2025 IEEE International Conference on Big Data and **Smart Computing** (**BigComp 2025**)

Kota Kinabalu, Malaysia 9-12 February 2025



IEEE Catalog Number: CFP2540X-POD ISBN:

979-8-3315-2903-1

Copyright © 2025 by the Institute of Electrical and Electronics Engineers, Inc. All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

 IEEE Catalog Number:
 CFP2540X-POD

 ISBN (Print-On-Demand):
 979-8-3315-2903-1

 ISBN (Online):
 979-8-3315-2902-4

ISSN: 2375-933X

Additional Copies of This Publication Are Available From:

Curran Associates, Inc 57 Morehouse Lane Red Hook, NY 12571 USA Phone: (845) 758-0400

Fax: (845) 758-2633

E-mail: curran@proceedings.com Web: www.proceedings.com



2025 IEEE International Conference on Big Data and Smart Computing (BigComp) BigComp 2025

Table of Contents

Message from the Chairs xvii Organizing Committee xvii Program Committee xix Steering Committee xxii Reviewers xxiii
Conference Program
Session 1: Big Data and AI Applications I
Examining the Performance of Mamba Tab for Software Defect Prediction
Optimizing PROTAC Design: A Novel Approach using Exposure Scores for Linker Site Selection 9 Yeonju Jeong (Chungnam National University, South Korea), Jae-Mun Choi (Calici Co., Ltd., South Korea), and Young-Kuk Kim (Chungnam National University, South Korea)
Refining Long-Term Predictions: Two-Stage Spatial-Temporal Feature Learning for 3D Human Motion Prediction
Session 2A: Generative AI
CubeDimAE: Automatic Autoencoder Generation Based on Dimension Estimation by Tessellation . 20 Byungrok Kim (KENTECH, South Korea), Myeong-Ha Hwang (KEPRI, South Korea), Jeonghyun Joo (KEPRI, South Korea), YooJin Kwon (KEPRI, South Korea), and Hyunwoo Lee (KENTECH, South Korea)
Diffusion-Based Portrait Image Generation Scheme using Plug-in Adapters

ntropy-Based Sentence-Level Hallucination Score in Large Language Models Eojin Joo (KAIST, South Korea), Young-Jun Lee (KAIST, South Korea), and Ho-Jin Choi (KAIST, South Korea)	.77
IE Text Encoder : Padding is Enough to Generate Image in Diffusion Based Text-to-Image Model	. 79
JongHyun Han (Seoul National Univ. of Sci. & Tech., Korea), SooHyun Lee (Seoul National Univ. of Sci. & Tech., Korea), and JongYoul Park (Seoul National Univ. of Sci. & Tech., Korea)	
Multimodal Learning for Just-In-Time Software Defect Prediction in Autonomous Driving ystems	88
Faisal Mohammad (Jeonbuk National University, South Korea) and Duksan Ryu (Jeonbuk National University, South Korea)	
DefectGRANDE: Hybrid Approach for Class Imbalance in Software Defect Prediction Eunjeong Ju (Jeonbuk National University, South Korea), Jeonghaw Lee (Jeonbuk National University, South Korea), and Duksan Ryu (Jeonbuk National University, South Korea)	. 90
llu as the Driver of CNV Formation: A Computational Approach to Reveal NAHR	92
ffective Detection Technique of Copyright Infringement Illegal Sharing Sites Gyu Hyeon Choi (Korea Advanced Institute of Science and Technology, Korea), Kyeong Seok Lee (Korea Advanced Institute of Science and Technology, Korea), and Ho Mook Cho (Korea Advanced Institute of Science and Technology, Korea)	.94
mooth Categorical Cross-Entropy: An Improved Approach to Overcome the Limitations of CCE Minseo Kwon (Chungnam National University, Republic of Korea), Sungyoon Jeong (Chungnam National University, Republic of Korea), Jeongmin Choi (Chungnam National University, Republic of Korea), Gwanghee Lee (Chungnam National University, Republic of Korea), and Kyoungson Jhang (Chungnam National University, Republic of Korea)	.96
nhancing the Software Defect Prediction using AutoInt	103
urvival Sequences: Win Prediction from a Strategy Sequence Approach	105
xploring Performance and Consistency in Mathematical Reasoning Across Problem Types	107
redicting Protein-Ligand Binding Residue with Token Classification Keumseok Kang (Pusan National University, Republic of Korea), Sanghun Sel (Pusan National University, Republic of Korea), Jeongmu Kim (Pusan National University, Republic of Korea), and Giltae Song (Pusan National University, Republic of Korea)	111

Enhancing Software Defect Prediction:Integrating SAINT with Ant Colony Optimization and	113
Explainable AI	113
Examining the Impact of Expression Fragmentation and Number Redundancy on Financial QA Task	115
Yoonjae Lee (Seoul National University, South Korea), Kyung Seo Ki (Seoul National University, South Korea), Gahgene Gweon (Seoul National University, South Korea), and Bugeun Kim (Chung-Ang University, South Korea)	113
TNIM: A Unified Text-to-Noise Influence Mapping Approach for Diffusion Models	117
Enhancing Gender Prediction Model Performance through Automatic Individual Entity Extraction and Class Balance	120
Chaeyun Kim (Ewha Womans University, South Korea), Eunseo Kim (Ewha Womans University, South Korea), Yeonhee Kim (Ewha Womans University, South Korea), Jaehyeong Sim (Ewha Womans University, South Korea), and Jongkil Kim (Ewha Womans University, South Korea)	
Session 3A: Text and Natural Language Processing	
Mixterm Topic Model: A Short Text Topic Model in Low-Resource Scenarios Chuangying Zhu (Guilin University of Electronic Technology, China), Yongyu Liang (Guilin University of Electronic Technology, China), Limiao Zhong (Guilin University of Electronic Technology, China), Xinyuan Liang (Guilin University of Electronic Technology, China), and Fei Xie (Guilin University of Electronic Technology, China)	122
Code Summarization using Mamba-SoTaNa Jeonghwa Lee (Jeonbuk National University, South Korea), Eunjeong Ju (Jeonbuk National University, South Korea), and Duksan Ryu (Jeonbuk National University, South Korea)	129
Sentence Embedding using Supervised Contrastive Learning on Hierarchical Categories	136
Session 3B: Infrastructure and Platform for Big Data I	
eCommTouch: A Benchmark Dataset for Touch-Based Continuous Mobile Device Authentication for e-Commerce	

CWODL: Cross-Platform Data Analytics Workflow Optimization with Deep Learning	8
CXLSim: A Simulator for CXL Memory Expander	6
Trusted and Privacy-Preserving Billing System for Smart Grids	0
Session 4A: Infrastructure and Platform for Big Data II	
Dynamic Computing First Network Multi-Dimensional Resource Collaborative Allocation Based on Federated Segmentation Learning	7
MSformer: Joint Prediction of Multi-Dimensional Resources in Computing Power Network Based on Multi-Granularity Sampling	5
Few-Shot Learning-Based Machine Fault Diagnosis using EMD-Gammatone Spectrogram with Limited Labeled Audio Dataset	3
Enhancing Speech Emotion Recognition through Segmental Average Pooling of Self-Supervised Learning Features	1

Session 4B: Security and Fraud Detection

FrAug: Enhanced Fraud Detection in Interbank Transfers via Augmented Account Features 19 Seonkyu Lim (Korea Financial Telecommunications and Clearings Institute, South Korea), Jeongwhan Choi (Yonsei University, South Korea), Jaehoon Lee (LG AI Research, South Korea), and Noseong Park (Korea Advanced Institute of Science and Technology, South Korea)	9
A Framework for Anti-Money Laundering Based on AI and SATP in CBDC Transactions)7
Automated Machine Learning for Predicting Corporate Tax Non-Compliance	.4
Advancing Android Malware Detection with Deep Learning & LLMs in Big Data Ecosystem: A Forward Vision	<u>?</u> 1
Session 5A: Learning Techniques I	
Continual Learning with Flexible Preservation	<u>2</u> 8
A Novel Feature Filtering Mechanism to Avoid Concept Shifts from Deteriorating Model Predictability	36
Yi-Cheng Shih (National Yang Ming Chiao Tung University, Taiwan), Tian-Shyr Dai (National Yang Ming Chiao Tung University, Taiwan),	
Yen-Wu Ti (Xiamen Ocean Vocational College, Taiwan), Yun Kuo (National	
Yang Ming Chiao Tung University, Taiwan), Wun-Hao Wang (National Yang Ming Chiao Tung University, Taiwan), and Ying-Ping Chen (National Yang Ming Chiao Tung University, Taiwan)	
An Approach to Enhancing Fairness in a Dynamically Growing Federated Learning Environment 24 Sean Vucinich (University of Michigan - Ann Arbor, USA), Qiang Zhu (University of Michigan - Dearborn, USA), and Niccolo Meneghetti (University of Michigan - Dearborn, USA)	!4
Dual-Branch Residual CNN-LSTM Model for Emotion Classification using PPG and GSR	52
Session 5B: Knowledge and Question Answering	
Attribute-Controlled Summarization: Enhancing Semantic Approach for Numerical Attribute Prediction on Knowledge Graphs	56

Multilingual Table Question Answering for Low-Resource Languages via Selective Question Translation and Cross-Lingual Adaptation
SWING: Weakly Supervised Table Question Answering with Self-Training via Reinforcement Learning
MERNet: Multimodal Emotion Recognition in Conversation Network via Cross-Modal Knowledge Distillation and Contrastive Learning
Session 6: Data Mining and Analysis I
CoDAQ: Congressional-Based Data Imputation for Aggregate Queries
Relaxed HLP: A Metric for Flexible High-Level Plan Evaluation in Embodied Environments 295 Suyeon Shin (Seoul National University, Korea) and Byoung-Tak Zhang (Seoul National University, Korea)
InsightCube: Accelerating Interactive Insight Discovery in Exploratory Data Analysis
Attention GCN-LSTM Model for Dementia Identification using Functional Brain Networks
Session 7A: Data Mining and Analysis II
IoT Anomaly Detection with Sound-Based Spectral Analysis and Lightweight Model

MUEU: Meta-Learned User Preference Estimation for Cold-Start Recommendation via Enriching the Representations of Users
An Improved Recommendation System using K-Means Clustering and Matrix Factorization 329 Parth N. Thakrar (Dhirubhai Ambani Institute of Information and Communication Technology (DAIICT), India), Vedant V. Patel (Dhirubhai Ambani Institute of Information and Communication Technology (DAIICT), India), and Abhishek Jindal (Dhirubhai Ambani Institute of Information and Communication Technology (DAIICT), India)
SNP Position Filtering and Multimodal Approach in Heart Failure Risk Prediction
Session 7B: Learning Techniques II
Big Data Time Series Prediction Based on Feature-Enhanced Transformer Model
Time-Series Forecasting in Healthcare: A Dual-Framework Approach with Latent Class
Integration
MOAT:An Improved Multi-Modal Matching Algorithm Based on Object Attention and Tensor
Fusion
BIGDAS 2025: Big Data Service Society 2025 International Academic Seminar
Session 1
A Comprehensive Deep Learning Approach for Real-Time Detection and Classification of Large and Small Defects in Industrial Applications

AI-Based Color Prediction Analysis for Recycled Polyester Dyeing Processes Kim Sung Hoon (Chungbuk National University, South Korea), F M Fahmid Hossain (Chungbuk National University, South Korea), Kwan Hee Yoo (Chungbuk National University, South Korea), Shimaa Kotb (Chungbuk National University, South Korea), Kim Ah Rong (DyTech Institute, South Korea), Kim Hyun Jong (Electronics and Telecommunications Research Institute, South Korea), and Chi Su Young (Electronics and Telecommunications Research Institute, South Korea)	365
An Automated Scenario Generation Model for Anti-Phishing using Generative AI	368
Comparison of Crop Image Classification Model Performance According to Image Augmentation Techniques	371
DVSMOT: Dual View Sharing for Enhanced Multi-Object Tracking and ID Switch Reduction3 Heon Jun Shin (Chungbuk National University, South Korea), Joohyung Oh (Chungbuk National University, South Korea), Woo-Seok Choi (Chungbuk National University, South Korea), Joong Hun Cho (Chungbuk National University, South Korea), Yeonseok Lee (Chungbuk National University, South Korea), and Tae-Kyung Kim (Chungbuk National University, South Korea)	375
Session 2	
Enhancement Strategies for the AI Platform for Accelerated Materials Development and Innovation	379
Estimation of Compensation Scheme Functions According to the Level of Image Data De-Identification	382
Framework for Effective Information Security Investments: A Data-Driven Model Recommendation for Information Security Investment Decision-Making	386
Proposing a Business Model Analysis Framework for IoT Bigdata	390

Ar Data Analysis for Efficient Green Hydrogen Production	
Development of a Construction Type Classification Model for Construction using the KLUE-BERT Seung-Hyeon Shin (Chungbuk National University, Republic of Korea), Jeong-Hun Won (Chungbuk National University, Republic of Korea), Moon Gyu Kim (Ministry of Employment and Labor, Republic of Korea), and Doo Hwan Shin (Gangdong University, Republic of Korea)	397
IWDS 2025: The Eighth International Workshop on Dialog System	S
Session 1: Natural Language Understanding	
Quality Assurance Framework for Multimodal Assessment Datasets on AI Risk Factors	401
An Analysis of Unsafe Responses with Magic Expressions Across Large Language Models Eojin Joo (KAIST, South Korea), Chae-Gyun Lim (KAIST, South Korea), Jeongyun Han (University of Seoul, South Korea), and Ho-Jin Choi (KAIST, South Korea)	406
Exploring the Visual Perspective Taking of Vision-and-Language Model	410
Session 2: Chatbots and Applications	
Analytical Approach for Truthfulness Detection using Multimodal Features: A Case Study with Vision and Audio Data Jangkyum Kim (Sejong University, South Korea), Ji-Won Choi (Sejong University, South Korea), Donghyeon Park (Sejong University, South Korea), and Young Dae Ko (Sejong University, South Korea)	412
Personalized Conversation Management with Emotional Memory	415

BigDefense2025

Time Series Data Classification using the Knowledge-Augmented Transformer Model
Evaluation of LLM-Based VQA Models for Enhancing Battlefield Situation Awareness
STP 2025: The Fifth International Workshop on Science and Technology Policy
Session 1: Energy Diffusion and the Climate Environment
Case Study and Qualitative Comparison of Global Legal Frameworks for Platform Labor
Decomposition Analysis of Household Electricity Consumption in South Korean Municipalities 430 Yeolhoon Sung (Ajou University, Republic of Korea), Jooyeoun Lee (Ajou University, Republic of Korea), and Myoungsug Chung (Ajou University, Republic of Korea)
Author Index