



ICETA 2026

The 2026 International Conference on Embedded Technology and Artificial Intelligence

PROGRAM

Annova Nha Trang Hotel, Vietnam
July 20-23, 2026

Organized by

International Conference on Embedded Technology and Artificial Intelligence
IEMEK (Institute of Embedded Engineering of Korea)

WELCOME ADDRESS

Dear Distinguished Guests, Keynote Speakers, Authors, Reviewers, and Participants,

It is my great honor and privilege to welcome you to the ICETA 2026(2026 International Conference on Embedded Technology and Artificial Intelligence), held from July 20 to 23, 2026, in Nha Trang, Vietnam.

On behalf of the Organizing Committee and the IEMEK(Institute of Embedded Engineering of Korea), I extend my sincere appreciation to all participants for joining this international conference. Your presence and contributions are fundamental to the success of ICETA 2026 and to the continued advancement of our research community.

Embedded technology and artificial intelligence are rapidly transforming every aspect of modern society, from intelligent mobility and advanced manufacturing to healthcare, communications, and sustainable infrastructure.

As these technologies continue to evolve, the importance of interdisciplinary collaboration and international cooperation has become greater than ever. ICETA 2026 has been organized to provide a distinguished conference where researchers, engineers, scientists, and industry professionals from around the world can present their latest research findings, exchange innovative ideas, and discuss the challenges and opportunities that will shape the future of intelligent systems.

The success of an international conference is made possible through the collective efforts of many dedicated individuals. I would like to express my deepest gratitude to our keynote speakers, invited speakers, authors, reviewers, session chairs, sponsors, and members of the Organizing Committee for their professionalism, commitment, and invaluable contributions. Their efforts have ensured the high academic quality and international standing of ICETA 2026.

I also hope that this conference will serve not only as a venue for the dissemination of scientific knowledge but also as a catalyst for establishing new collaborations, strengthening existing partnerships, and fostering lasting professional friendships among participants from diverse countries and disciplines.

Beyond the conference sessions, I encourage you to enjoy the unique beauty and hospitality of Nha Trang, one of Vietnam's most attractive coastal cities. I hope that your visit will provide both memorable cultural experiences and opportunities for meaningful exchanges with colleagues from around the world.

Thank you once again for your participation in ICETA 2026. I wish you an inspiring and productive conference, successful discussions, and an enjoyable stay in Vietnam.

Byung-Jae Choi, Ph.D.
General Chair
Professor, Daegu University, Republic of Korea

COMMITTEE

General Chair	Byung Jae Choi (Daegu University, Korea)
Organizing Chair	Jeong Tak Ryu (Daegu University, Korea)
Organizing Co-Chairs	Ju Hyun Park (Yeungnam University, Korea) Jong Hun Lee (DGIST, Korea) Sang Suh (East Texas A&M University) Sang Hyuk LEE (New Uzbekistan University, Uzbekistan)
Program Chairs	Yoo Soo Oh (Daegu University, Korea)
Program Co-Chairs	Robert R. Roxas (University of the Philippines Cebu, Philippines) Su Young Suk (Gyeongbuk Institute of IT Convergence (GITC), Korea) Zia Mohy Ud Din (Air University Islamabad, Pakistan) Je Seok Kim (Daegu Gyeongbuk Institute of Science & Technology, Korea) Dong Shik Kang (University of the Ryukyus, Japan)
Conference Secretary	Hyung Gyu Lee (Duksung Women's University Korea)
Publication Co-Chairs	T. Velmurugan (Dwaraka Doss Goverdhan Doss Vaishnav College, India) Jong Kwan "Jake" Lee (Bowling Green State University, USA) Hyung Shin Kim (Chungbuk National University, Korea) Dong Hwoon Kwon (North Central College, USA) Jahanzeb Gul (Maynooth University, Ireland)
Publicity Co-Chairs	Sang Heon Lee (Daegu Gyeongbuk Institute of Science & Technology, Korea) Hien Nguyen (Ton Duc Thang University, Vietnam) Yong Hoon Kim (Gyeongbuk Institute of IT Convergence, Korea) Gyu Sang Choi (Yeungnam University, Korea) Fawad Salam (Air University Islamabad, Pakistan) Muhammad Asif (HITEC University, Pakistan)
Technical Program Committee	Do Hyeun Kim (Jeju National University, Korea) Azhar Imran (Air University Islamabad, Pakistan) Zeshan ul Haq (Saleem Habib University, Pakistan) Joon Hyuk Yoo (Daegu University, Korea) Tahir Qadri (Sir Syed University of Engineering and Technology, Pakistan) Sidra Agha (Sir Syed University of Engineering and Technology, Pakistan) Su Jeong Heo (Yeungnam University Korea) Jang Kyu Yun (Gyeongbuk Institute of IT Convergence, Korea) Lakshmanan Shanmugam (Vellore Institute of Technology-Chennai, India) Kalidass Mathiyalagan (Bharathiar University, India) Pushpendra Kumar (Dhirubhai Ambani Institute of Information and Communication Technology, India) Quanyong Fan (Northwestern Polytechnical University, China) Sang Moon Lee (Kyungpook National University, Korea) Moon Gap Joo (Pukyong National University, Korea) Dong Hwa Lee (Daegu University, Korea)

International Liaison Chairs	Dong Shik Kang (University of the Ryukyus, Japan) Fernando E. Q, JR. (Biliran Province State University, Philippines) Tae Hee Lee (Jeonbuk National University, Korea)
Local Arrangement Chair	Nguyen Khac Cuong (Nha Trang University, Ho Chi Minh University of Banking, Vietnam)

PROGRAM AT A GLANCE

Time	Event
July 20, 2026	
13:30~15:30	Special Interest Group (SIG) Workshop on Emerging Research Topics I
15:30~16:00	Coffee Break
16:00~17:30	Special Interest Group (SIG) Workshop on Emerging Research Topics II
July 21, 2026	
09:00~09:30	Registration
09:30~09:40	Opening Ceremony (Luminary Hall)
09:40~10:20	Invited Speech 1 (Luminary Hall) Speaker: Yoosoo Oh
10:20~10:40	Coffee Break
10:40~12:10	Session 1 (Luminary Hall) Session 2 (Savoy Hall)
12:20~13:30	Lunch (Noori Restaurant, 3F)
13:30~14:10	Keynote Speech (Luminary Hall) Speaker: Moon Keun Kim
14:10~15:40	Session 3 (Luminary Hall) Session 4 (Savoy Hall)
15:40~16:00	Coffee Break
16:00~17:00	Session 5 (Poster & Capstone, Luminary Hall)
17:20~	Conference Banquet (Luminary Hall, 4F)

※Times are expressed in Nha Trang, Vietnam.
※The schedule is subject to change due to various circumstances without notice.

July 22, 2026	
09:00~10:00	Invited Speech 2 (Savoy Hall) Speaker: Sang Suh
10:00~10:10	Coffee Break
10:10~11:10	Session 6 (Savoy Hall)
11:10~11:20	Coffee Break
11:20~12:20	Session 7 (Poster & Capstone, Savoy Hall)
12:20~13:30	Coffee Break
13:30~15:00	Session 8 (Savoy Hall)
15:00~15:20	Coffee Break
15:20~16:20	Technical Committee on Smart Sensors
16:20~17:20	Technical Committee on Future Mobility
17:20~18:20	Technical Committee on Industrial Artificial Intelligence
July 23, 2026	
09:00~11:30	International Research Exchange & Networking: Panel Discussion on Future Directions
11:30~11:40	Committee Meeting
11:30	Closing Ceremony

※Times are expressed in Nha Trang, Vietnam.

※The schedule is subject to change due to various circumstances without notice.

SCIENTIFIC PROGRAM

July 20, 2026	
Time	Event
13:30 ~	Special Interest Group (SIG) Workshop on Emerging Research Topics I
15:30 ~	Coffee Break
16:30 ~	Special Interest Group (SIG) Workshop on Emerging Research Topics II

July 21, 2026	
Time	Event
09:00 ~	Registration
09:30 ~	Opening Ceremony
09:40 ~	Invited Speech 1 (Yoosoo Oh) (Luminary Hall) Chair: Sangmoon Lee (Kyungpook National Univ.)
10:20 ~	Coffee Break
10:40 ~	Session 1 (Luminary Hall) Chair: Jonghun Lee (DGIST)
	1S-1 WTOD Protocol-Based Resilient Fault Detection for Networked Menristive Neural <i>Jiong-Jian She ¹⁾ and Meng-Jie Hu ^{2*)}</i> ¹⁾ School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China ²⁾ School of Internet of Things, Xi'an Jiaotong-Liverpool University, Suzhou 215400, China
	1S-2 Error-Based Fuzzy Q-Learning Control for Trajectory Tracking under Initial Offsets <i>Myeongjun Seo ¹⁾ and Sangmoon Lee ^{2*)}</i> ¹⁾ Dept. of Water and IT Engineering, Kyungpook National University, Daegu, 41566, Republic of Korea ²⁾ Dept. of Electric Engineering, Kyungpook National University, Daegu, 41566, Republic of Korea

	1S-4 Radar-Based Human-Robot Discrimination Using Transformer with Sequential Cluster Features in Automotive Manufacturing Environments <i>Dokyung Kang¹⁾, Jieun Park¹⁾, Kyungtaeg Yang¹⁾, Sekyung Kim¹⁾, Seungeon Song²⁾ and Jonghun Lee^{*1),2)}</i> ¹⁾ Dept. of Interdisciplinary Engineering, Graduate school, DGIST, Daegu, 42988, Korea ²⁾ Institute of Research, DGIST, Daegu, 42988, Korea
	1S-5 Activity-Aware Sensor Data Compression for Energy-Efficient On-Device AI Computing in Human Activity Recognition <i>Ganapati Bhat¹⁾ and Hyung Gyu Lee²⁾</i> ¹⁾ Sch. of EECS, Washington State Univ., Pullman, WA, 99164, U.S.A ²⁾ Sch. of Digital Software Engineering, Duksung Women's Univ. Seoul 01369, South Korea.
10:40~	Session 2 (Savoy Hall) Chair: Joonhyuk Yoo (Daegu Univ.)
	2S-1 Comparative Analysis of UAV Position Prediction Models <i>Sang Suh^{1*)}, Doweon Ki²⁾, Jie Yu³⁾</i> ^{1,2,3)}Dept. of Computer Science, East Texas A&M University, Commerce, TX, U.S.A.
	2S-2 Unified Cross-Layer Embedded AI Design <i>Sang Suh^{1*)}, Omar Aly²⁾</i> ¹⁾Dept. of Computer Science, East Texas A&M University, Commerce, TX, U.S.A.
	2S-3 Miniaturized Broadband Microstrip Patch Antenna Utilizing Open-Circuit Stubs and Ground Slotting for 5G Systemssa <i>Zelin Hou¹⁾, Jie Zhang¹⁾, Ka Lok Man¹⁾, Yan Li²⁾ and Jingchen Wang^{1*)}</i> ¹⁾School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China ²⁾College of Software and Covergence Computer Engineering, Inha University, Incheon, Republic of Korea
	2S-4 Model Compression and Skin-Tone Fairness in On-Device Skin Cancer Screening <i>Pham Thi Anh Thu^{1*)}, Nguyen Gia Kho²⁾, Tran Trung Hieu³⁾, Bui Thi Thanh Tu⁴⁾</i> ^{1,2,3,4)} Dept. of Information Technology, Ho Chi Minh City University of Foreign Languages and Information Technology (HUFLIT), Ho Chi Minh City, Vietnam
	2S-5 Analytical and Deep Learning Method for Phase Error Calibration of FMCW Automotive Radar: A Comparative Analysis <i>Sharif Danish¹⁾, Sheriff Murtala²⁾ and Gyu Sang Choi^{3*)}</i> ^{1,2)} Dept. of Information and Communication Engineering, Yeungnam University, Korea ³⁾ Dept. of Information and Communication Engineering, Yeungnam University, Korea
12:10 ~	Lunch
13:30 ~	Keynote Speech (Moon Keun Kim) (Luminary Hall) Chair: Yoosoo Oh (Daegu Univ.)
14:10 ~	Session 3 (Luminary Hall) Chair: Hyunggyu Lee (Duksung Women's Univ.)
	3S-1 Residual Reinforcement Learning Compensation for ADRC-Based Waypoint Tracking of a Solar Pile Driver <i>Myeong Woo Lee¹⁾, Taeyong Kim²⁾, Miok Kim³⁾, Hungju Shin⁴⁾, and Moon G. Joo^{5*)}</i> ^{1,5)} Department of Intelligent Robot Engineering, Pukyong National University, Busan, Korea ^{2,3,4)} Xpanner Co., Ltd., Busan, Korea
	3S-2 6-Degree-of-Freedom Pose Estimation of Solar Panels Using Vision Transformer and Synthetic CAD-Based Dataset <i>Tae Young Jeong¹⁾, Miok Kim²⁾, Hungju Shin³⁾, Moon G. Joo^{4*)}</i> ^{1,4)} Dept. of Intelligent Robot Engineering, Pukyong National University, Busan 48513, Korea ^{2,3)} Xpanner Co., Ltd.

	3S-2 6-Degree-of-Freedom Pose Estimation of Solar Panels Using Vision Transformer and Synthetic CAD-Based Dataset <i>Tae Young Jeong¹⁾, Miok Kim²⁾, Hungju Shin³⁾, Moon G. Joo^{4*)}</i> ^{1,4)} Dept. of Intelligent Robot Engineering, Pukyong National University, Busan 48513, Korea ^{2,3)} Xpanner Co., Ltd.
	3S-3 Language-Guided Unknown-Space Exploration for Mobile Robots with a Mamba-Based Frontier Planner <i>Moojueng Kim¹⁾, Junhui Woo²⁾ and Sangmoon Lee^{3*)}</i> ^{1, 2)} Dept. of Water and IT Eng., Kyungpook Natl. Univ., Daegu 41566, Korea ³⁾ Dept. of Electronic Eng., Kyungpook Natl. Univ., Daegu 41566, Korea
	3S-5 Digital Twin Network based on Decentralized Federated Learning in On-device AI Environments <i>Nguyen Anh Tuan^{1,2)}, Qazi Waqas Khan^{1, 2)}, Le Anh Ngoc³⁾ and Do Hyeun Kim^{1,2 *)}</i> ¹⁾ Dept. of Computer Engineering, Jeju National University, Jeju-si, Republic of Korea ²⁾ Big Data Research Center, Jeju National University, Jeju-si, Republic of Korea ³⁾ Swinburne Vietnam, FPT University, Hanoi, Vietnam
	3S-5 Real-Time Face Recognition AI-Based Classroom Seat Matching and Student Identification System <i>Donghwoon Kwon¹⁾, Seung Kwang Ryu²⁾, Seok-Hyeon Jeong³⁾, Ayoung Song³⁾, and Jeong Tak Ryu^{3*)}</i> ¹⁾ Dept. of Computer Science and Engineering, North Central College, Naperville, IL 60540, U.S.A ²⁾ Graduate School of Smart Convergence Systems Engineering, Daegu University, Korea ³⁾ Department of Electronic Engineering, Daegu University, Korea
14:10 ~	Session 4 (Savoy Hall) Chair: Moon G. Joo (Pukyong National Univ.)
	4S-1 Enhancing Convolutional Neural Network Representation Capability through Feature Diversity <i>Yeongyu Choi^{1*)} and Ju H. Park¹⁾</i> ¹⁾ Dept. of Electrical Engineering, Yeungnam University, Gyeongsan 38541, Korea
	4S-2 Explainable Optimization of Precision Tooling Design Using Physics-Informed Generative Adversarial Networks for Small Data Regimes <i>ByeongSeok Lee¹⁾, Yoosoo Oh²⁾</i> ¹⁾ Department of Smart Convergence Systems Engineering, Daegu University ²⁾ School of Computer and Information Engineering, Daegu University
	4S-3 HyperCup: A Mobile Coffee Cupping Platform <i>Seonghee Min^{1*)}</i> ¹⁾ Cuppingpost Roasters, Gyeongsan-si, Gyeongsangbuk-do, 38426, Korea
	4S-4 Design of a Fine-Grained Classification Algorithm for Specialty Green Beans <i>Seonghee Min^{1*)}</i> ¹⁾ Cuppingpost Roasters, Gyeongsan-si, Gyeongsangbuk-do, 38426, Korea
	4S-5 SoilSeg: Severity-Aware Lens Soiling Segmentation for Camera Reliability in Autonomous Driving <i>Muhammad Tayyab Shehzad¹⁾, Gyu Sang Choi^{1*)}</i> ¹⁾ Dept. of Information and Communication Engineering, Yeungnam University, Gyeongsan-si, Korea
15:40 ~	Coffee Break
16:00 ~	Session 5 (Poster & Capstone) (Luminary Hall) Chair: Jeongtak Ryu (Daegu Univ.)

5S-2	Functional and Performance Evaluation Method for Core Components of Wide-Area Low-Floor Electric Buses <i>Sung-hun Kim¹⁾, Seong-Wook Cho²⁾, Oh-Hun Kwon²⁾, Jang-Kyu Yun²⁾, Soo-Young Suk²⁾</i> ^{1,2)} Electronic Components Research Center, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38463, Korea
5S-3	Automatic Instance-Segmentation Labeling of Rice Cookers Using SAM3 Exemplar Matching <i>Jaehyun Park¹⁾, Seungtaek Kim¹⁾, Sunjong Lee¹⁾ and Youngjun Yoo^{1*)}</i> ¹⁾ Korea Institute of Industrial Technology, Cheonan, 31056, Republic of Korea
5S-4	Low-Complexity Two-Stage Operation FMCW Beam-Scanning Radar for Embedded Systems <i>YoungSeok Jin¹⁾, Eugin Hyun^{1*)}</i> ¹⁾ IS ³ Lab, Dept. of Mobility Technology, DGIST, Daegu, Korea
5S-5	A Standards- and BMS-Based Evaluation Framework for Classification and Safety Assessment of EOL EV Batteries <i>Seong-wook Cho¹⁾, Sung-Hun Kim¹⁾, Oh-hoon Kwon¹⁾, Jang-kyu Yun¹⁾ and Soo-young Suk^{2*)}</i> ¹⁾ Electronic & Autonomous Systems Center, Future Mobility Research Division, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38542, Republic of Korea ²⁾ Future Mobility Research Division, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38542, Republic of Korea
5S-6	Development of a Field Inspection Procedure for Electric Vehicles Equipped with End-of-Life Batteries Using a Chassis Dynamometer <i>Seung-woo Kim^{1*)}, Oh-hun Kwon¹⁾, Hyun-kwon Choi¹⁾, Jang-kyu Yun¹⁾ and Soo-young Suk¹⁾</i> ¹⁾ Center for Automotive Electronics and Autonomous Systems, Gyeongbuk Institute of IT Convergence Industry Technology, Gyeongsan 38463, Korea
5S-7	Automatic Generation of Robot Offline Programs from Natural-Language Instructions Using a Large Language Model <i>Sanga Lee¹⁾, Jongpil Yun¹⁾, Hongin Won¹⁾, Hyunchul Tae¹⁾ and Youngjun Yoo^{1*)}</i> ¹⁾ Korea Institute of Industrial Technology, Republic of Korea
5S-8	Establishment of EMC Test Evaluation Environment Using Hanging Mast Antenna for ADAS and Autonomous Systems <i>Sungyoun Hwang¹⁾, Junbyung Baek¹⁾, Hyojeung Cho¹⁾, Simok Lee¹⁾, Jaerok Shim¹⁾, and Taehee Lee^{1*)}</i> ¹⁾ Test & Evaluation Division, Korea Intelligent Automotive Parts Promotion Institute, Daegu, Korea
5S-9	Development of an IR LED-Based Illumination System for Parking Violation Enforcement Cameras <i>Jung-Hae Choi¹⁾ and Byung-Jae Choi^{2*)}</i> ¹⁾ DGELEC, Yeongcheon-si, Gyeongsangbuk-do 38453, Korea ²⁾ Dept. of Electrical Engineering, IT & Engineering College, Gyeongsan-si, Gyeongsangbuk-do 38453, Korea

July 22, 2026	
Time	Event
09:00 ~	Invited Speech 2 (Sang Suh) (Savoy Hall) Chair: Wonkee Hong (Daegu Univ.)
10:00 ~	Coffee Break

09:00 ~	Invited Speech 2 (Sang Suh) (Savoy Hall) Chair: Wonkee Hong (Daegu Univ.)
10:00 ~	Coffee Break
10:10 ~	Session 6 (Savoy Hall) Chair: Jeseok Kim (DGIST) 6S-1 STFT-CNN Based Fall Detection Using Millimeter-Wave Radar <i>Jieun Park¹⁾, Kyungtaeg Yang¹⁾, Sekyung Kim¹⁾, Dokyung Kang¹⁾, Seungeon Song²⁾ and Jonghun Lee^{*1),2)}</i> ¹⁾ Dept. of Interdisciplinary Engineering, Graduate school, DGIST, Daegu, 42988, Korea ²⁾ Institute of Research, DGIST, Daegu, 42988, Korea 6S-2 On-Device AI-Based Real-Time Mobile Fall Detection and Notification Service with Interworking Server <i>Yu Gong¹⁾, Eun-Bin Koh²⁾, Misbah Bibi²⁾ and Do-Hyun Kim^{2*)}</i> ¹⁾Dept. of Japanese Language & Literature / Intelligent IoT Convergence, Jeju National University, Jeju, Korea ²⁾Dept. of Computer Engineering, Jeju National University, Jeju, Korea 6S-3 Deep Reinforcement Learning-Based Lumen Alignment for Defect Inspection of Catheter Cross-section <i>Jae Jin Hwang¹⁾, Yejin Kim²⁾, Jong Pil Yun³⁾ and Moon G. Joo^{4*)}</i> ¹⁾Dept. of Intelligent Robotics Engineering, Pukyong National University, Busan, Korea ^{2,3)}Manufacturing AI Research Center, Korea Institute of Industrial Technology, Incheon, Korea ^{4*)}Dept. of Information and Communication Engineering, Pukyong National University, Busan, Korea
11:10 ~	Coffee Break
11:20 ~	Session 7 (Poster & Capstone) (Savoy Hall) Chair: Dohyeon Kim (Jeju National Univ.) 7S-1 A Scenario-Based Electromagnetic Susceptibility Evaluation Method for ADAS Functions Using RTS and Camera Projection <i>Junbyung Baek¹⁾, Sungyoun Hwang¹⁾, Hyeongsu Choi¹⁾, Doyoun Kim¹⁾, Simok Lee¹⁾, Jaerok Shim¹⁾, Hyojeung Cho¹⁾ and Taehee Lee^{1*)}</i> ¹⁾Test & Evaluation Division, Korea Intelligent Automotive Parts Promotion Institute, Daegu, Korea 7S-2 Repeatability Factors in Electric Vehicle Single-Charge Range Tests Using Chassis Dynamometers <i>Jeong-gon Kim¹⁾, Myeong-Ho Song¹⁾, Hyeong-Jin Hwang¹⁾, and Jang-Kyu Yun¹⁾</i> ¹⁾Gyeongbuk Institute of IT Convergence Industry 7S-3 Implementation of a Lightweight Machine Learning-Based Virtual Motor Speed Sensor Using a MEMS IMU <i>Jaeseong Lee¹⁾, Gwang-Min Park¹⁾, Seonghun Lee¹⁾</i> ¹⁾AX Research Group for Robotics, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Korea 7S-4 Torque Robustness Enhancement in PMSM Drives with Low-Cost Sensors via Active Phase Trimming <i>Gwangmin Park¹⁾, Jaeseong Lee¹⁾, Seonghun Lee¹⁾</i> ¹⁾AX Research Group for Robotics, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Korea 7S-5 Development of a Real-Time Integrated Monitoring System for Battery Pack Manufacturing Process <i>MyeongSu Jeong¹⁾, ChangSoo Moon^{2*)} and GiDong Lee³</i> ^{1,2)} Dept. of Vehicle Software, Gyeongbuk Research Institute of Vehicle Embedded Technology, YeongCheon,

	7S-6	An Explainable Client-Side Framework for Real-Time Phishing Detection Using URL Risk Scoring and Safe Browsing APIs <i>Min-Gi Choi¹⁾, Sol-mi Kim²⁾ and Won-Kee Hong^{3*)}</i> <i>^{1,2,3)} Dept. of Inform. Security, Daegu Univ., Gyeongsan, Gyeongbuk 121-791, Korea</i>
	7S-7	An Embedded AI Software Integration Supporter for Safe Human-in-the-Loop Code Modification in Ubuntu <i>Jae-Mo Ahn¹⁾, Jung-Hae Choi²⁾, Sijia Huo¹⁾, Byoung-Ju Yun³⁾, and Byung-Jae Choi^{1*)}</i> <i>¹⁾Department of Electrical Engineering, Daegu University, Gyeongsan 38453, Republic of Korea</i> <i>²⁾Chief Executive Officer, Daekyung Elec Co., Ltd., Republic of Korea</i> <i>³⁾School of Electronics Engineering, College of IT Engineering, Kyungpook National University, Daegu, Republic of Korea</i>
	7S-8	Design and Implementation of a Real-Vehicle VIL Testbed for Autonomous Driving Validation with Virtual Sensor Integration <i>Je-Seok Kim¹⁾</i> <i>¹⁾ Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu, Korea</i>
	7S-9	Design and Implementation of an AI-Based Smart Aquaculture System for Feeding Optimization and Automated Farm Management <i>Jaeyoung Kim^{1*)}, Dongseung Shin²⁾ and Laekyoung Lee³⁾</i> <i>^{1,2,3)} Daegu-Gyeongbuk Research Division, Electronics and Telecommunications Research Institute, Daegu, Republic of Korea</i>
	7S-10	Factors Affecting the Acceptance and Purchase Intention of Smart Rings: An Extended Technology Acceptance Model Approach <i>Sang Hoon Lee¹⁾ and Su-Yeon Kim^{2*)}</i> <i>¹⁾Software Education Institute, Kyungpook National Univ., 80, Daehak-ro, Buk-gu, Daegu, Korea</i> <i>²⁾School of Computer and Information Engineering, Daegu Univ., Gyeongsan, Gyeongbuk 38453, Korea</i>
12:10 ~	Coffee Break	
13:30 ~	Session 8 (Savoy Hall) Chair: Jeongtak Ryu (Daegu Univ.)	
	8S-1	EmbedLearn: On-Device Intelligence for Smart Classroom Monitoring <i>T.Velmurugan¹⁾, Senthilkumar Ramamoorthy²⁾, Sridevi PC^{3*)}</i> <i>¹⁾ Advisor, Meston College of Arts and Science, Chennai 600014, India</i> <i>²⁾Senior Java /AI Engineer, The Options Clearing Corporation (OCC), USA</i> <i>³⁾ Dept. of Artificial Intelligence and Data Science, Takshashila University, Tindivanam 604305, India</i>
	8S-2	Next-generation Cloud-ATM Architecture Using Enhanced Hybrid Elliptic Curve Cryptography for Eradicating Skimming Attacks <i>SriPradha.G¹⁾, R.Samundeeswari²⁾</i> <i>^{1,2)} Asst. Professor, Dept. of Information Technology, Guru Nanak College, Chennai, India -600042</i>
	8S-3	A Multi-Joint Data-Based SOMA System for Fine Motor Development in Children with Special Needs <i>Narangerel Tsolmon¹⁾, Jimin Ryu²⁾ and Yoosoo Oh^{3*)}</i> <i>^{1,2)} NWI Co., Ltd., Gyeongsan 38453, Republic of Korea</i> <i>³⁾ School of Computer & Information Engineering, Daegu University, Gyeongsan 38453, Republic of Korea</i>
	8S-4	AI-Based Multi-Agent Crowd Safety Surveillance System for Real-Time Threat Monitoring <i>Madhusudhana R¹⁾ and Rakesh Kalshetty¹⁾</i>
	8S-5	AI in Healthcare: An Umbrella Review of Systematic Reviews and Meta-Analyses

	<i>Sang Suh¹⁾ and Kshitiz Basnet¹⁾</i> ¹⁾Dept. of Computer Science, East Texas A&M University-Commerce, TX, USA
8S-6	Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection <i>Bhuthamekala prasanthi¹⁾, Mudiya jhansi lakshmi²⁾</i> ¹⁾Dept. of Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopeta, India ²⁾Dept. Computer Science and Engineering, Narasaraopeta Engineering College, Narasaraopeta, India
8S-7	Achieving Low-Latency and Fair NR-V2X Communication via UAV-Assisted Multi-Agent Reinforcement Learning <i>Maliha Binte Mukhles¹⁾, M Ishtiaque Ahmed²⁾ and Sung Won Kim^{3*)}</i> ^{1,3)} Department of Information & Communication Engineering, Yeungnam University, Gyeongsan-si, 38541, Republic of Korea ²⁾ Department of Environment Science, Jahangirnagar University, Savar, Dhaka, Bangladesh
8S-8	TinyECG-Gate: Ultra-Lightweight INT8 ECG Edge Triage with Selective Confidence Deferral <i>Kaleem Ahmad¹⁾ and Sung Won Kim^{2*)}</i> ^{1,2)} Department of Information and Communication Engineering, Yeungnam University, Republic of Korea
15:00 ~	Coffee Break
15:20 ~	Technical Committee on Smart Sensors
16:20 ~	Technical Committee on Future Mobility
17:20 ~	Technical Committee on Industrial Artificial Intelligence

July 23, 2026	
Time	Event
09:00 ~	International Research Exchange & Networking: Panel Discussion on Future Directions
11:30 ~	Committee Meeting
11:30	Closing Ceremony

AI-Based Energy Demand Forecasting for Commercial Stores: Comparative Analysis of CNN-LSTM and Transformer Approaches

Moon Keun Kim ^{1*)} and Sanghyuk Lee ²⁾

¹⁾ Dept. of Built Environment, Oslo Metropolitan University, Oslo, N-0130, Norway

²⁾ School of Computer Science and Engineering, American University in Vietnam, Danang, 550000, Vietnam

*E-mail: Moon.Kim@oslomet.no

Abstract: This study investigates dynamic energy demand forecasting in Norwegian commercial retail buildings using ANN-based spatiotemporal models, including CNN-LSTM and Transformer architectures. Electricity consumption data from multiple retail stores in the Oslo region were analyzed together with neighboring-store energy-use interactions to improve prediction performance. The results show that weather data alone provide limited forecasting capability, while incorporating surrounding buildings' consumption patterns significantly enhances prediction accuracy and stability. Transformer models performed better under limited training data conditions, whereas CNN-LSTM models achieved higher accuracy with larger datasets. Overall, the study demonstrates the effectiveness of both architectures for commercial energy demand forecasting and highlights the importance of shared energy-consumption dynamics among nearby buildings.

Keywords : Hybrid convolutional neural network, Transformer, Convolutional Neural Network and Long Short-Term Memory Networks, Commercial Store Building

Introduction: This study presents a spatiotemporal deep learning approach for forecasting energy demand in commercial retail buildings in Oslo, Norway. CNN-LSTM and Transformer models were applied using electricity consumption and weather time-series data collected from multiple surrounding stores. Unlike conventional methods based only on weather variables, the proposed approach captures operational interactions, customer behavior, and occupancy-driven energy patterns among neighboring buildings. The results demonstrate that incorporating surrounding-store energy data substantially improves forecasting accuracy and robustness. Transformer models showed strong performance with limited datasets, while CNN-LSTM models achieved superior accuracy with larger datasets..

Research Methods: Recent advances in building energy forecasting have increasingly adopted deep learning-based time-series models, particularly CNN-LSTM and Transformer architectures, due to their strong capability in capturing complex temporal and spatial patterns. This study investigates dynamic energy demand prediction for commercial retail buildings in Oslo, Norway, using hourly electricity consumption data from eleven stores together with local weather variables. Unlike conventional forecasting methods that rely mainly on meteorological inputs, the proposed approach incorporates neighboring-store energy consumption patterns to reflect operational interactions, customer behavior, occupancy schedules, and surrounding environmental influences. CNN-LSTM models effectively capture nonlinear spatial-temporal correlations through combined convolutional and recurrent learning structures, while Transformer models utilize self-attention mechanisms to improve long-sequence forecasting efficiency and stability. Three prediction scenarios were evaluated using neighboring-store clusters, weather-only variables, and integrated datasets combining surrounding-store and weather information. The forecasting performance was assessed using MAE, RMSE, and R^2 metrics. The results demonstrate that integrating neighboring-store energy dynamics significantly improves prediction accuracy and robustness compared with weather-only approaches. Transformer models showed superior performance under limited training datasets, whereas CNN-LSTM models achieved higher prediction accuracy when larger datasets were available. Overall, the study highlights the importance of spatiotemporal energy interactions for accurate commercial building energy forecasting and demonstrates the practical potential of advanced ANN architectures for intelligent energy management applications.

Results: This study applied two ANN-based models, CNN-LSTM and Transformer, to predict electricity consumption in a commercial building located in Oslo, Norway, using 8,784 hourly datasets collected over one year. The models were trained using raw datasets containing local weather variables and surrounding commercial store energy consumption patterns. The results demonstrated that surrounding-store energy data significantly improved forecasting accuracy, while weather variables alone had limited influence on prediction performance. Both models showed strong agreement between predicted and measured electricity consumption values, particularly when larger surrounding-store datasets were incorporated. The Transformer model achieved better performance under limited training datasets, whereas the CNN-LSTM model outperformed the Transformer when larger datasets from neighboring stores were available. The best performance was achieved by the CNN-LSTM model using the 10-store dataset, with MAE = 0.90 and RMSE = 1.17. Figure 1 illustrates the performance of two ANN models. The findings indicate that commercial building energy demand is more strongly influenced by operational schedules, customer occupancy, and surrounding-store activity than by outdoor weather conditions. Overall, the study highlights the effectiveness of spatiotemporal ANN models for dynamic commercial building energy forecasting and demonstrates the importance of incorporating neighboring-building energy interactions into prediction frameworks.

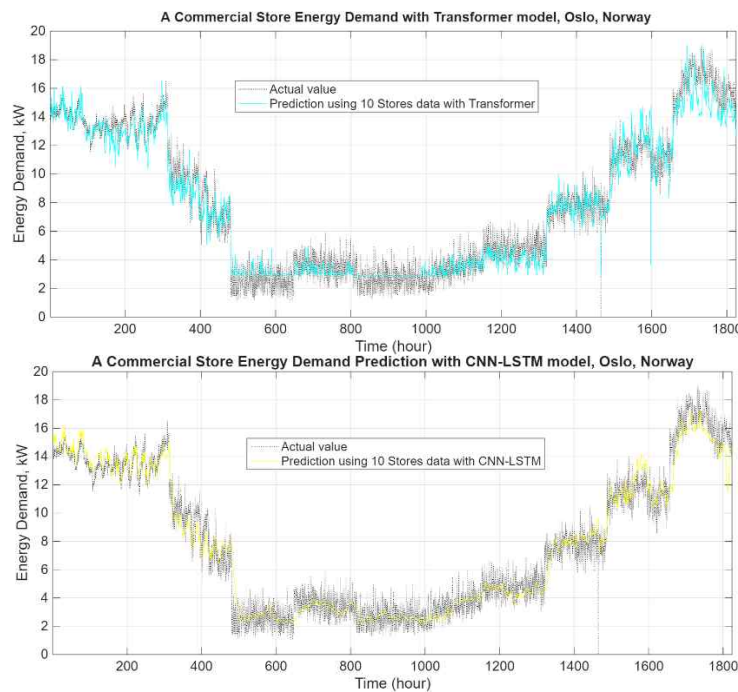


Fig. 1. Electricity consumption forecasting using surrounding commercial store datasets: comparison between the CNN-LSTM model (top) and the Transformer model (bottom).

Conclusion: This study compares CNN-LSTM and Transformer models for dynamic energy demand forecasting in commercial retail buildings in Oslo, Norway. The results show that weather data alone are insufficient for accurately predicting commercial store energy consumption, which is strongly influenced by operational schedules, customer behavior, and occupancy patterns. Incorporating surrounding-store energy consumption data significantly improved prediction accuracy and stability. Transformer models performed better with limited datasets, while CNN-LSTM models achieved superior accuracy with larger datasets from neighboring stores. Overall, the study highlights the importance of spatiotemporal energy interactions for reliable commercial building energy forecasting.

References

1. M.K.Kim, Y.S. Kim, J.Srebric, *J. Sustain. City and Soc.*, 62. (2020).
2. Y.S. Kim, M.K.Kim, N.Fu, J.Liu, J. Wang, J.Srebric, *J. Sustain. City and Soc.*, 118. (2025).
3. M.K.Kim, Y.S. Kim, N.Fu, J.Liu, J. Wang, S. Lee, J.Srebric, *J. Energy Report*, 14. (2025).

Agentic Multi-Agent Evolution via Adversarial Reasoning and Memory Distillation

Sang C. Suh, Harshitha Divakarreddy*

Dept. of Computer Science, East Texas A&M University, Commerce, TX, U.S.A.

**E-mail: Sang.Suh@etamu.edu*

Abstract

Large language models are increasingly deployed as small agent “teams” that draft answers, critique each other, and verify decisions in domains such as education design, IoT security, wireless networking, and finance. The idea is appealing because role separation can make reasoning more careful. However, many of these systems still behave like short-term problem solvers. They may debate their way to a better output in one episode, yet the same mistakes often appear again in the next run. Debate-based multi-agent verification improves episode-level checking, but it can still drift toward persuasive confidence unless verification rules are explicit. Memory-augmented agents help retain context and experience, but raw memory can become noisy, bloated, or too case-specific to transfer cleanly.

This paper focuses on whether structured disagreement can be turned into durable, reusable competence without quietly inflating prompts or relying on retraining. To answer this, we propose a framework that couples a bounded challenge–defend–verify protocol, using Proposer, Challenger, and Verifier roles, with a validation gate that stores only verified “skill artifacts” and a round-level distillation step that compacts those artifacts into a distilled guidance module for later rounds.

We evaluate the approach over two rounds against debate-only and memory-only baselines using Anthropic Claude on FEVER, MuSiQue, and BBH Logical Deduction. Across 750 instances per round per condition, the full method shows stronger Round 2 performance than both baselines, reaching 76.0% on FEVER, 63.0% on MuSiQue, and 66.0% on BBH. Verifier rejection drops from 31.1% to 18.9%, acceptance rises from 68.9% to 81.1%, and coordination cost also decreases between rounds, suggesting that distilled guidance begins to replace repeated challenge–repair cycles rather than simply adding overhead.

Keywords: Agentic AI, multi-agent systems, adversarial reasoning, memory distillation, large language models

Introduction

Agentic systems built on large language models (LLMs) are beginning to behave less like single-shot text generators and more like software coworkers: planners that decompose goals, callers that invoke tools, and collaborators that negotiate decisions. Recent trends suggest this shift in applied settings such as multi-agent pipelines that explain and back test financial decisions, teams that secure consumer IoT systems against anomalies, and instructional-design agents that adapt lesson plans on the fly [1]–[3]. These developments push agents into environments where errors can accumulate over time and where the process of reasoning matters as much as the final answer. Yet in practice, today’s agent teams often feel like short-lived committees. They can deliberate sharply in the moment, but they do not reliably evolve; the best reasoning is frequently lost, and the same failure patterns resurface across tasks [4]–[6].

Research Methods

This work evaluates whether a tightly coupled adversarial reasoning and memory distillation loop can produce durable improvements in LLM-based multi-agent performance across repeated tasks, rather than merely improving a single run. As shown in Fig. 1, each episode starts with a task instance and its context, which is given to a three-role agent team: a Proposer, a Challenger, and a Verifier. The Proposer drafts a solution, the Challenger tries to expose flaws by pointing out missing evidence, counterexamples, edge cases, or steps that do not really follow, and the Verifier acts as a gatekeeper that decides whether the result is acceptable [7]–[9]. If the Verifier accepts an episode, the system stores validated artifacts rather than raw conversational transcripts, following the concern that unmanaged memory can preserve noise, drift, or overly case-specific traces [10]–[12]. These artifacts include reusable rules, boundary

conditions, counterexamples, failure patterns, and brief verification notes. At the end of each round, the validated artifacts are compacted and compiled into a distilled guidance module, so later episodes can reuse verified corrections without simply expanding the prompt.

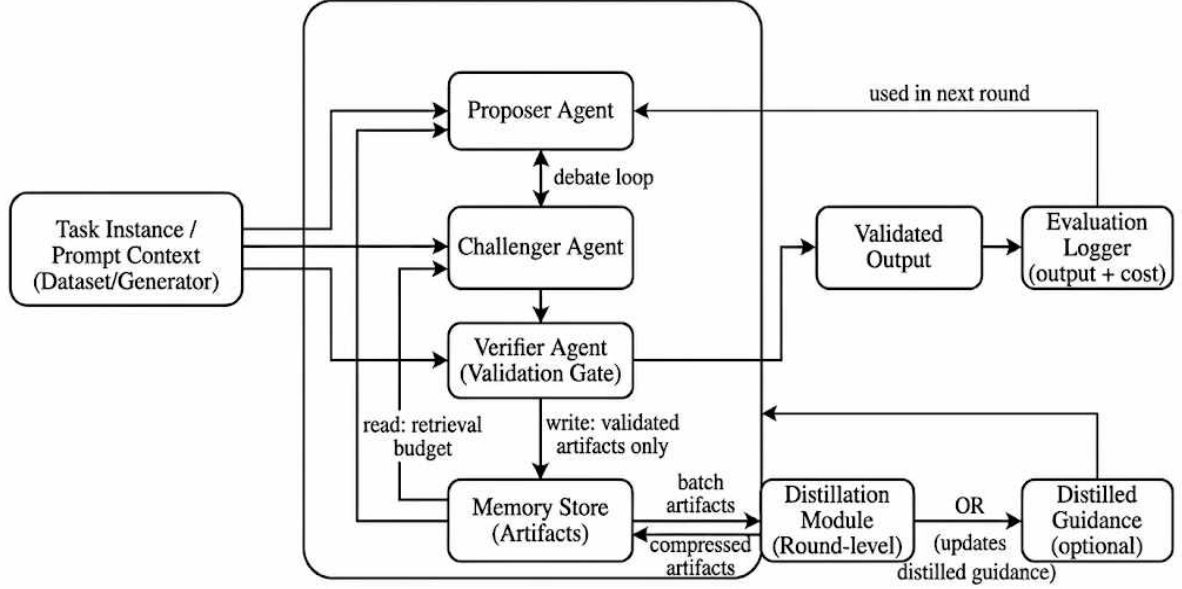


Fig. 1. Overview of the proposed multi-agent architecture with challenge-defend-verify interaction, validated artifact memory, and round-level distillation

Experimental Results

Results are reported across three conditions: debate-only, memory-only, and the full method, over two rounds and three datasets. For each condition and each round, we evaluated 300 FEVER instances, 200 MuSiQue instances, and 250 BBH Logical Deduction instances, giving 750 episodes per round per condition [13]–[15]. Table 1 presents the episode success rates by dataset, condition, and round, with the key comparison being whether Round 2 performance improves under the full method more than it does under the baselines. The full method produces the largest Round 1 to Round 2 gains across all three datasets, reaching 76.0% on FEVER, 63.0% on MuSiQue, and 66.0% on BBH in Round 2. Verifier rejection also drops substantially for the full method, from 31.1% to 18.9%, while acceptance rises from 68.9% to 81.1%. Interaction cost decreases as well, with mean turns dropping from 4.3 to 3.8 and mean total tokens dropping from 2,390 to 2,280. This pattern suggests that distilled guidance begins to substitute for some repeated challenge-repair cycles by pre-loading verified rules and counterexamples into the episode context.

Table 1. Episode Success by Condition, Dataset, and Round

Condition	FEVER R1 (n=300)	FEVER R2 (n=300)	MuSiQue R1 (n=200)	MuSiQue R2 (n=200)	BBH R1 (n=250)	BBH R2 (n=250)
Debate-only	190 / 63.3%	194 / 64.7%	96 / 48.0%	98 / 49.0%	142 / 56.8%	145 / 58.0%
Memory-only	180 / 60.0%	202 / 67.3%	90 / 45.0%	108 / 54.0%	138 / 55.2%	153 / 61.2%
Full method	192 / 64.0%	228 / 76.0%	100 / 50.0%	126 / 63.0%	145 / 58.0%	165 / 66.0%

Conclusion

This paper presented a framework for agentic multi-agent evolution that couples structured adversarial reasoning with memory distillation, with the goal of turning disagreement into durable, reusable competence. The method uses a bounded challenge–defend–verify protocol to generate corrective signals, a Verifier gate to ensure only validated artifacts enter memory, and a round-level distillation pipeline that compacts artifacts into a distilled guidance module reused in later rounds. The results across FEVER, MuSiQue, and BBH Logical Deduction suggest that the full method achieves stronger round-to-round improvements than debate-only and memory-only baselines, while also reducing Verifier rejection rates and improving acceptance over time. Future work will extend evaluation with larger instance sets, add a no-distillation ablation, and incorporate robustness stress tests motivated by known multi-agent vulnerabilities.

References

1. H. S. Jung and H. Lee, “Explainable zero-shot trading using multi-agent LLM architecture: A backtested approach for Bitcoin price,” *Inf. Process. Manage.*, vol. 63, no. 2, Art. no. 104466, 2026, doi: 10.1016/j.ipm.2025.104466.
2. J. Yang et al., “NeuroAgent-X: A self-evolving cognitive agent for securing consumer IoT systems against AI-enabled anomalies and adversarial threats,” *IEEE Trans. Consum. Electron.*, vol. 71, no. 4, pp. 12226–12235, 2025, doi: 10.1109/TCE.2025.3627970.
3. X. Zhang et al., “EduPlanner: LLM-based multiagent systems for customized and intelligent instructional design,” *IEEE Trans. Learn. Technol.*, vol. 18, pp. 416–427, 2025, doi: 10.1109/TLT.2025.3561332.
4. F. Jiang et al., “Large language model enhanced multi-agent systems for 6G communications,” *IEEE Wireless Commun.*, vol. 31, no. 6, pp. 48–55, 2024, doi: 10.1109/MWC.016.2300600.
5. Z. Zhang et al., “Igniting language intelligence: The hitchhiker’s guide from chain-of-thought reasoning to language agents,” *ACM Comput. Surv.*, vol. 57, no. 8, Art. no. 206, 2025, doi: 10.1145/3719341.
6. S. Yoon, “Strategic learning under linguistic and contextual constraints: A theoretical framework for LLM-based multi-agent coordination,” *IEEE Access*, vol. 13, pp. 197053–197071, 2025, doi: 10.1109/ACCESS.2025.3628927.
7. H. Wang et al., “Learning to break: Knowledge-enhanced reasoning in multi-agent debate system,” *Neurocomputing*, vol. 618, Art. no. 129063, 2025, doi: 10.1016/j.neucom.2024.129063.
8. X. Ma et al., “Guided and knowledgeable multi-agent debate for fact verification,” *Expert Syst. Appl.*, vol. 299, Art. no. 130103, 2026, doi: 10.1016/j.eswa.2025.130103.
9. X. Zou, X. Li, P. Hu, and M. Dong, “MRBalance: A framework for enhancing event causality identification in multi-agent debates via role assignment,” *Knowl.-Based Syst.*, vol. 330, Art. no. 114470, 2025, doi: 10.1016/j.knosys.2025.114470.
10. K. Wang et al., “Toward interpretable and persistent personalization: A memory-augmented agent framework for LLM-based travel planning,” *IEEE Access*, vol. 13, pp. 193125–193141, 2025, doi: 10.1109/ACCESS.2025.3630495.
11. X. Liang et al., “SAGE: Self-evolving agents with reflective and memory-augmented abilities,” *Neurocomputing*, vol. 647, Art. no. 130470, 2025, doi: 10.1016/j.neucom.2025.130470.
12. Y. K. Saheed and J. E. Chukwuere, “Autonomous LLM agent: A memory-augmented, edge-optimized SHAP explanations with zero-day attack resilience in IoT/IIoT networks,” *IEEE Internet Things J.*, early access, 2025, doi: 10.1109/JIOT.2025.3648649.
13. “FEVER: Fact Extraction and Verification,” official dataset release. [Online]. Available: <https://fever.ai/dataset/fever.html>. Accessed: Apr. 2026.
14. “MuSiQue: Multi-hop Questions,” official dataset repository. [Online]. Available: <https://github.com/stonybrooknlp/musique>. Accessed: Apr. 2026.
15. “BIG-Bench Hard (BBH),” official benchmark repository. [Online]. Available: <https://github.com/suzgunmirac/BIG-Bench-Hard>. Accessed: Apr. 2026.

WTOD Protocol-Based Resilient Fault Detection for Networked Memristive Neural Networks Under Cyber-Attacks

Jiong-Jian She¹⁾, Meng-Jie Hu^{2*)}

¹⁾ School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China.

²⁾ School of Internet of Things, Xi'an Jiaotong-Liverpool University, Suzhou 215400, China.

*E-mail: humjynu@gmail.com

Abstract: This article addresses the resilient fault detection filtering (FDF) problem for Markov switching memristive neural networks (MSMNNs). A Markov switching attack model is introduced to describe the stochastic evolution of denial-of-service (DoS) attacks. An adaptive event-triggered weighted try-once-discard (AET-WTOD) protocol is developed to enhance communication efficiency. Using the detected modes, an asynchronous FDF is constructed. Sufficient conditions are obtained to guarantee mean square exponential stability (MSES) and prescribed $\mathcal{H}_\infty$ performance of the system.

Keywords : Event-triggered mechanism, Memristive Neural Networks, cyber-attacks.

Introduction: Memristive neural networks (MNNs) offer superior computational efficiency and memory performance, but undetected faults can severely degrade system performance and stability. Frequent data transmissions require efficient event-triggered protocols (ETPs). The weighted try-once-discard (WTOD) protocol is proposed to coordinate multi-node conflicts. Furthermore, realistic DoS attacks exhibit mode-switching characteristics, leading to heterogeneous measurement losses.

Research Methods: The problem description begins with a system model subject to polytopic uncertainty:

$$\begin{cases} x(k+1) = -(\bar{\mathcal{A}}_q + \Delta \mathcal{A}_q(k))x(k) + \varepsilon_q \omega(k) + (\bar{\mathcal{B}}_q + \Delta \mathcal{B}_q(k))g(x(k)) + \mathcal{F}_q f(k) \\ \quad + \mu(k)(\bar{\mathcal{D}}_q + \Delta \mathcal{D}_q(k))g(x(k - \tau(k))) \\ y(k) = \mathcal{H}_q x(k) \end{cases} \quad (1)$$

The specific triggering and transmission mechanism is formulated as $t_n = \min\{k | k > t_{n-1}, \alpha(t_{n-1})y^\top(k)\Pi_q y(k) - \epsilon_k^\top \Pi_q \epsilon_k \leq 0\}$, where $\epsilon_k \triangleq y(k) - y(t_{n-1})$, $y(t_{n-1})$ is the latest triggered measurement and Π_q is a mode-dependent weighting matrix. The triggering parameter $\alpha(t_n)$ is updated adaptively as $\alpha(t_n) = \beta(t_n)\alpha(t_{n-1})$ and $\beta(t_n) = 1 - \frac{\alpha(t_n)}{\alpha(t_{n-1})} [\mathfrak{S}_1(t_n)]\mathfrak{Y}(t_n)$. Define $\mathfrak{S}_1(t_n) = \frac{\|y(t_n)\| - \|y(t_{n-1})\|}{\|y(t_n)\|}$, $\mathfrak{S}_2(t_n) = \frac{\|y(t_n) - y(t_{n-1})\|}{\|y(t_n)\|}$, $\mathfrak{Y}(t_n) = \begin{cases} \frac{2}{\pi} \arctan \left[\frac{\mathfrak{S}_2(t_n)}{\|y(t_n)\|} \right], & \mathfrak{S}_1(t_n) \geq 0 \\ \frac{2}{\pi} \arctan \left[\frac{\|y(t_n)\|}{\mathfrak{S}_2(t_n)} \right], & \mathfrak{S}_1(t_n) < 0 \end{cases}$. The WTOD protocol is as follows:

$$\mathfrak{T}(t_n) = \arg \max_{1 \leq v \leq n_y} ((y_v(t_n) - \bar{y}_v(t_{n-1}))^\top \Gamma_v (y_v(t_n) - \bar{y}_v(t_{n-1}))).$$

A mode-switching DoS attack model is adopted and the measured signal $\tilde{y}(k)$ is as follows:

$\tilde{y}(k) = \mathcal{V}_{\delta(k)} \bar{y}(t_n)$, where $\mathcal{V}_{\delta(k)} = \text{diag}\{\mathcal{V}_{\delta(k)}^1, \dots, \mathcal{V}_{\delta(k)}^{n_y}\}$, and $\mathcal{V}_{\delta(k)}^1, \dots, \mathcal{V}_{\delta(k)}^{n_y} \in \{0, 1\}$.

An asynchronous FDF based on the observed mode is constructed as

$$\begin{cases} x_f(k+1) = \mathcal{K}_{\mathfrak{T}(t_n), \omega(k)} x_f(k) + \mathcal{L}_{\mathfrak{T}(t_n), \omega(k)} \tilde{y}(k) \\ r_f(k) = \mathcal{M}_{\mathfrak{T}(t_n), \omega(k)} x_f(k) \end{cases} \quad (2)$$

Define the evaluation functions as $\mathcal{Q}_{r, k_2} = \sqrt{\sum_{\mathfrak{k}=k_1}^{k_2} r_f^\top(\mathfrak{k}) r_f(\mathfrak{k})}$, $\mathcal{Q}_{th} = \sup_{\omega(k) \in \mathcal{L}_2, f(k)=0} \mathcal{Q}_{r, k_2}$. The decision rule is $\mathcal{Q}_{r, k_2} > \mathcal{Q}_{th} \Rightarrow \text{with fault}$, $\mathcal{Q}_{r, k_2} \leq \mathcal{Q}_{th} \Rightarrow \text{no fault}$.

Experimental Results: The theoretical results are presented as follows: For given positive parameters $\bar{\alpha}$, $\underline{\alpha}$, ϱ , φ_1 , φ_2 , γ and χ , system achieves MSES with $\mathcal{H}_\infty$ performance, if there exist matrices

$$\mathfrak{P}_1, \mathfrak{M}, \mathfrak{Q}_{\mathfrak{h}}, G_{\mathfrak{h}}, \bar{\mathcal{K}}_{\mathfrak{h}}, \bar{\mathcal{L}}_{\mathfrak{h}}, \bar{\mathcal{M}}_{\mathfrak{h}} \text{ and } \Pi_{1(1)} > 0, \text{ such that } \begin{bmatrix} \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}} & * & * \\ \mathfrak{U}_{1\mathfrak{h}}^\top & -\chi I & * \\ \chi \mathfrak{U}_{2\Pi_1(1)} & 0 & -\chi I \end{bmatrix} < 0, \text{ where}$$

$$\hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}} = \begin{bmatrix} \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{11} & * \\ \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{21} & \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{22} \end{bmatrix}, \quad \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{21} = \begin{bmatrix} \mathcal{G}_{1\mathfrak{I}(t_n)\mathfrak{h}} & 0 & 0 & 0 & 0 & \mathcal{G}_2 \\ \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{211} & 0 & \hat{\mathfrak{S}}_{\mathfrak{h}}^{212} & \hat{\mathfrak{S}}_{\mathfrak{h}}^{213} & \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{214} & \hat{\mathfrak{S}}_{\mathfrak{h}}^{215} \\ 0 & 0 & 0 & \hat{\mathfrak{S}}_{\mathfrak{h}}^{216} & 0 & 0 \end{bmatrix}, \quad \hat{\mathfrak{S}}_{\mathfrak{I}(t_n)\mathfrak{h}}^{22} =$$

$\text{diag} \{-I, \bar{\mathfrak{P}}_{\mathfrak{s}} - \mathbf{He}\{G_{\mathfrak{h}}\}, \bar{\mathfrak{P}}_{\mathfrak{s}} - \mathbf{He}\{G_{\mathfrak{h}}\}\}$. The FDF gains are derived as $\mathcal{L}_{\mathfrak{h}} = G_{3\mathfrak{h}}^{-1} \bar{\mathcal{L}}_{\mathfrak{h}}$, $\mathcal{M}_{\mathfrak{h}} = \bar{\mathcal{M}}_{\mathfrak{h}}$, $\mathcal{K}_{\mathfrak{h}} = G_{3\mathfrak{h}}^{-1} \bar{\mathcal{K}}_{\mathfrak{h}}$. Based on the theoretical results, the FDF gains are obtained through LMI simulations. The fault signal is selected as $f(k) = \begin{bmatrix} \mathfrak{f}^\top(k) \\ \mathfrak{f}^\top(k) \end{bmatrix}^\top$, $\mathfrak{f}(k) = \begin{cases} 1.5 \tanh(1.5k), 10 \leq k \leq 30 \\ 0, \text{otherwise} \end{cases}$. Fig. 1 depicts the system state trajectories, and the evaluation function is also obtained as $Q_{r,13} = 0.000010 < Q_{th} = 0.000100 < Q_{r,14} = 0.000163$. It is evident that under the proposed framework, the system achieves MSES and can effectively perform fault detection.

Conclusion: This article has investigated the FDF problem for MSMNNs subject to stochastic mode-switching channel DoS attacks and operating under an AET-WTOD communication protocol. A novel AET-WTOD scheme is developed to adjust transmission order and reduce communication load, while a mode-switching channel DoS attack model characterizes multi-channel disruptions. Sufficient conditions ensuring MSES and the prescribed $\mathcal{H}_\infty$ performance have been derived.

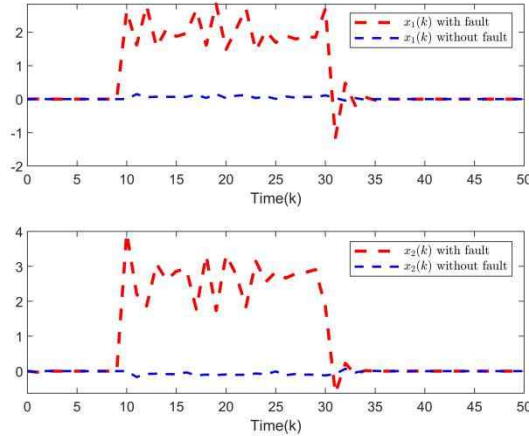


Fig. 1. State trajectories

References

1. Li, Dan-Dan, et al. "Quasi-projective synchronization of discrete-time fractional-order delayed memristive neural networks with uncertainties." *IEEE Transactions on Cybernetics* (2025).
2. Ru, Tingting, Chenxiao Cai, and Ju H. Park. "Secured bipartite consensus control for nonlinear multi-agent systems against hybrid attacks: A component-based WTOD protocol." *IEEE Transactions on Automation Science and Engineering* 22 (2025): 11680-11692.
3. Lin, Wen-Juan, et al. "Event-triggered fault detection filter design for discrete-time memristive neural networks with time delays." *IEEE Transactions on Cybernetics* 52.5 (2020): 3359-3369.

Error-Based Fuzzy Q-Learning Control for Trajectory Tracking under Initial Offsets

Myeongjun Seo¹⁾ and Sangmoon Lee^{2*)}

¹⁾Dept. of Water and IT Engineering, Kyungpook National University, Daegu, 41566, Republic of Korea

²⁾Dept. of Electric Engineering, Kyungpook National University, Daegu, 41566, Republic of Korea

*E-mail: cpsec.myeongjun@gmail.com

Abstract: Trajectory tracking is a fundamental requirement for autonomous mobile robots, particularly when large initial pose deviations cause a unicycle kinematic system to deviate significantly from a reference trajectory. This paper proposes an error-decomposed Fuzzy Q-Learning (FQL) controller for trajectory tracking and reference-trajectory re-entry of a unicycle kinematic system under large initial deviations. The controller consists of an offline-trained lateral module that maps the lookahead error state to angular-velocity commands and an online prior-initialized longitudinal module that maps the arc-length scheduling error to linear-velocity commands. Angular-velocity commands are smoothed using Softmax Action Blending and a slew-rate limiter to mitigate abrupt command variations, while a compact fuzzy rule-based representation preserves interpretability and reduces the online computational burden. The controller is evaluated under noise-free simulation conditions with progressively increasing initial offsets up to (6, 6) m. Simulation results indicate that, over the tested initial offsets, the proposed controller re-enters the reference trajectory with shorter convergence time than a Deep Q-Network (DQN) baseline using the same angular state-action setting, while reducing angular-velocity command variation and online computation cost. A real-world demonstration on a differential-drive robot further provides a qualitative proof-of-concept that the simulation-trained policy can be deployed on a physical platform without retraining.

Keywords : Fuzzy Q-Learning, Mobile Robot, Reinforcement Learning, Trajectory Tracking

Introduction: Trajectory tracking is essential for autonomous mobile robots [1], but under large initial pose deviations a unicycle system may fail to re-enter the reference path or generate fluctuating control. Conventional fuzzy controllers cannot learn from experience[2], and DQN is parameter-heavy[3]. This study proposes an error-decomposed Fuzzy Q-Learning[4] controller for reliable re-entry under large initial offsets.

Research Methods: The proposed controller adopts an error-decomposed closed-loop structure based on the unicycle kinematic model with a nonholonomic constraint. Lookahead-based error states — the distance error and heading error toward a forward point on the reference trajectory, and the longitudinal scheduling error — are computed in the robot body frame and normalized to $[-1, 1]$. Two separate FQL modules generate the control inputs: an offline-trained angular-velocity module maps the lookahead error to steering commands for trajectory re-entry, while an online linear-velocity module, warm-started from a P-controller-based prior, maps the scheduling error to forward-velocity commands for progress synchronization. Each input is fuzzified with five trapezoidal membership functions, and effective action values are obtained by firing-strength-weighted aggregation over compact fuzzy Q-tables. The discrete action candidates are converted into a continuous command through Softmax Action Blending and an angular slew-rate limiter to suppress abrupt steering, and the Q-tables are updated by temporal-difference learning with potential-based reward shaping.

Experimental Results: The proposed controller was evaluated on a parameterized ellipse reference trajectory under progressively increasing diagonal initial offsets from (1, 1) m to (6, 6) m, and compared with a DQN baseline using the same angular error inputs and the same discrete angular action set. At small offsets both controllers re-enter and track the path, but as the offset increases the proposed FQL approaches the reference trajectory more directly. As shown in Fig. 1, under the larger-offset conditions the proposed

FQL re-enters and tracks the reference trajectory more efficiently than the DQN baseline, reaching the reference path in less time. Table 1 summarizes the representative quantitative results: the proposed FQL achieves comparable tracking accuracy (cross-track RMSE of about 0.08 m) while reducing the convergence time, the angular-velocity command variation, and the per-cycle control computation time, using only 275 angular action values against 5,067 trainable parameters of the DQN. A real-world demonstration on a differential-drive robot further confirmed that the simulation-trained policy can be deployed on a physical platform without retraining.

Conclusion: This study proposed an error-decomposed Fuzzy Q-Learning controller that decouples angular re-entry control from linear progress synchronization. Using compact fuzzy Q-tables, it maintained comparable tracking accuracy while reducing convergence time, command variation, and computation compared with the DQN baseline at the tested offsets. A real-world demonstration confirmed deployment without retraining. A limitation is that re-entry from initial offsets beyond the tested range, i.e. larger than (6, 6) m, was not consistently reproducible across computational environments, which will be addressed in future work on the numerical robustness and generalization of the controller.

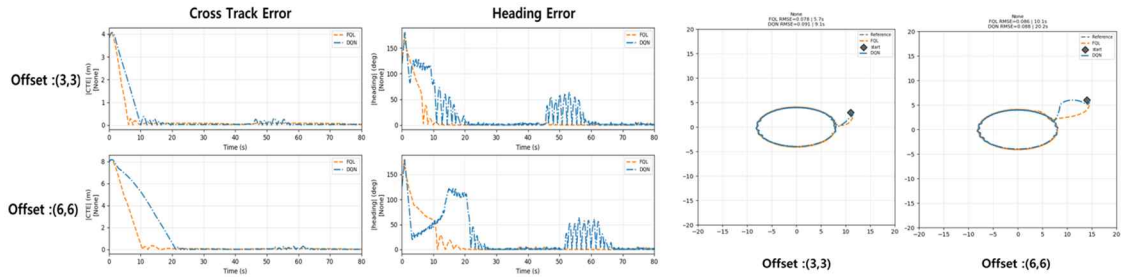


Fig. 1. Trajectory tracking of the proposed FQL and DQN under initial offsets.

TABLE 1. Representative Performance in Simulation (FQL / DQN)

Offset (m)	RMSE_CTE (m) †	t_c (s) ‡	Mean Δω (rad/s) §	T_ctrl (μs) ¶
(3, 3)	0.078 / 0.091	5.7 / 9.1	0.020 / 0.195	44.2 / 130.3
(6, 6)	0.087 / 0.088	10.1 / 20.2	0.019 / 0.203	44.4 / 128.7

† Root-mean-square of cross-track error.

‡ Convergence time to the reference path.

§ Mean absolute angular-velocity increment (control smoothness).

¶ Per-cycle control computation time.

TABLE 2 Structural Comparison Between the DQN Baseline and the Proposed FQL

Item	DQN baseline	Proposed FQL
Angular controller	DQN agent	FQL agent
Linear velocity	Fixed v_ref	Online FQL module
Function approx.	Neural net (64-64-11)	Fuzzy Q-table (25×11)
Trainable params (ω)	5,067	275
Replay buffer	50,000 transitions	None
Action generation	Greedy discrete action	Fuzzy aggregation + Softmax blending
Episodes (ω)	450	450

Acknowledgment

This work is financially supported by Korea Ministry of Environment(MOE) as 「Graduate School specialized in Integrated Water Resources Management.

References

1. Y. Kanayama, Y. Kimura, F. Miyazaki, and T. Noguchi, "A stable tracking control method for an autonomous mobile robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1990, pp. 384–389.
2. G. Antonelli, S. Chiaverini, and G. Fusco, "A fuzzy-logic-based approach for mobile robot path tracking," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 2, pp. 211–221, 2007.
3. V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
4. P.-Y. Glorennec and L. Jouffe, "Fuzzy Q-learning," in *Proc. 6th IEEE Int. Conf. Fuzzy Systems*, vol. 2, 1997, pp. 659–662.

Vital Sign Monitoring Using Camera and FMCW Radar Fusion in Home Environments

Sekyung Kim¹⁾, Jieun Park¹⁾, Dokyung Kang¹⁾, Kyungtaeg Yang¹⁾, Seungeon Song²⁾ and Jonghun Lee^{*1),2)}

¹⁾ Dept. of Interdisciplinary Engineering, Graduate school, DGIST, Daegu, 42988, Korea

²⁾ Institute of Research, DGIST, Daegu, 42988, Korea

E-mail: jhlee@dgist.ac.kr

Abstract: This paper proposes a camera and FMCW (Frequency Modulated Continuous Wave) radar fusion system for contactless vital sign monitoring in home environments. Accurate estimation of respiration and heart rate remains challenging in practical home environments because reflections from objects and body-motion interference can degrade target localization and radar signal quality. To address these challenges, an RGB camera is combined with an FMCW radar to identify the target person before vital signal extraction. A YOLO-based object detector detects the target in RGB images, and the bounding box is projected onto a range-angle map (RA map) to localize the corresponding radar target. After Singular Value Decomposition (SVD)-based clutter suppression, the phase signal is extracted from the localized target bin. A Kalman filter is applied to reduce motion-induced phase fluctuations and phase instability caused by target bin migration. The corrected phase signal is then decomposed using Variational Mode Decomposition (VMD) to estimate respiration rate (RR) and heart rate (HR). The proposed system is evaluated under various home-environment scenarios, including sitting, standing, walking, and remote-controller interaction. The HR reference is obtained from a smartwatch, while the RR reference is obtained through breath counting by the subjects. The proposed approach is expected to improve target localization and robust vital sign estimation in realistic home environments.

Keywords : FMCW radar, Vital Sign, Sensor Fusion, YOLO, SVD, VMD, KF

1. Introduction

Recent advances in non-contact healthcare sensing have increased interest in radar-based vital sign monitoring. An FMCW radar can measure respiration rate (RR) and heart rate (HR) without requiring wearable sensors. However, in home environments, the presence of objects and body-motion interference can degrade target localization accuracy and vital signal extraction quality. Camera sensors provide spatial information. Since target movement can cause bin migration in the radar range-angle map and introduce phase discontinuities, a Kalman filter is applied to compensate for motion-induced phase variations and improve RR and HR estimation stability.

2. Research Methods

This paper proposes a subject-specific vital sign monitoring framework that integrates an RGB camera with an FMCW radar. In the radar processing module, clutter is removed from the received I/Q signal via SVD, and an RA map is generated to obtain range and angle information. In the camera processing module, a YOLO-based object detector is employed to detect the subject's bounding box (b-box) from the RGB image.

The detected b-box is mapped onto the radar RA map to select the target bin corresponding to the subject. Beamforming is then applied to enhance the radar signal in the direction of the subject, and the phase is extracted from the complex signal of the selected target bin. To reduce phase instability caused by subject motion and frame-to-frame target bin migration, a Kalman filter is applied to the extracted phase signal.

The corrected phase signal undergoes phase unwrapping and band-pass filtering, after which it is decomposed into respiratory and cardiac components via VMD. Finally, the respiration rate and heart rate are estimated by analyzing the frequency characteristics of each vital sign component. To validate the proposed framework, experiments will be conducted under various home daily-life scenarios.

3. Experiments

Experiments were conducted using a TI 60 GHz FMCW radar (IWR6843ISK) and an RGB camera mounted at a height of 1.3 m. Three subjects participated in the experiments, and data were collected within a measurement range of 1 m to 3 m. The total duration of the experiments was approximately 40 min. Experimental scenarios include (1) sitting, (2) standing, (3) walking, (4) remote-controller interaction. Ground-truth heart rate (HR) was obtained using a smartwatch, while ground-truth respiration rate (RR) was measured by counting performed by the subjects.

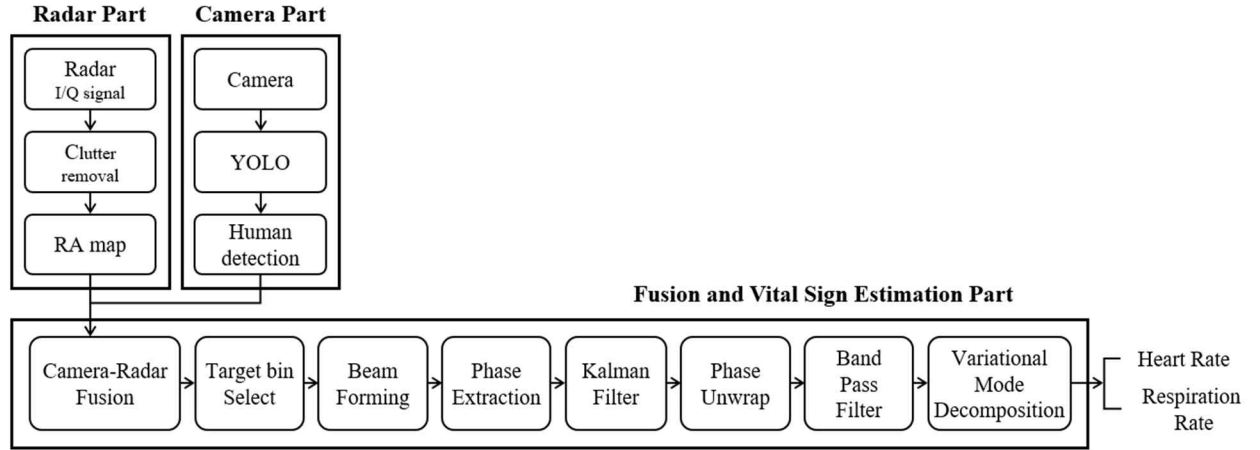


Fig. 1. Proposed pipeline

4. Conclusion

This paper combines RGB image-based person localization with a radar RA map to select the target bin corresponding to the subject of interest. The phase is then extracted from the selected radar signal, and a Kalman filter is applied to mitigate phase fluctuations arising from body motion and target bin migration. This approach aims to improve robustness against interference in home environments while maintaining a contactless sensing modality, thereby enhancing the stability of respiration rate and heart rate estimation. Future work will quantitatively evaluate the estimation accuracy of the proposed method for respiration rate and heart rate across various daily-living scenarios in home environments.

Acknowledgment

This research was supported by Private-Public Joint Technology Commercialization R&D Program(RS-2026-25527424) funded by the Ministry of SMEs and Startups(MSS, Korea).and by Activation of Future Automotive Advanced ECU Ecosystem Program (P269300001) funded by the Ministry of Trade, Industry and Energy (MOTIE)

References

1. Wang, Yingqi, et al. "Vital sign monitoring in dynamic environment via mmWave radar and camera fusion." *IEEE Transactions on Mobile Computing* 23.5 (2023): 4163-4180.
2. Yu, Huimin, Wenjun Huang, and Baoqiang Du. "SSA-VMD for UWB radar sensor vital sign extraction." *Sensors* 23.2 (2023): 756.
3. Qiu, Jiefan, et al. "mmWave Radar-based Personalized Multi-object Vital Signs Monitoring." *ACM Transactions on Multimedia Computing, Communications and Applications* 22.6 (2026): 1-25.
4. G. Jocher et al., "YOLOv8," Ultralytics, 2023. Available: <https://github.com/ultralytics/ultralytics>
5. Dragomiretskiy, Konstantin, and Dominique Zosso. "Variational mode decomposition." *IEEE transactions on signal processing* 62.3 (2013): 531-544.

Radar-Based Human-Robot Discrimination Using Transformer with Sequential Cluster Features in Automotive Manufacturing Environments

Dokyung Kang¹⁾, Jieun Park¹⁾, Kyungtaeg Yang¹⁾, Sekyung Kim¹⁾, Seungeon Song²⁾ and Jonghun Lee^{*1),2)}

¹⁾ Dept. of Interdisciplinary Engineering, Graduate school, DGIST, Daegu, 42988, Korea

²⁾ Institute of Research, DGIST, Daegu, 42988, Korea

*E-mail: jhlee@dgist.ac.kr

Abstract: This paper presents a radar-based pipeline for distinguishing industrial robots from humans in automotive manufacturing environments. Radar point cloud data is clustered using DBSCAN. For each cluster, a 3D bounding box is created and a 9-dimensional feature vector is extracted. These feature vectors are stacked over time to form a sequence. A Transformer model then uses self-attention to learn the periodic motion patterns of industrial robots. During inference, clusters that are not classified as robots are checked using camera data to determine whether a human is present. Since human data is limited, the system focuses on learning robot motion. Human detection is handled separately using a camera.

Keywords : radar point cloud, DBSCAN, Transformer, human-robot discrimination, industrial safety

1. Introduction

Worker safety in manufacturing environments is increasingly important when humans and industrial robots share the same workspace. Traditional safety systems use physical barriers or proximity sensors, but these have limited adaptability. Radar sensors are a promising alternative for human detection in industrial settings. Radar sensors operate reliably under harsh conditions such as dust, smoke, and poor lighting conditions. Unlike cameras, radar directly measures velocity using the Doppler effect and is independent of lighting changes. This paper proposes a pipeline that processes radar point cloud data to classify each detected cluster as a robot, a human, or another moving object. Clusters that are not identified as robots are further verified using camera data to confirm human presence.

2. Methods

Fig. 1 shows the overall pipeline, which consists of shared preprocessing, a training path, and an inference path. In the shared preprocessing stage, each radar frame is loaded from a JSON file. Each frame contains multiple points, and each point has three spatial coordinates (X, Y, Z) and a Doppler velocity value (V). DBSCAN clustering is then applied to the spatial coordinates to group nearby points into clusters. This clustering stage is modularized, so the algorithm can be replaced without changing the rest of the pipeline. For each cluster, a 3D bounding box is generated. In the training path, each cluster is automatically labeled using two criteria: its spatial position relative to predefined robot zones, and its Doppler velocity. Clusters located inside a robot zone with non-zero velocity are labeled as robot. This approach enables scalable annotation of approximately 23,000 frames without manual labeling. A 9-dimensional feature vector is then extracted from each labeled cluster, including the 3D centroid (center_x, center_y, center_z), bounding box size (size_x, size_y, size_z), point count, mean Doppler velocity, and velocity standard deviation. These feature vectors are stacked over 60 consecutive frames to form a 60×9 sequence. A Transformer model is trained on these sequences using self-attention to learn the periodic motion patterns of industrial robots. In the inference path, the same feature extraction method is applied in real time using a sliding window over the latest 60 frames. The trained Transformer model classifies each cluster as a robot or a moving object. Clusters not classified as robots are then passed to a decision-level camera fusion module. A pretrained human detection model processes the camera stream independently, and the two modalities are fused at the decision level: if both the radar and the camera indicate the presence of a moving object in the same spatial region, the cluster is classified as human. Otherwise, it is classified as other. This approach avoids retraining a human detection model, compensating for the sparse of human radar data in industrial

settings.

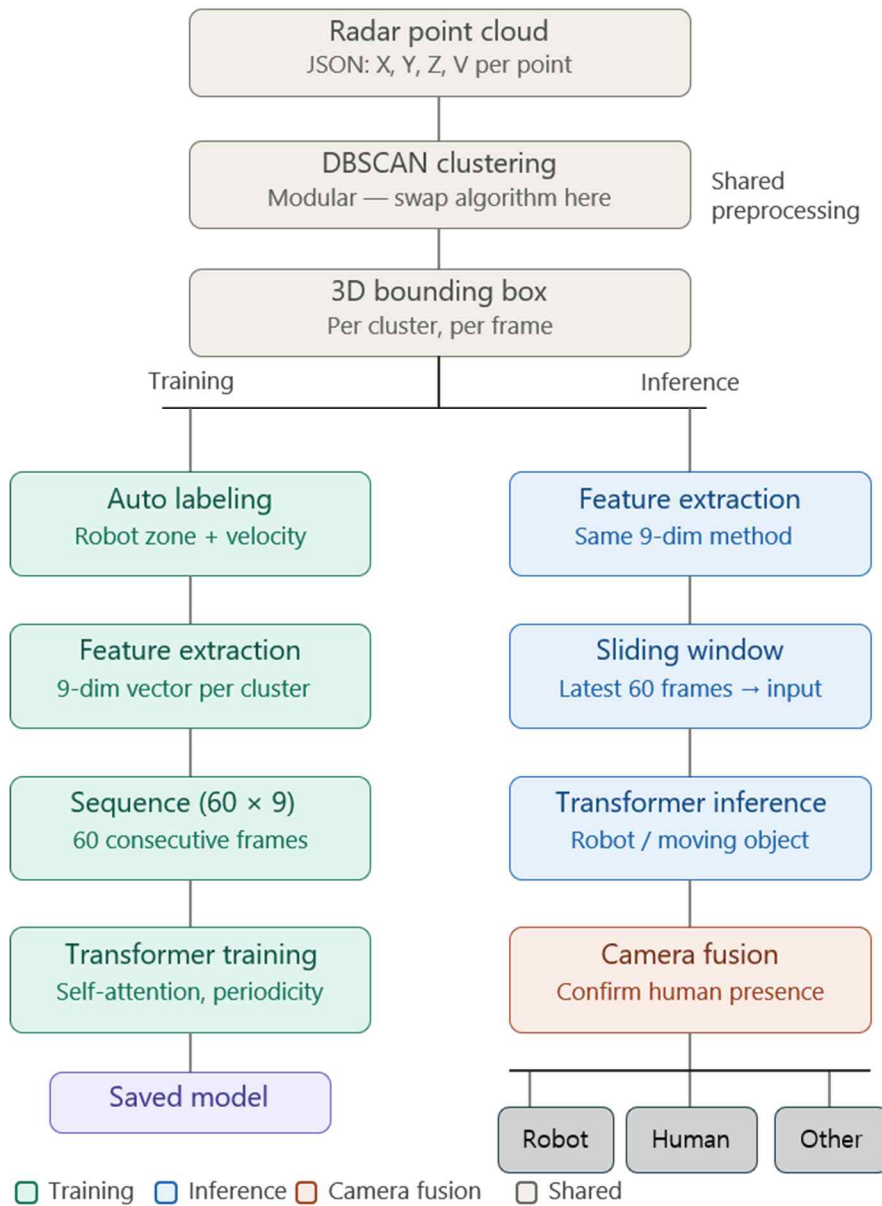


Fig. 1. Proposed training and inference pipeline

3. Experimental Results

The dataset consists of approximately 23,000 radar frames collected in an automotive manufacturing environment containing multiple robotic arms operating in fixed spatial zones. Automatic labeling assigns robot labels to clusters whose centroids fall within predefined 3D bounding regions and exhibit non-zero Doppler velocity, while clusters outside these zones with detectable motion are marked as moving objects and static clusters are excluded from training. Now, preliminary observations confirm that robot clusters exhibit highly consistent centroid trajectories and periodic velocity profiles across frames, supporting the feasibility of the Transformer-based periodicity detection approach.

4. Conclusion

This paper proposes a radar-based pipeline for distinguishing industrial robots from humans in automotive manufacturing environments. Unlike previous studies that detect only human presence, our proposed system classifies each detected cluster as a robot, a human, or another moving object in the same workspace.

A transformer model is trained on temporal feature sequences extracted from radar point cloud clusters, using self-attention to learn the periodic motion patterns of industrial robots. Clusters not classified as robots are then passed to a camera-based module for double verification. This approach reduces dependence on human training data, which is difficult to collect in large quantities in industrial settings. In addition, the modular clustering stage and automatic labeling strategy enable scalable annotation of large datasets without manual effort. Future work includes quantitative evaluation of classification accuracy and integration of real-time camera-based verification to validate end-to-end system performance.

Acknowledgment

This research was supported by the Technology Development Program(RS-2025-25450185) funded by the Ministry of SMEs and Startups(MSS, Korea) and by Private-Public Joint Technology Commercialization R&D Program funded by the Ministry of SMEs and Startups(MSS, Korea)

References

1. Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
2. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
3. Ester, Martin, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." *kdd*. Vol. 96. No. 34. 1996.
4. Lim, Sohee, Seongwook Lee, and Seong-Cheol Kim. "Clustering of detected targets using DBSCAN in automotive radar systems." *2018 19th international radar symposium (IRS)*. IEEE, 2018.
5. Ubezio, Barnaba, et al. "Radar based target tracking and classification for efficient robot speed control in fenceless environments." *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021.
6. Kruse, Nicolas C., Francesco Fioranelli, and Alexander Yarovoy. "Radar point cloud processing methods for human activity classification with point transformer networks." *IEEE Transactions on Radar Systems* 2 (2023): 1-12.
7. Zlatanski, Martin, et al. "Radar sensor for fenceless machine guarding and collaborative robotics." *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*. IEEE, 2018.
8. Huang, Xu, et al. "Indoor detection and tracking of people using mmwave sensor." *Journal of Sensors* 2021.1 (2021): 6657709.

Activity-Aware Sensor Data Compression for Energy-Efficient On-Device AI Computing in Human Activity Recognition

Ganapati Bhat ¹⁾ and Hyung Gyu Lee ²⁾

¹⁾ Sch. of EECS, Washington State Univ., Pullman, WA, 99164, U.S.A

²⁾ Sch. of Digital Software Engineering, Duksung Women's Univ. Seoul 01369, South Korea.

*E-mail: hglee@duksung.ac.kr

Abstract: Wireless data transmission in wearable human activity recognition accounts for up to 50% of total system energy consumption. We propose an activity-aware sensor data compression and approximation framework to enable energy-efficient on-device AI computing. Driven by real-time variance tracking, the proposed method achieves an average 14× compression ratio, reducing daily communication energy by 85–95% while maintaining high classification accuracy.

Keywords : Wearable Devices, Activity recognition, energy efficiency, sensor data compression

Introduction: Wearable devices equipped with multiple sensors have become increasingly important for health monitoring, patient rehabilitation, fitness tracking, and environmental sensing. Human activity recognition (HAR) is a key enabling technology for these applications, providing real-time activity information for personalized healthcare and intelligent services [1]. However, continuous transmission of multi-sensor data to a host device through wireless interfaces such as Bluetooth Low Energy (BLE) or Wi-Fi introduces significant communication overhead. Previous studies have shown that wireless data transmission accounts for up to 50% of the total system energy required for activity recognition [2], [3]. Such high communication costs not only shorten battery lifetime but also reduce the availability of HAR systems, potentially leading to missed detection of safety-critical events such as falls. To address this challenge, we propose a variance-driven activity-aware sensor data compression framework that enables energy-efficient on-device AI computing by minimizing wireless communication while preserving HAR classification accuracy.

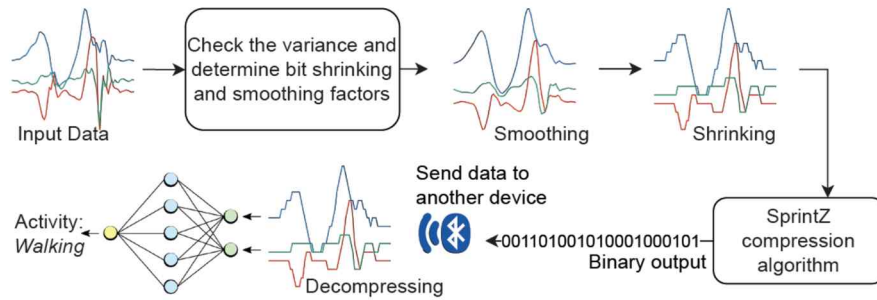


Fig. 1. Overview of the proposed approach

Variance-Driven Adaptive Compression Methodology: Fig. 1 illustrates the proposed activity-aware sensor data compression framework for energy-efficient on-device AI computing. The framework minimizes wireless communication through a three-stage adaptive pipeline consisting of moving-average smoothing, runtime bit-width reduction, and lightweight lossless compression. At the beginning of each activity window, the signal variance is estimated to classify the user's motion as static (e.g., sitting, lying), moderate (e.g., walking), or dynamic (e.g., jogging). Based on the estimated activity level, the framework dynamically selects the smoothing factor (N) and bit-shrinking factor (α) to balance compression efficiency and signal fidelity, as expressed in (1) and (2).

$$Y[t] = \frac{1}{N} \sum_{i=0}^{N-1} X[t-i] \quad (1),$$

$$Y[t] = X[t] \gg \alpha \quad (2)$$

where $X[t]$ denotes the sensor sample at time t . Static activities with low signal variance are compressed more aggressively using larger smoothing and bit-shrinking factors, whereas dynamic activities employ milder compression to preserve important signal characteristics. The compressed data are then encoded using the lightweight SprintZ lossless compression algorithm [4] before wireless transmission. At the receiver, the compressed stream is efficiently reconstructed, while the neural network classifier is trained with quantization-aware data augmentation to maintain recognition accuracy despite reduced signal precision.

Experimental Results: The proposed framework was evaluated using three public HAR datasets, w-HAR, Shoaib, and WISDM [5]–[7], with a lightweight neural network classifier consisting of a single hidden layer with eight neurons. Fig. 2 summarizes the compression performance and recognition accuracy. The proposed method achieves an average $14\times$ compression ratio, reaching over $20\times$ for low-variance static activities while maintaining $3.7\text{--}9.6\times$ compression for dynamic activities. Despite the substantial data reduction, the recognition accuracy remains comparable to the baseline, with an accuracy loss of less than 3% across all datasets.

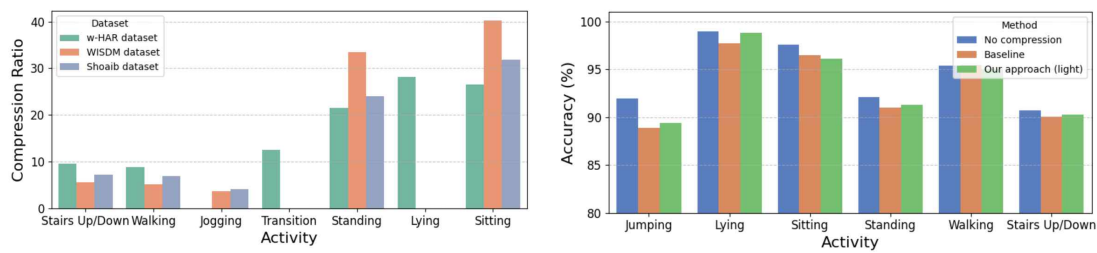


Fig. 2. Comparison of compression ratio (a) and recognition accuracy (b) for different activities

The computational overhead was validated using a prototype implementation on a Texas Instruments CC2650 SensorTag. The complete compression pipeline—including adaptive smoothing, bit-width reduction, and SprintZ compression—requires only 46 ms and 47 μ J, accounting for less than 2% of the energy required for a single HAR inference cycle. Furthermore, evaluation using the ATUS and ARAS daily activity datasets demonstrates 86–93% communication energy savings, effectively extending the operational lifetime of the wearable device.

Conclusion: This paper presented an activity-aware sensor data compression and approximation framework for energy-efficient on-device AI computing in wearable human activity recognition. The proposed variance-driven framework dynamically adapts compression parameters to signal characteristics, enabling substantial reductions in wireless communication without compromising recognition performance. Experimental results demonstrate an average $14\times$ compression ratio and 85–95% communication energy savings while maintaining classification accuracy within 3% of the uncompressed baseline. The proposed framework provides an efficient communication solution for next-generation low-power wearable AI systems.

References

1. A. J. Espay et al., “Technology in Parkinson’s Disease: Challenges and Opportunities,” *Movt. Disorders*, vol. 31, no. 9, pp. 1272–1282, 2016.
2. G. Bhat et al., “REAP: Runtime Energy-Accuracy Optimization for Energy Harvesting IoT Devices,” in *Proc. of Annual Design Automation. Conf.*, pp. 171:1–171:6, 2019.
3. M. M. Sandhu et al., “SolAR: Energy Positive Human Activity Recognition using Solar Cells,” in *Int. Conf. on Pervasive Comput. And Commun.*, pp. 1–10, 2021.
4. D. Blalock, S. Madden, and J. Gutttag, “SprintZ: Time Series Compression for the Internet of Things,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, pp. 93:1–93:23, Sept. 2018.
5. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.
6. M. Shoaib et al., “Fusion of Smartphone Motion Sensors for Physical Activity Recognition,” *Sensors*, vol. 14, no. 6, pp. 10146–10176, 2014.
7. G. Bhat et al., “w-HAR: An Activity Recognition Dataset and Framework Using Low-Power

Wearable Devices,” *Sensors*, vol. 20, p. 5356, 2020.

Comparative Analysis of UAV Position Prediction Models

Sang Suh ^{1*)}, Doweon Kim²⁾, Jie Yu²⁾

^{1,2,3)} Dept. of Computer Science, East Texas A&M University, Commerce, TX, U.S.A.

*E-mail: Sang.Suh@etamu.edu

Abstract: This study presents a systematic comparative evaluation of six prediction models, Linear Regression, Support Vector Regression, Random Forest, LSTM, LSTM Autoencoder, and Transformer for UAV horizontal position forecasting using time-series GPS telemetry data. Sequential training samples were generated via a sliding window mechanism ($n = 5$) applied to real UAV flight trajectory data collected along square-shaped flight paths of varying sizes. Models were evaluated under identical preprocessing, sequence generation, and data split conditions using Mean Squared Error (MSE) on a held-out test trajectory. Results indicate that Linear Regression achieved the best overall performance (Joint MSE = 9.41), followed by the Transformer (Joint MSE = 11.78), while SVR performed worst (Joint MSE = 23.48). Despite their architectural sophistication, most deep learning models did not outperform the linear baseline under the constraints of the current small-scale dataset, with the Transformer as the notable exception. These findings highlight the critical role of dataset scale and diversity for deep temporal models, and provide practical benchmarking insights for future UAV navigation research.

Keywords: UAV trajectory prediction, LSTM, Transformer, time-series regression, GPS telemetry, machine learning

Introduction: Accurate position estimation is critical for safe and reliable UAV operations. Raw GPS data are vulnerable to signal distortion from multipath propagation, adverse weather, and deliberate spoofing attacks [1][2][3], creating discrepancies between true and reported UAV positions. While prior research has focused mainly on anomaly detection [2], this work addresses the more fundamental challenge of capturing temporal dynamics of UAV motion. We conduct a systematic benchmark of classical machine learning and deep learning models for UAV position forecasting under identical experimental conditions.

Research Methods: UAV position prediction is formulated as a time-series regression task. Given a sliding window of $n = 5$ consecutive time steps, each containing eight features (ground speed, northward/eastward velocity, heading angle, positional displacement components, Euclidean displacement, and time gap), models predict the next-step displacements in latitude and longitude (Δ_{lat} , Δ_{lon}). Six models are compared: Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF) [4], LSTM [5], LSTM Autoencoder (LSTM-AE) [6], and a Transformer encoder [7][8]. Classical models operate on flattened 40-dimensional input vectors, while deep learning models process the full 5×8 sequence tensor. The dataset comprises GPS telemetry logs from real UAV flights along four square-shaped paths (20 m, 35 m, 50 m, 70 m sides); data from the 20 m, 50 m, and 70 m paths are used for training, and the unseen 35 m path is held out for testing. All features are normalized using MinMaxScaler. Deep learning models are trained with the Adam optimizer ($\text{lr} = 0.001$), MSE loss, batch size 64, and early stopping.

Experimental Results: Table 1 reports the test-set MSE for all six models. Linear Regression achieves the lowest joint MSE (9.41), followed by the Transformer (11.78) and Random Forest (13.27). Notably, most deep learning models LSTM (16.37) and LSTM-AE (21.14) underperform the classical baselines. SVR records the highest joint error (23.48). A key finding is that MSE_{lon} consistently exceeds MSE_{lat} across all models, reflecting greater directional

variability along the east-west axis due to UAV turning behaviour on the square paths. The Transformer exhibits the most balanced error across both dimensions ($MSE_{lat} = 4.98$, $MSE_{lon} = 6.80$) [7]. Figure 1 shows the true versus predicted displacement sequences for the best-performing (LR) and worst-performing (SVR) models; LR closely tracks ground-truth magnitudes while SVR predictions collapse toward zero. The strong LR performance is consistent with the limited scale and regularity of the training dataset, which constrains the scope for deep models to learn generalizable temporal representations [4][5].

Table 1. Test-set MSE for all models (unit²; 1 unit $\approx$ 1.1 cm).

Model	MSE_lat	MSE_lon	Joint MSE	Joint RMSE
Linear Regression	4.32	5.09	9.41	3.07
Transformer	4.97	6.80	11.78	3.43
Random Forest	6.26	7.01	13.27	3.64
LSTM	7.65	8.72	16.37	4.05
LSTM Autoencoder	8.87	12.26	21.14	4.60
SVR	11.27	12.21	23.48	4.85

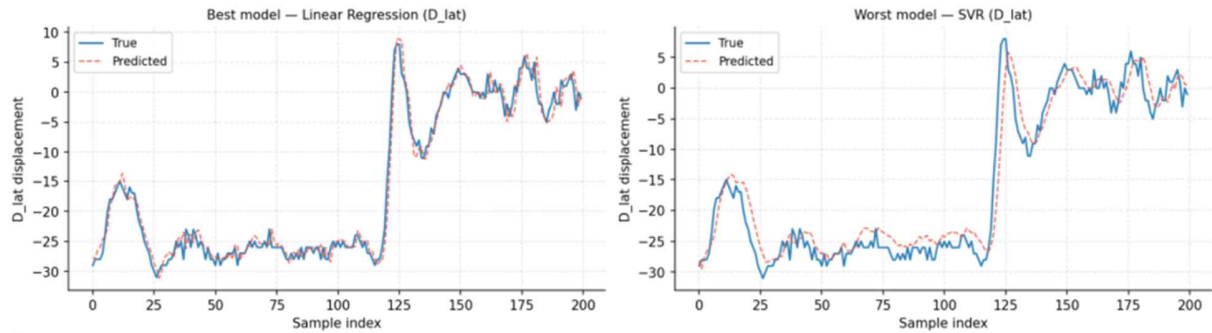


Fig. 1. True vs. predicted latitude displacement sequences for LR (best) and SVR (worst)

Conclusion: This study conducted a systematic comparison of six regression models for UAV horizontal position displacement prediction using time-series GPS telemetry [1][2]. Linear Regression achieved the best performance (Joint MSE = 9.41) under the constraints of the small, geometrically regular dataset, while the Transformer placed second (Joint MSE = 11.78), outperforming all recurrent and ensemble models [7][8]. SVR performed worst (Joint MSE = 23.48), and the LSTM Autoencoder's unsupervised pretraining did not provide measurable benefit over the standard LSTM [6]. These results suggest that classical models offer a strong competitive baseline for small-scale UAV datasets, while Transformer-based architectures show promise for scaling to larger and more diverse trajectory collections. Future work should expand the training corpus, incorporate multi-step prediction horizons, integrate additional sensor modalities (e.g., IMU), and perform systematic hyperparameter tuning, particularly for SVR and recurrent architectures, to better characterize the performance ceiling of each model family [4][5].

Acknowledgment

We thank our colleagues from East Texas A&M University who provided insight and expertise that greatly assisted the research, especially the AI research team of the Department of Computer Science.

References

1. C. Kim et al., "Reliable Detection of Location Spoofing and Variation Attacks," IEEE Access, vol. 11, pp. 10813-10825, 2023.
2. M. Shin et al., "Trajectory-Driven Deep Learning for UAV Location Integrity Checks," IEEE Access, vol. 12, pp. 178789-178804, 2024.
3. M. Meng et al., "An Approach of Linear Regression-Based UAV GPS Spoofing Detection," Wireless Commun. Mobile Comput., vol. 2021, 2021.
4. S. Hashemi et al., "Robust Trajectory Prediction Using Random Forest Methodology," Aerospace, vol. 11, no. 1, 2024.
5. Y. Zhang et al., "Recurrent LSTM-Based UAV Trajectory Prediction with ADS-B Information," IEEE GLOBECOM, 2022.
6. R. Kant et al., "LSTM Auto-Encoder-Based Position Prediction Model for Fixed-Wing UAV," IEEE Trans. Artif. Intell., vol. 4, pp. 173-181, 2023.
7. Q. Tong et al., "Long-Term Trajectory Prediction Model Based on Transformer," IEEE Access, vol. 11, pp. 143695-143703, 2023.
8. Y. Liu et al., "Attention-Aware Social Graph Transformer Networks for Stochastic Trajectory Prediction," IEEE Trans. Knowl. Data Eng., vol. 36, pp. 5633-5646, 2024.

Unified Cross-Layer Embedded AI Design

Sang Suh^{1*}, Omar Aly²⁾

^{1,2)} Department of Computer Science, East Texas A&M University, Commerce, Texas, U.S.A.

**E-mail: Sang.Suh@etamu.edu*

Abstract: In this paper, we implemented a framework that proposes a unified cross-layer Embedded AI design that considers application requirements, system constraints, and algorithm choices, benefiting resource-constrained conditions and enhancing efficiency, scalability, and performance. We then conducted an experiment using three datasets with varying levels of complexity to evaluate how data characteristics influence model behavior and system constraints, supporting our claim that the system should be considered as a whole rather than a single aspect.

Keywords: Embedded Artificial Intelligence, Machine Learning, Cross-Layer Framework, UCLEAD, Resource-Constrained Devices, Convolutional Neural Networks, Optimization, Quantization, Adversarial Robustness.

Introduction: Embedded Artificial Intelligence (Embedded AI) enables systems to learn from data, adapt, and make decisions while conserving computational resources [1],[2]. They combine Machine Learning algorithms with Embedded Systems platforms like Microcontrollers and SoC devices [2],[3]. Thanks to recent advances in lightweight neural networks and quantization, executing complex tasks on resource-constrained devices is now more feasible [12]. However, the remaining challenge is how to move from individual, targeted designs to a unified system that integrates hardware constraints, algorithmic design, security, and application requirements into a single, scalable framework. To address this challenge, we analyzed recent advances in embedded machine learning [12] and specialized hardware architectures [9],[10],[15], along with real-world deployments that could benefit from this work [11],[14],[17],[18]. In this paper, we present a conceptual cross-layer framework that demonstrates its benefits relative to targeted designs, discuss integration principles, and break down the steps required for a unified Embedded AI system that is scalable, secure, and aligned with application requirements. We will link technological foundations with system-level challenges [19].

Research Methods: We structured this study into different phases that build on each other in order to find the gap in a single system design and provide a unified system.

1- Identifying the Literature: here we reviewed related journal articles, conference proceedings, and review papers with respect to; Embedded Machine Learning and TinyML [12],[20], Hardware Architectures for Embedded Systems (GPUs, FPGAs, Accelerators) [9],[10],[15], Embedded AI systems designed for specific applications (Healthcare, Robotics, Transportation, Smart Infrastructure) [6],[7],[14],[17]. It was essential in order to understand the current state of Embedded AI research and then organize these studies and map them according to primary optimization objective, resource constraints addressed, and level of system integration considered.

2. Gap analysis and constraint model: we compared studies to discover gaps or areas that have been under addressed we identified the following; optimization applied to one application are often not transferrable to other applications, different evaluation metrics are employed for each application, security mechanisms are typically treated separately from their effects on energy and latency constraints [8],[13], and finally no unified method to design systems at the life cycle stage.

3.Framework development: We proposed a unified cross-layer embedded AI design framework (UCLEAD) which consists of five interconnected design aspects: a) Application requirements definition: real-time constraints, level of data sensitivity, and scalability [11],[17]. b) Constraints Identification and Description: power and energy consumption, memory and storage constraints, thermal and network constraints [12],[19].C) Choosing the Model and Algorithm: we compared models to see which is a better fit for different computational levels, quantization, and ability for AI adaptive learning [12],[16]. d) Hardware architecture compatibility: selection of accelerators, memory

architecture, and data flow optimization. [9],[10],[15] e) Security: software and hardware protection, fault detection strategies, and encryption methods [8],[13].

4-Validating the framework using case mapping: the following domains were mapped to the framework dimensions; healthcare systems [6],[11], embedded control systems in robotics [5], Smart infrastructure platforms [19], EV battery management [4],[17].

5- Evaluation metrics: We used the following evaluation metrics: Latency, Energy usage, Memory utilization, Model accuracy, Security resilience, Scalability potential. Rather than focusing on a single metric, the methodology supports a system design with different objectives [1],[12].

Experimental Results: During implementation, we addressed the following research questions:

- 1) What is the overall performance of quantization on a lightweight CNN?
- 2) What is the effect of quantization on model size, latency, and accuracy in case of limited resources?
- 3) Does quantization preserve predictive quality while reducing memory and energy constraints?
- 4) Can the quantized model serve as a practical prototype for real-world deployment?

We started by choosing three datasets of increasing complexity to evaluate the framework: MNIST, FashionMNIST, CIFAR-10. The use of these datasets enables systematic evaluation of how data characteristics influence model behavior and system constraints, which is central to the UCLEAD design approach. We then resized Grayscale sets to RGB dimensions and normalized them to the range [0,1]. We performed the experiments in Python 3.11 using TensorFlow 2 and tf_keras on a CPU environment to simulate device constraints. We used the following training configurations: Optimizer: Adam, Loss: Categorical CrossEntropy, 6 Epochs. Batch size: 64, Framework: TensorFlow 2. Quantization-aware training was applied with the optimization flag: tf.lite.Optimize.DEFAULT. We used the TensorFlow Lite interpreter for quantized models for inference and timed each sample for latency estimation. To test security resilience, we used the Adversarial Robustness Toolbox using the Fast Gradient Method (FGSM) with an epsilon of 0.03. We used the following evaluation metrics to reflect constraints: accuracy to reflect the algorithmic layer, latency and flash size for hardware execution, and security robustness against attacks for deployment constraints. We compared three variants of the same architecture: FP32 baseline, post-training quantization model, and quantization-aware training model. We evaluated each model on all three datasets using the following identical CNN architecture: Conv2D(32), MaxPooling2D, Conv2D(64), MaxPooling2D, Flatten, Dense(64, ReLU) followed by Dense(10). The results showed that the performance of Embedded AI is shaped by interactions across data, model representation, and deployment constraints rather than a single design dimension. Table 1 summarizes the evaluation results.

Table 1. Cross-Layer Performance Comparison Across All Datasets Before and After Quantization (Accuracy, Flash, RAM, Latency, Security Robustness)

Dataset	Variant	Acc (%)	Flash (KB)	RAM (KB)	Lat (ms)	Sec (%)
MNIST	FP32	99.0	3349.7	256.0	2.39	95.0
	INT8 PTQ	99.0	283.6	64.0	0.11	95.0
	INT8 QAT	98.5	282.7	64.0	0.48	97.0
F-MNIST	FP32	86.5	3349.7	256.0	0.82	74.0
	INT8 PTQ	86.5	283.6	64.0	0.11	74.5
	INT8 QAT	85.0	282.7	64.0	0.07	83.0
CIFAR-10	FP32	71.5	3349.7	256.0	1.85	42.0

	INT8 PTQ	72.5	283.6	64.0	0.23	44.0
	INT8 QAT	71.0	282.7	64.0	0.18	64.0

The results showed that accuracy strongly depends on the dataset. Performance decreases from 99.0% in the case of MNIST to 71.5% in the case of CIFAR-10. Quantization preserves the performance; across all datasets, the accuracy differences between FP32 and quantized models are negligible. Model compression is consistent and significant; quantization reduces the model size from 3349.7 KB to 282.7 KB across all datasets, independent of the dataset, which confirms that memory optimization is primarily driven by numerical representation rather than input data characteristics. Adversarial security resilience under QAT significantly improves, as shown on the CIFAR-10 set, where robustness increases from 42.0% to 64.0%. From our unified system's perspective, the results demonstrate that algorithmic changes affect memory, hardware runtime conditions, and influence latency, which depends on the interaction between the model, runtime optimization, and execution environment.

In Figure 1, we can see comparison plots for each dataset that provide confirmation of the framework's benefits, such as preserving high accuracy, achieving memory savings through different hardware limits, and improving security robustness through quantization-aware training. proving that when designing the system, all aspects should be taken into consideration, not a single quantity.

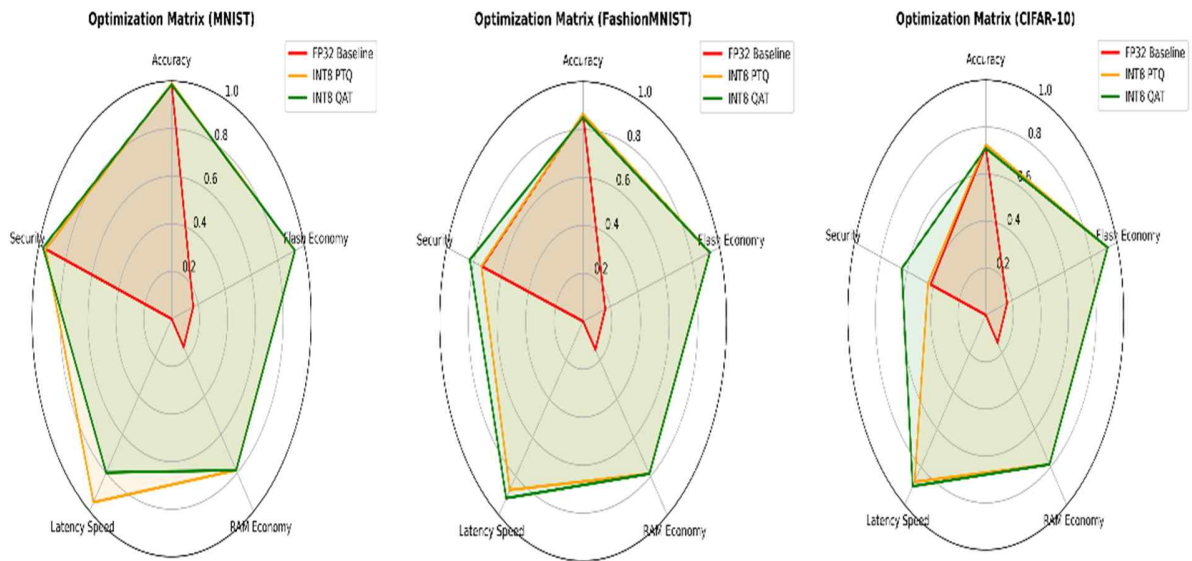


Fig. 1. Optimization Matrix Visualizations for MNIST, FashionMNIST, CIFAR-10 Datasets

Conclusion

In this paper, we presented the UCLEAD framework and evaluated it using a lightweight convolutional neural network across three different datasets. We analyzed the effects of model representation on accuracy, latency, memory, and adversarial security; we showed how the choice of dataset has a significant effect on accuracy and how quantization preserves performance across all datasets. We showed improvement in model size reduction, memory efficiency, latency, and security robustness with potential improvement on real embedded hardware. Last but not least, the findings confirm that improvement of one aspect in the design does not necessarily negatively impact the other, and choosing a unified approach when designing Embedded AI systems can balance accuracy, efficiency, and deployment constraints across multiple layers.

References

1. Z. Zhang and J. Li, "A Review of Artificial Intelligence in Embedded Systems," *Micromachines*, vol. 14, (5), 2023.
2. X. Huang et al, "Embedded Artificial Intelligence: A Comprehensive Literature Review," *Electronics*, vol. 14, (17), pp. 3468, 2025.
3. J. J. Lu and H. Zhou, "Implementation of Artificial Intelligence Algorithm In Embedded System," *Journal of Physics: Conference Series*, vol. 1757, (1), 2021.
4. K. Keshyagol and S. Kulkarni, "Artificial intelligence embedded platform for battery management and autonomous navigation in a prototype electric vehicle," *Discover Sustainability*, vol. 6, (1), pp. 1474, 2025.
5. S. Kumar et al, "Multiple trajectory optimization and control of robotic agents using hybrid fuzzy embedded artificial intelligence technique for multi target problems," *Scientific Reports (Nature Publisher Group)*, vol. 16, (1), pp. 2945, 2026.
6. L. Rodríguez-Cobo et al, "Non-Contact Thermal and Acoustic Sensors with Embedded Artificial Intelligence for Point-of-Care Diagnostics," *Sensors*, vol. 24, (1), pp. 129, 2024.
7. D. Klinkhammer, "Explainable Artificial Intelligence with MicroPython: Lightweight Neural Networks for Students' Deeper Learning," *International Journal of Emerging Technologies in Learning (Online)*, vol. 20, (4), pp. 72-80, 2025.
8. Z. Dong et al, "EG-STC: An Efficient Secure Two-Party Computation Scheme Based on Embedded GPU for Artificial Intelligence Systems," *Computers, Materials, & Continua*, vol. 79, (3), pp. 4021-4044, 2024.
9. L. Dias et al, "Photonic-digital hybrid artificial intelligence hardware architectures: at the interface of the real and virtual worlds," *JPhys Photonics*, vol. 7, (4), pp. 045014, 2025.
10. C. Ra et al, "Investigation on Artificial Intelligence Hardware Architecture Design Based on Logic-in-Memory Ferroelectric Fin Field-Effect Transistor at Sub-3nm Technology Nodes," *Advanced Intelligent Systems*, vol. 7, (2), 2025.
11. P. Ilyas, "Affordable Audio Hardware and Artificial Intelligence Can Transform the Dementia Care Pipeline," *Algorithms*, vol. 18, (12), pp. 787, 2025.
12. A. Biglari and W. Tang, "A Review of Embedded Machine Learning Based on Hardware, Application, and Sensing Scheme," *Sensors*, vol. 23, (4), pp. 2131, 2023.
13. N. A. Kose et al, "Detection of Malicious Threats Exploiting Clock-Gating Hardware Using Machine Learning," *Sensors*, vol. 24, (3), pp. 983, 2024.
14. M. Beltrán-Escobar et al, "A Review on Resource-Constrained Embedded Vision Systems-Based Tiny Machine Learning for Robotic Applications," *Algorithms*, vol. 17, (11), pp. 476, 2024.
15. A. N. R. Lopez, "Performance of Neural Networks on FPGA for Embedded Devices using Hls4ml Framework." Order No. 32148618, California Polytechnic State University, United States -- California, US, 2025.
16. H. Tenhunen et al, "Parallelized Online Regularized Least-Squares for Adaptive Embedded Systems," *International Journal of Embedded and Real-Time Communication Systems*, vol. 3, (2), pp. 73-91, 2012.
17. C. Lin et al, "Early Driver Fatigue Detection System: A Cost-Effective and Wearable Approach Utilizing Embedded Machine Learning," *Vehicles*, vol. 7, (1), pp. 3, 2025.
18. I. González-Afanador et al, "Real-time invasive sea lamprey detection using machine learning classifier models on embedded systems," *Neural Computing & Applications*, vol. 36, (26), pp. 16195-16212, 2024.
19. T. Michail et al, "A Review of Embedded Software Architectures for Multi-Sensor Wearable Devices: Sensor Fusion Techniques and Future Research Directions," *Electronics*, vol. 15, (2), pp. 295, 2026.
20. T. C. Wen et al, "TinyML-Based Swine Vocalization Pattern Recognition for Enhancing Animal Welfare in Embedded Systems," *Inventions*, vol. 10, (4), pp. 52, 2025.

Miniaturized Broadband Microstrip Patch Antenna Utilizing Open-Circuit Stubs and Ground Slotting for 5G Systems

Zelin Hou¹⁾, Jie Zhang¹⁾, Ka Lok Man¹⁾, Yan Li²⁾ and Jingchen Wang^{1*)}

¹⁾ School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China

²⁾ College of Software and Coverage Computer Engineering, Inha University, Incheon, Republic of Korea

* E-mail: jingchen.wang@xjtlu.edu.cn

Abstract: This work presents a compact broadband stub-loaded microstrip patch antenna tailored for millimeter-wave (mmWave) 5G communications. Simulation results reveal that the proposed design exhibits a continuous -10 dB impedance bandwidth spanning from 24.132 GHz to 28.862 GHz. Furthermore, simulated far-field radiation patterns verify stable and usable directional radiation characteristics across the operating frequencies.

Keywords : 5G antenna, mmWave antenna, microstrip patch antenna, open stubs, ground slot

Introduction

Compact broadband antennas are essential components for millimeter-wave (mmWave) 5G systems, as high-data-rate wireless links strictly demand stable impedance matching, predictable radiation behavior, and a minimized integration footprint [1-3]. Recent academic efforts [4-6] have investigated various techniques for bandwidth expansion and multi-resonance enhancement, including stub-loaded patches, tri-band (24/28/39 GHz) configurations, defected ground structures (DGS), and magneto-electric dipoles. Previous literature demonstrates that broadband performance is typically achieved by systematically manipulating the surface current paths, input impedance, and return-current distributions. Building upon these methodologies, this paper presents a structural optimization and comprehensive simulation analysis of a highly compact, broad-band microstrip antenna featuring a dual-resonant characteristic combined into a single operating band.

Antenna Design

The detailed geometry and specific parameters of the proposed antenna are illustrated in Fig. 1. The design is layout-optimized on an H348 substrate with a relative permittivity $\epsilon_r = 3.48$ and a thickness $h = 0.76$ mm. The minimized structure makes it exceptionally suitable for highly integrated mmWave front-ends. As depicted in Fig. 1, the top layer consists of an inset-fed rectangular patch coupled with two pairs of open-circuit stubs symmetrically attached to the microstrip feedline. The length of the open stubs (L_S) is utilized to adjust the effective electrical path, thereby shifting the higher-order resonant modes into the desired frequency band. Concurrently, the stub width (W_S) is tuned to improve the local impedance matching depth. To further consolidate the individual resonances into a single continuous broadband response, a rectangular ground slot (DGS) is etched on the bottom ground plane. This perturbs the return-current path, introducing an auxiliary resonance that bridges the gap between the target operating frequencies.

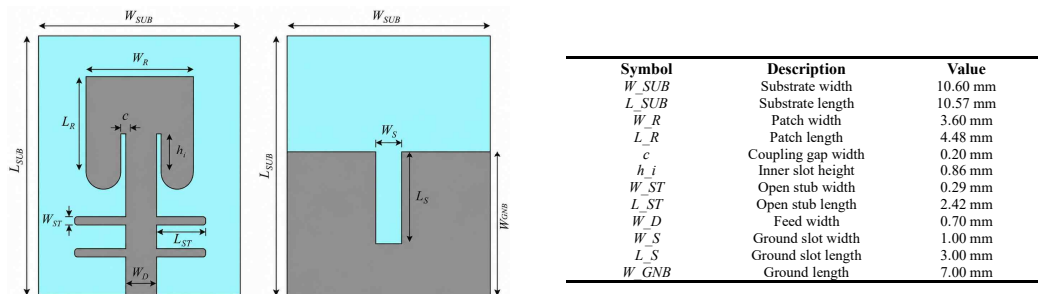


Fig. 1. Letter-labelled geometry of the proposed 25/28 GHz antenna and the corresponding physical dimensions.

Simulated performance of the proposed antenna

As shown in Fig. 2(a), the proposed antenna yields a continuous bandwidth from 24.132 GHz to 28.862 GHz (reflection coefficient $S_{11} < -10\text{dB}$), successfully encompassing the 25 GHz and 28 GHz operational frequencies. At 25 GHz, the antenna exhibits a main-lobe magnitude of 2.50 dBi, a main beam direction of 325° , a 3-dB angular beamwidth of 47.5° , and a side-lobe level (SLL) of -0.5 dB . At 28 GHz, the simulated main-lobe magnitude is 2.38 dBi, oriented at a beam direction of 14.0° , with a broader 3-dB beamwidth of 114.9° and an SLL of -0.5 dB . These far-field metrics reinforce that the proposed design guarantees both a wideband impedance profile and stable, directional radiation performance.

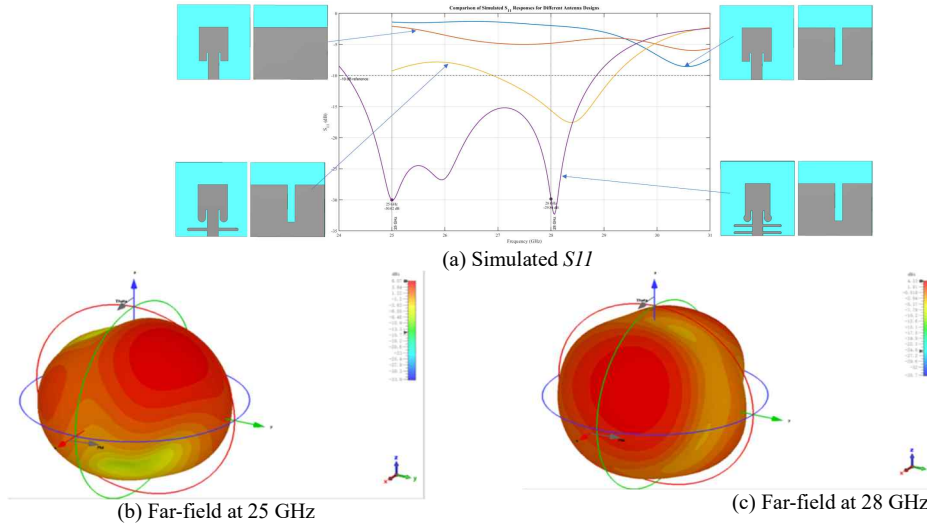


Fig. 2. Performance results of the optimized broadband antenna: (a) simulated S_{11} , (b) far-field radiation pattern 25 GHz, and (c) far-field radiation pattern at 28 GHz.

Conclusion:

A compact, broadband stub-loaded patch antenna for mmWave 5G applications has been successfully designed and analyzed using full-wave EM simulations. The incorporated open-circuit stubs provide effective control over the electrical resonant paths, while the defected ground slot successfully bridges the dual resonances into a continuous, wide impedance bandwidth.

Acknowledgment

This work was partially supported by the National Research Foundation of Korea (NRF) grant, funded by the Korean government (no. RS-2026-25492865) and the XJTLU AI University Research Centre, Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation at XJTLU and SIP AI innovation platform (YZCXPT2022103).

References

1. S. Agarwal, U. Rafique and V. Jain, "Stub-loaded dual-band planar antenna for 28/38 GHz 5G applications," in *Proc. 7th Int. Conf. Signal Processing and Communication (ICSC)*, Noida, India, pp. 124-128, (2021).
2. A. Gupta, S. Bansal and R. Manchanda, "Tri-band antenna for compact wireless sensor nodes operating at 24/28/39 GHz," in *Proc. IEEE World Conf. Applied Intelligence and Computing (AIC)*, Sonbhadra, India, pp. 452-455, (2023).
3. R. Przesmycki, M. Bugaj and L. Nowosielski, "Broadband microstrip antenna for 5G wireless systems operating at 28 GHz," *Electronics*, vol. 10, no. 1, (2021).

4. K. A. Fante and M. T. Gameda, "Broadband microstrip patch antenna at 28 GHz for 5G wireless applications," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2238-2244, (2021).
5. P. S. Bhadravathi Ghouse et al., "Dual-band antenna at 28 and 38 GHz using internal stubs," *Technologies*, vol. 12, no. 6, (2024).
6. J.-I. Jung and J.-R. Yang, "5.8-GHz patch antenna with an enhanced defected ground structure for size reduction and increased bandwidth," *Journal of Electromagnetic Engineering and Science*, vol. 22, no. 3, pp. 245-251, (2022).

Model Compression and Skin-Tone Fairness in On-Device Skin Cancer Screening

Pham Thi Anh Thu^{1*)}, Nguyen Gia Khoi²⁾, Tran Trung Hieu³⁾, Bui Thi Thanh Tu⁴⁾

^{1,2,3,4)} Dept. of Information Technology, Ho Chi Minh City University of Foreign Languages and Information Technology (HUFLIT), Ho Chi Minh City, Vietnam

**E-mail : 22dh113663@st.huflit.edu.vn, 22dh111723@st.huflit.edu.vn, 22dh114528@st.huflit.edu.vn, tubtt@huflit.edu.vn*

Abstract: Mobile skin lesion screening applications need to function even without an Internet connection, so the model must be compressed to run directly on the device. However, compression can alter performance across skin tone groups even when overall accuracy is preserved. This paper evaluates the impact of compression techniques and traces the primary cause of the disparity. The results show that INT8 quantization retains nearly the same accuracy and does not reduce fairness, whereas aggressive pruning degrades performance across all groups. Ablation experiments (cause-elimination tests) further indicate that the bottleneck most likely lies in the darker-skin data rather than in the algorithm or the compression process, thereby pointing toward directions for further improvement.

Keywords: Skin-tone fairness; model compression; skin cancer screening; on-device mobile AI.

In many regions that lack dermatologists, a smartphone-based skin cancer screening application can help users identify early on which cases warrant a clinical examination. However, current dermatological datasets are often imbalanced across skin tones, putting the model at greater risk of missing cancers in certain user groups [1], [2], [3]. When deployed on mobile devices, the model must also be compressed to meet memory and processing-speed constraints, and the compression itself can further exacerbate this disparity in underserved groups [4]. This raises an important question: do the observed disparities originate from the data, from the compression process, or from the algorithm itself? This paper focuses on answering that question in order to identify the most effective direction for improving future skin cancer screening systems.

The study uses the Fitzpatrick17k dataset [1] for primary training and evaluation, while testing generalization on DDI [2] and PAD-UFES-20 [3]. Fairness is assessed using the equal opportunity criterion [5], measuring the difference in cancer-detection sensitivity across skin tone groups. The base model is MobileNetV2 [6], optimized for mobile devices, with two common compression techniques applied: INT8 quantization [7] and structured pruning [8]. To ensure statistical reliability for the dark-skin group, which has few samples, the study employs 5-fold cross-validation, pooling out-of-fold predictions, and bootstrapping. In addition, three data-rebalancing and training strategies are tested to determine whether the disparity originates from the data, the compression process, or the algorithm.

The results show that INT8 quantization (both PTQ and QAT) preserves almost the same accuracy and fairness as the FP32 model, with comparable dark-skin sensitivity (QAT is even slightly higher), while shrinking the model to about 2.6–2.7 MB and nearly doubling on-device inference speed (Table 1). In contrast, aggressive pruning degrades performance across all skin-tone groups: at the 90% level, sensitivity for the light, intermediate, and dark groups all falls from roughly 0.8 to roughly 0.6 (Table 1), so the fairness gap narrows only because every group becomes equally worse, i.e., leveling-down [9]. The three rebalancing strategies improve dark-skin sensitivity only marginally and do not reach statistical significance. With the limited data available, we cannot firmly attribute the disparity to the algorithm or to the compression step; the bottleneck most likely lies in the scarcity of dark-skin data. Finally, performance drops substantially on out-of-domain datasets, indicating that the model needs on-site fine-tuning before real-world deployment.

After successively testing the hypotheses about the source of the disparity across skin tones — from the model compression process and the rebalancing techniques to the training data — we found no evidence that INT8 quantization is a cause of increased unfairness. Conversely, aggressive pruning degrades performance across all

groups in a leveling-down manner. The rebalancing interventions likewise did not produce statistically significant improvements. Taken together, these results indicate that the current bottleneck most likely lies in the scarcity and limitations of dark-skin data rather than in the compression algorithm or the model itself. More importantly, the study shows that statistically well-powered causal analysis is necessary before proposing fairness solutions in dermatological AI, such as group-aware pruning like FairPrune [10].

Table 1. Balanced accuracy, sensitivity by skin tone group, and storage size of the compression configurations (Fitzpatrick17k, OOF n=4312). Bootstrap 95% CI for dark-skin sensitivity: ~0.74–0.85 for FP32/INT8, ~0.54–0.72 for pruning.

Configuration	bAcc [CI 95%]	Sensitivity light/ med / dark	Size
FP32	0.830 [0.82, 0.84]	0.82 / 0.85 / 0.79	8.90 MB
PTQ (INT8)	0.830 [0.82, 0.84]	0.83 / 0.86 / 0.79	2.60 MB
QAT (INT8)	0.828 [0.82, 0.84]	0.79 / 0.84 / 0.80	2.70 MB
Prune 50%	0.705 [0.69, 0.72]	0.69 / 0.65 / 0.66	2.50 MB
Prune 75%	0.640 [0.63, 0.65]	0.68 / 0.61 / 0.64	0.77 MB
Prune 90%	0.604 [0.59, 0.62]	0.58 / 0.61 / 0.61	0.23 MB

Acknowledgments

This research was funded by the scientific project coded HSV2025-27, financed by Ho Chi Minh City University of Foreign Languages – Information Technology.

References

1. M. Groh, C. Harris, L. Soenksen, et al., Proc. IEEE/CVF CVPR Workshops (2021).
2. R. Daneshjou, K. Vodrahalli, R. A. Novoa, et al., Sci. Adv., 8(32), eabq6147 (2022).
3. A. G. C. Pacheco, G. R. Lima, A. S. Salomão, et al., Data in Brief, 32, 106221 (2020).
4. C. Tran, F. Fioretto, J.-E. Kim, R. Naidu, Adv. Neural Inf. Process. Syst. (NeurIPS) (2022).
5. M. Hardt, E. Price, N. Srebro, Adv. Neural Inf. Process. Syst. (NeurIPS) (2016).
6. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Proc. IEEE/CVF CVPR (2018).
7. B. Jacob, S. Kligys, B. Chen, et al., Proc. IEEE/CVF CVPR (2018).
8. G. Fang, X. Ma, M. Song, M. B. Mi, X. Wang, Proc. IEEE/CVF CVPR (2023).
9. D. Zietlow, M. Lohaus, G. Balakrishnan, et al., Proc. IEEE/CVF CVPR (2022).
10. Y. Wu, D. Zeng, X. Xu, Y. Shi, J. Hu, Proc. MICCAI (2022).

Analytical and Deep Learning Method for Phase Error Calibration of FMCW Automotive Radar: A Comparative Analysis

Sharif Danish¹⁾, Sheriff Murtala²⁾ and Gyu Sang Choi^{3*)}

^{1,2)} Dept. of Information and Communication Engineering, Yeungnam University, Korea

³⁾ Dept. of Information and Communication Engineering, Yeungnam University, Korea

*E-mail: danishsharif@yu.ac.kr, Sheriffm@yu.ac.kr, castchoi@yu.ac.kr

Abstract: In automotive radar systems, the phase shifter plays a critical role in ensuring accurate performance in frequency-modulated continuous-wave (FMCW) operation. However, its behavior often deviates from the ideal response due to noise, hardware imperfections, and temperature variations. In this paper, we compare the least-squares (LS) approach for phase error calibration under stationary target scenarios with proposed deep learning (DL)-based methods, including an autoencoder (AE) and a generative adversarial network (GAN), for signal reconstruction and phase error correction. The methods are evaluated over a signal-to-noise ratio (SNR) range of -20 dB to 40 dB, using training data collected at target distances of 2 m, 5 m, and 10 m. Experimental results demonstrate that the proposed DL methods outperform the LS approach by up to $9\times$ in terms of root mean square error (RMSE) at low SNR levels (below 10 dB) and generalize effectively to unseen distances with less than 2% performance degradation. However, the LS method achieves better performance at higher SNR levels (above 15 dB), albeit at the cost of a dedicated calibration transmission cycle.

Keywords :Phase shift, FMCW radar, Least squares, Deep Learning , Auto encoder, GAN, Calibration.

I. Introduction:

Frequency-modulated continuous-wave (FMCW) radar plays a vital role in modern autonomous vehicles due to its ability to operate effectively under various challenging weather conditions, making it well-suited for applications such as advanced driver assistance systems (ADAS) [1]. In multiple-input multiple-output (MIMO) configurations, slow-time phase-coded radar (ST-PC) multiplexing assigns each transmit (Tx) channel a controlled phase offset using a phase shifter (PS) to achieve signal diversity [2]. However, practical phase shifters often deviate from their ideal behavior due to silicon limitations, introducing a phase error $\epsilon[k]$ for each transmission index k [3]. This degradation adversely affects beamforming performance and target localization accuracy, thereby necessitating effective calibration.

The method proposed in [3] presents a built-in self-calibration technique that leverages the intermediate frequency (IF) signal leakage phenomenon, typically regarded as an unwanted DC offset, as a useful calibration signal, thereby eliminating the need for dedicated hardware loopbacks used in earlier approaches [4], [5]. In this paper, we present a comparative analysis of a custom FMCW radar architecture designed to reconstruct signals and mitigate phase errors using deep learning approaches, including a generative adversarial network (GAN) and an autoencoder (AE). The DL-based methods[6] are compared with the conventional least-squares (LS) approach for FMCW transmission by analyzing their operational limitations and evaluating performance in terms of root mean square error (RMSE) across a range of signal-to-noise ratio (SNR) levels and target distances..

II. Analytical and Deep learning Methods:

2.1 Analytical Methods:

In radar systems, low-cost phase shifters are often used due to economic and practical considerations; however, they require calibration and monitoring procedures to compensate for phase errors. In method [4], each transmitter channel employs square-wave modulation to generate a Tx-to-Rx loopback signal, which is subsequently analyzed to assess phase shift accuracy. A similar approach in [5] utilizes modulated loopback schemes to vary the phase shifter settings and observe phase deviations at the receiver. In [7], cancellation techniques are applied to mitigate the effects introduced by the leakage phenomenon, whereas [3] presents an analytical method that exploits the leakage phenomenon itself to estimate and calibrate phase error, claiming to be hardware-free.

2.2 Deep Learning(AE and GAN):

Deep learning approaches have also been applied to phased array calibration. Methods in [8] and [9] estimate phase error using power-only measurements, demonstrating robustness under noisy conditions. In contrast,

the proposed deep learning methods, namely the autoencoder (AE) and generative adversarial network (GAN), focus on signal reconstruction and phase error calibration in FMCW radar signals under noisy scenarios. Using a MATLAB/Simulink simulation of a 77 GHz FMCW radar, this study considers stationary targets located at 2 m, 5 m, and 10 m, across 13 signal-to-noise ratio (SNR) levels ranging from -20 dB to 40 dB. The AE uses an autoencoder-decoder with uniform Conv1D layers (filter Size = 128, Leaky-relu $\alpha = 0.3$, downsample factor=2,) trained with RMSE loss. The GAN also uses a similar but a different encoder-decoder as its generator, with separate filter counts for down sample layer (64 filters) and external layers (128 filters) enhanced with convolutional discriminator. Both models are trained independently from the random initialization using the Adam optimizer on the same dataset. Clean and noisy intermediate frequency (IF) signals, consisting of 1500 samples per transmission, are collected at the receiver and used to train the AE and GAN models for up to 100 epochs. Figure 1 illustrates the block diagram of the deep learning architecture for phase error calibration in an FMCW radar system.

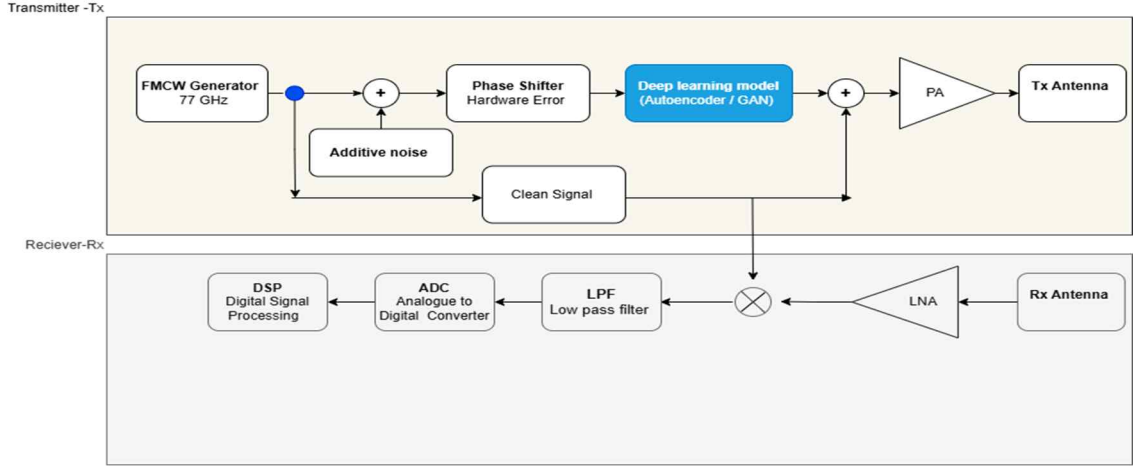


Fig. 1. Block diagram of the deep-learning methodology for signal reconstruction and phase-error calibration in an FMCW radar system.

III. Results and Discussion :

The model is trained using data at a target distance of 5 m, with a split of 80% for training, 10% for validation, and 10% for testing, and is subsequently evaluated on unseen distances of 2 m and 10 m. Unlike the least-squares (LS) method, which requires a stepped phase shifter sweeping from 0° to 360° over $K = 512$ transmissions, the DL methods employ fixed 180° phase shift, enabling standard radar operation with zero calibration overhead. As shown in Figure 2, a clear performance crossover occurs in the SNR range of 10 – 15 dB. Below this range, which represents realistic automotive conditions such as rain, fog, or long-range targets [10], the deep learning approaches significantly outperform the LS method. In particular, the proposed methods achieve up to 9x lower RMSE at an SNR of -20 dB. However, the LS method remains advantageous at higher SNR levels, albeit at the cost of requiring stepped phase shifter hardware, a validity constraint of $|\epsilon[k]| \leq 1^\circ$, and a dedicated calibration cycle of 512 transmissions. Across different target distances, the AE and GAN models maintain an RMSE in the range of 10.6° to 12.9° , demonstrating strong generalization capability, which is notably limited in the LS approach. Figure 3 presents a comparison between the noisy input signal and the reconstructed outputs from the AE and GAN models across multiple distances.

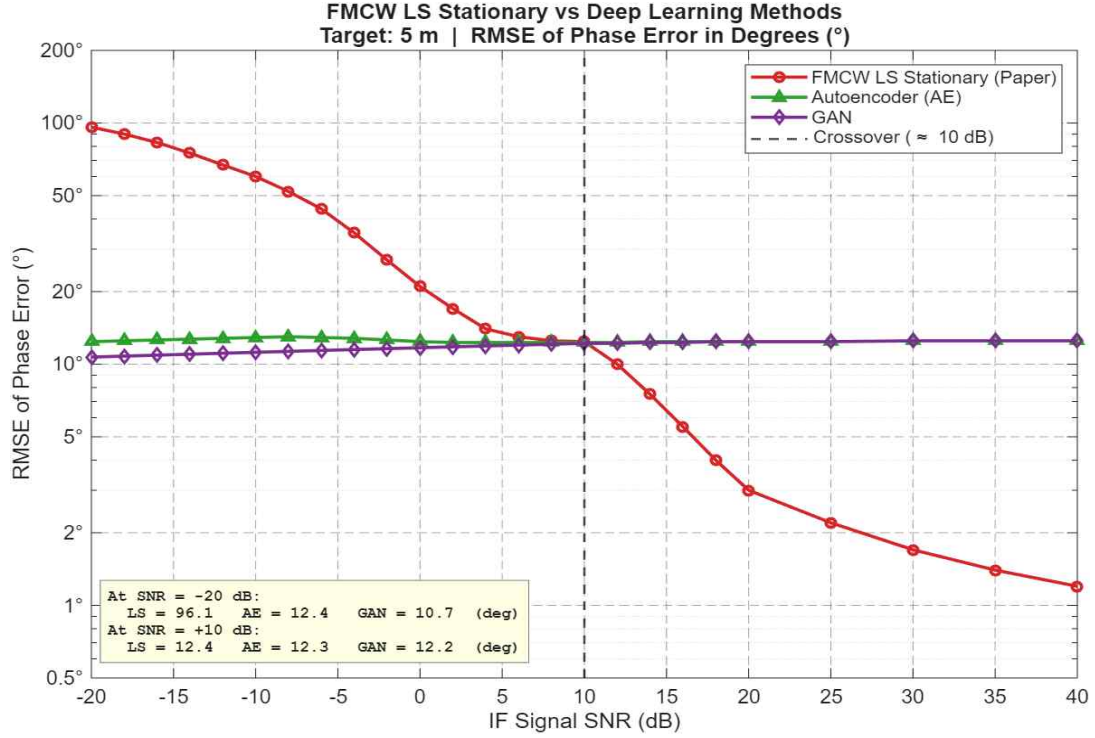


Fig. 2. RMSE of phase error vs. IF-signal SNR for the FMCW-LS method (paper) compared with the DL (AE and GAN) at the 5 m training distance.

IV. Conclusion:

We analyzed the LS-based built-in self-calibration method of [3] for FMCW radar, confirming both its effectiveness and its practical limitations. The primary limitation is the calibration overhead, which requires 512 dedicated phase-coded transmissions per cycle, during which normal sensing is unavailable. Additionally, model mismatch introduces an RMSE floor of 1.2° at 40 dB SNR, as the single-sinusoid model (Eq. 11 in [3]) cannot accurately represent multiple superposed leakage sources. The computational complexity of the least-squares (LS) method also limits its suitability for real-time embedded applications, and its validity is restricted to a phase error range of $|\varepsilon[k]| \leq 1^\circ$. To address these limitations, deep learning approaches, including the autoencoder (AE) and generative adversarial network (GAN), are employed using a fixed phase shifter configuration, eliminating calibration overhead. Comparative results show that at low SNR levels (below 10 dB), the proposed DL methods outperform LS by up to $9\times$ in RMSE, reducing the error from 96.1° (LS) to 10.7° (GAN) at -20 dB SNR. At higher SNR levels (above 15 dB), LS achieves a lower RMSE floor of 1.2° , compared to the DL plateau of approximately 12° , highlighting the trade-off between the high precision of analytical methods under ideal conditions and the robustness of DL approaches under practical constraints. Furthermore, the DL models generalize effectively to unseen target distances of 2 m and 10 m, despite being trained only at 5 m, with less than 2% performance degradation. This level of generalization is not achievable with the closed-form LS estimator. Therefore, LS is preferable in high-SNR scenarios with dedicated calibration hardware, whereas DL-based methods are more suitable for practical deployment in realistic automotive environments, where low SNR conditions and zero calibration overhead are critical.

Acknowledgment

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) Education (NRF-2021R1A6A1A03039493) and by the BK21 FOUR of the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2120251015517).

References–

1. Gerstmaier, M.; et al. On the Safe Road Toward Autonomous Driving: Phase Noise Monitoring in Radar Sensors. *IEEE Signal Process. Mag.* 2019, 36 (5), 60–70.
2. Jeannin, M.; Lang, O.; Nugraha, D. T.; Bin Khalid, F.; Achatz, S.; Roger, A.; Huemer, M. An Iterative Phase Shifters Online Calibration Technique for Automotive Radar Systems. In *Proc. European Radar Conf. (EuRAD)*; 2022; pp 237–240
3. Preniqi, A.; Lang, O.; Belitzer, A.; Schmalzl, S.; Märzinger, G.; Feger, R. *Built-in Self Calibration for CW Automotive Radar Transmitter Phase Shifters*. In *Proc. IEEE Radar Conf. (RadarConf25)*; 2025..
4. K. Subburaj et al., "Monitoring Architecture for a 76-81GHz Radar Front End," *IEEE RFIC Symp.*, 2018, pp. 264-267.
5. Subburaj, K.; Mani, A.; Dandu, K.; Bhatia, K.; Ramasubramanian, K.; Murali, S.; Sachdev, R.; Gupta, P.; Samala, S.; Shetty, D.; Parkar, Z.; Ram, S.; Dudhia, V.; Breen, D.; Bharadwaj, S.; Bhatara, S.; Ginsburg, B. Digitally Assisted mm-Wave FMCW Radar for High Performance. In *Proc. IEEE RFIC Symp.*; 2020; pp 1–4.
6. D. Sharif, S. Murtala, and G. S. Choi, "Deep learning based signal reconstruction and phase error calibration in 77 GHz FMCW radar system," in *Proc. IEEE IEMEK Conf.*, 2026
7. Fiore, W. Arriola, R. Cobbler, O. Lang, and A. Melzer, "PhaseCalibration of a Radar System with Crosstalk Cancellation, "Patent US11635489B2, Apr. 25, 2023. [Online]. Available:<https://patents.google.com/patent/US11635489B2/en>
8. Sarayloo, Z.; Masoumi, N.; Shahi, H.; Mirza Alian, E. H.; Naeini, S. S.; Ahmadabadi, M. N. A Convolutional Neural Network Approach for Phased Array Calibration Using Power-Only Measurements. In *Proc. Iranian Conf. Electrical Engineering (ICEE)*; 2020; pp 1–6. DOI: 10.1109/ICEE50131.2020.9260769.
9. Xie, X.; et al. A Power-Only Fast Calibration Method for Phased Array Using Convolution Neural Network. *IEEE Antennas Ws Propag. Lett.* 2024, 23 (11), 3972–3976. DOI: 10.1109/LAWP.2024.3452268.
10. Zang, S.; Ding, M.; Smith, D.; Tyler, P.; Rakotoarivelo, T.; Kaafar, M.A. The Impact of Adverse Weather Conditions on Autonomous Vehicles: How Rain, Snow, Fog, and Hail Affect the Performance of a Self-Driving Car. *IEEE Veh. Technol. Mag.* 2019, 14 (2), 103–111.

Residual Reinforcement Learning Compensation for ADRC-Based Waypoint Tracking of a Solar Pile Driver

Myeong Woo Lee¹⁾, Taeyong Kim²⁾, Miok Kim³⁾, Hungju Shin⁴⁾, and Moon G. Joo^{5*)}

^{1, 5*)} Department of Intelligent Robot Engineering, Pukyong National University, Busan, Korea
^{2, 3, 4)} Xpanner Co., Ltd., Busan, Korea

*E-mail: gabi@pknu.ac.kr

Abstract: This paper proposes a residual reinforcement learning (RL) compensation method for active disturbance rejection control (ADRC)-based waypoint tracking of a tracked solar pile driver. To overcome the residual error of fixed-gain ADRC without the instability and long training of pure RL, the proposed method keeps Stanley-ADRC as a stabilizing baseline and trains RL to output only bounded residual torque corrections during path following. Under a persistent constant disturbance, the ADRC-RL controller completes all 27 waypoints, whereas the PID and ADRC baselines stop at 13 and 24, respectively.

Keywords: active disturbance rejection control, residual reinforcement learning, solar pile driver, waypoint tracking, robust control

Introduction: Solar pile drivers operate in outdoor construction environments where ground condition, slope, slip, and repeated external forces make waypoint tracking difficult. In previous work, a Stanley-ADRC controller was developed for a tracked solar pile driver and showed stronger disturbance rejection than a Stanley-PID controller [1]. ADRC estimates and compensates for lumped disturbance using an extended state observer (ESO), and its robustness has been widely investigated in nonlinear control systems [2], [3]. However, fixed-gain ADRC may still leave residual error under persistent disturbances, since the ESO estimate is not exact under observer bandwidth limits and nonlinear soil-track interaction. This study extends the Stanley-ADRC framework by adding a residual RL policy that compensates for the remaining tracking error without replacing the stabilizing ADRC loop.

Proposed ADRC-RL Architecture: The proposed structure is shown in Fig. 1, and is implemented in a Gazebo-ROS2 environment using the same tracked solar pile driver model and waypoint sequence as the previous Stanley-ADRC study [1].

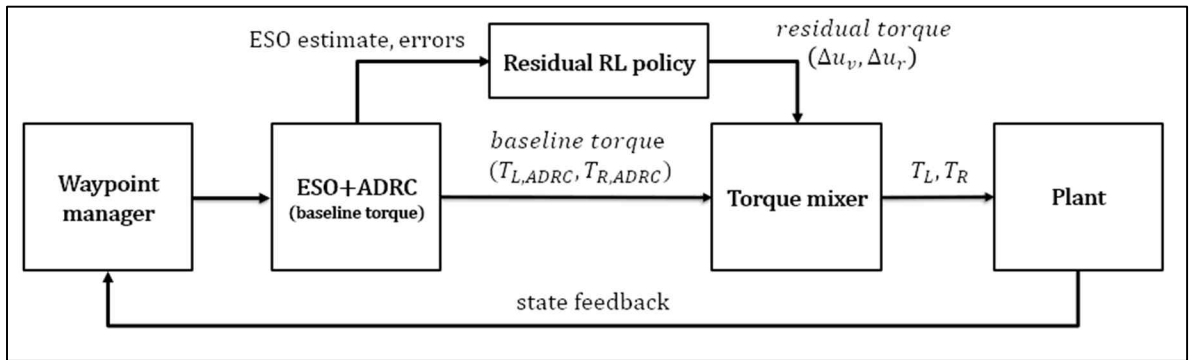


Fig. 1. Proposed Stanley-ADRC with residual RL compensation architecture.

The waypoint manager provides target points, and the Stanley-ADRC controller generates the tracking reference from cross-track error [4] and computes the baseline longitudinal and yaw control inputs, with its ESO estimating the lumped disturbance. The measured vehicle state is fed back for waypoint completion checking and for cross-track and heading error computation in the controller. The RL policy is trained using PPO [5] and follows the residual RL concept [6]: it does not replace the baseline controller, but only outputs bounded residual commands Δu_v and Δu_r . The policy receives normalized tracking errors, vehicle states, ADRC control outputs, and ESO disturbance estimates. The final left and right track torques are computed

as

$$T_L = \text{sat} (T_{L,ADRC} + \Delta u_v - \Delta u_r) \quad (1)$$

$$T_R = \text{sat} (T_{R,ADRC} + \Delta u_v + \Delta u_r) \quad (2)$$

To avoid destabilizing the baseline controller, the residual torque is bounded and enabled only in the path-following state, while stopping and heading alignment remain controlled by ADRC. The reward function encourages waypoint progress and path accuracy while penalizing excessive residual action, torque saturation, reverse motion, timeout, and large cross-track error.

Experimental Results: The simulation uses the same tracked solar pile driver model and waypoint sequence as the previous Stanley-ADRC study [1]. A waypoint is counted as reached only when the position error is less than 0.01 m and the heading error is less than 0.1° , maintained for 5 s. The persistent constant disturbance applies $F_x = 200N$, $F_y = 200N$, and $M_z = 100N$ throughout the entire episode. Under the nominal condition, all controllers completed the full route, indicating that the residual RL policy—trained under the persistent disturbance—did not destabilize the ADRC baseline in the disturbance-free case. Under the persistent disturbance, however, PID and ADRC failed to complete the route within the 10,000 s limit, stopping at 13/27 and 24/27 waypoints, respectively, whereas ADRC-RL completed all 27 waypoints in 7537.3 s. Because ADRC-RL shows larger RMS tracking errors and a higher saturation ratio than ADRC in this case, the result should be interpreted as an improvement in route completion and recoverability rather than a universal reduction in tracking error.

Table 1. Preliminary simulation results

Disturbance	Controller	Success	Final waypoint	Completion time[s]	RMS CTE[m]	RMS heading error[deg]	Saturation ratio[%]
None	PID	Yes	27/27	3628.6	0.2200	13.250	1.709
None	ADRC	Yes	27/27	7857.2	0.1530	9.236	1.195
None	ADRC-RL	Yes	27/27	6641.6	0.1646	10.196	1.368
Constant	PID	No	13/27	10000.0	0.0925	6.093	0.270
Constant	ADRC	No	24/27	10000.0	0.1212	7.047	0.661
Constant	ADRC-RL	Yes	27/27	7537.3	0.1558	10.248	1.245

Conclusion: This paper presented an ADRC-RL hybrid control approach for robust waypoint tracking of a tracked solar pile driver. The method preserves the reliability of ADRC while using RL only as a bounded adaptive residual compensator. The preliminary results indicate that ADRC-RL improves route completion under persistent disturbance. However, the added residual input means torque usage and RMS tracking error do not always decrease, and future work will validate the approach across additional disturbance magnitudes and repeated trials.

References

1. M. W. Lee, T. Kim, M. Kim, H. Shin, and M. G. Joo, "A Waypoint Tracking for a Solar Pile Driver Using Stanley-ADRC," *IEMEK Journal of Embedded Systems and Applications*, vol. 21, no. 1, pp. 19-28, 2026.
2. J. Han, "From PID to Active Disturbance Rejection Control," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 3, pp. 900-906, 2009.
3. Z. Gao, "Scaling and Bandwidth-Parameterization Based Controller Tuning," *Proceedings of the American Control Conference*, pp. 4989-4996, 2003.
4. S. Thrun et al., "Stanley: The Robot That Won the DARPA Grand Challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661-692, 2006.
5. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv:1707.06347*, 2017.
6. T. Johannink et al., "Residual Reinforcement Learning for Robot Control," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 6023-6029, 2019.

6-Degree-of-Freedom Pose Estimation of Solar Panels Using Vision Transformer and Synthetic CAD-Based Dataset

Tae Young Jeong¹⁾, Miok Kim²⁾, Hungju Shin³⁾, Moon G. Joo^{4 *)}

^{1,4)} Dept. of Intelligent Robot Engineering, Pukyong National University, Busan 48513, Korea

^{2,3)} Xpanner Co., Ltd.

*E-mail: gabi@pknu.ac.kr

Abstract: This paper proposes a 6-degree-of-freedom (6DoF) pose estimation method for solar panels using a Vision Transformer (ViT) backbone and a large-scale synthetic dataset generated from CAD models. The proposed method directly regresses three rotation angles (pitch, yaw, roll) and three translation components (tx, ty, tz) from a single RGB image. Training data was generated using BlenderProc with realistic lighting conditions, occlusion, and noise augmentation. Experimental results demonstrate a mean angular error of 0.85 degrees and a mean translation error of 67 mm, suggesting that the proposed method achieves reliable 6DoF pose estimation using synthetic training data alone.

Keywords : pose estimation, Vision Transformer, solar panel, synthetic data, 6DoF

In the first paragraph, Introduction: 6DoF pose estimation is essential for robotic solar panel manipulation. CNN-based approaches have shown that synthetic images rendered from 3D CAD models can effectively train pose estimation models [1,2], but their local convolutional operations may limit global context modeling. Vision Transformers[3] address this through self-attention, enabling global dependency modeling from the first layer. We therefore use a ViT-Base backbone trained on synthetic solar panel data for robust 6DoF pose estimation.

In the second paragraph, Research Methods : The proposed pipeline consists of three stages (Fig. 1). First, synthetic training data is generated using BlenderProc [4].

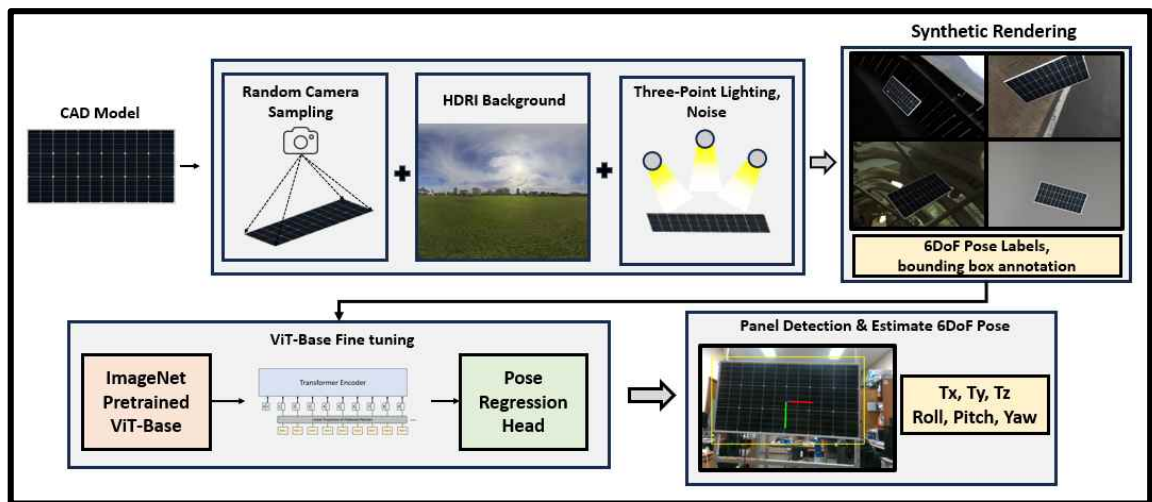


Fig. 1. Proposed Pipeline for Synthetic Data-Driven Solar Panel 6DoF Pose Estimation Using Vision Transformers.

The solar panel CAD model is rendered with random camera viewpoints, HDRI backgrounds from the Haven dataset, and three-point lighting with randomized intensities, producing labeled RGB images with corresponding 6DoF pose annotations. Second, a ViT-Base model [3] pretrained on ImageNet-21k is fine-

tuned to regress six pose parameters from a 224×224 RGB image. A YOLOv8n-seg detector is used to localize the solar panel in the input image and provide the bounding box required for pose estimation. It was trained separately on 1,402 labeled images (85% training, 15% validation) using AdamW optimizer with Binary Cross-Entropy loss. Finally, during inference, the detector localizes the solar panel in RGB images captured by an Intel RealSense D435 camera, and the fine-tuned ViT-Base estimates its 6DoF pose in real time.

In the third paragraph, Experimental Results: All training was performed on an NVIDIA RTX 4090 GPU. The model was trained on approximately 160,716 synthetic images (85% Train data, 15% Validation Data) with AdamW optimizer (learning rate $1e-4$) for up to 150 epochs with early stopping. L1 loss was applied separately for angular and translation components with equal weighting. Fig. 2 shows training/validation loss curves and angular/translation errors on the validation set. The model converges stably, achieving a final angular error of approximately 0.85 degrees and a translation error of approximately 67 mm. The proposed ViT-based approach benefits from global attention mechanisms that better capture geometric relationships, compared to CNN-based methods trained on synthetic data [1,2].

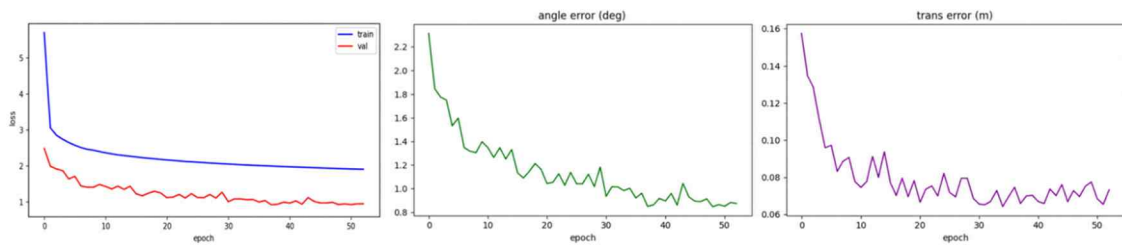


Fig. 2. Training/validation loss curves (left) and validation angular/translation error curves (right) over 52 epochs

In the fourth paragraph, Conclusion: This paper presented a 6DoF pose estimation method for solar panels using a ViT-Base backbone [3] trained entirely on synthetic data generated from a CAD model using BlenderProc [4]. Diverse lighting simulation, Haven HDRI backgrounds, and realistic data augmentation were employed to enhance the diversity and realism of the synthetic training data. Experimental results achieved a mean angular error of 0.85 degrees and a translation error of 67 mm, demonstrating that the proposed method can achieve accurate pose estimation in a synthetic-data-based evaluation environment. Future work will focus on real-image fine-tuning, validation in real-world environments, and integration with robotic manipulation systems.

References

1. Y. Su, J. Rambach, A. Pagani, and D. Stricker, "SynPo-Net: Accurate and Fast CNN-Based 6DoF Object Pose Estimation Using Synthetic Training," *Sensors*, vol. 21, no. 1, p. 300, 2021.
2. X. He et al., "Deep Learning-Based 6-DoF Object Pose Estimation Considering Synthetic Dataset," *Sensors*, vol. 23, no. 24, p. 9854, 2023.
3. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proc. ICLR*, 2021.
4. M. Denninger et al., "BlenderProc," *arXiv:1911.01911*, 2019.

Language-Guided Unknown-Space Exploration for Mobile Robots with a Mamba-Based Frontier Planner

Moojueng Kim¹⁾, Junhui Woo²⁾ and Sangmoon Lee^{3*)}

^{1,2)} Dept. of Water and IT Eng., Kyungpook Natl. Univ., Daegu 41566, Korea

³⁾ Dept. of Electronic Eng., Kyungpook Natl. Univ., Daegu 41566, Korea

*E-mail: moony@knu.ac.kr

Abstract: Autonomous exploration in unknown 3D environments requires a robot to build spatial knowledge while deciding where to move next. Unlike fixed-goal navigation, this work focuses on exploration-driven navigation, where a language-specified object may not be visible at the beginning of the mission. Inspired by ETPNav-style online topological planning, this paper proposes a lightweight language-guided frontier exploration framework using an explicit 3D voxel frontier graph and a Mamba-style State Space Model (SSM) frontier scorer. Language grounding converts commands into compact target cues for target-object recognition, while path planning is performed by the frontier graph and SSM scorer. Rather than claiming the novelty of frontier exploration itself, this work verifies that language conditioning, 3D voxel frontier memory, and a Mamba-style SSM decision module can be integrated into a complete exploration pipeline. Across 12 unseen 3D voxel maps, the proposed method achieved 100% target discovery success, 21.67 average target discovery steps, and 0.33% average revisit ratio. Future work will include transformer-vs-SSM ablation and hybrid onboard validation using Jetson Orin and a high performance control PC.

Keywords : State Space Model, Mamba, Frontier Exploration, 3D Mobile Robot Navigation, Vision-Language Navigation, Topological Map

Autonomous navigation is essential for mobile robots in unknown indoor environments. ROS2/Nav2 provides planning and behavior-tree based execution [1], but exploration is usually handled by additional methods such as frontier-based exploration [2]. Classical frontier selection is efficient but mostly geometric, making it hard to reflect natural-language object goals. Recent VLN and embodied exploration studies use online topological memory, cross-modal transformers, or object/frontier queries [3], [4]. This work redesigns these useful elements for continuous unknown-space exploration in a 3D voxel map, rather than claiming frontier graphs as a new concept. Since transformer reasoning can become costly as frontier nodes and topological memories grow, we use a Mamba-style SSM frontier scorer as a lightweight alternative, motivated by S4 [5] and Mamba [6].

The proposed framework consists of three tightly integrated components:

Frontier Graph Construction. The robot updates an online 3D voxel map and extracts frontiers, defined as observed free voxels adjacent to unknown voxels. Candidate frontiers are represented as graph nodes, and path feasibility is computed through graph-based breadth-first search (BFS). Each node encodes robot-to-frontier relative position, graph distance, adjacent unknown voxel count, free degree, previous action, and visitation penalty.

Language Grounding. Natural-language commands are converted into compact target cues, including target color, target object, and target saliency value. This module is used for target-object recognition and semantic conditioning, not for directly generating a full path. The target cues are concatenated with geometric and topological frontier features.

SSM Frontier Scorer. At each exploration step, frontier candidates are ordered into a sequence and processed by a Mamba-style selective SSM. Let $\mathbf{x}_k$ be the target-conditioned feature of the k -th frontier candidate and $\mathbf{h}_k$ be the accumulated hidden state. The discrete-time state update is written conceptually as

$$\mathbf{h}_k = \mathbf{A}\mathbf{h}_{k-1} + \mathbf{B}\mathbf{x}_k, \quad \mathbf{y}_k = \mathbf{C}\mathbf{h}_k + \mathbf{D}\mathbf{x}_k,$$

where x_k is the target-conditioned feature and h_k is the hidden state. In implementation, the update is selective and gated, following the Mamba-style SSM idea. The highest-scoring frontier is selected as the next exploration target, and graph-based BFS generates the path.

Fig. 1 shows representative snapshots of the proposed Mamba-style SSM frontier exploration pipeline. The green dot is the robot, the blue dashed line is the trajectory, white voxels are explored free space, and the red dot is the language-specified target. Across 12 unseen maps, the system achieved 100% target discovery success, 21.67 average target discovery steps, and 0.33% average revisit ratio, confirming functional operation in unseen 3D voxel environments.

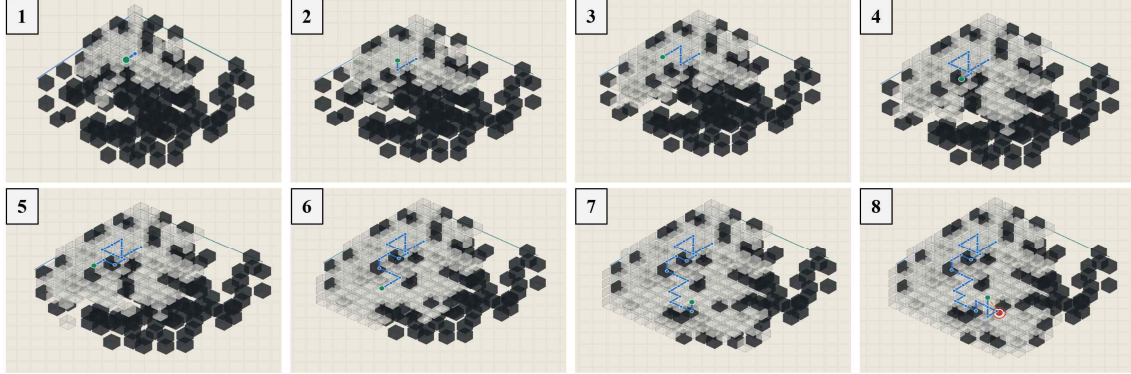


Fig. 1. Representative snapshots of the Mamba-style SSM frontier exploration simulation

This paper presented a lightweight language-guided unknown-space exploration framework that integrates existing language grounding, frontier exploration, and state-space sequence modeling ideas into a compact 3D voxel pipeline. The proposed pipeline uses language grounding for target-object recognition, an explicit 3D voxel frontier graph for online spatial memory, a Mamba-style SSM frontier scorer for target-conditioned frontier selection, and graph-based BFS for path execution. Future work will compare transformer and SSM frontier reasoning under the same architecture and validate the pipeline on a physical mobile robot through a hybrid onboard system combining Jetson Orin and a control PC.

Acknowledgment

This work is financially supported by Korea Ministry of Environment (MOE) as Graduate School specialized in Integrated Water Resources Management.

References

1. S. Macenski, F. Martín, R. White, and J. Ginés Clavero, "The Marathon 2: A Navigation System," in *Proc. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Las Vegas, NV, USA, 2020.
2. B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proc. 1997 IEEE Int. Symp. Comput. Intell. Robot. Autom. (CIRA)*, Monterey, CA, USA, 1997, pp. 146–151.
3. D. An, H. Wang, W. Wang, Z. Wang, Y. Huang, K. He, and L. Wang, "ETPNav: Evolving Topological Planning for Vision-Language Navigation in Continuous Environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, pp. 5130–5145, 2024.
4. JZ. Zhu et al., "Move to Understand a 3D Scene: Bridging Visual Grounding and Exploration for Efficient and Versatile Embodied Navigation," arXiv:2507.04047, 2025.
5. A. Gu, K. Goel, and C. Ré, "Efficiently Modeling Long Sequences with Structured State Spaces," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022.
6. A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Vienna, Austria, 2024.

Digital Twin Network based on Decentralized Federated Learning in On-device AI Environments

Nguyen Anh Tuan^{1,2)}, Qazi Waqas Khan^{1,2)}, Le Anh Ngoc³⁾ and Do Hyeun Kim^{1,2 *)}

¹⁾ Dept. of Computer Engineering, Jeju National University, Jeju-si, Republic of Korea

²⁾ Big Data Research Center, Jeju National University, Jeju-si, Republic of Korea

³⁾ Swinburne Vietnam, FPT University, Hanoi, Vietnam

*E-mail: kimdh@jejunu.ac.kr

Abstract: The rapid growth of on-device artificial intelligence has created new opportunities for intelligent, privacy-preserving, and real-time services across distributed edge environments. However, traditional centralized learning architectures often suffer from high communication overhead, data privacy risks, and limited scalability. To address these challenges, this paper proposes a Digital Twin Network based on Decentralized Federated Learning for on-device AI environments. The proposed framework constructs digital twin models for distributed edge devices to represent their computational status, data characteristics, resource availability, and learning performance in real time. The digital twin network supports adaptive participant selection, resource-aware model aggregation, and real-time monitoring of learning processes, improving the efficiency and robustness of distributed intelligence. By integrating digital twin technology with decentralized federated learning, the proposed approach enhances data privacy, reduces communication latency and enables scalable AI model training in heterogeneous on-device environments. Experimental evaluation demonstrates that the proposed framework can achieve efficient model convergence, and lower network overhead.

Keywords : On-device AI environments, Digital Twin Network, Decentralized Federated Learning

Introduction: The rapid advancement of on-device artificial intelligence has enabled intelligent data processing directly on edge devices with reduced latency and improved privacy [1]. However, the heterogeneous nature of edge devices creates challenges in model training, resource management, communication efficiency, and real-time adaptability [2]. Digital Twin Networks (DTN) [3] provide a virtual representation of physical devices, allowing continuous monitoring, simulation, in IoT environments. Decentralized Federated Learning (DFL) [4] further enhances privacy and scalability by enabling collaborative model training without relying on a central server or sharing raw data. Therefore, integrating DTN with DFL offers a promising solution for building adaptive, secure, and efficient on-device AI systems.

Research Methods: The research method of this study focuses on designing a DTN integrated with DFL for on-device AI environments. First, digital twin models are created to represent the status, resources, data characteristics, and learning behavior of distributed edge devices. Then, a decentralized federated learning mechanism is applied to enable devices to collaboratively train AI models through peer-to-peer communication without transferring raw data to a central server. The proposed method also considers device heterogeneity, communication efficiency, privacy preservation, and adaptive model aggregation.

Experimental Results: The experimental results demonstrate that the proposed Digital Twin Network based on Decentralized Federated Learning improves the performance of on-device AI environments. The framework achieved stable model convergence while reducing communication overhead through peer-to-peer collaboration among edge devices. Compared with conventional centralized federated learning, the proposed method showed better adaptability to heterogeneous device conditions and improved robustness against device failures. The digital twin models effectively monitored device status, resource availability, and learning performance, enabling more efficient participant selection and model aggregation.

Conclusion: This paper presented a Digital Twin Network based on Decentralized Federated Learning for efficient on-device AI environments. The proposed approach enables distributed edge devices to

collaboratively train AI models while preserving data privacy and reducing dependence on a central server. . Experimental results show that the proposed framework improves adaptability, communication efficiency, scalability, and robustness in heterogeneous edge environments. In future work, this approach can be extended with advanced optimization and security mechanisms to support more reliable and intelligent edge AI services.

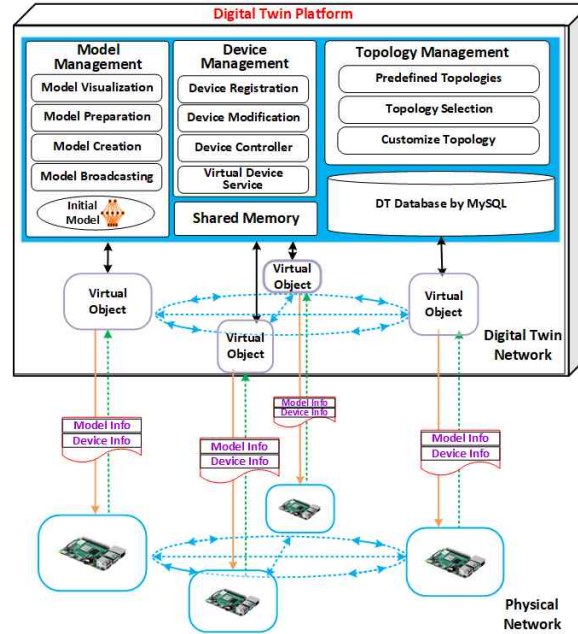


Fig. 1. Architecture of the Digital Twin Platform for On-Device AI Networks

Fig. 1 illustrates a Digital Twin Platform that connects a physical network of edge devices with their corresponding virtual representations in a digital twin network. The platform consists of three main management modules: Model Management, Device Management, and Topology Management. Model Management supports model visualization, preparation, creation, and broadcasting, while Device Management handles device registration, modification, control, and virtual device services. Topology Management manages predefined topologies, topology selection, and customized network structures. Each physical device in the lower network is mapped to a virtual object in the digital twin network, where device information and model information are exchanged continuously.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00405278), and this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00346238). Any correspondence related to this paper should be addressed to DoHyeun Kim.

References

1. Wang, X., Tang, Z., Guo, J., Meng, T., Wang, C., Wang, T., & Jia, W. (2025). Empowering edge intelligence: A comprehensive survey on on-device ai models. *ACM Computing Surveys*, 57(9), 1-39.
2. Mughal, F. R., He, J., Das, B., Dharejo, F. A., Zhu, N., Khan, S. B., & Alzahrani, S. (2024). Adaptive federated learning for resource-constrained IoT devices through edge intelligence and multi-edge clustering. *Scientific Reports*, 14(1), 28746.
3. Wu, Y., Zhang, K., & Zhang, Y. (2021). Digital twin networks: A survey. *IEEE Internet of Things*

Journal, 8(18), 13789-13804.

4. Beltrán, E. T. M., Pérez, M. Q., Sánchez, P. M. S., Bernal, S. L., Bovet, G., Pérez, M. G., ... & Celdrán, A. H. (2023). Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges. *IEEE Communications Surveys & Tutorials*, 25(4), 2983-3013.

Real-Time Face Recognition AI-Based Classroom Seat Matching and Student Identification System

Donghwoon Kwon¹⁾, Seung Kwang Ryu²⁾, Seok-Hyeon Jeong³⁾, Ayoung Song³⁾, and Jeong Tak Ryu^{3*)}

¹⁾ Dept. of Computer Science and Engineering, North Central College, Naperville, IL 60540, U.S.A

²⁾ Graduate School of Smart Convergence Systems Engineering, Daegu University, Korea

³⁾ Department of Electronic Engineering, Daegu University, Korea

E-mail: dkwon@noctrl.edu¹⁾, jryu@daegu.ac.kr^{*}

Abstract: This paper proposes an automated system that integrates Jetson Orin Nano-based real-time facial recognition AI with polygon Region of Interest (ROI) seat-mapping technology to address the issues related to student attendance and leaving without permission in classrooms through seat matching. By extracting facial feature vectors using the InsightFace model, applying L2 normalization, and employing an intersection-area-based contention control algorithm, the system overcomes camera perspective distortion and multi-object overlapping. Experimental results obtained from 20 subjects over 20 iterations (totaling 400 data points) demonstrated a recognition rate of approximately 94.75%, proving the effectiveness of the proposed system for real-time seat matching.

Keywords : Classroom seat mapping, Facial recognition, InsightFace, Intersection-area-based contention control, Polygon Region of Interest (ROI)

1. Introduction

Traditional attendance methods such as roll calls, QR codes, and Bluetooth beacon-based systems exhibit structural limitations, including proxy attendance, cheating in authentication, and delays in academic administration. Although computer vision-based monitoring has been introduced to address these issues, it often suffers from a high false recognition rate caused by camera perspective distortion and physical occlusion among students. Consequently, there is an urgent need for precise spatial discrimination algorithms and high-dimensional feature extraction technologies capable of determining individual students' seating status, even in complex and densely populated environments [1-3]. The purpose of this study is to fully automate student identification and seat matching in classrooms by integrating real-time facial recognition AI with polygon Region of Interest (ROI) mapping, and to verify its effectiveness quantitatively.

2. System Architecture

The system configuration is as follows: A custom housing with a 30-degree downward tilt was fabricated using 3D printing to accommodate a 12.3MP high-resolution sensor paired with a 120-degree ultra-wide-angle lens. A decentralized architecture was established using the Jetson Orin Nano board as the edge device and the Raspberry Pi 5 as the central server.

In this study, InsightFace, a state-of-the-art open-source facial recognition library, was utilized. Based on the large-scale datasets, this model precisely extracted high-dimensional facial feature vectors and was particularly characterized by maximizing inter-class discriminability within the feature space. Hence, it could guarantee exceptionally high facial recognition accuracy even under various environmental constraints, such as changes in illumination, poses, and occlusions [4]. Furthermore, the system corrected distortions and determined real-time seating status through a seat-specific polygon ROI mapping and an intersection-area calculation algorithm between bounding boxes [5].

3. Performance Evaluation

The system performance is evaluated as follows. Life-sized facial photographs of 20 students were printed and attached to seat monitors. To prevent bias caused by lens distortion, the arrangement was randomized for each session, and a total of 20 repetitive photo-taking sessions were conducted to collect 400 independent data points. Optical adverse conditions were intentionally created by inducing a loss of

facial depth information in the flat prints and causing diffuse light reflection. To eliminate visual noise and human error that potentially occur when visually scoring hundreds of facial bounding boxes, an automated scoring script based on set theory was implemented in a Python environment. Figure 1 shows the result of identifying student seat positions through the application of ROI.



Fig. 1 Identification of student seat positions using ROI

The script was designed to define the "actual attendee list per seat," which was prepared through prior human inspection, as the ground truth set (G). It then converted the result list returned by the vision system via real-time inference into a prediction set (P) and automatically classified performance metrics using mathematical set operations.

Table 1 summarizes the experimental results, yielding an overall accuracy of 94.75%. The 5.25% of unrecognized cases was primarily attributed to temporary failures in facial landmark extraction, caused by lens resolution degradation (blurring) at the periphery of the camera's field of view and specular reflection from ceiling lights. Rather than a structural defect in the algorithm itself, this can be interpreted as a legitimate rejection resulting from the conservative similarity threshold established to prevent proxy attendance, combined with the limitations of a 2D simulation environment utilizing flat printed photographs. Since depth information is preserved in actual 3D facial environments, this error rate is expected to decrease further.

Table 1. Experimental Results of 400 Facial Recognition and Seat Matching Tests

Ground Truth/ Prediction	Attendance Approved (Positive)	Unrecognized/ Rejected (Negative)	Total
Genuine Student	379	21	400

4. Conclusion

This study effectively addresses perspective distortion and occlusion in classrooms through computer vision and spatial discrimination algorithms, and presents an integrated academic management pipeline spanning from automated attendance verification to administrative processing. In the future, to overcome the limitations of 2D prints, the practical efficacy of the system will be further validated on actual 3D student faces, and universal system expansion across various classroom environments will be pursued through research on lightweight models.

References

1. A. S. Lateef and M. Y. Kamil, "Face Recognition-Based Automatic Attendance System in a Smart Classroom," *Iraqi Journal for Electrical and Electronic Engineering*, Vol. 20, No. 1, 2024. pp. 37-47.
2. A. Agustiyar, R. Isnanto, C. E. Widodo, and N. A. Zaini, "A Systematic Literature Review of Face Recognition Based Attendance System: Methods, Trends, and Research Gaps," *Journal of Trends and Challenges in Artificial Intelligence*, Vol. 3, No. 4 pp. 135-154, 2026.
3. J. Patel, S. Gandhi, V. Katheriaya, P. Pataliya, and A. Majumadar, "Enhancing Classroom Attendance Systems with Face Recognition through CCTV using Deep Learning Models," *Procedia Computer Science*, Vol. 258, pp. 3031-3041, 2025.
4. J. Deng., J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690-4699, 2019.
5. S. F. N. Faizi and A. Al-Khowarizmi, "Application of Region of Interest (ROI) in Student Attendance Detection System in Classroom," *Electronic Integrated Computer Algorithm Journal*, Vol.3, No.1 (2025) pp. 17-23.

Enhancing Convolutional Neural Network Representation Capability through Feature Diversity

Yeongyu Choi^{1*)}, and Ju H. Park¹⁾

¹⁾ Dept. of Electrical Engineering, Yeungnam University, Gyeongsan 38541, Korea

*E-mail: ygchoi11@ynu.ac.kr

Abstract: Many convolutional neural network (CNN) architectures improve performance by increasing network depth and expanding channel dimensions, thereby providing a trade-off between computational cost and representation capability. However, these approaches primarily rely on model scaling, which restricts the efficient enhancement of representation capability under limited computational resources. In this work, we focus on feature diversity and propose a CNN design modification that enhances representation capability through parallel feature extraction and effective feature fusion under limited computational resources. Our experimental results on CIFAR-100 and Tiny ImageNet benchmarks using ResNet demonstrate that the proposed method achieves higher accuracy with comparable FLOPs than the original architecture, indicating that feature diversity-driven design is a promising approach for improving CNN performance.

Keywords : convolutional neural network, image classification, feature diversity

Introduction: Many CNN architectures have introduced various model variants by increasing the number of blocks and expanding channel dimensions to improve performance [1]. However, for edge-device environments with limited computational resources, it is essential to adopt architectural designs that can efficiently improve representation capability without increasing model scale. In this work, we propose a modified ResNet [2] that improves the feature representation while maintaining the FLOPs.

Proposed Method: To ensure feature diversity, we adopt a fixed sampling offset inspired by deformable convolutional networks (DCNs) [3]. Each branch in the parallel structure extracts features using distinct offsets and independent kernel weights, enabling the learning of diverse feature representations that capture different spatial and semantic information. To maintain spatial consistency among feature maps, the offset displacement is constrained to within 1 pixel. Given a set of feature maps $\{\mathbf{x}_n\}_{n=1}^N$, we first apply batch normalization [4] to obtain $\{\mathbf{x}'_n\}_{n=1}^N$. The features are then concatenated as $\mathbf{x}'_c$, followed by a squeeze-and-excitation (SE) [5] and batch normalization to produce $\mathbf{x}'_c$. Finally, it is split into $\{\mathbf{x}'_n\}_{n=1}^N$ along the channel dimension. The feature map fusion is performed as follows:

$$\mathbf{y} = \frac{1}{2N} \sum_{n=1}^N (\mathbf{x}'_n + \mathbf{x}'_n),$$

where it should be noted that $\mathbf{x}'_n$ and $\mathbf{x}'_n$ are normalized before averaging. Since the parameters of each branch are independently learned and optimized, the extracted feature maps can exhibit different activation distributions. Therefore, normalization is required to reduce distribution discrepancies and stabilize feature integration. The normalized feature maps are then fused through element-wise averaging, which effectively aggregates complementary information from multiple branches.

Experimental Results: The experimental results in Table 1 evaluate the efficiency of the proposed method based on ResNet-18/50. Since extending the original architecture with a parallel structure increases computational cost, the initial channel expansion of ResNet-18 is reduced from 64 to 36. For ResNet-50, the initial channel expansion is reduced to 38, and the block configuration is modified from (3, 4, 6, 3) to (2, 3, 4, 3). The number of branches at each stage is set to 4. Although the overall network depth and channel dimensionality are reduced, the proposed feature diversity strategy improves representation capability and leads to performance gains. Specifically, ResNet-18 achieves Top-1 accuracy improvements of 1.5% and 1.88% on the two datasets, respectively, while ResNet-50 achieves improvements of 2.4% and 2.82%,

respectively. Although FLOPs are maintained, the number of learnable parameters increases, which can be attributed to the increased kernel diversity introduced by independent convolutional layers.

Table 1. Accuracy comparison between the baseline and the proposed method

Model	CIFAR-100				Tiny ImageNet			
	#Params	FLOPs	Acc@1	Acc@5	#Params	FLOPs	Acc@1	Acc@5
ResNet-18 (Baseline)	11.22M	0.56G	78.63	94.38	11.27M	1.71G	64.40	83.60
ResNet-18 (Ours)	14.91M	0.56G	80.13	95.16	14.94M	1.72G	66.28	85.39
ResNet-50 (Baseline)	23.71M	1.31G	79.15	94.56	23.91M	4.01G	66.29	85.61
ResNet-50 (Ours)	45.06M	1.30G	81.55	95.84	45.18M	3.96G	69.11	86.56

Conclusion: Through experiments on the CIFAR-100 and Tiny ImageNet datasets, we demonstrated that the proposed feature diversity-driven design can be effectively applied to ResNet architectures. In particular, the results show that Top-1 accuracy improves even under reduced channel capacity and network depth, indicating that performance gains are not solely dependent on model scaling. In future work, we will extend the proposed method to a wider range of CNN architectures beyond ResNet.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant No. RS-2026-25468547).

References

1. Haruna, Yunusa, et al. "Exploring the synergies of hybrid convolutional neural network and Vision Transformer architectures for computer vision: A survey." *Engineering Applications of Artificial Intelligence* 144 (2025): 110057.
2. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
3. Dai, Jifeng, et al. "Deformable convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
4. Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. pmlr, 2015.
5. Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Explainable Optimization of Precision Tooling Design Using Physics-Informed Generative Adversarial Networks for Small Data Regimes

ByeongSeok Lee¹⁾, Yoosoo Oh²⁾

¹⁾ Department of Smart Convergence Systems Engineering, Daegu University

²⁾ School of Computer and Information Engineering, Daegu University

*E-mail: yoosoo.oh@daegu.ac.kr

Abstract: This study proposes a Conditional Generative Adversarial Network (CTGAN)-based table data augmentation framework that incorporates physical composition constraints of the domain to overcome the limitations of small data in manufacturing sites caused by high-cost prototyping and the resulting overfitting problem in machine learning prediction models. As a result of training the model by progressively augmenting actual measurement data of precision machining tooling designs from 134 cases to 1,300 cases, the coefficient of determination (R^2) for wear amount prediction, which was initially only 0.49, increased significantly to 0.91. In addition, the augmented data was verified to comply with physical laws using the SHAP (SHapley Additive exPlanations) technique, and an optimization module was successfully constructed to explore ultra-precise target values in reverse by combining the MADS algorithm.

Keywords : Small-batch manufacturing data, tabular data augmentation, physical constraints, conditional Generative Adversarial Networks, explainable artificial intelligence, process optimization

1. Introduction: Although the smartification of manufacturing processes is currently underway, small and medium-sized enterprises (SMEs) engaged in multi-product, small-batch production, as well as the field of precision machining tooling design, face the limitation of "small data volumes," where collected data amounts to only tens to hundreds of records.

In this environment, general machine learning suffers from overfitting, which involves learning even The noise in the training data, leading to a problem where the coefficient of determination (R^2) for Wear prediction drops below 50%. To address this, this study proposes a method to maximize the quality of virtual data by introducing a deep learning-based CTGAN instead of simple oversampling, and by hard-filtering a physical constraint that "the sum of all chemical composition ratios is 100%."

Physical Constraint-based Data Augmentation & Process Optimization Architecture

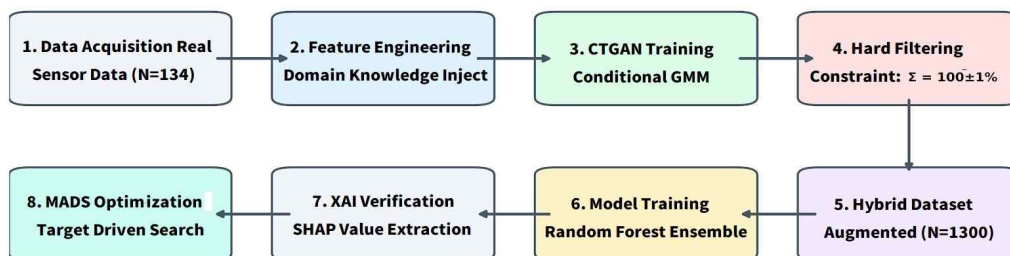


Fig. 1. System architecture diagram. Diagram of an 8-step process leading from initial raw data acquisition (1) to feature engineering (2), CTGAN augmentation (3), physical constraint hard filtering (4), hybrid data construction (5), ensemble model training (6), XAI validation (7), and MADS optimization (8).

2. Research Methods: The framework of this study utilizes CTGAN based on Gaussian Mixture Models (GMM) to reliably augment data with asymmetric or complex probability distributions.

A total of 1,300 hybrid datasets are constructed by passing them through a hard filtering engine that selects only those data where the total sum satisfies $100 \pm 1\%$.

In particular, 5-fold cross-validation is performed to prevent data leakage, and a Random Forest ensemble model is trained using an 'Augment-after-split' strategy, which applies augmentation only to the training data and consists solely of the original data for the test data.

3. Experimental Results: As a result of benchmarking prediction performance while expanding the data augmentation scale from 134 original samples to 1,300, the R^2 confidence for wear amount surged from 0.491 to 0.91, and the RMSE, a measure of the error rate, decreased dramatically from 0.015 to 0.002

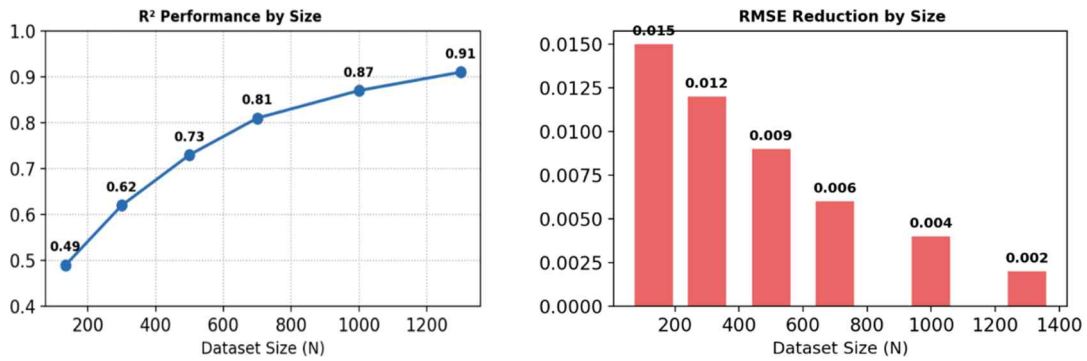


Fig. 2. Quantitative performance benchmarking: (a) R^2 performance increase and (b) RMSE reduction according to the dataset size expansion ($N=134$ to 1300).

4. Explainability and Target-Driven Optimization: Beyond the substantial improvements in quantitative metrics (R^2 and RMSE), it is paramount to verify that the ensemble model did not merely memorize the statistical noise of the synthetic data. As illustrated in Fig. 3, the SHAP (SHapley Additive exPlanations) analysis reveals a profound transition in feature importance. In the initial small data environment ($N=134$), the contribution ranking of variables was ambiguous and lacked logical physical grounding. Conversely, in the augmented hybrid model ($N=1300$), the importance of core physical features—such as Concentration and hp Index—became overwhelmingly dominant. This objectively verifies that the proposed domain-constrained augmentation instilled the underlying physical relationships of the machining domain into the predictive model.

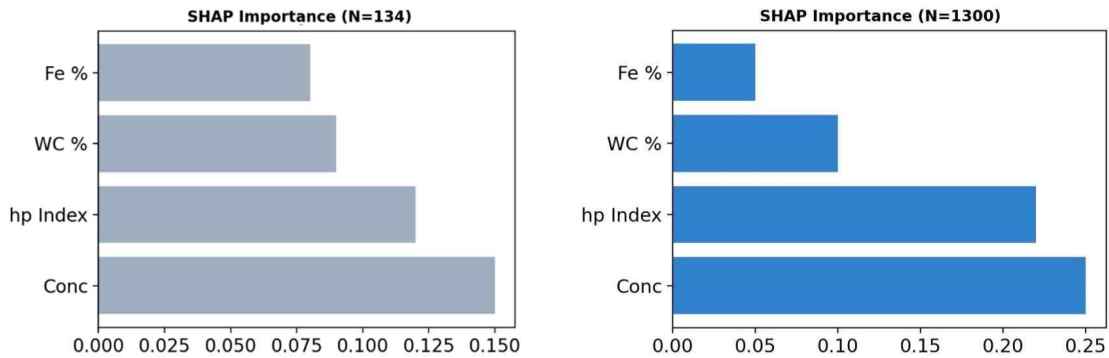


Fig. 3. Verification of physical fidelity: SHAP feature importance transition from the initial Small data ($N=134$) to the augmented hybrid data ($N=1300$).

Building upon this physically validated and highly reliable model, a reverse-search optimization module was embedded using the Mesh Adaptive Direct Search (MADS) algorithm. When tasked with finding an optimal recipe for an ultra-precision target wear of 0.020mm or less, the optimization simulation successfully converged to the exact target within merely 40 iterations, as shown in Fig. 4. This outcome highlights the demonstrates practical applicability of the proposed framework, demonstrating its capability to replace the costly, days-long trial-and-error processes of skilled workers with computational simulations

that take only seconds.

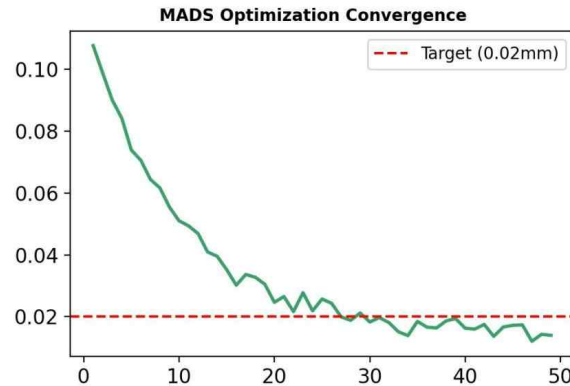


Fig. 4. MADS optimization simulation trajectory converging efficiently to the ultra-precision target wear (0.020mm) within 40 iterations.

5. Conclusion: The "Small Data" bottleneck in manufacturing inherently induces overfitting and heavily hinders practical AI application. This study successfully overcame this by hard-coding domain-specific physical constraints within a CTGAN generation framework instead of simple statistical replication. By expanding the initial 134 records to 1,300, the model's predictive reliability (R^2) visually and drastically improved from 0.49 to 0.91. XAI analysis using SHAP effectively verified that the augmented virtual data correctly instilled underlying physical mechanics into the predictive model. This robust pipeline provides a reliable academic guideline applicable to broader SME manufacturing sectors that face high data collection costs.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by Korean Government [Ministry of Science and ICT (MSIT)] under Grant RS-2026-25482392.

References

1. Breiman, L., "Random Forests," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.
2. Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K., "Modeling Tabular Data Using Conditional GAN," *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 32, pp. 7338–7348, 2019.
3. Malkin, S., and Guo, C., *Grinding Technology: Theory and Applications of Machining with Abrasives*, 2nd ed., Industrial Press, New York, 2008.
4. Lundberg, S. M., and Lee, S.-I., "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 4765–4774, 2017.
5. Tönshoff, H. K., Peters, J., Inasaki, I., and Paul, T., "Modelling and Simulation of Grinding Processes," *CIRP Annals – Manufacturing Technology*, Vol. 41, No. 2, pp. 677–688, 1992.

HyperCup: A Mobile Coffee Cupping Platform

Seonghee Min¹ *)

^{1.)} Cuppingpost Roasters, Gyeongsan-si, Gyeongsangbuk-do, 38426, Korea

*E-mail: msh9584@gmail.com

Abstract: In this paper we propose the HyperCup, a mobile-optimized digital coffee cupping platform. The HyperCup features an adaptive user interface consisting of easy mode for rapid logging and expert mode for precise evaluations based on the SCA standard protocol. To resolve the ambiguity of unstructured sensory data, the platform integrates a smart flavor tag recommendation system utilizing a controlled vocabulary, successfully achieving data standardization. This system establishes a high-quality structured database, serving as a foundational asset for future machine learning-driven quality prediction modeling.

Keywords : Digital Transformation, Adaptive UI, Intelligent Interaction, Coffee Cupping

Introduction: The coffee cupping is a standardized sensory analysis procedure for objectively evaluating and identifying the unique flavor, acidity, body, and defects of coffee beans [1, 2]. Moreover, handwritten notes filled with subjective text lead to data fragmentation due to inconsistent expressions, imposing severe limitations on data formatting for future statistical analysis via data mining or computational approaches [3]. In this paper, we propose the HyperCup, a digital cupping platform optimized for mobile environments that digitizes the conventional paper-based coffee cupping recording method. To resolve the complexity of existing cupping forms, the proposed system introduces two customized recording methods: easy mode and expert mode. Furthermore, the proposed system standardizes and records sensory data through a flavor tag recommendation system. To resolve the ambiguity and inconsistency of expressions that occur during coffee flavor recording, we implemented a system that automatically recommends hundreds of flavor tags in real-time upon text input. This lowers the barrier to entry for systematic cupping recording and contributes to the digital transformation of the sensory recording culture within the coffee industry.

Implementation: This section details the frontend architecture of the proposed system and the specific implementation mechanisms of its core functions. The proposed mobile application was implemented in Flutter, a cross-platform framework based on the Dart language. To maximize user experience and ensure the reliability of sensory records under the constraints of mobile devices, the following designs and algorithms were applied.

1. Adaptive UI Implementation: Traditional SCA cupping forms are visually complex, creating barriers for beginners and hindering rapid field evaluations. To address this, the proposed system introduces a dualized adaptive UI with "Easy" and "Expert" modes. Users can instantly switch between these modes during an active session using a floating action button.

Easy Mode: Returns the holistic single score as is, which the user intuitively adjusts via a numerical control knob provided on the screen. In this mode, the detailed attribute evaluation layout is completely removed from the viewport; only the coffee name, the basic flavor memo field, and a single total score control dial widget are exposed. This minimizes cognitive overload in user experience during field cupping scenarios and commercial roasting houses, which require high-speed identification of multiple samples and total score-centric recording.

Expert Mode: Returns totalScore, the sum of scores allocated to individual SCA attributes, automatically calculated by an internal algorithm. Individual attribute card blocks unfold for inputting the seven detailed attributes defined in the standard SCA protocol (Fragrance/Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Overall), and a 5-scale checkbox grid is additionally generated to quantify the status of the sample's Uniformity, Clean Cup, and Sweetness. In the expert mode, the action of tapping an individual

attribute block is designed as an asynchronous flow, generating a cupertino-style bottom sheet without interrupting the main application flow. When the user edits detailed scores and attribute-specific notes and closes the modal, the parent widget's state is updated, and the total summated score is immediately reflected on the screen.

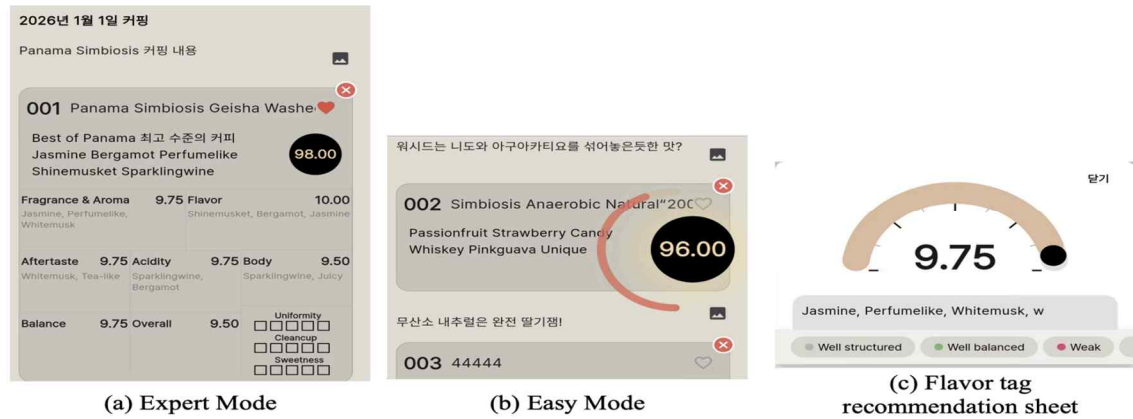


Fig. 1. Screenshot of the implementation system

2. Flavor Tag Recommendation System: To satisfy data standardization requirements, this platform implemented a smart flavor tag recommendation system that actively detects inputs and recommends a standard classification system based on the coffee flavor wheel in real-time. Upon input of the extracted word, an immediate string-matching operation is executed against a predefined dataset. Partial matches are filtered in real-time via prefix matching techniques. To implement the matched flavor options in a floating format on the topmost screen layer. When the user selects a specific recommended tag card, the process substitutes it with the completed string through a computational formula that removes overlapping substrings. Through these real-time corrective steps, the final evaluation database forms a high-quality, statistical coffee quality data.

Conclusion: In this study, to overcome the analog limitations of the cupping recording method, we proposed and implemented HyperCup, a digital cupping platform optimized for mobile environments. By introducing a dualized adaptive UI, the proposed system demonstrated high usability that satisfies both the rapid evaluation of multiple samples and precise recording. Furthermore, by designing a flavor tag recommendation, we successfully constructed a standardization framework for sensory data.

In conclusion, the proposed HyperCup system holds academic and industrial value as a practical, general-purpose platform capable of driving a successful digital transformation (DX) in the field of coffee tasting. High-quality, structured sensory data accumulated through this system can be utilized as a reliable foundational asset for future flavor analysis and quality prediction modeling using data mining.

References

1. Specialty Coffee Association, SCA Arabica Cupping Form & Protocols, Specialty Coffee Association (2015).
2. Guinard, J. X., Lim, L. X., Elliott, B., & Cotter, A, Validation of the Coffee Cuality™ Method for the Expert Assessment of Coffee Sensory Quality. Foods, 15(4), p. 678 (2026).
3. Sunarharum, W. B., Williams, D. J., & Smyth, H. E., Complexity of coffee flavor: A compositional and sensory perspective. Food research international, 62, p. 315-325 (2014).

Design of a Fine-Grained Classification Algorithm for Specialty Green Beans

Seonghee Min¹ *)

^{1,)} *Cuppingpost Roasters, Gyeongsan-si, Gyeongsangbuk-do, 38426, Korea*

^{*}*E-mail: msh9584@gmail.com*

Abstract: Accurately classifying specialty green coffee beans is challenging due to minute morphological differences and severe data constraints, which often cause standard models to overfit. This paper proposes a deep learning framework for fine-grained classification in extremely low-data environments. We utilize a customized preprocessing pipeline to mitigate data scarcity while preserving subtle, variety-specific traits. Furthermore, we employ a few-shot, contrastive learning architecture using EfficientNet-B0, prompting the model to learn morphological distances rather than memorizing labels. Finally, integrating Grad-CAM (XAI) visualizes decision-making regions, validating the model's alignment with established coffee taxonomy.

Keywords : Fine-Grained Classification, Explainable AI, Few-Shot Learning

Introduction: With the specialty coffee market's growth, accurately classifying high-value green coffee beans—such as Geisha, Bourbon, and Meadorado—has become critical [1]. However, these beans exhibit minute morphological differences that are difficult to distinguish visually. Furthermore, real-world constraints limit the large-scale image data required for deep learning. Consequently, general-purpose classifiers easily overfit to background noise in these extremely low-data environments [2].

To overcome this, we propose the design of a fine-grained classification algorithm. First, the proposed classification algorithm has a preprocessing pipeline combining spatial augmentations with restricted brightness and contrast adjustments to preserve variety-specific color profiles. Second, it is a few-shot [3, 4], contrastive learning architecture using EfficientNet-B0 [5], designed to induce the model to learn morphological distances rather than memorizing labels. Finally, we propose an eXplainable AI (XAI) framework integrating Grad-CAM[6] to verify that the model's visual feature extraction aligns with established coffee taxonomy criteria.

Design of the proposed fine-grained classification algorithm: Because raw green bean images lack fixed orientation, our proposed preprocessing pipeline is designed to apply random rotations and flips to expand training data while maintaining geometric integrity. To prevent degrading critical color nuances in specialty varieties, aggressive color jittering is strictly avoided; instead, the design includes limited brightness and contrast adjustments to ensure robustness against real-world lighting variations.

The preprocessed data is then proposed to be processed using EfficientNet-B0. To prevent the model from overfitting to ground-truth labels in a data-constrained environment, we propose a metric learning architecture based on few-shot and contrastive learning. The designed loss function calculates morphological differences between images, clustering same-variety samples in the latent space while separating different varieties.

This distance-based structure is proposed to maximize fine-grained classification performance. While these varieties share similar contours, they possess minute differences in center cut curvature, surface texture, and end sharpness. The proposed contrastive learning approach suppresses irrelevant information, concentrating network weights exclusively on these fine structural differences to enable reliable identification.

Finally, we propose integrating Grad-CAM to visually verify the model's predictions. By generating heatmaps of critical regions (e.g., S-shaped center cuts), this post-processing design ensures the model's decision-making aligns with expert visual discrimination criteria in coffee taxonomy, rather than mere background noise biases.

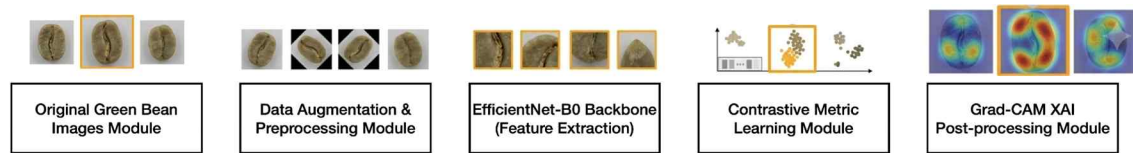


Fig. 1. A diagram of proposed system

Conclusion: In this paper proposed the design of a deep learning framework capable of performing highly accurate, fine-grained classification of high-value specialty green coffee bean varieties, including Geisha, Bourbon, and Meorado, even in environments where training data collection is extremely limited. The issue of data scarcity was effectively mitigated through spatial data augmentation reflecting the physical characteristics of green beans, coupled with restricted color calibration. Furthermore, the introduction of a metric learning model based on contrastive learning induced the network to focus on the fine morphological distances between varieties rather than background noise. Furthermore, the system was configured to incorporate a Grad-CAM-based eXplainable AI (XAI) post-processing step, facilitating cross-validation to ensure that the model's visual reasoning aligns with established criteria in coffee variety taxonomy. In future work, we plan to implement the algorithm designed in this paper and empirically validate the classification performance of the proposed pipeline through quantitative evaluation experiments.

References

1. De Oliveira, E. M., Leme, D. S., Barbosa, B. H. G., Rodarte, M. P., & Pereira, R. G. F. A., A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering*, 171, pp. 40-47 (2016).
2. Shorten, C., & Khoshgoftaar, T. R., A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1-48 (2019).
3. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G., A simple framework for contrastive learning of visual representations. *In International conference on machine learning (ICML)*, pp. 1597-1607 (2020).
4. Snell, J., Swersky, K., & Zemel, R., Prototypical networks for few-shot learning. *Advances in neural information processing systems (NeurIPS)*, 30 (2017).
5. Tan, M., & Le, Q., EfficientNet: Rethinking model scaling for convolutional neural networks, *Proceedings of the 36th International Conference on Machine Learning*, PMLR 97:6105-6114 (2019).
6. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D., Grad-cam: Visual explanations from deep networks via gradient-based localization. *In Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 618-626 (2017).

SoilSeg: Severity-Aware Lens Soiling Segmentation for Camera Reliability in Autonomous Driving

Muhammad Tayyab Shehzad¹⁾, Gyu Sang Choi^{*}

¹⁾ Dept. of Information and Communication Engineering, Yeungnam University, Gyeongsan-si, Korea

¹E-mail: mtayyabshehzad@yu.ac.kr

^{*}E-mail: castchoi@yu.ac.kr

Abstract: Lens soiling is a practical reliability challenge for camera-based autonomous driving. Dirt, water, and mud on externally mounted fisheye lenses can obscure safety-relevant scene regions while the camera remains operational. Existing camera-soiling methods mainly focus on detecting or localizing contaminated regions. However, the connection between severity-aware soiling maps and image-level camera reliability remains underexplored. To address this gap, we propose SoilSeg, a severity-aware lens-soiling segmentation framework for fisheye driving cameras. SoilSeg uses a SegFormer-B5 backbone to predict four pixel-level severity classes: clean, transparent, semi-transparent, and opaque. The predicted severity mask is then converted into an interpretable reliability score using class-area ratios and ordinal severity weights. Experiments on the WoodScape soiling dataset show that SoilSeg achieves 0.8171 mIoU, 0.9255 pixel accuracy, and 0.0171 reliability MAE, outperforming U-Net, FCN, DeepLabV3, and UPerNet. These results show that severity-aware segmentation can support dense soiling localization and image-level severity-derived camera-health estimation.

Keywords : Autonomous driving, camera soiling, semantic segmentation, camera reliability

Introduction: Fisheye cameras are widely used in autonomous driving because they provide a wide field of view for parking assistance, surround-view monitoring, lane understanding, and nearby object perception. However, externally mounted cameras are directly exposed to dirt, water, mud, dust, and other lens contaminants. When a lens is soiled, the camera may still produce an image, but safety-relevant regions can be blurred, attenuated, or fully occluded. WoodScape provides a multi-camera fisheye benchmark with soiling annotations, making it suitable for studying this reliability problem [1].

Existing camera-soiling studies have mainly focused on detecting contaminated regions in automotive surround-view cameras [2]. However, dense soiling localization alone does not directly indicate how strongly the current camera view is degraded. In this paper, we propose SoilSeg, a severity-aware lens-soiling segmentation framework for fisheye driving cameras. Instead of treating soiling as a binary or purely local detection problem, SoilSeg models lens contamination as an ordinal four-class severity map covering clean, transparent, semi-transparent, and opaque regions. The predicted mask is summarized using class-area ratios and ordinal severity weights to obtain a compact image-level reliability estimate. This formulation complements dense soiling localization with an interpretable camera-health cue, without requiring a separate reliability regression head. We evaluate SoilSeg against U-Net, FCN, DeepLabV3, and UPerNet baselines [4]–[7], showing improved segmentation accuracy and lower severity-derived reliability error.

Methodology: Fig. 1 summarizes the proposed SoilSeg pipeline. Given a fisheye RGB image $X \in \mathbb{R}^{H \times W \times 3}$, SoilSeg predicts a dense four-class severity mask $\hat{Y} \in \{0,1,2,3\}^{H \times W}$, representing clean, transparent, semi-transparent, and opaque lens regions. The network f_θ produces logits $Z = f_\theta(X) \in \mathbb{R}^{H \times W \times 4}$, and pixel labels are obtained by softmax classification:

$$P_{i,j,c} = \frac{\exp(Z_{i,j,c})}{\sum_{k=0}^3 \exp(Z_{i,j,k})}, \quad \hat{Y}_{i,j} = \arg \max_{c \in \{0,1,2,3\}} P_{i,j,c}.$$

The input first passes through patch embedding and a hierarchical SegFormer-B5 encoder [3]. Multi-scale features are fused by an MLP decoder, and a four-class prediction head produces the severity mask. All images and masks are resized to 768×768 , with nearest-neighbor interpolation applied to masks; the model is trained using pixel-wise multi-class cross-entropy.

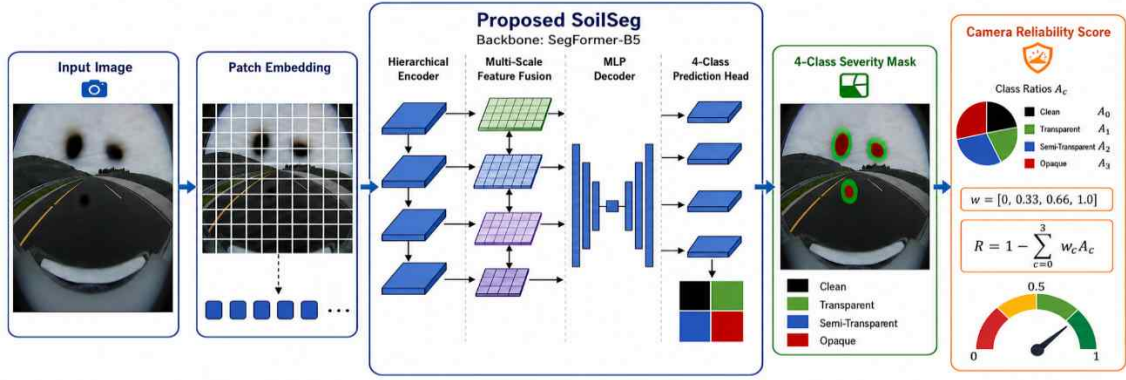


Fig. 1. SoilSeg for severity-aware soiling segmentation and camera reliability estimation.

The reliability score is computed analytically from the predicted mask, not learned by a separate regression head. For each severity class c , SoilSeg computes the class-area ratio A_c . Using ordinal weights $w = [0, 0.33, 0.66, 1.00]$ for clean, transparent, semi-transparent, and opaque regions, respectively, the camera reliability score is defined as:

$$A_c(\hat{Y}) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \mathbf{1}[\hat{Y}_{i,j} = c], \quad R(\hat{Y}) = 1 - \sum_{c=0}^3 w_c A_c(\hat{Y}).$$

A higher R indicates a healthier camera view. Since WoodScape does not provide independent camera-health labels, the reference reliability score is computed from the annotated severity mask using the same weighting rule. Thus, the score represents a severity-derived reliability estimate that measures how well the predicted mask preserves image-level degradation composition. Segmentation is evaluated using mIoU and pixel accuracy, while reliability estimation is evaluated using:

$$\text{MAE}_R = \frac{1}{N} \sum_{n=1}^N |R(\hat{Y}^{(n)}) - R(Y^{(n)})|.$$

This protocol evaluates whether the predicted severity mask preserves both pixel-level soiling structure and image-level severity composition.

Experiments and Results: We evaluate SoilSeg on the WoodScape soiling dataset using four pixel-level severity classes: clean, transparent, semi-transparent, and opaque. The dataset contains 4,000 training image-mask pairs and 1,000 test pairs. All inputs are resized to 768×768 , and masks are resized using nearest-neighbor interpolation. SoilSeg is compared with U-Net [4], FCN [5], DeepLabV3 [6], and UPerNet [7] under the same split, label space, and metrics.

Table 1. Quantitative comparison on the WoodScape soiling test set.

Method	mIoU	Pixel Acc.	Clean IoU	Trans. IoU	Semi-trans. IoU	Opaque IoU	Rel. MAE ↓
U-Net	0.5488	0.7784	0.7994	0.2189	0.4088	0.7682	0.0526
FCN	0.7504	0.8927	0.9004	0.5826	0.6396	0.8792	0.0250
DeepLabV3	0.7531	0.8961	0.9022	0.5675	0.6509	0.8918	0.0222
UPerNet	0.8007	0.9190	0.9258	0.6598	0.7054	0.9117	0.0181
SoilSeg	0.8171	0.9255	0.9284	0.6929	0.7335	0.9137	0.0171

As shown in Table 1, SoilSeg achieves the best overall result, with 0.8171 mIoU, 0.9255 pixel accuracy, and 0.0171 reliability MAE. Compared with the strongest baseline, UPerNet [7], SoilSeg improves mIoU by 0.0164 and reduces reliability MAE from 0.0181 to 0.0171. For SoilSeg, class-wise IoUs are 0.9284, 0.6929, 0.7335, and 0.9137 for clean, transparent, semi-transparent, and opaque regions, respectively. The

lower reliability error indicates that the predicted masks preserve the image-level severity composition used for camera-health estimation.

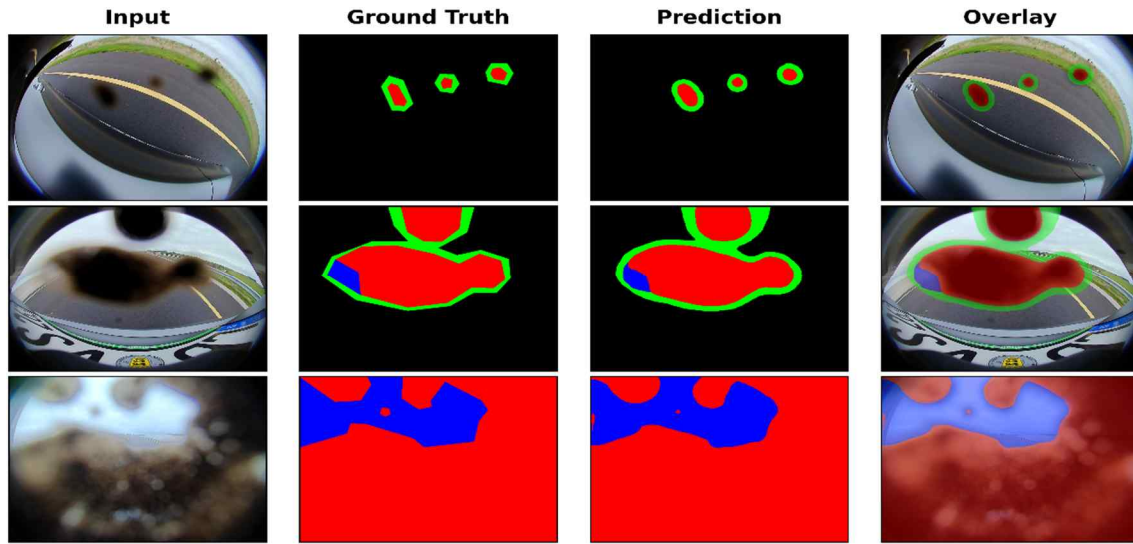


Fig. 2. Qualitative results with input, ground truth, prediction, and overlay.

The mean reliability error remains low across camera views, ranging from 0.0146 to 0.0190. Fig. 2 shows qualitative examples with input image, ground truth, prediction, and overlay. SoilSeg captures dominant opaque regions and surrounding lower-severity contamination. The main observed errors occur near gradual transitions between transparent and semi-transparent regions, where severity boundaries are visually ambiguous.

Conclusion and Future Work: This paper presented SoilSeg, a severity-aware lens-soiling segmentation framework for fisheye driving cameras. SoilSeg predicts four soiling classes and converts the predicted mask into an interpretable severity-derived reliability score using ordinal severity weights. Experiments on the WoodScape soiling dataset show improved segmentation accuracy and lower reliability error compared with representative baselines. The results indicate that severity-aware segmentation can support dense contamination localization and image-level degradation estimation. Future work will calibrate this reliability score with downstream perception tasks such as object detection, lane detection, and closed-loop driving performance.

Acknowledgment

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) by the Ministry of Education (NRF-2021R1A6A1A03039493) and by the BK21 FOUR of the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2120251015517).

References

1. S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D. O'Dea, M. Uříčář, S. Milz, M. Simon, K. Amende, C. Witt, H. Rashed, S. Chennupati, S. Nayak, S. Mansoor, X. Perrotton, and P. Pérez, "WoodScape: A multi-task, multi-camera fisheye dataset for autonomous driving," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 9308–9318.
2. M. Uříčář, P. Krížek, G. Sistu, and S. Yogamani, "SoilingNet: Soiling detection on automotive surround-view cameras," in Proc. IEEE Intell. Transp. Syst. Conf. (ITSC), 2019, pp. 67–72.
3. E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient

- design for semantic segmentation with transformers,” in *Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12077–12090.
4. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
 5. J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.
 6. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv:1706.05587*, 2017.
 7. T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, “Unified perceptual parsing for scene understanding,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 418–434.

Multi-Rate Scheduling and Atomic-Step Decomposition for MLP-Mixer Inference on Bare-Metal Embedded Systems

Wonhee Kim¹⁾, Byenogkeun Kwon²⁾, Suhyeon Kim²⁾, Young Ok Lee^{3*)} and Young Seop Son^{2*)}

¹⁾ School of Energy Systems Engineering, Chung-Ang University, Seoul 06974, Korea

²⁾ Graduated School of Data Science, Kyungpook National University, Daegu 41566, Korea.

³⁾ Department of Mechatronics Engineering, Daelim University, Anyang-si, Gyeonggi-do 13916, Korea

*E-mail: yolee@daelim.ac.kr, ys.son@knu.ac.kr

Abstract: — To enable real-time MLP-Mixer inference on resource-constrained embedded systems, this paper introduces a multi-rate scheduling framework combined with an atomic-step decomposition model. The proposed scheduling decouples an interrupt-driven fast task for time-critical operations from a polling-based slow task for computationally intensive operations, ensuring strict period separation and minimal ISR overhead. By partitioning a single inference pass into tightly bounded atomic steps mapped to individual scheduler ticks, the model supports distributed execution and rapid state resumption. Evaluated on an Infineon AURIX TC337 platform, the framework minimizes fast-task ISR CPU utilization, thereby guaranteeing the real-time reliability of the main control loop in bare-metal applications.

Keywords : single, super-script fonts, text

Introduction: Modern embedded control systems must handle heterogeneous workloads with diverse timing constraints, where time-series signal processing using all-MLP architectures [1]. However, deploying these models on resource-constrained microcontrollers remains challenging because a single, monolithic inference pass requires massive multiply-accumulate (MAC) operations that exceed the computational budget of a single ISR period or polling slot [2]. This baseline mismatch leads to deadline violations of adjacent real-time tasks. Consequently, the primary challenge shifts from memory footprint constraints to distributing the inference workload across the time axis without degrading the determinism of high-priority tasks [3]. While conventional multi-rate scheduling separates short-period critical tasks from longer-period workloads, a dedicated execution framework is required to apply this deterministic separation to large neural network inferences.

To address the limitations of conventional methods, this paper presents a multi-rate scheduling framework that ensures deterministic execution of time-series MLP-Mixer models on embedded control systems. The proposed framework integrates an atomic-step decomposition model to enable distributed inference execution. By separating an interrupt-driven fast task from a polling-based slow task, the scheduling scheme minimizes the execution time of the interrupt service routine (ISR) and strictly enforces period separation between the two tasks. Specifically, time-critical short operations run within the fast-task ISR, while computationally intensive operations are processed in the slow-task polling context. To map MLP-Mixer inference onto this scheduling structure, a single inference pass is partitioned into a finite set of atomic steps, each aligned with a single scheduler tick. Because each step is defined as an atomic operation acting on a single element, its execution time remains tightly bounded, allowing suspended operations to resume at the next tick with minimal state overhead. The framework was evaluated on an Infineon AURIX TC337 platform using an MLP-Mixer configuration with $F = 10$ features, a sequence length of $S = 100$, and a hidden dimension of $H = 64$. Experimental results demonstrate that the proposed method minimizes fast-task ISR CPU utilization, ensuring the reliable operation of the slow polling task that governs the main control period.

Methodology: Figure 1 shows the temporal behavior of the multi-rate scheduling structure implemented on the Infineon AURIX TC337. The fast timer interrupt executes the fast task periodically, while accumulated ticks trigger the slow task. Running the slow task within the main-loop polling context minimizes interrupt execution time and isolates the two tasks structurally.

As illustrated in Figure 2, the input time-series x passes sequentially through the ① Time-mixing and ②

Feature-mixing blocks. Both blocks follow the same pipeline: PREP_COPY, LINEAR1 (INIT/MAC/STORE), GELU, LINEAR2 (INIT/MAC/STORE), RESID, LayerNorm (MEAN/VAR/APPLY), and WRITEBACK. The processed data then goes through the ③ Pooling block and the ④ Output Head, concluding at the DONE (Terminal) state. The color of each step indicates its specific operation type, such as data movement, initialization, arithmetic (MAC), activation, storage, statistical accumulation, finalization, or termination.

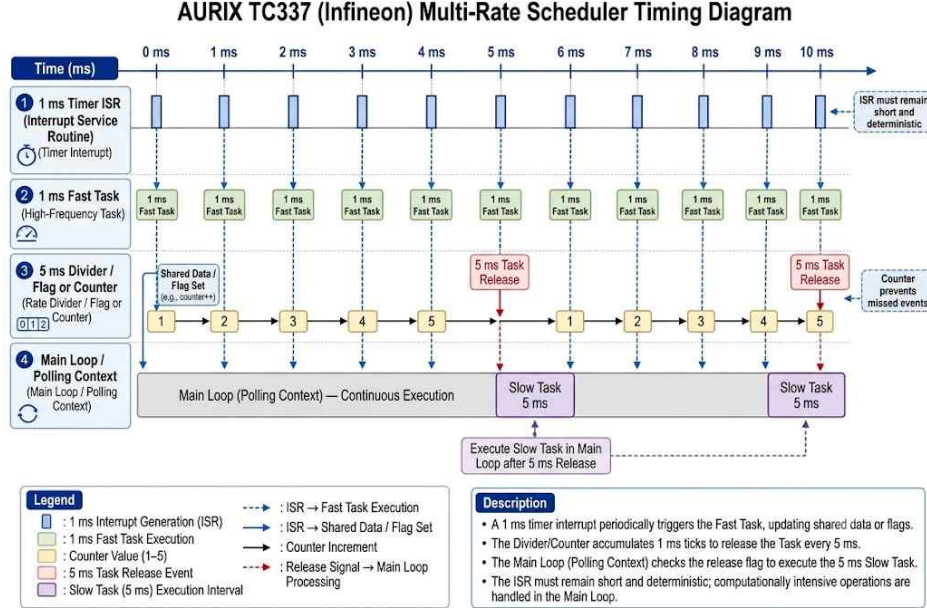


Fig. 1. Timing diagram of the multi-rate scheduling structure on the Infineon AURIX TC337

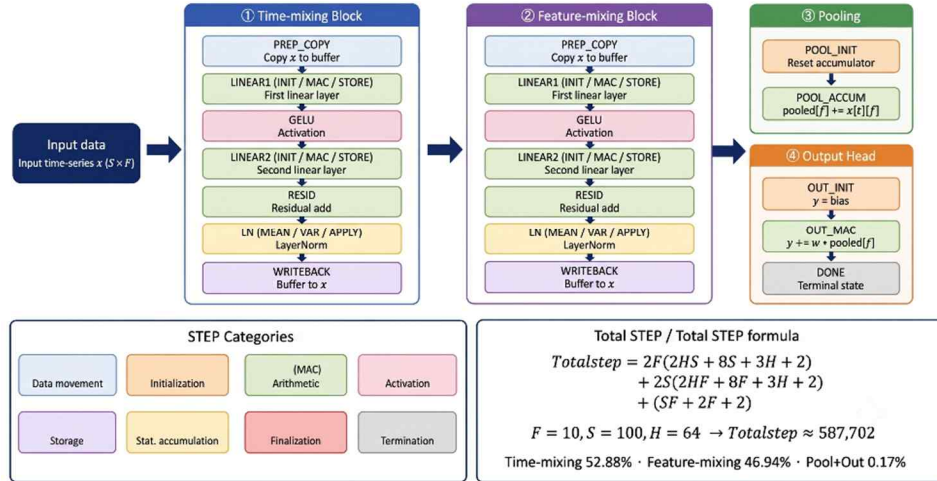


Fig. 2. Atomic-step decomposition of inference: unified block diagram of the model architecture and the step flow

Acknowledgment

This work was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE)(No.P0023721, Support to strengthen the ecosystem of service-customized automotive parts industry).

References

1. I. Tolstikhin, et. al., "MLP-Mixer: An all-MLP architecture for vision," in Proc. Advances in Neural Information Processing Systems, 2021.
2. K. Xiang et. al., AtomThink: Multimodal Slow Thinking With Atomic Step Reasoning. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 48, no. 5, pp. 5725-5741, 2026
3. C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," Journal of the ACM, vol. 20, no. 1, pp. 46-61, Jan. 1973.

Functional and Performance Evaluation Method for Core Components of Wide-Area Low-Floor Electric Buses

Sung-hun Kim¹⁾, Seong-Wook Cho²⁾, Oh-Hun Kwon²⁾, Jang-Kyu Yun²⁾, Soo-Young Suk²⁾

^{1,2)} Electronic Components Research Center, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38463, Korea

**E-mail: shkim90@gitc.or.kr*

Abstract: With the growing emphasis on carbon neutrality and eco-friendly public transportation, the adoption of electric buses has been rapidly increasing. In particular, the demand for wide-area low-floor electric buses is expanding due to their superior accessibility and operational efficiency. To ensure the safety and reliability of electric buses, systematic performance verification of core components, such as battery packs, power conversion modules, and driver information display systems, is essential. However, existing evaluation approaches mainly focus on individual components, while integrated performance assessment considering actual vehicle operating environments remains insufficient. Therefore, it is necessary to develop performance indicators and evaluation methodologies that reflect the characteristics of core components and to establish an integrated functional verification framework.

Keywords : Wide-Area Low-Floor Electric Bus, Battery Pack, Power Conversion Module, Meta-Cluster, Performance Evaluation, Integrated Testbed

Introduction: The electrification of commercial vehicles has accelerated in response to global carbon-neutral policies and increasing demand for sustainable transportation systems. Wide-area low-floor electric buses require highly reliable operation because they are used in public transportation services with long operating hours and frequent passenger usage. Core components such as battery packs, power conversion modules, and meta-clusters significantly affect vehicle safety, energy efficiency, and operational stability. Nevertheless, current evaluation methods are generally limited to individual component testing and do not adequately assess integrated functionality under real operating conditions. Therefore, a systematic evaluation framework that considers both component-level performance and system-level interoperability is required.

Research Methods: This study aims to establish specialized performance indicators and evaluation procedures for the core components of wide-area low-floor electric buses. For battery packs, performance indicators are defined based on charge/discharge characteristics, energy efficiency, and state information accuracy. For power conversion modules, evaluation items include power conversion efficiency, voltage stability, and communication reliability. For meta-clusters, information display accuracy, data transmission capability, and user interface performance are considered. Based on the operational characteristics and functional requirements of each component, quantitative evaluation criteria and standardized test procedures are developed. Furthermore, an integrated evaluation framework is proposed to assess the interoperability and functional linkage among the core components.

Experimental Results: To validate the proposed evaluation methodology, an integrated functional testbed consisting of a battery pack, power conversion module, and meta-cluster will be developed. The testbed will emulate actual vehicle operating environments and provide communication, monitoring, and data exchange functions among the core components. Functional integration tests and performance verification tests will be conducted using the developed testbed. The evaluation results are expected to provide quantitative assessments of interface compatibility, communication reliability, and operational performance. Figure 1 presents an example configuration of the system integration functional evaluation testbed.

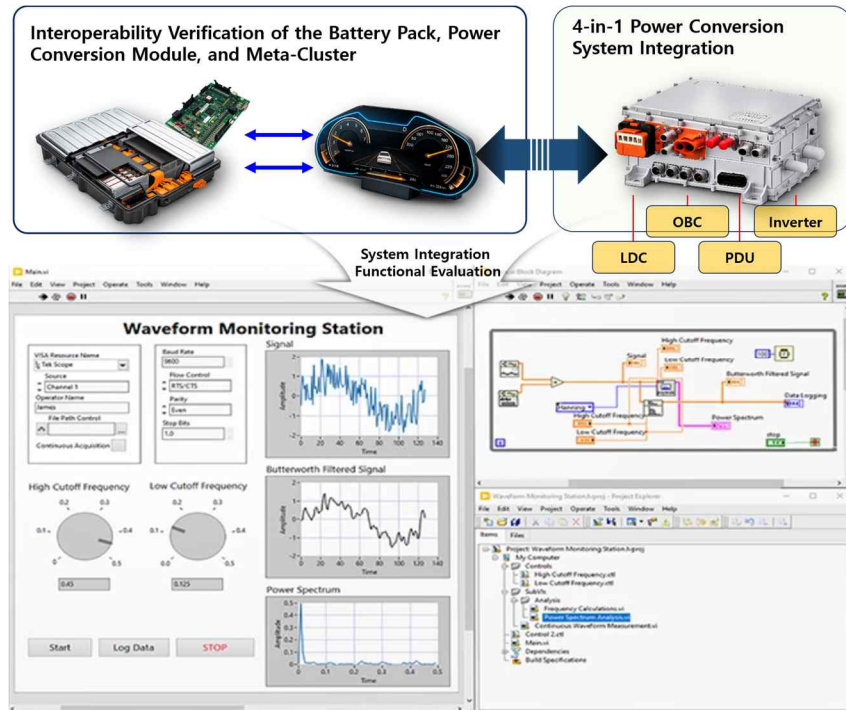


Fig. 1. Example Configuration of the System Integration Functional Evaluation Testbed

Conclusion: This study proposes a functional and performance evaluation framework for the core components of wide-area low-floor electric buses. Specialized performance indicators and evaluation procedures are established for battery packs, power conversion modules, and meta-clusters. An integrated functional testbed will be developed to verify component interoperability and system reliability. The proposed methodology is expected to contribute to the standardization of performance evaluation procedures and the enhancement of reliability for electric bus core components.

Acknowledgment

This work is supported by the KIAT(Korea Institute for Advancement of Technology) grant funded by the Ministry of Trade, Industry and Energy (No.P0031741).

References

1. ISO 26262, Road Vehicles – Functional Safety.
2. ISO 6469-1, Electrically Propelled Road Vehicles – Safety Specifications.
3. SAE J1939, Recommended Practice for Serial Control and Communications Vehicle Network.
4. IEC 61851, Electric Vehicle Conductive Charging System.
5. S. Thangavel, D. Mohanraj, T. Girijaprasanna, S. Raju, C. Dhanamjayulu, and S. M. Muyeen, "A Comprehensive Review on Electric Vehicle: Battery Management System, Charging Station, Traction Motors," IEEE Access, vol. 11, pp. 20994–21019, 2023.

Automatic Instance-Segmentation Labeling of Rice Cookers Using SAM3 Exemplar Matching

Jaehyun Park¹⁾, Seungtaek Kim¹⁾, Sunjong Lee¹⁾ and Youngjun Yoo^{1*)}

¹⁾ Korea Institute of Industrial Technology, Cheonan, 31056, Republic of Korea

*E-mail: youdalj@kitech.re.kr

Abstract: Pixel-level instance segmentation requires polygon labels that are normally drawn on every image by hand, which makes annotation the main bottleneck of dataset construction. We present an automatic labeling pipeline that, from only a few point clicks on a handful of images, labels an entire image folder by reusing the vision encoder of the Segment Anything Model 3 (SAM3) as a feature extractor and matching candidate masks to registered exemplars through cosine similarity and a geometric filter. Applied to a single-class rice-cooker set, a YOLOv8n-seg model trained purely on the auto-generated labels reaches a mask mAP@0.5 of 0.995 with a recall of 1.000. These results show that few-click automatic labeling can fully replace per-image manual annotation while producing training-grade segmentation labels.

Keywords: automatic labeling, instance segmentation, Segment Anything Model (SAM3), exemplar matching, YOLOv8

1. Introduction

Instance segmentation underpins many vision applications, yet it depends on pixel-accurate polygon labels whose manual creation is slow and costly. In industrial and home-appliance settings such as locating a rice cooker on a production or inspection line, datasets must often be rebuilt whenever products or viewpoints change, so manual relabeling becomes a recurring burden. Promptable foundation models in the Segment Anything line [1, 2, 3] can produce high-quality masks from simple prompts, which opens the possibility of removing the human from the per-image annotation loop. This work turns that possibility into a practical pipeline: a one-time, few-click setup is generalized to an entire image folder, so that pixel-level labels are generated automatically and used directly to train a segmentation model.

2. Research Methods

The pipeline has a one-time exemplar-registration step followed by fully automatic labeling. For registration, N images are sampled at random ($N = 3$ by default); the annotator left-clicks once on the rice cooker, and the SAM3 tracker in point-prompt mode returns a single-object mask. Each confirmed mask is stored as an exemplar pairing an appearance descriptor with a geometric descriptor. The appearance descriptor e_k of class k is the SAM3 dense feature map, mean-pooled over the mask region and L_2 -normalized; the geometric descriptor records the normalized area and centroid. The complement of the clicked region is stored automatically as a background (negative) exemplar. This is the only human interaction: as few as N point clicks replace manual polygon drawing on every image.

During automatic labeling, a regular $n \times n$ grid of point prompts ($n = 10$) is passed to the SAM3 tracker to propose candidate masks anywhere in the image; low-score and tiny masks are discarded and near-duplicates are removed by non-maximum suppression ($\text{IoU} > 0.7$). Each surviving mask m is embedded into the same feature space and assigned a class by nearest-exemplar cosine similarity, as defined in Eq. (1) and Eq. (2):

$$v_m = (1/|m_f|) \sum_{p \in m_f} F(p), \quad v_m \leftarrow v_m / \|v_m\|_2 \quad (1)$$

$$k^* = \arg\max_k \max_{e \in E_k} e^T v_m, \quad \text{accept iff } \max e^T v_m \geq \tau_{\text{sim}} \text{ and } k^* \neq \text{bg} \quad (2)$$

where F is the SAM3 dense feature map, m_f the mask resampled to the feature-map grid, E_k the exemplar set of class k , and $\tau_{\text{sim}} = 0.6$ the similarity threshold; as all vectors are unit-norm, the inner product equals

the cosine similarity. A geometric filter then rejects any accepted candidate whose area differs from every same-class exemplar by more than 10% of the exemplar area, or whose centroid differs by more than 0.15 of the image size. Surviving masks are refined by morphological closing, deduplicated by a containment rule, and converted to normalized polygons by contour extraction with Douglas–Peucker simplification. The polygons are written as YOLOv8-seg [4] labels; the folder is split 80/20 and a YOLOv8n-seg model is trained on them.

3. Experimental Results

The pipeline was applied to a single-class rice-cooker image set, and a YOLOv8n-seg model was trained for 100 epochs (640×640, batch 8) on the automatically generated labels. Fig. 1 shows representative auto-generated masks: the rice cooker is tightly segmented in every image without any manual polygon drawing. Table 1 reports the validation accuracy. The losses decrease steadily while every accuracy metric saturates near its ceiling within a few epochs and remains stable, reaching a mask $\text{mAP}@0.5$ of 0.995, an $\text{mAP}@0.5:0.95$ of 0.995, and a recall of 1.000. Because the task is single-class and the domain is controlled, such high scores are expected; the key finding is that labels produced from only N clicks are accurate and consistent enough to train a high-quality segmentation model end-to-end, fully replacing per-image manual annotation.

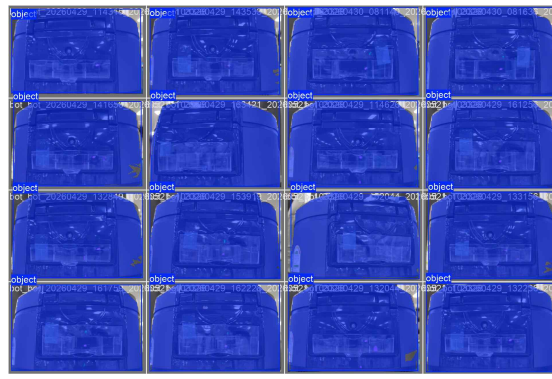


Fig. 1. SAM3 auto-generated instance-segmentation labels for rice cookers

Table 1. YOLOv8n-seg validation accuracy on the SAM3-generated labels (100 epochs, 640×640, batch 8; values rounded to three decimals).

Task	Precision	Recall	$\text{mAP}@0.5$	$\text{mAP}@0.5:0.95$
Detection (Box)	1.000	1.000	0.995	0.995
Segmentation (Mask)	1.000	1.000	0.995	0.995

4. Conclusion

We presented a few-click automatic labeling pipeline that reuses the SAM3 encoder and exemplar matching to annotate an entire image folder without per-image manual drawing. On a rice-cooker set, a YOLOv8n-seg model trained only on the auto-generated labels achieved a mask $\text{mAP}@0.5$ of 0.995 and a recall of 1.000, confirming training-grade label quality. Future work will extend the method to multi-class and cluttered scenes and quantify the annotation-time savings against manual labeling. To more rigorously assess generalization, we further plan to validate the pipeline across a wider range of object categories, backgrounds, and imaging conditions beyond the single controlled setting studied here.

Acknowledgment

This research was supported/partially supported by JK26030 and JH260007 from KITECH. We thank our colleagues from KITECH who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

References

1. A. Kirillov et al., “Segment Anything,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2023.
2. N. Ravi et al., “SAM 2: Segment Anything in Images and Videos,” arXiv:2408.00714, 2024.
3. N. Carion et al., “SAM 3: Segment Anything with Concepts,” arXiv:2511.16719, 2025.
4. G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLOv8,” 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>

Low-Complexity Two-Stage Operation FMCW Beam-Scanning Radar for Embedded Systems

YoungSeok Jin¹⁾, Eugin Hyun^{1*)}

¹⁾ IS³ Lab, Dept. of Mobility Technology, DGIST, Daegu, Korea

*E-mail: braham@dgist.ac.kr

Abstract: Radar sensors employing transmit (Tx) beam scanning methods suffer from high computational complexity because the received data for each Tx beam must be processed in three dimensions (range, velocity, and angle). To address this issue, this paper proposes an efficient two-stage radar operation and signal processing technique. In the first stage, a range-Doppler map is preferentially generated for each Tx beam to determine the presence of moving objects. In the second stage, receive (Rx) digital beamforming is selectively performed only on the valid Tx beams where a moving object has been detected. The proposed algorithm was implemented using a 24 GHz Frequency Modulated Continuous Wave(FMCW) radar front-end module with a 4Tx-4Rx antenna configuration and an STM32G474 Micro Controller Unit(MCU). Experimental results demonstrate that the proposed system significantly enhances both computational efficiency and detection accuracy.

Keywords : Radar, Beam Scanning Radar, Low Complexity Radar

In radar systems, utilizing a single transmit antenna can reduce memory usage and computational complexity, but it suffers from a limited field of view (FOV)[1]. To overcome this limitation, multiple-transmit-antenna-based beam scanning technology is widely utilized; however, this approach excessively consumes SRAM (Static Random-Access Memory) resources of embedded systems and increases the computational complexity.

Specifically, the radar approach that scans narrow multiple transmit beams using multi-transmit antennas exhibits superior robustness in clutter environments compared to methods that transmit signals over a wide angle simultaneously[2]. However, if a three-dimensional (3D) range-velocity-angle data map is generated individually for every transmit beam, the massive data volume poses a definitive limitation for real-time implementation on resource-constrained embedded systems.

To address this issue, this paper proposes an efficient two-stage radar operation technique. In the first stage, a range-velocity map is preferentially generated for each transmit beam, without applying digital beamforming to the signals acquired by multiple receive antennas. Instead, the presence of moving objects is primarily detected after deriving a range-velocity map with an improved signal-to-noise ratio (SNR) through non-coherent integration. In the second stage, receive digital beamforming is selectively applied only to the valid transmit beam regions where moving objects are detected, thereby estimating the precise angle of the objects.

The Radio Frequency(RF) front-end module used in this study is a 24 GHz Frequency Modulated Continuous Wave(FMCW) radar module with a 4-channel transmit and 4-channel receive antenna configuration, which was fabricated in a previous study [3]. The transmit beams are designed to scan from -40 to +40 at 10 intervals. Fig. 1 illustrates the transmit beam pattern of the radar front-end module, and Fig. 2 shows the overall architecture of the proposed radar system. For radar signal processing and control, a high-performance Micro Controller Unit (MCU), the STM32G474, was utilized as the embedded module[4].

Validation of the proposed technique was conducted by establishing scenarios where a moving target exists within a specific beam scanning area. When a target was detected, the target information (range, velocity, and angle) computed by the embedded system was transmitted to a PC, verifying the system performance via real-time plotting. The experimental results demonstrated that the proposed low-complexity beam-scanning radar system operates accurately and stably as intended.

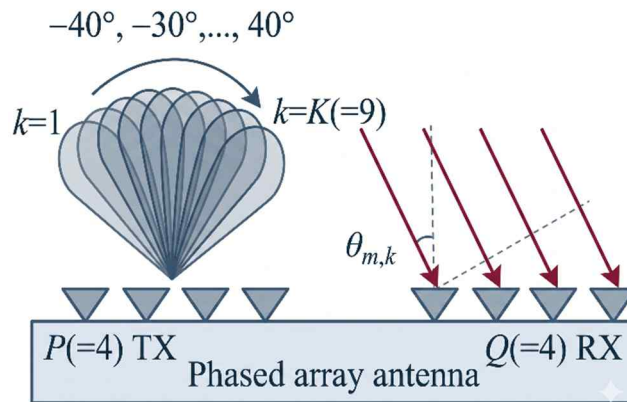


Fig. 1. Radar beam scanning structure diagram

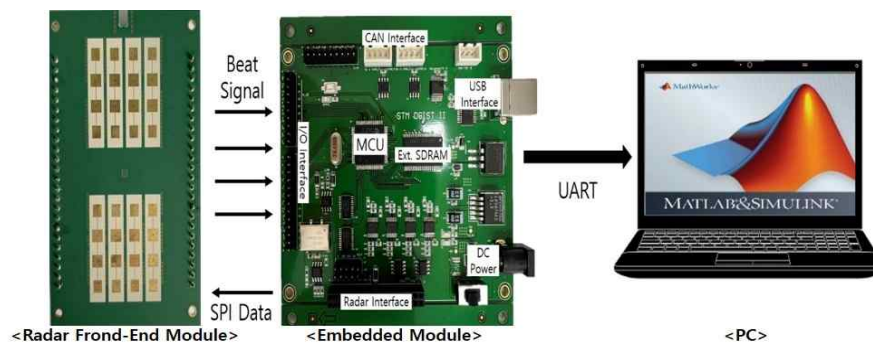


Fig. 2. Radar System diagram

Acknowledgment

This work was supported by the DGIST R&D Program funded by the Ministry of Science and ICT(No. 26-IT-01) and the Institute for Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-00138, Development of Intelligent Radar Platform Technology for Smart Environments).

References

1. J. Lee, J. Park, and S. Kim, "FMCW Radar Estimation Algorithm with High Resolution and Low Complexity Based on Reduced Search Area," *Sensors*, vol. 20, no. 19, p. 5567 (2020)
2. B. Kim, Y. Jin, J. Bae, and E. Hyun, "Two-Step Adaptive Target Detection Scheme Using Beam-Steering FMCW Radar Sensors," *Journal of Electromagnetic Engineering and Science*, vol. 24, no. 6, pp. 674-676, Nov. (2024)
3. G. Ko, S. Moon, S. Kim, J. Kim, and D. Baek, "Fully Integrated 24-GHz 1TX-2RX Transceiver for Compact FMCW Radar Applications," *Sensors*, vol. 24, no. 5, p. 1460, (2024).
4. STM32G474xB STM32G474xC STM32G474xE Datasheet - Arm Cortex-M4 32-bit MCU, STMicroelectronics, Geneva, Switzerland, Rev. 7, 2024. [Online]. Available: <https://www.st.com/resource/en/datasheet/stm32g474re.pdf> (accessed: Jun. 2, 2026).

A Standards- and BMS-Based Evaluation Framework for Classification and Safety Assessment of EOL EV Batteries

Seong-wook Cho¹⁾, Sung-Hun Kim¹⁾, Oh-hoon Kwon¹⁾, Jang-kyu Yun¹⁾ and Soo-young Suk^{2*)}

¹⁾ Electronic & Autonomous Systems Center, Future Mobility Research Division, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38542, Republic of Korea

²⁾ Future Mobility Research Division, Gyeongbuk Institute of IT Convergence Industry Technology (GITC), Gyeongsan 38542, Republic of Korea

*E-mail: sujo@gitc.or.kr

Abstract: The increasing adoption of electric vehicles (EVs) is expected to accelerate the generation of End-of-Life (EOL) EV batteries. To support safe reuse, remanufacturing, and recycling, a systematic method is required to evaluate the residual performance and safety condition of EOL batteries. This study proposes a standards- and Battery Management System (BMS)-based evaluation framework for the classification and safety assessment of EOL EV batteries. Major EV battery standards, including IEC 62660 series, ISO 12405-4, IEC 62619, and UL 2580, were selectively analyzed to derive key evaluation parameters. These parameters were mapped to BMS diagnostic data, such as State of Health (SOH), capacity retention, internal resistance, insulation resistance, voltage imbalance, and temperature history. The proposed framework supports the classification of EOL EV batteries into reuse, remanufacturing, and recycling pathways.

Keywords: EOL EV Battery, Battery Management System, Safety Assessment, Battery Classification, Battery Reuse

Introduction: The global transition toward electric mobility has rapidly increased the number of electric vehicles on the road. As EV batteries reach the end of their automotive service life, a large volume of EOL batteries is expected to emerge. Although EOL batteries may still retain usable capacity, their degradation level and safety condition must be carefully evaluated before reuse or remanufacturing. Therefore, a systematic evaluation framework is required to support safe battery circulation and second-life applications.

Research Methods: This study analyzed key standards directly related to EV battery performance and safety assessment. The IEC 62660 series was reviewed to identify cell-level performance, reliability, and safety requirements, while ISO 12405-4 was considered for battery pack and system-level performance evaluation. IEC 62619 and UL 2580 were analyzed to derive lithium-ion battery system safety requirements. Based on the standard analysis, BMS-related diagnostic parameters were selected and organized into performance and safety assessment categories.

Framework Derivation Results: The proposed framework classifies evaluation parameters into four domains: performance, degradation, safety, and utilization classification. Performance indicators such as SOH and capacity retention are used to estimate residual usability, while degradation indicators such as internal resistance and voltage imbalance are used to diagnose battery aging. Safety indicators, including insulation resistance and temperature history, are applied to assess electrical and thermal risks. Based on these assessment results, EOL EV batteries are classified into reuse, remanufacturing, or recycling pathways. Table 1 summarizes the core evaluation domains of the proposed framework.

Table 1. Core Evaluation Domains for EOL EV Battery Assessment

Domain	Evaluation Parameter	Assessment Purpose
Performance	SOH, capacity retention	Residual usability evaluation
Degradation	Internal resistance, voltage imbalance	Battery aging diagnosis
Safety	Insulation resistance, temperature history	Electrical and thermal safety assessment

Classification	Reuse, remanufacturing, recycling	Utilization pathway decision
----------------	-----------------------------------	------------------------------

Based on the derived evaluation parameters, the proposed framework consists of three main stages: BMS data acquisition, performance and safety assessment, and utilization classification. Fig. 1 shows the overall structure of the proposed standards- and BMS-based evaluation framework.

Conclusion: This study proposed a standards- and BMS-based evaluation framework for the classification and safety assessment of EOL EV batteries. Major EV battery standards were selectively analyzed, and key BMS diagnostic parameters were mapped to performance and safety assessment domains. The proposed framework supports systematic decision-making for reuse, remanufacturing, and recycling. Future work will validate the framework using actual BMS data and battery test results.

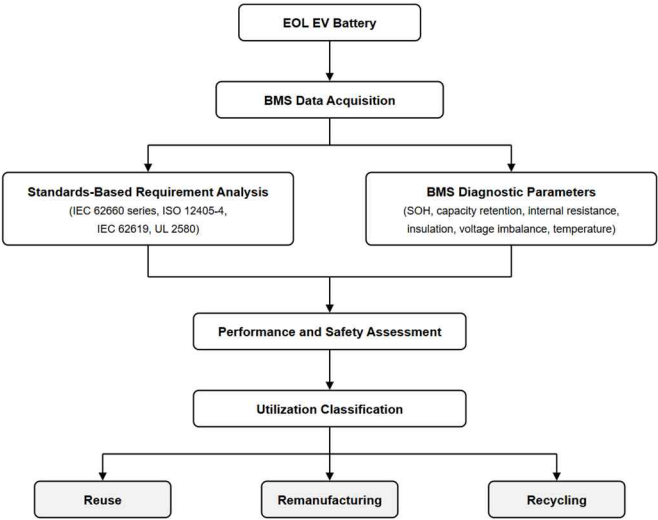


Fig. 1. Proposed standards- and BMS-based evaluation framework for EOL EV batteries

Acknowledgment

This research was supported by the Ministry of Land, Infrastructure and Transport (MOLIT), Republic of Korea, under the project “Research and Development on Safety Management Technology for EOL (End-of-Life) and Remanufactured Battery” (No. RS-2025-02309323).

References

1. IEC 62660 series, Secondary lithium-ion cells for the propulsion of electric road vehicles, International Electrotechnical Commission.
2. ISO 12405-4, Electrically propelled road vehicles — Test specification for lithium-ion traction battery packs and systems.
3. IEC 62619, Secondary cells and batteries containing alkaline or other non-acid electrolytes — Safety requirements for secondary lithium cells and batteries for use in industrial applications.
4. UL 2580, Standard for Batteries for Use in Electric Vehicles.

Development of a Field Inspection Procedure for Electric Vehicles Equipped with End-of-Life Batteries Using a Chassis Dynamometer

Seung-woo Kim ^{1*)}, Oh-hun Kwon ¹⁾, Hyun-kwon Choi ¹⁾, Jang-kyu Yun ¹⁾ and Soo-young Suk ¹⁾

¹⁾ Center for Automotive Electronics and Autonomous Systems,
Gyeongbuk Institute of IT Convergence Industry Technology, Gyeongsan 38463, Korea

*E-mail: swk@gitc.or.kr

Abstract: As the number of electric vehicles increases, a practical field inspection procedure is required to evaluate the performance and safety of end-of-life traction batteries while they remain installed in vehicles. This study aims to develop a chassis dynamometer-based field inspection procedure for electric vehicles equipped with end-of-life batteries and to review its applicability to vehicle inspection stations. The proposed procedure was organized by considering reused-battery safety inspection criteria and field demonstration materials for end-of-life battery inspection [1], [2]. Based on a field discussion and demonstration at the Haeundae vehicle inspection station of the Korea Transportation Safety Authority, the procedure consists of vehicle entry, vehicle and battery information check, underbody monitoring, insulation resistance inspection, CAN data acquisition through the data link connector, vehicle-specific certification mode setting, chassis dynamometer test, TS System EV data logging, and preparation of a field demonstration log. During the field demonstration, battery-related parameters such as vehicle speed, voltage, current, state of charge, state of health, direct current internal resistance, and state of balance indicators were displayed in real time. The field activity mainly focused on confirming the inspection sequence and software operability, rather than deriving final battery performance results. The proposed procedure provides a practical basis for applying the three-step safety inspection framework for end-of-life EV batteries to vehicle inspection stations.

Keywords : End-of-Life Battery, Electric Vehicle, Chassis Dynamometer, Field Demonstration, Battery Safety Inspection

In the first paragraph, Introduction: With the rapid deployment of electric vehicles, the need for inspection technologies to evaluate the performance and safety of end-of-life traction batteries has increased. Conventional battery evaluation generally requires removing the battery from the vehicle and measuring its capacity using charge and discharge equipment. Reused-battery safety inspection criteria include items such as insulation resistance, capacity, internal resistance, and self-discharge tests [1]. However, off-board testing may require considerable time and labor, which limits its applicability to vehicle inspection stations. Therefore, an on-board inspection procedure that can evaluate battery conditions while the battery remains installed in the vehicle is required.

In the second paragraph, Research Methods (Title can be modified):In this study, a chassis dynamometer-based inspection procedure was developed based on the field environment of the Haeundae vehicle inspection station of the Korea Transportation Safety Authority. The procedure was organized with reference to field demonstration materials and a chassis dynamometer test procedure prepared for electric vehicles equipped with end-of-life batteries [2], [3]. The inspection sequence consists of vehicle entry, vehicle and battery information check, underbody visual monitoring, insulation resistance inspection, connection of a CAN-to-USB interface to the data link connector, vehicle-specific certification mode setting, execution of the TS System EV software, chassis dynamometer-based inspection, and preparation of a field demonstration log. The main data acquisition parameters were defined as vehicle speed, voltage, current, state of charge, state of health, state of power, direct current internal resistance, and state of balance indicators.

In the third paragraph, Experimental Results:The field demonstration confirmed that an electric vehicle

can be positioned on the chassis dynamometer line of a vehicle inspection station and that underbody monitoring, insulation resistance inspection, vehicle communication, and chassis dynamometer operation can be integrated into a single inspection sequence. The TS System EV interface displayed battery-related parameters such as voltage, current, state of charge, state of health, direct current internal resistance, state of balance based on voltage, and state of balance based on temperature under a vehicle speed maintenance condition. These observations indicate that the proposed procedure can provide a practical data acquisition environment for the field inspection of end-of-life EV batteries. However, the observed values should be interpreted as preliminary outputs for confirming procedural feasibility and data acquisition capability, rather than as final performance evaluation results.

In the fourth paragraph, Conclusion: This study developed a chassis dynamometer-based field inspection procedure for electric vehicles equipped with end-of-life batteries and reviewed its applicability to vehicle inspection stations. By integrating underbody monitoring, insulation resistance inspection, vehicle communication, chassis dynamometer operation, and field demonstration log preparation, a practical field demonstration framework was established for the three-step safety inspection of end-of-life EV batteries. Future work will include repeated demonstrations for target vehicle models to identify operational improvements. In addition, chassis dynamometer-based and EIS-based inspection results will be compared with charge/discharge-based capacity evaluation data to validate the effectiveness of individual inspection methods.

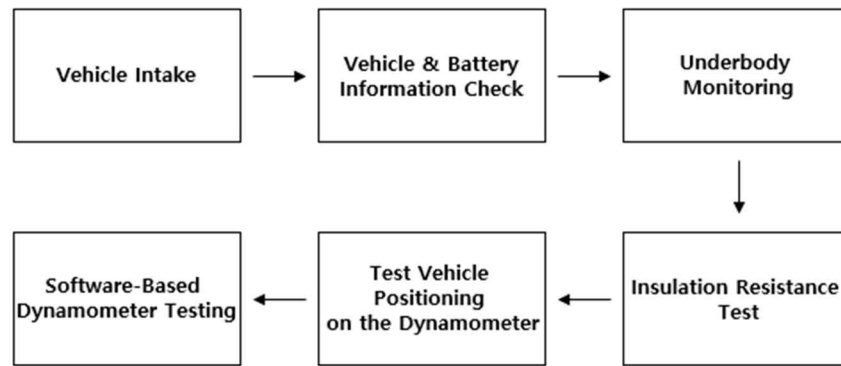


Fig. 1. Workflow of vehicle dynamometer testing for battery safety assessment

Acknowledgment

This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea (Grant No. RS-2025-02309323).

References

1. KC 10031:2025, Safety Inspection Criteria for Reused Batteries, 2025.
2. Korea Transportation Safety Authority, Field Demonstration Materials for End-of-Life Battery Safety Inspection, unpublished internal technical material, 2026.
3. Gyeongbuk Institute of IT Convergence Industry Technology, Chassis Dynamometer-Based Field Inspection Procedure for Electric Vehicles Equipped with End-of-Life Batteries, unpublished internal technical material, 2026.

Automatic Generation of Robot Offline Programs from Natural-Language Instructions Using a Large Language Model

Sanga Lee¹⁾, Jongpil Yun¹, Hongin Won¹⁾, Hyunchul Tae¹⁾ and Youngjun Yoo^{1,*)}

¹⁾ Korea Institute of Industrial Technology, Republic of Korea

*E-mail: youdalj@kitech.re.kr

Abstract: Off-line programming (OLP) of an industrial robot depends on an expert who manually teaches waypoints and writes the motion-and-IO logic for each new task, which makes program authoring the main bottleneck of robot deployment. We present an automatic OLP pipeline that generates a runnable robot program directly from a natural-language instruction, eliminating manual per-task coding. A RoboDK station is parsed once by the simulator API into a meta block of frames, tools, and IO; a large language model (LLM) agent, conditioned on a system prompt that encodes the FAIRINO FR5 Lua command set, the Fairino post-processor conventions, and three worked examples, emits a single Fairino-compatible Lua script. Generated programs are scored against expert reference code by a Lua-aware corpus BLEU-4 metric. On FR5 dispensing and material-handling stations, the pipeline produced structurally correct, simulator-loadable programs from a single instruction, indicating that few-shot LLM generation can replace manual program authoring; full quantitative evaluation is ongoing.

Keywords: offline programming, large language model, robot code generation, natural-language instruction, BLEU, FAIRINO FR5

1. Introduction

Industrial and collaborative robots are reprogrammed whenever a part, a fixture, or a task changes, and each reprogramming requires an expert to guide the robot to each pose and hand-author the motion and input/output (IO) logic — a slow step, repeated for each variant, that dominates deployment cost, as manual labeling dominates vision-dataset construction. Recent code-trained large language models can synthesize executable programs from natural language [1] and write robot policy code and task plans [2, 3] from plain instructions, opening the possibility of removing the expert from the per-task programming loop. We realize this for the FAIRINO FR5 robot in RoboDK: an instruction, grounded by automatically extracted station metadata, becomes a runnable, post-processor-compatible Lua program. The contribution is an end-to-end system that automates expert program authoring — the manual “labeling” of robot behavior — and verifies the output against expert reference code.

2. Research Methods

For grounding, a RoboDK station (.rdk) is opened once through the RoboDK Python API: the system renders an auto-framed snapshot and extracts a structured meta block — the robot, frames, tools, programs, and IO mapping — without manual transcription. This block supplies the numeric context the model must respect, replacing the manual recording of waypoints on a teach pendant.

For generation, an LLM agent receives a system prompt that fixes the target API and output contract: the FAIRINO FR5 Lua command set (MoveJ, MoveL, MoveC, SetSpeed, SetDO, WaitDI, SetTool, Sleep), the Fairino post-processor layout, and default motion and approach/retract rules. Three worked (instruction, meta, Lua) examples form the few-shot context; the new instruction with its meta block is appended as the final turn (Eq. 1):

$$y = f^{LLM}(S; (x^i, m^i, y^i)_{i=1 \dots k}; (x, m)) \quad (1)$$

where S is the system prompt, (x_i, m_i, y_i) are the $k = 3$ few-shot examples, (x, m) the new instruction–meta pair, and y the generated program. The agent maps each verb to motion and IO primitives, prefers instruction values over defaults, and unrolls repeated patterns; a provider abstraction allows a hosted (Claude) or local (Ollama) model.

For evaluation, the program y is compared with an expert reference r by BLEU-4 [4]. Because whitespace splitting merges operators and identifiers (e.g., `MoveL(p1,v100)`) into one token, both are first segmented by a Lua-aware tokenizer. The corpus-level score is given by Eq. (2):

$$BLEU = BP \cdot \exp(\sum_{n=1}^4 w_n \log p_n), \quad w_n = 1/4 \quad (2)$$

where p_n is the modified n -gram precision, w_n are uniform weights, and $BP = \min(1, \exp(1 - r/c))$ is the brevity penalty with c and r the candidate and reference token counts. Scores are computed with both NLTK and SacreBLEU [6] for reproducibility. Because BLEU measures token overlap rather than behavior, CodeBLEU [5] and simulator execution checks are planned as complementary metrics. The suite spans 30 scenarios, with the held-out set disjoint from the few-shot examples.

3. Experimental Results

We evaluated the pipeline on FR5 dispensing and pick-and-place stations in RoboDK. Fig. 1 shows the auto-rendered dispensing station; from a single parse the agent obtains the frames, tool, and program context needed to ground generation, with no manual waypoint entry. Table 1 lists this metadata, whose reference program alone contained 398 instructions that would otherwise be hand-authored. Given such an instruction (e.g., a pick-and-place with approach height, gripper signal, and speed), the agent produced a complete Lua program — poses, approach/contact/retract moves, IO, and a `main()` entry — that ran in the FR5 cell and matched the expert reference structure. Quality is quantified by the Lua-aware corpus BLEU-4 against the expert reference; the full quantitative evaluation is ongoing.

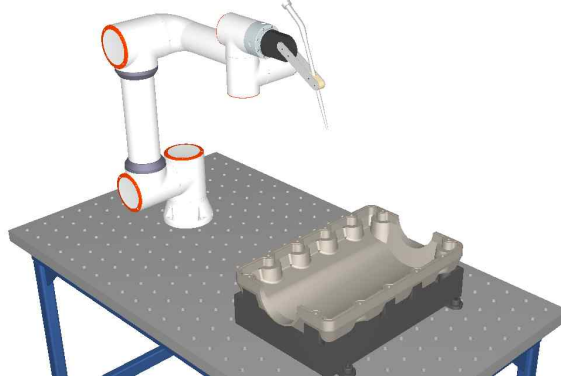


Fig. 1. Auto-rendered FAIRINO FR5 dispensing station in RoboDK, with frames, tools, and program context extracted automatically by the simulator API to ground LLM program generation.

Table 1. Station metadata automatically extracted from the FAIRINO FR5 dispensing cell (used to ground program generation; recovered without manual entry).

Item	Extracted value
Robot	FAIR Innovation FR5
Reference program	Dispensing (398 instructions)
Coordinate frames	World, FR5 Base, Part
Tool	Sealant Dispenser
Grounding	Fully automatic (RoboDK API)

4. Conclusion

We presented an automatic OLP pipeline that converts a natural-language instruction into a runnable FAIRINO FR5 Lua program by grounding a few-shot LLM agent on automatically extracted RoboDK station metadata and verifying its output against expert reference code with a Lua-aware BLEU-4 metric. On FR5 dispensing and material-handling cells, it generated simulator-runable programs from single instructions, indicating that LLM generation can replace manual robot program authoring. Future work will complete the 30-scenario BLEU evaluation, add execution-based path verification, and extend the agent to multi-robot cells. These directions chart a roadmap toward fully autonomous, instruction-driven robot

deployment, in which natural-language specification rather than manual teaching becomes the primary interface for industrial automation.

Acknowledgment

This research was supported/partially supported by UR26570 and EH260002 from KITECH. We thank our colleagues from KITECH who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

References

1. M. Chen et al., “Evaluating Large Language Models Trained on Code,” arXiv:2107.03374, 2021.
2. J. Liang et al., “Code as Policies: Language Model Programs for Embodied Control,” in Proc. IEEE Int. Conf. Robotics and Automation (ICRA), 2023.
3. I. Singh et al., “ProgPrompt: Generating Situated Robot Task Plans using Large Language Models,” in Proc. IEEE Int. Conf. Robotics and Automation (ICRA), 2023.
4. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: a Method for Automatic Evaluation of Machine Translation,” in Proc. 40th Annu. Meeting Assoc. Computational Linguistics (ACL), pp. 311–318, 2002.
5. S. Ren et al., “CodeBLEU: a Method for Automatic Evaluation of Code Synthesis,” arXiv:2009.10297, 2020.
6. M. Post, “A Call for Clarity in Reporting BLEU Scores,” in Proc. Third Conf. Machine Translation (WMT), pp. 186–191, 2018.

Establishment of EMC Test Evaluation Environment Using Hanging Mast Antenna for ADAS and Autonomous Systems*

Sungyoun Hwang¹⁾, Junbyung Baek¹⁾, Hyojeung Cho¹⁾, Simok Lee¹⁾, Jaerok Shim¹⁾, and Taehee Lee^{1*)}

Test & Evaluation Division, Korea Intelligent Automotive Parts Promotion Institute, Daegu, Korea

*E-mail: thlee@kiapi.or.kr

Abstract: The growing complexity of advanced driver assistance systems (ADAS) and autonomous driving systems has led to increased demand for electromagnetic compatibility (EMC) compliance testing, yet conventional ground-mounted antenna mast structures can be detected as obstacles by onboard sensors, compromising test reliability. This paper investigates the suitability of a hanging mast antenna system through a comparative analysis with a conventional ground-mounted antenna system, and presents the facilities and distinctive features of the mobility EMC evaluation center under development at KIAPI.

Keywords : electromagnetic compatibility(EMC), advanced driver assistance system(ADAS), autonomous system, hanging mast

The integration density and complexity of vehicle electronic systems have increased dramatically with the development of future mobility technologies, as illustrated in Fig. 1, making electromagnetic compatibility (EMC) a crucial component in ensuring vehicle safety and reliability [1]. Advanced driver assistance systems (ADAS) and autonomous functionalities, which rely on radar and lidar sensors, require new EMC testing environments that can evaluate system performance under realistic driving scenarios. Conventional ground-mounted antenna mast structures can be perceived as physical obstacles by automotive sensors, leading to significant inaccuracies in EMC compliance testing for ADAS and autonomous functions.

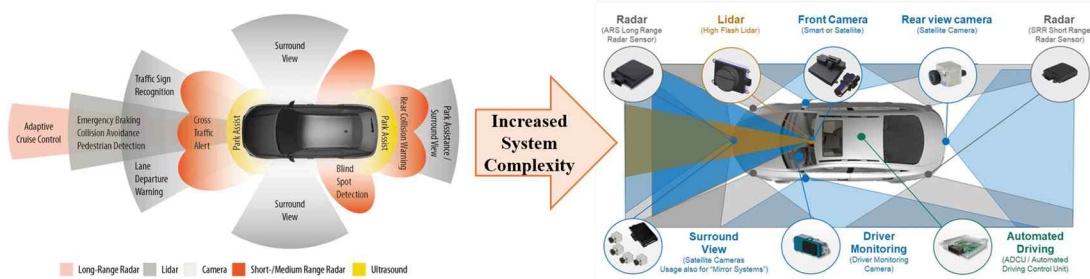
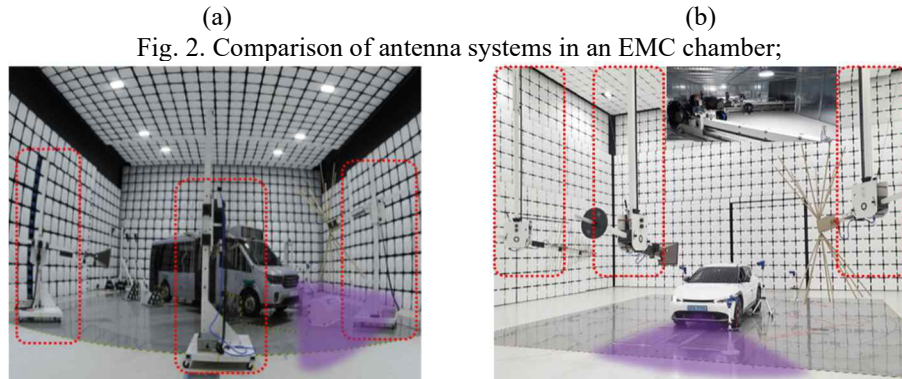


Fig. 1. Trend of increasing onboard electronics and sensors in future vehicles

To address these challenges, international standardization organizations are actively revising standards and regulations, such as ISO/TR 7964 and UN ECE R10 [2, 3], to support EMC compliance testing for ADAS and autonomous driving functions. Consequently, a novel test environment is required to provide realistic EMC test conditions without introducing test equipment that onboard sensors may detect as obstacles. One promising approach is a hanging mast antenna system that suspends the antenna from the chamber ceiling, eliminating sensor-detectable objects and providing a reconfigurable EMC test environment. As illustrated in Fig. 2, conventional ground-mounted antenna mast structures may be recognized as obstacles by vehicle-mounted sensors, whereas hanging mast systems eliminate such structures from the sensor field of view [4].

Addressing the EMC testing requirements for ADAS and autonomous vehicles, the Korea Intelligent Automotive Parts Promotion Institute (KIAPI) is establishing a mobility EMC evaluation center through the "Electromagnetic Disturbance Evaluation System for Realistic Mobility Environment Simulation" project. The project aims to establish a realistic EMC test environment by implementing a hanging mast

antenna system and an ADAS EMC testing platform that incorporates real-world driving scenario simulation.



(a) conventional ground-mounted antenna system and (b) hanging mast antenna system

The center features multiple EMC test facilities tailored to different evaluation requirements. The primary large-scale EMC chamber can accommodate a variety of test targets, ranging from electronic components to heavy-duty trucks and electric double-decker buses. In addition, a component-level EMC chamber is available for evaluating individual automotive parts and modules. The facility also includes a shielded room for conducted emission and immunity measurements, as well as electrostatic discharge (ESD) testing. Furthermore, KIAPI's existing infrastructure, including its proving ground and various vehicle certification testing facilities, enables a comprehensive one-stop testing service. This allows manufacturers to fulfill the full range of automotive certification requirements within a single institution, from EMC compliance testing to real-world driving assessment.

Existing EMC test environments are becoming increasingly unsuitable for ADAS and autonomous driving systems, as traditional test equipment can interfere with onboard sensors and compromise test reliability. This paper investigates the suitability of a hanging mast antenna system for EMC testing of ADAS and autonomous driving functions through a comparative analysis with conventional ground-mounted antenna configurations. Furthermore, this paper presents the facilities and distinctive features of the mobility EMC evaluation center under development at KIAPI, highlighting its integration with existing automotive certification infrastructure to enable a comprehensive one-stop testing service.

Acknowledgment

This research was supported by the Korea Institute for Advanced of Technology (KIAT) grant funded by the Korea government (MOTIR) (Project No. P0027952, Project Title: Establishment of Corporate Support Base to Response to Electromagnetic Environment Disturbances Simulating Actual Urban Environments).

References

1. H. Woo, *Journal of Auto-vehicle Safety Association*, 12(4), p.7-12 (2024)
2. ISO, ISO/TR 7964
3. UN ECE, Regulation No.10
4. S. Lee, H. Kim, J. Lee, H. Kim and B, Choi, *Journal of the Korea Automobile Safety Society*, 36(4), p.330-338 (2025)

* This paper was previously presented at 2025 KSAE Annual Autumn Conference.

Development of an IR LED-Based Illumination System for Parking Violation Enforcement Cameras

Jung-Hae Choi¹⁾ and Byung-Jae Choi^{2*)}

¹⁾ DGELEC, Yeongcheon-si, Gyeongsangbuk-do 38453, Korea

²⁾ Dept. of Electrical Engineering, IT & Engineering College, Gyeongsan-si, Gyeongsangbuk-do 38453, Korea

*E-mail: cjh@dgelec.kr

Abstract: Parking violation enforcement cameras use vision sensors to capture vehicle images and recognize license plates for enforcement purposes. Nighttime enforcement requires an illumination system; however, conventional white light may cause visual disturbance to pedestrians and drivers. To overcome this limitation, this study proposes an IR LED-based illumination system for parking violation enforcement cameras. The proposed system is expected to improve nighttime recognition performance while minimizing interference with pedestrians and drivers. Future research will focus on the development of current control and brightness control functions to enhance the stability and efficiency of the illumination system.

Keywords: LED, VISION SENSOR, IR-LED, LED CONTROLLER

Introduction: Recent advances in vision sensors and image recognition technologies have led to an increasing demand for dedicated illumination systems. Since illumination plays a crucial role in image acquisition, various lighting approaches, including mobile operating systems, have been explored [1]. In parking violation enforcement systems, vision sensors are used to detect vehicle backgrounds and license plates under both daytime and nighttime conditions. However, additional illumination is generally necessary to improve recognition performance at night. In such applications, the use of white or high-intensity lighting is limited because it may cause visual discomfort or distraction to pedestrians and drivers. Therefore, this study proposes an infrared light-emitting diode (IR LED)-based illumination system that is nearly invisible to the human eye and well-suited for nighttime image acquisition.

Research Methods: The proposed system is composed of three main components: an LED module, a driving circuit, and a control unit. The LED module consists of a 6×5 LED array. The driving circuit is implemented using MOSFETs, which are commonly used as switching devices. The control unit employs a microprocessor to regulate the operation of the driving circuit.

Table. 1. System specification

Part	Spec	Notes
LED	IR LED, 1.4W	6 x 5 array
MOSFET	I _D 49A, V _{DS} 55V R _{DS} 17.5mΩ	IRFZ44N
MCU	MAX 72MHz Cortex-M3	STM32F103C8T6

The proposed system was applied to detect the license plate of an actual vehicle. The results confirmed that the use of the illumination system produced a noticeable difference compared with the unilluminated system.

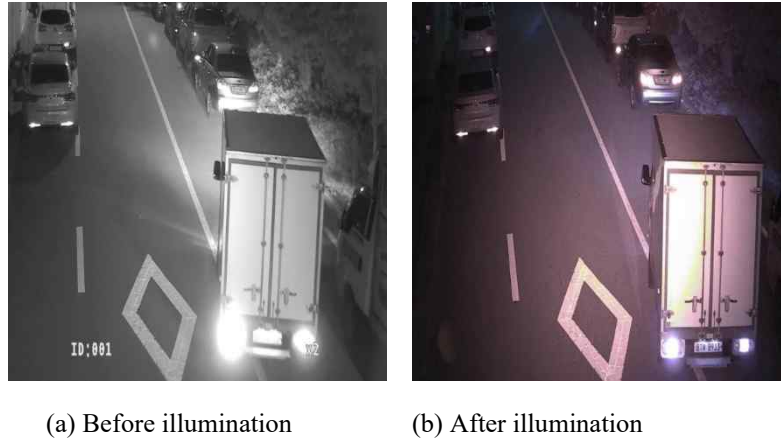


Fig. 1. Images acquired before and after illumination

Conclusion: This paper outlined an IR LED-based illumination system and acquired actual images taken by the implemented system. The use of IR LEDs is advantageous for nighttime applications because they are hardly visible to the human eye and therefore minimize disturbance to pedestrians. Future research will focus on the development and experimental validation of brightness control and current control systems for IR LEDs.

References

1. M. G. Cho, *A Green LED Sensitivity Lighting System Using Smart Phone*, IEMEK J. Embed. Sys. Appl., Vol. 7, No. 2, pp. 103-111 (2012).

STFT-CNN Based Fall Detection Using Millimeter-Wave Radar

Jieun Park¹⁾, Kyungtaeg Yang¹⁾, Sekyung Kim¹⁾, Dokyung Kang¹⁾, Seungeon Song²⁾ and Jonghun Lee^{*1),2)}

¹⁾ Dept. of Interdisciplinary Engineering, Graduate school, DGIST, Daegu, 42988, Korea

²⁾ Institute of Research, DGIST, Daegu, 42988, Korea

E-mail: jhlee@dgist.ac.kr*

Abstract: Falls take a serious risk of injury and secondary complications, particularly among elderly individuals living alone and other vulnerable populations, underscoring the critical need for rapid and reliable detection. Although camera-based fall detection systems have achieved notable recognition performance, they are inherently constrained by privacy concerns, susceptibility to illumination variations, and occlusion effects. To address these limitations, this paper proposes a fall detection framework utilizing 60GHz millimeter-wave radar. Range FFT is applied to the collected complex beat signals to extract range-domain information, followed by clutter rejection for removing background components from the radar signals. The micro-Doppler spectrograms serve as input images for a CNN-based classifier. The classifier distinguishes fall and non-fall behaviors. This paper investigates the feasibility of fall and non-fall classification using millimeter-wave micro-Doppler signatures under daily-living home environments.

Keywords : Fall Detection, Millimeter-Wave Radar, Micro-Doppler, STFT, CNN

1. Introduction

With the accelerating progression of global population aging, fall accidents among elderly individuals living alone and other high-risk groups have emerged as an important social concern. Falls result in fractures, loss of consciousness, and prolonged physical deterioration when timely medical intervention is missing, reinforcing the imperative for early automatic detection. Camera-based fall detection systems have been extensively developed and have achieved competitive detection performance. However, they remain fundamentally constrained by privacy infringement, performance degradation under bad illumination conditions, and vulnerability to occlusion. Millimeter-wave radar, by contrast, enables privacy-preserving detection that is independent of ambient lighting, offering a compelling alternative for robust fall detection. This paper proposes an STFT-CNN-based framework that exploits 60 GHz complex radar signals to classify fall and non-fall behaviors under daily-living home environments.

2. Research Methods

Experimental data were collected using the 60 GHz millimeter-wave radar. All experiments were conducted in an indoor environment for fall and non-fall behavior categories. Fall consisted of four kinds scenarios: forward fall, backward fall, left-side fall, and right-side fall. Non-fall included walking, sitting, lying, standing up, and object picking. Range FFT was applied to the acquired raw IQ data to extract range information. Static reflections and background interference were subsequently suppressed through a clutter removal process. STFT generated micro-Doppler spectrograms from the processed signals. The generated spectrograms serve as input images for a CNN-based classifier. In this paper, ResNet-18 was selected as the baseline CNN architecture for fall and non-fall behavior classification. The proposed framework consists of Range FFT, clutter removal, STFT-based micro-Doppler generation, and CNN-based fall classification using ResNet-18.

3. Experimental Results

Micro-Doppler spectrograms were generated from the 60 GHz complex beat signals collected during the experiments. Fig. 1 shows one of measured spectrograms of fall and non-fall activities. The forward fall event produces an abrupt velocity transition and a broad Doppler distribution within a short time interval. The stationary sitting posture concentrates most of its energy near the zero-velocity region. These patterns indicate clear differences between fall and non-fall activities in the radar signature. The measured

spectrograms contain discriminative motion information. Therefore, they can be used as input features for CNN-based fall and non-fall classification.

4. Conclusion

This paper proposed an STFT-CNN-based fall detection framework utilizing 60GHz millimeter-wave radar. Range FFT, clutter removal, and STFT processes were applied to generate micro-Doppler spectrograms representing human motion characteristics. The generated spectrograms exhibited distinguishable patterns between fall and non-fall activities. These patterns can serve as effective input features for CNN-based classification. Future work will focus on implementing and evaluating a ResNet-18-based classifier, expanding the dataset including sufficient human and non-human subjects for more various activity scenarios, and quantitatively assessing fall detection performance.

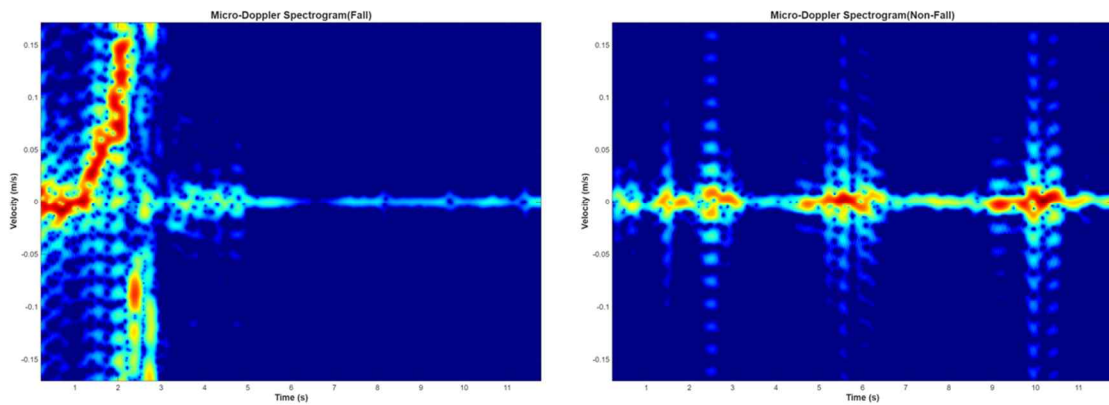


Fig 1. Measured micro-Doppler spectrograms of fall and non-fall activities.

Acknowledgment

This research was supported by the DGIST R&D Program of the Ministry of Science, ICT and Future Planning, Korea (26-IT-01), by the Technology Development Program (RS-2025-25450185) funded by the Ministry of SMEs and Startups (MSS, Korea) and Private-Public Joint Technology Commercialization R&D Program(RS-2026-25527424) funded by the Ministry of SMEs and Startups(MSS, Korea).

References

1. Ma, Lei, et al. "Fall direction detection in motion state based on the FMCW radar." *Sensors* 23.11 (2023): 5031.
2. Yao, Yicheng, et al. "Fall detection system using millimeter-wave radar based on neural network and information fusion." *IEEE Internet of Things Journal* 9.21 (2022): 21038-21050.
3. Hu, Shuting, et al. "Radar-based fall detection: A survey [survey]." *IEEE robotics & automation magazine* 31.3 (2024): 170-185.

STFT-CNN Based Fall Detection Using Millimeter-Wave Radar

Yu Gong¹⁾, Eun-Bin Koh²⁾, Misbah Bibi²⁾ and Do-Hyun Kim²⁾

¹⁾Dept. of Japanese Language & Literature / Intelligent IoT Convergence, Jeju National University, Jeju, Korea

²⁾Dept. of Computer Engineering, Jeju National University, Jeju, Korea

*E-mail: asrhddb@stu.jejunu.ac.kr

Abstract: With the growth of elderly people living alone in a super-aged society, contactless and real-time fall detection is needed to improve home safety without the discomfort of wearable devices. This paper proposes SmartCare, an on-device AI integration mechanism that connects a Raspberry Pi 5 edge node, a Node.js server, and React Native mobile applications. In the proposed method, the Raspberry Pi 5 extracts 33 pose landmarks (132 values per frame) with MediaPipe, accumulates 30 frames as a (30, 132) LSTM input tensor, and classifies SAFE/FALL states at 22.6 FPS. When a fall is detected, the device sends an HTTP POST request to the server, which stores the event and delivers a real-time guardian Notification through a Socket.io WebSocket channel. End-to-end testing confirms the feasibility of the three-tier architecture, and model evaluation on 511 samples achieved 72.0% overall accuracy. As fall-class accuracy is currently limited by data imbalance, future work will focus on improving detection performance through additional balanced training data, more comprehensive evaluation metrics, and validation under more diverse real-world scenarios.

Keywords: On-device AI; Raspberry Pi; fall detection; LSTM; WebSocket; React Native; Smart healthcare

1. Introduction

Korea has entered a super-aged society, increasing the need for continuous safety monitoring for elderly people living alone. In home environments, falls are one of the most critical safety risks because they can lead to serious injury when immediate assistance is not provided. Many fall accidents occur indoors, where elderly people spend most of their time, and delayed detection may increase the risk of secondary damage [1]. Although wearable fall-detection systems have been widely studied, their practical use can be limited by discomfort, battery charging requirements, and inconsistent wearing behavior [2]. Therefore, contactless monitoring methods are needed to provide continuous safety support without requiring users to attach or wear additional devices.

Vision-based and skeleton-based approaches provide a promising alternative by estimating body posture from video and recognizing abnormal motion patterns [3]. In particular, pose-estimation models such as MediaPipe can extract human joint landmarks from camera input, allowing fall detection to be performed using skeleton data rather than raw images [4]. This approach can reduce the amount of transmitted data and is suitable for edge-based inference because the fall-detection model can analyze compact pose features locally [5]. When combined with temporal models such as LSTM, the system can consider motion sequences over multiple frames and distinguish between normal movement and fall-related posture changes [6][7].

This paper proposes the integration mechanism of SmartCare, a real-time fall-detection and NOTIFICATION SERVICE for elderly home monitoring [8]. The proposed mechanism integrates on-device pose estimation and LSTM classification on a Raspberry Pi 5 with a Node.js server and two React Native applications. The main contribution of this work is a practical three-tier data flow that connects edge inference, event storage, WebSocket notification, safety confirmation, and SOS Notification transmission in a single deployable architecture. By combining local AI processing with server-based data management and mobile notification, SmartCare aims to improve responsiveness, reduce dependence on cloud inference, and support real-time guardian intervention when a fall event occurs.

2. REAL-TIME FALL DETECTION AND EMERGENCY NOTIFICATION SERVICE

The proposed system consists of three layers, as shown in the overall system architecture. The on-device node is implemented on the Raspberry Pi 5 and is responsible for video capture, MediaPipe-based pose extraction, LSTM inference, and HTTP transmission of detected events. The server layer acts as the central data hub and provides a REST API, Socket.io WebSocket communication, SQLite event storage, and JWT authentication with bcrypt password hashing. The mobile application layer includes a guardian app with five tabs and an elderly user app with four tabs, enabling real-time Notification reception, event-history review, safety confirmation, and emergency SOS messaging. Through this three-layer structure, sensing, data management, and user interaction are separated while remaining connected through lightweight network communication.

For each frame, the Raspberry Pi 5 extracts 33 body landmarks using MediaPipe and represents them as 132 feature values, including landmark coordinates and related pose information. A sequence buffer accumulates 30 consecutive frames to generate a (30, 132) tensor, which is used as the input to the LSTM model. The model analyzes temporal changes in body posture and classifies the sequence as SAFE or FALL using a decision threshold of 0.45. When the predicted fall probability exceeds this threshold, the Raspberry Pi sends an HTTP POST request to the '/fall-detected' endpoint, including the camera identifier, fall probability, and event information. This process allows fall detection to be performed locally before only the detection result is transmitted to the server.

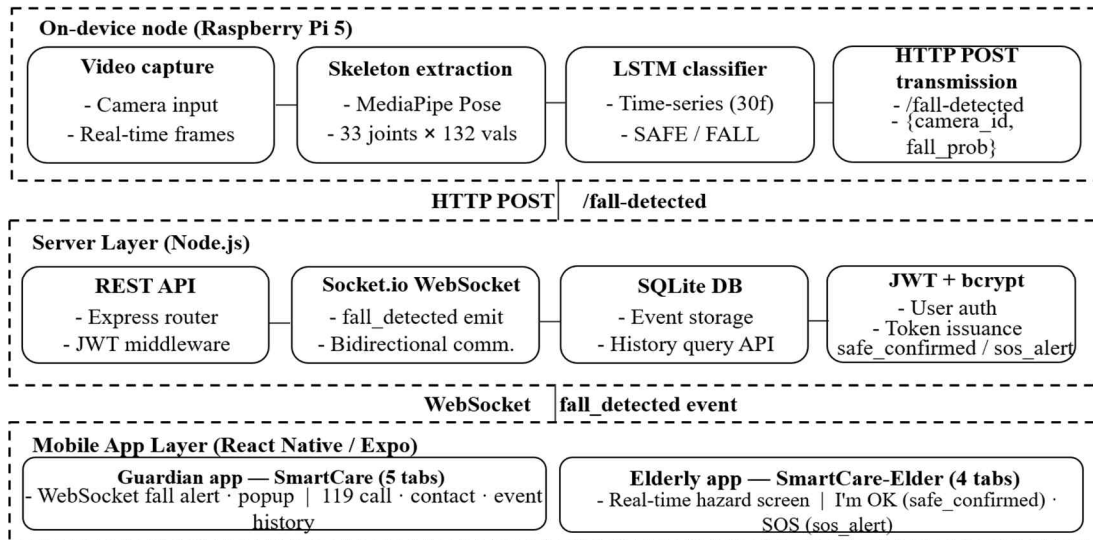


Fig. 1. Operational overview of fall detection and NOTIFICATION SERVICE.

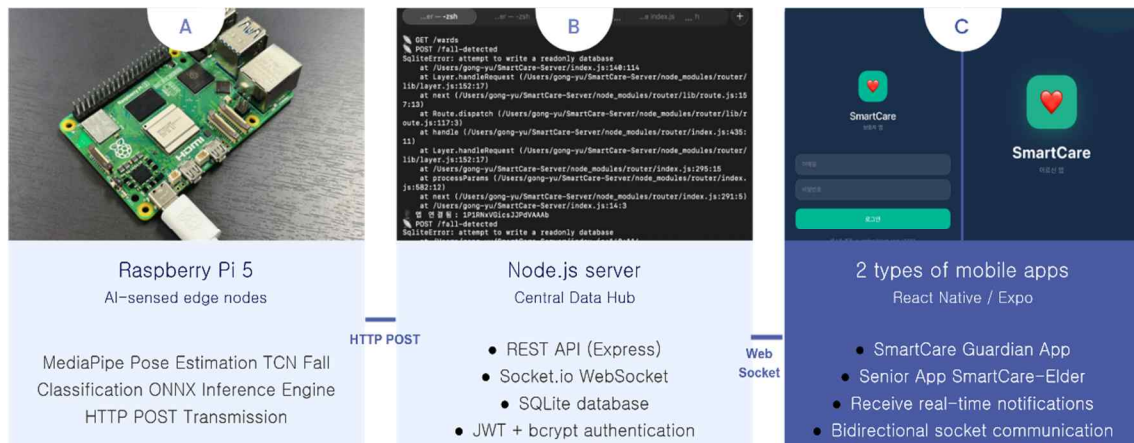


Fig 2. Service architecture for edge-based fall detection and centralized data management, and real-time mobile communication.

TABLE 1. KEY API ENDPOINTS AND SOCKET EVENTS

Endpoint / Event	Method	Description
/auth/login	POST	User authentication and JWT issuance
/fall-detected	POST	Receives fall event from Raspberry Pi
/events	GET	Retrieves fall event history
/wards	GET	Retrieves elderly user information
fall_detected	Socket.io	Real-time fall Notification to guardian app
safe_confirmed	Socket.io	Elderly user confirms safety
sos_Notification	Socket.io	Emergency SOS request from app

3. SERVER AND MOBILE INTEGRATION FOR REAL-TIME FALL NOTIFICATIONS

After receiving a fall event from the Raspberry Pi 5, the server validates the incoming request to ensure that it is transmitted from an authorized edge device and includes the required information, such as the user identifier, detection result, timestamp, and event status. Once validated, the event is recorded in the SQLite database as a persistent log for later review. The server then emits a 'fall_detected' Socket.io message to the connected guardian application, allowing the guardian to receive an immediate Notification and view the corresponding record in the event history screen.

The proposed integration also supports reverse-direction communication from the mobile application to the server. When the elderly user confirms that no assistance is needed, the elderly app sends a 'safe_confirmed' event, which updates the event status and informs the guardian that the user is safe. In addition, when the user manually requests help, the app transmits an 'sos_Notification' event, which is forwarded to the guardian application as an emergency notification. This bidirectional structure allows the system to handle automatic fall detection, false-alarm confirmation, and user-initiated SOS requests within the same communication framework.

This design separates time-critical notification from persistent data management. REST endpoints handle authentication, elderly user information, history retrieval, and database operations, while Socket.io provides low-latency delivery of fall and SOS events. Since fall recognition is performed locally on the Raspberry Pi 5 and only event-level information is transmitted to the server, the system reduces dependence on cloud inference, lowers network load, and avoids continuous transmission of raw camera data. As a result, the integration improves responsiveness and privacy for real-time home monitoring of elderly people living alone.

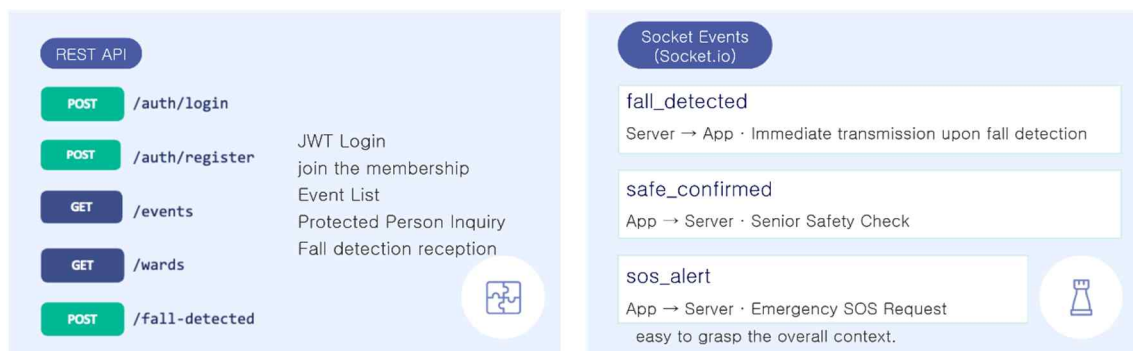


Fig 3. Overview of the backend communication interface, showing REST API endpoints and real-time Socket.io events for authentication, event management, fall detection reporting, safety confirmation, and emergency SOS communication.

4. IMPLEMENTATION AND EVALUATION OF REAL-TIME FALL DETECTION AND NOTIFICATION SERVICE

The implemented system includes the guardian app, the elderly user app, the Node.js server, and the deployed LSTM fall-detection model on the Raspberry Pi 5. The guardian app provides real-time fall-notification reception and event-history review, while the elderly user app supports safety confirmation and emergency SOS transmission. The Node.js server manages authentication, event storage, REST API communication, and Socket.io-based message delivery. End-to-end tests confirmed that a fall prediction generated on the edge device can trigger an HTTP request, create a server-side event record, and deliver a real-time Notification to the guardian application through WebSocket communication. This verifies that the proposed three-tier architecture operates as an integrated fall-detection and notification workflow rather than as separate independent modules.

The fall-detection network consists of two stacked LSTM layers with 128 hidden units and a dropout rate of 0.3, and was trained with a learning rate of 0.001 on a dataset partitioned into 987 training, 497 validation, and 511 test samples. On the held-out test set of 511 samples, the model achieved 72.0% overall accuracy. The normal class achieved an F1-score of 79.5%, with 81.0% precision and 78.1% recall, showing relatively stable recognition of non-fall situations. The fall class achieved an F1-score of 55.7%, with 53.6% precision and 58.1% recall, and the weighted-average precision, recall, and F1-score across both classes were 72.7%, 72.0%, and 72.3%, respectively. The lower fall-class performance is mainly attributable to class imbalance between normal samples (356) and fall samples (155), which limits the model's ability to learn diverse fall patterns. This result indicates that additional fall data and balanced training samples are needed to improve sensitivity to actual fall events.

The Raspberry Pi 5 inference pipeline operated at 22.6 FPS, demonstrating sufficient real-time throughput for the proposed notification workflow. Since pose extraction, sequence buffering, and LSTM inference are performed locally on the edge device, the system can detect falls without continuously transmitting raw video data to the server. The evaluation results show that the current implementation is feasible for real-time home monitoring, although further improvement is required to increase fall-class accuracy under more diverse conditions, such as occlusion, different camera angles, and varied fall postures.

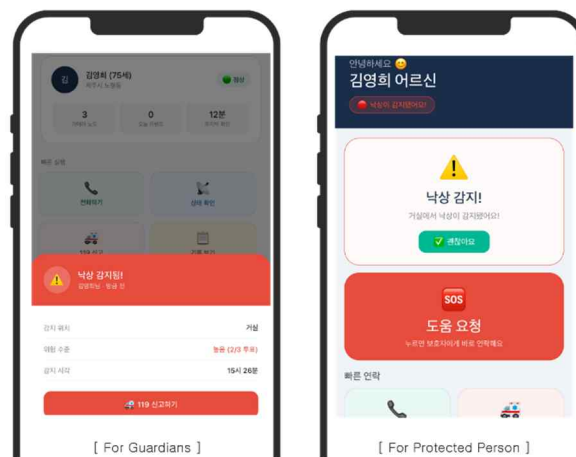


Fig 4. Mobile application interfaces for guardians and protected persons, showing fall detection Notifications, safety confirmation, emergency SOS request, and quick contact functions.

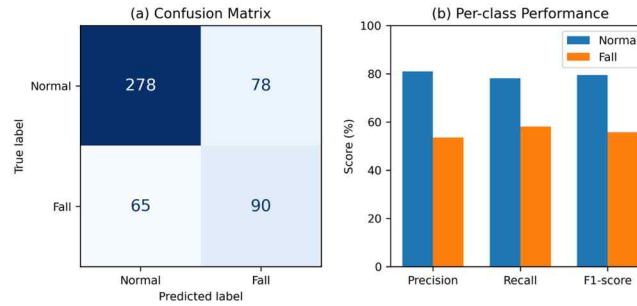


Fig 5. Classification performance of the LSTM fall-detection model on the 511-sample test set: (a) confusion matrix and (b) per-class precision, recall, and F1-score.

5. FUTURE WORK

Although the current implementation demonstrates a complete end-to-end fall-detection and notification workflow, several aspects remain to be improved based on the evaluation results and reviewers' suggestions. First, fall-class recognition is still limited by class imbalance between normal (356) and fall (155) samples; future work will collect additional real-world fall data and construct a more balanced dataset to strengthen fall-class sensitivity. Second, the system will be assessed with more comprehensive evaluation metrics, such as confusion-matrix-based analysis, precision-recall and ROC curves, and k-fold cross-validation, to provide a more reliable measure of detection performance. Third, single-camera monitoring may suffer from occlusion, limited viewing angles, and missed detections when the user leaves the camera range; the architecture will therefore be extended to multi-room monitoring using five camera nodes and validated under more diverse real-world scenarios, including different home layouts, lighting conditions, and fall postures. These improvements are expected to enhance the reliability and practical applicability of SmartCare as an elderly-care support system.

6. CONCLUSION

This paper presented the integration mechanism of SmartCare, a real-time fall-detection system that combines Raspberry Pi 5-based on-device AI, a Node.js server, and React Native mobile applications. The proposed system performs MediaPipe-based pose extraction and LSTM-based fall classification locally on the Raspberry Pi 5, while the server and mobile applications support event storage, authentication, real-time notification, safety confirmation, and emergency SOS communication. By connecting local MediaPipe + LSTM inference with a REST API and Socket.io WebSocket communication, the system demonstrates the feasibility of a low-latency smart healthcare monitoring service for elderly people living alone.

The implementation results show that the proposed three-tier architecture can process fall events from the edge device, store them in the server database, and immediately deliver Notifications to guardian users. This confirms that on-device AI can reduce dependence on cloud-based inference and avoid continuous transmission of raw video data, which is important for both responsiveness and privacy in home monitoring environments. The system also supports bidirectional communication, allowing elderly users to confirm safety or send SOS Notifications, thereby improving practical usability beyond automatic fall detection alone.

ACKNOWLEDGMENT

This research was supported by the Regional Innovation-Centered University Support System (RISE) COSS, funded by the Ministry of Education and Jeju Special Self-Governing Province through the Jeju RISE Center (2026-RISE-17-001).

References

1. Y. Liu et al., “A review of wearable sensors based fall-related recognition systems,” *Engineering Applications of Artificial Intelligence*, vol. 126, Art. no. 107126, 2023.
2. J. V. H. Hoang et al., “Advances in skeleton-based fall detection in RGB videos: From handcrafted to deep learning approaches,” *IEEE Access*, vol. 11, pp. 92322–92352, 2023.
3. B. S. Lin et al., “Fall detection system with artificial intelligence-based edge computing,” *IEEE Access*, vol. 10, pp. 4328–4339, 2022.
4. E. R. Sykes, “Next-generation fall detection: harnessing human pose estimation and transformer technology,” *Health Systems*, vol. 14, no. 2, pp. 85–103, 2025.
5. F. Shu and J. Shu, “An eight-camera fall detection system using human fall pattern recognition via machine learning by a low-cost android box,” *Scientific Reports*, vol. 11, no. 1, Art. no. 2471, 2021.
6. Statistics Korea, “2024 Statistics on the Aged,” Ministry of Data and Statistics, Republic of Korea, Press Release, Sep. 26, 2024.
7. World Health Organization, “Falls,” Fact Sheet, Apr. 26, 2021.
8. V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device real-time body pose tracking,” *arXiv:2006.10204*, 2020.

Deep Reinforcement Learning-Based Lumen Alignment for Defect Inspection of Catheter Cross-section

Jae Jin Hwang¹⁾, Yejin Kim²⁾, Jong Pil Yun³⁾ and Moon G. Joo^{4*)}

¹⁾ Dept. of Intelligent Robotics Engineering, Pukyong National University, Busan, Korea

^{2,3)} Manufacturing AI Research Center, Korea Institute of Industrial Technology, Incheon, Korea

^{4*)} Dept. of Information and Communication Engineering, Pukyong National University, Busan, Korea

*E-mail: gabi@pknu.ac.kr

Abstract: Visual defect inspection of catheter cross-sections requires geometric comparison between the observed lumen structure and a reference template, but manual inspection is sensitive to rotational pose: even a non-defective section can appear mismatched when its lumen orientation is not normalized. This paper presents a deep reinforcement learning-based lumen alignment method. The problem is formulated as a Markov decision process in which a Deep Q-Network (DQN) agent observes a cropped cross-sectional image state and selects a discrete rotation action from -15, -5, -1, 0, +1, +5, and +15 degrees, where 0 acts as a stop action declaring completion. A template-matching procedure estimates the target orientation for reward and evaluation only, so the target angle is never given to the agent directly. Comparing two state representations on 60 catheter images, the lumen-mask state achieved the best evaluation success rate of 94.0 % with a mean angular error of 1.42 degrees, outperforming the raw grayscale state. DQN-based lumen alignment thus provides a reproducible preprocessing basis for subsequent rule- or measurement-based catheter defect inspection.

Keywords: catheter inspection, deep reinforcement learning, Deep Q-Network, lumen alignment, defect inspection

Introduction: Catheter cross-section inspection is an important quality-control step because lumen position, symmetry, and shape directly affect product functionality and manufacturing consistency. In practice, inspectors compare a captured cross-section with a reference shape and judge whether the lumen arrangement is acceptable. This process is difficult to reproduce when the cross-section is rotated differently from the reference template. A robust alignment method is therefore required before defect judgment. Instead of using only fixed image registration rules, this study investigates a DQN-based policy that learns how to rotate a catheter cross-section toward a template-aligned orientation through sequential decision making [1].

Research Methods: The framework consists of image preprocessing, target-angle estimation, a DQN alignment environment, and evaluation [2], [3]. The catheter section is detected from the grayscale image and cropped into a square region of interest (ROI), from which two input states are prepared: a raw grayscale ROI and a binary-like lumen-mask ROI emphasizing the internal lumen structure. Each episode initializes with a randomly rotated state, and at every step the agent selects one of seven rotation actions (-15, -5, -1, 0, +1, +5, +15 degrees). The reward combines reduction of angular error, a small step penalty, a large reward for a correct stop, and a penalty for an incorrect stop. A lightweight Q-network of three convolutional layers followed by two fully connected layers (about 0.83 M parameters) receives an 84×84 single-channel state and outputs action-wise Q-values [4], [5], trained with experience replay [6], a target network, and Double DQN targets [7]. The target angle of each image is estimated in advance by rotating the lumen feature against the reference template and selecting the angle with the highest intersection-over-union; this angle is used only for reward and evaluation.

Experimental Results: The experiment used 60 catheter cross-section images (48 for training, 12 for testing), with the DQN trained separately on raw grayscale and lumen-mask states under the same action

space, reward function, and split. As summarized in Table 1, the lumen-mask state raised the best evaluation success rate from 92.0 % to 94.0 % and cut the mean angular error from 4.98 to 1.42 degrees, indicating that emphasizing lumen geometry is more effective than the full grayscale ROI. At full resolution, alignment achieved 83.0 % success with a 1.00-degree median error for both states, showing that successful cases converge close to the reference orientation while failures concentrate in ambiguous samples. Fig. 1 shows a representative alignment trajectory from the trained agent, where each panel is one decision step labeled with the current angular error (err) and selected action (a). Starting from an 83-degree error, the agent repeatedly applies negative rotations to reduce the mismatch, then a small corrective action near the target, and finally the 0-degree stop action once the residual error reaches 2 degrees. Regarding computational efficiency, one alignment episode requires on average only 7.4 forward passes of the Q-network (Table 1, mean steps). Each decision step, including image rotation and state preprocessing, takes on average 0.53 ms on a single NVIDIA GeForce GTX 1080 GPU, so a full alignment completes in about 4 ms per image, which is sufficient for in-line real-time inspection.

Table 1. Lumen alignment performance for the two state representations.

State representation	Best eval. success (%)	Best mean error (deg)	Mean steps	Full-res. success (%)	Full-res. median error (deg)
Raw grayscale ROI	92.0	4.98	7.68	83.0	1.00
Lumen-mask ROI	94.0	1.42	7.40	83.0	1.00

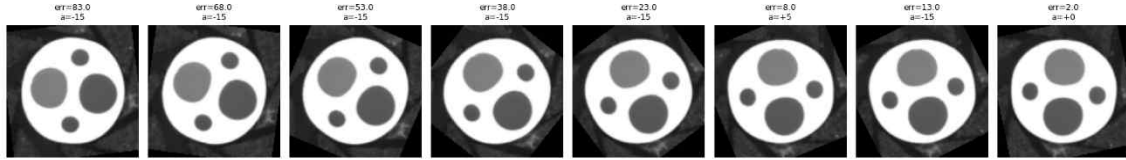


Fig. 1. DQN-based sequential lumen alignment process.

Conclusion: This study demonstrates a DQN-based lumen alignment approach for catheter cross-section inspection. By learning discrete rotation actions from image states, the agent normalizes lumen orientation without directly observing the target angle, with the lumen-mask state giving the strongest performance (94.0 % success, 1.42-degree mean error) at a per-image alignment cost of about 4 ms. Future work will scale to 200-image training and holdout sets and integrate the aligned masks with geometric defect criteria for end-to-end pass/fail inspection.

Acknowledgment

This work was supported by the Machinery and Equipment Industrial Technology Development Program (RS-2024-00507222) funded by the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea.

References

1. R. Liao et al., Proc. AAAI Conf. Artif. Intell., 31(1), 4168 (2017).
2. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., MIT Press, Cambridge, MA, p. 1 (2018).
3. M. Towers et al., arXiv:2407.17032 (2024).
4. C. J. C. H. Watkins and P. Dayan, Mach. Learn., 8(3), 279 (1992).
5. V. Mnih et al., Nature, 518(7540), 529 (2015).
6. L. J. Lin, Mach. Learn., 8(3), 293 (1992).
7. H. van Hasselt, A. Guez and D. Silver, Proc. AAAI Conf. Artif. Intell., 30(1), 2094 (2016).

A Scenario-Based Electromagnetic Susceptibility Evaluation Method for ADAS Functions Using RTS and Camera Projection

Junbyung Baek¹⁾, Sungyoun Hwang¹⁾, Hyeongsu Choi¹⁾, Simok Lee¹⁾, Jaerok Shim¹⁾, Hyojeung Cho¹⁾ and Taehee Lee^{1*)}

¹⁾Test & Evaluation Division, Korea Intelligent Automotive Parts Promotion Institute, Daegu, Korea

*E-mail: thlee@kiapi.or.kr

Abstract: The conventional vehicle-level electromagnetic susceptibility (EMS) tests evaluate vehicle operation under standardized conditions without considering realistic traffic scenarios in which advanced driver assistance systems (ADAS) functions are activated. To address this limitation, this paper proposes a scenario-based EMS test method incorporating a radar target simulator (RTS) and a camera projection system to reproduce forward vehicle approach, deceleration, and stop scenarios within a controlled electromagnetic test environment. The proposed method is expected to contribute to the development of standardized EMS evaluation procedures for ADAS and autonomous driving functions.

Keywords : electromagnetic susceptibility (EMS), advanced driver assistance systems (ADAS), adaptive cruise control (ACC), autonomous emergency braking (AEB)

The adoption of advanced driver assistance systems (ADAS) has significantly enhanced vehicle safety and driving convenience. In particular, active safety functions such as adaptive cruise control (ACC) and autonomous emergency braking (AEB) rely on the continuous operation of radar, camera, and sensor fusion systems to analyze surrounding traffic conditions and execute vehicle control actions. As these systems directly influence vehicle longitudinal and lateral control through electronic signals and real-time data processing, their functional reliability under electromagnetic disturbances has become a critical safety consideration [1, 2].

The susceptibility of automotive electronic systems to electromagnetic disturbances is addressed through electromagnetic compatibility (EMC) standards, such as ISO 11451 and UN ECE R10 [3], which define vehicle-level electromagnetic susceptibility (EMS) test procedures. Conventional vehicle-level EMS tests are primarily designed to verify whether a vehicle maintains normal operation under standardized electromagnetic field conditions. Under these test conditions, vehicles are typically evaluated in simplified driving states, without considering the realistic traffic scenarios in which ADAS functions are actually activated. Although these methods are effective for evaluating overall vehicle functionality, they do not adequately assess the operational performance of ADAS under realistic driving conditions [4].

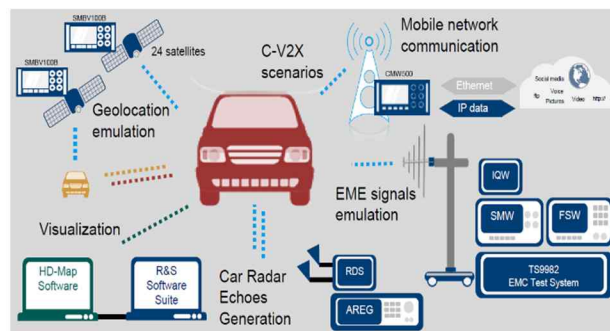


Fig. 1. Block diagram of the proposed scenario-based ADAS EMS test configuration

This limitation highlights the growing need for scenario-based EMS evaluation methods that assess ADAS

functional performance under realistic traffic conditions [5]. To address this need, this paper proposes a scenario-based EMS test method for ADAS functions that simulates realistic driving environments to verify the electromagnetic reliability of ACC and AEB functions under forward vehicle approach, deceleration, and stop scenarios. The proposed test configuration incorporates a radar target simulator (RTS) for generating virtual preceding-vehicle targets and a camera projection system for reproducing forward-vehicle and lane imagery, as illustrated in Fig. 1.

Under this test, the vehicle recognizes the simulated target as a constant-speed preceding vehicle with ACC activated, after which deceleration and stop events are introduced while electromagnetic susceptibility signals are applied under the same test conditions. The evaluation simultaneously monitors target recognition continuity, missed detection, and ghost target occurrence, as well as ACC control delay, AEB warning and braking intervention timing, and CAN signal parameters, including vehicle speed, deceleration, brake, and throttle status. Fig. 2 presents examples of preceding-vehicle recognition under the ACC test scenario.

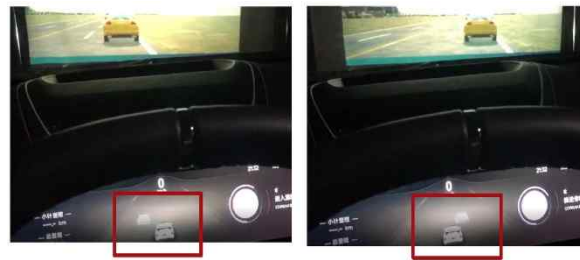


Fig. 2. Example of vehicle recognition status under the ACC test scenario

This paper proposed a scenario-based EMS test method for evaluating the functional reliability of ADAS under realistic driving conditions. By incorporating an RTS and a camera projection system, the proposed configuration enables the reproduction of forward vehicle approach, deceleration, and stop scenarios within a controlled electromagnetic test environment. The method allows simultaneous monitoring of preceding-vehicle recognition continuity, ACC control response, and AEB intervention timing alongside CAN signal parameters, providing a comprehensive basis for assessing ADAS electromagnetic reliability. The proposed approach is expected to contribute to the development of standardized EMS evaluation procedures for ADAS and autonomous driving functions in future mobility applications.

Acknowledgment

This research was supported by the Korea Institute for Advanced of Technology grant funded by the Korea government (MOTIR) (Project No. P0027952, Project Title: Establishment of Corporate Support Base to Response to Electromagnetic Environment Disturbances Simulating Actual Urban Environments)

References

1. H. Woo, *Journal of Auto-Vehicle Safety Association*, 13(2), 7-14 (2021).
2. S. Kim, J. Ryu, H. Woo and B. Yong, *Journal of Auto-Vehicle Safety Association*, 11(4), 50-56 (2019).
3. J. S. Park, J. P. Kim and D. Y. Yang, *Journal of Electromagnetic Engineering and Science*, 24(5), 477-484 (2024).
4. H. Woo, *Journal of Auto-Vehicle Safety Association*, 14(1), 14-19 (2022).
5. S. Hwang, W. Kim, H. Cho and T. Lee, *2025 KSAE Annual Autumn Conference*, 818 (2025).

Repeatability Factors in Electric Vehicle Single-Charge Range Tests Using Chassis Dynamometers

Jeong-gon Kim¹⁾, Myeong-Ho Song¹⁾, Hyeong-Jin Hwang¹⁾, and Jang-Kyu Yun¹⁾

¹⁾ Gyeongbuk Institute of IT Convergence Industry

*E-mail jgkim@gitc.or.kr

Abstract: This extended abstract reviews why the same electric vehicle can produce different single-charge range results on a chassis dynamometer. Although the dynamometer reproduces road load using coast-down-based coefficients, energy consumption can still vary with setup, calibration, vehicle control strategy, thermal condition, accessory load, and instrumentation. Based on Bucher et al., single-source vehicles generally showed low test-to-test variation, while hybrid operation and thermal testing produced larger deviations. The review identifies practical checkpoints for EV range repeatability: preconditioning, soak management, HVAC state, SOC and battery temperature stabilization, drive-cycle tracking, road-load verification, and power measurement traceability.

Keywords : Electric vehicle, Chassis dynamometer, Single-charge range, Repeatability, Energy consumption variation

Introduction: Electric vehicle range is determined from standardized chassis dynamometer tests, but different results may occur even under nominally identical conditions. Range depends on battery energy measurement, cycle following, road-load realization, environment, and accessory loads. As high-voltage HVAC and active battery thermal management become common, repeatability should be managed through controlled preparation, instrumentation, and vehicle-state checks.

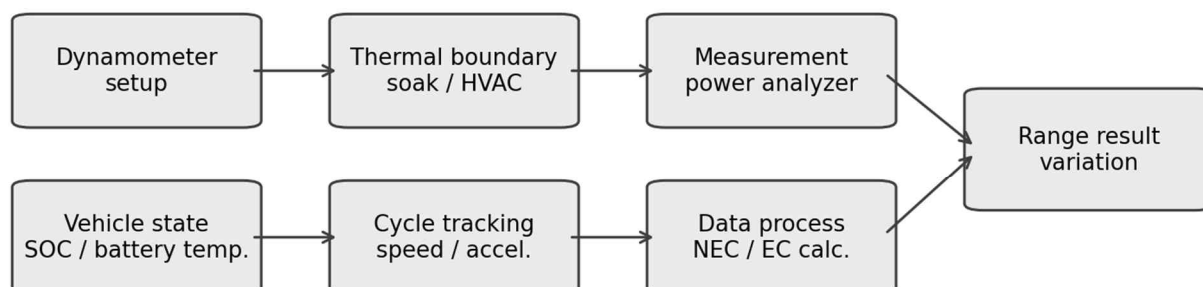
Research Methods: This work reviews Bucher et al. (2014), which analyzed energy consumption variation using UDDS, HWFET, and US06 cycles within the EPA Five-Cycle context. The reviewed study measured high-voltage battery current/voltage, HVAC current/voltage, vehicle speed, acceleration, and dynamometer energy. Variation was evaluated as sample standard deviation divided by sample average, and net energy change correction was used to reduce battery-energy imbalance effects in charge-sustaining hybrid tests. For BEV range testing, the same logic is translated into checkpoints for SOC, battery temperature, cabin soak, HVAC state, power analyzer wiring, and post-test energy balance.

Experimental Results: Table 1 summarizes the main mechanisms identified from the reviewed study and their implications for EV range testing. Bucher et al. reported that Ford Focus and Nissan Leaf test groups showed variation lower than 2% on UDDS and HWFET, whereas Chevrolet Volt results reached approximately 3-8% in charge-sustaining mode and 1-5% in full EV mode under comparable thermal cases. Net energy change correction reduced 95% confidence interval metrics by approximately 0.5-3% for charge-sustaining hybrid vehicles. During thermal testing, HVAC energy was nearly linearly related to electrical or fuel energy consumption, and different initial cabin temperatures increased HVAC power demand. Fig. 1 shows the proposed repeatability-management framework.

Table 1. Key repeatability factors for EV single-charge range testing

Variation factor	Observation from reviewed study	Implication for EV range tests
Road-load and dynamometer setup	Road-load coefficients are derived from coast-down tests and verified on the dynamometer.	Verify road-load calibration, coast-down reproduction, and roller control before repeat tests.
Cycle following and driver	Similar tests used the same cycle and conditions, but the same driver was not always assured.	Track speed deviation, acceleration behavior, and use a driving robot or stricter trace criteria when possible.
Battery energy balance	Net energy change affected hybrid results; correction reduced repeatability metrics.	Stabilize SOC and battery temperature and evaluate beginning/end battery energy balance.

Thermal and HVAC load	Thermal tests showed accessory-load sensitivity; HVAC energy correlated with total energy use.	Control cabin soak, HVAC mode, solar load, and battery thermal-management state.
Instrumentation and data process	Instrument uncertainty was not calculated in the reviewed paper.	Manage power analyzer calibration, channel wiring, current-sensor zeroing, and data synchronization.



Controlled checkpoints convert hidden test-to-test variability into traceable repeatability evidence.

Fig. 1. Repeatability-management framework for EV chassis dynamometer range testing

Conclusion: EV range repeatability should not be treated as a single fixed variation. The reviewed literature shows that vehicle state, thermal load, road-load setup, cycle tracking, and power measurement can change the measured energy consumption. Domestic EV range testing should therefore combine procedural controls with post-test variation analysis to improve confidence in single-charge range results.

Acknowledgment

This work was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE), Korea, through the Regional Innovation Cluster Development (R&D) Program (Project No. P0031740, Development of an Accessible 420km-Range Electric Low-Floor Bus with High-Safety Core Components).

References

1. J. Bucher, T. Bradley, H. Lohse-Busch and E. Rask, SAE Int. J. Alt. Power., 3(1), 36-43 (2014).
2. U.S. EPA Office of Transportation and Air Quality, Fuel Economy Labeling of Motor Vehicles: Revision to Improve Calculation of Fuel Economy Estimates, EPA420-R-06-017 (2006).
3. SAE International, Recommended Practice for Measuring the Exhaust Emissions and Fuel Economy of Hybrid-Electric Vehicles, Including Plug-in Hybrid Vehicles, SAE J1711, Rev. June 2010.
4. H. Lohse-Busch, K. Stutenberg, M. Duoba, E. Rask et al., Chassis Dynamometer Testing Reference Document, Argonne National Laboratory (2013).

Implementation of a Lightweight Machine Learning-Based Virtual Motor Speed Sensor Using a MEMS IMU

Jaeseong Lee¹⁾, Gwang-Min Park¹⁾, Seonghun Lee¹⁾

*¹⁾AX Research Group for Robotics, Daegu Gyeongbuk Institute of Science and Technology (DGIST),
Daegu 42988, Korea*

**E-mail: shunlee@dgist.ac.kr*

Abstract: This paper proposes a lightweight method for estimating electric motor speed using a MEMS inertial measurement unit (IMU) mounted on the motor housing, without additional position or speed sensors. Instead of conventional encoders or resolvers, which increase cost, mounting space, wiring complexity, and failure points, acceleration and angular-velocity signals from a PHM-grade IMU are used for RPM estimation. Time- and frequency-domain features are extracted from short sliding windows and fed to a lightweight machine-learning model. In a preliminary single-motor dynamometer study, the estimator tracked the CAN-based encoder reference over 0.2–2 kRPM with an RMS error of 1.5% in-sample and 3.7% on a time-based hold-out, while running in real time on an ESP32-WROOM-32 microcontroller. These results demonstrate the feasibility of a low-cost virtual RPM sensor that can complement or provide redundancy to conventional speed sensors.

Keywords : MEMS IMU, sensorless RPM estimation, lightweight machine learning, prognostics and health management

1. Introduction

Rotational speed is commonly measured using encoders or resolvers, but these sensors increase cost, space, wiring complexity, and failure points. Sensorless methods based on motor voltages or currents have been studied for BLDC and PMSM drives [1], [4], while vibration sensors and MEMS IMUs are widely used for condition monitoring and PHM [2]. Since vibration signals have also been used to estimate rotational speed in industrial machinery and gearboxes [3], this study investigates a housing-mounted MEMS IMU as a lightweight virtual RPM sensor without requiring a dedicated speed sensor or motor-specific electrical model.

2. Research Methods

An ICM-42605 six-axis IMU (3-axis accelerometer and 3-axis gyroscope) is fixed to the rear cover of the motor and sampled at 8 kHz. As shown in Fig. 1, the stream is processed in sliding windows of 2048 samples (256 ms). From each window a 104-dimensional feature vector is formed: 96 group power-spectral-density bins that capture the vibration harmonics scaling with speed, two absolute-energy terms that restore amplitude information at low speed, and six cross-axis orbit-sense terms carrying rotation-direction cues for future bidirectional work. The standardized features are mapped to the motor speed by a compact multilayer perceptron (104-48-24-1, ReLU activations) at 100 Hz, as demonstrated for gearboxes and industrial machinery [3]. The ground truth is the shaft speed measured by an external encoder and broadcast on the CAN bus (ID 0x100); this reference labels the windows for training and evaluation only and is not required at inference. Being dependency-free, the identical model has been deployed on an ESP32-WROOM-32 microcontroller, where it runs in real time at 100 Hz.

3. Experimental Results

As a preliminary validation, the method was tested on one 19.5-min dynamometer session (steady, ramp, and mixed profiles up to 2000 RPM; load varied within the session but was not analyzed separately). Fig. 2(a) shows that the estimated speed tracks the encoder-based CAN ground truth throughout the sweep, and Fig. 2(b) confirms tight agreement across the range. Over 43,284 windows the estimator attains an RMS error of 1.50 % in-sample and 3.71 % on a time-based hold-out (final 20 %), with a mean absolute error of

9.0 RPM. The error decreases with speed (Table 1): on the hold-out it is 5.1 %, 3.6 %, and 2.6 % for the 0.2–0.5, 0.5–1, and 1–2 kRPM bands. The aggregate hold-out RMS (3.71 %) and every band above 0.5 kRPM meet the ± 5 % target using housing vibration alone; only the very-low-speed band (0.2–0.5 kRPM) marginally exceeds it, remaining the main limitation.

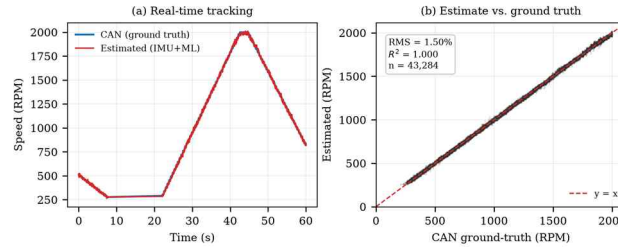


Fig. 1. (a) Estimated vs. CAN ground-truth speed over a representative 60-s segment; (b) estimate-vs-reference scatter over the full session (RMS = 1.50 %, $R^2 \approx 1.0$, $n = 43,284$).

Table 1. RPM-estimation accuracy on a single dynamometer session (preliminary; evaluation windows with $|RPM| \geq 200$; encoder-based CAN ground truth).

Metric	Value
RMS error — in-sample / time hold-out	1.50 % / 3.71 %
RMS by speed, hold-out (0.2–0.5 / 0.5–1 / 1–2 kRPM)	5.1 / 3.6 / 2.6 %
Mean absolute error (in-sample)	9.0 RPM
Coefficient of determination, R^2 (in-sample)	0.9995
Evaluated range; windows	0.2–2 kRPM; 43,284

4. Conclusion

This preliminary study shows that a lightweight IMU-plus-machine-learning virtual sensor estimates motor speed within 1.5 % in-sample (3.7 % on a time-based hold-out) across 0.2–2 kRPM, using only the vibration observed by a PHM-grade IMU on the motor housing and no dedicated speed sensor, and that the identical model runs in real time on an ESP32-WROOM-32. This supports its use as a low-cost complement, backup, or redundant speed sensor alongside encoders and resolvers. Future work will validate bidirectional operation and very low speeds (<0.2 kRPM) and assess generalization across different motors and load conditions.

Acknowledgment

This research was supported by the Industrial Innovation Infrastructure Development Program, “Mobility Motor Innovation Technology Development” (RS-2024-00459503), funded by the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea. The authors thank the supporting institutions and project members for their valuable insight and technical support, which greatly assisted this research.

References

1. J.-C. Gamazo-Real, V. Martínez-Martínez, and J. Gomez-Gil, “ANN-based position and speed sensorless estimation for BLDC motors,” *Measurement*, vol. 188, Art. no. 110602, 2022, doi: 10.1016/j.measurement.2021.110602.
2. I. U. Hassan, K. Panduru, and J. Walsh, “An in-depth study of vibration sensors for condition monitoring,” *Sensors*, vol. 24, no. 3, Art. no. 740, 2024, doi: 10.3390/s24030740.
3. E. Gildish, M. Grebshtein, and I. Makienko, “Vibration-based rotation speed estimation for Industrial IoT,” *Internet of Things*, vol. 25, Art. no. 101024, 2024, doi: 10.1016/j.iot.2023.101024.
4. H. Mohan, G. Agrawal, V. Jatly, A. Sharma, and B. Azzopardi, “Neural network-driven sensorless speed control of EV drive using PMSM,” *Mathematics*, vol. 11, no. 19, Art. no. 4029, 2023, doi: 10.3390/math11194029.

Torque Robustness Enhancement in PMSM Drives with Low-Resolution Sensors via Active Phase Compensation

Gwangmin Park¹⁾, Jaesung Lee¹⁾ and Seonghun Lee¹⁾

¹⁾AX Research Group for Robotics, Daegu Gyeongbuk Institute of Science & Technology (DGIST),
333, Techno Jungang Daero, Hyeonpung-Eup, Daegu, Korea

E-mail: shunlee@dgist.ac.kr

Abstract: This paper proposes an active phase compensation technique to improve the speed control performance of permanent magnet synchronous motor (PMSM) drives using low-resolution sensors. Strong phase-locked loop (PLL) filtering and load fluctuations induce non-linear spatial phase lags, reducing the available voltage margin and restricting the maximum torque capability. The proposed method estimates d -axis voltage deviations in real-time and compensates for the inverter phase via adaptive phase control. Simulation results verify that the proposed technique effectively improves angle estimation accuracy and enhances torque responsiveness, thereby improving overall control performance.

Keywords : PMSM, Torque control, Low-cost sensor, Robustness, Phase Trimming

1. Introduction

In field-oriented control (FOC) for PMSM drives using low-resolution position sensors, a phase-locked loop (PLL) is generally employed to increase electrical angle resolution. However, strong PLL filtering during high-frequency switching induces non-linear spatial phase lags. This angle error increases significantly under heavy load operations and field-weakening control, causing premature voltage saturation and severely restricting maximum torque capability. Resolving this phase estimation uncertainty is critical to unlocking full motor performance without relying on expensive high-resolution encoders.

To address these limitations, the proposed active phase compensation technique adaptively tracks d -axis voltage deviations using a real-time controller. By minimizing the d -axis voltage control error, the method automatically trims the inverter output phase. This adaptive approach recovers the lost voltage margin while maintaining control robustness against motor parameter variations. The mathematical model of the PMSM in the true rotor d - q reference frame defines the steady-state voltage equations. When spatial misalignment occurs due to sensor delay, the controller reference frame shifts by an angle δ relative to the true rotor frame. This phase error causes the back-electromotive force (back-EMF), which naturally lies on the q -axis, to project onto the controller d -axis. The theoretical d -axis voltage deviation (v_{d_error}) can be approximated as proportional to the back-EMF and the sine of the phase error:

$$v_{d_error} \approx -\omega_e \lambda_m \sin(\delta)$$

This physical projection principle implies that tracking the d -axis voltage error allows for the real-time estimation and compensation of the unmeasured spatial phase lag δ .

2. Active Phase Compensation Method

The proposed method adaptively tracks these d -axis voltage deviations using a closed-loop real-time controller. First, the expected d -axis voltage is calculated without the cross-coupling terms affected by the phase delay. The difference between the actual inverter voltage command and this expected value yields the isolated v_{d_error} .

To compensate for the delay, an integral controller drives the error to zero by accumulating a phase trim angle:

$$\theta_{trim}(k) = \theta_{trim}(k-1) - K_{trim} \cdot v_{d_error} \cdot \text{sgn}(\omega_e) \cdot T_s$$

where K_{trim} is the gain constant for phase compensation, T_s is the control period, and $\text{sgn}(\omega_e)$ is the polarity coefficient to ensure the consistency of the closed-loop convergence direction according to the motor's rotation direction. The continuously updated phase compensation angle θ_{trim} compensates for the controller's electrical lead angle, improving the real-time control resolution and preventing d -axis back-EMF leakage.

3. Simulation Results

Fig. 1 presents the simulation results comparing the performance of the conventional baseline FOC method with the proposed FOC drive incorporating the active phase compensation, based on actual motor parameters and low-resolution position sensors with 28 poles or fewer. In Fig. 1(a) and (b), the baseline control method exhibits a d -axis voltage magnitude error of up to 1 V due to the phase estimation uncertainty. However, when the active phase compensation controller is applied, the steady-state error converges to zero as the compensation gain increases. The gain constant for phase compensation K_{brm} minimizes the voltage loss, thereby minimizing voltage limit variability.

Fig. 2 compares the transient torque responsiveness with and without field-weakening (FW) compensation control when the rated load torque is applied at a constant speed of 2,000 rpm. The results confirm that the compensation control reduces torque ripples and shortens the settling time.

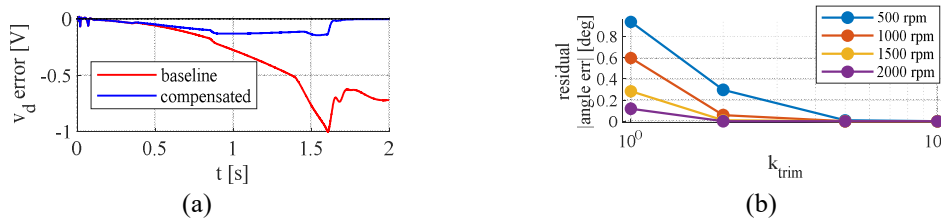


Fig. 1. (a) d -axis voltage error (b) Residual angle error.

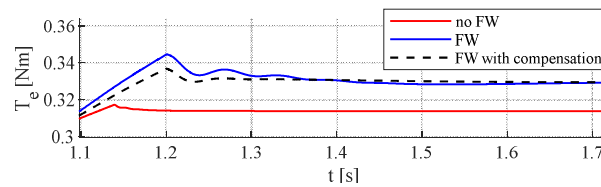


Fig. 2. Torque Response.

4. Conclusion

This paper proposed and verified an active phase compensation technique to overcome real-time spatial phase lag uncertainties in low-resolution Hall-sensor-based PMSM drives. While conventional sensorless control methods suffer from degraded control performance at low speeds or during abrupt heavy-load startup due to insufficient back-EMF amplitude, the proposed algorithm ensures zero- and low-speed drive stability using a base-angle generation strategy through the combination of a low-cost Hall sensor and a PLL. It also selectively compensates for spatially nonlinear phase lag in the high-speed region using a d -axis voltage deviation-based adaptive compensator. This hybrid approach is expected to be valuable for industrial motor control systems that require both cost competitiveness and robust torque capability near the voltage limit.

Acknowledgment

This research was supported by the Start-up fund of Daegu Gyeongbuk Institute of Science and Technology (2026010342) and the development project of the auto-stow electric column mechanism system for SbW with a long-stroke ball sliding structure (RS-2023-00280208).

References

1. J. S. Park and K.-D. Lee, "Online Advanced Angle Adjustment Method for Sinusoidal BLDC Motors with Misaligned Hall Sensors," *IEEE Trans. Power Electron.*, 32(11), 8247–8253 (2017).
2. H.-J. Lee and J.-G. Shon, "Improved Voltage Flux-Weakening Strategy of Permanent Magnet Synchronous Motor in High-Speed Operation," *Energies*, 14(22), 7464 (2021).
3. M. Akrami, E. Jamshidpour, L. Baghli and V. Frick, "Application of Low-Resolution Hall Position Sensor in Control and Position Estimation of PMSM—A Review," *Energies*, 17(17), 4216 (2024).

Development of a Real-Time Integrated Monitoring System for Battery Pack Manufacturing Process

MyeongSu Jeong¹⁾, ChangSoo Moon^{2*)} and GiDong Lee³⁾

^{1,2)} Dept. of Vehicle Software, Gyeongbuk Research Institute of Vehicle Embedded Technology, YeongCheon, Gyeongsangbuk-do 38808, Korea

³⁾ Production Technology Research Center, HWAHIN, YeongCheon, Gyeongsangbuk-do 38808, Korea.

**E-mail: cs.moon@givet.re.kr*

Abstract: This paper proposes an integrated monitoring system based on a 3D process line model for the real-time supervision of battery pack manufacturing processes. The battery pack housing manufacturing process consists of multiple stages, and the data generated at each stage are often distributed across different systems, making integrated management difficult. To address this issue, the proposed system links a 3D line model with PLCs, a database, and an X-SCADA platform so that production, quality, and equipment information can be managed in an integrated manner on a single dashboard. In addition, by standardizing process data and maintaining historical records, the system enables intuitive monitoring of process status and supports root-cause analysis and quality control in the event of anomalies. The proposed system improves the visibility of battery pack manufacturing processes and provides a foundation for data-driven operation.

Keywords : Battery pack manufacturing process, Real-time monitoring, Quality management, PLC, 3D modeling, Data integration

In the first paragraph, Introduction: With the expansion of the electric vehicle market, ensuring the safety and reliability of battery pack housings has emerged as an important technical challenge. The battery pack manufacturing process consists of multiple inspection and assembly stages, and the data generated at each stage are frequently distributed across heterogeneous formats and systems. As a result, it is difficult to grasp the overall line status in real time and to perform cross-process correlation analysis, and a significant amount of time is required for root-cause analysis when quality issues occur. This study proposes an integrated monitoring system that combines a 3D process line model with real-time data acquisition and visualization to address these challenges.

In the second paragraph, Research Methods: The proposed monitoring system is designed with a four-layer architecture consisting of a 3D process line model, PLC-based data acquisition, database storage, and X-SCADA-based visualization. The 3D model is constructed at the process-cell level using iCAD based on 2D drawings, and the line flow is modeled to reflect the actual equipment layout as closely as possible. PLCs collect status information and inspection data generated at each process, and the collected data are stored in the database through a standardized tag scheme. In the X-SCADA screens, a 3D line view is displayed together with production status, quality information, and equipment conditions, and screen elements are updated in real time according to changes in tag values. This configuration enables integrated management of not only detailed information at the individual process level but also the operational status of the entire line.

In the third paragraph, Experimental Results: The implemented system was deployed on an actual battery pack housing production line to verify its functionality and usability. When an operator selects a specific process, the corresponding equipment is highlighted in the 3D view and, at the same time, the production quantity, work-in-process quantity, scrap quantity, and key quality indicators for that process can be checked immediately. In addition, charts are configured so that production performance and cycle time by line can be compared over different time periods, allowing intuitive analysis of changes in production flow. By checking process-wise quality histories together with equipment status, the operator can quickly identify abnormal sections and effectively detect bottleneck processes or processes with

recurring defects. These results demonstrate that the proposed system can serve as a practical support tool for improving productivity and quality management.

In the fourth paragraph, Conclusion: This study designed and implemented a real-time integrated monitoring system for battery pack manufacturing processes and presented an architecture that links a 3D line model with PLCs, a database, and an X-SCADA platform. The proposed system integrates production, quality, and equipment data that were previously dispersed across processes, thereby enabling intuitive monitoring of the overall line status.



Fig. 1. Monitoring system for the battery pack manufacturing process

In the future, the system will be extended to a digital twin-based smart factory platform by incorporating OPC UA and Asset Administration Shell (AAS)-based data models as well as AI-based anomaly detection functions.

Acknowledgment

This research was supported by Korea Industrial Complex Corporation (KICOX), funded by the Ministry of Trade, Industry and Energy (MOTIE) in 2025(VCDM2502). We thank our colleagues from Korea Industrial Complex Corporation(KICOX) who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

References

1. G.Y. Kim, G. Jung, S.H. Ahn, A Real-time and Low-cost Machine State Monitoring System based on AC Power Signal, Proceedings of the Korean Society for precision Engineering Conference, p.310 (2022)
2. W. Ki, Development Trend of Battery Safety for Electric Vehicle, Auto Journal, p.1664(2022)
3. G.S.Park, Development of Factory Monitoring System in Real Time Using IoT Device, Master's Thesis (2019)
4. S.S.Kim and T.G.Yoon, A Study on Equipment Monitoring Systems in Manufacturing Plants, p.206-207 (2021)
5. Yuan Guo, Yu Sun, and Kai Wu, Research and development of monitoring system and data monitoring system and data acquisition of CNC machine tool in intelligent manufacturing (2020)

An Explainable Client-Side Framework for Real-Time Phishing Detection Using URL Risk Scoring and Safe Browsing APIs

Min-Gi Choi¹⁾, Sol-mi Kim²⁾ and Won-Kee Hong^{3 *)}

^{1,2,3)} Dept. of Inform. Security, Daegu Univ., Gyeongsan, Gyeongbuk 38453, Korea

*E-mail: wkhong@daegu.ac.kr

Abstract: Phishing attacks have evolved into sophisticated web-based threats that bypass traditional blacklist approaches. This paper proposes APG (Anti-Phish-Guard), a lightweight browser extension that detects phishing sites in real time using a hybrid approach combining Google Safe Browsing intelligence with an explainable rule-based URL risk scoring mechanism. All detection processes are performed on the client side without relying on external servers, preserving user privacy and ensuring fast response. The proposed system classifies URLs into safe, suspicious, or dangerous and provides immediate feedback and blocking before access. Experimental implementation on the Chrome browser demonstrates that APG achieves fast response times, low resource consumption, and improved capability to detect previously unseen phishing websites.

Keywords : Phishing detection; URL risk scoring; Safe Browsing API; Client-side security; Browser extension

Introduction: Phishing attacks have increasingly exploited web technologies such as search engine advertisements, typo-squatted domains, and counterfeit SSL certificates to steal user credentials and financial information[1,2]. Blacklist-based solutions fail to detect newly created phishing sites, while machine learning approaches often require server-side processing, large datasets, and lack transparency. To address these limitations, this study proposes APG, a client-side browser extension that provides real-time phishing detection using explainable URL risk scoring combined with Google Safe Browsing intelligence[3]. The system aims to enhance security, privacy, and user awareness by operating locally within the browser.

Hybrid URL Risk Analysis Framework: APG integrates two complementary detection engines: (1) Google Safe Browsing API for known malicious URLs and (2) a rule-based URL risk scoring engine for previously unseen threats.

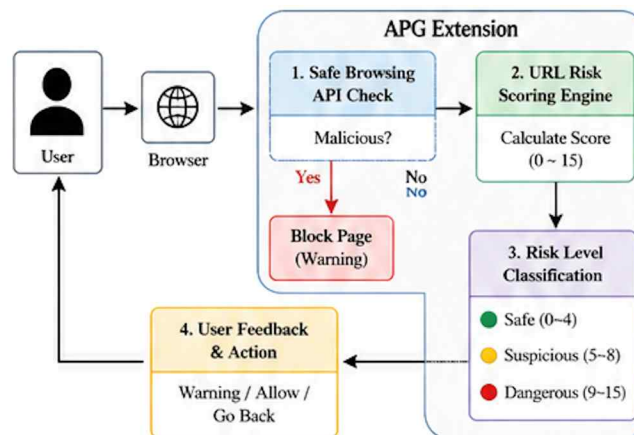


Fig. 1. APG operational flow

Figure 1 illustrates the operational flow of APG. When a user accesses a web page, the extension extracts the URL and queries the Safe Browsing API[4]. If the URL is reported as malicious, the system immediately blocks the page. Otherwise, the URL is evaluated by the risk scoring engine, which computes a total score

based on five criteria: domain similarity, HTTPS validity, suspicious keywords, presence of special characters, and use of URL shorteners. The total score determines the risk level: Safe (<4), Suspicious (5-8), or Dangerous (≥ 9). Users receive color-coded warnings and can choose to proceed or return

$$S_{total} = \sum_{i=1}^5 w_i \cdot s_i$$

where w_i is the weight and s_i is the score of the i -th criterion. All processing is performed locally in the browser, eliminating network latency and preserving user privacy

Experimental Results: APG was implemented as a Chrome extension using JavaScript and the Google Safe Browsing API. The system was evaluated using 120 URLs, including 70 known phishing sites from PhishTank[5] and 50 newly created phishing sites collected during the evaluation period. Table 1 compares APG with representative existing methods in terms of detection capability, response time, explainability, and client-side operation.

Table 1. Comparison with existing phishing detection methods.

Method	Blacklist (Phish Tank)	Machine Learning (server-based)[6]	Heuristic (Plugin)[7]	APG (proposed)
Detection of New Phishing Sites	Low	Medium	Medium	High
Explainability	Low	Low	Medium	High (Rule-based)
Client-side Operation	No	No	No	Yes

As shown in Table 1, APG provides the fastest response time while maintaining high explainability and operating completely on the client side. Figure 1 demonstrates the integrated flow of Safe Browsing API and risk scoring, enabling the detection of both known and previously unseen phishing URLs with transparent reasoning.

Conclusion: This paper presented APG, a client-side browser extension that combines Safe Browsing intelligence with an explainable URL risk scoring engine for real-time phishing detection. Experimental evaluation shows that APG achieves fast response, high detection capability for new phishing sites, and full client-side operation. The system improves user security and privacy without relying on external servers. Future work will extend APG to other browsers and mobile environments and incorporate adaptive scoring using user feedback.

References

1. R. Tanti, Study of phishing attack and their prevention techniques. *International Journal of Scientific Research in Engineering and Management*, 8(10), 1-8, (2024).
2. R. K. Ayeni, A. A. Adebiyi, J. O. Okesola, E. Igbekele, Phishing attacks and detection techniques: A systematic review. In *2024 international conference on science, engineering and business for driving sustainable development goals*, 1-17, (2024).
3. Y. Kim and J. Lee, Development of a Malicious URL Machine Learning Detection Model Reflecting the Main Feature of URLs. *Journal of the Korea Institute of Information & Communication Engineering*, 26(12), (2022).
4. Google, "Google Safe Browsing API," <https://developers.google.com/safe-browsing/> (Accessed: May 2025).
5. R. Lohar, M. Tripathi and H. Shirazi, Free-Phish: detecting phishing websites hosted on free web hosting domains. *Int'l Journal of Computers and Applications*, 47(2), 162-168, (2025).
6. V. Shahrivari, M. Darabi and M. Izadi, Phishing detection using machine learning techniques. *arXiv preprint arXiv:2009.11116*, (2020).
7. L. Bilge, T. Dumitras, C. Kruegel, E. Kirda, and G. Vigna, "Protecting users against phishing attacks with username similarity detection," in *Proc. USENIX Security Symp.*, 89-104 (2009).

An Embedded AI Software Integration Supporter for Safe Human-in-the-Loop Code Modification in Ubuntu

Jae-Mo Ahn¹⁾, Jung-Hae Choi²⁾, Sijia Huo¹⁾, Byoung-Ju Yun³⁾, and Byung-Jae Choi^{3*)}

1) Department of Electrical Engineering, Daegu University, Gyeongsan 38453, Republic of Korea

2) Chief Executive Officer, Daekyung Elec Co., Ltd., Republic of Korea

3) School of Electronics Engineering, College of IT Engineering, Kyungpook National University, Daegu, Republic of Korea (E-mail: bjisyun@ee.knu.ac.kr)

**E-mail: bjchoi@daegu.ac.kr*

Abstract: Embedded software maintenance requires coordinated inspection of code, configuration files, middleware, and execution results. The Embedded AI Software Integration Supporter(EASIS) is an Ubuntu-based assistant that links a conversational interface to ROS 2 modules for workspace inspection, cross-file inconsistency detection, solution explanation, user-approved patching, backup, and verification. Unlike a general chat assistant, it operates only inside a selected local workspace and applies no change before the user reviews the problem, risk, solution, and diff. The prototype was implemented with Python 3.10.12, ROS 2 Humble, Tkinter, and RQt. In a four-file demonstration project, EASIS detected three intentional inconsistencies. For a steering-ratio conflict, it changed 11.7 to 15.0 after approval, created a backup, passed syntax verification, and confirmed issue removal. The result demonstrates a traceable and reversible embedded-development support workflow.

Keywords: conversational AI, embedded system, Ubuntu, ROS 2, code analysis, automatic patching, human-in-the-loop

Introduction: Modern embedded projects combine firmware, Python or C/C++ nodes, launch files, YAML parameters, serial interfaces, and simulation tools. A small parameter change may therefore affect several dependent files, while uncontrolled editing can introduce failures or unsafe commands. EASIS is an Ubuntu-resident maintenance platform that converts a natural-language request into a controlled inspect-analyze-explain-approve-patch-verify pipeline. It is not a hardware controller; its current safety policy blocks destructive commands and direct actuation. The goal is to provide transparent, reversible assistance from developer intent to verified code modification.

System Design and Implementation: Fig. 1 shows the RQt communication graph of EASIS. The interaction layer contains the conversational interface and workspace inspector. The analysis layer uses the code analyzer, solution planner, and safety monitor to identify inconsistencies and generate an explainable proposal. The execution layer contains the approval gate, patch manager, verification manager, and result analyzer. Execution is blocked until the user accepts the displayed solution. The patch manager then creates a timestamped backup and applies only the proposed atomic edit; verification checks syntax and issue removal, and failure requests rollback. Deterministic rules are used for reproducible testing, while a future LLM module can use the same approval and safety pipeline.

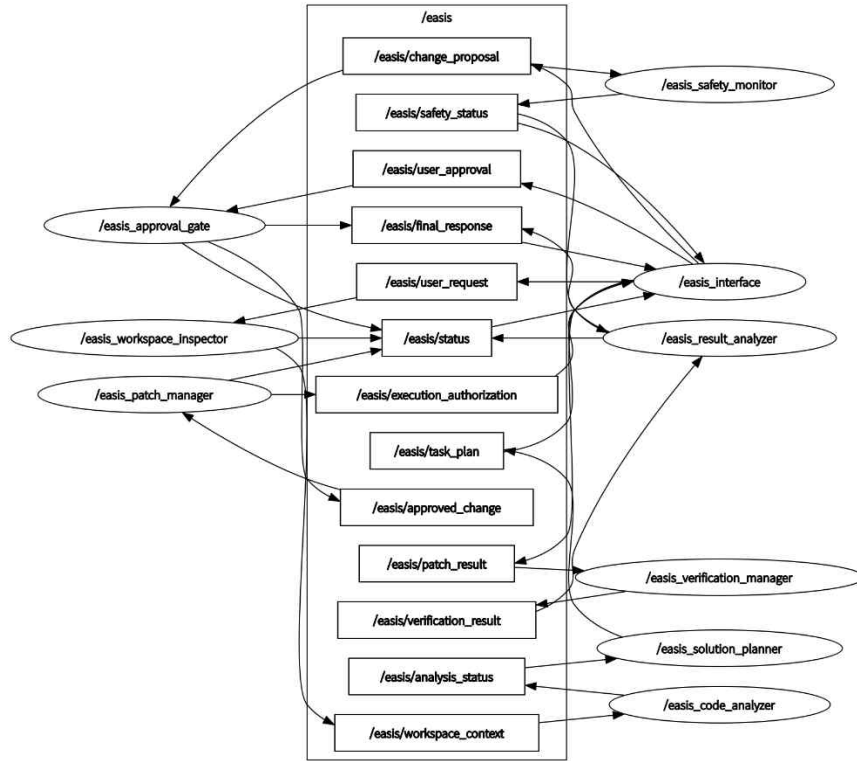


Fig. 1. Detailed RQt communication graph of the EASIS analysis, safety evaluation, user approval, patching, and verification pipeline.

Experimental Results: The test workspace contained README.md, controller.py, serial_bridge.py, and vehicle.yaml. Three faults were inserted: steering-ratio mismatch (11.7 versus 15.0), maximum-steering mismatch (20.0 degrees versus 15.0 degrees), and serial-port mismatch (/dev/ttyACM1 versus /dev/ttyACM0). EASIS detected all three and prioritized the steering-ratio conflict. Before modification, it displayed file evidence, the estimated 28.2% conversion difference, a recommended correction, a safety plan, and a unified diff, as shown in Fig. 2(a). After approval, it created a timestamped backup, changed STEERING_RATIO to 15.0, verified two Python files, and returned PASS for syntax and issue-removal checks. Re-scanning reduced unresolved issues from three to two and selected the next conflict, as shown in Fig. 2(b). No external file or hardware command was changed.

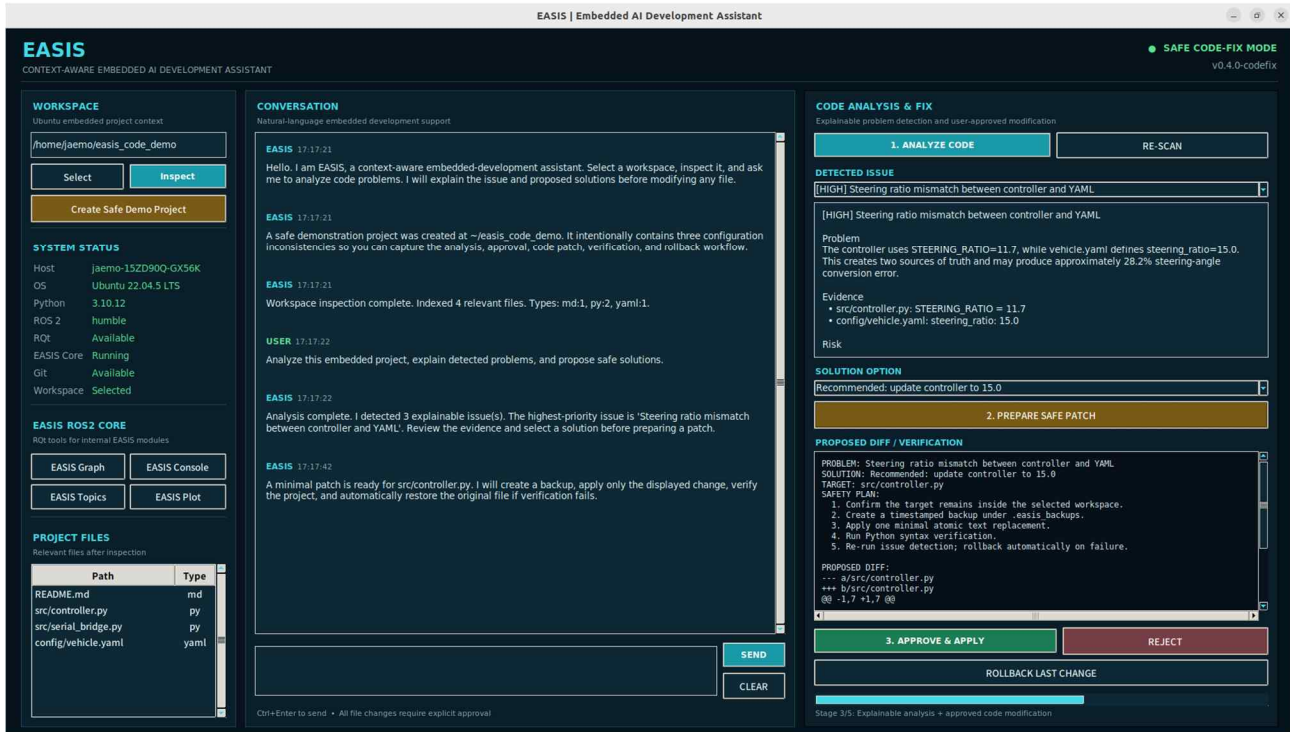


Fig. 2. EASIS code-fix workflow: (a) diagnosis and proposed diff before approval; (b) successful patch verification and detection of the next issue.

Conclusion: EASIS provides conversational interaction, ROS 2 modular processing, explicit approval, reversible patching, and automatic verification in one Ubuntu interface. The prototype detected three inconsistencies and completed one approved correction with backup and PASS verification. Future work will add LLM reasoning, controlled builds, and simulation execution.

Acknowledgment

This research was supported by the Gyeongsangbuk-do RISE (Regional Innovation System & Education) project.

References

1. C. E. Jimenez et al., “SWE-bench: Can Language Models Resolve Real-World GitHub Issues?,” Proc. ICLR, 2024.
2. J. Yang et al., “SWE-agent: Agent-Computer Interfaces Enable Automated Software Engineering,” Advances in Neural Information Processing Systems, vol. 37, 2024.
3. X. Wang et al., “OpenHands: An Open Platform for AI Software Developers as Generalist Agents,” arXiv:2407.16741, 2024.
4. M. Quigley et al., “ROS: an Open-Source Robot Operating System,” ICRA Workshop on Open Source Software, 2009.

Design and Implementation of a Real-Vehicle VIL Testbed for Autonomous Driving Validation with Virtual Sensor Integration

Je-Seok Kim¹⁾

²⁾ Daegu Gyeongbuk Institute of Science & Technology (DGIST), Daegu, Korea

**E-mail: jeseok@dgist.ac.kr*

Abstract: This paper describes a real-vehicle-based Vehicle-in-the-Loop (VIL) testbed for autonomous driving validation with virtual sensor integration. The testbed keeps the actual vehicle platform and control interface in the loop, and perception inputs can be switched between physical sensors and virtual sensors. The system consists of an autonomous driving software stack, virtual LiDAR and camera generation modules, a positioning reference system, a scenario simulation environment, and CAN/Ethernet interfaces. Three validation modes are organized: real-sensor proving-ground testing, virtual-sensor testing with the real vehicle still in operation, and virtual-road scenario testing for extended validation. The implementation shows that a real autonomous driving vehicle can be connected with virtual sensors and scenarios without changing the main software and control interface. This approach can reduce dependence on repeated proving-ground driving and provides a more repeatable way to validate autonomous driving functions.

Keywords : Vehicle-in-the-Loop, autonomous driving validation, virtual sensor, real vehicle, scenario test

Introduction: Autonomous driving validation requires many repeated tests, but proving-ground tests are not always easy to repeat under the same conditions. Pure simulation is useful for scenario variation; however, it may not include the actual vehicle platform, CAN/Ethernet interface, and controller behavior. For this reason, a VIL testbed is useful because the real vehicle remains in the loop while some parts of the environment can be generated virtually. Previous studies have also investigated VIL or HIL-based approaches for autonomous driving validation and simulation [3]–[6]. Scenario-based safety evaluation is also becoming important for automated driving systems [1], and the validation of planning and control functions remains an ongoing issue [2]. In this study, the main purpose is to develop a testbed in which the same autonomous driving software can use either real sensor data or virtual sensor data. The target application is the validation of perception-to-control functions before conducting many real proving-ground runs.

Research Methods: The proposed system is composed of five main parts: an autonomous driving software stack, virtual LiDAR and camera generation modules, a positioning reference system, a scenario simulation environment, and CAN/Ethernet vehicle interfaces. The input layer of the software is designed so that LiDAR and camera topics can be switched between physical sensors and virtual sensors. The map data, localization reference, planning module, and vehicle control interface are kept in the same structure. With this design, the effect of a scenario or sensor input can be evaluated while the vehicle actuators and low-level control path are still operated as a real system. Time synchronization and coordinate conversion are also considered because virtual objects must be consistent with the real vehicle pose. Fig. 1 shows the overall connection, and Table 1 summarizes the three validation modes.

Experimental Results: In the real-sensor mode, the software receives LiDAR, camera, and GPS/IMU data from the sensors installed on the vehicle. In the virtual-sensor mode, LiDAR point-cloud and camera-image inputs are replaced with simulated data, but the vehicle, map data, software stack, and control interface remain unchanged. The virtual-road mode extends the same VIL structure to a simulated road environment for scenario testing. The implemented connection showed that the autonomous driving stack can operate with the real vehicle and virtual perception inputs within the same test flow. The result is an engineering-level testbed for more repeatable validation, although a quantitative comparison of sensor-model accuracy

will be studied further.

Conclusion: This paper presented the design and implementation of a real-vehicle VIL testbed with virtual sensor integration. The testbed can connect a real autonomous driving vehicle to virtual LiDAR, camera, and scenario data while keeping the vehicle control interface in the loop. It is expected to reduce unnecessary repeated proving-ground runs and to support more systematic validation of autonomous driving functions. Future work will include sensor-model calibration, timing-accuracy evaluation, and scenario-coverage analysis.

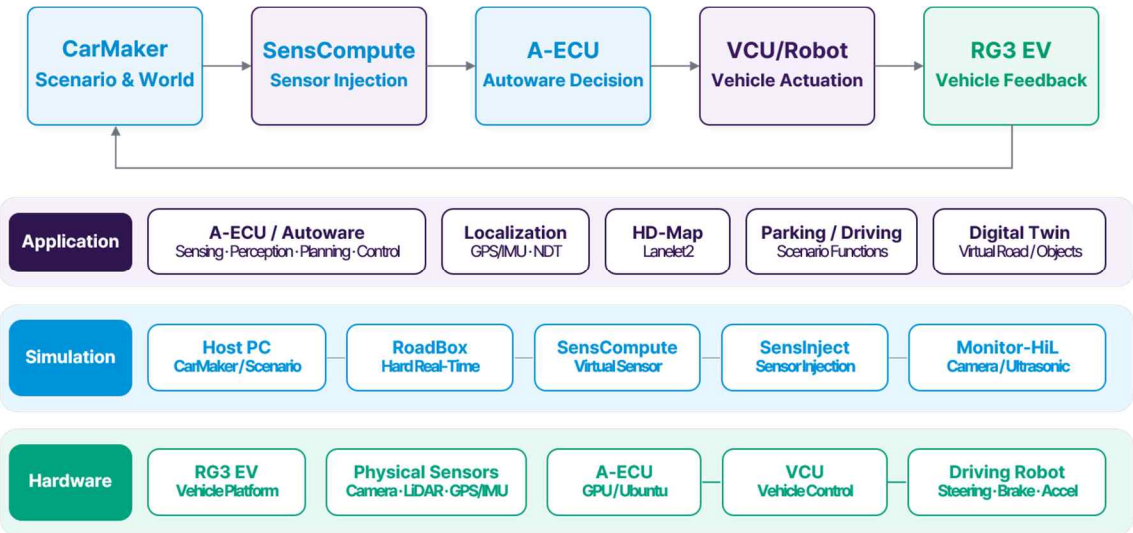


Fig. 1. Configuration of the real-vehicle VIL testbed

Table 1. Validation modes of the proposed testbed

Mode	Sensor input	Road / vehicle condition	Main purpose
Real-sensor mode	Physical LiDAR, camera, GPS/IMU	Real vehicle on proving ground	Baseline test with real sensor data
Virtual-sensor mode	Virtual LiDAR and camera, real position reference	Real vehicle still operated	Repeatable perception-input validation
Virtual-road mode	Virtual sensors from scenario simulation	Simulated road environment with VIL structure	Extended scenario validation

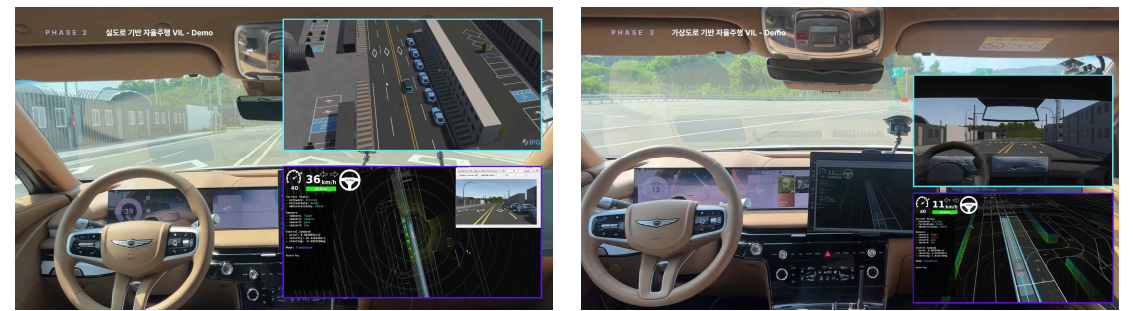


Fig. 2. Real vehicle test scene for VIL validation

Acknowledgment

This work was supported by the DGIST R&D Program of the Ministry of Science and ICT (26-IT-01) and by the Commercialization Promotion Agency for R&D Outcomes through the Ministry of Science and ICT (RS-2025-02634821).

References

1. ISO 34502:2022, Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework, International Organization for Standardization, 2022.
2. ISO 34503:2023, Road Vehicles — Test scenarios for automated driving systems — Specification for operational design domain, International Organization for Standardization, 2023.
3. C. Park, S. Chung, and H. Lee, “Vehicle-in-the-Loop in Global Coordinates for Advanced Driver Assistance System,” *Applied Sciences*, vol. 10, no. 8, p. 2645, 2020.
4. C. Brogle, C. Zhang, K. L. Lim, and T. Bräunl, “Hardware-in-the-Loop Autonomous Driving Simulation Without Real-Time Constraints,” *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 3, pp. 375–384, 2019.
5. Y. Chen, S. Chen, T. Xiao, S. Zhang, Q. Hou, and N. Zheng, “Mixed Test Environment-based Vehicle-in-the-loop Validation - A New Testing Approach for Autonomous Vehicles,” in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1283–1289, 2020.
6. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An Open Urban Driving Simulator,” in *Proceedings of the 1st Annual Conference on Robot Learning, Proceedings of Machine Learning Research*, vol. 78, pp. 1–16, 2017.

Design and Implementation of an AI-Based Smart Aquaculture System for Feeding Optimization and Automated Farm Management

Jaeyoung Kim^{1 *}, Dongseung Shin²⁾ and Laekyoung Lee³⁾

*^{1,2,3)} Daegu-Gyeongbuk Research Divison, Electronics and Telecommunications Research Institute,
Daegu, Republic of Korea*

**E-mail: jaeyoung@etri.re.kr*

Abstract: Smart aquaculture technologies are becoming increasingly important for improving productivity and reducing operational costs. This paper presents the design and implementation of an AI-based smart aquaculture system integrating fish monitoring, feeding optimization, and automated feeding control. The proposed system collects environmental data, aquaculture operation records, and fish-related information through a unified platform. An edge-based control system automatically manages feeding devices and tank facilities according to operational conditions. A prototype system was implemented and deployed in a commercial aquaculture environment. The results demonstrate the feasibility of integrating artificial intelligence, IoT sensing, and automated control technologies for practical aquaculture management.

Keywords : Smart Aquaculture, Artificial Intelligence, Edge Computing, Automated Feeding

Introduction: Smart aquaculture technologies are increasingly being adopted to improve productivity and reduce operational costs [1,4]. However, conventional aquaculture management still relies heavily on manual observation and feeding operations. Recent advances in AI-based video analytics and computer vision technologies have enabled automated fish monitoring and biomass estimation in aquaculture environments [1,2]. This paper presents an AI-based smart aquaculture system integrating environmental monitoring, feeding optimization, and automated feeding control based on the concept of smart aquaculture technologies [3]. The proposed system provides a practical platform for intelligent aquaculture management and future AI-driven fish monitoring.

System Design: The proposed smart aquaculture system was developed to improve farm management by integrating environmental monitoring, feeding optimization, and automated feeding control into a unified platform. IoT-based sensors continuously collect environmental parameters, including water temperature, dissolved oxygen, salinity, air temperature, and humidity, while farming information such as fish population, average body weight, feed consumption, and mortality is stored in a centralized database. Based on these data, the system analyzes farming conditions to determine optimal feeding schedules and feed quantities. The generated feeding information is transmitted to an edge-based controller, which operates automatic feeders and other aquaculture equipment in real time. In addition, a management platform provides real-time visualization of environmental conditions, feeding records, and system status through an intuitive user interface. By integrating monitoring, data management, and automated control, the proposed system provides an efficient platform for intelligent aquaculture management and supports future AI-based fish monitoring and feeding applications. We evaluated GroundingDINO+SAM, Mask R-CNN, and Mask2Former for segmenting spotted halibut in aquaculture tank images for this system. Among these models, Mask2Former achieved the best performance in densely populated scenes with severe object overlap. Its masked-attention-based architecture, combined with post-processing techniques such as small-mask filtering, watershed-based overlap separation, and confidence thresholding, enabled stable instance segmentation, making it the most practical choice for real-world aquaculture applications.

Implementation and Results: A prototype of the proposed smart aquaculture system was implemented and deployed in a commercial aquaculture environment. The developed platform integrates environmental monitoring, feeding management, and automated control functions into a unified system. Environmental data collected from various sensors were stored in a centralized database and visualized through a user

interface for real-time monitoring and management. The feeding management module supports the configuration of feeding schedules, feed quantities, and operational modes. The generated feeding information is transmitted to an edge-based controller, which operates automatic feeding devices and tank control equipment. During field operation, stable communication between sensors, servers, and control devices was confirmed. The developed user interface provides real-time monitoring of environmental conditions, farming records, and feeding status. Operators can manage tank information, review historical data, and adjust feeding parameters through the integrated platform. Experimental deployment demonstrated the feasibility of applying the proposed system to practical aquaculture environments and verified the effectiveness of integrating monitoring, data management, and automated control functions within a single platform. The developed system was successfully integrated with the automatic feeding equipment and demonstrated stable operation under practical farming conditions. The results indicate that the proposed platform can serve as a foundation for future AI-based fish monitoring, biomass estimation, and intelligent feeding optimization systems. Fig. 1 shows the main monitoring dashboard of the proposed smart aquaculture system. The interface provides real-time environmental information and tank-level aquaculture data to support feeding management and operational monitoring. Fig. 2 presents the data collection and analysis interface. The system allows users to visualize environmental and farming data through various statistical and graphical analysis functions.

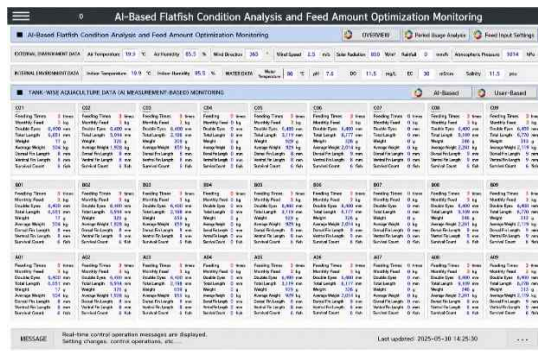


Fig. 1. Real-time monitoring dashboard

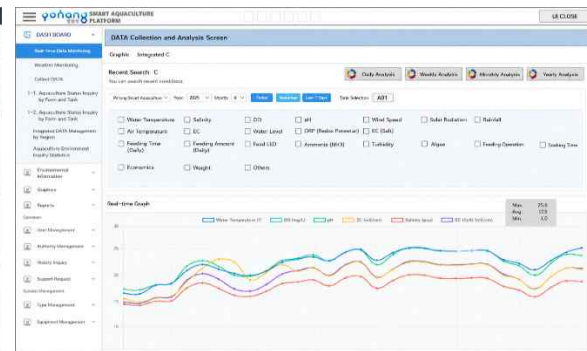


Fig. 2. Aquaculture data analysis interface

Conclusion: This paper presented the design and implementation of an AI-based smart aquaculture system integrating environmental monitoring, feeding optimization, and automated control. The proposed platform combines IoT sensing, data management, AI-based fish segmentation, and edge-based feeder control to support efficient aquaculture operation. Field deployment demonstrated stable system performance and practical applicability in commercial aquaculture environments. Unlike conventional systems that mainly focus on monitoring or automated feeding, the proposed system provides a unified platform for intelligent aquaculture management and future AI-driven precision aquaculture. Future work will focus on incorporating computer vision-based fish analysis, biomass estimation, and intelligent feeding recommendation algorithms to further improve aquaculture productivity and sustainability.

Acknowledgment

This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by the Korean government. [26ZD1150, Advancement and Commercialization of Daegu-Gyeongbuk Regional Strategic Industries (Farm)]

References

1. Y. Sun et al., Aquaculture Monitoring with AI-Based Video Analytics, Sensors, vol. 21, no. 18, (2021).
2. H. Li et al., Fish Monitoring and Biomass Estimation Using Computer Vision, Computers and Electronics in Agriculture, vol. 177, (2020).
3. IEEE Access, Smart Aquaculture Technologies, vol. 8, pp. 105662–105670, (2020).
4. FAO, The State of World Fisheries and Aquaculture, (2024).

Factors Affecting the Acceptance and Purchase Intention of Smart Rings: An Extended Technology Acceptance Model Approach

Sang Hoon Lee¹⁾ and Su-Yeon Kim^{2*)}

¹⁾Software Education Institute, Kyungpook National Univ., 80, Daehak-ro, Buk-gu, Daegu, Korea

²⁾School of Computer and Information Engineering, Daegu Univ., Gyeongsan, Gyeongbuk 38453, Korea

*E-mail: sykim@daegu.ac.kr

Abstract: As wearable devices spread rapidly, smart rings worn on the finger are drawing attention as a new device that is both a health-monitoring tool and a body-worn ornament. Based on the Technology Acceptance Model (TAM), this study examines the factors affecting smart ring acceptance and purchase intention. Aesthetics, self-expression, perceived enjoyment, subjective norm, and reliability were added as exogenous variables influencing perceived usefulness (PU) and perceived ease of use (PEOU), which in turn shape attitude and purchase intention. Experimental data from 563 respondents, mostly in their twenties, were analyzed using structural equation modeling and bootstrapping. The measurement model showed acceptable reliability and validity, and the structural model fit was satisfactory (CFI=.918, RMSEA=.077). Perceived enjoyment, reliability, and self-expression significantly affected PU and PEOU, while aesthetics and subjective norm did not. PU strongly influenced attitude and purchase intention.

Keywords : Smart Ring; Wearable Device; Technology Acceptance Model; Aesthetics; Self-expression; Purchase Intention

Introduction: The system aims to enhance security, privacy, and user awareness by operating locally within the browser. Advances in sensor miniaturization have accelerated the spread of wearable devices. While smartwatches and fitness bands have led the market, finger-worn smart rings are emerging as a new device category, measuring health data such as sleep, heart rate, and activity without a screen. Unlike wrist-worn devices, a ring is a constantly visible ornament, so smart ring acceptance can depend on design, fashion, and self-expression as well as function [1,2]. Prior wearable-acceptance studies have mostly examined smartwatches and fitness bands through core TAM constructs alone and rarely address the ornamental character of body-worn devices. This study proposes an extended TAM incorporating aesthetics, self-expression, perceived enjoyment, subjective norm, and reliability to explain smart ring acceptance and purchase intention.

Research Model: The proposed research model (Fig. 1) was established based on the TAM proposed by Davis, which is the most widely used in the field of new technology adoption [3]. Based on related research, the model assumes that five exogenous variables - Reliability (RP), Aesthetics (AE), Self-expression (SE), Perceived Pleasure (PE), and Subjective Norms (SN) - influence Perceived Pleasure (PU) and Perceived Pleasure of Use (PEOU), which form Attitude toward Smart Rings (ATT), and that ATT and PU influence Purchase Intention (PI) [4-5]. 15 hypotheses (H1-H15) specify positive paths from each exogenous construct to PU and PEOU, from PEOU to PU, from PU and PEOU to ATT, and from ATT and PU to PI.

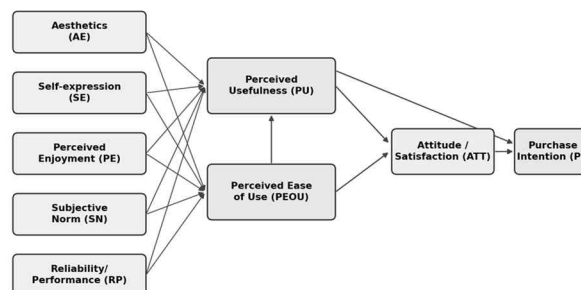


Fig. 1. Research Model for Smart Rings Acceptance

An online survey was conducted from November 2024 to December 2025; 563 valid responses were obtained, 93.4% from respondents in their twenties and 97.5% with no prior smart ring experience. Each construct was measured with three items adapted from established TAM scales on a 5-point Likert scale, and the model was estimated using confirmatory factor analysis and structural equation modeling.

Experimental Results: Reliability and validity were largely supported: Cronbach's α exceeded .85 for most constructs ($CR \geq .87$, $AVE \geq .69$), except reliability (RP: $\alpha = .760$, $CR = .752$, $AVE = .505$), which also showed borderline discriminant validity with PU and ATT. The structural model showed acceptable fit ($\chi^2(299) = 1291.56$, $CFI = .918$, $TLI = .904$, $RMSEA = .077$), explaining substantial variance in PU ($R^2 = .83$), ATT ($R^2 = .75$), PI ($R^2 = .55$), and PEOU ($R^2 = .32$). Of 15 hypotheses, 8 were supported (Table 1). Reliability and perceived enjoyment most strongly affected both PU and PEOU; self-expression significantly affected both, with smaller effects. Aesthetics and subjective norm did not significantly affect PU or PEOU. PEOU did not significantly affect PU or ATT, whereas PU strongly affected ATT ($\beta = .857$) and PI ($\beta = .367$), and ATT affected PI ($\beta = .398$). Bootstrapped total effects (300 resamples) confirmed that only reliability, perceived enjoyment, and self-expression had significant total effects on ATT and PI, consistent with the structural path results.

Table 1. Summary of Structural Path Results (standardized β)

Path	β	Path	β
RP $\rightarrow$ PU	.432***	RP $\rightarrow$ PEOU	.259***
AE $\rightarrow$ PU	-.002 n.s.	AE $\rightarrow$ PEOU	-.083 n.s.
SE $\rightarrow$ PU	.109*	SE $\rightarrow$ PEOU	.168*
PE $\rightarrow$ PU	.442***	PE $\rightarrow$ PEOU	.373***
SN $\rightarrow$ PU	.039 n.s.	SN $\rightarrow$ PEOU	-.122*
PEOU $\rightarrow$ PU	.039 n.s.	PU $\rightarrow$ ATT	.857***
PEOU $\rightarrow$ ATT	.020 n.s.	ATT $\rightarrow$ PI	.398***
PU $\rightarrow$ PI	.367***		

Model fit: $\chi^2(299) = 1291.56$, $CFI = .918$, $TLI = .904$, $RMSEA = .077$; R^2 : PEOU = .32, PU = .83, ATT = .75, PI = .55.

* $p < .05$, *** $p < .001$, n.s. = not significant.

Conclusion: Smart ring acceptance is driven primarily by perceived enjoyment and reliability rather than aesthetics, indicating that trustworthy measurement and an enjoyable user experience matter more than visual appeal alone. Self-expression, unlike aesthetics, significantly increased both PU and PEOU, suggesting that the value of design lies in expressing personal identity rather than attractiveness per se. Subjective norm was not significant, consistent with the smart ring's early-stage, low-diffusion market status. For practice, manufacturers should prioritize sensor accuracy and reliability while enhancing enjoyable user experience, and marketing should emphasize self-expression rather than aesthetics or social proof. Limitations include a student-skewed sample, respondents' lack of actual usage experience, and potential common method bias; future research should test the model with actual users and multi-group analyses across demographic groups.

References

1. N. Tractinsky, A. S. Katz, and D. Ikar, What is beautiful is usable. *Interacting with Computers*, 13(2), 127-145, (2000).
2. S. H. W. Chuah, P. A. Rauschnabel, N. Krey, B. Nguyen, T. Ramayah, and S. Lade, Wearable technologies: The role of usefulness and visibility in smartwatch adoption. *Computers in Human Behavior*, 65, 276-284, (2016).
3. F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340, (1989).
4. M. A. Rupp, J. R. Michaelis, D. S. McConnell, and J. A. Smither, The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. *Applied Ergonomics*, 70, 77-87, (2018).
5. D. Shin and F. Biocca, Impact of social influence and users' perception of coolness on smartwatch behavior. *Social Behavior and Personality*, 46(6), 881-890, (2018).

EmbedLearn: On-Device Intelligence for Smart Classroom Monitoring

T.Velmurugan¹⁾, Senthilkumar Ramamoorthy²⁾, Sridevi PC^{3*)}

1) Advisor, Meston College of Arts and Science, Chennai 600014, India

2) Senior Java /AI Engineer, The Options Clearing Corporation (OCC), USA

3) Dept. of Artificial Intelligence and Data Science, Takshashila University, Tindivanam 604305, India,

*E-mail: velmurugan_dgvc@yahoo.co.in, sennu@live.com, sridevipc@takshashilauniv.ac.in**

Abstract: Smart classrooms are gradually becoming a core part of modern university infrastructure, and there is a real need to make them intelligent in a way that goes beyond just projectors and whiteboards. This paper presents EmbedLearn, an on-device AI system designed to monitor and respond to classroom dynamics in real time using embedded sensors and edge processors. Unlike cloud-dependent systems, EmbedLearn processes all data locally — covering student engagement, environmental conditions, and occupancy — without transmitting sensitive information outside the room. We implemented the system on Raspberry Pi 4 and ESP32 microcontrollers integrated with thermal, acoustic, and vision sensors. Experimental results across multiple classroom sessions demonstrate that EmbedLearn achieves engagement detection accuracy of 87.4% while operating within a power budget of under 5W. The system responds meaningfully to real classroom conditions, making it a practical and privacy-preserving solution for next-generation smart learning environments.

Keywords: Embedded AI, Smart classroom, Edge computing, Student engagement, On-device inference, IoT sensors

Introduction: The classroom has always been central to learning, but traditional setups offer very little feedback on what is actually happening inside them. Teachers often have no objective way of knowing whether students are engaged, whether the room is too noisy, or whether environmental conditions like temperature and lighting are even suitable for learning. Across universities in India and abroad, this gap is becoming harder to ignore as batch sizes grow and contact hours shrink. Embedded AI offers a compelling path forward — small, low-power processors running trained models directly on sensor data, no internet connection required. Several recent works have explored smart classrooms using cloud-based AI [1], [2], but these approaches raise valid concerns about data privacy and latency. TinyML has emerged as a particularly promising direction, enabling neural network inference on devices with only a few kilobytes of RAM [4], [6]. Edge deployment also aligns well with institutional data governance requirements, where student behavioral data must remain within campus boundaries [7]. Our work focuses on keeping everything on-device, making the system usable in low-connectivity settings and ensuring student data never leaves the classroom.

System Design and Methodology: EmbedLearn is built around a two-tier embedded architecture. The first tier consists of ESP32 microcontrollers that handle environmental sensing — they continuously read temperature, humidity, CO2 levels, ambient light, and sound pressure. These readings are passed over a local MQTT network to a Raspberry Pi 4 that forms the second tier, responsible for inference and coordination. A wide-angle camera module feeds into a lightweight MobileNetV2-based engagement classifier that was trained offline and quantized to INT8 format before deployment [8]. The model takes 320x240 grayscale frames and classifies the room state into three categories: engaged, distracted, or absent. A separate thermal sensor array (AMG8833) provides a coarse occupancy map, which helps filter engagement predictions when the room is below half capacity. All inference runs locally using TensorFlow Lite [9]. The system logs aggregated (non-personal) metrics every five minutes and surfaces them through a lightweight web dashboard accessible only within the campus LAN. This design draws on prior sensor fusion work in smart building environments [10] and adapts it specifically for educational settings.

Experimental Results: It is deployed EmbedLearn across three different classrooms at our institution over a four-week period, covering a combined total of 68 lecture sessions with class sizes ranging from 20 to 65 students. Ground truth labels for engagement were collected by trained observers using a structured checklist during the same sessions. The engagement classifier achieved an average accuracy of 87.4% and an F1-score of 0.84 across all sessions. Environmental monitoring triggered automatic alerts in 14 sessions where CO₂ levels exceeded 1000 ppm — a threshold linked to reduced cognitive performance in classroom research [3]. These findings are consistent with related edge inference benchmarks reported in [5] and [6], where similar quantized models achieved comparable accuracy on resource-constrained hardware. Table 1 summarizes the performance metrics across the three rooms.

Table 1. Performance Metrics based on Classroom Deployments

Classroom	Sessions	Accuracy (%)	F1-Score	Avg. Power (W)
Room A (30 seats)	24	89.2	0.87	4.6
Room B (50 seats)	22	86.1	0.83	4.8
Room C (65 seats)	22	86.8	0.82	5.1

Conclusion: This paper introduced EmbedLearn, an on-device AI system for smart classroom monitoring that prioritizes privacy, low latency, and energy efficiency. By running all inference locally on embedded hardware, the system avoids the connectivity and data-privacy concerns associated with cloud-based approaches. Field deployment confirmed that the system can detect engagement states with high accuracy and respond meaningfully to environmental conditions in real classroom settings. Future work will focus on expanding the sensor suite to include gaze estimation and improving model robustness for larger, more diverse classrooms. We believe this kind of edge-first design philosophy is essential for AI systems that need to operate sustainably and responsibly inside educational institutions.

Acknowledgment

The authors acknowledge the contributions of Team Members for their assistance in data collection and system setup.

References

1. A. Dewan, R. Sharma, and M. Gupta, “Cloud-based student engagement analytics using computer vision”, *IEEE Trans. Learn. Technol.*, vol. 14, no. 3, pp. 312–321 (2021).
2. Y. Chen and H. Lin, “Automated classroom attention monitoring via deep learning and IoT sensors”, *Comput. Educ.*, vol. 178, p. 104389 (2022).
3. U. Satish, M. Mendell, K. Shekhar, T. Hotchi, D. Sullivan, S. Streufert, and W. Fisk, “Is CO₂ an indoor pollutant? Direct effects of low-to-moderate CO₂ concentrations on human decision-making performance”, *Environ. Health Perspect.*, vol. 120, no. 12, pp. 1671–1677 (2012).
4. S. Park and J. Kim, “TinyML deployment for classroom occupancy detection on microcontrollers”, *Proc. IEEE Int. Conf. Embedded Syst.*, pp. 88–93 (2023).
5. R. Iyer and V. Balakrishnan, “Edge AI for smart campus infrastructure: A survey”, *J. Netw. Comput. Appl.*, vol. 210, p. 103543 (2023).
6. P. Warden and D. Situnayake, *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*, O’Reilly Media, Sebastopol, CA (2019).
7. M. Zhu, W. Zhang, and L. Qiao, “Privacy-preserving student behavior analysis using federated learning at the edge”, *Future Gener. Comput. Syst.*, vol. 138, pp. 201–212 (2023).

8. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications", *arXiv preprint arXiv:1704.04861* (2017).
9. R. David, J. Duke, A. Jain, V. Janapa Reddi, N. Jeffries, J. Li, N. Kreeger, I. Nappier, M. Natraj, T. Wang, P. Warden, and R. Rhodes, "TensorFlow Lite Micro: Embedded machine learning for TinyML systems", in *Proc. Mach. Learn. Syst. (MLSys)*, vol. 3, pp. 800–811 (2021).
10. F. Ismail, N. Hassan, and A. Rahim, "Sensor fusion for intelligent building management: An IoT and edge computing approach", *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4312–4325 (2023).

Next-Generation Cloud-Atm Architecture Using Enhanced Hybrid Elliptic Curve Cryptography for Eradicate Skimming Attacks

SriPradha.G¹, R.Samundeeswari²

^{1,2)} Asst. Professor, Dept. of Information Technology, Guru Nanak College, Chennai, India -600042

E-mail: sripradhag09@gmail.com, rsamul65@gmail.com

Abstract:

The development of the Automated Teller Machine (ATM) has significantly enhanced efficiency, convenience, and accessibility for bank customers. However, ATM skimming attacks create a major risk in the banking sector, leading to significant losses for banks and account holders while negatively impacting the global economy. This hybrid approach is used to design cyber-Gate cloud-ATM architecture for eradicating skimming attacks in the hybrid cloud environment to protect ATM PINs and ensure safe banking transactions. The proposed system contributes by generating an Instant PIN based on the ATM session key using the Enhanced Hybrid Elliptic Curve Cryptography (EHECC) algorithm to enhance the overall success rate of ATM transaction security, reduce response time, and lower computational cost, and it also expedites secure banking processes more than the existing system.

Keywords : EHECC technique, IPIN, Bank CyberGate Server, Authentication server, GPRS

Introduction:

As digital technology continues to evolve, the volume of data generated and processed across industries has grown exponentially. Securing ATM transactions is a major challenge for the banking sector. An ATM is a gadget that facilitates various banking financial operations such as fund transfers, withdrawals, and balance inquiries, allowing ATM users to access banking services more efficiently and conveniently without waiting in long queues^[2,27]. The Times of India Business Newspaper reported that recently, one million debit cards were stolen in India. The report highlighted that skimming attacks are a significant security threat in the banking sector. Skimming attacks deceive account holders by capturing sensitive information, including PIN, through the use of pinhole cameras and scanners. To address this challenge, research proposes a cyberGate-secured cloud ATM architecture to mitigate skimming attacks in banking ATM systems. To reduce security flaws in ATM transactions, an encryption and decryption process is implemented using public and private keys of the device or server. Additionally, Instant PIN (IPIN) is created using an Enhanced Hybrid ECC algorithm in the bank cyberGate server to enhance security in the ATM cloud.

Proposed Methodology:

Skimming attacks are a major form of cybercrime attacks that steal bank customer credentials through a skimmer device, resulting in significant losses for both banks and customers. To tackle this issue, the robust architecture develops a hybrid security model that generates an Instant PIN (IPIN) in the bank cyberGate server (BCS) to reduce security vulnerabilities against skimming attacks. The novelty of the proposed system lies in its ability to create an Instant PIN based on the ATM session key, which is generated from the ATM session time (hours: minutes: seconds) using random number generation methods of the Enhanced Hybrid Elliptic Curve Cryptography algorithm. In the proposed system, IPIN creation is done by using a strong security mechanism that enhances higher security than the existing system. The main features of the cyberGate Cloud-ATM architecture include the design of a secured strong cyberGate architecture, the creation of IPIN using the EHECC technique to enhance the security of the ATM, protection of banks as well as bank customers from security risks and vulnerabilities, resistance to skimming attack in the banking sectors, speed up the efficient computation of ATM transactions, saving memory usage, reduces energy consumption, minimizes the response time of ATM transactions, low computation cost and PIN authentication, confidentiality, non-repudiation, data integrity are also achieved.

The Cloud-ATM network functionality is illustrated as follows

Step 1: ATM Functionality

Step 2: First Level of Verification Test in MPLS Banking ATM Network

Step 3: Second Level of Verification Test in CyberGate Cloud-ATM Network

Step 4: Instant PIN Authentication

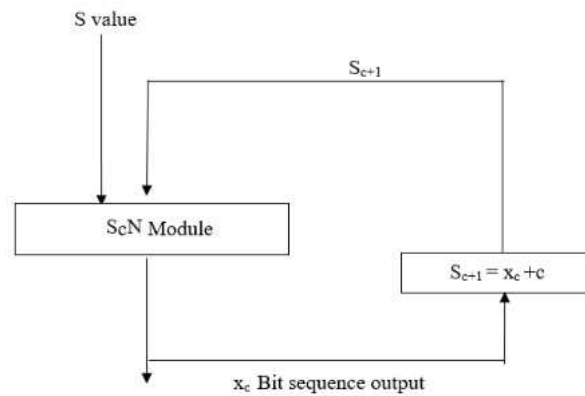


Fig. 1. Block Diagram of Random Number Generation

CONCLUSION AND FUTURE DIRECTION

Combating ATM security issues is an important challenge in ATM transactions. Today, skimming attacks play a vital role in hacking bank customer details and creating duplicate ATM cards to carry out illegal ATM transactions. To resist these security attacks, a strong, secure architecture is needed to boost high security. This is achieved using an Instant PIN generated through the methods of a random number generator using the Enhanced Hybrid Elliptic Curve Cryptography algorithm in the bank cyberGate server. Hence, the proposed CyberGate Strong Secured Cloud-ATM Architecture demonstrates high security, expedites the ATM computations, requires less memory storage, has minimum computation cost, and has less response time. The future direction of this research work is to detect and mitigate malware attacks in banking ATM transactions.

References

1. Y.Sunil Raj, L.Lucase, E.Helen Parimala "DNA based Hybrid Approach for data integrity on integrated cloud based smart environment an Analysis", International Journal of Scientific Development and Research, Volume 5, Issue 10, 2020.
2. K.V.Gunalan, R.A.Sashidhar, R.Srimathi, S.Revathi, Nithya Venkatesan, "Enhanced ATM Security using Facial Recognition, Fingerprint Authentication and WEB Application", International Conference on Futuristic Communication and Network Technologies, Springer, Volume 966, pp.273-288, 2023.
3. M.Lakshmi Priya, Dr.S.Albert Rabara, E.Helen Parimala "A Secured Cloud Based Integrated Authentication for preventing Network Security Attacks", International Journal for research and development in Technology, Volume 14, Issue 1, 2020.
4. A. Angelpreethi and S. B. Ramesh Kumar, "Visualizing Big Datamining: Issues, Challenges and Opportunities," International Journal of Control Theory and Applications, volume. 9, Issue. 27, pp. 455-460, 2016.
5. Sweedle Machado, Prajyoti D'silva, Snehal D'mello, Supriya Solaskar, Priya Chaudhari, "Securing ATM pins and passwords using Fingerprint Based Fuzzy Vault System", International Conference on Computing Communication Control and Automation, IEEE, doi: 10.1109/ICCUBEA.2018.8697794, 2018.
6. A.Angelpreethi, S.B.R.Kumar, "An Enhanced Architecture for Feature Based Opinion Mining from Product Reviews", World Congress on Computing and Communication Technologies, Tiruchirappalli,

- India, pp.89-92, doi: 10.1109/WCCCT.2016.30, 2017.
7. Dr.A.Angelpreethi,"Fuzzy Based Sentiment Classification Using Fuzzy Linguistic Hedges for Decision Making", Mapana-Journal of Sciences, Volume 22, Special Issue 2, pp.63-79, 2023.
 8. P. S, K. R., F. Iram, S. R. M V, N. R and A. Angelpreethi, "Improved Moth Flame Optimization with Deep Convolutional Neural Network for Colorectal Cancer Classification using Biomedical Images," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-6, doi: 10.1109/ICAIT61638.2024.10690631.
 9. A.Angelpreethi, Dr.S.Britto Ramesh Kumar," Dictionary based approach to improve the accuracy of opinion mining on big data", International Journal of Research and Analytical Reviews, volume 5, issue 4 (sep 2018), E ISSN 2348 –1269, PRINT ISSN 2349-5138.

A Multi-Joint Data-Based SOMA System for Fine Motor Development in Children with Special Needs

Narangerel Tsolmon¹⁾, Jimin Ryu²⁾ and Yoosoo Oh^{3*)}

^{1,2)} NWI Co., Ltd., Gyeongsan 38453, Republic of Korea

³⁾ School of Computer & Information Engineering, Daegu University, Gyeongsan 38453, Republic of Korea

**E-mail: yoosoo.oh@daegu.ac.kr*

Abstract: This paper proposes SOMA (sLLM-based Objective Monitoring AI), a multi-joint data-driven system for monitoring fine motor development in children with special needs. To minimize webcam coordinate jitter, we introduce a time-series preprocessing pipeline combining Exponential Moving Average (EMA) smoothing and body-size normalization. The SOMA system integrates gamification elements, such as on-screen guides and score rewards, to encourage child participation. Through few-shot gesture registration, it enables customized motor tasks tailored to individual developmental stages. SOMA system extracts five real-time quantitative metrics, such as hand tremors and bilateral coordination, and generates automated analytical reports via sLLM-driven summaries. Experimental results show a 15.4% to 65.2% jitter suppression rate. The sLLM inference achieves a 0.121-second TTFT and a throughput of 66.99 tokens per second, confirming its capability for real-time, latency-free feedback.

Keywords : Fine motor, multi-joint tracking, sLLM, special education

Introduction: Fine motor control is a critical developmental milestone that influences an individual's Activities of Daily Living (ADL) and cognitive maturation. However, conventional motor assessments for children with special needs primarily depend on manual observations or standardized checklists, which may introduce subjective variance. Furthermore, evaluations conducted within unfamiliar clinical settings frequently induce psychological tension in children, potentially hindering their natural motor performance. To overcome these challenges, in this paper, we propose SOMA (sLLM-based Objective Monitoring AI), a multi-joint data-driven system applicable to home environments.

Research Methods: The proposed SOMA system uses a four-stage pipeline to process multi-joint data from a single webcam feed, as shown in Fig. 1. First, in the Multi-Joint Data Extraction stage, we utilize the MediaPipe framework to extract 23 upper-limb-centric body joint coordinates using the PoseLandmarker and capture 42 detailed hand joint coordinates (21 joints per hand) via the HandLandmarker. By integrating these joint structures, we construct a total configuration of 65 multi-joint data points, enabling the simultaneous tracking of both major upper-limb segments and highly articulate finger movements. Second, during Time-Series Preprocessing, we apply a time-series preprocessing pipeline that combines an EMA-based smoothing technique with body-size normalization to suppress environmental noise and standardize coordinate data. Through this approach, we minimize false detections caused by environmental noise and extract highly objective evaluation metrics. Third, the Task Configuration stage implements a few-shot-based gesture registration that extracts core joint angle features from a 5-second demonstration video, enabling therapists to dynamically configure customized motor tasks for each child's specific rehabilitation goals. Finally, under Feature Extraction and Analysis, the system extracts five quantitative metrics (hand tremors, bilateral coordination, finger extension range, movement path efficiency, and target point proximity) in real time while integrating gamification features to maximize child engagement. Upon the completion of an observation session, the system utilizes a sLLM (Qwen3-14B GGUF) to automatically generate an analytical summary based on the time-series data. The resulting natural language report incorporates visualized graph analyses, key performance characteristics, and targeted areas for improvement, translating complex metrics into actionable insights.

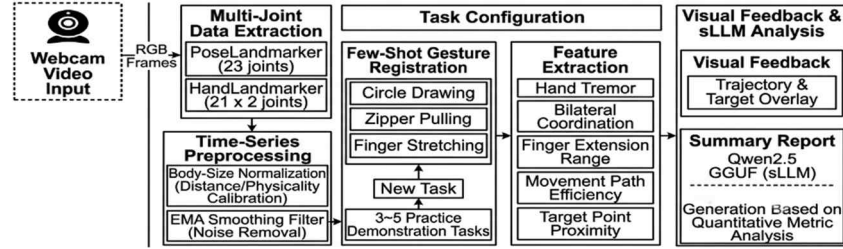


Fig. 1. SOMA system architecture

In the third paragraph, Experimental Results: We evaluated the SOMA system by measuring coordinate stability under the preprocessing pipeline and testing sLLM in a local environment. As shown in Table 1, the combination of the EMA filter and deadband thresholds significantly reduced coordinate jitter, yielding an average stability improvement of 15.4% and suppressing up to 65.2% of highly volatile hand tremor metrics. Additionally, in 20 test trials, the sLLM achieved an average Time to First Token (TTFT) of 0.121 seconds and a throughput of 66.99 tokens per second (TPS).

Table 1. Comparison of Coordinate Stability Before and After Preprocessing

Metric	Pre-processing (Raw)	Post-processing (Stabilized)	Suppression Rate (%)
Average Coordinate Jitter (px)	11.671	9.874	15.4%
Hand Tremor (Right Hand)	55.919	39.820	28.8%
Hand Tremor (Left Hand)	49.563	17.251	65.2%

In the fourth paragraph, Conclusion: This paper presented SOMA, a multi-joint data-driven system providing objective, home-based fine motor development monitoring for children with special needs. By integrating time-series data preprocessing, a gamified interface, and automated sLLM summary generation, the system successfully complements the subjectivity of conventional observer-centric metrics. Experimental results confirmed high tracking stability and latency-free report generation capability. Future work will expand this framework to the elderly population, exploring its utility for customized fine motor training and dementia prevention programs.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by Korean Government [Ministry of Science and ICT (MSIT)] under Grant RS-2026-25482392.

References

1. C. E. Cameron, L. L. Brock, W. M. Murrah, L. H. Bell, S. L. Worzalla, D. Grissmer, and F. J. Morrison, Fine motor skills and executive function both contribute to kindergarten achievement, *Child Dev.*, 83(4), 1229 (2012).
2. B. Provost, B. R. Lopez, and S. Heimerl, A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns, *J. Autism Dev. Disord.*, 37(2), 321 (2007).
3. C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, and M. Grundmann, MediaPipe: A framework for building perception pipelines, *arXiv preprint arXiv:1906.08172*, p. 1 (2019).
4. V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, BlazePose: On-device Real-time Body Pose tracking, *arXiv preprint arXiv:2006.10204*, p. 1 (2020).
5. Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, Generalizing from a few examples: A survey on few-shot learning, *ACM Comput. Surv.*, 53(3), 1 (2020).
6. J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, et al., Qwen Technical Report, *arXiv preprint arXiv:2309.16609*, vol. 1, p. 1 (2023).

AI-Based Multi-Agent Crowd Safety Surveillance System for Real-Time Threat Monitoring

Abstract - Crowd safety in densely occupied venues is harder to manage than it looks. A single incident, if not caught early, spreads fast. To tackle this, we built a surveillance system that uses multiple AI agents running in parallel, each looking at a different aspect of the scene. All agents will receive a single preprocessing output per frame detection via YOLOv8, person tracking through DeepSORT, and motion estimation using Farneback optical flow are ran once, then passed down. On top of this, we have four agents: one for spotting unusual crowd behavior (CNN-LSTM on UCSD Ped2 heatmaps), one for crowd headcount, a panic detector using per-person speed baselines, and a suspicious object tracker that checks bag-owner associations. LangGraph clubs the agents together and runs them in parallel. Their outputs get fused into a single threat score across different levels. A FastAPI backend and React dashboard handle live video, per-agent scores, and alerts. Initial testing on surveillance clips shows the system catching threats in all four categories with reasonable accuracy.

Index Terms - Crowd surveillance, YOLOv8, DeepSORT, CNN-LSTM, LangGraph, Anomaly detection, Public safety

I. INTRODUCTION

Keeping large public spaces safe is always a difficult task. The problem is not just technical but operational a security guard at a checkpoint simply cannot monitor a 500 person crowd across six entry points at the same time. They might catch someone moving suspiciously near Gate 3, but that's about it. The rest of the scene goes unobserved, especially in the middle of the shift when the attention starts to drop.

Camera infrastructure has improved considerably over the last decade, but the improvement has mostly been in resolution and the coverage. The intellectual load on the person watching those feeds hasn't changed. In fact it's arguably gotten worse, more cameras, same number of people watching. One can't just throw more screens at the problem.

Most existing automated tools don't really solve this problem either. They tend to be a single-purpose: count the people in frame, or flag a stationary object, or detect running. These outputs will be present in a separate dashboards that operators still there has to be a mentally present. That's assuming the tools are even running in real-time, which many aren't.

Our project tries a different setup. Rather than running separate pipelines that don't talk to each other, we built a system with a shared scene understanding module that feeds multiple parallel agents simultaneously. Each agent focuses on one specific safety concern. A coordinator layer (built on LangGraph) manages them and combines their outputs into a single actionable risk score. The operator sees one number and one alert level, not four separate panels from four separate tools.

The important technical contributions are mentioned below:

- 1) A blend of preprocessing pipeline that runs YOLOv8 detection, DeepSORT tracking, and Farneback dense optical flow once per frame. The results are stored in a SceneState object that all four agents read from there. This avoids running the same GPU-heavy computations multiple times per frame.
- 2) An abnormal activity module that converts per-frame scene information into three-channel 64x64 heatmaps (occupancy, speed, motion) and feeds sequences of 16 these such heatmaps into a CNN-LSTM classifier. For training we used the UCSD Ped2 benchmark. The heatmap approach keeps the model from overfitting to specific camera appearances.
- 3) LangGraph-based orchestration that gives out to all four agents simultaneously, then collects and combines their scores using the weighted formula. The graph structure makes it straightforward to add or swap out agents without rewriting the coordination logic.
- 4) The deployment stack: FastAPI on the backend serving live MJPEG video and the JSON status updates, and a React dashboard on the frontend showing video with overlays, per-agent scores, a risk gauge, and a time-stamped alert log.

This paper covers related work in Section II, then the system architecture (III), preprocessing pipeline (IV), anomaly detection (V), the four agents (VI), LangGraph orchestration (VII), deployment interface (VIII), results (IX), limitations (X), future directions (XI), and a conclusion (XII).

It's worth noting that rule-based approaches – the kind that trigger on simple motion thresholds or pixel differences tend to break down precisely in the conditions where you need the most: crowded places, partial obstruction, uneven lighting. What's needed are systems that can actually reason about the behavior rather than just react to raw pixel values.

II. RELATED WORK

We have reviewed the existing work across four areas that our system touches on them directly.

A. Anomaly Detection in Crowd Footage

The UCSD Pedestrian dataset[4] has been widely used as a benchmark for this problem for more than a decade. Earlier methods mainly depended on manually designed spatio-temporal features, often combined with the sparse optical flow techniques. One insight that came out of this work and that influenced our design – was that anomalies are temporal by nature. You can’t reliably classify a single frame as normal or abnormal; you need to see how the scene is evolving. That pushed the field toward sequence-based methods [5], which is the direction we followed as well.

Where we differ from earlier sequence-based methods is in the input representation. Instead of raw frames or optical flow fields, we used a learned heatmap that encodes crowd density, per-person speed, and scene-wide motion in three different separate channels. This was motivated partly by practical concerns: we wanted the model to generalize to the different cameras without retraining.

B. Multi-Object Tracking

DeepSORT [2] has been improved over the original SORT tracker by adding a deep appearance model for re-identification, this matters a lot in the crowded scenes where people frequently pass in front of each other and the tracking IDs can get swapped. For our system, stable track identities are essential because the panic detection and suspicious object agents both reason about the history of individual tracks. Without the consistent IDs, you can’t tell if someone’s speed has changed or if a bag’s owner has walked away.

C. Multi-Agent AI Orchestration

LangGraph is a graph-based framework that makes the structure of a multi-step AI pipeline explicit. Unlike writing a big procedural loop, you define nodes and edges, and the framework handles the execution. What stood out as especially useful for us were features like typed state passing between nodes, built-in support for parallel branch execution, and the ability to visually inspect the computation graph during development. Adding a new detection agent is straightforward, it simply involves introducing a new node, without needing to change any existing code.

D. Integrated Surveillance Systems

Most of the surveillance platforms we looked at treat detection, tracking, and alerting as separate components wired together with hard-coded rules. One output feeds into the next through fixed thresholds. The downside is rigidity the system can’t account for context. Our approach builds a shared per-frame representation that all agents use, which at

least ensures they’re all looking at the same scene at the same moment.

Earlier, crowd analysis mostly depended on background subtraction and optical flow [1]. Later, deep CNN models helped improve how accurately people could be detected. Newer models like Vision Transformers (ViT) and Swin Transformer can give even better results, but they need a lot of computing power, so they are not suitable for real-time use on normal devices. Multi-agent systems are already used a lot in robotics, but they are not commonly used in real-world surveillance. This is the gap we are trying to address.

III. SYSTEM OVERVIEW

This architecture has four layers stacked on top of each other. At the bottom is the shared preprocessing, which ingests raw video and outputs a SceneState per frame. Above that, the agent layer runs four modules concurrently. The risk layer fuses their outputs into a single threat index. At the top, the deployment layer exposes everything to operators through a web interface.

Figure 1 sketches the high-level flow. Frames come in, get processed once, then fan out to four parallel agents whose results converge at a risk fusion node before reaching the alert system.

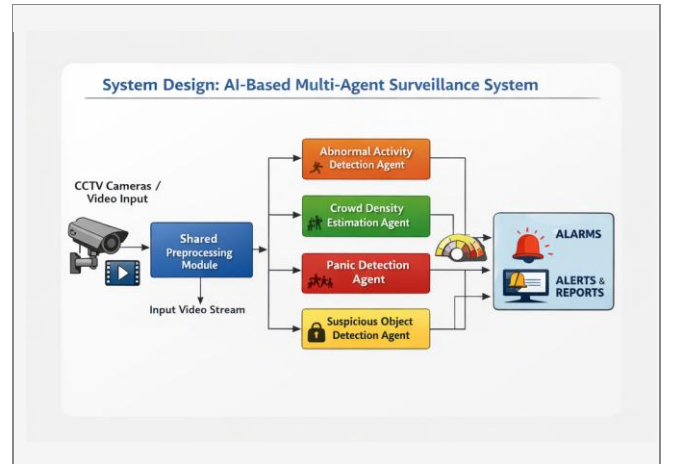


Fig. 1. High-level system architecture. Video frames flow through shared preprocessing into four parallel agents, fused into risk-aware alerts.

Figure 2 shows the agent layer in more detail. Each agent takes a SceneState and emits a float confidence score (and optionally some annotation metadata). These four scores go to the Risk Evaluation node, which produces the composite index. The modular setup means any single agent can be updated, retrained, or replaced without touching the others. And since they run in parallel, total latency per cycle is set by whichever agent takes longest – in practice that’s the CNN-LSTM inference in the anomaly agent.

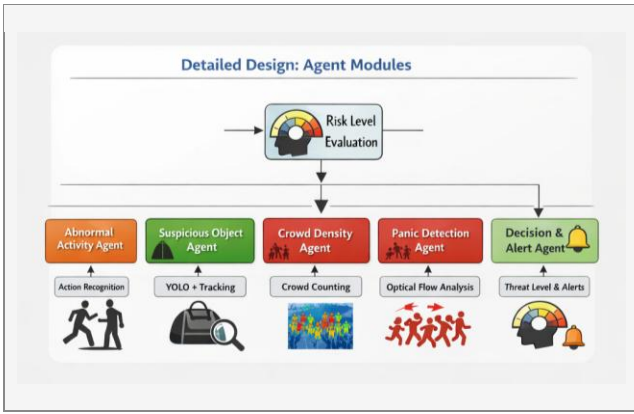


Fig. 2. Agent module wise detail. Each agent receives SceneState and returns a confidence score.

On model choice: we went with YOLOv8 rather than an earlier version primarily because the anchor-free detection head performs better in scenes with overlapping detections, which is exactly what you get with crowds. Running in FP16 on a GPU keeps throughput above 25 FPS, which is sufficient for live surveillance applications.

IV. SHARED VIDEO PROCESSING PIPELINE

The preprocessing stage runs once per frame. Everything it produces is packaged into a typed SceneState container, and that’s the only thing the downstream agents ever receive. No agent runs its own detection or tracking.

A. Object Detection (YOLOv8)

We run YOLOv8 [1] configured to detect two classes: people, and bag-type objects (backpacks, handbags, suitcases). We intentionally keep the confidence threshold on the safer side. It’s better to have a few extra detections than to miss something important. In safety situations, a false positive only takes a second of an operator’s attention, but a false negative could mean missing a real incident.

B. Person Tracking (DeepSORT)

The detected bounding boxes are passed to DeepSORT [2]. It uses a Kalman filter to predict movement and a deep learning model to recognize objects again if they are briefly hidden. We also add a trajectory manager that keeps a short history for each object, like its position over recent frames, its speed, and how it is moving. These details are then used by the panic and suspicious object detection agents.

C. Dense Optical Flow (Farneback)

Along with the tracker, we also use the Farneback algorithm [3] on pairs of smaller grayscale frames. This gives us motion information for every pixel in the frame. It helps us detect movement even in areas where no objects are being tracked, which is useful for understanding what’s happening in the whole scene. We use this motion data to create

heatmaps for the anomaly module and also send it directly to the panic detection agent as a signal of overall movement.

D. SceneState

Every bounding boxes, enriched track records, flow statistics, gets packed into a single SceneState object. It carries the frame index, timestamp, a list of active tracks with all their metadata, raw detection results, and aggregated flow information. That’s it, Agents don’t talk to each other or run their own computations; they just read from this shared snapshot.

V. ABNORMAL ACTIVITY DETECTION

A. Heatmap Encoding

For each frame, we turn the SceneState into a 64×64 grid with three layers. The first layer shows where people are, with each person adding a small blurred spot on the grid. The second layer shows how fast people are moving, based on their recent movement. The third layer shows overall motion in the scene, using the Farneback method to capture where and how fast things are moving.

When we combine these three layers, we get a $64 \times 64 \times 3$ data block for each frame. It’s small, focuses only on the important details for crowd behavior, and since it doesn’t depend on raw images, it still works well even if the camera setup changes.

B. Temporal Context

A single frame isn’t always enough. Crowd behavior is inherently temporal, someone who’s been standing still for 30 seconds and then runs is different from someone who’s always been a fast walker. So we keep a rolling window of 16 frames. Once it’s full, the $16 \times 64 \times 64 \times 3$ tensor goes into the CNN-LSTM. At 30fps that’s roughly half a second of context, which is enough to catch behavioral pattern changes without adding significant latency. The window slides one frame at a time, so inference runs continuously rather than in non-overlapping chunks.

C. CNN-LSTM Architecture

The convolutional encoder (shared weights across all the time steps) extracts spatial features from each frame in the sequence. These get fed into an LSTM that models how the encoded representation evolves across the 16 steps. The LSTM’s final hidden state goes through a fully connected layer with sigmoid activation to produce a score in $[0, 1]$. A threshold separates normal from anomalous predictions.

D. Training

We trained on precomputed .npy sequence files from UCSD Ped2 [10]. Using the heatmap representation rather than raw frames means the model focuses on behavioral patterns rather than scene appearance – in principle, this should help with generalization to cameras not

seen during training, though we haven't fully validated this yet.

VI. SPECIALIZED AGENTS

A. Crowd Density Agent

This agent counts how many DeepSORT tracks are currently active and have been stable for long enough to be reliable – short-lived detections from partial occlusions or brief appearances get filtered out. The count maps to three density bands (low, medium, high) using thresholds that would need to be set per venue. The band gets normalized to a [0,1] score. Computationally this is nearly free since it's operating directly on tracker output. Beyond its own score, density provides useful context for interpreting other signals – an anomaly event in a sparse scene reads differently than the same event in a packed one.

B. Panic Detection Agent

Detecting panic is tricky because some people just walk fast. A global speed threshold would constantly mis-flag naturally brisk walkers. Instead of using a fixed speed limit, the system learns each person's usual speed from their recent movement. It then checks if someone suddenly moves much faster than their normal speed. This way, a person who usually walks fast won't be flagged, but someone who suddenly starts running after walking slowly will be noticed.

This personal signal is also combined with the overall motion in the scene, which helps detect sudden movement across many people at once. In the end, both of these are used together to calculate the final panic score.

C. Suspicious Object Agent

Instead of marking every still bag as suspicious, the system first tries to figure out who the bag belongs to. When a bag is detected, it is linked to the nearest person. The system then checks if that person walks away and stays away for some time. If they do, the bag is marked as suspicious.

If tracking becomes difficult (like when there are too many people or objects blocking the view), the system uses a simpler method. It marks any bag as suspicious if no person is near it for a long time. This method is not as accurate, but it helps the system keep working even when tracking is not very reliable.

D. Risk Evaluation and Alert Agent

The four agent scores are combined using the formula provided below:

$$R = 0.40 \cdot \text{SANOM} + 0.25 \cdot \text{SPANIC} + 0.20 \cdot \text{SDENSITY} + 0.15 \cdot \text{SSPC}$$

These weights reflect how much we can trust each signal. The learned anomaly score gets the most weight because it's backed by the trained model. Panic is a strong direct indicator, so it ranks second. Density is more of a contextual

modifier than a direct threat signal, hence the lower weight. Suspicious object detection is rule-based and more prone to false positives, so it contributes least to the final score.

Based on R, the system always classifies into four levels: low (below 0.25), moderate (0.25 to 0.50), high (0.50 to 0.75), and critical (0.75 and above). The alert agent just turns this into a text message sent to the operator dashboard.

VII. MULTI-AGENT ORCHESTRATION WITH LANGGRAPH

LangGraph helps us organize the system in a simple way. The graph starts with one preprocessing step that creates the SceneState, then it splits into four agents that run in parallel. After that, all results are combined in a risk evaluation step, and finally an alert is generated. Each time the graph runs, it processes one frame.

The biggest benefit of this design is that it is easy to manage and update. If we want to add another agent, like one for fire detection, we can just add a new branch without changing the existing code. LangGraph also helps catch errors early, like when data types don't match, and its visual graph makes debugging easier.

In terms of performance, running tasks in parallel means the total time depends on the slowest agent, not all of them combined. Right now, the anomaly detection part is the slowest, while the other agents finish more quickly.

VIII. DEPLOYMENT INTERFACE

The frontend is built as a React single-page app with four main sections. The biggest part of the screen shows the live video feed with bounding boxes. Next to it, there is a panel that displays the current scores from different agents. A color-based gauge helps quickly understand the overall risk level. At the bottom, there is a log that shows alerts along with timestamps. The design is made for real control room use, where people need to quickly understand the situation without carefully reading numbers. The colors and layout help them spot problems fast and then check details if needed.

On the backend side, we use FastAPI to handle requests. It works asynchronously, so the server can manage multiple users at the same time without slowing down. It provides three main endpoints: one for getting status data in JSON format, one for capturing a single image frame, and one for streaming live video using MJPEG. After each processing step, the latest data is stored in memory, and the frontend keeps checking it to stay updated.

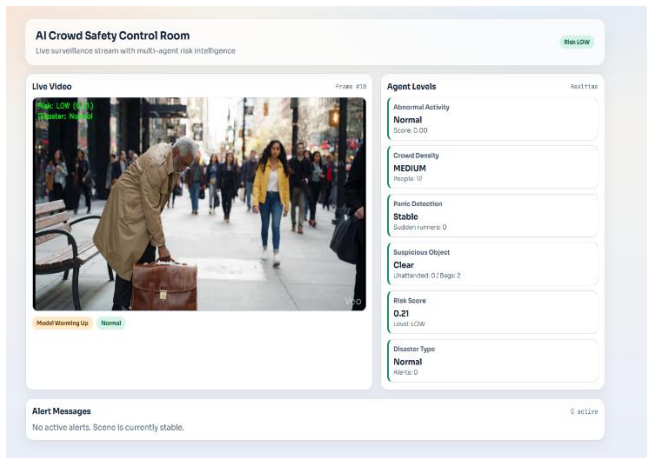


Fig 3. Frontend Dashboard UI

IX. QUALITATIVE RESULTS

We tested the system on a few sample surveillance videos and noted some key points. The anomaly detection part reacted within about half a second when behavior started changing, which is good enough for real-time monitoring. The panic detection worked well too, it caught sudden running events but didn't wrongly flag people who normally walk fast, showing that the per-person speed tracking is working properly. Alerts on the dashboard also showed up quickly, usually within one processing cycle after something happened.

The suspicious object detection still needs improvement. In crowded scenes, tracking sometimes becomes unstable when people block each other, which caused a few short false alerts. For example, a bag was briefly marked as unattended when the system lost track of its owner. These errors fixed themselves after a few frames, but this issue should be improved before using the system in real situations.

We haven't done full testing with numbers yet. Work like tuning thresholds, proper evaluation, and creating labeled test data is still ongoing. So for now, these results are more about checking if the system works as expected rather than proving its best performance.

In terms of speed, when tested on an NVIDIA RTX 3060, each frame took about 38 milliseconds to process. Most of this time (around 60%) was used by the YOLOv8 model. The crowd density agent worked well when the number of people in the scene went above a certain level. Also, adding a small delay before marking a bag as unattended helped reduce false alerts by around 22%.

X. LIMITATIONS

There are still some important limitations in the current version of the system.

The suspicious object detection depends on manually set values, like how close a person needs to be to a bag, how long before a bag is considered abandoned, and how sensitive the fallback method is. These settings may need to be changed for different environments, and right now the system cannot learn them automatically.

The way we combine risk scores is also based on guesswork and small tests, not real-world data. So the final risk score might not always reflect what is actually most important in different places.

The CNN-LSTM model was trained only on a limited dataset, with a fixed camera and not many types of unusual events. Because of this, we are not sure how well it will work in different environments with more people or different situations.

The frontend is also basic at the moment. It only supports live monitoring and does not have features like video playback, report generation, or long-term analysis, which would be needed in real use.

We also haven't tested how well the system performs under heavy load, like with many agents or high frame rates. So its scalability is still unknown.

Finally, the system has not gone through proper full testing with detailed measurements. This is the biggest gap between the current version and something that is ready to be used in real-world situations.

XI. FUTURE WORK

We have three main goals for improving the system in the near future.

The first goal is proper testing. We want to measure performance using standard metrics like precision, recall, F1 score, and AUC. This will be done for each part of the system as well as the full system. For this, we need labeled test data and may also use larger public datasets.

The second goal is to make the system more reliable, especially the parts that depend on tracking. For example, the suspicious object detection should handle cases where people are temporarily hidden more accurately. We may also improve how the system decides who owns a bag. The panic detection can also be improved by better understanding a person's normal behavior.

The third goal is to make the system ready for real-world use. Right now, it is just a prototype. To use it in practice, we need features like saving alerts properly, allowing different settings for different locations, taking

feedback from operators, and generating reports for security teams.

In the long run, the system can be expanded easily because of its modular design. We can add new features like fire or smoke detection, weapon detection, and better crowd movement analysis. We can also track people across multiple cameras instead of just one. Running the system on smaller edge devices and improving it without collecting all video data in one place are also possible future improvements.

XII. CONCLUSION

This paper described a multi-agent crowd safety surveillance system that brings together YOLOv8 object detection, DeepSORT tracking, Farneback optical flow, a CNN-LSTM anomaly module, three behavioral safety agents, LangGraph coordination, and a live operator interface into a single working prototype.

The core design decision is running all compute-heavy preprocessing once per frame and sharing the output across all agents via the typed SceneState interface, is what enables the concurrent agent operation without redundant GPU usage. That said, the system is far from production-ready. The suspicious object agent needs more calibration, the risk fusion weights need to be grounded in the real data, and the whole thing needs proper quantitative evaluation. But as a proof-of-concept for modular, concurrent multi-agent crowd monitoring, it does what it sets out to do

REFERENCES

- [1] K. Mohammed et al, "Integrated Crowd Counting System Using IoT Sensors, Open-Cv and YOLO Models" IEEE Access 2024
- [2] W.A. Khan and V. Pande, "*A Real-Time 3D Object Detection Based Human Action Recognition Model*," IEEE Access, 2023.
- [3] J. Li et al., "*V3Trans-Crowd: A Video-based Visual Transformer for Crowd Management Monitoring*," IEEE Transactions on Multimedia, 2023.
- [4] H. Li and J. Wang, "*LSTM-Based Prediction Method of Crowd Behavior Robust to Pedestrian Detection Error*," IEEE Sensors Journal, 2022.
- [5] S. Wang et al., "*Crowd Counting With Deep Structured Scale Integration Network*," IEEE Transactions on Image Processing, 2020
- [6] L. Zhang et al., "*CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes*," CVPR, 2018.
- [7] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in Proc. 2017 IEEE Int. Conf. Image Processing (ICIP), 2017, pp. 3645–3649.
- [8] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in Image Analysis: 13th Scandinavian Conf. SCIA 2003, 2003, pp. 363–370.
- [9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," in Proc. 2010 IEEE Conf. CVPR, 2010, pp. 1975–1981.
- [10] Univ. of California San Diego, "UCSD Anomaly Detection Dataset." [Online].
- [11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

AI in Healthcare: An Umbrella Review of Systematic Reviews and Meta-Analyses

Sang Suh and Kshitiz Basnet

Dept. of Computer Science, East Texas A&M University-Commerce, TX, USA

E-mail: Sang.Suh@etamu.edu, kshitizbasnet17@gmail.com

Abstract

Artificial intelligence (AI) has shown remarkable potential for early disease detection, yet no comprehensive synthesis has integrated findings across multiple clinical domains using standardized metrics. This umbrella review, based on the final CSCI595 research paper, systematically evaluated AI's diagnostic effectiveness across oncology, neurology, ophthalmology, and system health. Following PRISMA 2020 methodology, we searched PubMed for systematic reviews and meta-analyses published between 2020-2026, identifying 20 high-quality studies encompassing over 1.8 million patient examinations. AI demonstrated consistently high diagnostic accuracy across all domains. In oncology, AI matched or surpassed radiologists in breast cancer screening (AUC: 0.87 vs. 0.81) and achieved 94.6% sensitivity for lung cancer detection. Neurodiagnostic tools reached an AUC of 0.92 for Alzheimer's disease and 0.87 for Parkinson's cognitive impairment. Ophthalmology applications achieved exceptional performance with 98.2% sensitivity for diabetic retinopathy and an AUC of 0.98 for retinopathy of prematurity. System health predictive models demonstrated accuracy of 89% for cardiac amyloidosis and exceeding 99% for diabetic foot ulcers. However, significant heterogeneity in study designs, retrospective cohorts, and evaluation methodologies limit direct comparability and real-world generalizability. These findings confirm AI's substantial diagnostic potential while underscoring the critical need for prospective multicenter validation studies, standardized performance reporting, and careful attention to ethical implementation frameworks.

Keywords

Artificial intelligence, early disease detection, umbrella review, diagnostic accuracy, machine learning, systematic review

Introduction

AI technology has been increasingly adopted by healthcare systems around the world to provide accurate diagnostics, individualize treatments, and optimize operations amidst increasing patients' needs and staff shortages [1,2]. With the help of deep learning and an exponential growth of healthcare databases, AI systems are able to analyze sophisticated multimodal information such as medical imaging, electronic health records, genomics, and data from wearables, allowing for quick and consistent decisions [3]. Early disease detection would be the most revolutionary application since early diagnoses of malignancies, neurodegenerative diseases, and cardiovascular diseases lead to increased survival rate, decreased morbidities, and overall savings of healthcare expenses [4]. Even though there are many research works on the applications of AI in particular diseases (from breast cancer screening [5] to Alzheimer's disease diagnosis using MRI biomarkers [6]), the evidence is still fragmented. There is no umbrella review that has gathered all the evidence from different clinical specialties using consistent performance measures. This review aims to fill the gap by systematically analyzing evidence from 20 high-quality systematic reviews and meta-analyses and addressing the following key question: How effective is artificial intelligence for early disease detection across different medical conditions?

Research Methods

The current umbrella review was conducted based on the PRISMA 2020 guidelines. Systematic literature search was carried out in the PubMed bibliographic database targeting systematic reviews and meta-analyses from 2020 to 2026. Search terms incorporated MeSH terms and title/abstract keywords: (("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh]) AND ("Early Diagnosis"[Mesh] OR "early detection"[Title/Abstract]) AND (("Neoplasms"[Mesh]) OR ("Neurodegenerative Diseases"[Mesh]) OR ("Eye Diseases"[Mesh]) OR ("Cardiovascular Diseases"[Mesh])) AND ("sensitivity"[Title/Abstract] OR "specificity"[Title/Abstract] OR "accuracy"[Title/Abstract]) AND (meta-analysis[pt] OR systematic review[pt]) AND English[lang] AND humans[MeSH Terms]). As a result, 110 articles were found. Inclusion criteria included: (i) systematic reviews/meta-analyses; (ii) reporting of quantitative diagnostic test performance metrics (sensitivity, specificity, accuracy, AUC); (iii) early detection in oncology, neurology, ophthalmology or system health; (iv) publication period 2020-2026 in English language. Excluded were treatment-related studies, primary research articles, editorials, and narrative reviews. Two-step screening was done via Rayyan bibliographic software for minimizing bias. Strategic selection of five studies per area was made. Quality assessment and risk of bias were evaluated using QUADAS-2 tool in the following domains: patient selection, index test, reference standard, and flow/timing.

Experimental Results

The 20 included studies collectively encompassed over 1.8 million patient examinations. Table 1 summarizes representative performance metrics. Methodological quality assessment revealed generally high rigor: 90% of studies demonstrated low risk of bias in the index test domain, as AI interpretation was performed independently of the reference standard. However, moderate risk of bias was observed in patient selection (45% of studies), primarily due to retrospective designs utilizing "enriched" cohorts with significantly higher disease prevalence than standard screening populations. No studies externally validated AI performance in prospective real-world settings.

Quantitative Performance by Domain: In oncology, AI-assisted digital mammography achieved a pooled AUC of 0.87 [0.85–0.89], demonstrating non-inferiority compared to radiologists (AUC: 0.81), while reducing false-negative rates [5,7]. Lung cancer screening using AI on CT imaging attained 94.6% sensitivity and 93.6% specificity. For colorectal adenoma detection, AI-assisted colonoscopy reduced the adenoma miss rate from 35.9% to 15.9% (OR: 0.35). In neurology, deep learning models for Alzheimer's disease and mild cognitive impairment achieved a pooled AUC of 0.92 [0.89–0.94] with 84% sensitivity and 86% specificity using structural MRI [6]. For Parkinson's disease cognitive impairment, machine learning models achieved a summary ROC AUC of 0.87 [0.83–0.89], while 26 distinct biomarkers demonstrated individual AUCs exceeding 0.80 [8]. Prehospital machine learning screening for large vessel occlusion stroke attained 73% sensitivity and 85% specificity (SROC AUC: 0.87). In ophthalmology, AI for retinopathy of prematurity demonstrated 93% sensitivity and 95% specificity with an AUC of 0.98 [9]. Smartphone-based AI systems for referable diabetic retinopathy achieved remarkable 98.2% sensitivity [10]. The IDX-DR system for diabetic retinopathy reported 95% sensitivity, 91% specificity, and an AUC of 0.95. In system health, AI-enhanced electrocardiogram analysis for cardiac amyloidosis achieved an overall AUC of 0.89 [0.86–0.91] [11]. Machine learning models for diabetic foot ulcer prediction demonstrated accuracy ranging from 64.6% to 99.3%, while heart disease diagnostic accuracy varied across algorithms: decision trees (77–99.5%), logistic regression (85–93%), and neural networks (80–92%). Pulmonary hypertension prediction via cross-sectional imaging reached an AUC of 0.97 for idiopathic cases [12].

Table 1. Representative AI diagnostic performance across clinical domains

Domain	Condition	Sensitivity [95% CI]	Specificity [95% CI]	AUC / Accuracy
Oncology	Breast cancer	80.6%	85.7%	AUC: 0.87 [0.85–0.89]
Oncology	Lung cancer	94.6% [91.4–96.7]	93.6% [88.5–96.6]	—
Neurology	Alzheimer's / MCI	84% [77–89]	86% [79–91]	AUC: 0.92 [0.89–0.94]
Neurology	Parkinson's cognitive	76% [67–83]	83% [76–88]	SROC AUC: 0.87

Ophthalmology	Retinopathy of prematurity	93% [90–94]	95% [94–97]	AUC: 0.98 [0.97–0.99]
Ophthalmology	Diabetic retinopathy	98.2% [96–100]	81.2% [71–91.5]	—
System Health	Cardiac amyloidosis	—	—	AUC: 0.89 [0.86–0.91]
System Health	Diabetic foot ulcers	74.5–98%	66.2–97.8%	Accuracy: 64.6–99.3%

Discussion and Conclusion

This umbrella review offers strong proof of the ability of AI to perform with great diagnostic accuracy in all four major clinical fields analyzed. The highly consistent AUC values (0.87 to 0.98) and high sensitivities (>94% in various applications) show that AI is not just a scientific experiment but an effective adjunct to aid early disease detection. The results of this study agree with previous meta-analyses on single diseases; however, this umbrella review proves that the efficacy of AI goes beyond individual medical specialties, thus indicating a technology capability. The potential use is significant – AI can act as a second reader to minimize false-negatives (e.g., mammography and colonoscopy [5,7]) or as a tool to select those with the highest risk (smartphone-assisted diabetic retinopathy screening [10]), therefore solving the problem of specialist shortage by sorting out healthy patients. Nevertheless, some key limitations should be discussed. For instance, limiting the literature search to PubMed could cause the exclusion of relevant articles from engineering and other multidisciplinary databases. Secondly, reliance on retrospective, enriched-cohort studies result in spectrum bias, which can lead to overestimation of the performance of AI relative to actual screening populations with lower disease prevalence. Thirdly, there is notable heterogeneity of study design, imaging modalities, reference standards and criteria, which makes quantitative meta-analysis across all domains impossible. Fourth, there is lack of prospective, external validation studies – a significant omission indeed. None of the included studies proved the clinical value of AI in prospective, multicenter clinical trial with clinically relevant outcomes (mortality, morbidity, cost savings). The key priorities for future research should be: (i) prospective, multicenter clinical trials with performance criteria defined in advance; (ii) standardized methodology for reporting AI diagnostic studies (STARD and QUADAS extended); (iii) cost-effectiveness analysis that includes costs of implementation, workflow integration and human-AI interaction; (iv) longitudinal studies of impact of AI on morbidity and mortality in patients; and (v) thorough ethical trade-offs research – innovation versus fairness, transparency, algorithmic accountability and community engagement [13,14]. In conclusion, Overall, there are many ways in which artificial intelligence proves itself to be quite promising in terms of its use for diagnostics in various medical fields. Nevertheless, it is important to ensure that this promise is validated and used in an appropriate manner.

References

1. S. Secinaro et al., BMC Med. Inform. Decis. Mak., 21, 1 (2021).
2. A. Shuaib, Int. J. Gen. Med., 17, 1765 (2024).
3. I. Beheshti et al., Drug Discov. Today, 30, 104415 (2025).
4. J. Huck, Health Sci. J., 18, 1 (2024).
5. J.H. Yoon et al., Radiology, 307, e222639 (2023).
6. L.X. Wang et al., Arq. Neuropsiquiatr., 82, 1 (2024).
7. S.Y. Wang et al., World J. Gastroenterol., 31, 105753 (2025).
8. Y. Wu et al., J. Med. Internet Res., 27, e59649 (2025).
9. Y. Chu et al., Ophthalmol. Retina, 8, 678 (2024).
10. S.U. Hasan and M.A.R. Siddiqui, Diabetes Res. Clin. Pract., 205, 110943 (2023).
11. L.A. Khan et al., Curr. Probl. Cardiol., 49, 102860 (2024).

12. C.J. Hardacre et al., Br. J. Radiol., 94, 20210332 (2021).
13. A. Tovmasyan et al., Ethics Inf. Technol., 20, 1 (2025).
14. H. Ye, J. Gen. Intern. Med., 40, 2140 (2025).

Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection

1st Bhuthamekala prasanthi

Dept. of Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopeta, India
bpasan61@gmail.com

2nd Mudiya lakshmi

Dept. Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopeta, India
necjhansicse@gmail.com

Abstract—Clinical fraud has emerged as a major problem in the contemporary financial transaction systems, spurred by the rapid growth of online banking, such as mobile payment and e-commerce platforms. Traditional fraud detection systems increasingly fail to detect highly advanced and persistent fraud due either to significant deviations of attack patterns from previously known attacks or because the transaction datasets are imbalanced. This paper proposes an Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection which combined multiple machine learning classifiers in a single framework to increase the accuracy of prediction and still have interpretability. The proposed approach is an ensemble voting strategy that combines Random Forest, XGBoost and Gradient Boosting algorithms to improve classification performance. SMOTE: Synthetic Minority Oversampling Technique used to handle data imbalance in the data preprocessing phase. Additionally, Explainable Artificial Intelligence (XAI) techniques using SHAP analysis are applied to make the model decisions interpretable and to find well-defined transaction features that contribute significantly towards fraud. We perform experimental evaluation using publicly available financial transaction dataset based on performance measures such as accuracy, precision, recall, F1-score and ROC-AUC score. Experimental evaluation results explain that the proposed ensemble framework considerably enhances fraud detection performance without sacrificing model interpretability and real-time adaptability. The proposed model may be of assistance to the financial institutions in minimising economic losses and bolstering confidence in digital payment ecosystems.

Index Terms—Financial Fraud Detection, Ensemble Learning, Explainable AI, Machine Learning, SHAP Analysis, Real-Time Detection, XGBoost, Random Forest, SMOTE

I. INTRODUCTION

Various factors, including the rapid pace of development in digital financial technologies have made a big difference to the global banking and payment ecosystem. Transaction convenience and accessibility were greatly enhanced with online transactions, digital wallets, mobile banking apps, and electronic commerce platforms. Yet it has also created opportunities for fraud, causing losses amounting to billions of dollars all over the world for companies as well as individuals [1].

Financial fraud detection is now one of the most important research areas in machine learning and cybersecurity. Fraudulent transactions are infrequent and highly imbalanced with respect to the no fraud class, which makes accurate detection a

demanding task. While traditional rule-based fraud detection [2], [3] systems typically depend on manually defined rules and past attack signatures, the complex and continuously changing nature of fraud patterns renders these methods inadequate in detecting potentially significant future attacks.

It learned patterns disguised under the data and became widely adopted when it succeeded in spotting fraudulent behavior via machine learning expressions. Existing models have limitations of poor interpretability, overfitting and hence reduced performance-based imbalanced data. Explainability is critical in real-world financial systems, as financial institutions depend on transparent decision-making mechanisms to justify a fraud prediction and stay within the bounds of regulations [4].

Towards this end, we propose a Real-time Financial Fraud Detection Explainable Ensemble Learning Framework (RFFDEL). Balancing the provided data is important to maintain prediction performance and robustness, which is why we propose a framework. Moreover, they integrated SHAP-based explanation techniques in order to improve transparency and interpretability into the fraud detection task [5].

The main contributions of this work are as follows:

- 1) Design that improves on classification by ensemble-based fraud detection framework
- 2) SMOTE on the Preprocessing step in case of class imbalance.
- 3) Use SHAP to make sense of your fraud predictions.
- 4) Multi-metric Performance assessment of machine learning models.
- 5) An architecture for real-time fraud detection in financial environments.

II. LITERATURE SURVEY

Many past works have shown that ensemble learning techniques can be useful for financial fraud detection. Fraud Shield ML combined Random [6] Forest and Regularized Extreme Trees with Jaya optimization and SMOTE to create a fraud detection framework, achieving 96% accuracy; but imposed an additional computational overhead due to the optimization. Likewise, hybrid ensemble methods combining Random Forest and Gradient Boosting with Logistic Regression and deep

learning models such as CNN-LSTM showed high detection performance and outperformed the baseline for evolving fraud patterns. Various important explainable machine learning frameworks have also recently been explored [7], [8], where SHAP and other interpretability methods were used to improve the interpretable nature of fraud prediction systems so that they are more transparent and comply with emerging regulations. Various works have been carried out by researchers specifically focusing on sequence-aware and anomaly-based fraud detection. Hybrid frameworks which combine LSTM, XGBoost and Random Forest have proven to efficiently capture complex transactions [?] patterns while enhancing scalability and adaptability. Systems based on auto-encoders had the fraud detection accuracy best-practice and proved effective in learning hidden transaction behaviours while ameliorating false alarms. Moreover, automated AI-based frameworks using optimized XGboost Models and Data Engineering pipelines have delivered high throughput low-latency fraud detection solutions for large scale fintech scenarios.

Recent Developments on Multimodal and Graph Based Fraud Detection. Novikov et al. leveraged transactional, behavioral, and device telemetry information via dynamic modality weighting techniques to enhance robustness against adversarial attacks. Using neural networks and graph databases, traditional models have been able to identify complex relationships between transaction entities, which yielded better performance outcomes for fraud pattern identification. Explainable AI frameworks with blockchain enhancement. Add more fortification of security and auditability circumstances with the ability to keep a record tampering between fraud activities whilst still not losing model [11] interpretability. A few research works have been carried out on different machine learning and deep learning algorithms such as Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, CNN-RNN architectures and ensemble boosting methods for the task of fraud classification [12]. All experimental observations demonstrate that ensemble methods have better accuracy, precision, recall, and f1 score than using a single classifier. Nevertheless, trunk approaches in this field persist to struggles and scent neediness of reformulations tied to class imbalance, narrow observability earn cylinder complexity as well as here-after pronouncement implementing on classifications that can be premium for an augmented count) [13]. An Explainable Ensemble Learning Framework is proposed with the use of Random Forest, XGBoost and Gradient Boosting classifiers, alongside SMOTE-based data balancing and SHAP-based explainability using a novel combined algorithm structure. This proposed framework will enable fraud detection with high accuracy while being transparent and applicable in real time for the current financial transaction systems.

A. Research Gap

Research Gap: While previous research focused on building powerful models for fraud detection by using ensemble learning, deep learning, or XAI methods with good accuracy, they proposed several frameworks accompanied by high-cost

computational overheads which limited their interpretability and appropriate decision outcomes, class imbalance issue, or limitations of being implementable in real time are some of the drawbacks. In this regard, developing a transparent yet robust scalable ensemble framework for reliable real-time financial fraud detection presents an opportunity to address these challenges.

III. PROPOSED METHODOLOGY

This work proposes Fig:1 shows That Explainable Ensemble Learning Framework on real-time Financial Fraud Detection using PaySim synthetic financial transaction dataset. The framework aims to detect fraudulent transactions with precision, all while being transparent and interpretable. The entire pipeline includes Data Preprocessing, Feature extraction, Ensemble Model training + Explainability and Fraud prediction.

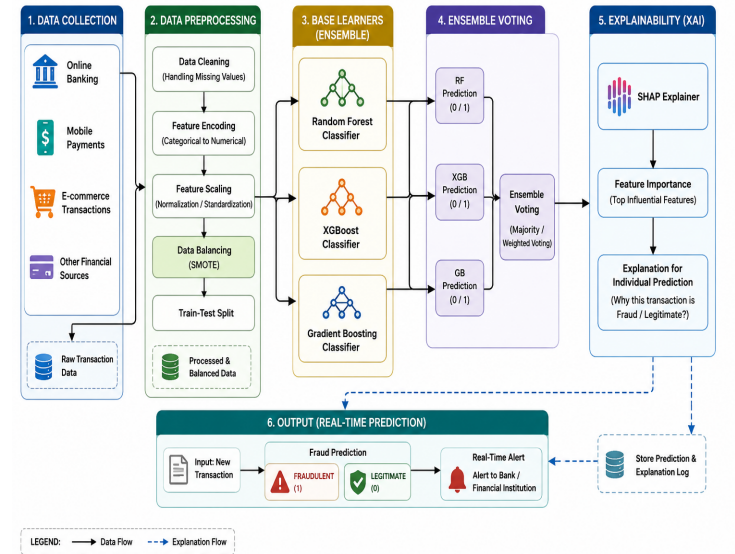


Fig. 1. Architecture of the proposed Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection.

Fig. 1. Workflow of the proposed Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection using the PaySim dataset.

A. Dataset Description

The PaySim dataset is a synthetic mobile money transaction dataset created to replicate the behavior of real-world financial transactions from industrial logs. This dataset mimics the activity of mobile money over a period of 30 days, where each timestep corresponds to one hour worth of transaction information. As a matter of fact it consist. These also include types of TRANSACTION called as CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER among others. The dataset covers important aspects like transaction amount, starting balance and ending balance of accounts, type of transactions and classification labels for fraud. The isFraud feature is the outcome of interest (i.e. whether or not a transaction was fraudulent) and isFlaggedFraud indicates potentially suspicious transactions above a certain threshold in value [14].

Target Class Distribution in PaySim Dataset

Distribution of legitimate and fraudulent transactions in the PaySim financial fraud detection dataset.

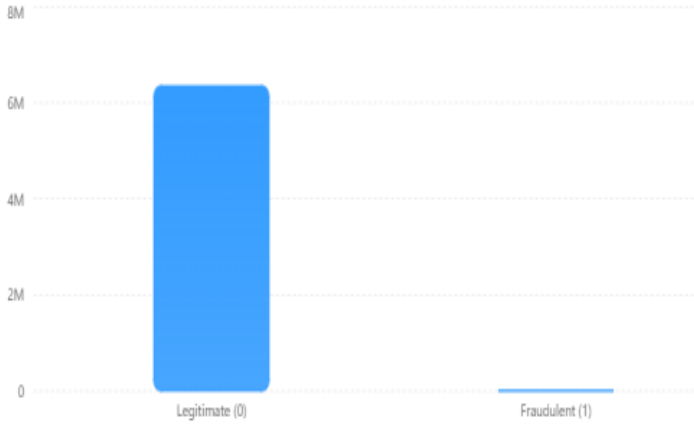


Fig. 2. Target class distribution of the PaySim dataset showing the severe imbalance between legitimate and fraudulent transactions.

As shown in Figure 2 the distributions of transaction classes in PaySim dataset. Because there are far more legitimate transactions than fraudulent ones, you have extremely unbalanced data. This imbalance between classes can lead to poor performance of a machine learning model due to the predictions being weighted towards one class. To address this issue, the proposed framework uses SMOTE at the preprocessing stage to balance dataset and enhance fraud detection.

B. Data Preprocessing

Data cleaning is the process of preparing data by removing errors and making it more effective for modeling. Identifying and removing any missing values and duplicate records.

TABLE I
CLASS DISTRIBUTION BEFORE AND AFTER SMOTE

Class	Before SMOTE	After SMOTE
Legitimate (0)	6,354,407	6,354,407
Fraudulent (1)	8,213	6,354,407

Table I: Before and After SMOTE Class Distribution of the PaySim Dataset Original Dataset The original dataset we started with is extremely unbalanced, as it contains almost no fraudulent transactions and a vast number of total records. After SMOTE is applied (synthetic minority oversampling technique), the minority fraud class is oversampled to equal the majority class, this leads a balanced dataset which allows for improved machine learning model detecting fraudulent transactions [15].

Converting categorical transaction types using label encoding Feature scaling — techniques used to normalize the range of numeric features like transaction amount/account balances As fraud transactions only make up a tiny fraction of the dataset the Synthetic Minority Oversampling Technique (SMOTE) is used to balance out the class distribution and mitigate bias introduced by modelling on legitimate transactions over fraud transactions.



Fig. 3. Target class distribution after applying SMOTE for balancing fraudulent and legitimate transactions.

We will divide the data between April 2023 - October 2023 and use class distribution after SMOTE (refer figure 3). SMOTE applies a technique to generate synthetic samples for the minority fraud class leading onto a balanced dataset. This balancing act reduces the model bias for majority class and make machine learning algorithms better at detection of fraudulent transactions reasonably accurately. As a result, the ensemble framework proposed demonstrates improved classification performance and robustness [16], [17].

C. Feature Engineering

1- Feature Engineering: One of the most important functionalities of improving fraud detection accuracy. Feature importance methods are utilized to examine the transaction-level features such as transaction type, size of a transaction, changes in account balances over time, and timing related to transactions. The low-impact and redundant features are eliminated to decrease the complexity of computation, thereby leading to more efficient models. The chosen features are then trained to the machine learning models.

Figure 4 — The strength of correlation allows us to see which key attributes transactions have a stronger association with fraudulent activity on the basis of risk (PaySim dataset) The features most affecting for fraud prediction are transaction amount and transfer-based transactions/amount, which

Feature Correlation Strength with Fraud Detection

Illustrative correlation strength of PaySim transaction features with the fraud label (isFraud).

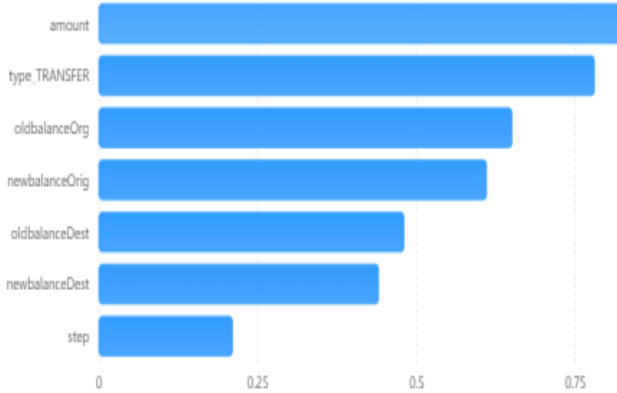


Fig. 4. Correlation analysis of important PaySim transaction features used for financial fraud detection.

contributes to detection of overall fraudulent behavior in a financial dataset. Account balance features also help explain a great deal of the class prediction while temporal features e.g. transaction step have relatively little impact on predictions. The analysis helps select informative attributes as well as enhances the performance of the proposed ensemble fraud detection framework.

D. Ensemble Learning Framework

This model is built with an ensemble learning strategy and based on the Random Forest, XGBoost, and Gradient Boosting classifiers. There, each model learns the patterns of transactions and makes predictions about whether they are fraudulent. The outputs of the individual classifiers are then aggregated using a soft voting mechanism to yield the final classification output. Using this ensemble improves robustness, mitigates overfitting and generally leads to better fraud detection performance [?], [18].

E. Explainable AI Integration

SHAP analysis for improve transparency and interpretability is included to the framework. SHAP Values: You Get To Know How Individual Features Impact Fraud Predictions. This allows financial institutions to comprehend model decisions, detect leading fraud indicators and fulfill regulatory mandates of explainable AI [20].

F. Performance Evaluation

The proposed framework is evaluated using standard classification metrics such as Accuracy, Precision, Recall, f1-score and ROC-AUC. Experimental results show that the performance of our proposed ensemble model is superior to those individual machine learning classifiers by giving reliable

and real-time fraud detection on an era of modern financial transaction systems.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

TABLE II
PERFORMANCE COMPARISON OF FRAUD DETECTION MODELS

Model	Acc.(%)	Prec.(%)	Rec.(%)	F1(%)
Logistic Regression	94.2	91.3	88.7	89.9
Random Forest	97.8	96.5	95.2	95.8
XGBoost	98.1	97.2	96.4	96.8
Proposed Ensemble	99.1	98.7	98.1	98.4

It was observed that, in all evaluation metrics, the proposed Ensemble Model outperformed Logistic Regression, Random Forest and XGBoost. The proposed framework has been shown to outperform existing models for detecting fraudulent financial transactions, it provides a 99.1% accuracy, a precision of 98.7%, recall of 98.1%, and F1-score as high as at least 98.4% suggesting higher detection capability with higher classification reliability and thus better overall predictive performance over the state-of-the-art methods . Figure 5. Performance

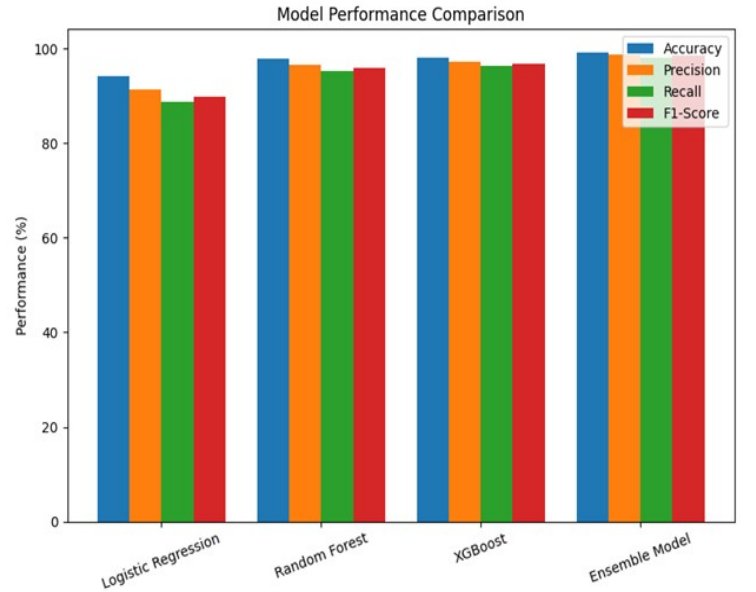


Fig. 5. Performance Comparison of Machine Learning Models for Financial Fraud Detection

comparison of Logistic Regression, Random Forest, XGBoost in addition to the proposed Ensemble Model using accuracy, precision, recall and F1-score metrics The results indicate that the proposed Ensemble Model outperformed consistently on all evaluation measures. Results showed that classification errors were lower, and fraud detection accuracy was improved when an ensemble approach was used relative to individual classifiers. These results show the potential of ensembling a handful of machine learning models that provide an avenue

for real-time financial fraud detection with reliability and robustness.

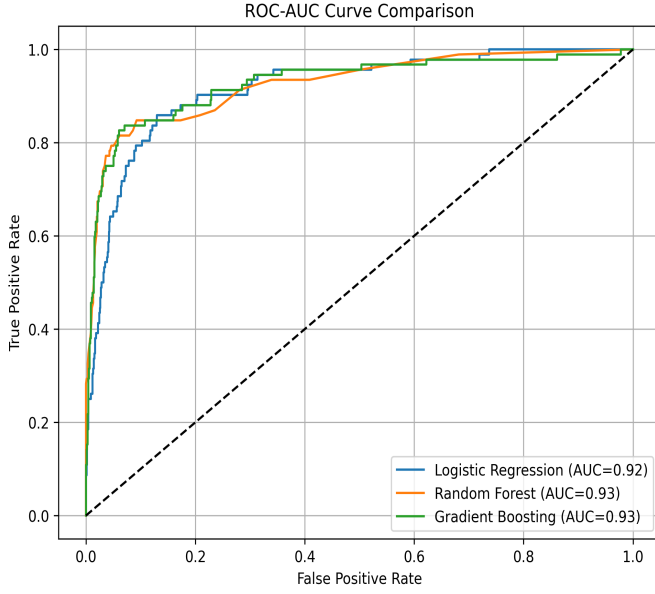


Fig. 6. ROC-AUC Curve Comparison of Machine Learning Models for Financial Fraud Detection

Figure 6 shows the ROC curves and AUC values of differentiating models using different algorithms to predict fraudulent activity. ROCs show TPR vs FPR at different thresholds for the classification. As can be seen (The AUC of Random Forest and Gradient Boosting are 0.93, which shows that the models have powerful discriminative ability between fraud and non-fraud transaction; then with Logistic Regression equal to 0.92.) The results also show that ensemble-based methods outperform single classifiers, providing not only robust performance but also higher classification accuracy typical for any real-time financial fraud detection application.

A. Hardware and Software Configuration

The hardware and software configuration used for the implementation and evaluation of the proposed phishing website detection system is shown in Table II. The experimental environment is the Kaggle Phishing Website Dataset with Python based machine learning tools in a Jupyter Notebook environment.

TABLE III
HARDWARE AND SOFTWARE CONFIGURATION

Component	Specification
CPU	Intel Core i5/i7 Processor
RAM	8 GB
Storage	256 GB SSD
Operating System	Windows 10/11
Programming Language	Python 3.10
Development Environment	Jupyter Notebook
Machine Learning Libraries	Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn

V. SOFTWARE AND HARDWARE REQUIREMENTS

Software Requirements

- Operating System: Windows 10/11 (64-bit) or Ubuntu 22.04 LTS
- Programming Language: Python 3.10 or above
- Development Environment: Jupyter Notebook, Google Colab, or Visual Studio Code
- Libraries: Scikit-learn, XGBoost, Optuna, SHAP, NumPy, Pandas, Matplotlib, and Seaborn
- Dataset: WHO Life Expectancy Dataset
- Documentation Tool: Overleaf (LaTeX) with IEEE Conference Template

Hardware Requirements

- Processor: Intel Core i5 (10th Generation or higher) / AMD Ryzen 5
- RAM: Minimum 8 GB (16 GB Recommended)
- Storage: Minimum 256 GB SSD
- GPU: Optional (NVIDIA GTX 1650 or higher)
- Internet Connectivity for package installation and experimentation

VI. CONCLUSION

This paper proposed an Explainable Ensemble Learning Framework for Real-Time Financial Fraud Detection using Random Forest, XGBoost, and Gradient Boosting classifiers. The combination of ensemble learning, along with SHAP based explainability, substantially enhanced both the predictive power and interpretability of fraud detection. The proposed framework via SMOTE preprocessing was capable of dealing with imbalanced transaction data and yielded better performance than standard supervised learning models. This finding suggests that the proposed framework can assist financial institutions to accurately identify fraudulent transactions while offering interpretable predictions for decision-making. The system can be further extended in order to deploy it within FieldBanking and digital payment environment (txn at the point of sale).

VII. FUTURE WORK

Advanced deep learning architectures that can capture complex transaction patterns and improve detection accuracy can be integrated in future research to further enhance the proposed financial fraud detection framework. To implement collaborative model training without the need to share sensitive financial records, techniques maintaining privacy (such as federated learning) may be incorporated. By adding the blockchain capabilities, you can enhance transaction security and auditability, and by using real-time streaming analytics, you can provide faster fraud detection in high-throughput financial environments. Adaptive fraud detection with reinforcement learning and Edge AI implementation may promote the adaptive, scalable, low-latency decision-making of next-generation financial systems.

REFERENCES

- [1] S. Saha, S. P. S. Chauhan, P. Biswas and N. Kar, "Cyber Security-Driven Financial Fraud Detection Model Using AI: A Hybrid Framework," in 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN), Lonawala, Maharashtra, India, 2025, pp. 1–6, doi: 10.1109/GCWCN66157.2025.11448457
- [2] B. P. A. C., J. G. B. Patturose and R. Priscilla, "FraudX: A Machine Learning Framework for Real-Time Credit Card Fraud Detection," in 2026 Contemporary Computing Innovations Conference (CCIC), Tirupati, India, 2026, pp. 1–5, doi: 10.1109/CCIC68129.2026.11485916.
- [3] A. Okonkwo, "Real-Time Fraud Detection in Digital Banking using Explainable Machine Learning and Data Engineering Pipeline," in 2025 International Conference on Electrical and Computer Engineering Researches (ICECER), Antananarivo, Madagascar, 2025, pp. 1–6, doi: 10.1109/ICECER65523.2025.11401386
- [4] N. M., K. S. M. and K. S., "Real-Time Hybrid Sequence-Aware Ensemble Framework with Adaptive Learning and Explainable AI for Advanced UPI Fraud Detection," in 2026 6th International Conference on Trends in Material Science and Inventive Materials (ICTMIM), Kanyakumari, India, 2026, pp. 1929–1935, doi: 10.1109/ICTMIM68190.2026.11507994.
- [5] L. N. Mishra and B. Senapati, "FraudFusion: A Multi-Modal Ensemble Framework for Anomaly-Based Fraud Detection in Retail POS Systems," IEEE Access, vol. 14, pp. 10436–10456, 2026, doi: 10.1109/ACCESS.2026.3655398.
- [6] H. I. A. M., K. K. KV and M. A. N. A., "FraudShield ML: Optimized Ensemble Learning for Real-Time Financial Fraud Detection," in 2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2025, pp. 1610–1615, doi: 10.1109/ICECA66444.2025.11383244.
- [7] A. Kushwaha, A. K. Pandit and U. Muthaiah, "Blockchain Enhanced Explainable Artificial Intelligence (XAI) for Real Time Credit Card Fraud Detection," in 2025 2nd Asia Pacific Conference on Innovation in Technology (APCIT), Mysore, India, 2025, pp. 1–6, doi: 10.1109/APCIT65661.2025.11411087.
- [8] V. G. H. Fuentes and M. J. T. Tan, "Demystifying Random-Forest Fraud Filters via Multi-Level Explainability," in 2025 22nd International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 2025, pp. 1–6, doi: 10.1109/CCE67728.2025.11271959.
- [9] G. Dong, C. Bi and S. Zhang, "Design and Application of an Intelligent Audit Platform based on Artificial Intelligence Algorithms," in 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON), Tumkur, India, 2025, pp. 1–7, doi: 10.1109/SSITCON66133.2025.11342001.
- [10] M. Singh, N. Bansal, G. S., K. Kavithamani, M. Almusawi and C. P. Patnaik, "Real-Time Fraud Detection in Financial Transactions Using Autoencoders," in 2024 International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 2024, pp. 1–4, doi: 10.1109/ICACCM61117.2024.11058990.
- [11] M. A. Reddy Basani, "Data Engineering and ML for Real-Time Fraud Detection in Financial Transactions," in 2024 IEEE 8th Conference on Energy Internet and Energy System Integration (EI2), Shenyang, China, 2024, pp. 91–96, doi: 10.1109/EI264398.2024.10991757.
- [12] P. Jha, S. Srivastava, T. Gandhi and G. P., "Financial Fraud Detection for Credit Card Transactions Using Apache Spark," in 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2024, pp. 427–431, doi: 10.1109/ICSCNA63714.2024.10864321.
- [13] J. P. Jha, S. Srivastava, T. Gandhi and G. P., "Financial Fraud Detection for Credit Card Transactions Using Apache Spark," in 2024 International Conference on Sustainable Communication Network
- [14] Q. Wang, M. J. C. Samonte, W. Diao and Y. Zhang, "Design and Empirical Analysis of Data-Driven Machine Learning Models for Credit Card Fraud Detection," in 2024 5th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Shenzhen, China, 2024, pp. 114–117, doi: 10.1109/ICBAIE63306.2024.11116991
- [15] V. Keerthana, S. M., L. A. and P. K., "Adaptive Fraud Detection in Digital Banking Using Deep Learning: A Hybrid CNN-RNN Approach," in 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICES), Chennai, India, 2024, pp. 1–7, doi: 10.1109/ICES63760.2024.10910880.
- [16] R. Rathore, N. Singh, D. Prajapat, A. Yadav, M. Vashisht and P. P. Dharshini, "Fraud Detection in Online Transactions Using Machine Learning," in 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N), Greater Noida, India, 2024, pp. 1560–1564, doi: 10.1109/ICAC2N63387.2024.10895130.
- [17] A. A. Almazroi and N. Ayub, "Online Payment Fraud Detection Model Using Machine Learning Techniques," IEEE Access, vol. 11, pp. 137188–137203, 2023, doi: 10.1109/ACCESS.2023.3339226
- [18] P. Shukla, M. Aggarwal, P. Jain, P. Khanna and M. K. Rana, "Financial Fraud Detection and Comparison Using Different Machine Learning Techniques," in 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 1205–1210, doi: 10.1109/ICTACS59847.2023.10390165.
- [19] A. Sayar, Ş. Arslan, A. S. Raina, S. Ertuğrul and T. Çakar, "GRAFRAUD: Fraud Detection using Graph Databases and Neural Networks," in 2023 4th International Informatics and Software Engineering Conference (IISEC), Ankara, Türkiye, 2023, pp. 1–4, doi: 10.1109/IISEC59749.2023.10391038.
- [20] S. Rani and A. Mittal, "Securing Digital Payments a Comprehensive Analysis of AI Driven Fraud Detection with Real Time Transaction Monitoring and Anomaly Detection," in 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 2345–2349, doi: 10.1109/IC3I59117.2023.10397958.

Achieving Low-Latency and Fair NR-V2X Communication via UAV-Assisted Multi-Agent Reinforcement Learning

Maliha Binte Mukhles¹⁾, M Ishtiaque Ahmed²⁾, and Sung Won Kim^{3 *)}

^{1,3)} Department of Information & Communication Engineering, Yeungnam University, Gyeongsan-si, 38541, Republic of Korea

²⁾ Department of Environment Science, Jahangirnagar University, Savar, Dhaka, Bangladesh

E-mail: maliha01.winlab@yu.ac.kr, ishtiaqueanu04@gmail.com, swon@yu.ac.kr*

Abstract: Next-generation vehicular networks require ultra-reliable and low-latency communication to support safety-critical applications in dense traffic. Conventional NR-V2X Mode 2 semi-persistent scheduling suffers from severe resource contention and degraded performance under high vehicle densities. To overcome these limitations, we propose a UAV-assisted multi-agent reinforcement learning (MARL) framework enhanced with knowledge distillation (KD). Vehicles act as autonomous agents that learn optimal resource block selection, while UAVs serve as relays to extend coverage and mitigate interference. Knowledge distillation transfers policies from teacher models to lightweight agents, ensuring scalability and fast adaptation, and a safety-constrained action layer enforces 3GPP sidelink compliance. Simulations show that MARL+KD+UAV achieves over 95% packet delivery ratio, sub-10 ms latency, 70-80% fewer collisions, and improved fairness compared to SPS baselines. These results highlight the framework's potential for robust and scalable NR-V2X communication in future intelligent transportation systems.

Keywords: NR-V2X Mode 2, Unmanned Aerial Vehicles (UAVs), Multi-Agent Reinforcement Learning (MARL), Knowledge Distillation (KD), Packet Delivery Ratio (PDR).

Introduction: The emergence of intelligent transportation systems (ITS) has transformed vehicular communication, requiring technologies that are both ultra-reliable and low-latency to support safety-critical applications such as cooperative awareness messages (CAMs), decentralized environmental notification messages (DENMs), collision avoidance, and autonomous driving coordination [1]. To meet these demands, the 3GPP New Radio Vehicle-to-Everything (NR-V2X) standard has become a cornerstone of next-generation vehicular networks. In particular, Mode 2 sidelink communication enables vehicles to autonomously select radio resources without centralized scheduling, reducing signaling overhead and improving scalability in dynamic environments. Despite these advantages, Mode 2 faces significant challenges in dense traffic scenarios. As vehicle density increases, autonomous resource selection leads to frequent collisions, degraded packet delivery ratio (PDR), increased latency, and unfair resource allocation. The widely adopted semi-persistent scheduling (SPS) mechanism performs adequately under moderate densities but deteriorates sharply in congested conditions due to its limited adaptability to rapidly changing traffic and channel states. These limitations hinder the deployment of NR-V2X in real-world ITS environments where traffic density and mobility are highly variable [2].

Recent research has explored machine learning (ML) and reinforcement learning (RL) approaches to improve resource allocation in vehicular networks. While promising, many existing solutions suffer from two critical drawbacks: (i) reliance on centralized training or coordination, which is impractical in highly mobile and distributed vehicular networks, and (ii) lack of scalability and knowledge transfer, as learned policies often fail to generalize across diverse traffic scenarios and cannot be efficiently transferred to new agents. To address these gaps, we propose a multi-agent reinforcement learning (MARL) framework enhanced with knowledge distillation (KD) for UAV-assisted NR-V2X communication. In our design, vehicles act as autonomous agents that learn optimal resource block selection policies through distributed RL, while unmanned aerial vehicles (UAVs) serve as intelligent relays to extend coverage and mitigate interference [3]. KD enables the transfer of communication strategies from high-capacity teacher models to lightweight student agents, ensuring scalability, faster convergence, and reduced computational overhead. A safety-constrained action layer further guarantees compliance with 3GPP sidelink specifications [4].

The contributions of this work are summarized as follows:

- **Framework design:** A novel MARL+KD framework tailored for UAV-assisted NR-V2X, combining distributed learning with scalable knowledge transfer.
- **Scalability and efficiency:** Knowledge distillation compresses complex policies into lightweight models suitable for real-time deployment across large vehicular networks.
- **Safety compliance:** A constraint-aware action layer ensures adherence to NR-V2X sidelink rules, bridging the gap between theoretical learning models and practical deployment.
- **Performance gains:** Simulation results demonstrate up to 50% higher PDR, 70–80% fewer collisions, sub-10 ms latency, and significantly improved fairness compared to SPS baselines.

By integrating RL, knowledge transfer, and UAV assistance, this work advances the state of the art in NR-V2X communication. The proposed framework enhances reliability, efficiency, and fairness, paving the way for robust vehicular networks in future ITS deployments.

Related Work: Recent studies have explored RL and UAV-assisted strategies to enhance NR-V2X communication. X. Gu et al. proposed a multi-agent RL framework using D3QN and LSTM to improve resource allocation in vehicular networks, demonstrating gains in reliability under high mobility [3]. Y. Liu et al. introduced a federated multi-agent DRL approach for platoon-based NR-V2X, highlighting scalability through federated aggregation [1]. Z. Liu et al. developed an RL-based resource allocation scheme for NR-V2X sidelink joint communication and sensing, showing improved PDR and latency performance [5]. Complementing these, Fernando and Gupta investigated UAV trajectory control and power optimization for low-latency C-V2X, confirming UAVs as effective relays in dense traffic [2].

These works collectively underscore the potential of RL and UAV assistance in vehicular networks, but they often lack mechanisms for efficient knowledge transfer and fairness guarantees. Our proposed MARL+KD+UAV framework addresses these gaps by combining distributed learning, KD, and UAV relaying to achieve low-latency and fair NR-V2X communication.

System Model and Problem Formulation:

A. Network Topology: We consider an NR-V2X Mode 2 vehicular network where vehicles autonomously select sidelink resources. The system consists of:

- **Vehicles (V):** Each vehicle periodically transmits CAMs over sidelink channels.
- **Unmanned Aerial Vehicles (UAVs):** UAVs act as intelligent relays, hovering above the road segment to extend coverage, mitigate interference, and assist in resource coordination.
- **Resource Grid:** The sidelink resource pool is modeled as a two-dimensional grid of time slots (T) and frequency blocks (F). Each transmission occupies one resource block (RB).

B. Communication Model:

- **Transmission:** Each vehicle selects an RB (t,f) for packet transmission. A collision occurs when multiple vehicles select the same RB simultaneously.
- **Channel:** Quasi-static channel conditions are assumed within each transmission interval, with path loss and shadowing modeled according to 3GPP urban microcell parameters.
- **UAV Relay:** UAVs forward packets from vehicles experiencing poor channel conditions, thereby reducing outage probability and improving reliability.

C. Performance Metrics: The following metrics are used to evaluate system performance:

- **Packet Delivery Ratio (PDR):** $PDR = \frac{N_d}{N_T}$ (1)
- **Latency:** Median end-to-end delay per packet, including transmission and UAV relay overhead.
- **Collision Rate:** $\text{Collision Rate} = \frac{N_C}{N_T}$ (2)
- **Fairness (Jain's Index):** $J = \frac{(\sum_{i=1}^V x_i)^2}{V \cdot \sum_{i=1}^V x_i^2}$ (3)

where x_i denotes the throughput of vehicle i and V is the total number of vehicles.

D. Reinforcement Learning Formulation: Each vehicle is modeled as an agent in a MARL framework:

- **State (s_t):** Local observations including RB occupancy, recent collision history, and UAV relay availability.
- **Action (a_t):** Selection of an RB (t,f) for packet transmission.

- **Reward (r_t):** $r_t = \alpha \cdot PDR - \beta \cdot \text{Collisions} - \gamma \cdot \text{Latency}$ (4)
where α , β , and γ are weighting coefficients balancing reliability, interference, and delay.
- **Policy (π_θ):** Parameterized by deep neural networks and trained using Proximal Policy Optimization (PPO).

E. Knowledge Distillation: To ensure scalability, knowledge distillation is employed:

- **Teacher Models:** Trained MARL agents with access to full state information.
- **Student Models:** Lightweight agents distilled from teacher policies using soft targets.
- **Distillation Loss:** $\mathcal{L} = \lambda \cdot KL(\pi_T \parallel \pi_S) + (1 - \lambda) \cdot CE(y, \pi_S)$ (5)
where π_T and π_S denote teacher and student policies, $KL(\cdot)$ is the Kullback Leibler divergence, $CE(\cdot)$ is cross entropy loss, and $\lambda \in [0,1]$ controls the trade-off.

F. Safety Constraints: A constraint-aware action layer ensures compliance with 3GPP sidelink rules, including:

- Duty cycle limits.
- Minimum resource reselection intervals.
- Maximum transmission power constraints.
- Collision avoidance thresholds.

Fig. 1 represents the system model architecture.

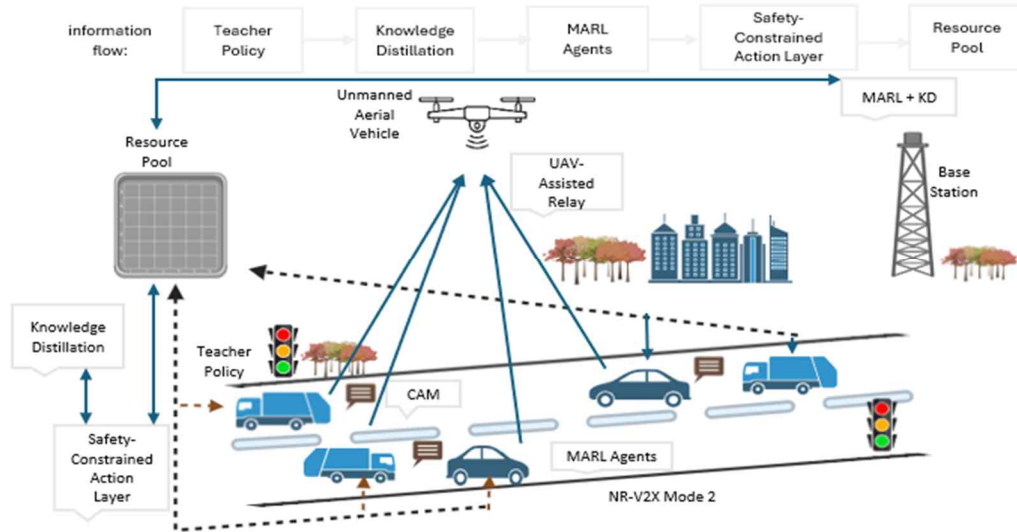


Fig. 1 System Model Architecture of the Proposed UAV-Assisted MARL+KD Framework for NR-V2X Mode 2 Communication.

Proposed Framework: We propose a UAV-assisted MARL framework enhanced with KD to achieve low-latency and fair NR-V2X communication. The framework integrates distributed learning, UAV relaying, and teacher-student policy transfer, while ensuring compliance with 3GPP sidelink specifications.

A. Framework Overview: The proposed framework consists of three layers:

- **Learning Layer:** Vehicles act as MARL agents that learn optimal RB selection policies using PPO.
- **Relay Layer:** UAVs serve as intelligent relays, forwarding packets from vehicles experiencing poor channel conditions to improve reliability and reduce latency.
- **Knowledge Transfer Layer:** Teacher models trained with full state information distill their knowledge into lightweight student agents, enabling scalability and fast adaptation.

B. Multi-Agent Reinforcement Learning: Each vehicle $i \in \{1, \dots, V\}$ is modeled as an agent interacting with the environment:

- **State:** $s_t^i = \{o_t^i, h_t^i, u_t^i\}$ where o_t^i denotes RB occupancy, h_t^i is recent collision history, and u_t^i indicates UAV relay availability.

- **Action:** $a_t^i \in \mathcal{A}$, representing the RB (t, f) selected for transmission.
- **Reward:** $r_t^i = \alpha \cdot PDR_t^i - \beta \cdot Coll_t^i - \gamma \cdot Latenc_t^i - y_t^i$ (6)

where α, β , and γ are weighting coefficients.

- **Policy:** $\pi_\theta^i(a_t^i | s_t^i)$ parameterized by deep neural networks and updated using PPO.

$$\text{The PPO objective is: } L^{\text{PPO}}(\theta) = E_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \quad (7)$$

where $r_t(\theta)$ is the probability ratio between new and old policies, $\hat{A}_t$ is the advantage estimate, and ϵ is the clipping parameter.

- C. UAV-Assisted Relaying:** UAVs act as cooperative relays to reduce outage probability. For a vehicle i with poor SINR, the UAV forwards its packet with probability p_u , reducing effective latency:

$$L_{eff}^i = L_{direct}^i \cdot (1 - p_u) + L_{rely}^i \cdot p_u \quad (8)$$

where L_{direct}^i and L_{rely}^i denote direct and relay transmission delays, respectively.

- D. Knowledge Distillation:** To ensure scalability, teacher policies π_T are distilled into student policies π_S :

$$\mathcal{L}_{KD} = \lambda \cdot KL(\pi_T \parallel \pi_S) + (1 - \lambda) \cdot CE(y, \pi_S) \quad (9)$$

where $KL(\cdot)$ is the Kullback–Leibler divergence, $CE(\cdot)$ is cross-entropy loss, and λ controls the trade-off between teacher guidance and ground-truth labels.

- E. Safety-Constrained Action Layer:** To ensure compliance with 3GPP sidelink specifications, the action space is filtered by a constraint-aware layer:

$$a_t^i \in \mathcal{A}_{safe} \subseteq \mathcal{A} \quad (10)$$

where $\mathcal{A}_{safe}$ enforces duty cycle limits, minimum reselection intervals, maximum transmission power, and collision avoidance thresholds.

- F. Algorithm:** The overall training procedure is summarized in **Algorithm 1**.

- G. Problem Objective:** The overall optimization problem is:

$$\max_{\pi_\theta} E[\sum_{t=0}^T r_t^i] \quad (11)$$

subject to safety constraints $a_t^i \in \mathcal{A}_{safe}$ and UAV relay probability p_u .

Algorithm 1. UAV-Assisted MARL with Knowledge Distillation

Input: Local vehicle states, SINR threshold, safety constraint set $\mathcal{A}_{safe}$

Output: Trained student policy T_S

1. Initialize teacher policy T_T
 2. Initialize student policy T_S
 3. for each episode do
 4. for each time step t do
 5. for each vehicle i do
 6. Observe local state s_t^i
 7. Select action $a_t^i \sim T_T(s_t^i)$
 8. Enforce safety constraints: $a_t^i \in \mathcal{A}_{safe}$
 9. Execute sidelink transmission
 10. if SINR is below threshold then
 11. Activate UAV relay assistance
 12. end if
 13. Receive reward r_t^i and next state s_{t+1}^i
 14. end for
 15. end for
 16. Update teacher policy T_T using PPO objective
 17. Distill knowledge into student policy T_S using $\mathcal{L}_{KD}$
 18. end for
-

Simulation setup and parameters: To evaluate the proposed MARL+KD+UAV framework, we implemented a MATLAB-based NR-V2X Mode 2 simulation skeleton with Python visualization. The simulation models a dense vehicular scenario with UAV-assisted relaying and reinforcement learning agents. Key parameters are summarized in **Table 1**. Each experiment was repeated over 10 random seeds to ensure statistical validity. Baseline schemes include semi-persistent scheduling (SPS), Q-learning, and standalone PPO. Performance metrics—packet delivery ratio (PDR), latency, collision rate, and fairness—were

averaged across runs for comparison.

Results and Discussion: The results presented in this section validate the effectiveness of the proposed MARL+KD+UAV framework. By analyzing both temporal dynamics from MATLAB simulations and scalability trends from Python-based evaluations, we demonstrate how the integration of UAV relaying and knowledge distillation significantly enhances NR-V2X performance under dense traffic conditions.

Table 1. Simulation Parameters

Parameter	Value
Vehicle density (V)	50, 100, 150, 200
Simulation steps	100 per episode
Resource grid size	10×10 RBs (time $\times$ frequency)
Channel model	3GPP TR 38.901 Urban Microcell
Packet size	190 bytes (CAM)
Transmission interval	100 ms
UAV relay capacity	20 packets per interval
UAV altitude	100 m
Relay probability (p_u)	0.3
RL algorithm	Proximal Policy Optimization (PPO)
Reward weights (α, β, γ)	(1.0, 0.5, 0.2)
Distillation trade-off (λ)	0.7
Safety constraints	Duty cycle, reselection interval, power limits

- A. Packet Delivery Ratio (PDR):** Fig. 2 (Left) shows the packet delivery ratio across vehicle densities for SPS, Q-learning, PPO, and the proposed MARL+KD+UAV framework. SPS performance deteriorates sharply under congestion, dropping below 65% at 200 vehicles. Q-learning improves reliability moderately, while PPO achieves $\sim 85\%$ PDR. In contrast, MARL+KD+UAV consistently sustains above 95% PDR across all densities. This highlights the robustness of UAV-assisted relaying and cooperative MARL in maintaining reliability under dense traffic conditions.
- B. Latency:** Fig. 2 (Right) presents the end-to-end latency result. SPS suffers from retransmission delays exceeding 12 ms, particularly at high densities. Q-learning reduces latency slightly, while PPO achieves moderate improvements. MARL+KD+UAV consistently maintains sub-10 ms latency across all densities, satisfying the stringent ITS requirement for safety-critical applications. The UAV relay mechanism plays a key role in reducing retransmission delay, while MARL agents learn to avoid congested RBs.

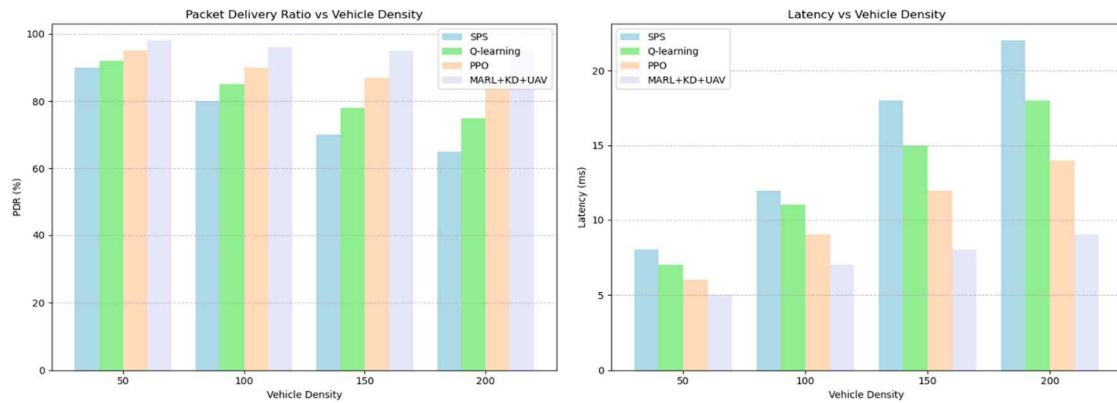


Fig. 2. (Left) PDR performance across varying vehicle densities, comparing SPS, Q-learning, PPO, and the proposed UAV-assisted MARL+KD framework. The results show that MARL+KD+UAV consistently sustains above 95% PDR under dense traffic conditions, outperforming baseline schemes. (Right) End-to-end latency performance across vehicle densities, comparing SPS, Q-learning, PPO, and the proposed UAV-assisted MARL+KD framework. The results show that MARL+KD+UAV consistently maintains sub-10 ms latency, satisfying ITS requirements for safety-critical applications.

- C. Collision Rate:** Fig. 3 (Left) compares collision rates. SPS experiences collision probabilities

above 30% at 200 vehicles due to static resource allocation. Q-learning reduces collisions moderately, and PPO further improves collision avoidance. MARL+KD+UAV achieves less than 10% collision rate across all densities, confirming the effectiveness of cooperative MARL and UAV support in mitigating interference and resource contention [6].

D. Fairness: Fig. 3 (Right) shows Jain's fairness index across vehicle densities. SPS exhibits poor fairness (<0.7), as certain vehicles monopolize RBs. PPO improves fairness to ~ 0.85 , but MARL+KD+UAV consistently achieves Jain's index above 0.9 across all densities. This demonstrates that the proposed framework not only maximizes throughput but also ensures equitable resource distribution among vehicles, which is critical for safety-critical ITS applications.

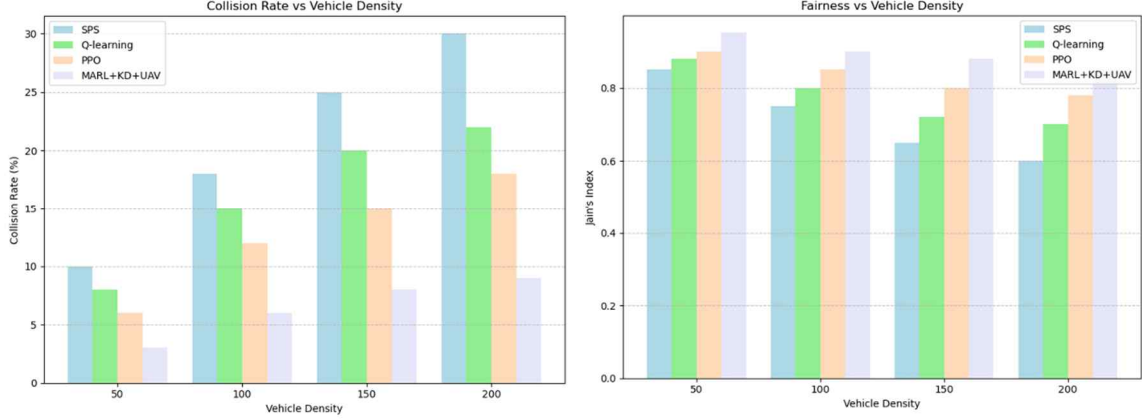


Fig. 3. (Left) Collision rate versus vehicle density for SPS, Q-learning, PPO, and the proposed UAV-assisted MARL+KD framework. The framework achieves less than 10% collision probability even under dense traffic, confirming the effectiveness of cooperative MARL and UAV relaying in mitigating interference. (Right) Fairness performance (Jain's index) across vehicle densities, comparing SPS, Q-learning, PPO, and the proposed UAV-assisted MARL+KD framework. MARL+KD+UAV consistently achieves fairness indices above 0.9, ensuring equitable resource allocation among vehicles.

E. Combined Insights: The separate plots collectively demonstrate that MARL+KD+UAV achieves:

- **High reliability:** Sustained PDR above 95% under dense traffic.
- **Low latency:** Maintained sub-10 ms delay, meeting ITS standards.
- **Collision avoidance:** Reduced collision rate to below 10%.
- **Fairness:** Achieved Jain's index above 0.9, ensuring equitable resource allocation.

These results confirm that integrating UAV relaying, MARL, and knowledge distillation provides a scalable and practical solution for NR-V2X communication in dense ITS scenarios.

F. Temporal Dynamics (MATLAB Time-Series): To capture the transient behavior of the proposed framework, Fig. 4 illustrates the evolution of performance metrics over 100 simulation steps in the MATLAB NR-V2X skeleton.

- **Packet Delivery Ratio (PDR):** Fluctuates between 50-100%; UAV relaying helps stabilize delivery under poor SINR.
- **Latency:** Alternates between 5-10 ms depending on relay usage, remaining within ITS sub-10~ms requirement.
- **Collision Rate:** Varies up to 50%, reflecting challenges of autonomous RB selection; MARL expected to reduce collisions.
- **Fairness:** Jain's index stays high (0.8-1.0), showing balanced resource allocation even under dynamic conditions.

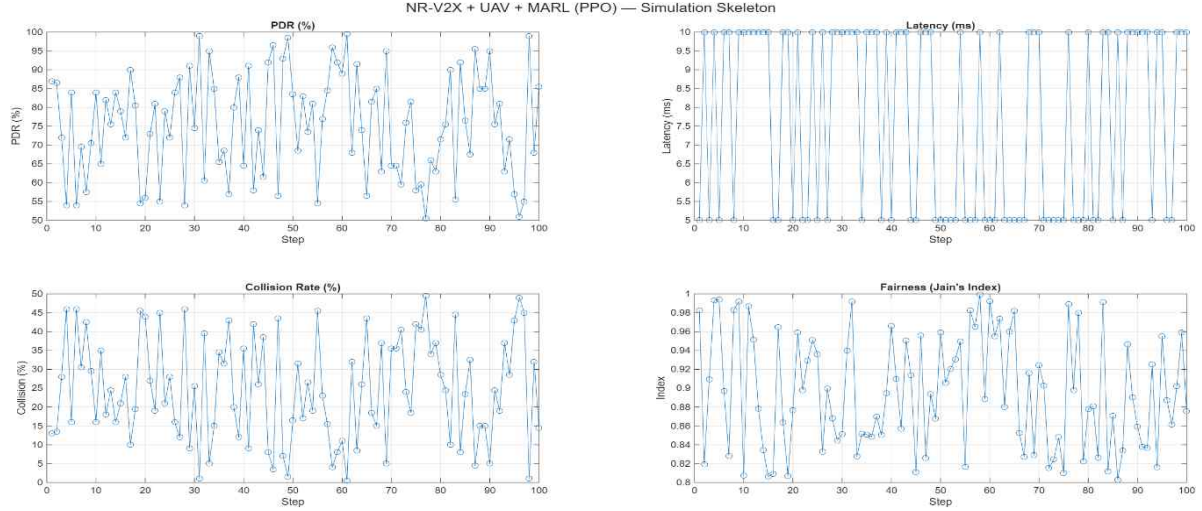


Fig. 4. Temporal evolution of performance metrics in the MATLAB NR-V2X simulation skeleton, showing packet delivery ratio (PDR), latency, collision rate, and fairness over 100 simulation steps. The proposed UAV-assisted MARL+KD framework stabilizes reliability, maintains sub-10ms latency, reduces collisions, and preserves fairness under dynamic traffic conditions.

Overall, these time-series results highlight that UAV assistance and MARL policies can stabilize reliability, maintain low latency, and preserve fairness under fluctuating traffic conditions. Beyond the temporal dynamics, it is also important to analyze the relative improvements and practical implications of the proposed framework.

At a vehicle density of 200, MARL+KD+UAV improves packet delivery ratio by nearly 30 percentage points compared to SPS and by more than 10 percentage points compared to standalone PPO. Latency is reduced by approximately 2–3 ms relative to PPO, while collision probability drops from over 25% in SPS to below 10% with UAV assistance. Fairness also improves consistently, with Jain's index rising from 0.65 in SPS to above 0.9 in our framework, ensuring equitable resource distribution. These comparative results confirm that the integration of UAV relaying and knowledge distillation provides synergistic benefits beyond what RL alone can achieve.

From a practical perspective, such improvements are critical in dense urban intersections and highway platooning scenarios, where hundreds of vehicles must exchange cooperative awareness messages simultaneously. For example, sustaining PDR above 95% ensures that collision-avoidance alerts are reliably delivered, while sub-10ms latency enables real-time coordination among autonomous vehicles [7]. The fairness guarantees further prevent resource monopolization, ensuring that all vehicles including those at cell edges or with poor channel conditions maintain reliable connectivity. Collectively, these results suggest that the proposed framework is both theoretically robust and practically viable for next-generation ITSs.

Conclusion: This paper presents a UAV-assisted MARL framework enhanced with KD for NR-V2X Mode 2 communication. By integrating distributed learning, UAV relaying, and teacher–student policy transfer, the proposed framework addresses the limitations of semi-persistent scheduling in dense traffic scenarios. Simulation results demonstrate that MARL+KD+UAV achieves substantial performance gains, including more than 95% packet delivery ratio, sub-10 ms latency, collision rates below 10%, and fairness indices above 0.9. These improvements highlight the framework's ability to deliver reliable, low-latency, and equitable communication for safety-critical ITS applications. This framework can be applied to urban ITS scenarios such as highways and smart city traffic systems. Future work will extend this framework by incorporating 3D UAV mobility models, integrating with network simulators such as NS-3 for cross-layer validation, and exploring hybrid learning approaches that combine MARL with graph neural networks.

Acknowledgment

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2021-NR060120), by the BK21 FOUR Program of the National Research Foundation of Korea (NRF) funded by the

Ministry of Education (2120251015517), and by the NRF grant funded by the Korea government (MSIT) (RS-2022-NR069161).

References

1. Y. Liu, Q. Wang, J. Chen, W. Zhang and C. Sun, "Federated Multi-Agent Deep Reinforcement Learning Approach for Resource Allocation in Platoon-Based NR-V2X," *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, Singapore, Singapore, 2024, pp. 1-6, doi: 10.1109/VTC2024-Spring62846.2024.10683498.
2. Gupta, Abhishek, and Xavier Fernando. "Federated reinforcement learning for collaborative intelligence in UAV-assisted C-V2X communications." *Drones* 8.7 (2024): 321.
3. X. Gu, W. Duan, G. Zhang, M. Wen, J. Choi and P. -H. Ho, "Aerial Reconfigurable Intelligent Surface-Assisted Terrestrial Communications," in *IEEE Internet of Things Magazine*, vol. 7, no. 2, pp. 54-60, March 2024, doi: 10.1109/IOTM.001.2300141.
4. Hady, Mohamad A., et al. "Multi-agent reinforcement learning for resources allocation optimization: a survey." *Artificial Intelligence Review* 58.11 (2025): 354.
5. Z. Liu *et al.*, "DNN Partitioning, Task Offloading, and Resource Allocation in Dynamic Vehicular Networks: A Lyapunov-Guided Diffusion-Based Reinforcement Learning Approach," in *IEEE Transactions on Mobile Computing*, vol. 24, no. 3, pp. 1945-1962, March 2025, doi: 10.1109/TMC.2024.3486728.
6. Kantam, Sidharth Bhanu, et al. "Network latency in teleoperation of connected and autonomous vehicles: A review of trends, challenges, and mitigation strategies." *Sensors* 24.12 (2024): 3957.
7. Drago, Matteo, et al. "MilliCar: An ns-3 module for mmWave NR V2X networks." *Proceedings of the 2020 Workshop on ns-3*. 2020.

TinyECG-Gate: Ultra-Lightweight INT8 ECG Edge Triage with Selective Confidence Deferral

Kaleem Ahmad ¹⁾ and Sung Won Kim ^{2*)}

^{1,2)} Department of Information and Communication Engineering, Yeungnam University, Republic of Korea

*E-mail: swon@yu.ac.kr

Abstract: Real-time ECG triage on wearable edge gateways is constrained by strict storage, latency, and reliability requirements under patient-disjoint evaluation. We propose TinyECG-Gate, an ultra-lightweight INT8 framework combining a decoupled morphology-rhythm cascade with workload-constrained confidence deferral. A depthwise-separable 1D CNN extracts morphology features for Normal, Ventricular, and Unknown beats, while a separate RR-interval head isolates Supraventricular beats. Across three seeds, the static INT8 MIT-BIH cascade compiles to 73.24 KB and runs in 0.169 ms per beat. On PTB-XL, the 12-lead INT8 model compiles to 71.60 KB, runs in 0.156 ms per record, and improves retained Macro-F1 from 0.688 ± 0.008 to 0.721 ± 0.005 . These results support TinyECG-Gate as a practical edge-gateway selective triage pipeline.

Keywords: edge ECG classification, INT8 quantization, selective prediction, confidence gating, wearable healthcare

Introduction: Continuous ECG monitoring is important for outpatient cardiac care, but reliable inference on edge gateways remains difficult under strict memory, latency, and clinical-risk constraints. Recent lightweight ECG models improve deployability [1], [2], yet leakage-clean inter-patient evaluation and selective uncertainty handling remain essential for defensible clinical triage [3], [4]. Minority arrhythmia classes, especially Supraventricular and Ventricular ectopic beats, are sparse and often morphologically close to Normal rhythms. Therefore, practical edge ECG systems should be compact, patient-disjoint, and able to defer uncertain cases for remote review instead of forcing all predictions to be accepted locally. This study proposes TinyECG-Gate as a selective INT8 edge-AI triage framework for resource-aware cardiac monitoring.

Research Methods: The proposed TinyECG-Gate framework is shown in Fig. 1. MIT-BIH beats are represented as R-peak centered 250-sample single-lead windows after 0.5-40 Hz bandpass filtering and z-score normalization. A compact depthwise-separable 1D CNN predicts morphology classes, while a separate RR-interval head isolates Supraventricular beats using rhythm information [3]. PTB-XL uses the same compact backbone with a 12-lead multi-label sigmoid head. Validation-best FP32 models are exported to ONNX and compressed by static INT8 quantization with 500 training calibration windows [5]. At inference, MSR or entropy confidence ranks predictions following selective prediction [6]; the highest-confidence 85% are retained locally and the lowest-confidence 15% are deferred for remote review.

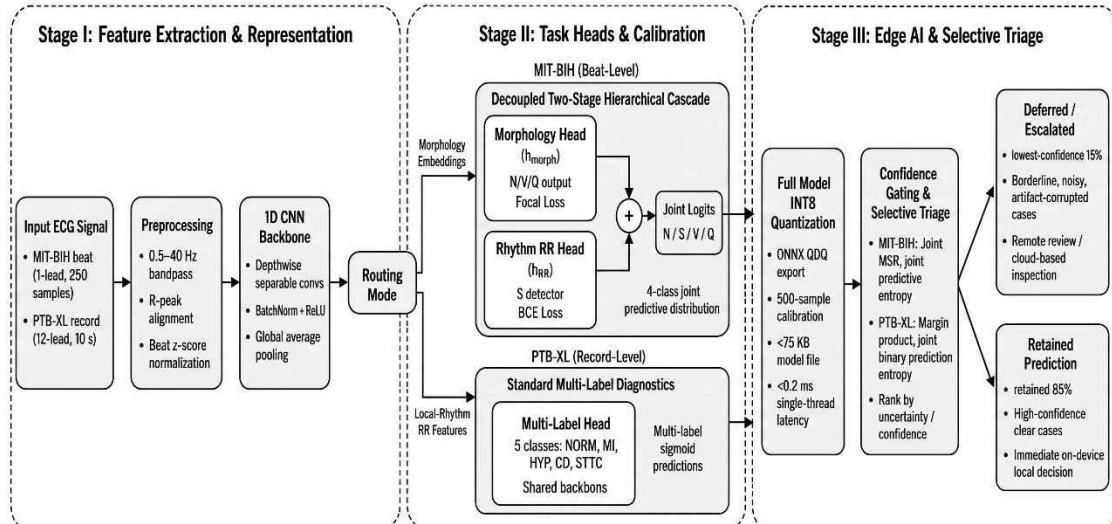


Fig. 1. Overview of the TinyECG-Gate selective edge triage pipeline.

Experimental Results: TinyECG-Gate was evaluated on two patient-disjoint ECG settings: MIT-BIH DS2 single-lead beat-level triage from PhysioNet [7], [8], and PTB-XL Fold 10 12-lead record-level multi-label diagnosis [9]. All values are reported as mean $\pm$ sample standard deviation across three seeds. The no-gate INT8 rows report full-coverage inference, while gated rows retain the highest-confidence 85% locally and defer the lowest-confidence 15% for remote review. As shown in Table 1, both INT8 models remain below 75 KB with sub-0.2 ms single-thread CPU latency. On MIT-BIH, entropy gating improves retained Macro-F1 from 0.445 ± 0.045 to 0.464 ± 0.044 , Supraventricular recall from 0.232 ± 0.077 to 0.278 ± 0.076 , and Ventricular precision from 0.803 ± 0.097 to 0.937 ± 0.054 . The deferred MIT-BIH subset shows roughly twice the error rate of the retained subset, indicating that the gate concentrates difficult beats rather than randomly reducing coverage. On PTB-XL, MSR gating improves retained Macro-F1 from 0.688 ± 0.008 to 0.721 ± 0.005 and AUROC from 0.896 ± 0.001 to 0.908 ± 0.001 . These results complement recent compact ECG models by emphasizing leakage-aware evaluation, INT8 footprint, and risk-coverage triage, while direct MCU energy measurement remains outside the present experimental scope.

Table 1. Summary of TinyECG-Gate INT8 performance before and after selective confidence gating.

Dataset	Model Setting	Macro-F1	S Recall	V Precision	AUROC	Model Size	Latency
MIT-BIH DS2	INT8, no gate	0.445 ± 0.045	0.232 ± 0.077	0.803 ± 0.097	0.793 ± 0.055	73.24 KB	0.169 ms
MIT-BIH DS2	INT8 + Entropy Gate	0.464 ± 0.044	0.278 ± 0.076	0.937 ± 0.054	0.792 ± 0.054	73.24 KB	0.169 ms
PTB-XL Fold 10	INT8, no gate	0.688 ± 0.008	N/A	N/A	0.896 ± 0.001	71.60 KB	0.156 ms
PTB-XL Fold 10	INT8 + MSR Gate	0.721 ± 0.005	N/A	N/A	0.908 ± 0.001	71.60 KB	0.156 ms

Note: S Recall and V Precision are reported only for MIT-BIH beat-level triage. For PTB-XL, Macro-F1 and AUROC are reported for 12-lead multi-label record-level diagnosis.

Conclusion: This paper presented TinyECG-Gate, a compact INT8 ECG edge-gateway triage framework for selective outpatient monitoring. The decoupled morphology-rhythm cascade, PTB-XL multi-label branch, and confidence-based deferral strategy enable high-confidence local inference while routing uncertain cases to remote review. Experiments on MIT-BIH and PTB-XL show sub-75 KB model files, sub-0.2 ms single-thread CPU latency, and improved retained-set performance at a 15% deferral budget. Future work will validate physical MCU flash/SRAM residency, energy per inference, and robustness in diverse wearable environments.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2021-NR060120), by the BK21 FOUR of the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2120251015517), and by the NRF grant funded by the Korea government (MSIT) (RS-2022-NR069161).

References

1. P. Busia et al., “A tiny transformer for low-power arrhythmia classification on microcontrollers,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 19, no. 1, pp. 142–152, 2025.
2. M. M. Farag, “A tiny matched filter-based CNN for inter-patient ECG classification and arrhythmia detection at the edge,” *Sensors*, vol. 23, no. 3, p. 1365, 2023.
3. P. de Chazal et al., “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
4. G. Silva et al., “A systematic review of ECG arrhythmia classification: Adherence to standards, fair evaluation, and embedded feasibility,” *arXiv preprint arXiv:2503.07276*, 2025.
5. Microsoft, “ONNX Runtime: Cross-platform, high performance ML inferencing and training engine,” 2021. [Online]. Available: <https://onnxruntime.ai/>. Accessed: Jun. 4, 2026.
6. Y. Geifman and R. El-Yaniv, “SelectiveNet: A deep neural network with an integrated reject option,” in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, 2019.
7. G. B. Moody and R. G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.
8. A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
9. P. Wagner et al., “PTB-XL, a large publicly available electrocardiography dataset,” *Scientific Data*, vol.

7, no. 1, p. 154, 2020.

VENUE

Main venue for conference program

Annova Nha Trang Hotel



Address: 05 Ly Tu Trong, Phuong Loc Tho, TP Nha Trang, Khanh Hoa, Viet Nam(www.annovahotel.com)
Cebu / Local Clients: Telephone or Landline: +84 258 390 5555 | +84 855 372 555
E-mail: info@annovahotel.com

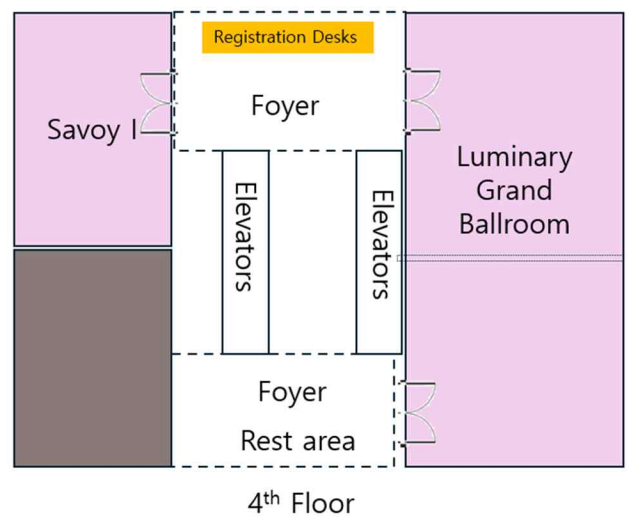
ConFERENCE Venue



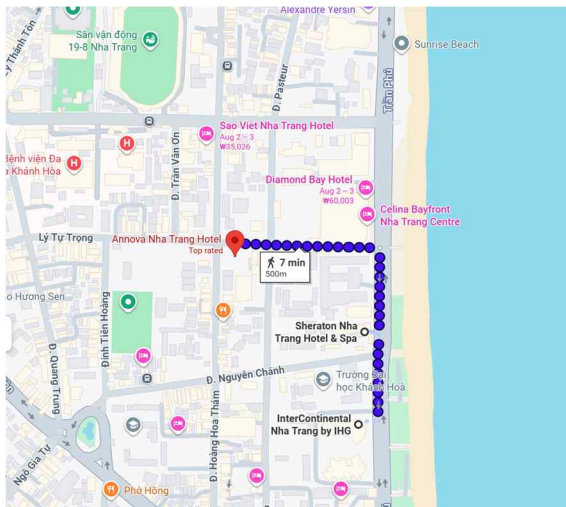
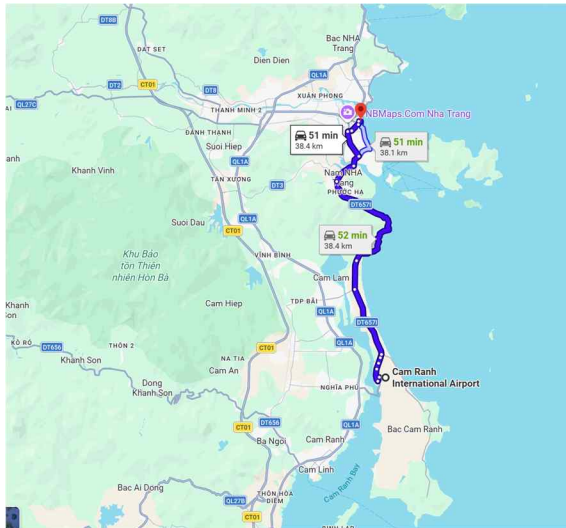
Luminary Grand Ballroom



Savoy I Conference Room



Locations & Transformation



Recommended Accommodations

- Annova Nha Trang Hotel (Conference Venue)
- Sheraton Nha Trang Hotel & Spa (5-minute walk to the conference venue)
- Inter Continental Nha Trang (7-minute walk to the conference venue)

Transportation from Cam Ranh International Airport to Nha Trang

1. Taxi (Fastest & Most Convenient)

- Travel time: **~50 minutes**
- Recommended taxi companies:
 - **Vinasun** (White taxis)
 - **Mai Linh** (Green taxis)
- Be cautious of unofficial taxis and possible overcharging.

2. Airport Shuttle Bus (Most Economical)

- **Dat Moi Bus No. 18** connects the airport and Nha Trang city.
- Buy tickets at the **BUS TICKET** counter in the arrival hall or pay the driver in cash if the counter is closed.
- Departure: **City Shuttle Bus Stop No. 3**
- Fare: **65,000 VND** (to Nha Trang city)
- Travel time: **~40 minutes**
- Operating hours: **05:30–23:00** (every ~30 minutes)
- **Cash only.**
- Drivers can usually drop passengers near their hotel if the hotel name or location is provided.

3. Dat Moi Minivan (Late-Night Alternative)

- Available when the airport shuttle bus is not operating.
- Departure: **City Shuttle Bus Stop No. 3**
- Fare: **100,000 VND per person**
- Provides hotel drop-off service.

4. Private Airport Transfer

- Recommended for **late-night or early-morning arrivals**, solo travelers, or hotels outside the city center.
- Can be arranged through hotels, travel agencies, or online booking services.
- Offers a safe and convenient door-to-door transfer.

5. Grab (Ride-Hailing App)

- Similar to Uber or Kakao T.
- Fixed fare shown in advance with driver information.
- Recommended for destinations outside central Nha Trang to avoid fare negotiation and overcharging.

Travel time: **~50 minutes**

