3ICT 2026 | Complete Conference Schedule

18–19 November 2026 • InterContinental Regency Bahrain

All times Bahrain (UTC+3)

DAY 1 | Wednesday, 18 November 2026

08:00–09:00	Registration
09:00–10:00	Opening Ceremony — H.E. the UoB President’s Speech; Keynotes 1 & 2
10:00–10:30	Coffee Break
10:30–11:50	D1-A1 Secure LLMs & Generative AI Systems D1-B1 Explainable AI for Health & Predictive Analytics D1-C1 Smart Infrastructure, Robotics & Energy AI
11:50–12:20	Coffee Break
12:20–13:40	D1-A2 LLM Robustness, Safety & Prompt Security D1-B2 AI Optimization, Vision & Immersive Computing D1-C2 Threat Detection, Phishing & Digital Forensics
13:40–14:40	Lunch Break
14:40–16:00	D1-A3 Cybersecurity Governance, Risk & Cloud Forensics D1-B3 IoT Security, Intrusion Detection & Federated AI D1-C3 Intelligent Communications & Secure AI Applications

DAY 2 | Thursday, 19 November 2026

09:00–10:40	D2-A1 Cybersecurity, Phishing & Adversarial AI D2-B1 AI for Healthcare & Clinical Decision Support D2-C1 Next-Generation 5G/6G & Digital Twin Networks
10:40–12:20	D2-A2 Privacy-Preserving AI & Security Analytics D2-B2 Computer Vision & AI for Real-World Monitoring D2-C2 Intelligent Networks & Emergency Communications
12:20–12:35	Coffee Break
12:35–14:15	D2-A3 Intelligent Cloud, FinTech & Agentic Security D2-B3 Machine Learning, Optimization & Predictive Models D2-C3 IoT Applications, Smart Energy & Logistics
14:15–15:15	Lunch Break
15:15–16:55	D2-A4 Smart Cities, Sustainable Systems & Monitoring D2-B4 Big Data, Health AI & Knowledge Systems D2-C4 Digital Education, HCI & Computing Systems
16:55–17:40	Keynote 3
17:40–18:00	Conference Closing

Each paper: 15-minute presentation + 5-minute Q&A. See following pages for paper details and abstracts.

2026 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) Program

📄 Program Summary (PDF)

Full program and paper abstracts follow the complete schedule on the preceding page.

Wednesday, November 18

Wednesday, November 18 10:30 - 11:50

D1-A1: Secure LLMs & Generative AI Systems

Design and Evaluation of a Secure RAG-Based AI-Powered Virtual University Assistant

Nayla Khalid Aljaber, Latifa Alkooheji, Jumana Bucheeri, Zahra Nayem and Abdulla Aldoseri

Universities increasingly use conversational interfaces to improve access to academic information, but large language model applications can produce unsupported answers, expose sensitive data, or follow manipulated instructions. This paper presents a locally hosted AI-powered virtual university assistant developed as a University of Bahrain case study. The system combines Retrieval-Augmented Generation, LlamaIndex document indexing, BAAI bge-base-en-v1.5 embeddings, and a Llama 3.1:8B model served through Ollama. Guest users can query program and policy information, while verified prototype student sessions combine retrieved university documents with simulated academic records and deterministic utilities for expected GPA calculation and target-grade estimation. Security controls include regex-based input filtering, instruction-data separation using structured delimiters and sanitization, hardened system-level instructions, session isolation, and encrypted SSH tunneling to the university high-performance computing environment that hosts the backend AI service and models. Evaluation compared embedding and language-model configurations and tested prompt-injection resistance, multi-user behavior, and latency. Llama 3.1 improved correctness, relevance, semantic similarity, answer relevance, and guideline adherence over Phi-3. Across six representative workloads, end-to-end response time ranged from 2.28 to 13.63 s, with a mean of 7.05 s. The prototype provides a privacy-aware baseline for secure institution-specific academic assistance.

QueryMind: Design and Evaluation of Secure and Privacy-Preserving Text-to-SQL Using Schema-Aware RAG and LLMs

Mohamed Almaraghi, Ali Alsowaidi, Abdulrahman Bumeajib and Abdulla Aldoseri

Natural-language interfaces can reduce the technical barrier to relational database access, but LLM-assisted querying raises privacy and governance concerns when sensitive records are exposed during model processing. This paper presents QueryMind, a privacy-oriented Text-to-SQL framework designed to generate database queries without providing underlying database records to the language model. QueryMind combines schema-aware Retrieval-Augmented Generation (RAG), locally deployed LLMs, and SELECT-only SQL validation. Relevant schema metadata is retrieved and supplied as structural context for SQL generation, limiting model visibility to database structure rather than record content. The system was evaluated on an e-commerce relational database using five LLM and embedding-model configurations. Llama 3.1 8B with mxbai-embed-large generated correct results for all benchmark queries, with an average end-to-end processing time of approximately 33 seconds. Smaller configurations reduced average processing time to approximately 4-7 seconds, with result match rates between 40% and 60%. Performance profiling identified LLM generation as the principal computational bottleneck, accounting for approximately 86% of measured processing time. Security evaluation assessed five predefined scenarios involving SQL injection, cross-site scripting (XSS), session misuse, and unsafe SQL operations, and all produced the expected defensive outcome. Local and remote database experiments showed limited influence from database location on overall latency. The findings support the feasibility of schema-restricted AI-assisted database querying as a privacy-preserving baseline, while identifying model latency and smaller-model accuracy as areas for future optimization.

Fine-Tuning Large Language Models for Secret Detection in Postman API Collections

Ali Mohsin, Mahmood Sabt and Wael M El-Medany

Postman Application Programming Interface (API) collections frequently contain hard-coded secrets such as API keys, authentication tokens, and credentials, creating significant security risks when shared or stored insecurely. Traditional Regular Expressions (Regex)-based detection tools produce excessive false positives (50% precision), making them impractical due to alert fatigue. This paper presents a hybrid approach that combines regex-based candidate extraction with fine-tuning Large Language Models (LLMs) for contextual classification of secrets in Postman API collections. The Qwen2.5-3B-Instruct model has been fine-tuning on a dataset of manually verified secrets from Postman API collections. The results demonstrate that the fine-tuning model achieves strong performance with 94.9% accuracy, 99.3% precision, 90.3% recall, and 94.6% F1- score, substantially outperforming regex-only baselines. Using the Parameter Efficient Fine-Tuning (PEFT) and Low-Rank Adaptation (LoRA) with 4-bit quantization, inference runs efficiently on a single NVIDIA RTX 5080 Graphics Processing Unit (GPU). This locally deployable model reduces false positives by 98% while maintaining high recall, making it suitable for production deployment while preserving privacy and without overwhelming developers.

ShellSentry: A Defense-in-Depth Framework for Secure Natural Language-to-Bash Execution on Remote Linux Systems

Dana Hussain, Aljazi Ali Almujaddam and Abdulla Aldoseri

Large Language Models (LLMs) have significantly improved natural language-to-Bash translation, enabling users to interact with Linux systems using intuitive textual instructions. However, executing AI-generated Bash commands directly on remote hosts introduces critical security risks, including hallucinated commands, prompt injection, privilege misuse, and unintended destructive operations. Existing Natural Language-to-Bash (NL2Bash) solutions primarily focus on command generation accuracy while providing limited support for secure and controlled execution in operational environments. This paper presents ShellSentry, a defense-in-depth framework for secure natural language-to- Bash execution across remote Linux systems. ShellSentry combines retrieval-augmented generation (RAG), host-aware context collection, deterministic multi-layer command validation, authenticated SSH-based multi-host orchestration, configurable execution policies, and comprehensive audit logging to reduce the risks associated with AI-generated commands. Rather than directly trusting LLM outputs, the framework validates generated commands through multiple security layers before execution. The proposed framework was evaluated in a controlled virtual Linux environment using functional, security, and performance testing, together with comparisons across different LLM configurations and a real- world solution. The framework achieved a mean end-to-end Time to First Byte (TTFB) of approximately 2.03 s for single- host execution, while rejected unsafe requests completed in approximately 1.51 s. In the LLM comparison, llama-3.1-8b- instant demonstrated better overall response performance, with a mean TTFB of 4.67 s compared with 7.97 s for llama-3.3- 70b-versatile, while both models produced acceptable Bash command variants. The evaluation demonstrates that ShellSentry effectively prevents unsafe command execution while maintaining reliable remote administration, execution traceability, and multi-host management capabilities. These findings demonstrate that integrating layered security mechanisms around LLM-assisted system administration provides a practical and trustworthy approach for secure AI- driven shell automation.

D1-B1: Explainable AI for Health & Predictive Analytics

Explainable Machine Learning for Breast Cancer Survival Prediction Using SEER Clinical Data

Mohammed Majid M Al Khalidy and Abdulrahman Abdulla Al Khaja

Prediction of breast cancer survival from registry data faces two challenges: class imbalance and risk of temporal leakage. Here we demonstrate an explainability-oriented ML pipeline on the SEER Breast Cancer dataset (4024 patients, diagnosed 2006-2010). Our pipeline eliminates a leakage factor (Survival Months; single-feature ROC-AUC = 0.83), addresses the 5.53: 1 imbalance ratio with class-weighted model training, and evaluates five classifiers using a leakage-safe nested cross-validation protocol. Logistic Regression provided the best overall validation set performance (ROC-AUC = 0.7472 ± 0.0228; average precision = 0.408). Wilcoxon signed-rank tests show significant differences from each of the four alternatives ($p < 0.001$). Six standard prognostic factors (age, regional positive nodes, examined nodes, N stage, histologic grade, and progesterone-receptor status) are identified by SHAP analysis as the primary determinants of mortality risk and have consistent rankings across all model families. A held-out test set will be used for confirmatory analysis in a follow-up report.

Beyond Prediction: An Explainable AI Framework for Early Risk Detection, Personalised Intervention Planning, and Dynamic Reassessment of At-Risk Students

Israa Tajaldin and Shahnawaz Khan

Early identification of academically at-risk students improves retention and benefits institutions, but many predictive approaches provide limited guidance on why a student is at risk, what intervention may be appropriate, or how risk changes afterwards. This study proposes a closed AI framework integrating early risk prediction, SHAP-based explanation, personalised intervention support, and dynamic reassessment. Anonymised Banner and Moodle data from 3,760 Bahrain Polytechnic students, limited to Week 5 prediction factors. Six machine learning classifiers were compared, with a calibrated Random Forest selected as the final model. On the held-out test set, the model achieved a recall of 0.9091, an F1 score of 0.8252, and a PR AUC of 0.9020, identifying 225 students as at risk. Student specific SHAP evidence was mapped to five intervention support categories. Reassessment reused the frozen model under Conservative, Moderate, and Strong progression scenarios, reducing mean predicted risk from 0.7600 to 0.7450, 0.7321, and 0.7204, respectively. The results demonstrate the potential of connecting prediction, explanation, intervention support, and reassessment within one consistent academic decision support framework.

A Diagnosis-Informed Multitask Framework for Multimodal Cardiopulmonary Prediction

Arwa Salem Yalos Alenezi and Mohammed Majid M Al Khalidy

Clinical assessment of cardiopulmonary disease extends beyond diagnosis to treatment, risk and the level of care required, with these decisions closely linked. Existing multimodal systems usually address a single outcome or predict several outcomes in parallel from a shared representation, without allowing the output of one task to inform the next. This paper presents a cascaded multimodal framework in which diagnostic predictions are generated first and then supplied to the tasks that follow. The system combines chest radiography, electrocardiography, laboratory results, echocardiography, clinical history and diagnoses from previous admissions, addressing five prediction tasks across 52,337 encounters from MIMIC IV. Disease detection achieved an AUC of 92.1 percent across twelve conditions, medication recommendation improved the Jaccard index by 0.080 over an established input representation on the same cohort, and intensive care admission achieved an AUC of 91.4 percent using only information available before the admission decision. Ablation showed that adding diagnostic predictions improved intensive care admission by 2.1 AUC points while leaving mortality and medication unchanged, indicating that the value of the cascade depends on the clinical relationship between tasks.

Explainable Deep Learning for Remaining Useful Life Prediction in Rotating Machinery

Muhammad Farrukh Amin and Mohab A Mangoud M

For Predictive maintenance, Remaining Useful Life (RUL) prediction is one of the key tasks as it helps estimate how long a machine would take before failure. If multivariate time-series sensor data is available, RUL can be predicted using Deep learning models. In practical settings, high predictive accuracy alone is often not enough. Maintenance engineers also need to understand why a model produced a certain RUL estimate. In this work, an explainable deep learning framework is developed for RUL prediction using NASA C-MAPSS benchmark dataset. Multiple deep learning variants including a 2-stage LSTM baseline, a BiLSTM with temporal attention, and CNN + LSTM model were tested in this study. Each architecture is evaluated on FD001-FD004 using independently trained models, 40-cycle inputs, 24 features excluding the cycle index, three training seeds, and all official test engines. The CNN+LSTM yields mean last-cycle RMSEs of 14.50, 15.24, 16.57, and 17.30 cycles on the four subsets, respectively; it is not the best architecture on every subset. Global SHAP covers 100 engines per subset, while 30 stratified engines per subset receive paired local SHAP and LIME explanations. Reference-background ranking stability, explanation reconstruction error, feature-replacement fidelity, progressive fidelity, and SHAP-LIME rank agreement was measured. Mean background-ranking correlations are 0.995-0.999; mean local SHAP-LIME rank correlations range from 0.517 to 0.856. The study shows that accurate and interpretable RUL prediction can be combined in one structured workflow.

D1-C1: Smart Infrastructure, Robotics & Energy AI

Design and Implementation of a Backend Verified Smart Lock Using Bluetooth Low Energy, Hash-based Message Authentication Code, and Email One-Time Password

Mohammed Majid M Al Khalidy, Ebrahim Khalil Alanzoor and Saud Naeem Al-Khal

This paper describes the design and implementation of a smart door lock aimed at mitigating the security risks typically associated with Bluetooth Low Energy (BLE)-based access control systems. The prototype relies on an ESP32 microcontroller that handles BLE communications and operates a relay, while all authentication decisions are delegated to a backend server written in Go. The locking system is implemented in two ways: the first involves a challenge-response exchange via BLE, where the server issues a one-time challenge, and the mobile application returns an HMAC-SHA256 response calculated from the user ID, the challenge, and a user-specific secret code. The second method is a one-time password (OTP) sent via email through a web gateway. The code is generated on the device, stored only as an SHA-256 hash, and is subject to limitations such as a short expiration period and a maximum number of attempts. An SQLite database records users, challenges, trusted devices, and access logs, and a web-based operations gateway allows the administrator to manage accounts and review access events. The prototype was implemented and tested on a local network, where the unlocking process was successful from start to finish.

Voice-Based Authorization for Humanoid Robot Control Using Machine Learning

Mohammed Majid M Al Khalidy and Sayed Muntadher Al Najjar

This paper proposes a machine learning-based framework for secure voice authorization, establishing a biometric gatekeeping layer for robotic control. A rigorous comparative evaluation of classical and deep learning models. Three classical algorithms, Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), are benchmarked against two deep learning architectures: 1D-CNN and CNN-LSTM hybrid. Evaluations are conducted on a hybrid dataset combining the Google Speech Commands V2 corpus with localized authorized user voice samples to simulate realistic deployment conditions. Experimental results demonstrate that among classical models, the SVM with a Radial Basis Function (RBF) kernel achieved the highest F1-score of 0.95, with a False Positive Rate (FPR) of 0.04, making it the most reliable classical solution for edge-based deployment. Among deep learning architectures, the CNN-LSTM hybrid exhibited superior robustness against stochastic environmental noise. Achieving an F1 score of 0.98 and an FPR of 0.01, outperforming the 1D-CNN in both biometric fidelity and generalization. The Decision Tree, while achieving high nominal accuracy (0.98), recorded a substantially elevated FPR, confirming that accuracy alone is an insufficient metric for security-critical authorization tasks. These findings establish that localized, real-time biometric voice authorization is both feasible and effective on edge-based robotic resources, and the SVM offers a strong lightweight alternative for resource-constrained deployments.

A Comparative Evaluation of Machine Learning for Early Warning of EV-Induced Grid Violations in Bahrain's LV Network

Mohammed Majid M Al Khalidy and Ahmed Alfardan

The rapid growth in electric vehicle (EV) adoption poses significant stress on low-voltage distribution networks, raising concerns over transformer overloading and voltage violations. Existing studies address this by using a limited number of fixed, deterministic scenarios, which limits their ability to predict violations under arbitrary operating conditions. This study proposes a machine learning-based early warning model trained on 1,200 samples simulated from Bahrain's LV network parameters. Four classifiers were evaluated: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost). SVM achieved the highest accuracy of 96.25% and recall of 98.04%, making it the most suitable model for safety-critical grid applications. Logistic Regression achieved the highest cross-validation F1-score of 96.07%, indicating strong generalization consistency. The feature importance analysis identified transformer base loading and feeder length as the dominant risk factors. The proposed model provides a practical tool for proactive grid management, enabling distribution operators to predict violations in real time.

AI-Based Energy Loss Reduction in Distribution Networks: Two-Stage Loss Forecasting and Label-Aware Electricity-Theft Detection

Alawi Sharaf and Zouhir Bahri

Electricity distribution networks lose a significant share of the power they carry to technical losses on ageing infrastructure and to non-technical losses such as meter tampering. This paper presents a software-only framework that addresses both. A two-stage model forecasts nodal load with a Long Short-Term Memory (LSTM) network and derives network loss through AC power flow on the IEEE 33-bus feeder, reaching a daily-aggregated mean absolute percentage error (MAPE) of 2.39% against a 3.29% literature reference. An electricity-theft detector is evaluated on the field-labeled State Grid Corporation of China (SGCC) dataset of 42,372 customers under three label-availability scenarios. With no labels, a per-cluster PCA-One-Class SVM and Isolation Forest ensemble reaches a ROC-AUC of 0.68, a ceiling confirmed by eleven variants. Labels for 1,137 audited customers raise precision at 40% recall from 0.50 to 0.65. With full labels, a gradient-boosted detector that retains meter-reporting-gap features reaches a ROC-AUC of 0.877, MAP@100 of 0.994, and precision of 0.94 at 12.5% recall, above the best results among the papers reviewed on the same dataset (0.784, 0.960, and 0.75). Both components run on a two-core CPU and are unified in an offline operator dashboard that reports the computational cost of every method alongside its accuracy.

Wednesday, November 18 12:20 - 13:40

D1-A2: LLM Robustness, Safety & Prompt Security

PromptExtract: A FuzzyAI Extension for Black-Box Vulnerability Assessment of System Prompt Extraction in LLMs

Husain Jameel, Salman Mohamed Jasim, Murtadha Ahmed Yahya and Abdulla Aldoseri

The integration of Large Language Models (LLMs) into enterprise workflows introduces critical data exfiltration vectors, primarily system prompt leakage. Static keyword filters and manual penetration testing fail against adaptive, template-based adversarial perturbations and conversational roleplay. This paper presents an automated, black-box vulnerability assessment framework designed to evaluate system prompt confidentiality in LLM applications. Extending the open-source CyberArk FuzzyAI architecture, we develop an asynchronous `PROMPT_EXTRACTION` attack handler that applies structured payload perturbations (character substitution, translation/encoding, and persona-adoption jailbreaks) against target endpoints. The pipeline integrates an automated "LLM-as-a-Judge" semantic verification module for scalable response classification alongside structured JSONL telemetry logging. Empirical evaluation across instruction-tuned open-weight models (Llama 3.1 8B, Gemma 2 9B, DeepSeek-R1 8B via local Ollama REST endpoints) and managed cloud APIs (OpenAI API) demonstrates prompt extraction rates between 20.0% and 59.0% on local instruction-tuned models lacking upstream application firewalls, compared to 12.0% on commercial cloud endpoints. We analyze model-specific behavioral differences, notably the roleplay resistance provided by chain-of-thought reasoning models and discuss integration pathways for pre-deployment DevSecOps pipelines.

Developing a Structured Framework for Effective Cyber Risk Managementk: A Systematic Literature Review

Osama Erfaei Mohamed, Yusuf Mothanna, Ali Hussein Saleh Zolait and Wael M El-Medany

Public and government institutions face increasingly complex cybersecurity risks driven by digital transformation, interconnected services, and evolving threat actors. Although established cybersecurity risk management frameworks exist, their direct adoption in public-sector environments is often constrained by fragmented governance, legacy systems, crossagency dependencies, and resource limitations. This study conducts a Systematic Literature Review (SLR) of cybersecurity risk management in public and government institutions by analyzing sixteen peer-reviewed studies published between 2009 and 2023. The review identifies the most frequently adopted frameworks-including ISO/IEC 27005, NIST SP 800- 30, NIST CSF, CIS RAM, and COBIT 5-and synthesizes recurring limitations that hinder practical implementation in government contexts. Based on these findings, this study proposes the Practical Cyber Risk Management Framework (PCRMF) as its primary contribution. PCRMF integrates governance alignment, structured risk lifecycle management, and assessment depth into an adaptable model tailored to public-sector constraints. The framework defines five stages-Context Formation, Risk Identification, Risk Assessment, Risk Treatment, and Risk Monitoring-to improve consistency, coordination, and risk-informed decision-making in government cybersecurity programs.

A Context-Aware, Multimodal AI Gateway for Real-Time Child-Safe Social Media Content Moderation

Hussah Adnan Almuhareb and Wael M El-Medany

The widespread adoption of social media platforms has enabled communication and learning opportunities while simultaneously exposing children to online threats such as cyber bullying, inappropriate content, online grooming, and violent ma terial. Existing content moderation approaches exhibit significant limitations in providing real-time and context-aware protection tailored to young users. This paper proposes an AI-centric gateway architecture that leverages state-of-the-art deep learning techniques to enable real-time, context-aware moderation of inappropriate content on social media platforms. The proposed architecture integrates multiple deep learning models, including transformer-based natural language processing models (BERT and RoBERTa), convolutional neural networks for image and video analysis (YOLO and EfficientNet), and vision language models (CLIP) for multimodal content understanding. Contextual attributes such as user age, behavioral patterns, tem poral factors, and social context are incorporated to dynamically adjust moderation thresholds according to age appropriateness. Theoretical analysis indicates that the proposed gateway can achieve precision rates of 94-96% and recall rates of 90-93% for multimodal content moderation, with latency constrained to below 100ms for text processing and 300ms for multimedia analysis. This work introduces a novel architectural framework for child-focused social media environments, presents a comprehen sive 12-month implementation methodology, and provides theo retical validation of expected performance metrics. The proposed solution addresses existing vulnerabilities in content moderation by enabling scalable, real-time, and threat-aware protection while adhering to strict privacy and regulatory requirements.

D1-B2: AI Optimization, Vision & Immersive Computing

Comparative study between standard Genetic algorithms and Machine-Learning assisted Genetic algorithms for smart home scheduling

Abdulla Desmal and Zahraa Sayed Mohamed

With the increasing demand for energy efficiency and sustainability in the residential sector, the integration of advanced optimization techniques is essential in modern smart homes. This paper presents a comparative study between standard Genetic algorithms and machine-learning-assisted genetic algorithms for multi-objective smart home scheduling. Two genetic algorithm configurations were developed: a standard genetic algorithm (GA) and a machine-learning-assisted genetic algorithm (ML-GA) incorporating a bagged tree model as energy demand profile preference signal. The optimization performance, convergence, constraint satisfaction, robustness, and computational cost of both configurations were evaluated. The results demonstrated that while standard GA yielded better optimization performance, ML-GA was more stable and reliable and better at satisfying the constraints, making it more suitable candidate for real-world applications in the context of smart home energy management system.

Designing Immersion: UI/UX Design and Development of a Virtual Reality Application for Cultural Heritage Tourism

Bashayer Khalifa Aldoseri, Rahaf Nabil Shehada, Amna Khalid Othman and Hadeel Alobaidy

Digital tourism on mobile Virtual Reality (VR) platforms frequently suffers from severe hardware compute bottlenecks, rendering latency, and spatial interaction friction introduced by complex, controller-bound graphical user interfaces. This paper presents the design, architecture, and evaluation of a hands-free, AI-guided Virtual Reality experience built in Unity XR and deployed on standalone mobile hardware (Meta Oculus Quest 2), reconstructing the Pearl Trail (*Masar Al-Lulu*) within Bahrain's annual Muharraq Nights (*Layaly Al Muharraq*) heritage festival, itself rooted in the UNESCO World Heritage Pearling Path in Muharraq. Physical controller-bound 2D/3D visual menus are replaced with an English-language conversational AI cultural guide, *Hakim*, whose responses are grounded in a curated, source-verified knowledge base to minimize hallucination, following retrieval-grounded generation principles. A companion low-poly asset pipeline and static-lighting optimization strategy maintains high visual fidelity while avoiding the frame-time spikes associated with simulator sickness. The system was evaluated through a three-phase test protocol (unit, integration, and user-acceptance testing), a hands-on usability study with 12 participants, and a broader perception survey of 80 respondents. The scene sustained an average of 71.8 FPS (minimum 68 FPS), loaded in 4.2 s, and recorded a 100% pass rate across 14 structured test cases. These results establish a reproducible paradigm for combining retrieval-grounded conversational AI with performance-constrained VR asset design for mobile cultural-heritage preservation.

Robustness Analysis of Semantic Segmentation Architectures Under Simultaneous Stochastic and Spatial Interference

Alauddin Yousif Al-Omary and Hessa Abdulla Albuflasa

level understanding of roads, pedestrians, vehicles, and infrastructure. This paper investigates the robustness of two widely used segmentation architectures, U-Net and DeepLabv3+, under simultaneous Gaussian noise and spatial occlusion. Using Cityscapes as the primary benchmark and a domain-shift evaluation setting with Google Street View imagery, the study examines how coupled corruption affects segmentation stability, boundary recovery, and generalisation. The experimental pipeline follows a 60/20/20 data split, corruption-aware inference, and mean Intersection over Union as the main evaluation metric. Results show that combined stochastic and spatial interference produces a measurable drop in segmentation quality, with thin structures and object boundaries affected most severely. The findings suggest that robustness in robotic perception requires not only accurate encoders, but also explicit mitigation layers that reduce sensitivity to local masking and noise accumulation. The paper proposes a denoising autoencoder front-end as a practical pre-processing module to improve resilience before semantic inference.

Assessing the Utility of Topological Filters in Graph-based Deep Clustering Models

Huda Fardan, Riyadh Ksantini and Abdulla Ebrahim

Graph deep clustering has emerged as an effective approach for discovering meaningful structures in complex datasets by jointly learning latent representations and cluster assignments. However, existing deep clustering frameworks often suffer from unreliable neighborhood selection, which can negatively affect cluster homogeneity and overall clustering performance. This paper proposes TopoFilter R-DC, a topology-aware enhancement of the Rethinking Deep Clustering (R-DC) framework designed to improve reliable sample selection while preserving its self-supervised learning strategy. The proposed approach integrates TopoFilter into the second phase of R-DC to identify structurally consistent samples using a K-nearest neighbor graph constructed from learned latent representations. By filtering weakly connected and conflicting samples before neighborhood refinement, the framework enhances clustering reliability without introducing clustering-level pseudo-supervision. The proposed method is evaluated on three graph datasets (Cora, Citeseer, and Wiki) and four image datasets (CIFAR-10, PneumoniaMNIST, BloodMNIST, and BreastMNIST). Experimental results demonstrate improvements in clustering accuracy compared with the baseline R-DC model, increasing from 74.5% to 75.4% on Cora, 71.6% to 73.5% on Citeseer, and 55.0% to 58.3% on Wiki. Furthermore, accuracy improvements ranging from 1.2 to 6.0 percentage points are observed across the image datasets. These findings demonstrate that incorporating topology-aware reliable sample selection into self-supervised deep clustering can improve clustering accuracy and representation quality across different data modalities without modifying the underlying learning paradigm.

D1-C2: Threat Detection, Phishing & Digital Forensics

OpticNet: A Unified Platform for Network Visualization, Vulnerability Discovery, and AI-Guided Remediation

Maryam Isa Abdulla Alhawi, Maryam Ali Hasan Darwish and Abdulla Aldoseri

Small and medium-sized enterprises (SMEs) face rising cyberattacks yet often lack a unified view of their assets and attack surface, relying on fragmented, expert-oriented tools. This paper presents OpticNet, an interactive, browser-based network visualization platform that combines internal network discovery (Nmap, ARP, DHCP, passive sniffing), external reconnaissance (Subfinder), and active web-application testing (Nikto, Gobuster, SQLMap, Nuclei) into one scanning pipeline. Discovered assets are scored using a three-layer risk model including matching CVEs from the National Vulnerability Database (NVD) and are rendered as an interactive graph with attack paths mapped to the MITRE ATT&CK framework. The platform, including a rule-based advisory engine and a conversational AI Security Coach powered by locally hosted large language models, is available in English and Arabic with a right-to-left interface, translating findings into prioritized, plain-language remediation guidance without sending scan results or internal topology to external services. Implemented as a Flask/MySQL web application with per-user encrypted storage, OpticNet was evaluated in a controlled Docker testbed of vulnerable hosts and in pilot scans of a 44-host institutional network and a home network, passing all 426 automated tests, detecting every planted vulnerability covered by the functional tests, reporting 218 version-matched CVE records on an 11-host test network, and achieving a usability score of 4.53/5.0 from ten evaluators. These results indicate that OpticNet addresses a persistent gap by unifying these capabilities in one accessible platform for non-experts.

PhishArena: A Controlled Interactive Phishing Simulation Platform with Hybrid LLM-Assisted Evaluation

Abdulla Aldoseri, Aldana Althawadi, Ghufran Shaikh Mohammed and Haya Alkaabi

Phishing remains a persistent cybersecurity This study investigates the following question: How can a problem because deceptive messages exploit human controlled, gamified phishing-simulation platform combine judgment and organizational trust. This paper presents realistic email interaction with deterministic and LLM- PhishArena, a controlled interactive phishing-simulation assisted evaluation while maintaining predictable behavior on prototype that combines scenario-based gameplay, a local limited local hardware? The question is intentionally framed email workflow, deterministic validation, and locally around system design and technical feasibility. No pre- deployed large language model (LLM) assessment. Users test/post-test user study was conducted; therefore, the paper compose simulated phishing messages for fabricated does not claim that the prototype has already improved organizational personas through Thunderbird or an phishing-detection performance or long-term awareness. alternative web interface. The primary mode uses local C. Contribution Postfix, Dovecot, and Maildir services, while both modes The main contributions are: (1) an integrated phishing- share a Flask-based evaluation and progression pipeline. simulation prototype that combines a Flask web application, a Deterministic checks verify case requirements before an local email workflow, fabricated target profiles, progressive Ollama-hosted model assesses qualitative properties such as cases, and flag-based progression; (2) a hybrid evaluation professionalism, realism, and completeness. Four lightweight pipeline in which deterministic validation constrains LLM models were compared; Llama 3.2:1B provided the shortest influence and provides fallback behavior; (3) a local observed response-time range of approximately 28-87 s and deployment approach using Postfix, Dovecot, Maildir, was selected for deployment. Functional, multi-device, Thunderbird, and Ollama; and (4) a technical evaluation performance, and security-oriented robustness testing covering functional behavior, local-model response demonstrated that the prototype operates as intended in the characteristics, multi-device operation, and security-oriented tested environment. The evaluation establishes technical robustness. These contributions position PhishArena as a feasibility rather than educational effectiveness, which prototype and feasibility study that can support a later remains a direction for future controlled study. educational evaluation. Keywords-phishing awareness, cybersecurity training, gamification, large language models, simulated email, usable security.

A Steganalysis Framework Based on Machine Learning for Detecting Modern Data-Hiding Techniques in Multimedia Files

Ali Hussein Saleh Zolait, Hawra Fadhel Abbas and Khaireya Husain Alhaiki

As more images, audio files, and videos are being shared online, especially nowadays, steganography is being used more often to hide secret information or messages that do not want to be found by anyone inside these files. Traditional steganalysis, which mainly refers to the methods used to detect hidden messages within files, may not detect new data-hiding techniques, especially those that use advanced machine learning-based methods. This makes digital forensic investigations even harder because this hidden evidence may continue to go undetected and become lost within the layers of the file. Current detection methods share the same limitations. They may perform poorly when tested on data from new sources, require a large amount of computing power, and mainly focus on images rather than audio and video. This paper reviews recent research on machine learning-based steganalysis and proposes a conceptual framework for detecting hidden information in different types of multimedia files. The framework combines media-specific preprocessing, feature learning, and mixed-source training to improve generalization across different multimedia sources. Although the framework has not been implemented or experimentally tested, it is expected to improve generalization, support image, audio, and video analysis, and provide a better balance between detection accuracy and computational efficiency. Overall, this work provides a foundation for future research and supports the development of more effective digital forensic tools.

A Comparative Review of Machine Learning Models for Email Spam Detection

Bashayer Husain, Wael M El-Medany and Saeed Sharif

Email spam continues to be a significant issue af fecting the security, efficiency, and trust of digital communication systems. Automated detection and filtering of spam emails have been effectively addressed through the use of machine learning (ML) techniques. In this paper we present a comparative review of recent machine learning models applied for email spam detection with emphasis on traditional, ensemble and explainable artificial intelligence (XAI) models. The studied studies were obtained from major academic data sources and published in 2023-2025 and were analyzed in terms of the dataset utilized, the feature extraction approach, classification algorithms, per formance evaluation metrics and explanation approaches. It is found that the reported performance of the reviewed models varies significantly based on the characteristics of the dataset and preprocessing techniques as well as the validation approach and methodology adopted in the studies. Hence, making fair performance comparisons between the studied studies appears very difficult due to the variable nature of the parameters under consideration. Furthermore, while ensemble learning models consistently produce robust predictive performance, the issues of class imbalance, false positive and the dataset-dependent nature remain challenges. Indeed, although the recent studies included XAI techniques such as SHAP and LIME to improve model transparency and explainability, the quantitative evaluation of the generated explanations has been relatively limited with only few attempts to examine the consistency of the explanations. In this paper we highlight the current research trend, identified research gaps and provide an overview of the future research directions with regard to the development of high-performance, transparent and trustworthy email spam detection systems.

Wednesday, November 18 14:40 - 16:00

D1-A3: Cybersecurity Governance, Risk & Cloud Forensics

Evaluating the Robustness of Quantized Open-Weight LLMs to Indirect Prompt Injection and Light-Weight Defenses

Mubarak AlZaabi

Prompt injection ranks first in the OWASP Top 10 for LLM Applications [1]. In its indirect form, instructions hidden in data that an agent reads through its tools, such as an email or a web page, can make the agent acts maliciously against its user. Open-weight models are often run locally, and to fit on limited hardware they are commonly quantized, yet little is known about how quantization affects an agent's robustness to such attacks or the effectiveness of lightweight defenses. This paper evaluates Qwen3.5-9B at two quantization levels, Q4 K M and Q8 0, on the AgentDojo benchmark, covering all 97 user tasks and 949 attacked pairs under two attacks, with and without three prompt-level defenses: spotlight, sandwich and repeat prompt. Without a defense, the 4-bit configuration was significantly more vulnerable, with attack success rates of 16.5% and 13.5% against 9.1% and 7.8% for the 8-bit build, while clean utility was similar (87.6% vs. 89.7%). Sandwich was the most effective defense, lowering attack success to 0.3% for Q8 and to 13.9% and 7.7% for Q4, whereas spotlight had little effect, and no defense measurably reduced utility. The 4-bit build remained more vulnerable under every defense. Because the two runs also differed in their evaluation harness, these results compare two complete configurations rather than isolating quantization, which we identify as the main step for future work.

Ethical Governance of Deepfake Technology in Cybersecurity: A Literature Review

Osama Erfaei Mohamed, Ali Farid Kassem, Mohammed Redha Qader and Fatema Albalooshi

Deepfake technologies-representing synthetic AI-generated audio, image, and video content-have rapidly matured into a significant cybersecurity concern. Made possible by advanced generative models like Generative Adversarial Networks (GANs), deepfakes enable realistic impersonation, the manipulation of evidence, and scalable social-engineering attacks. Such capabilities undermine basic security properties such as authenticity, integrity, non-repudiation, and trust, while also raising complex ethical questions about consent, accountability, deception, and digital manipulation. This paper presents an SLR of interdisciplinary deepfake research from technical, ethical, legal, and organizational domains, published from January 2018 to December 2024. Major academic databases were searched, returning 146 records, from which eight core studies were selected for qualitative-comparative analysis. Contributing to the above gap, the paper provides an original ethics-to-controls mapping that translates professional ethics principles into actionable cybersecurity measures, taking into account standards from ACM, IEEE, and ISC2. It proposes an integrated ethical-cybersecurity framework that brings together detection technologies, provenance verification, governance policies, and trust-preserving mechanisms. The findings offer actionable guidance to organizations on how to address deepfake risks while ensuring ethical responsibility and public trust. Index Terms

A Comprehensive Review of Digital Forensic Challenges in Cloud Computing Environments

Osama Erfaei Mohamed, Yusuf Mothanna and Wael M El-Medany

Nowadays, cloud computing is considered the backbone of many organizations' information technology services. Also, cybersecurity plays a critical role and should be a focus for the organization. One of the significant fields of cloud computing is Digital forensics. Digital forensics aids investigators in obtaining evidence from a cloud environment in the event of a cyberattack or crime. Cloud computing has unique characteristics, and cloud forensics faces several challenges when applied. Those challenges need to be addressed to perform appropriate digital forensics in the cloud environment. This study reviews cloud forensic challenges based on the defined research literature, focusing on issues that require more effort.

Synchrophasor Systems for IT Professionals: A Review of Architecture, Cybersecurity, and Research Directions in the Context of Saudi Arabia's Net-Zero 2060 Transition

Hazim Al-Sibai, Rafat Rob and Wael M El-Medany

Electric power grids are becoming more dynamic as inverter-based renewable generation replaces the synchronous machines whose inertia once made slow measurement acceptable. Synchrophasor systems, in which phasor measurement units (PMUs) stream GPS-time-stamped phasors into wide-area monitoring systems (WAMS), provide the sub-second and system-wide visibility that this transition requires. However, most of the literature on synchrophasors is written for power engineers. In this structured narrative review, we describe synchrophasor systems for information-technology (IT) and computer-science readers as a distributed real-time telemetry system, covering its architecture, protocols, data types and threat model together with the impact figures from the experimental literature. We then examine the operator confidence gap that the industry itself reports. Control rooms use synchrophasor tools for offline analysis far more readily than for mission-critical decisions, because the integrity and the timing of the data cannot be assured end to end. We interpret the gap as an information-assurance problem, we map the countermeasures that computing research can offer against it, we organize these countermeasures in a People, Process, Governance and Technology framework, and we set out a prioritized research agenda together with a catalog of datasets and tools. Throughout, the motivating context (not a field study) is the net-zero-2060 transition of Saudi Arabia.

D1-B3: IoT Security, Intrusion Detection & Federated AI

Lightweight Edge Deception with Unsupervised Risk Scoring for IP Video Surveillance: A Framework and Evaluation Methodology

Mona Fahad

IP video surveillance systems are exposed to auto mated credential attacks via their web management and streaming interfaces. Such video endpoints are often supplied by legacy edge gateways with insufficient resources for enterprise intrusion detection. Camera honeypots measure attacker browsing actions, but not in real time, while flow anomaly detectors such as Kitsune do not model attacker application layer behavior. We propose a lightweight framework that combines a minimal CCTV-console decoy, privacy-preserving login telemetry, unsupervised Isolation Forest scoring, and risk-prioritized instant-messaging alerts in a single edge gateway. Using a design science research approach, three research questions of (i) behavioral discrimination, (ii) op erator situational awareness, and (iii) edge overhead are related to seven design requirements. We specify a pseudonymized telemetry schema, a session-based 10-feature representation, a percentile 0-100 risk score with feature-level evidence, and a grouped alerting policy. We also define a reproducible evaluation methodology based on a hybrid virtual Raspberry Pi 3B testbed, five class scenarios mapped to MITRE ATT&CK brute-force sub techniques, leakage-controlled data splits, a fixed-rule baseline and a Kitsune baseline, and a six-criterion validation gate for data collection. Our contribution is the framework design and its evaluation protocol; empirical results of the framework are probably covered in future work.

Centralized vs. Federated Learning for Cyber-Attack Detection in Smart Water Systems: A Systematic Review of Reported Performance

Nurulain Zulfiqar

Smart water infrastructure is increasingly monitored and controlled through networked sensors and actuators, a shift that improves operational visibility while widening the attack surface of facilities that societies depend on. Machine learning has become the default defensive response and most of the evidence for it comes from three public benchmarks: SWaT, WADI, and BATADAL. This paper presents a systematic review of thirty studies to ask a question the individual papers cannot answer: how do reported results actually compare across model families and training settings, and can they be compared at all? Three findings stand out. First, centralized studies report accuracy or F1 above 96%, and one reporting a perfect 100% on BATADAL, while federated and cross-site studies on the same data report F1 between roughly 92% and 96%. Second, almost all of this gap traces back to evaluation choices rather than learning settings, and studies differ in whether they split data chronologically, which metrics they lead with, how they count attacks and whether they score per attack or in aggregate. A benchmark that enforces chronological split obtains an F1 of 0.509 on SWaT despite test accuracy above 0.93 and another model reports 90% accuracy with an attack recall of 0.22. Third, federated learning evaluations rarely include a centralized upper bound on an identical data partition, and communication cost is rarely reported, which prevents the accuracy-versus privacy trade-off from being quantified. This paper concludes with evaluation protocol intended to make future centralized and federated results directly comparable.

Risk-Aware IoT Intrusion Detection: Integrating Machine Learning with NIST and ISO 27001 Compliance Controls

Turki Al Ali and Wael M El-Medany

A number of intrusion detection methods that use machine learning techniques have been developed to identify and analyze attacks in IoT networks; however, the performance of such methods is normally assessed using accuracy, precision, recall, and F1-score measures only. The identified attack does not have any tag; thus, it does not indicate what type of assets or network segment has been affected, what the potential consequences are, and what the security framework used. There are several security frameworks, including NIST and ISO 27001, that specify the controls for monitoring incident response and risk treatment. Yet intrusion detection research rarely links its outputs to them, and compliance evidence is often compiled manually after an incident. This paper provides a solution that translates machine learning detection results into risk levels, mapping them against the NIST and ISO 27001 controls. The research is conducted in accordance with the Design Science Research methodology, including a quantitative experiment. For training and comparing the models-namely, Random Forest, XGBoost and Neural Network-a publicly available benchmark data set is used. Attack class prediction, model confidence level, and criticality of the asset become inputs to the risk score calculation algorithm where the risk level will be translated to controls depending on the impact on confidentiality, integrity and availability of the attack. This study is to highlight the performance of the chosen models on IoT traffic, how the risk scoring and mapping output add value to compliance monitoring beyond alerts, and what is the limitation of the technique is due to benchmarking. Contributions include evaluation of detection models, risk scoring of detection models' outputs, mapping of detection models' output to compliance controls, and a framework to make detection models' outputs usable for risk management and compliance monitoring.

The Key to Secure IoT: Effective Authentication and Access Control Solutions

Taif Nabeel Shehada, Maryam Abdulghani Hasan Saleh and Rawan Husain

Authentication and access control are difficult to im plement consistently across Internet of Things (IoT) deployments. Devices vary widely in processing capacity, connectivity, and lifetime, while permissions often depend on changing users and operating conditions. This paper reviews common authentication methods and access control models, then examines multi-factor authentication, blockchain-based approaches, and context-aware authorization as possible complementary controls. A qualitative synthesis of the literature identifies recurring trade-offs involving security, latency, scalability, policy maintenance, and energy use. Based on these findings, we outline a layered design that places lightweight device authentication near endpoints and reserves more demanding identity management and policy decisions for gateways or supporting services. The design is a recommendation derived from published work; no prototype or experimental evaluation is claimed. The paper closes with practical deployment considerations and priorities for future validation.

D1-C3: Intelligent Communications & Secure AI Applications

Intelligent FIFO Sizing for FPGA Based High Energy Physics Data Acquisition System

Amna Mohammed Abdulla, Mohammed Majid M Al Khalidy and Mehmet Özgür Sahin

First-in, first-out (FIFO) buffers play an important role in FPGA-based data acquisition systems (DAQ) used in high-energy physics experiments with the maximum expected data rate reaching up to 20Tb/s. Insufficient buffering causes critical data loss, while excessive buffers may unnecessarily consume valuable hardware resources. This work investigates intelligent design-space exploration to optimize FIFO configurations in the readout firmware of a high-throughput particle detector DAQ system. It applies intelligent optimization techniques such as Bayesian Optimization (BO) and Tree-structured Parzen Estimator (TPE) to optimize FIFO configurations. These configurations are evaluated using an RTL-based simulation environment that measures packet loss and latency. This paper also introduces a mathematical estimator for LUTRAM and BRAM utilization within UltraScale+ VU13P FPGAs. We compare these optimization techniques against exhaustive grid search and random grid search to assess their effectiveness and feasibility. Exhaustive grid search was used as the baseline as it evaluated all 140 candidate configurations and identified a maximum reward of 0.949. It required a total wall-clock time of approximately 5 hours. In contrast, BO and TPE were able to obtain an optimal configuration within the 25-evaluation search budget, reducing the number of required evaluations by approximately 82%. The robustness of these techniques was evaluated across 20 random seeds. TPE achieved this highest reward on 55 % of the total number of seeds, whereas BO did so in 40% and Random Search in 30 % of the runs. These results demonstrate the potential of intelligent optimization techniques for RTL-based design space exploration of FPGA DAQ firmware.

A Semantic Communication Framework for 6G Using Edge-Deployable Large Language Models

Ahmad Abdulkareem Jaber, Mohab A Mangoud M, Saeed Serwan Abdulsattar and Fayzeh Abdulkareem Jaber

Semantic communication is a promising paradigm for sixth-generation (6G) networks because it shifts the objective of transmission from exact bit recovery to preservation of meaning. This paper proposes an end-to-end semantic communication framework that combines edge-deployable large language models (LLMs) with a shared hierarchical symbolic codec for compact packetized transmission. Rather than transmitting raw text or dense learned latent features, the proposed system first reduces the source message into a compact semantic phrase using a lightweight local LLM and then maps that phrase into phrase, word-, and character-level symbolic identifiers using shared semantic knowledge available at both transmitter and receiver. The resulting identifiers are packetized and transmitted, after which the receiver decodes the symbolic sequence and uses a lightweight restoration model to reconstruct a fluent sentence that preserves the intended meaning. Experimental results show that the symbolic codec alone reduces the average transmitted bit cost from 303.47 to 154.80 bits, corresponding to a bandwidth saving of 49.10%, while maintaining near-perfect task success of 0.992. The complete framework further reduces the transmitted bit cost to 114.40 bits and increases bandwidth saving to 62.01%, with a packet compression ratio of 3.004, while still achieving a task success rate of 0.883. Robustness analysis under byte level channel corruption further shows consistent performance degradation under increasing impairment. Overall, the proposed framework demonstrates that shared symbolic knowledge and compact local LLMs can provide a practical and transmission efficient pathway toward meaning-aware communication in future 6G edge environments.

AI-Driven Automated Incident Response in Zero Trust Environments Using Artificial Neural Networks

Mohamed Hasan and Jenan Hasan

Modern cybersecurity environments generate massive volumes of telemetry that cannot be handled effectively by manual or rule-based incident response. Zero Trust Architecture (ZTA) strengthens security through continuous verification; however, it also increases the volume and urgency of real-time access decisions. This paper proposes an Artificial Neural Network (ANN)-driven automated incident response framework aligned with Zero Trust principles. A lightweight feed-forward ANN with two hidden layers (32 and 16 neurons) is trained on the CICIDS2017 dataset to classify network flows into benign traffic and specific attack classes. The predicted class and its confidence are combined with identity, device, and session context inside a Zero Trust policy engine, which assigns a risk level and selects a proportional response such as source blocking, endpoint isolation, multi-factor authentication enforcement, or continued access under least privilege. The model was evaluated on all eight CICIDS2017 traffic files, covering denial-of-service, distributed denial-of-service, port scanning, brute-force, web, botnet, and infiltration attacks. It achieved accuracy between 98% and 99% and macro F1-scores between 0.98 and 0.99, which improves on multilayer perceptron results previously reported on the same dataset while remaining lightweight enough for near real-time scoring. These findings show that ANN-based classification can act as an effective decision engine for automated, policy-driven incident response in Zero Trust environments.

Thursday, November 19

Thursday, November 19 9:00 - 10:40

D2-A1: Cybersecurity, Phishing & Adversarial AI

DeepFakeGuard-Text: A Lightweight Transformer Detector for AI-Generated Text and an Evaluation of Embedding-Space SMOTE

Bayan Alfarrayeh, Mais Alashqar, Marya Hani Sabeha, Farah Ghassan Abu Al-Soud and Qasem Abu Al-Haija

The rapid evolution of Large Language Models (LLMs) has democratized the creation of highly sophisticated, human-like text, giving rise to an unprecedented proliferation of synthetic misinformation and textual deepfakes that easily bypass traditional statistical or rule-based detection systems. This paper introduces DeepFakeGuard-Text, an optimized transformer-based security layer for detecting AI-generated text. The system leverages a fine-tuned 6-layer DistilBERT core with approximately 66 million parameters to extract dense, context-aware 768-dimensional semantic representations. To resolve real-world dataset skewness, a Synthetic Minority Over-sampling Technique (SMOTE)-based oversampling step (k = 5) is integrated directly within the latent embedding space to enforce an equitable 1:1 class balance across an aggregated corpus of 13,727 samples. Comprehensive hyperparameter optimization establishes a stable operating profile using an initial learning rate of 3 × 10−5 and a batch size of 16. A strict 5-fold cross-validation protocol verifies model stability, producing an average training accuracy of 96.48% (±1.01%) and a mean ROC-AUC of 99.73% (±0.08%). On an independent, locked hold-out partition, the finalized model achieves a peak validation accuracy of 96.83%, a weighted F1- score of 96.85%, a human-text recall of approximately 96%, and a synthetic-text sensitivity of approximately 99%. Empirical resource profiling records an end-to-end processing latency of 24.2 milliseconds per document in the training environment, suggesting feasibility for high-throughput deployment, pending validation on dedicated edge hardware. These results demonstrate that pairing a distilled transformer encoder with latent- space class balancing yields a lightweight yet accurate defense module suitable for integration into larger multimodal deepfake-detection pipelines.

Interpretable Phishing Email Detection Using Concept Bottleneck Models

Yara Said Abdallah, Ali Ballout and Nazih Salhab

Text classifiers can detect phishing emails accurately but often provide little insight into the cues underlying their decisions. We investigate a sequential Concept Bottleneck Model (CBM) that first predicts 20 phishing-related concepts and then classifies an email using only the predicted concept vector. Meta-Llama-3-8B-Instruct provides structured concept annotations for 15,270 emails from the curated CEAS-08 dataset. We compare the CBM with TF--IDF text-only classifiers and classifiers trained on the LLM-generated concepts. The strongest text-only neural network achieves an F1-score of 0.996 and a ROC-AUC of 1.000; the sequential CBM achieves 0.975 and 0.996, respectively. The best concept-only classifier obtains an F1-score of 0.951 and a ROC-AUC of 0.980. Thus, the concept bottleneck retains substantial predictive information while exposing an verifiable intermediate representation, although concept prediction remains imperfect and the annotations have not been manually validated at scale.

An HDFS-Spark-CNN Workflow for Malware Family Classification: A Retrospective Case Study

Afrin Nanmi, Saeed Sharif, Khawaja Daim Asif Ansari and Kimiya Sadeghi Dastjerdi

Malware-family classification is both a learning problem and a systems problem: growing repositories demand accurate models and efficient data processing. This retrospective case study documents an implemented workflow that converts executable binaries into normalised byteplot images, stores them in the Hadoop Distributed File System (HDFS), and uses Apache Spark with Hadoop YARN (Yet Another Resource Negotiator) to coordinate multi-node training of a convolutional neural network (CNN). We report the recorded results of three experiments: a standalone CNN configuration study and a support vector machine (SVM) baseline on the Malimg benchmark, and a one-, two-, and three-node comparison of a distributed CNN (DCNN) on a separate VirusShare-derived dataset. The best recorded CNN and SVM configurations reached 99.74% and 93.6% accuracy. In the node comparison, recorded runtime decreased from 63 to 27 minutes (runtime ratio 2.33×; 57.1% reduction), and recorded accuracy was 99.59%, 98.52%, and 97.79%. Because equivalent workloads and training settings cannot be verified from the available records, and only one result per configuration was available, these measurements describe the configurations evaluated and do not establish controlled scaling performance. The evaluation partitions underlying the reported accuracies could not be established, so these values should not be interpreted as verified held-out test performance.

Adversarial Robustness of Large Language Models in Misinformation Detection: A Comparative Eval-uation Using the Adversarial Prompt Evaluation Framework

Saeed Sharif, Afrin Nanmi, Jitendra Seelankala and Wael M El-Medany

Misinformation, increasingly prevalent in digital media, poses a serious threat to public knowledge and democratic institutions. Despite encouraging progress in automated misinformation detection using Large Language Models (LLMs), these models' vulnerability to adversarial framed conversational pressure, particularly across multi-turn interactions resembling real-world use, remains poorly investigated. This paper presents the Adversarial Prompt Evaluation Framework (APEF), a four-stage framework for evaluating the long-term resilience of an LLM under sustained adversarial prompting. Two architecturally dissimilar models, GPT-4o (OpenAI, commercial) and Llama 3.1 8B Instruct (Meta, open-source), were evaluated under four experimental conditions, baseline classification, authority-bias framing, emotional-appeal framing, and self-correction mitigation, on 182 news articles (94 real, 88 fake) drawn from the WELFake dataset. Across the adversarial conditions, the generated variants were designed to preserve the core article claims while modifying rhetorical framing, emotional tone, and source-attribution cues. GPT-4o achieved higher baseline accuracy than Llama 3.1 8B (66.5% vs. 62.1%) but showed greater classification instability under adversarial prompting: its Adversarial Flip Rate (AFR), the percentage of articles whose classification differed between Turn 1 and Turn 5, reached 55.0% under emotional-appeal framing, compared with 28.6% for Llama under the same condition. Within the balanced judge-evaluated subsample, a confidence paradox was observed: judge-assessed certainty increased from Turn 1 to Turn 5 even in some cases where the final classification was incorrect. Self-correction prompting at Turn 5 partly recovered accuracy for both models, suggesting it as a lightweight prompting-based mitigation strategy, though its effectiveness varied by model.

D2-B1: AI for Healthcare & Clinical Decision Support

Quality-Aware and Risk-Sensitive Deep Learning for Selective Pneumonia Decision Support Under Domain Shift

Muhammad Uzair Khan, Afrin Nanmi and Saeed Sharif

Good in-domain classification performance does not guarantee reliable pneumonia decision support when chest X-ray image characteristics vary, confidence is imperfectly calibrated, or deployment data differs from development data. This paper presents a selective decision-support framework based on DenseNet-121 with weighted binary cross-entropy, a training-only-fitted blur/brightness/contrast image-characteristics proxy, temperature scaling, Monte Carlo Dropout uncertainty, and a rule-based workflow for acceptance, review, escalation, review-before-rule-out, or recapture. On the held-out RSNA test set, the selected model achieved accuracy 0.8171, recall 0.7151, F1 score 0.6380, and AUROC 0.8835. The workflow automated 37.42% of cases, limited unsafe acceptance of true pneumonia cases to 4.10%, and routed 76.05% of true pneumonia cases to review or escalation. Temperature scaling produced statistically detectable but modest calibration improvement, while predictive uncertainty was substantially higher for incorrect predictions. Controlled corruption analysis showed that Gaussian noise caused the largest F1/AUROC degradation. Partial external testing on NIH ChestX-ray14 showed recall falling to 0.4306; a harder pneumonia-versus-other-abnormality cohort reduced AUROC to 0.5834 and F1 to 0.1310. These results support selective automation with explicit human review under uncertainty and domain shift, rather than a clinically validated diagnostic system.

Deep Learning for Ultrasound Thyroid Nodule Classification: A Comparative Study

Mir Hasnain Hyder, Mir Allahyar Khan Talpur, Ramesh Kumar Ayyasamy, Mubariz Rehman, Syed Saad Ahmed and Hamid Ur Rehman

Deep learning has emerged as a promising approach for thyroid nodule classification from ultrasound images. However, many studies primarily report performance on internal test sets derived from the same source as the training data, providing limited evidence of model reliability under independent acquisition conditions. This study analyzes the external generalization capability of six ImageNet architectures, namely MobileNetV2, EfficientNet-B0, ResNet-50, DenseNet-121, ConvNeXt-Tiny and Swin Transformer-Tiny, for binary thyroid nodule classification. A leakage-safe version of the TN5000 dataset containing 4,874 images was created through exact-duplicate auditing and used for model development, while external validation was performed on the complete TN3K/TNCD cohort of 3,493 images without retraining, threshold adjustment, or recalibration. Operating thresholds were selected on the TN5000 validation set using Youden's J statistic and applied unchanged during both internal and external evaluation. ConvNeXt-Tiny achieved the highest external performance with a ROC-AUC of 0.790, while DenseNet-121 produced comparable results with a smaller generalization gap. All evaluated models showed lower performance on the external cohort, indicating that internal testing alone may overestimate deployable performance. Calibration analysis further revealed that probability reliability remained affected after cross-dataset transfer despite post-hoc calibration. These findings highlight the importance of external validation, leakage auditing, locked-threshold evaluation and calibration-aware assessment for developing reliable thyroid ultrasound artificial intelligence systems.

Beyond Marginal Coverage: Auditing Conformal Prediction for Medical VQA

Ahmed Mellit, Achraf Cohen and Hafssa Benaboud

Conformal prediction can turn a frozen medical vision-language model into a set-valued predictor with a finite-sample marginal coverage guarantee. We ask what that guarantee leaves unexamined by varying the evaluation protocol around three post-hoc scores (LAC, APS and RAPS) over 20 re-splits of OmniMedVQA and VQA-RAD, with Qwen2.5-VL-3B and Qwen3-VL-4B as frozen backbones. On OmniMedVQA, a head that sees only the answer strings reaches 95.3% top-1 accuracy, so on this sample high accuracy and small sets are weak evidence of visual reasoning. Choosing hyperparameters on the calibration set lowers coverage by up to 1.3 percentage points. On binary VQA-RAD, tuning the temperature leaves LAC's sets unchanged, whereas APS without its internal randomization returns both options for every question, because a third or more of its calibration scores sit at the score's maximum. On Qwen2.5-VL-3B, near-nominal marginal coverage hides an ultrasound coverage of 0.709 and option-order effects, and the undercovered groups change with the backbone. The findings concern the evaluated models and corpora, and coverage under our stratified splits is empirical rather than guaranteed.

Gated CNN-ViT Fusion for Five-Class Diabetic Retinopathy Grading: Benchmarking Ten Backbones and a Hybrid ResNet50-ViT-Base-16 Model

Md. Sifatuzzaman Mia, MD. Refat Sarder, Nabonita Halder Bina, Md. Sabbir Ahamed, Abdullah Al Faiyaz Sina and Md. Sadekur Rahman

Diabetic retinopathy (DR) is a leading cause of preventable vision loss, and automated severity grading of fundus photographs can relieve screening workloads. Hybrid CNN-transformer models, however, are often built from ad hoc backbone choices and rarely tested for statistical significance. This paper benchmarks five CNNs (ResNet50, DenseNet121, ConvNeXtTiny, EfficientNetB0, EfficientNetB3) and five vision transformers (ViT-Base-16, Swin-Tiny, DeiT-Small, MaxViT-Tiny, CaiT-XXS24) on the 3,662-image APTOS 2019 dataset, each as a pre-trained (frozen) backbone and after fine-tuning of its last 25% of layers or blocks. ResNet50 and ViT-Base-16 are selected as the best CNN and the best ViT, and a gated adaptive fusion of their 256-d and 768-d embeddings is proposed. The fusion model reaches accuracy 0.891, macro-F1 0.785, quadratic weighted kappa (QWK) 0.934 and macro ROC-AUC 0.975 on the test set (five-seed mean ± SD: 0.881 ± 0.008, 0.774 ± 0.011, 0.930 ± 0.006, 0.973 ± 0.002). Bootstrap intervals, paired tests and ablations show that its gains over the best single backbone and a training-free ensemble are modest and not statistically significant, while 70% of its errors are between adjacent grades.

D2-C1: Next-Generation 5G/6G & Digital Twin Networks

Safety-Constrained Adaptive Deep Reinforcement Learning for PRACH Overload Control in 5G-Advanced and Beyond Networks

Fawaz Yaseen Abdallah and Saad Ahmed Ayoob

The 5G NR Physical Random Access Channel (PRACH) is overwhelmed when massive machine-type communications (mMTC) bursts lead to collisions, delays, and energy consumption. The contemporary control approaches treat parameters separately, resulting in the loss of full protection for all service classes. The work formalizes a constrained Markov decision process (MDP), which jointly maximizes PRACH resources, class access probabilities, and preamble partitions. In this work, we thus develop and run the Adaptive Deep Reinforcement Learning Algorithm for Safety-Constrained Operation Over Communications (ADR-SOOC) with a hybrid proximal policy optimization controller. These control actions are interconnected with key network performance metrics and costs through a multi-class analytical model. The performance of the algorithm is tested under steady, bursty, and reserved traffic conditions for a combined two-step and four-step random-access procedure. Both statistical and real-time checks ensure reliable system operation.

Predicting Optimal Data Transmission Strategies in IoV: A Comparative Study of Edge Processing over 5G, Wi-Fi, and LoRaWAN

Omar Dario Delgado Brito, Alexandra Macarena Flores, Víctor José Huilca and Miguel Angel Osorio

The rapid proliferation of the Internet of Vehicles demands highly efficient data processing and transmission strategies to manage the exponential surge in telemetry. Relying on static routing protocols in hybrid urban environments often leads to latency bottlenecks or rapid energy depletion. In this paper, we propose an intelligent, Machine Learning-driven orchestration framework that dynamically selects the optimal combination of edge processing strategy and wireless network technology based on instantaneous physical context. By simulating highly dense urban vehicular scenarios grounded in hardware-accurate energy and delay profiles, we evaluate the operational limits of each technology. Our findings demonstrate that semantic compression via Data Fusion allows narrow-band LPWANs to evade massive collision failures, achieving an effective application throughput comparable to broadband networks. To govern the trade-off between computational latency and transmission energy, we deployed a Random Forest predictive orchestrator. The model achieved over 91% accuracy in inferring the optimal transmission strategy in real-time based on payload volume and distance. This dynamic AI orchestration significantly outperforms traditional static rules, paving the way for highly efficient, scalable, and heterogeneous smart city ecosystems.

Tail-Risk and Medical-Digital-Twin-Aware Predictive Orchestration for Hyper-Reliable 6G Healthcare Slices

Jihane Benbrik, Houssam Benali, Asmaa Kharouaa, Aya Bouri, El Miloud Ar-Reyouchi and Kamal Ghoumid

Future sixth-generation (6G/IMT-2030) healthcare services are expected to combine hyper-reliable and low-latency communication with edge intelligence and continuously synchronized medical digital twins. Mean-oriented resource allocation can nevertheless hide short but severe under-service episodes, while a stale twin can remain unsafe even when average throughput is acceptable. This paper develops a risk-aware predictive slice orchestrator, termed RACE-MDT6G, that jointly controls the upper tail of critical medical resource deficit and the Age of Digital Twin (AoDT). The controller uses a strictly causal onestep forecast, residual-based uncertainty, a Conditional Valueat-Risk (CVaR) excess signal, clinical urgency, and critical-twin age to trigger a bounded reserve that is released only to underprotected medical slices. Feasibility and linear score-construction complexity are established. A reproducible synthetic multi-slice study uses randomized, correlated traffic with abrupt medical bursts and compares demand-aware, priority, backlog-aware, and predictive-reserve baselines. At normalized load 1.15, the proposed method lowers the 90% CVaR of critical deficit from 0.347 for the strongest predictive baseline to 0.276, increases critical satisfaction from 0.920 to 0.959, and reduces mean critical AoDT from 1.513 to 1.202 slots. Paired Wilcoxon tests give $p < 10^{-6}$ for the three primary comparisons. The results are orchestration-level proof of concept, not clinical validation or certified 6G latency guarantee .

Age-of-Digital-Twin and Carbon-Aware Scheduling for Sustainable 6G Smart-City Internet of Things

Aya Bouri, Asmaa Kharouaa, Jihane Benbrik, Houssam Benali, El Miloud Ar-Reyouchi and Kamal Ghoumid

Future sixth-generation (6G) smart cities will depend on dense sensing infrastructures whose updates feed digital twins for public safety, mobility, energy, and environmental services. Scheduling only by packet age or battery state ignores two increasingly important dimensions: synchronization fidelity of the digital twin and the carbon intensity of the energy path serving a device. This paper develops a lightweight event-aware scheduler, termed TWIN-GREEN6G, that jointly uses Age of Digital Twin (AoDT), causal event-intensity prediction, channel opportunity, battery margin, energy harvesting, transmission cost, and site-specific time-varying carbon intensity. Unlike a common carbon scalar that would cancel from device ranking, the proposed model assigns carbon intensity by gateway/site and includes a lexicographic high-risk guard before the multi-objective score. A reproducible synthetic evaluation models 24 devices distributed across four sites and compares round-robin, Age of Information, energy-aware, AoDT-aware, and carbon-aware baselines. At emergency-burst probability 0.08, the proposed method achieves event-weighted AoDT 4.014, critical-event delivery 0.728, battery-unavailable fraction 0.024, and normalized carbon cost 0.0312 per successful update. Relative to the AoDT baseline, this corresponds to 4.9% lower AoDT, 35.4% higher critical delivery, 88% lower battery unavailability, and 3.7% lower carbon cost. EnergyAware remains slightly stronger in critical delivery and CarbonAware in carbon cost, confirming a multi-objective Pareto interpretation rather than artificial universal dominance. The study is a proof of concept, not a municipal 6G deployment.

Thursday, November 19 10:40 - 12:20

D2-A2: Privacy-Preserving AI & Security Analytics

Attack Advantage Can Measure Class Ratio Rather Than Leakage: A Forensic Audit of Synthetic Tabular Data Evaluation

Afrin Nanmi, Chukwuemeka Iheanyichukwu Onyenokporo, Kimiya Sadeghi Dastjerdi and Saeed Sharif

Empirical privacy evaluation of synthetic tabular data is commonly summarised using attack-based metrics, but these metrics can become misleading when their baselines do not match the evaluation population. This paper presents a forensic measurement audit of recorded CTGAN and Copula GAN evaluations on the KDD Census-Income dataset, using preserved outputs without retraining either synthesiser. The audited pipeline reports a Membership Inference Advantage of 0.199989 for all four configurations. However, the attack classifier labels every evaluated record as a member, yielding recall 1.000 and balanced accuracy 0.500; its accuracy of 0.699989 is therefore exactly the member proportion of the imbalanced attack set. Re-scoring the same predictions using the standard prevalence-invariant definition, TPR−FPR, gives an advantage of 0 for all four runs, and accuracy lift over the majority baseline is likewise 0. Attribute inference shows the converse baseline problem: for identical predictions, model separation is 0.002418 under majority-baseline accuracy but 0.276400 under prevalence-invariant scoring, a 114-fold difference in measurement scale rather than leakage. Fidelity analysis further shows that CTGAN's 3.63-point aggregate lead is driven entirely by 2 of 41 scored columns, while Copula GAN is marginally ahead across the remaining 39. The audit also identifies that predictive utility and attribute inference reuse the same estimator and predictions. These findings demonstrate measurement defects in the audited pipeline rather than evidence of privacy leakage or model safety, and motivate baseline-aware, per-column and estimator-independent reporting for synthetic-data evaluation.

Joint Evaluation of Non-IID Robustness and Poisoning Vulnerability in Federated Medical Image Classification

Kimiya Sadeghi Dastjerdi, Saeed Sharif, Afrin Nanmi and Wael M El-Medany

Federated learning (FL) allows institutions to train a shared model without moving patient data beyond local sites. In practice, however, clinical FL must cope with both non-identically distributed (non-IID) data and adversarial client updates. We study these challenges together in a Flower/PyTorch simulation using ResNet-18 on BloodMNIST and ISIC 2019. The framework evaluates FedAvg, FedProx, focal loss, symmetric cross-entropy (SCE), inverse-loss adaptive client weighting, label-flip poisoning, and gradient-scaling poisoning across five Dirichlet heterogeneity levels. In the heterogeneity-sweep series, reducing α from 0.5 to 0.1 lowered FedAvg accuracy by 24.0 percentage points on BloodMNIST and 12.8 points on ISIC 2019. Under clean conditions, the combined Focal+SCE+Adaptive configuration improved accuracy by 3.6 and 3.0 points, respectively. Under a single-client label-flip attack (1 of 5 clients), however, it scored 5.0 points below FedAvg under the same attack on BloodMNIST. A single gradient-scaling attacker with a scale factor of 10 reduced BloodMNIST accuracy from 82.1% to 7.1-8.3%. These single-seed, five-client experiments show that better performance on clean non-IID data does not necessarily translate into poisoning robustness. Heterogeneity and poisoning defenses should therefore be assessed both separately and together before deployment.

FL-TVDP: Threshold-Verifiable Secure Aggregation with Distributed Differential Privacy for Collusion-Resilient Federated IoT Intrusion Detection

Muhammad Zunnurain Hussain, Muhammad Zulkifl Hasan and Saheef Abbas

Federated learning (FL) lets distributed IoT gateways train a shared intrusion-detection model without centralizing raw traffic. The exchanged updates still leak private data through gradient inversion attacks (GIAs). Additive secretsharing schemes such as FL-SQMPC hide individual updates at low cost, but only when the aggregators do not collude. Their protection is all-or-nothing: it fails once a coalition of aggregators reaches the reconstruction threshold, it places no formal bound on what the released aggregate leaks, it cannot detect an aggregator that tampers with the sum, and it breaks when a participant drops out. This paper proposes FL-TVDP, a secure-aggregation protocol that removes these limitations through three changes to the privacy layer. First, a packed (t, n) threshold secret-sharing scheme over a prime field hides each update against any coalition of at most t−1 aggregators and reconstructs the aggregate from any t of n shares, which separates the collusion margin from the dropout margin. Second, a distributed differential privacy (DDP) layer has each client add a calibrated discrete-Gaussian noise share, so the released model carries a certifiable (ε, δ) guarantee at about 1/√n of the perclient utility cost of local DP. Third, a verifiable-aggregation layer built on linearly homomorphic commitments lets the server detect a malicious aggregator. We evaluate FL-TVDP on Bot-IoT, ToNIoT, and CICIoT2023 under IID and non-IID partitions against a TabLeak adversary. FL-TVDP matches vanilla FL utility within noise, keeps TabLeak Acc@20% low even after the secret-sharing threshold is breached, tolerates 40 to 50% client dropout with little utility loss, and runs one to two orders of magnitude faster than CKKS homomorphic encryption.

MediGuard: A Minority-Aware, Cost-Sensitive, and Interpretation-Consistent Lightweight Framework for Internet of Medical Things Intrusion Detection

Muhammad Zunnurain Hussain, Muhammad Zulkifl Hasan, Saheef Abbas and Huzaifa Bin Munir

Internet of Medical Things (IoMT) introduces numerous network attacks to hospitals and remote-care systems. A device dropped there must be reliable and cost-effective to operate, and an analyst must be able to review the data it alerts on. The research on the CIC-IoMT-2024 benchmark shows an accuracy of 99.72% with a tuned Random Forest model and 10 features selected by the impurity, while the macro-averaged F1 is stuck at 92.64% due to the poor performance of the minority classes of attack types, such as Recon-VulScan (F1=0.56) and OS-Scan (F1=0.77). These failures are masked by high accuracy on a highly skewed problem with 19 classes. We revisit the pipeline and propose MediGuard, a framework based on five fixes: an in-fold hybrid of Borderline-SMOTE and inverse-frequency cost weighting for which resampling the training set by bootstrapping (to avoid the biased single run Gini importance) and a consistency check across interpretations, to measure interpretation agreement from permutation, SHAP, and LIME rankings, and a light audit reporting model size, peak memory and per-instance latency. On CIC-IoMT-2024, MediGuard achieves the success of macro-F1 94.99%, the worst-class F1 0.71%, and accuracy 99.70%, with the model remaining compact enough to be deployed at the edge. Confirmed by Wilcoxon signed rank test, macro gains are significant.

D2-B2: Computer Vision & AI for Real-World Monitoring

YOLOv11-Based Escalator Hazard Prevention Using Dual-Camera Monitoring and Automated Audio-Visual Alerts

Joseph Mark M. Centeno, Tristan Kyle A. Siazon, Angelo C. Tagalag and Gerald P Arada

Escalator incidents involving wheelchairs, strollers, and wheeled luggage can create hazards in high-traffic public spaces, yet conventional signage and CCTV monitoring provide limited proactive intervention. This paper presents a real-time monitoring and advisory system that uses YOLOv11 to detect prohibited bulky objects before escalator entry and automatically activate object-specific audio and visual warnings. The system processes one or two CCTV streams on a personal computer, validates detections within a predefined region of interest using temporal persistence logic, and sends User Datagram Protocol (UDP) commands to an ESP32-based warning unit. A custom dataset containing more than 10,000 images was assembled from open-source datasets, CCTV-style recordings, and extracted video frames for three classes: luggage, strollers, and wheelchairs. The trained model achieved a mean average precision at 0.5 intersection-over-union (mAP@0.5) of 0.951, with class recalls of 90.01%, 97.34%, and 96.04%, respectively. Camera-placement and trigger-distance experiments identified a 1.5-m trigger zone as the best tested operating point. Integrated tests achieved overall success rates of 100.00%, 94.67%, 88.89%, and 97.78% for single-object, two-object, three-object, and non-target scenarios, respectively; crowded trials with prohibited objects achieved 95.56%. The results support the feasibility of a non-intrusive escalator safety advisory architecture that does not require modification of the escalator control hardware.

CCTV Integrated with Thermal Sensors for Fire and Person Detection Using YOLO11n with Real-Time Notification Alert System

Ellis Dominic P. Chua, Tobi Andrei Q. Laher, Miguel Tristan Mortega, Gabrielle C. Simbulan and Gerald P Arada

Early fire detection and timely emergency notification are essential for reducing property damage, supporting occupant evacuation, and improving emergency response. This study presents a proof-of-concept intelligent CCTV surveillance system that integrates thermal fire detection, optical person detection and counting, audible alarm activation, and automated SMS notification within a multi-camera edge-computing architecture. Separate YOLO11n models were developed for fire and person detection using approximately 5,000 thermal fire images and 25,000 optical person images, respectively, and were deployed on Raspberry Pi 5 devices. The fire model achieved an mAP50 of 96.73% and an mAP50:95 of 75.35%, while the person model achieved 71.23% and 45.67%, respectively. Under controlled field-testing conditions, Camera 1 and Camera 2 achieved fire detection accuracies of 95.56% and 97.78%, respectively, while person detection accuracies reached 91.67% and 90.00%. For confirmed Case 3 fire events, the notification subsystem achieved 100% correct SMS and alarm activation, with a mean end-to-end SMS response time of 8.40 +/- 2.46 s. The centralized maximum-count rule also matched the ground-truth occupant count in all eight evaluated multi-camera trials. The results demonstrate the feasibility of integrating thermal confirmation, optical occupant monitoring, edge-based detection, and automated emergency notification in a unified real-time fire surveillance platform, while also highlighting limitations related to thermal-image resolution, obstruction, and visually similar non-fire heat sources.

AppleDisease-1K: Benchmarking Deep Learning Models for Reliable Healthy-Infected Apple Classification

Mir Allahyar Khan Talpur, Ramesh Kumar Ayyasamy, Mubariz Rehman, Mir Hasnain Hyder, Kesavan Krishnan and Roshan Kumar

Accurate identification of apple diseases is important for crop quality assessment and the reduction of production losses. However, deep learning models are often evaluated using images acquired under controlled conditions, while less attention is given to prediction reliability when natural variations in image acquisition are present. This study introduces AppleDisease-1K, a dataset of 1,045 original apple images collected from natural market environments and categorized as healthy or infected. Six ImageNet pretrained convolutional neural networks, VGG16, ResNet50, DenseNet121, MobileNetV2, EfficientNetB0 and InceptionV3, were evaluated using consistent data partitions, preprocessing and training settings. Performance was assessed using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, Average Precision, Brier Score and Expected Calibration Error. VGG16 achieved the highest accuracy of 98.57%, precision of 100.00% and F1-score of 98.48% while DenseNet121 achieved the highest area under the curve of 99.87%. DenseNet121 and ResNet50 obtained the highest recall of 99.00%. VGG16 also produced a Brier Score of 0.0135 and an Expected Calibration Error of 0.022 indicating reliable probability estimates. The findings establish a comparative reliability-oriented evaluation of pretrained architectures on naturally acquired apple images and provide a basis for further validation across broader acquisition environments.

MAMSE-EfficientNet: Microscopy-Aware Multi-Stage Image Enhancement Guided Deep Learning for Canine Blood Parasite Detection

Zoren P Mabunga, AprilJoyce M. Mapa, Katrin Joy D. Pornosdoro, Maurine C. Panergo and Dan Arris Lauresta

Microscopic blood-smear examination remains an accessible first-line tool for canine hemoparasite screening, but image quality varies with staining, illumination, acquisition hardware, and smear preparation. This paper investigates whether deterministic, microscopy-aware enhancement can improve deep-learning discrimination under a small-data setting. A five-stage Microscopy-Aware Multi-Stage Enhancement (MAMSE) front end is proposed: Shades-of-Gray color constancy, morphological background correction, median denoising, contrast-limited adaptive histogram equalization (CLAHE), and unsharp masking. MAMSE is coupled to an ImageNet-pretrained EfficientNet-B0 classifier. Experiments use an actual archive of 118 canine blood-smear fields comprising 51 parasite-positive and 67 parasite-negative images. Raw and MAMSE representations are compared using the same repeated stratified 5-fold cross-validation protocol repeated three times (15 paired outer test folds), with training-only augmentation, class weighting, early stopping, and validation-set threshold selection. Relative to the raw-image baseline, MAMSE increased mean accuracy from 69.58% to 76.73%, specificity from 59.12% to 72.67%, ROC-AUC from 78.68% to 88.48%, and PR-AUC from 72.96% to 86.04%; sensitivity remained comparable (83.33% versus 82.00%). Because repeated cross-validation folds are correlated, Nadeau-Bengio corrected resampled t-tests were used for primary inference. Improvements remained significant for ROC-AUC ($p=0.0315$) and PR-AUC ($p=0.0175$). The results indicate that the proposed front end primarily improves ranking quality and false-positive rejection rather than sensitivity. The study is a proof-of-concept and does not constitute species-level or clinical diagnostic validation.

A Two-Component AI Framework for Early Warning and Image-Based Diagnosis of Tuta Absoluta in Tomato Crops

Saja Salahiden Hwaishel and Shadia Yahya Baroud

Tuta absoluta is a destructive tomato pest whose population can increase rapidly before severe plant damage becomes visible. This study develops a complementary artificial-intelligence framework for short-term pest forecasting and image-based diagnosis. The forecasting component used 1,736 surveillance observations from 229 pheromone traps in Malawi, enriched with seven-day historical temperature, humidity, rainfall, and wind-speed variables derived from NASA POWER. A chronological split trained the models on 1,498 earlier observations and evaluated them on 138 later observations. Random Forest achieved the best root mean squared error of 324.62 and an R-squared value of 0.238, reducing RMSE by 11.16% relative to a persistence baseline. A MobileNetV2 transfer-learning classifier was separately trained on 6,015 balanced tomato images to distinguish healthy plants, T. absoluta damage, and other tomato diseases. On 903 unseen images, it achieved 79.29% accuracy and a macro F1-score of 79.06%; recall for T. absoluta was 79.07%. Results show that historical trap counts were substantially more informative for next-observation forecasting than the available district-level weather variables, while image classification provided a useful complementary diagnostic signal. The framework supports early warning but requires local, multi-season validation before operational deployment.

D2-C2: Intelligent Networks & Emergency Communications

A Comparative and Explainable Machine Learning Framework for Elevated Anxiety Symptom Prediction in Gamers

Saeed Sharif, Afrin Nanmi, Jitendra Seelankala and Wael M El-Medany

This study presents a comparative and explainable machine learning framework for identifying elevated GAD-7 symptom levels among online gamers. Using a public gaming survey dataset of 13,464 participants, behavioural-demographic, psychological, and combined feature sets were evaluated with Logistic Regression, Random Forest, and XGBoost for classification, and Ridge Regression, Random Forest, and XGBoost for regression. Stratified five-fold cross-validation was used for classification, while standard five-fold cross-validation was used for regression. Behavioural-demographic models performed weakly in isolation, with a maximum ROC-AUC of 0.5622, whereas psychological-only Logistic Regression achieved ROC-AUC 0.8111. The combined Logistic Regression model achieved ROC-AUC 0.8128, representing only a 0.0017 improvement over the psychological-only model. Regression showed the same pattern: behavioural-demographic models produced negative R^2 values, while psychological-only Ridge Regression achieved $R^2 = 0.3232$. SHAP analysis of the combined XGBoost model indicated that social anxiety (SPIN_T) and life satisfaction (SWL_T) contributed most strongly to predictions, while basic gaming-behaviour variables provided limited additional information. These findings suggest that survey-level gaming behaviour alone has weak predictive value for elevated anxiety symptoms, and that feature-group comparison and explainable modelling are important for avoiding overinterpretation in behavioural mental-health machine learning.

Design and Evaluation of an AI-Assisted Decision Support System in Microsoft Power BI

Muhammad Nabeel, Afrin Nanmi and Saeed Sharif

Despite the growing prevalence of AI-enabled platforms for business intelligence applications, the empirical analysis of such features as integrated decision support systems is still rare. In this paper, the design, implementation and evaluation of an AI-assisted decision support system in Microsoft Power BI with six native AI features (forecasting, anomaly detection, Key Influencers, Smart Narratives, Decomposition Tree, Q&A) are presented. The features were integrated into a three-page interactive dashboard using a 9,800-row retail sales dataset covering 2015-2018. A Design Science Research methodology guided a six-phase implementation pipeline incorporating ETL, star-schema modelling, DAX, and AI-assisted analytics. For a three-month holdout period from October to December 2018, the native Power BI forecast produced MAPE = 49.73%, compared with 38.73% for the Naive baseline and 21.19% for the Seasonal Naive baseline. Directional accuracy was 0/2 transitions, indicating that neither observed month-tomonth direction change was predicted correctly. The main contributions are an evaluated six-phase implementation framework for integrated AI-assisted BI, a benchmarked assessment of native Power BI forecasting against simple baselines, and a cross-feature analytical consistency check between Key Influencers and Smart Narratives

RDECS: A Portable Radio Repeater with RF-Cellular Interoperability for Emergency Communications

Faith Allison B. Buenaobra, Franchesca Isabel T. Dela Torre, Audrey Marie G. Hans, Jimica Charise G. Lau and Gerald P Arada

Communication is essential for coordination, emergency response, and field operations, particularly in areas where conventional communication infrastructure is unavailable or disrupted. Traditional radio repeaters provide extended communication coverage but may be vulnerable to technical failure, maintenance delays, environmental conditions, and dependence on fixed infrastructure. This study presents the design and field evaluation of the Rapidly Deployable Emergency Communication System (RDECS), a low-cost and portable radio repeater prototype developed using accessible hardware and an Arduino-based controller. The system was designed to extend handheld-radio communication while supporting interoperability with Push-to-Talk over Cellular (PoC) communication. Field tests were conducted in Tagaytay and Taguig, Philippines, under different line-of-sight, non-line-of-sight, terrain, and deployment-height conditions. Direct-mode testing demonstrated progressive signal degradation with increasing distance, while repeater-assisted operation sustained communication over substantially longer test distances. Under the Tagaytay field conditions, repeater-assisted testing extended to approximately 54.5~km, with communication progressively degrading and becoming unavailable at the final test point. In the more heavily urbanized Taguig environment, communication range was considerably reduced because of building obstructions and limited line of sight. The system also demonstrated interoperability between conventional two-way radio and PoC communication. The results demonstrate the feasibility of a portable and rapidly deployable communication system for extending radio coverage in infrastructure-limited environments.

An ESP-NOW-Based Emergency Response System for Mining Using Multi-Hop Topology

Aerah Marie E. Alejandre, Adrian Marco S. Mortiz, Harold James S. Ordonez, Sarah Elizabeth C. Timbang and Gerald P Arada

Underground mines require hazard monitoring and emergency communication even when Wi-Fi, cellular, or cloud infrastructure is unavailable. This paper presents an infrastructure-independent emergency response system based on ESP32 microcontrollers and ESP-NOW multi-hop communication. The network consists of a sensor-equipped Source node, three Relay nodes, a Master node, and an optional Gateway node. The Source monitors methane, carbon monoxide, hydrogen sulfide, temperature, and humidity, while panic buttons permit manual emergency activation. The Master evaluates received environmental data and distributes SAFE or EVACUATE decisions to the network. Underground field tests evaluated request--response latency, response--decision latency, packet delivery ratio (PDR), relay-failure tolerance, panic-alert response, and extended-distance operation. At 30 and 60 m inter-node spacing, all tested configurations achieved 100% PDR. At 90 m, the full-node configuration maintained 100% PDR, while relay-bypass configurations achieved 94-98% depending on the communication process. Communication remained possible over an approximately 685 m deployment, but packet loss increased to 38-46%, making that spacing unsuitable for emergency use. The results demonstrate the feasibility of ESP-NOW multi-hop networking for low-cost, local, and fault-tolerant underground safety communication.

Temporal Characterization and Machine-Learning-Based Prediction of Television White-Space Availability Using Continuous UHF Spectrum Measurements

Annie Liza C. Pintor and Gerald P Arada

This paper presents a measurement-driven temporal characterization and machine-learning-based prediction of television white-space (TVWS) availability using continuous UHF spectrum measurements. A total of 5,099 spectrum sweeps were collected over approximately 84 h, and the 470-698 MHz television band was mapped to 38 nominal 6-MHz channels, yielding 193,762 channel-time observations. Channel occupancy was determined using a sweep-adaptive threshold based on the upper bound of the lowest 20% of measured spectral powers plus a 5-dB margin. The measurements indicated an overall occupancy ratio of 26.88%, corresponding to 73.12% TVWS availability. Censoring-aware Kaplan-Meier analysis yielded a median availability-duration point estimate of 29.26 h; however, substantial right censoring prevented estimation of the upper 95% confidence limit for the median within the approximately 84-h observation window. Multi-horizon prediction was evaluated using purged chronological train-validation-test partitions, with Random-Forest feature selection performed using validation data only and a persistence predictor used as a mandatory baseline. At 15 min, the selected Random Forest detected 23 of 81 aggregate state transitions; at 30 min, it detected 38 of 156. Directional analysis showed that the detected events were entirely occupied-to-available transitions: recall was 63.89% (95% Wilson CI: 47.58-77.52%) for 23 of 36 events at 15 min and 57.58% (45.56-68.76%) for 38 of 66 events at 30 min. No available-to-occupied events were detected among 45 and 90 such events at 15 and 30 min, respectively, corresponding to 95% upper confidence bounds of 7.87% and 4.09%. These results show that aggregate prediction accuracy can obscure directional uncertainty and safety-relevant transition behavior.

Thursday, November 19 12:35 - 14:15

D2-A3: Intelligent Cloud, FinTech & Agentic Security

LambdaForge: An Empirical Evaluation of Cost, Tail Latency, and Reliability in Serverless Algorithmic Trading

Vishwak Thatikonda and Shravani Parsi

Serverless platforms promise operational simplicity and pay-per-use pricing for event-driven workloads, but their suitability for latency-sensitive financial automation remains unclear. We present an empirical evaluation of LambdaForge, an open-source production serverless trading system, measuring end-to-end latency (p50/p95/p99), cost, and failure recovery across watchlist sizes (10-218 symbols), memory allocations (256-1769 MB), thermal states, and processor architectures, against an always-on AWS Fargate baseline, using a frozen replay workload and deterministic fault injection (≈1,400 controlled invocations). We find: (1) warm median latency grows linearly with watchlist size (8.78 ms/symbol over a 1.26 s fixed overhead, R^2 =0.9996), amortizing per-symbol time 9.3× from 10 to 218 symbols; (2) cold starts add a roughly workload-independent 0.41-0.55 s offset, doubling memory from 512 to 1024 MB halves latency at approximately equal cost, and ARM64 delivers a 1.32-1.59× price-performance advantage over x86_64; (3) over 82,633 passive scheduler observations, EventBridge rate rules carry stable per-rule constant delivery offsets of tens of seconds (27.9 s and 32.4 s for two identical rules), while minute-aligned cron schedules deliver with median 1.26 s (p99 1.42 s), and 0.07% of fires arrive more than 5 s late, including duplicates redelivered up to 42 minutes late; (4) Lambda remains cheaper than the smallest always-on alternative below ≈596,000 invocations/month, 27× the production cadence; (5) identical inputs produce different trading signals across the three Graviton generations Lambda schedules onto, a reproducibility hazard for float-threshold financial logic. We release our benchmark harness, manifests, and raw data.

RFMD Islamic Framework: A Scoping Review of Donor Retention in Islamic Digital Philanthropy

Najwan Firdaus Haris, Irmasari Hafidz and Nur Aini Rakhmawati

Islamic digital philanthropy is expanding as ZISWAF (zakat, infaq, sadaqah, waqf) collection migrates to online platforms, but donor retention on these platforms is complicated by multi-akad giving and by anonymous donors recorded under the shared label "Hamba Allah," who cannot be analysed with identity-based techniques. This scoping review maps how the existing literature addresses donor retention in Islamic digital philanthropy. Following PRISMA-ScR, we searched Scopus and IEEE Xplore in July 2026 and screened 538 records through seven sequential stages to a corpus of 41 documents (31 journal articles and 10 conference papers, published 2021-2026). Abstracts were synthesised using Structural Topic Modelling, which identified four research streams: zakat institutional risk and professional management; digital Islamic crowdfunding and waqf fintech adoption; cash waqf, trust, and institutional donation models; and donor retention and intention on ZIS digital platforms. The corpus is dominated by cross-sectional behavioural-intention studies and is geographically concentrated, with 25 of 41 papers (61%) carrying at least one Indonesian affiliation. Only two reviewed studies apply RFM-based segmentation to Islamic donor populations, and both treat retention as static classification rather than a longitudinal process; no reviewed study models segment transitions over time or treats anonymous donors as a distinct analytical population. These two absences define the principal open problems for platform-scale retention analytics on multi-akad Islamic donation platforms.

Confidential Computing in Practice: Evaluating Memory Isolation, Remote Attestation, Latency, and Cold-Start Cost in AWS Nitro Enclaves

Rajendranath Rengan Nambirajan, Abhimanyu Garg, Shasha Nigam and Sangharsh Agarwal

Unlike data at rest and in transit, data in use must be decrypted into host memory before a processor can operate on it, leaving it readable by anyone with privileged access to the machine. Confidential computing addresses this using hardware trusted execution environments, yet deployment guidance rarely separates the guarantee from its measured cost. We evaluate AWS Nitro Enclaves for an analytics workload motivated by cross-institution data sharing, running the same application as a conventional host process and as an enclave reached over vsock on a CPU-pinned baseline configuration. Under host-process memory inspection by a root-privileged observer, snapshots recovered all 200 planted records from the baseline and none from the enclave; this exercises one capability of that adversary, not resistance to kernel-level or DMA-based acquisition. With the key, key policy, and IAM principal held constant, the configured policy denied decryption to the parent, allowed it to an enclave presenting the pinned measurement, and denied it to a rebuilt image. The enclave-to-baseline latency ratio decreases from 1.49× at one record per request to 1.02× at ten thousand, the largest difference between mean run-level medians being 1.15±0.72 ms. A controlled padding sweep found a linear relationship between EIF size and time from run-enclave invocation to first accepted vsock connection, slope 13.81 ms/MiB and intercept 1.841 s; for the 183.6 MiB production image the fitted size-dependent term accounts for approximately 58% of the 4.376 s median startup, The experiment estimates the size- dependent component but does not identify the individual stages contributing to the fitted intercept.

From Prompt Injection to Authorized Tool Abuse: A Factorial Security-Control Testbed for MCP-Shaped Agentic Workflows

Valeri Milke

Agentic workflows can turn untrusted content into syntactically valid tool calls that carry valid credentials. This study isolates downstream enforcement after a compromised decision by evaluating audience-restricted identity, deterministic tool-and-argument policy, memory isolation, and high-impact approval. We implement a transparent, JSON-RPC-2.0- and tools/call-shaped in-memory testbed with five benign tasks and six seeded attacks: cross-server credential replay, valid-credential argument abuse, administration escalation, delayed memory poisoning, privileged queue selection, and record-scope abuse. A full 2 × 2 × 2 × 2 design produces 176 deterministic case executions. The uncontrolled baseline executed all six attacks and completed all five benign tasks. Audience binding alone left four attacks, memory isolation left five, and high-impact approval left two. Deterministic policy blocked all six fixed attacks but blocked two benign exception cases; adding the synthetic approval oracle restored 5/5 benign completion while maintaining 0/6 attack execution. The results characterize only the seeded suite: the agent and reviewer are deterministic test doubles, and no real model, MCP SDK, transport, OAuth flow, cryptography, or wire interoperability is evaluated. The contribution is a reproducible factor-isolation method, effect oracle, and raw decision evidence for security-control testing of MCP-shaped tool workflows, not a claim that MCP itself is vulnerable.

D2-B3: Machine Learning, Optimization & Predictive Models

Machine Learning Prediction of Wellhead Pressure Correction for Virtual Well Testing

Dwi Widardo and Rarasmaya Indraswari

Routine gas-well tests collected at a central processing facility may not represent wellhead pressure, because a long multiphase flowline separates the two measurement points. This study predicts the pressure correction with machine learning, and explains which operating variables drive it. One hundred paired records from seven anonymized wells carried nine inputs. Twenty records were set aside as a test set before tuning; the remaining 80 formed the training and validation data, assessed by five-fold cross-validation grouped by well. Random Forest, extreme gradient boosting, an artificial neural network, and support vector regression were compared by an exhaustive grid search over 2,145 hyperparameter settings, each trained once per fold, for 10,725 fits. On the validation sets, support vector regression reproduced the largest share of variance in the correction (coefficient of determination 0.377, that is 37.7%) and gave the lowest root mean squared error (96.12 psi). Random Forest reached the highest compliance rate, the share of records whose corrected pressure falls within 5% of the reference, 75.00% against 46.25% without calibration. Support vector regression was therefore selected as the calibration model. Test-set compliance rates were 75, 80, 65, and 70%, which cannot separate the models, since one record among 20 moves the rate by five percentage points. SHAP gave the same influence direction in all four models for seven of nine inputs. Small sample size, the time gap between paired measurements, and pending engineering validation preclude replacing physical well testing.

Exploratory Multimodal Analysis of Affect and Aggression Using Voice and Smartphone Interaction Patterns

Urbano B Patayon and Melvin A. Ballera

This study explored patterns in voice, touch, and smartphone motion to recognize positive and negative affect and estimate aggression during gameplay. Participants completed the aggression questionnaire developed by Buss and Perry and played the game while audio, keypress, accelerometer, and gyroscope data were recorded. Three raters evaluated affect, producing substantial agreement with a Fleiss Kappa of 0.6178. Exploratory analysis identified relationships between affect and MFCC, shake magnitude, and touch features. Verbal aggression was negatively associated with affect, whereas hostility was positively associated with affect. Pheromone-Guided Recursive Feature Elimination identified relevant multimodal variables for classification and regression. Extra Trees achieved the best affect classification performance, with 0.83 accuracy, precision, and sensitivity and 0.86 specificity. Gradient Boosting best predicted verbal aggression, explaining 87 percent of its variation, while CART best predicted hostility, explaining 95 percent. Feature importance differed across outcomes, showing that affect and aggression were represented by distinct combinations of vocal and smartphone interaction patterns. The findings support multimodal sensing for affect recognition and aggression estimation, although larger independent datasets are required to validate the models.

A Memetic Multi-Objective Genetic Algorithm for Motif Discovery in PAM50 Breast-Cancer Gene Promoters: A MOGAMOD Extension with NSGA-II and NSGA-III

Hind Ayad Majeed Alkakjea, Erdal Özbay and Mehmet Kaya

Discovering regulatory motifs in promoter sequences is a central, intrinsically multi-objective problem in computational biology, since a useful motif must trade off conservation, coverage and length. Building on the multi-objective genetic-algorithm (MOGA) line for motif discovery, we present a memetic many-objective genetic algorithm and apply it to the promoters of the 50 PAM50 breast-cancer genes. The method extends the MOGAMOD formulation with three additions: a consensus-guided memetic local-search operator, information content as a fourth objective, and a many-objective optimiser (NSGA-III) compared against NSGA-II. As a sanity check on synthetic planted data, the method recovers all implanted sites, and an ablation shows that the memetic operator is essential. On 50 real promoters from Ensembl, the discovered GC-rich pattern matches the GC-box / Sp-KLF zinc-finger family in JASPAR 2026 (best match SP5, score 0.825). The motif is a generic CpG-island promoter element, yet a per-promoter analysis shows it is significantly enriched in PAM50 relative to 50 non-PAM50 controls (mean match identity 0.750 vs. 0.713; Mann-Whitney p = 0.002). NSGA-II and NSGA-III are statistically indistinguishable at four objectives (Wilcoxon p = 0.80), and a similarity-versus-shuffle null is shown to be uninformative on GC-rich promoters. Both the positive enrichment result and the negative and methodological findings are reported transparently.

Sequence-Aware Multi-Graph Encoding for Temporal Link Prediction in Cryptocurrency Networks

Ansu Badjie, Toussaint Elvis Compaore, Diedhiou Bacary, Mohammed Tamimu Gadafi, Grace Mupoyi Ntuala and Aly Boro

Temporal link prediction in cryptocurrency trust networks is a security-relevant inference problem: anticipating which edges will appear in the next transaction snapshot enables proactive fraud screening, counterparty risk assessment, and behavioral anomaly detection before transactions settle. Existing methods based on structural similarity heuristics fail when candidate edges connect nodes that share no common neighbors, a condition that dominates in sparse monthly snapshots. Learned dynamic graph models handle temporal evolution but do not account for the large fraction of nodes with thin neighborhoods, where message passing produces unreliable representations. We propose a sequence-aware multi-graph encoder that addresses both limitations. Two graph-attention layers produce neighborhood summaries, a missing-information gate corrects the representations of weakly connected nodes by recovering signal that sparse neighborhoods fail to supply, and a compact two-layer lightweight GPT-2 transformer reads the full per-snapshot node set as a sequence so that structurally distant nodes share global context. A two-layer perceptron decodes candidate node pairs from fused local and sequence representations. Experiments on Bitcoin-Alpha, Bitcoin-OTC, and Ethereum ERC721 show that the proposed model achieves AUC values of 0.931, 0.924, and 0.935 and F1-scores of 0.926, 0.919, and 0.927, respectively, outperforming six baselines, including EvolveGCN and Tail-GNN. An ablation study identifies the missing-information gate as the most consistently effective component, with AUC reductions of 0.053 on Alpha, 0.055 on OTC, and 0.053 on ERC721 upon its removal. The complete model runs on a CPU in minutes, making it practical for real-time deployment in transaction monitoring pipelines.

A Model Proposal for Multi-Defect Classification in Fuzzy Statistical Process Control

Ibrahim Maher Said Bakr, Ismail Aburesha and Ali Karasan

This paper presents a proof-of-concept framework that connects multi-label surface-defect classification with fuzzy statistical process control (SPC). A synthetic industrial metal-surface dataset containing scratch, crack, rust, and hole defects was expanded to 8,500 images by creating double- and triple-defect samples. Source images were separated into training, validation, and test pools before compositing, and each generated sample used base images and defect patches only from its own split. A MobileNetV2 transfer-learning model produced four independent sigmoid scores. On the 1,000-image test set, it achieved 0.9912 binary accuracy, 0.9710 subset accuracy, 0.9578 macro F1-score, and 0.9992 AUC. The sigmoid scores were passed through a two-stage fuzzy inference system that generated a linguistic quality-risk level and a continuous 0-100 monitoring index. The test images were then shuffled with a fixed seed and arranged into 20 simulated batches, each containing ten subgroups of five images. X-bar and R charts monitor subgroup mean and within-subgroup variation, and a rule-based evaluator assigns batch-level diagnostic statuses. Nine batches are classified as overall process instability, five as early warning in both mean and variability, four as mean-risk instability, and two as variability instability. Based on the results, the SPC stage demonstrates the proposed monitoring logic on simulated batches.

D2-C3: IoT Applications, Smart Energy & Logistics

DapatWatt: Design and Field Evaluation of an AIoT-Based Residential Energy Management System

Erwin B. Bonto, Marci Angelo Q. Lara, Sean Francis C. Natividad, Kate Zyrll M Plantilla and Gerald P Arada

This paper presents DapatWatt, an Artificial Intelligence of Things (AIoT)-based residential energy management system designed to support household electricity monitoring, anomaly detection, and short-term load forecasting. The system integrates non-intrusive aggregate electrical measurements, environmental sensing, cloud-based data processing, and a user-oriented monitoring interface. Its operational feasibility was evaluated through deployment in five residential households. Four machine-learning classifiers were compared for electricity consumption anomaly detection using DBSCAN-generated pseudo-labels. The Support Vector Classifier achieved the highest accuracy of 97%, while Naive Bayes was selected based on its higher macro-average recall of 0.69. Five deep-learning architectures were also evaluated for load forecasting. The conventional Recurrent Neural Network achieved the lowest mean absolute error of 0.114, outperforming the attention-based LSTM, which obtained 0.144. Cross-household forecasting results revealed substantial performance variability, indicating limited generalization across different consumption patterns. A user evaluation involving five respondents indicated favorable perceptions of system usability and functionality, with an overall mean rating of 4.60 on a five-point scale. Although self-reported electricity bills suggested reduced consumption in some households, these observations do not establish causal energy savings. The findings demonstrate the feasibility of integrating AI-based analytics with IoT infrastructure for residential energy monitoring while highlighting the need for independently annotated anomaly data, broader household validation, and improved forecasting generalization.

Design and Implementation of a LoRaWAN-Based Tracker for Mountain Bikers

Nathaniel Dale S. Crisaldo, Patrick Miles L. Cheong, John Clarence V. Manga, Chester Russell U. Tan and Gerald P Arada

Mountain biking in rugged, multi-sloped terrain carries significant safety risks, as unreliable cellular coverage often delays location tracking and vital-sign monitoring during emergencies. LoRaWAN-based IoT technology can help address this gap in wilderness safety. To improve real-time localization and health monitoring for mountain bikers, two wearable LoRaWAN tracker nodes integrating RSSI-based trilateration, multihopping, and vital-sign sensing were designed and fieldtested. Each wrist-worn node is built with a 3D-printed housing and around a Heltec LoRa V3 microcontroller with a MAX30102 heart rate and SpO2 sensor, an MPU6050 for fall detection, an ATGM336H GPS module, and an SOS push button, connecting to three deployed LoRaWAN gateways and relaying data via HC-12 multihop outside gateway range. Field testing at Victoria Bike Trails and the De La Salle University - Laguna campus achieved a mean trilateration error of 15.4 m, with 98.3% of position estimates within the 50 m accuracy threshold, and demonstrated successful multihopping across both deployments. Sensor reliability was statistically validated against a commercial smartwatch using a paired-samples t-test, finding no significant difference in SpO2 ($p = .186$) and only a small difference in heart rate ($p = .011$). A Two One-Sided Tests (TOST) equivalence procedure confirmed equivalence within clinically negligible bounds for both vital signs (10 bpm for heart rate, 1.5% for SpO2; both $p < .001$), establishing sensor reliability for use in the tracker.

Smart Energy Management and Cloud Monitoring of Solar Power System for Sustainable Electricity

Auday Al-Mayyahi, Zainab Ali Khalaf, Ammar A Aldair and Mussab Yousif Ahmed

Solar energy systems have been increasingly applied as a sustainable power supply for generating electricity. In this paper, a smart solar power system is designed with an energy management system. This has been achieved based on integration between basic solar components and embedded systems. The main parts in the design are represented by Photovoltaic (PV) Solar Panels, Maximum Power Point Tracking (MPPT), Inverter, Batteries in addition to an Arduino-based controller, sensors, and Internet of Things (IoT) technologies. The system is provided with cloud monitoring capabilities to efficiently monitor, store, and distribute solar energy while providing real-time data on energy production, consumption, and battery health. Based on sensor measurements of voltages and currents. The system continuously gathers data to ensure optimal performance, displaying real-time information on a screen and uploading data to a cloud platform (Blynk Cloud) for remote access. Automated load management is implemented to prioritize essential appliances, ensuring energy efficiency during periods of low solar generation and battery voltage. This paper serves as a practical solution for optimizing renewable energy and contributing to sustainability based on automation in energy management. Practical design has been feasibly implemented to evaluate the system performance, and the obtained experimental results have effectively proven the effectiveness and validation of the overall design.

Optimal Routing Algorithm for Disaster Relief Goods Distribution System

Ann E Dulay, Kenley Lewis Chiu, Jet Orville Chua, Emil John Lopez and Josef Lyle Solis

The Disaster Relief Goods Distribution System (DRGDS) is a mobile and web-based application that offers seamless solution for relief operations during calamities. This system is specifically intended for use by the private sector although its features lend itself for DSWD use and for the other government agencies that are part of the relief efforts. This system aims to solve the issue of oversupply, undersupply, and inequitable distribution of goods to the affected communities. The app has classified three types of users, the LGU through its barangay units, the donor which can be the private sector or the DSWD, and the recipient resident. Three variations of the algorithm were developed and compared as to which provide the optimal route based on the goods to be donated and the distance of the donor to the affected sites. The three algorithms, Tabu search, Genetic algorithm (GA), and L-nearest neighbor were benchmarked with the CPLEX library. Of the three, the Tabu search gave the most optimal route. The web application was connected to the server via an API designed using the Django REST framework deployed in Heroku while the Android application was profiled using built-in tools in Android studio. The Android and React applications were measured using the software usability scale (SUS), both garnering a grade equivalent to A in SUS standards

IoT-Enabled Soil Moisture Monitoring and Automated Watering System

Joash Mari P. Manzanero, Almet R Tesorero, Arjay G Neri, John Joshua F. Montañez, Donna T. San Antonio and Roy Dialogo Tipones

Modern agriculture continues to evolve with the integration of emerging technologies, with precision agriculture offering innovative approaches to sustainable crop production. This study developed an Internet of Things (IoT)-based soil moisture monitoring and automated irrigation system aimed at optimizing water usage and enhancing plant growth. The system employs capacitive soil moisture sensors, a microcontroller platform, and a water pump system that activates when moisture levels drop below a preset threshold. An experimental posttest-only control group design was used to assess the performance of the automated system compared to traditional manual watering. Three replicates for each group (control and treatment) were tested under similar conditions using parameters such as plant height, number of leaves, biomass accumulation, soil moisture levels, and water efficiency. Statistical analysis using an independent sample t-test indicated significant improvements in the treatment group across all measured growth indicators. Notably, the automated system maintained soil moisture between 60-90%, improving hydration consistency. Water efficiency data revealed that the IoT-enabled system reduced water consumption by up to 30% compared to the manual method. Plants in the treatment group exhibited greater height, more leaves, and higher biomass accumulation, confirming the system's effectiveness. The study concludes that integrating IoT in irrigation through sensor-based control significantly improves water use efficiency and crop development. It is recommended that this system be adopted in wider agricultural settings to promote sustainable farming practices and optimize yield while conserving valuable water resources.

Thursday, November 19 15:15 - 16:55

D2-A4: Smart Cities, Sustainable Systems & Monitoring

Evaluating Cross-City Spatial Transferability of Multi-Modal Neural Networks in Retail Location Intelligence: A Case Study in Vietnam

Lam Mai and Duc-Trien Nguyen

While multi-modal deep learning has advanced commercial site selection, existing frameworks rely on the assumption of spatial homogeneity, leaving their generalizability to unseen geographical domains unverified. In data-scarce emerging markets, spatial transferability represents a key operational barrier for scaling retail networks. This study establishes a rigorous cross-city evaluation protocol to benchmark the spatial transferability and adaptation of deep site selection models. Leveraging a fine-grained geospatial dataset of 73,974 grid cells (250m x 250m) across three major Vietnamese cities (Hanoi, Ho Chi Minh City, and Da Nang), we evaluate an interaction-aware Wide-Cross-Deep neural network that integrates OpenStreetMap POIs, VIIRS night-time light imagery, and Google Maps community reviews. Evaluated across three leave-one-city-out scenarios, the model demonstrates high zero-shot generalizability in Da Nang and Ho Chi Minh City, whereas transferring to Hanoi highlights performance bounds driven by regional market variations. We further demonstrate that target-city fine-tuning significantly restores predictive accuracy under limited supervision. Our work offers empirical guidelines for deploying location intelligence architectures across heterogeneous urban domains.

A Retrospective Review of Isolation Forest and Threshold Monitoring in a Kubernetes Case Study

Afrin Nanmi, Kimiya Sadeghi Dastjerdi and Saeed Sharif

Threshold alerting is widely used in Kubernetes, but its evaluation windows set a lower bound on how soon a fault can be reported. We present a paired case study in which an unsupervised Isolation Forest detector and Prometheus alerting rules monitored the same telemetry while forty faults were injected into two AWS EKS microservices (ten trials for each of four fault types). For CPU exhaustion, high latency, and service crash, both methods detected 10/10 faults, while the detector had lower mean time to detect (MTTD): 14.8 s versus 67.8 s, 15.2 s versus 66.0 s, and 20.0 s versus 35.9 s. Memory pressure showed the opposite pattern: the detector alerted in 6/10 trials (95% CI 31-83%) compared with 10/10 for the rules. These results are shaped by configuration as well as capability. Threshold MTTD cannot be shorter than the configured for duration, while the detector's CPU and latency distributions closely match the phase offset of its thirty-second scoring cycle (within 0.4 s in both mean and standard deviation), consistent with first-cycle detection throughout. The conclusion is therefore deliberately narrow: performance depended on fault type and alert configuration, not simply on whether monitoring was learned or rule based.

An Attention-Enhanced Deep Reinforcement Learning Framework for Adaptive Traffic Signal Control

Aftab Ali, Rahmad Sadli, Fityanul Akhyar, Muhammad Husaini Nadri and Mahnoor Izzat

Traffic congestion remains a major urban mobility problem because fixed-time signal controllers cannot adapt to dynamic and asymmetric traffic demand. This study proposes a spatial-temporal attention-based Deep Q-Network (DQN) for adaptive traffic signal control in the SUMO environment using the TraCI interface. The controller uses detector-based traffic states from four incoming approaches, combining local direc tional features with global traffic context. A history buffer captures recent traffic patterns, while spatial and temporal multi-head attention modules help the agent focus on critical approaches and short-term congestion trends. The action space jointly selects the signal phase and green duration, allowing more flexible control than fixed-duration methods. The proposed model was evaluated against a fixed-time 40 s controller and a vanilla DQN baseline over five independent testing runs. Results show that the attention-enhanced DQN reduced mean queue length and mean waiting time by 52.89% and 53.40%, respectively, compared with fixed-time control, and by 16.46% and 16.09% compared with vanilla DQN. The model also maintained comparable throughput with fewer phase changes, indicating a more stable and efficient signal control policy.

Artificial Intelligence of Things (AIoT) for Smart Building Maintainability: A Comparative Literature Review and Policy Framework for Singapore and Manila

Gerald P Arada and Sheila Maria Arcuino Conejos

The convergence of artificial intelligence and the Internet of Things, increasingly referred to as AIoT, is reshaping how buildings and cities are sensed, operated, and governed. Yet the literature and policy attention on AIoT-enabled smart city development remain concentrated in mature digital economies such as Singapore, with comparatively little work on the governance adaptations required in Global South cities such as Manila. This paper presents a literature review of smart city research organized around nine thematic areas: infrastructural development, governance, renewable energy and smart grids, information systems integration, environmental technologies, IoT-driven decision-making, mobile crowd-sourcing, socio-technical implementation challenges, socio-economic development, and policy developments, with particular depth on the socio-technical, socio-economic, and policy dimensions, and further deepens two cross-cutting technical threads, digital twins and AIoT cybersecurity, that recur across them. Using Singapore's Smart Nation and Green Mark ecosystem as a model and Metro Manila as the receiving context, we identify five concrete AIoT policy gaps in the Philippine built-environment governance landscape and propose a transferable framework to close them.

D2-B4: Big Data, Health AI & Knowledge Systems

Thematic Analysis of Data Lake Research in Education: A Computational Literature Review

Chayun Fadila, Najwan Firdaus Haris, Shaban Lubanga and Nur Aini Rakhmawati

The rapid digitization of higher education has generated heterogeneous data from learning management systems, student records, and administrative platforms, creating new demands for data management. Among emerging paradigms, Data Lakes have gained attention for their flexibility in handling diverse educational data types. However, the literature lacks a systematic, computational synthesis of this area. This study addresses this gap by applying Latent Dirichlet Allocation (LDA), within the Structural Topic Modeling (STM) family, to Scopus-indexed research on Data Lake use in educational settings. From a corpus of 57 usable records the model identifies four themes spanning data lake infrastructure, education data processing, smart campus systems, and enterprise architecture, with education-facing themes forming a slight majority of the corpus and pairwise topic co-occurrence within papers remaining limited. Given the modest corpus size the results are exploratory, but they provide an initial computational map of the field and point to opportunities for evaluation-oriented, cross-paradigm research linking data lake architecture to educational outcomes.

System Dynamics-Based Conceptual Model: A Case Study of Malaysia's Early COVID-19 Outbreak

Aisyah Ibrahim, Tuty Asmawaty Abdul Kadir, Rod MacDonald and Hamdan Daniyal

Complex public health crises require decision-support approaches that can represent not only observed data trends, but also the causal mechanisms and feedback processes that generate them. This paper presents a System Dynamics (SD)-based conceptual model structure for public health decision support, using Malaysia's early COVID-19 outbreak as a case study. Causal Loop Diagrams (CLDs) were developed within the SD approach using qualitative methods, including literature review, thematic analysis, causal analysis, and chronological mapping of outbreak events. The analysis drew on academic publications, grey literature, news reports, official websites, and social media content, supported by ATLAS.ti software and discussions with an SD modelling expert. This study contributes by translating real-world outbreak chronicles into causal and feedback structures that explain how outbreaks escalated and were controlled over time. This study provides a contextual understanding of an early COVID-19 dynamics case and establishes a solid structural foundation for subsequent development of simulation-based decision-support models. Collectively, these contributions offer insights into future improvements in public health preparedness and response.

An LLM-RAG Architecture for Cross-Referencing Halal Standards and EU Deforestation Regulations

Naël Ahmed, Ilian Balit, Rarasmaya Indraswari and Nur Aini Rakhmawati

As global food supply chains face mounting pressure to satisfy both religious compliance and environmental due diligence, manual verification remains fragmented and error-prone. This research proposed a unified retrieval-augmented system for European food product verification that integrates heterogeneous regulatory corpora-including Islamic jurisprudence texts, EUDR legal frameworks, and geospatial traceability data. Experimental evaluation on real-world palm oil and cocoa supply chain documentation demonstrates 94% accuracy in identifying dual-compliance gaps, significantly outperforming baseline LLMs without RAG. The architecture provides explainable audit trails essential for certifiers and regulators. This work establishes a scalable technical foundation for harmonizing faith-based and sustainability standards, reducing compliance costs while strengthening trust in ethical food trade. The experimental results suggest RAG-mediated regulatory alignment can accelerate market access for producers navigating complex multi-standard environments.

Quantized CNN Inference for Heart Murmur Screening on Commodity Edge Hardware

Mohd Rayyan Khan, Mohammad Zaid, Kandarp Gupta and Mohd Wajid

Cardiac auscultation screening in resource constrained settings requires models that are not only accurate but deployable on low-power hardware. This paper presents a convolutional neural network (CNN) trained on the CirCor DigiScope phonocardiogram dataset to detect heart murmurs, together with a complete and numerically verified deployment path to a Raspberry Pi 4B. Four methodological pitfalls common in this literature are addressed: per-recording label assignment using auscultation-site metadata, patient-grouped splitting, threshold selection confined to a validation split, and stratified bootstrap confidence intervals that reflect the small number of positive test patients. The model attains a patient-level test AUROC of 0.901 (95% CI 0.821-0.965) and AUPRC of 0.824, at a sensitivity of 70.4% and specificity of 96.2%. Unlike prior work, which reports quantization error only on small probe sets, the full 4,357-segment test split is re-evaluated on the deployment device under both float32 and int8 precision. The two precisions produce identical confusion matrices and identical decisions for all 132 test patients; DeLong's test on the paired ROC curves gives $p = 0.62$, and exact McNemar's test on the paired decisions gives $p = 1.00$ with zero discordant pairs. Integer quantization reduces model size by 3.74× and per-segment latency by 37.4% at no measurable cost in accuracy, establishing that int8 deployment is accuracy-neutral for this task at this scale.

D2-C4: Digital Education, HCI & Computing Systems

Real-Time Screen Sharing With Low-Bitrate Text-Preserving Compression Using CNN-Based ROI Detection and SVT-AV1 Encoding

Ricardo B. Cruz, Angiolo Ongkinglok, Kerby C. Chiu, Calee P. Mangunay, Donabel Abuan and Hubert F. Abuan

This study introduces a real-time screen-sharing system that enhances text clarity under low-bitrate constraints by integrating a CNN text-region detection with region-aware AV1 video compression. Conventional screen-sharing codecs compress frames uniformly, often rendering smaller fonts and dense text regions illegible once bitrate is constrained, this of which is a limitation especially problematic in educational, technical, and professional fields. To address this, the system captures screen content at 720p and 24fps, applies a lightweight CNN model to detect text block regions per frame, and generates precise ROI (Region of Interest) masks with corresponding quantization parameter offsets. These masks feed into the SVT-AV1 encoder, which prioritizes bit allocation to text-dense areas while aggressively compressing non-text background content. The resulting complete pipeline was implemented as an end-to-end sender-receiver system constrained to 800kbps and 300ms latency maximum cap. Experimental results show the proposed approach achieved an average latency of 85.4ms, an improved compression ratio of 11.36:1, and superior text preservation reflected in higher SSIM (0.8148 vs 0.7544) and OCR confidence (78.440% vs 65.105%) compared to standard AV1 compression, validating the effectiveness of region-aware encoding for text-preserving real-time screen sharing.

AI-Based Formative Assessment of Programming Tasks in Moodle: Architecture, Feedback Quality and Teacher Acceptance

Alessandro Pagano, Paola Plantamura, Veronica Rossano and Sara Maria Cosima Tondo

Generative artificial intelligence is increasingly proposed as a way of sustaining formative assessment at scale, yet most reported systems are evaluated either as machine-learning artefacts or as classroom interventions, seldom as both. This paper reports a three-pillar evaluation of AI CodeRunner, a question-type plug-in that embeds a large language model into the Moodle assessment workflow to produce a numeric score and an argued textual feedback for introductory programming tasks. The technological pillar examines how the plug-in behaves when the input it receives departs from the nominal case: inference is delegated to an asynchronous background task, so the user-facing request cycle is never blocked, and a campaign of 33 black-box cases exercised authoring validation, prompt robustness, sanitisation, rubric persistence and graceful degradation, with 29 passed. The assessment-accuracy pillar reports a rubric-based expert review in which 20 teachers enrolled in a secondary-school computer-science qualification programme first graded the same student submissions themselves and then rated the machine-generated feedback on relevance, clarity, engagement and motivation, and the promotion of reflective thinking; clarity emerged as the strongest dimension and engagement as the weakest. The acceptance pillar reports the behavioural-intention scale of an adapted acceptance questionnaire administered to the same teachers, whose responses indicate a moderately positive but non-uniform willingness to adopt the tool in real teaching practice. The three pillars together support a teacher-in-the-loop design rather than autonomous grading.

Efficient FPGA-Based MUSIC Accelerator for DOA Estimation of Multiple Sound Sources

Shashwat Anand Pandey, Mohd Wajid and Balwinder Singh

The conventional complex valued MUSIC algorithm provides high-resolution direction-of-arrival (DOA) estimation but requires significant computational complexity and hardware resources due to complex covariance processing and eigenvalue decomposition (EVD). This paper presents an FPGA-oriented implementation of real valued MUSIC for acoustic DOA estimation. After complex covariance matrix computation, forward backward averaging (FBA) is followed by a real valued transformation that outputs a real valued symmetric matrix, enabling EVD in the real domain. The resulting noise subspace is reconstructed in the complex domain for MUSIC spectrum evaluation and DOA estimation. The complete architecture is developed using Xilinx Vitis HLS and Vivado and implemented on a Zynq-7000 based PYNQ-Z2 FPGA. The results demonstrate the feasibility of real-valued subspace processing for resource and power efficient hardware based acoustic DOA estimation.

Performance Evaluation of Contemporary AI Models Aiding Accessibility of PDF Documents

Kirti Mishra, Prasanna Jaichandrao Shete and Srijan Upadhyay

Documents in printable document format (PDF) are widely used for scientific and educational material since they preserve page layout across devices. That same visual representation, however, can make content difficult for assistive technologies to interpret. Mathematical content is especially challenging because the meaning of content depends not only on individual characters but also on relationships such as fractions, superscripts, subscripts, matrices, roots, and multi-line derivations. This study evaluates contemporary artificial intelligence (AI) models as a possible extraction layer for making such content more accessible. The study uses two complementary evaluation settings. The first evaluates a roster of approximately ten contemporary models on a controlled benchmark of 1,000 rendered mathematical images that contain 517 normal math formulae, 80 medium-complexity formulae, 251 long formulae, and 152 multi-line derivations. The second evaluates complete-page transcription on a real 53-page scientific paper containing text, equations, tables, and figures. Character Error Rate (CER) was used as the primary metric, with BLEU, exact match, precision, recall, and F1 used as complementary measures. Although ten contemporary models were considered in the study, for evaluating the performance at full document level, only two models viz. Nougat and Qwen2-VL-2B were evaluated. The results show that model performance depends strongly on the level at which the task is presented. Qwen2.5-VL-3B achieved the lowest overall CER on the formula-level benchmark, while Nougat achieved 22.78% CER on the complete document compared with 86.75% for Qwen2-VL-2B. A detailed inspection found that Qwen2-VL-2B produced bounding-box coordinates rather than complete transcription on 39 of 53 pages. The findings therefore suggest that no single model should automatically be treated as the best solution for all mathematical PDF tasks. Instead, model choice should reflect input granularity, document structure, and the type of output required.