List of Provisionally Accepted Papers

IETE's Second International Conference on Smart Mobility Systems (ICSMS 2026)
Organized by
The Institution of Electronics and Telecommunication Engineers (IETE) and
Osmania University, Hyderabad
at Osmania University, Hyderabad
in Hybrid Mode

Sl No	Paper ID	Paper Title
1	4	AI-Driven Intelligent Transportation Systems for Predictive Road Safety and Accident Prevention
2	6	CitiFine: AI-Enabled Citizen-Assisted Traffic Violation Detection System
3	25	A Portable IoT- Based Ambulatory Health Monitoring System with Real-Time Vital Signs Tracking for Remote Patient Care
4	26	A Deep Learning Framework for Sign Language Recognition using Multi-Modal Facial Expression and Hand Gesture Feature Fusion
5	30	DESIGN AND ANALYSIS OF METAMATERIAL-LOADED COMPACT ANTENNA FOR BANDWIDTH ENHANCEMENT IN WIDEBAND WIRELESS COMMUNICATION
6	42	A 3D Non-Stationary Spatiotemporal Channel Model for Next-Generation 6G Wireless Communication Systems
7	45	Design of a Frequency-Reconfigurable Microstrip Patch Antenna for Multi-Band 5G Wireless Communication Applications
8	48	Reinforcement Learning–Driven Pilot Scheduling to Reduce Pilot Contamination in Cell-Free Massive MIMO for Sixth-Generation Wireless Networks
9	52	CyberShield AgentMesh: Mobility-Aware Signed Response Bundles and Safe-Flow SDN Containment for Cyber-Resilient Intelligent Transportation Systems
10	57	Cognitive Digital Twin for Human-Centric Automated Driving: Modeling Driver Attention, Override Behavior, and Trust Dynamics at the IoT Edge
11	62	Digital Twin-Driven Autonomous AI Agents for Predictive Fault Propagation in Vehicle-to-Everything (V2X) IoT Meshes
12	73	Design and Performance Analysis of a Squirrel- Cage Induction Motor for Electric Two-Wheeler Applications using AltairFlux motor
13	78	Vedic Mathematics Driven Reversible ALU for Low-Power Applications
14	79	Smart ECG Monitoring Interface with Controller Support for Real Time Cardiac Monitoring
15	89	Automated Road Damage Detection Using UAV Imagery and Deep Learning Techniques
16	98	Optimizing Retrieval-Augmented Generation with Parameter-Tuned Local LLMs for Real-Time Disaster Response and Early Warning System
17	100	Optimizing for the Driver: A Predictive Positioning System to Enhance Welfare in Gig Networks
18	103	SPACE SUSTAINABILITY: AVERTING COLLISIONS BETWEEN SATELLITES AND SPACE DEBRIS

19	104	Design and Verification of AMBA APB Protocol with CRC-8 Error Detection and Fault Injection for SoC Communication
20	105	A Lightweight Learning-Assisted Congestion-Aware Routing Scheme for 8×8 Mesh Network-on-Chip
21	107	A GDPR-Compliant Hierarchical Memory Architecture for Autonomous AI Agents
22	108	A Transformer-Based Framework for Robust Underwater Image Enhancement Using Restormer
23	111	An Integrated Machine Learning Framework for Invoice Payment Delay Prediction Using Temporal Decay Features and Survival Analysis
24	164	Explainable AI for Adaptive Intrusion Detection in Intelligent Transportation Systems: A Closed-Loop System Architecture
25	169	Hybrid AI Model for Predictive Cyber Threat Detection using GNN and Transformers
26	171	LandmarkFusion-ST: Multi-Stream Spatial-Temporal Landmark Network with Semantic Refinement to Sign language recognitions and translations in real-time
27	173	Efficient Driver Drowsiness Detection Using a Lightweight Vision Transformer with Temporal Dynamics and Meta-Learning
28	175	A Human-Centric Adaptive Edge Intelligence Framework for Multimodal Traffic Learning and Decision Optimization
29	187	CogniFlow: An Explainable, Real-Time, Hardware-Free Multimodal Framework for Cognitive Focus Monitoring
30	190	Real-Time Driver Drowsiness Detection Using YOLOv8 and CNN-Based Eye State Analysis
31	198	Adaptive DoDAG Formation for Smart Mobility IoT Networks Using RPL
32	201	Dynamic Control of Grid-Connected EV Charging Systems with V2G and G2V Support
33	214	GPS/IMU Sensor Fusion Using the Error-State Extended Kalman Filter for Waypoint-Based Autonomous Mobile Robot Navigation
34	222	CogniLens: An AI-Powered multi-modal Smart Glass System for Memory Assistance and Adaptive Navigation in Elderly and Visually Impaired Users
35	245	Indian Sign Language Recognition Using an Ensemble of Deep Learning Models with Optimized Weighted Voting
36	248	Jal Netra: Ferry Optimization System
37	264	MLFDN: A Trusted Waterfall Machine Learning Based Secure Data Transmission Framework for Highly Dynamic Networks
38	265	Semantic-Cognitive 6G Networks for Meaning-Centric Terahertz Communication
39	268	Multi-Spectral Cloud Nowcasting Using Conditional Diffusion Models on INSAT-3D Satellite Data

LandmarkFusion-ST: Multi-Stream Spatial-Temporal Landmark Network with Semantic Refinement to Sign language recognitions and translations in real-time

Animesh Palui[1], Sahil Ghosh[1], Aditya Kumar[1] and Koushik Dutta[2]

[1] Department of Information Technology, Institute of Engineering and Management, University of Engineering and Management, Kolkata, India
[2] Department of Computer Science and Engineering, Institute of Engineering and Management, University of Engineering and Management, Kolkata, India
paluianimesh31@gmail.com , sahilghosh0512@gmail.com , adii.utsav@gmail.com , koushik.dutta@iem.edu.in

Abstract: The main communication medium of more than 70 million global deaf and hard-of-hearing people is the sign language, but the autocratized recognitions are unable to survive the continuous signing, variability in the signers, and require that communication be real time. For overcoming these shortcomings, LandmarkFusion-ST is an innovative multi-stream fusion framework that simultaneously predicts hand, facial, and upper-body trajectories of landmarks. The system is based on a combination of holistic pose estimation, the use of Graph Attention Networks to encode spatial dependence and the use of Temporal Convolutional Networks with multi-head self-attention to encode temporal dynamics. A semantic refinement system is based on a pre-trained language encoder that provides grammatically consistent output of translation. Presented with three benchmarks (ISL isolated gestures, PHOENIX-2014 continuous signing and a corpus of ISL sentences) this framework demonstrates 95.6% isolated sign accuracy, a word error rate of 19.3% (14.6% relative to baselines) and is maintainable at over 30 FPS on then-consumer hardware. LandmarkFusion-ST also has a bidirectional, multilingual translation tool and serves as a feasible stepping stone to an all-encompassing real-time assistive communication.

Keywords: Sign Language Recognition, Graph Attention Network, Temporal Convolution Network.

1 Introduction

Sign languages are completely developed natural languages with their own grammar, syntax, and vocabulary, and are spoken by about 70 million deaf people on the planet as the main method of communication [1]. The Indian Sign Language (ISL), used by an estimated 1.8 to 7 million deaf individuals in the Indian subcontinent [2], is grossly underrepresented in computational research as compared to ASL or BSL - forming a pronounced technological void in terms of accessibility tools to users of the ISL. The use of human interpreters cannot be scaled nor reached universally to facilitate deaf-hearing communication. A promising alternative is the Automated Sign Language Recognition (SLR) and Translation (SLT) systems that have significant obstacles in the variability of signers in hand shape, speed, and placement; context interpretation of multi-temporal horizons; and multi-modal combination of hand signs, facial expression, and upper-body posture. The real-time processing is also required to be low-latency and practical deployment. Although deep learning has already reached a level of over 95 percent accuracy in single sign benchmarks [3][4], there is a gap to bridge between single sign benchmarks and continuous recognition and grammatically coherent translation. This paper suggests logical combination of spatial landmark modeling, temporal sequence knowledge and linguistic post-processing in a unified architecture.

1.1 Research Gap

A critical analysis of the current literature indicates that there are a number of outstanding gaps. First, most recognition systems are only used on isolated static gestures, an omission to deal with continuing sentence level signing that is practical communication [5][6]. Second, the majority of architectures are based on pixel-level raw features processed using conventional convolutional networks which is prone to variations in lighting, the complexity of skin tones and the backgrounds [7][8]. Third, the geographical connections among hand joints, on which discriminative information