuss a few recent research activities related to hyper-realistic and immersive imaging. I will first consider point clouds, a very promising type of representation. One major distinguishing feature of point clouds is that, unlike images, they do not have a regular structure. Moreover, they can also be very sparse. For these reasons, point cloud processing presents significant challenges. Here, I will present recent learning-based approaches for point cloud compression and quality assessment. In a second phase, I will discuss high dynamic range imaging and in particular tone mapping operators (TMO). TMOs are used to compress the dynamic range with the aim of preserving the perceptual cues of the scene. Here, I will show how we can leverage semantic information as well as contextual cues from the scene to drive a TMO in a way similar to how expert photographers retouch images.

Bio: Dr. Frederic Dufaux is a CNRS Research Director at Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes (L2S, UMR 8506), where he is head of the Telecom and Networking research hub. He is a Fellow of IEEE. Frederic received the M.Sc. in physics and Ph.D. in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1990 and 1994 respectively. He has over 30 years of experience in research, previously holding positions at EPFL, Emitall Surveillance, Genimedia, Compaq, Digital Equipment, and MIT.

Frederic was Vice General Chair of ICIP 2014, General Chair of MMSP 2018, and Technical Program co-Chair of ICIP 2019 and ICIP 2021. He is Technical Program co-Chair of ICIP 2025 and MMSP 2025, and General Chair of ICME 2026. He served as Chair of the IEEE SPS Multimedia Signal Processing (MMSP) Technical Committee in 2018 and 2019. He was a member of the IEEE SPS Technical Directions Board from 2018 to 2021. He was Chair of the Steering Committee of ICME in 2022 and 2023. Since 2025, he is IEEE SPS Vice President Technical Directions, and member of the IEEE SPS Board of Governors and Executive Committee. He was also a founding member and the Chair of the EURASIP Technical Area Committee on Visual Information Processing from 2015 to 2021.

He was Editor-in-Chief of Signal Processing: Image Communication from 2010 until 2019. Since 2021, he is Specialty Chief Editor of the section on Image Processing in the journal Frontiers in Signal Processing.

In 2022, he received the EURASIP Meritorious Service Award, "for his leadership and contributions for the development of visual information processing within EURASIP".

Frederic is on the Executive Board of Systematic Paris-Region since 2019, a European competitiveness cluster which brings together and drives an ecosystem of excellence in digital technologies and DeepTech.

He has been involved in the standardization of digital video and imaging technologies for more than 15 years, participating both in the MPEG and JPEG committees. He was co-chairman of JPEG 2000 over wireless (JPWL) and co-chairman of JPSearch. He is the recipient of two ISO awards for these contributions.

His research interests include image and video coding, 3D video, high dynamic range imaging, visual quality assessment, video surveillance, privacy protection, image and video analysis, multimedia content search and retrieval, video transmission over wireless network. He is author or co-author of 3 books, more than 250 research publications (h-index=53, 11000+ citations) and more than 25 patents issued or pending. He is in the « World's Top 2% Scientists » list from Stanford University.

Keynote Speaker

Ali Mohammad-Djafari
Shaoxing University, China
IDT-EIT, China

Speech Time:10:40-11:20
Venue: Auditorium, Eiffel 4, 43 Quai de Grenelle, 75015 Paris

Speech Title: Physics Informed Neural Network (PINN), New Bayesian approach (BPINN), and the applications

Abstract: Inverse problems are present across scientific and engineering domains, where we seek to infer hidden parameters or fields from indirect, noisy observations. Classical methods, such as regularization and Bayesian inference, provide theoretical foundations for addressing ill-posedness, but face limitations in high-dimensional or computationally expensive problems.

Physics-Informed Neural Networks (PINNs) offer a promising data-driven approach by embedding physical laws within neural networks. In this Keynote talk, first the basics of PINN is introduced, then a new Bayesian approach (BPINN) is presented which extends the classical PINNs, accounting for training data modeling and measurement errors via priors and giving the possibility to quantify uncertainties in inverse problems.

To show the performances of the proposed approach, a few examples of applications, such as the inverse problems of infrared imaging, deconvolution and super-resolution, fusion of visual and infrared images are presented.

Bio: Ali Mohammad-Djafari (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Polytechnic of Tehran, Tehran, Iran, in 1975, the Diploma degree (M.Sc.) from École Supérieure d'Electricité (SUPELEC), Gif-sur-Yvette, France, in 1977, and the Docteur-Ingénieur (Ph.D.) and Doctorat d'État degrees in physics from the University of Paris Sud 11 (UPS), Orsay, France, in 1981 and 1987, respectively.,He supervised more than 22 Ph.D. students and has organized or co-organized more than ten international workshops and conferences. He has been an expert for a great number of French national and international projects. He published more than 300 papers in national and international journals, conferences, and workshops, and has also participated in and managed many industrial contracts with many French national industries, such as EDF, RENAULT, THALES, SAFRAN, and great research institutions, such as CEA, INSERM, INRIA, and the regional (Digiteo), national (ANR), and European projects (ERASYSBIO).,Dr. Mohammad-Djafari is a member of SIAM.

13

Invited Speaker

Sergo Shavgulidze,
Georgian Technical University, Georgia

Speech Time: 11:20-11:40
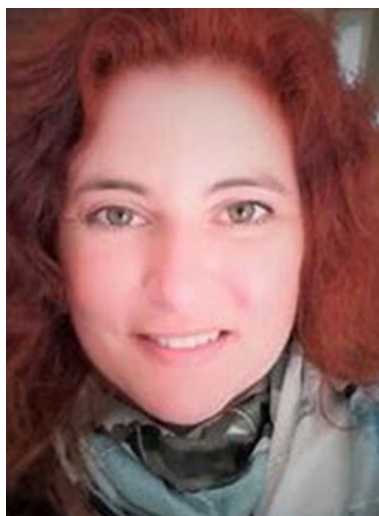**Venue: Auditorium, Eiffel 4, 43 Quai de Grenelle, 75015 Paris**

Speech Title: Bit Error Rate of MIMO with One Active Antenna Over Nakagami-m Fading

Abstarct: This article discusses the issue of constructing a new and straightforward multiple-input multiple-output (MIMO) system, in which only one of several transmitting antennas is active (MIMO-1 system). A regular method for constructing such systems is described and is based on forming so-called partner signals of different types, but of the same size and distance cha-racteristics for each antenna (from some base signal) by partitioning. For example, using the presented method, new MIMO-1 systems with spectral efficiency SE $\in$ {3, 4, 5, 6} bpcu (bits per channel use) are constructed; their error rate characteristics, which were obtained by computer simulation, are presented in the form of simulation curves. The Gaussian channel is represented by Nakagami-m fading. Four-dimensional hybrid signals with frequency-phase modulation were used as base signals. Comparison of the obtained results with other known results showed the high efficiency of the new MIMO-1 system at different fading depths m $\in$ {0.5, 0.6, 0.75, 1, 1.3, 2.5} with different spectral efficiencies.

Bio: Sergo Shavgulidze (Member, IEEE) received the Diploma with Excellence in Communication Engineering from the Georgian Technical University, Tbilisi, Georgia, in 1980, the Candidate of Technical Sciences degree from the Institute for Control Problems, Moscow, Russia, in 1984, and the Doctor of Technical Sciences degree from the Institute for Information Transmission Problems, Moscow, Russia, in 1991. He is currently a Professor at the Georgian Technical University and at the University of Georgia. He also serves on the Communications Commission of Georgia, where he is a member of the Advisory Board. From 2007 to 2008, he was an Associate Member of the Information and Communications Security Panel under the NATO Science for Peace and Security Program. He headed the Georgian delegation to RA-2007, WRC-2007, RA-2015, WRC-2015, RA-2019, and RA-2023, and was elected Vice-Chairman of ITU-R Study Group 5 (Terrestrial Services) at RA-2015 and RA-2019, as well as Vice-Chairman of ITU-R Study Group 7 (Science Services) at RA-2023.

His research interests include communication systems with a special emphasis on (generalized) concatenated codes, woven codes, coded continuous phase modulation, spatial modulation, and space-time coding, as well as flash memory endurance and data retention. He has authored over 190 publications on these subjects. On leave from the Georgian Technical University, he has held research positions at Linköping University and Lund University in Sweden; Darmstadt Technical University, Konstanz University of Applied Sciences, and Ulm University in Germany; the Technical University of Denmark in Lyngby, Denmark; and Lancaster University and HW Communications Ltd. in the UK.

Invited Speaker

Ioanna Miliaresi
Ionian University, Greece

Speech Time: 11:40-12:00
Venue: Auditorium, Eiffel 4, 43 Quai de Grenelle, 75015 Paris

Speech Title: Artificial Intelligence in Audio-Based Voice Pathology Diagnosis: Challenges and Advances

Abstract: Advances in artificial intelligence have opened new possibilities for detecting and classifying voice pathologies directly from audio signals, offering the potential for faster, non-invasive, and scalable diagnosis. This talk explores the integration of advanced machine learning techniques with audio signal processing to improve the accuracy, robustness, and generalizability of voice pathology classification systems. We will introduce methods for feature extraction, multimodal data fusion, and deep learning architectures capable of handling diverse datasets and variable-length recordings. Key challenges—such as limited data availability, variability in patient phonation, and the need for clinically interpretable models—will be addressed alongside recent breakthroughs, including the use of fully convolutional networks, data augmentation strategies, and the incorporation of complementary modalities like electroglottographic signals.

Bio: Dr. Ioanna Miliaresi studied Physics at the National and Kapodistrian University of Athens, where she also earned a Master of Science degree in Electronics and Telecommunications. She holds a Ph.D. from the University of Piraeus, specializing in digital audio processing.

Her professional background includes extensive experience in the telecommunications industry. From 1998 to 2004, she worked as a Senior System Test Engineer at Siemens AE in Greece, contributing to the development and testing of carrier-grade voice services over IP/ATM networks and intelligent network platforms. During this period, she also collaborated with Siemens AG in Munich, Germany, where she participated in verification activities for Voice over IP and Remote Access Server applications.

In academia, Dr. Miliaresi served as a Lecturer of Applications at the Department of Sound and Musical Instruments Technology at the Technological Educational Institute of Ionian Islands from 2003 to 2018. Since 2018, she has been a Lecturer of Applications in the Department of Audio and Visual Arts at the Ionian University, focusing on digital audio processing and sound technology.

Her research interests include artificial intelligence, digital signal processing, and complex systems modeling, with applications in audio analysis and healthcare diagnostics. She has authored scientific publications, participated in European Union–funded research projects, and serves as a reviewer for major international conferences such as the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), the IEEE Workshop on Machine Learning for Signal Processing (MLSP), and the International Joint Conference on Neural Networks (IJCNN). She is also actively involved in the organization of international scientific conferences, including the International Conference on Frontiers of Signal Processing (ICFSP) and the Digital Culture & Audiovisual Challenges (DCAC) conference series.

Onsite Session 1

Digital Image Analysis and Processing Technology

- ➢ **Session Chair:** Jacques BlancTalon, Université Paris XI, France
  Ngoc Tam-Bui, Shibaura Institute of Technology, Japan
- ➢ **Time:** 14:00-15:45, September 11, 2025
- ➢ **Room:** G019, Eiffel 4, 43 Quai de Grenelle, 75015 Paris
- ➢ **Papers:** FI105, FI2019, FI2017, FI2026, FI1012, FI2018, FI619-A

FI105 14:00-14:15	Real-World Logistics Object Detection with YoloV12-OBBBiFPNCBAM: A Lightweight and Accurate Approach **Author(s) :** Thien-Van Tran, Ngoc-Tam Bui, Phan Xuan Tan, Hiroshi Hasegawa **Presenter:** Thien-Van Tran, Shibaura Institute of Technology, Japan **Abstract:** With the rapid advancement of logistics, we propose a method for measuring object dimensions and detecting damaged parcels to improve operational quality. In this study, we build our model based on a dataset collected in a real-world logistics environment and address data scarcity by employing Stable Diffusion for data augmentation. To enhance accuracy while maintaining computational efficiency and inference speed, we optimize the standard YOLO architecture by integrating a BiFPN neck and CBAM attention mechanism. Our approach improves detection accuracy significantly—from 0.809 to 0.927. Furthermore, the proposed YOLOv12-BiFPN-CBAM model not only surpasses the current state-of-the-art YOLOv12-nano in accuracy (from 0.917 to 0.927) but also achieves this with only 65% of the parameters and FLOPs. These results demonstrate the model's strong potential for practical deployment in real logistics scenarios.
FI2019 14:15-14:30	Cross-Frame Semantic Alignment Network: Towards Zero-Shot Video Action Recognition **Author(s):** Yue Huang, Xiaodong Gu **Presenter:** Xiaodong Gu, Fudan University, China **Abstract:** Recent advances in vision-language models (VLMs) like CLIP have revolutionized zero-shot image understanding by aligning visual and textual semantics. However, extending this success to zero-shot video action recognition remains challenging due to the inherent temporal complexity of videos. Critical actions are often diluted by redundant frames, while subtle interframe variations are overlooked by global pooling-based methods. In this paper, we propose Cross-Frame Semantic Alignment Network (CFSAN), a novel framework that jointly models local fine-grained action dynamics and global semantic consistency. We develop a Local Adjacent Frame Semantic Alignment Module to capture inter-frame semantic changes by aggregating multi-scale semantic differences between adjacent frames. In addition, the Global Semantic Aware Spectral Cluster Module is designed to divide frames into semantic clusters, and a gate mechanism is employed to filter noise, which retains only action key segments, to reduce the dilution effect of irrelevant frames on the global pool. We conduct a zero-shot video action recognition evaluation on HMDB51 and UCF101 datasets and experimental results show that the CFSAN model achieves competitive results.
FI2017 14:30-14:45	Color Image Protection via Quantum Encryption and HVS-Based Mini-Block Data Hiding **Author(s):** Luis Angel Olvera-Martinez, Manuel Cedillo-Hernandez, Carlos Adolfo Diaz-Rodriguez, Leonardo Faustinos-Morales

	Presenter: Luis Angel Olvera-Martinez, Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME Culhuacan) del Instituto Politécnico Nacional, México **Abstract:** With the growing demand for secure and efficient digital image processing, the integration of data hiding and encryption techniques has become essential to ensure both confidentiality and integrity of visual information. This paper proposes a hybrid approach that combines a low complexity steganographic scheme for binary and color images with a computationally efficient encryption technique inspired by quantum operators. In the data hiding component, an RGB channel of a color image is selected, and its least significant bit plane is extracted. This bit plane is then divided into suitable blocks, and HVS-based encoding tables are used to embed the information within these blocks, achieving high embedding capacity while preserving visual quality. In parallel, the encryption stage operates on color images and is based on a novel operator model that simulates quantum behaviors such as creation, exchange, and annihilation, enabling effective protection without requiring intensive mathematical processing. Experimental validation confirms the robustness of the combined approach through quality metrics such as PSNR, SSIM and entropy.
FI2026 14:45-15:00	Hyperspectral Image-Based Feature Point Matching for High-Performance Novel View Synthesis **Author(s):** Tomohiro Yuguchi, Kaito Houjo, Ryota Fuse, Terumasa Aoki **Presenter:** Tomohiro Yuguchi, Tokyo University of Technology, Japan **Abstract:** In recent years, Novel View Synthesis (NVS) has paid attention since it can apply to filmmaking, gaming, and virtual reality etc. Thus, the performances of NVS strongly depend on the performance of feature point matching method. Under this situation, in this paper, we present a novel hyperspectral (HS) image-based feature point matching method to achieve high-performance NVS. In our method, a HS image is generated from an RGB image first. Next, feature points are extracted from each channel in a HS image. After that, all feature points from each channel are aggregated. In this process, similar feature points at the same position are removed. The number of feature points extracted this way is much greater than that of a gray-scale or a RGB image. Our method detects 4 to 6 times more feature point matches than traditional SIFT, enhancing SfM quality and significantly improving 3D Gaussian Splatting for NVS.
FI1012 15:00-15:15	Densely Connected Residual U-Net: Computed Tomography Reconstructed Image Enhancement and Artifact Removal **Author(s):** Rekha, Santosh Singh, S.K. Neogy **Presenter:** Rekha, Shiv Nadar Institution of Eminence, India **Abstract:** Historically, in the computed tomography (CT) image reconstruction technique, X-ray dose reduction has always been the central research focus with the best quality reconstructed images. Most of the efforts are centered on reducing the number of projection data required for reconstruction. Many algorithms have been proposed, and results have been compared with most of the commercially successful CT image reconstruction algorithms, i.e., the filter backprojection algorithm. Still, there is a gap in the image quality of all the algorithms in comparison to the filter backprojection algorithm. To overcome such a gap, another step as a part of postprocessing is performed to enhance the image quality and remove the artifacts in the reconstructed CT images. In this paper, we have proposed a U-Net-based model with an added local residual learning technique for image enhancement for few-view CT images. We have shown our enhancement results applied to the reconstructed images, and for reconstruction, the filtered backprojection algorithm is used with minimal possible projection data. We have quantified

	our results in comparison to ground-truth reconstructed images by using the peak signal-to-noise ratio (PSNR) and structural similarity index(SSIM). The results clearly show that our model is capable of capturing the essential features by enhancing the reconstructed image quality and reducing the artifacts.
FI2018 15:15-15:30	**Ownership Authentication and Clinical Data Backup in DICOM Images via Robust Watermarking and Binary Data Hiding** **Author(s):** Angela Axin Lopez-Martinez, Eduardo Fragoso-Navarro, Manuel Cedillo-Hernandez, Francisco Garcia-Ugalde **Presenter:** Angela Axin Lopez-Martinez, Instituto Politecnico Nacional (Esime Unidad Culhuacan), Mexico **Abstract:** Medical information is very important and delicate due to its relevance in clinical diagnosis, treatments, research and other commercial and non-commercial applications of governmental and private organizations. Recent advances in information and communication technologies have transformed the paradigm of medical data management, but also introduced new risks related to unauthorized access, manipulation, and distribution. To solve part of these problems, this paper proposes a method that employs two information security levels based on robust watermarking and binary data hiding, respectively. Firstly, clinical data is embedded into a logo using binary data hiding. Then, this logo is inserted into the spatial domain of a medical image through block segmentation, statistical information, and convolutional encoding for ownership authentication. Experimental results demonstrate logo robustness against JPEG DICOM compression in conjunction with aggressive attacks like cropping and copy-move, using bit error rate as a metric. The imperceptibility has been measured via peak signal-to-noise ratio and structural similarity index metrics.
FI619-A 15:30-15:45	Multiclass 3D U-Net for Tumor and Tissue Segmentation in PET/CT Imaging **Author(s):** Karla Michelle Martin-Sanchez, Francisco J Alvarez-Padilla, Diego Oliva, Teresa Romero-Gutiérrez, Rodolfo Aguayo- Jiménez, Gerardo Mendoza-Cuautla, Balam Octavio Huerta-Gomez **Presenter:** Karla Michelle Martin Sanchez, University of Guadalajara, Mexico. **Abstract:** Accurate segmentation of tumor lesions in PET/CT images remains a challenge due to anatomical variability, physiological uptake, and motion-related artifacts. Manual and semi-automatic approaches, though still widely used, are time-consuming and subject to inter- and intra-observer variability. Furthermore, many automatic methods oversimplify the task by reducing it to binary classification, overlooking the complexity of tumor morphology and physiological context. In this work, we propose an AI-based approach using a three-dimensional convolutional neural network (3D U-Net) trained on multimodal FDG PET/CT data. Our framework performs anatomically informed multiclass segmentation, leveraging the synergy between functional (PET) and anatomical (CT) imaging. The model incorporates sequential multiclass classification layers to differentiate among normal hypermetabolic tissue, excretory zones, and abnormal uptake regions (malignant or inflammatory). It was trained on 66 anonymized PET/CT studies, using data augmentation techniques to enhance generalization across varied anatomical sites and uptake patterns. Preliminary results demonstrate improved accuracy in delineating oncological lesions and their physiological context, suggesting the potential of this approach for clinical decision support.

Onsite Session 2

Multimodal Signal Detection, Estimation and Analysis Method

- ➢ **Session Chair:** Sergo Shavgulidze, University of Georgia, Georgia
- ➢ **Time:** 14:00-15:45, September 11, 2025
- ➢ **Room:** G024, Eiffel 4, 43 Quai de Grenelle, 75015 Paris
- ➢ **Papers:** FI106, FI2022, FI3036, FI3035, FI3031-A, FI4043, FI4042

FI106 14:00-14:15	Acoustic Analysis of the Underwater Baseline Sound Level at the PLOCAN Test Site **Author(s):** José Antonio Díaz, Silvana Neves and Eric Delory **Presenter:** José Antonio Díaz, PLOCAN, Spain **Abstract:** Anthropogenic Underwater Radiated Noise (URN) is an external agent that contributes to marine pollution and can affect marine ecosystems. It has the potential to alter the behaviour and distribution of marine species. In recent years, this topic has received more attention, and several projects began to study and analyze it. One of the key aspects in understanding the issue of Underwater Radiated Noise is the characterization of the baseline sound, this is, the ambient or background level in the absence of ships or any other anthropogenic sound source. Baseline sound would allow us to set a reference level with which other sounds can be compared. The aim of this study is to analyze one month of continuous underwater sound recordings at the PLOCAN test site to characterize its baseline sound level, in order to prepare the stage for a later comparison with different sources of anthropogenic noise, for example, the one produced by wind generators. These sound recordings were analyzed with a custom MATLAB based software and the following metrics were generated: SPL, Leq, L1, L5, L10, L50, L90, L95, and Lpeak. All these metrics were generated for the unfiltered signals, filtered signals from 20 Hz to 20 kHz, and one-third-octave band signals from 20 Hz to 20 kHz. The results were analyzed in numerical and graphical form, and a summary of the results that characterize the baseline sound of the PLOCAN test site were generated. Some of the findings of this study are the following: 1-Processing time is long for the dataset under analysis. It took 10 full days of non-stop processing to generate the above metrics, not including the post-processing time. 2- The amount of data generated was relatively large, requiring automation of the analysis process. 3- Some metrics like SPL and Leq tend to be better suited to characterize the baseline level. 4- A wide spread of values was found in the metrics for all the signals under analysis.
FI2022 14:15-14:30	Improving Efficiency of Hybrid Flexible Streaming Accelerator Systems via Memory and Interrupt Optimizations **Author(s):** Alexander Strachan, Nigel Topham **Presenter:** Alexander Strachan, University of Edinburgh, United Kingdom **Abstract:** Hybrid accelerator systems have the potential to increase the useful lifespan of hardware designs. These systems allow parts of a hardware accelerator to be replaced with software running on a general purpose CPU. This allows the design to be extended or fixed in the field, after the fact, compared to traditional fixed function hardware accelerators.However, when implemented naively, such systems have sig- nificant ove rheads due to cache and interrupt management, limited memory bandwidth, and single threaded operation. This paper details three mechanisms that can reduce these overheads, and evaluates their impact using full-system simulation of an ARM-based embedded multicore architecture, extended with a cycle-accurate hardware accelerator. We

	show that efficiency can be increased by 5-20%, leading to decreased power and energy consumption, extending the potential applications of such systems.
FI3036 14:30-14:45	What Your Brain Activity Says About You: EEG-based Subject Verification Across Sessions **Author(s):** Mahzad Gharleghi, Kay Fuhrmeister, Thomas Köllmer, Arne Pelzer, Joanna Scanlon, Insa Wolf, Julia Lechinger, Robert Göder **Presenter:** Mahzad Gharleghi, Fraunhofer IDMT, Ilmenau, Germany **Abstract:** Electroencephalography (EEG) signals contain rich person-specific information that can be used for biometric applications. This raises both opportunities and concerns in areas such as personalized health care and data privacy. This paper presents an approach for EEG based subject verification, where a transformer-based model learns to generate subject specific embeddings to later determine whether two EEG recordings belong to the same person or not. The proposed method generates subject-specific embeddings using pre-trained models and performs subject verification using a distance-based matching strategy with threshold tuning. To assess the models generalization, cross-session and cross-dataset experiments are conducted using two EEG datasets with different recording montages with one being made up of subjects with medical pathologies. Results indicate high performance even when trained and evaluated across different subject groups, and recording configurations, reaching accuracies above 80%. These findings show that identity-related information can be extracted from EEG signals independent of health status or recording protocols, while also highlighting the need for privacy preservation of EEG data to protect sensitive subject-specific information.
FI3035 14:45-15:00	A Robust Mutual Information Method for Time-Delay Estimation Based on Multiscale Feature Extraction **Author(s):** Hanlin Gao, Yong Wang, Xiaoyuan Li **Presenter:** Hanlin Gao, Northwestern Polytechnical University, China **Abstract:** Accurate parameter estimation from noisy, nonlinear time-series is a fundamental challenge in signal processing. Time-delay estimation, a critical step in techniques like time-delay embedding, is particularly susceptible to noise when using standard methods like mutual information (MI). This paper introduces a novel signal processing framework to overcome this limitation. We propose a multi-stage feature extraction front-end, termed GFMIEME, that transforms the raw signal into a more informative and noise-resilient feature space. This is achieved through a fine-grained multiscale analysis combined with the extraction of local statistical features (mean, standard deviation, root-mean-square). By computing MI in this enhanced feature domain, our method significantly boosts estimation accuracy. We benchmark our approach using chaotic signals, a canonical example of complex nonlinear data. Results show that our method successfully estimates the optimal time-delay at SNRs as low as -25 dB, a regime where traditional MI fails, demonstrating its superior robustness for practical signal analysis.
FI3031-A 15:00-15:15	An Early Predictive Model for Breast Cancer through Medical Record Signal Analysis Using Machine Learning and Digital Signal Processing **Author(s):** Mahgoub Hammad, Suliman Alrufai, Mozan Hammad **Presenter:** Mahgoub Hassan Hammad Mohamed, Neolt Institute, Ksa **Abstract:** Breast cancer remains one of the most pressing global health challenges affecting women, with statistics showing a steady increase in annual incidence rates. Early detection significantly enhances the chances of recovery and reduces mortality rates. This study proposes the development of an early predictive model that combines Digital Signal Processing (DSP) and Machine Learning (ML) to analyze signals extracted from medical

	records, including mammogram images, histopathology reports, and blood test data. Initially, these unstructured data are converted into analyzable signals using DSP tools. Subsequently, ML algorithms are applied to uncover patterns and predict the likelihood of cancer. The proposed model achieved a predictive accuracy exceeding 91% when tested on diverse clinical datasets. This paper recommends integrating the model as a clinical decision support tool to enhance diagnostic accuracy and minimize human error. Future work involves expanding the input data range to further improve model precision.
FI4043 15:15-15:30	C-Proper Centralized Fusion Estimation for Quaternion-Valued Signals under DoS Attacks and Noise Correlation **Author(s):** Felicidad Gutiérrez-Trujillo, José D. Jiménez-López, Rosa M. Fernández-Alcalá **Presenter:** Felicidad Gutiérrez Trujillo, University of Jaén, Spain **Abstract:** This paper addresses the problem of centralized fusion estimation for signals in the quaternion domain, based on observations from multiple sensors. It is assumed that the observations from each sensor may be affected by possible Denial-of-Service (DoS) attacks, which can block the transmission of sensor data. In such cases, to mitigate these threats, a compensation strategy is employed, where the missing observation is replaced by its predictor, based on the information available up to the previous time instant. Additionally, the correlation between the additive noise of the state and the observations at the same time instant is considered. To combine information from multiple sensors, a centralized fusion method is used. Under C-properness conditions, a recursive centralized fusion filtering algorithm is proposed to obtain the optimal estimator and corresponding error variance, reducing the problem dimension by half and thus significantly decreasing the computational load. Numerical simulation examples are presented to demonstrate the effectiveness and advantages of the proposed algorithm.
FI4042 15:30-15:45	Pitch Accent Pattern Transcription in Tokyo and Osaka Japanese Using Whisper-Derived Features **Author(s):** Hiroto Noguchi **Presenter:** Hiroto Noguchi, Institute of Education/Sophia University Institute of Science Tokyo/ Graduate School of Languages and Linguistics, Chiba/Tokyo, Japan **Abstract:** This paper introduces an automated method for Japanese pitch-accent transcription that combines Whisper-derived hidden-state embeddings with neural network classifiers. Three architectures—multilayer perceptron (MLP), convolutional neural network (CNN), and bidirectional long short-term memory (BiLSTM)—are evaluated on corpora of Tokyo and Osaka dialects under identical preprocessing and modeling protocols. In within-dialect experiments, the MLP attains the best mora-level accuracy for Tokyo, whereas the CNN performs best for Osaka. A cross-dialect transfer condition is also tested in which all Tokyo data are used for training and all Osaka data for evaluation with an unchanged pipeline. Under this setting, the best configuration reaches 69.13% mora accuracy, 0.678 macro-F1, and 22.82% word-level accuracy, with the other configurations slightly lower. These findings indicate that cross-dialect transfer remains challenging without explicit adaptation, motivating future work on simple, reproducible strategies for target-dialect robustness.

Onsite Session 3

Intelligent Communication System and Key Technologies

- **Session Chair:** José Antonio Díaz, PLOCAN, Spain
- **Time:** 16:00-17:30, September 11, 2025
- **Room:** G019, Eiffel 4, 43 Quai de Grenelle, 75015 Paris
- **Papers:** FI103, FI2028, FI2014, FI2021, FI3033, FI2023

FI103 16:00-16:15	Risk assessment in Wireless Network using Bayesian graphs **Author(s):** A.Adamova, T. Zhukabayeva, Y. Mardenov, N. Karabayev, D. Satybaldina **Presenter:** Aigul Adamova, Astana IT University, Kazakhstan **Abstract:** Nowadays, ensuring the security of wireless networks has become a serious problem because of the constant emergence of various cyber threats, which in turn require effective mechanisms for their detection and prevention. Bayesian graphs have shown notable advantages, particularly due to their capability to reflect complex interdependencies, adapt to dynamically changing conditions, and provide predictive insights for risk management. In this research, a Bayesian graph-based approach is proposed for systematically structuring and analyzing security states within wireless networks. The practical applicability of the suggested model was assessed through simulations involving common network attacks, specifically man-in-the-middle, denial of service, and node substitution scenarios. The obtained outcomes highlight the efficiency of Bayesian graphs in determining vulnerabilities and shaping proactive defensive measures. The presented method is particularly relevant in the context of smart city infrastructure, offering integration possibilities with other artificial intelligence tools, as well as potential applications across broader cybersecurity domains.
FI2028 16:15-16:30	Effect of Signal Quantization on Performance Measures of a 1st Order One-Dimensional Differential Microphone Array **Author(s):** Shweta Pal, Arun Kumar, Monika Agrawal **Presenter:** Shweta Pal, Indian Institute of Technology Delhi (IIT Delhi), India **Abstract:** In practical systems, recorded analog signals must be digitized for processing, introducing quantization as a critical aspect of data acquisition. While prior studies have examined quantization effects in various signal processing contexts, its impact on differential microphone arrays (DMAs), particularly in one-dimensional (1D) first-order configurations, remains unexplored. This paper investigates the influence of signal quantization on the performance of first-order 1D-DMAs across various beampatterns. An analytical expression for quantized beamformed output for a first-order 1D-DMA has been formulated. The effect of signal quantization has been studied on array performance measures such as the Beampattern, Directivity Factor (DF), Front-to-Back Ratio (FBR), and suppression depth at null points (SDN). Simulation results reveal that the beampattern shape remains structurally invariant across quantization bit depths, with quantization primarily affecting SDN. DF and FBR remain constant with the varying number of quantization bits. Additionally, SDN is shown to be frequency independent, however, it increases with increasing quantization bit depths, enhancing interference suppression. The study also examines the effect of steering nulls across the azimuthal range, showing that SDN degrades as the null moves closer to the source's look direction ($0°$), indicating reduced interference suppression.

FI2014 16:30-16:45	Fixed vs Mobile IRS Placements **Author(s):** Ndeye Penda Fall, Cherif Diallo, Michel Marot, Adel Mounir Said, Hossam Afifi, Hassine Moungla **Presenter:** Ndeye Penda Fall, Universté Gaston Berger, Sénégal **Abstract:** The Intelligent Reflecting Surface (IRS) is an enabling technology for beyond 5G and 6G. IRSs are passive reflective surfaces that can modify signals in a controlled manner. They are utilized to enhance coverage and spectrum with high energy efficiency and, therefore, are referred to as passive relays. This energy efficiency makes them a promising technology. The reflective nature of these surfaces allows for a line-of-sight path through the IRS between a transmitter and receiver when the direct path is obstructed with low-energy consumption. However, this requires careful positioning. Proper IRS positioning is crucial for optimizing network performance. Most studies consider a fixed placement of the IRS, but the IRS can also be mobile to better adapt to the dynamics of certain networks. In this paper, we propose to compare the fixed and mobile placements of several IRSs. The problem is modeled using an IRS activation strategy for each period, employing different comparison criteria.
FI2021 16:45-17:00	A Non-invasive Measurement System for Evaluating 3D Indoor Sound Source Localization Techniques **Author(s):** Laurens Sillekens, Oskar Rudolf, Martin Thißen, Ina Penner, Stefan Seyfarth, Elke Hergenröther, Jens-Peter Akelbein **Presenter:** Laurens Sillekens, University of Applied Sciences Darmstadt, Germany **Abstract:** Indoor sound source localization enables innovative and cost-effective applications, such as needs-based automation of building functions. This paper presents a measurement system for 3D positional sound source localization in the near-field using a single, distributed microphone array with non-invasive geometry. Two localization techniques are evaluated, both relying on Time-Difference-of-Arrival (TDOA) between microphone pairs, which are estimated using the well-known β-GCC-PHAT frequency weighting. The first algorithm is a closed-form analytical 3D Sound Source Localization (3D-SSL) solution, while the second employs a 3D grid-based Steered Response Power (SRP) method. Experimental results using both human speech and clapping signals in a highly reverberant environment show high localization accuracy, indicating strong potential for practical deployment. The 3D-SSL algorithm, when combined with stability and spatial filtering, achieves a localization error standard deviation of less than 30cm in both the x- and y-directions. The SRP algorithm exhibits lower localization accuracy, with errors below 70cm in the x- and y-directions, but offers greater robustness and does not require filtering. Both methods are particularly well-suited for 3D indoor applications that require high planar accuracy.
FI3033 17:00-17:15	An Efficient Implementation method for Superdirective Beamforming with Linear Arrays **Author(s):** Xiaoyuan Li, Yong Wang, Yixin Yang, Yichen Liu **Presenter:** Xiaoyuan Li, Northwestern Polytechnical University, China **Abstract:** Superdirective beamforming can achieve a narrower mainlobe width compared to conventional methods. However, its high sensitivity to errors and interferences limits its practical applications. This paper proposes an efficient implementation method for robust superdirective beamforming in linear arrays. First, a tridiagonal matrix is constructed based on the noise covariance matrix, and the optimal superdirective beam for the linear array is decomposed by exploiting the structural properties of Toeplitz matrices. The resulting eigenbeams exhibit different directivity and robustness. Subsequently, robust superdirective beampatterns of arbitrary orders are synthesized by selecting specific

	eigenbeams. Simulation results demonstrate that the proposed method achieves robust superdirectivity with a favorable trade-off between robustness and directivity.
FI3023 17:15-17:30	Smart Manufacturing Monitoring System Based on Industrial IoT **Author(s):** Yanping Zhang, Jianjun Yang **Presenter:** Yanping Zhang, Gonzaga University, USA **Abstract:** The Industrial Internet of Things (IIoT) has been extensively studied and applied across various industries. The integration of IoT, cloud computing, sensor technologies, and data processing has significantly advanced the digitization and automation of manufacturing processes. In automated manufacturing, the smooth operation of production along the line is essential for the overall efficiency and effectiveness of the system. This research proposes a smart factory monitoring system based on IIoT. We focus on the roles and functions of each component within the automated manufacturing environment, referred to as smart manufacturing. The proposed system integrates cloud services, smart sensors, data processing, and actuators/intelligent agents to work collaboratively, ensuring seamless operation of the automated production line. Simulation results demonstrate the effective and efficient design of the system and offer practical insights for implementing such a system across a wide range of industrial applications.

Onsite Session 4

Applications of Artificial Intelligence in Information Systems

- ➢ **Session Chair:** Ioanna Miliaresi, Ionian University, Greece
- ➢ **Time:** 16:00-17:15, September 11, 2025
- ➢ **Room:** G024, Eiffel 4, 43 Quai de Grenelle, 75015 Paris
- ➢ **Papers:** FI2024, FI614, FI620, FI724, FI3030-A

FI2024 16:00-16:15	OF-NeRF: Outward-Facing Cameras-based Neural Radiance Field for High-Resolution View Synthesis **Author(s):** Kaito Houjho, Aoki Terumasa **Presenter:** Kaito Houjho, Tokyo University of Technology, Japan **Abstract:** Novel view synthesis is a technique for generating free-view omni-directional images using multiple cameras. Traditionally, it has primarily been applied to synthesize novel views from inward-facing cameras placed around the center. This paper discusses whether novel view synthesis can also be achieved using outward-facing cameras placed around the center. In outward-facing camera configurations, the density of rays that can be captured is significantly lower compared to that in inward-facing cameras. As a result, achieving comparable reconstruction accuracy requires high-resolution cameras. The simplest approach to address this issue is to combine novel view synthesis with super-resolution techniques. In this paper, we present OF-NeRF (Outward-Facing cameras-based Neural Radiance Field), which is a novel method for synthesizing views from sparsely-placed outward-facing cameras. OF-NeRF integrates novel view synthesis and super-resolution techniques to achieve high-accuracy view generation.
FI614 16:15-16:30	Grid-Shift: An Image Preprocessing Approach to Reduce Overfitting in AI Training **Author(s):** Kevin Kocon, Michel Kraemer, Lina Emilie Budde **Presenter:** Kevin Kocon, TU Darmstadt, Germany **Abstract:** We present Grid-Shift, a lightweight image preprocessing approach to counteract overfitting when training Convolutional Neural Networks. Grid-Shift solves the problem that tiling large images for training disrupts coherent features (i.e. an object may be split at the edge of a sub-image) and thus leads to information loss. Existing augmentation methods that reduce overfitting do not solve this problem explicitly. In our case study of Land Use and Land Cover Classification, Grid-Shift outperforms all other approaches tested (a raw UNet, a UNet with Batch Normalization, and various augmentation methods). Grid-Shift achieves a Categorical Accuracy of 95%, which is almost 20% better than a raw UNet and still 4% better than the best augmentation approach tested.
FI620 16:30-16:45	Automation of Sperm Positioning Inside the ICSI Needle Using Computational Techniques **Author(s):** Celina Méndez Zavala, Estefania Hernandez Morales, Gerardo Mendizabal Ruiz **Presenter:** Celina Candelaria Méndez Zavala, Universidad de Guadalajara, México **Abstract:** Intracytoplasmic Sperm Injection (ICSI) is a widely used assisted reproduction technique involving the precise injection of a single spermatozoon into a mature oocyte. One of the most critical and challenging steps in this process is the manual manipulation

	and positioning of the sperm inside the injection pipette, which is prone to human error, variability, and potential damage to biological material. This work proposes a novel automated system based on Artificial Intelligence (AI) techniques to optimize and control sperm movement and positioning inside the ICSI needle. The system employs real-time AI-driven control algorithms, including reinforcement learning strategies, to enable adaptive and precise manipulation of sperm cells. This approach reduces operator dependency, enhances reproducibility, and improves the accuracy of sperm aspiration and delivery. By integrating AI and automation, the system aims to minimize procedural variability, reduce the risk of cellular damage, and ultimately improve fertilization success rates. This development represents a significant advancement in the automation of micromanipulation tasks within assisted reproductive technologies, opening new avenues for improving clinical outcomes in infertility treatments.
FI724 16:45-17:00	Optimizing Dataset Quality and Diversity in Neural Networks: A Study of the Vendi Score **Author(s):** Sebastian Krebs, Alexander Bernhardt, Bastian Goldlücke, Christopher Knievel **Presenter:** Sebastian Krebs, University of Applied Science Konstanz, Germany **Abstract:** We investigate methods for evaluating and optimizing dataset quality in the context of image-based object detection tasks, focusing on the Vendi Score and its enhanced variant, the quality-weighted Vendi Score. These metrics provide a robust framework for assessing and improving dataset diversity, a critical factor in effectively training neural networks. In particular, the quality-weighted Vendi Score allows for a customizable quality measure to account for dataset-specific characteristics. In this paper, we introduce a specific weighting function based on the ratio of objects to total images to ensure balanced dataset composition and to emphasize the importance of diverse object representations. Feature extraction techniques are employed to represent image objects in a unified feature space, enabling similarity calculations and diversity assessments. Additionally, a multi-objective genetic algorithm is utilized to optimize datasets across multiple classes, maximizing diversity while maintaining class balance. Our experimental results demonstrate that models trained on subsets optimized for diversity using the quality-weighted Vendi Score achieve improved performance. Diverse subsets lead to higher precision and generalization capabilities compared to randomly selected or less diverse datasets of the same size. Especially when sampling minibatches during training, this method may prove beneficial, as it allows for a more representative sample of the dataset. These findings underscore the pivotal role of dataset diversity in enhancing neural network performance and highlight the utility of the Vendi Score and the quality-weighted Vendi Score as valuable tools for strategically shaping dataset composition to improve outcomes.
FI3030-A 17:00-17:15	Signal characteristics for the identification of the asymmetric A0 and symmetric S0 Lamb wave modes **Author(s):** Lina Draudvilienė **Presenter:** Lina Draudviliene, Kaunas University of Technology, Lithuania **Abstract:** The study analyses the distribution of higher and lower frequency components and phase velocity values across different parts of the Lamb wave signals for the identifications of the A0 and S0 modes. Signals of Lamb wave A0 and S0 modes at different frequency ranges were used in the theoretical study, revealing five different trends of the phase and group velocities. The results indicate that the group velocity primarily affects the distribution and redistribution of frequency components. It was found that the higher and lower frequency components across the signal parts are distributed differently for both A0 and S0 mode signals, depending on the trend of the group velocity. Thus, the A0 and S0 modes cannot be distinguished solely based on their time–frequency (t–f)

distributions. It was found that the estimation of the phase velocity of the frequency values can be used as a characteristic for the identification of the A0 and S0 modes. The higher-frequency components of the A0 mode propagate faster than the lower-frequency components, regardless of their location in the signal, meanwhile the lower-frequency components of the S0 mode propagate faster than the higher-frequency components, regardless of their location in the signal. Based on these observed regularities, an algorithm for identifying the A0 and S0 modes is proposed. The experimental study was confirmed the determined regularities using signals of the A0 and S0 modes propagating in an aluminium plate, as well as using A0 mode signals propagating in a Glass Fibre Reinforced Plate (GFRP). The obtained results present the main regularities that aid in the interpretation and evaluation of A0 and S0 mode signals. The establish regularities can be used as a tool for determining the characteristics of higher An and Sn modes.

Online Session 1

Image Analysis and Computing

➢ **Time:** 10:00-11:45, September 12, Paris Standard Time, UTC/GMT+2

➢ **Zoom link:** https://us02web.zoom.us/j/88386795743

➢ **Passcode:** 091012

➢ **Session Chair:** Umesh Chandra Pati, National Institute of Technology, India

➢ **Papers:** Invited Speaker-Umesh C. Pati, Invited Speaker-Herminiño C. Lagunzad, FI3029-A, FI506, FI4039

Invited Speaker 10:00-10:20	**Umesh Chandra Pati** **National Institute of Technology Rourkela, India** *Speech Title: Video-based Intrusion Detection System using Deep Learning Techniques* **Abstract:** There is a huge demand for video surveillance-based intelligent security systems that can automatically detect unauthorized entry or mal-intentional intrusion into the unattended sensitive areas and notify the concerned authorities in real-time. A novel video-based Intrusion Detection System (IDS) using deep learning techniques has been proposed. In this work, You Only Look Once (YOLO) algorithm is used for object detection. Then, the intrusion is detected using the proposed algorithm based on the Shifted centre of the bounding box of the detected object. Further, Simple Online and Realtime Tracking (SORT) algorithm is used for the tracking of the intruder in real-time. The developed system is implemented and tested for live video stream using NVIDIA Jetson TX2 development platform. The proposed IDS is a generic one where the user can select the region of interest (the area to be intrusion-free) of any size as well as shape from the reference (starting) frame and potential intruders, such as a person, vehicle, etc., from the list of trained object classes. It can have a wide range of applications, like person intrusion-free zone, no vehicle entry zone, no parking zone, smart home security, etc.

Invited Speaker 10:20-10:40	**Herminiño C. Lagunzad** **National University, Philippines** *Speech Title: Securing Employee Records via 7th-Bit Manipulation in RGB Images for Improved Information Hiding* **Abstract:** The exchange of data across platforms often faces risks such as unauthorized access, underscoring the need for advanced methods to enhance information security. Image steganography in the RGB color space particularly through 7th-bit manipulation offers a promising solution for improving data concealment. Steganography, the technique of embedding information within other data, is increasingly recognized for its potential in secure communication and data protection. This study presents a method for securing employee records by embedding encrypted information into the least significant bits of RGB channels in widely used image formats, including PNG, JPG, and JPEG. The proposed approach ensures minimal perceptible changes to the visual quality of images while effectively safeguarding sensitive data. Experimental results confirm the method's robustness against common attacks, its high embedding capacity, and strong imperceptibility, highlighting the effectiveness of 7th-bit manipulation in advanced steganographic applications. These findings support the development of more secure communication protocols and strengthened data protection mechanisms for managing sensitive employee information in contemporary digital environments.
Invited Speaker 10:40-11:00	**San Gultekin** **Meta Platforms, Inc., United States** *Speech Title: Securing Employee Records via 7th-Bit Manipulation in RGB Images for Improved Information Hiding* **Abstract:** Since the seminal work of R. E. Kalman in 1960, Kalman Filters have found applications across a wide range of domains, from radar target tracking to robotics, from collaborative filtering to natural language processing. In this talk, we will revisit Kalman filtering through a Bayesian lens, which provides a probabilistic framework for formulating a broad class of problems. In particular, we will highlight the divergence minimization framework, offering a unified perspective that connects seemingly disparate models such as collaborative filtering and sensor network tracking. Finally, we will discuss extensions based on α-divergences, leveraging automatic differentiation with reparameterized gradients to implement nonlinear Kalman filters in a manner reminiscent of neural networks.

FI3029-A 10:00-10:15	Novel Generative Adversarial Network Architecture for Efficient Underwater Image Enhancement **Author(s):** Gordon Hung **Presenter:** Gordon Hung, Independent Researcher, Taiwan **Abstract:** Underwater image enhancement is becoming increasingly vital due to advancements in marine robotics and machine learning. Images captured underwater often suffer from color distortions, low contrast, and blurred details, which are detrimental to later analysis. As researchers have become more reliant on unmanned underwater vehicles (UUVs), the need for a robust image pre-processing framework has grown. While the existing literature consists of several strong image enhancement models, they are often unable to balance computational cost with model accuracy. This study proposes a novel lightweight Generative Adversarial Network (GAN) architecture specifically designed for underwater image enhancement; the model has been shown to achieve a superior balance between accuracy and computational efficiency. We provide a comprehensive evaluation of state-of-the-art generative models on the EUVP dataset and demonstrate that our GAN can perform high-quality enhancement while requiring significantly fewer computational resources. The contributions of this study would significantly enable efficient edge development, allowing enhanced image pre-processing to be performed directly on hardware with limited resources.
FI506 10:15-10:30	An Efficient and Lightweight Way of Instance Level Image Retrieval: Chain Retrieval with Visual Transformers **Author(s):** Yagmur Cigdem Aktas, Jorge Garcia Castano **Presenter:** Yagmur Cigdem Aktas, VICOMTECHl, Spain **Abstract:** —Instance-level image retrieval (ILIR) requires fine grained discrimination to identify all images depicting the same object or scene as a query, regardless of variations in viewpoint, scale, or background. While recent methods based on convolutional and transformer architectures have achieved high accuracy, most rely on supervised training and dataset specific tuning, limiting their generalization to real-world, open set scenarios. Pretrained Vision Transformers such as CLIP [1] offer a promising zero-shot alternative, but current approaches for image-to-image retrieval yet far from the supervised method's performances for benchmark datasets, as well as they focus only to the ranking task, not offering a solution for the amount of the images to retrieve, which is an important point for real world retrieval task to prevent false retrievals and obtaining unnecessary amount of retrieved data at the end. In this paper, we propose Chain Retrieve, a fully zero-shot retrieval framework that iteratively refines the query embedding by integrating information from top-k retrieved images. Unlike existing methods, our approach does not assume a fixed result count or rely on handcrafted thresholds. Instead, it expands the query representation in a retrieval loop and stops automatically when similarity scores fall below a fixed, interpretable threshold. Built entirely on pre-trained CLIP [1] ViTs and requiring no training or fine-tuning, Chain Retrieve achieves superior performance over existing zero-shot image-to-image methods on both custom and benchmark datasets, narrowing the gap with supervised models and offering a scalable, plug-and-play solution for ILIR.

FI4039 11:30-11:45	Research and Development of FPGA Layout and Routing Algorithms: From Traditional Methods to Open Source Frameworks and 3D Integration **Author(s):** Ma Shurui, Ye Tong, MingJun Xu, ZhiXuan Liang **Presenter:** Shurui Ma, China University of Petroleum (East China), China **Abstract:** This article provides an overview of the evolution of FPGA layout and routing algorithms, covering traditional methods (such as K-L algorithm), open-source frameworks (Yosys+NextPNR), and 3D integration technologies. Research has shown that traditional K-L algorithms suffer from local convergence defects and computational efficiency bottlenecks; The open-source Yosys+NextPNR framework significantly reduces the design threshold and promotes transparency innovation; The 3D layered hybrid algorithm (partitioning+simulated annealing) outperforms traditional 2D solutions in terms of online length ($\downarrow$ 21% and latency ($\downarrow$ 24%), but requires addressing thermal management and vertical through-hole optimization issues. In the future, it is necessary to integrate reinforcement learning, multiobjective optimization, and heterogeneous acceleration technologies to unleash the potential of next-generation reconfigurable computing.

Online Session 2

Artificial Intelligence Theory and Innovative Applications

- ➤ **Time:** 10:00-11:45, September 12, Paris Standard Time, UTC/GMT+2
- ➤ **Zoom link:** https://us02web.zoom.us/j/85941126636
- ➤ **Passcode:** 091012
- ➤ **Session Chair:** Ruksar Fatima, Khaja Bandanwaz University, India
- ➤ **Papers:** Invited Speaker-Ruksar Fatima, Invited Speaker-Warren B. Panizales, Invited Speaker-Bo Sun, FI617, FI826, FI616

Invited Speaker 10:00-10:20	**Ruksar Fatima** **Khaja Bandanwaz University, India** *Speech Title: Artificial Intelligence in Healthcare: Transforming Diagnosis, Personalization, and Predictive Medicine for a Smarter Future* **Abstract:** Artificial Intelligence (AI) is revolutionizing healthcare by offering data-driven solutions that enhance diagnostic accuracy, enable predictive analytics, and promote personalized medicine. From convolutional neural networks (CNNs) detecting early-stage cancers in radiology to AI-driven tools for managing chronic diseases via wearable sensors, AI is becoming an indispensable part of clinical practice. Predictive models are helping clinicians anticipate patient deterioration, while AI algorithms in genomics and pharmacology are optimizing treatment decisions based on individual profiles. Furthermore, AI chatbots and sentiment analysis tools are transforming mental healthcare delivery, offering scalable and accessible psychological support. Despite these advancements, challenges such as data privacy, algorithmic bias, and the need for interpretability persist. Ethical and regulatory frameworks must evolve to ensure equitable deployment and integration of AI systems in diverse clinical settings. This paper explores the current landscape, real-world applications, and future potential of AI in healthcare, highlighting the transformative shift towards smarter, patient-centric care.

Invited Speaker 10:20-10:40	**Warren B. Panizales** **National University Fairview, Philippines** *Speech Title: AI-Enhanced Flood Hazard Identification Using Predictive Analytics for Philippine Communities* **Abstract:** Flooding is one of the most frequent and destructive natural hazards in the Philippines, causing extensive economic losses, infrastructure damage, and displacement of vulnerable communities each year. Conventional flood hazard mapping and early warning systems, while widely implemented, often face limitations in predictive accuracy, spatial coverage, and timeliness due to reliance on historical hydrological data and static models. In response, Artificial Intelligence (AI) and predictive analytics have emerged as promising tools to enhance flood hazard identification by processing large, heterogeneous datasets such as rainfall, river discharge, topography, land use, and satellite imagery. This research proposes an AI-enhanced flood hazard identification framework tailored to Philippine communities. The system integrates predictive models such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNNs), and ensemble learning with geospatial data and local hydrological records. Data sources include PAGASA hydrometeorological reports, Project NOAH hazard maps, and real-time sensor networks deployed in flood-prone river basins. Preprocessing involves extraction and normalization to ensure accurate model training. Predictive outputs are visualized through dynamic hazard maps that update with real-time inputs, enabling more effective risk communication to local governments and residents. The expected contribution of this study is twofold: (1) advancing the application of AI-driven predictive analytics for localized flood hazard identification in the Philippines, and (2) supporting disaster risk reduction and climate resilience efforts by providing an adaptive, data-driven decision-support tool for communities and policymakers. This framework is envisioned to strengthen early warning capabilities, minimize flood-related damages, and improve preparedness in high-risk areas.

Invited Speaker 10:40-11:00	**Bo Sun** **Chang'an University, China** *Speech Title: Dimensionally Unified Metric Model for Multisource and Multiview Scene Matching* **Abstract:** The core challenge of multiview scene matching is to effectively extract features from multiview images, which is a key factor in achieving robust scene matching. This article delves into methodologies for enhancing the multiview robustness of scene matching, with the following key contributions: 1) this article pro_x0002_poses the metric feature consistency principle which emphasizes the necessity of implementing accurate correspondence between drone and satellite images. To verify the principle, consistent feature enhancement for channel, spatial, and hybrid dimensions is explored. As a counterexample, the shuffle operation is used to break the consistency of dimension semantic information. Experiments show that the mAP is reduced by about 20% after breaking the consistency relationship of metric features; 2) built upon the principle, this article introduces a scene-matching framework named dimensionally unified metric model (DUMM). Utilizing the multidimensional feature enhancement model effec_x0002_tively enhances the correspondence of the dual-branch features within the siamese network; and 3) this article is the first to introduce the feature dimension and frequency-domain feature. By the upper and lower bounds of the cosine transform function, the negative effects of cross-view variations can be effectively mit_x0002_igated, thus enhancing the overall robustness. The introduction of frequency features improves the mAP by 2.54% compared with the method of improving the dimensional consistency of $H \times W \times C$, verifying the positive impact of the fusion of frequency domain features on enhancing the robustness of scene matching. Nevertheless, it remains imperative to adhere to the principle of feature dimension consistency across all additional frequency dimensions.
FI617 11:00-11:15	Agentic AI for Explainable Data Analysis in District Heating Systems: An AutoGen-Based Approach **Author(s):** Matija Speletic, Stevica Cvetkovic, Milan Zdravkovic **Presenter:** Stevica Cvetkovic, University of Nis Faculty of Electronic Engineering, Serbia **Abstract:** This paper introduces a methodology for automated, comprehensive data analysis and report generation using a multi-agent AI system powered by Large Language Models and AutoGen framework. To orchestrate the collaboration, a Round-Robin communication pattern coordinates three specialized AI agents: a Data Scientist for analytical reasoning, a Coder Agent for source code generation, and a Code Executor for safe execution in a Docker environment. Applied to a real-world District Heating System dataset, the system successfully identifies patterns, trends, anomalies, performs correlation analysis, and assesses system efficiency. The iterative process, constrained to textual outputs between agents, leads to a final LLM that synthesizes the analysis into a detailed Jupyter Notebook report, including source code and automatically generated visualizations. The results demonstrate the effective and high-quality autonomous execution of complex data science tasks by this agentic AI system.

FI826 11:15-11:30	A Holistic AI Approach for Secure and Scalable Exam Registration **Author(s):** Pal Amutha K, Sethukkarasi C, Mehanathen N, Ethirajan D **Presenter:** Pal Amutha K, Centre for Development of Advanced Computing, India **Abstract:** As Computer Based Testing (CBT) grows, examination registration systems have an increased exposure to identity fraud, duplicate registration, and a reliance on manual checking. This Proposed System presents a cognitive training system that utilizes AI technologies, focusing on the enhancement of security, accuracy, and scaling in CBT registration procedure. Through the use of computer vision for facial recognition and image quality analysis, the tool carries out real-time automated ID checks. It also includes an unsupervised machine learning to identify anomalies like duplicated and suspicious patterns of registration. The proposed system reduces human errors and the amount of operation for a registrar, and builds a high degree of transparency and trust into the registration process. With a focus on both security and the student experience, it enables secure exam taking at scale.
FI616 11:30-11:45	AI-Powered Phishing Detection In Email Forensics: A Machine Learning Approach For Cyber Threat **Author(s):** Jayanthi Ganapathy, Purushothaman Ramachandran, Vasantha Kumar D, Aadhavan P, Suganthini **Presenter:** Jayanthi G, Sri Ramachandra Institute of Higher Education and Research, India **Abstract:** Phishing remains one of the most widespread and evolving cyber threats, often delivered through deceptive emails that imitate trusted entities to steal sensitive information. Traditional email security solutions such as blacklists and rule-based filters are insufficient against modern, context-aware phishing attacks. This project presents an AI-powered phishing detection system that leverages both Machine Learning and Deep Learning techniques to analyze and classify email content. A Random Forest classifier is employed for fast, interpretable detection using TF-IDF features, while a Convolutional Neural Network (CNN) is used for deeper contextual understanding of email semantics. The system was trained and tested on a labeled dataset of phishing and legitimate emails, achieving 75% accuracy with Random Forest and 87% accuracy with CNN. A command-line interface (CLI) is implemented for real-time predictions, and a Flask-based web deployment is proposed for future usability. This dual-model approach provides a scalable, efficient, and accurate solution to detect phishing emails and enhance cybersecurity defenses.

Online Session 3

Advanced Information Technology and Data Management

➢ **Time:** 13:00-14:40, September 12, Paris Standard Time, UTC/GMT+2
➢ **Zoom link:** https://us02web.zoom.us/j/88386795743
➢ **Passcode:** 091012
➢ **Session Chair:** Ronakkumar N Patel, Charotar University of Science and Technology, India
➢ **Papers:** Invited Speaker-Ronakkumar N Patel, Invited Speaker-Xinguo Yu, FI3037, FI4041, FI502, FI827

Invited Speaker 13:00-13:20	**Ronakkumar N Patel** **Charotar University of Science and Technology, India** *Speech Title: Generative AI and LLMs in Multimedia: Towards Smarter Analysis, Interaction, and Recommendation* **Abstract:** The emergence of Generative AI and Large Language Models (LLMs) is transforming the field of multimedia processing by enabling smarter analysis, richer interaction, and more effective recommendations. While traditional multimedia systems have primarily leveraged deep learning for image recognition, video analysis, and content delivery, the integration of LLMs introduces new opportunities for multimodal intelligence, where text, vision, and audio can be seamlessly understood in unison. This session examines how LLMs, when combined with deep learning and recommender systems, are reshaping multimedia technology. LLMs provide enhanced semantic understanding and contextual reasoning, which strengthen content description, cross-modal retrieval, and adaptive recommendation. Generative AI further enriches these capabilities by creating context-aware, human-centric interactions that go beyond static analysis toward dynamic and personalized multimedia experiences. The talk will also address the critical challenges facing this convergence, including scalability in large-scale multimedia systems, interpretability of AI-driven recommendations, ethical and privacy considerations, and the need for responsible deployment. By highlighting current advances and open research directions, the keynote envisions a future where multimedia platforms are not only technically intelligent but also transparent, trustworthy, and user-focused, unlocking new possibilities in both academic research and real-world applications.

Invited Speaker 13:20-13:40	**Xinguo Yu** **Central China Normal University, China** *Speech Title: Advancing System 2 Reasoning in LLMs for Complex Problem-Solving* **Abstract:** This talk delves into how large language models (LLMs) can harness System 2 reasoning—a deliberate, analytical form of human cognition—to solve complex problems. Rooted in dual-process theory from cognitive psychology, we contrast System 2's slow, reflective processes with System 1's fast, heuristic-driven thinking. Through real-world examples, including mathematical proofs and multi-step logical challenges, we showcase advanced techniques like chain-of-thought prompting and tree-of-thought architectures, which enhance LLMs' ability to reason, self-correct, and generalize. Yet, critical challenges remain: computational overhead, alignment with human reasoning patterns, and ensuring trustworthiness in high-stakes applications. Join us to explore how integrating insights from cognitive science can pave the way for next-generation AI systems that seamlessly blend machine intelligence with human-like analytical rigor.
FI3037 13:40-13:55	Cross-Subject and Cross-Session Continual Adaptation of EEG Signals Using Temporal Context Encoding **Author(s):** Saba Aslam, Musarat Hussain, Hongyan Wu, Tianyuan Ru **Presenter:** Saba Aslam, Chinese Academy of Sciences, China **Abstract:** Electroencephalogram (EEG) based emotion recognition faces significant challenges in real-world deployment due to non-stationary data distributions across sessions and subjects, leading to catastrophic forgetting in static deep learning models. This study proposes a unified continual learning (CL) framework combining a temporal context encoder (LSTMAttention) with two CL strategies: Fisher-Guided Synaptic Stability (EWC) for parameter regularization and Episodic Experience Integration (Replay) for memory retention. Evaluated on the SEED-IV dataset under session-based (SeCL) and subjectbased (SuCL) incremental learning scenarios, our framework outperforms Sequential Adaptation (SA) and Naive Fine-tuning, with Replay excelling in sessions (92% accuracy, 0.02 forgetting) and EWC dominating subject adaptation (0.955 stability ratio, 0.045 forgetting). Both strategies significantly reduce forgetting (>50%) versus baselines, demonstrating robust crosssession/subject generalization for real-world EEG applications.
FI4041 13:55-14:10	Recursive Covariance-Based Quadratic Filtering under Random Linear False Data Injection Attacks **Author(s):** Raquel Caballero-Águila, M. Pilar Frías-Bustamante, Josefa Linares-Pérez **Presenter:** Raquel Caballero-Águila, University of Jaén, Spain **Abstract:** This paper addresses the least-squares (LS) quadratic filtering problem for linear discrete-time non-Gaussian systems subject to random linear false data injection attacks, which compromise the integrity of the transmitted measurement data at each

	sampling instant. A sequence of Bernoulli-distributed random variables is incorporated into the observation model to characterize the stochastic nature of these attacks. A recursive LS quadratic filtering algorithm is proposed, based on the construction of an augmented observation model that combines the original signal and observation vectors with their second-order powers, defined through the Kronecker product. This transformation allows the LS quadratic filtering problem —originally defined in terms of the observations and their second-order powers— to be reformulated as a linear LS filtering problem based on the augmented observations. Unlike traditional approaches, the proposed algorithm does not require explicit use of the signal evolution equation. Instead, it relies on a covariance-based estimation framework, assuming the signal is zero-mean and that the covariance and cross-covariance functions of the signal and its second-order powers admit a separable representation. The proposed LS quadratic filter outperforms conventional linear filters in terms of estimation accuracy, while preserving computational tractability and enabling recursive implementation.
FI502 14:10-14:25	An Senegalese Legal Texts Structuration Using LLM-augmented Knowledge Graph **Author(s):** Oumar Kane, Mouhamad M. Allaya, Dame Samb, Mamadou Bousso **Presenter:** Oumar Kane, Iba Der Thiam University, Senegal **Abstract:** This study examines the application of artificial intelligence (AI) and large language models (LLM) to improve access to legal texts in Senegal's judicial system. The emphasis is on the difficulties of extracting and organizing legal documents, highlighting the need for better access to judicial information. The research successfully extracted 7,967 articles from various legal documents, particularly focusing on the Land and Public Domain Code. A detailed graph database was developed, which contains 2,872 nodes and 10,774 relationships, aiding in the visualization of interconnections within legal texts. In addition, advanced triple extraction techniques were utilized for knowledge, demonstrating the effectiveness of models such as GPT-4o, GPT-4, and Mistral-Large in identifying relationships and relevant metadata. Through these technologies, the aim is to create a solid framework that allows Senegalese citizens and legal professionals to more effectively understand their rights and responsibilities.
FI827 14:25-14:40	Study on the Indicators Evaluating Innovation Abilities of High-end Equipment Manufacturing Industry in Sichuan Province **Author(s):** Wei Liu, Yijiang Hong **Presenter:** Wei Liu, Chengdu Jincheng College, China **Abstract:** This paper aims to establish an evaluation index system for the innovation capacity of the high-end equipment manufacturing industry in Sichuan Province, providing a scientific basis for the evaluation and cultivation of innovation capacity for related enterprises. Through a comprehensive analysis of the characteristics of the high-end equipment manufacturing industry in Sichuan Province and drawing on existing research results, an initial evaluation index system was constructed, which includes four primary indicators, ten secondary indicators, and twenty-five tertiary indicators. After expert-interviewing and indicator screening, a simplified index system was finally formed, including three primary indicators, six secondary indicators, and nine tertiary indicators. Then ordering relation analysis method was applied to determine the weights of each indicator, among which the weight of innovation input capacity is the highest, indicating its crucial and fundamental role in innovation capacity. To provide high-quality development references for Sichuan's high-end equipment manufacturing industry, this study also proposed countermeasures and suggestions such as increasing innovation input, elevating innovation management level, optimizing innovation output and value realization, promoting the construction of innovation cooperation mechanisms, and strengthening policy support and guidance.

Notes