15th International Conference on Data Science, Technology and Applications (DATA 2026)

Volume 1

Porto, Portugal
16-18 July 2026

Part 1 of 2

ISBN: 979-8-3313-3992-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2026) by SCITEPRESS – Science and Technology Publications, Lda.
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda. at the address below.

SCITEPRESS – Science and Technology Publications, Lda.
Avenida de S. Francisco Xavier, Lote 7 Cv. C,
2900-616 Setúbal, Portugal

Phone: +351 265 520 185
Fax: +351 265520 186

info@scitepress.org

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
Email: curran@proceedings.com
Web: www.proceedings.com

CONTENTS

INVITED SPEAKERS

KEYNOTE SPEAKERS

Property Graph Transformations in Action: From Data Integration to Causal Analysis 5
Angela Bonifati

The Future of Agriculture through Data Science: Challenges and Opportunities for Autonomous Field Systems 7
João P. Papa

Data Management Meets Its AI Partner 9
Alon Halevy

DATA SCIENCE & MACHINE LEARNING

FULL PAPERS

Customer Churn Prediction in the Telecommunications Sector Using Explainable Artificial Intelligence 15
David Veríssimo, Joana Leite and Maryam Abbasi

Evaluating Heterogeneous Graph Neural Networks for Fashion Retail Recommendation: A Comparative Study on the H&M Dataset 25
Davi Rocha de Santana, Leila Weitzel, Anderson Amendoeira Namen and Graziela Ferreira Guarda

Structural-Temporal Anomaly Detection in Scholarly Networks Using Hyperbolic Representations 32
Renata Avros, Shoval Ben Shushan, Adar Budomski, Dvora Toledano-Kitai and Zeev Volkovich

Deep Representative Text Modeling for the Consistent Identification of SARS-CoV-2 Anomalies 44
Renata Avros, Valery Kirzner, Dvora Toledano-Kitai and Zeev Volkovich

Community-Aware Spectral Anomaly Detection in Scholarly Graphs 54
Itay Mohabati, Shai Yosef, Renata Avros and Zeev Volkovich

Imputation Strategies for Predicting Tropospheric Ozone 65
Miraceli Bonjardim Waldemar, Bruno Nazario Alves, Fernando Augusto Silveira Armani and Yuri Kaszubowski Lopes

An Optimized Ensemble Framework with Explainable AI for Proactive Credit Card Fraud Detection in Banking 74
Henrique Barros, Francisco Antunes and Maryam Abbasi

Continuous Operation of a Myoelectric Hand Prosthesis Using Reinforcement Learning 86
Carlos Ayala-Tipan, Jose-Fuertes Bustos, Lorena Isabel Barona López, Ángel Leonardo Valdivieso Caraguay and Marco E. Benalcázar

SecureBreak: A Dataset towards Safe and Secure Models 98
Marco Arazzi, Vignesh Kumar Kembu and Antonino Nocera

Prompting vs Ensemble Architectures for Arabic English Code-Switched Classification 110
Donia Ali, Salma Haytham, Sandra George and Caroline Sabty

ResSET: Enhanced Motor Imagery Classification Using Hybrid Transformer-ResNet Architectures and Generative Data Augmentation 121
Duc Thien Pham and Roman Mouček

ContrastSkill: Task-Oriented Contrastive Intermediate Training for Skill Extraction 133
Aleksander Bielinski, David Brazier, Dimitra Gkatzia and Alistair Lawson

Enabling Efficient Domain Adaptation via Noise-Enhanced Flow Matching 145
Aitian Ma, Dongsheng Luo and Mo Sha

A Faithful Multi-Head Framework for Explainable Toxicity Detection with Structured LLM-Based Synthetic Supervision 154
Linh Nguyen and Tran Anh Tuan

Evaluating Local Explainability Metrics for Machine Learning Models on Tabular Data 166
Tomás Pereira, João Vitorino, Eva Maia and Isabel Praça

Actionable Explainable AI for Credit Risk: A SAFE-AI Framework Integrating SHAP and Counterfactual Explanations with DiCE 178
Matheus Francelino Bezerra da Silva and Carlos Henrique Quartucci Forster

Dysgraphia Detection from Handwritten Text Using a Hybrid ConvNeXt-Transformer Model 190
Wallace Primo Freire Junior, Simone Bello Kaminski Aires and Maria Eduarda Guedes Pinto Gianisella

Optimising Wheel Replacement Schedules for Freight Trains Based on Machine Learning Models 202
Max Strang, Shaibal Barua, Mobyen Uddin Ahmed and Shahina Begum

A Two-Level Machine Learning Based-Intrusion Detection System for IoT Healthcare Application Based on Blockchain 214
Manel Boujelben and Manel Hentati

Query Adaptive Rank Fusion: A Training Free per Query Weighting Scheme for Sparse–Dense Hybrid Retrieval 223
Mohammad Saad Rafeeq, Chandramani Chaudhary, Nirmal Kumar Boran and Ashok Kumar Suhag

Explaining Rankings with Prototypes and Criticisms 234
Lars Schütz and Korinna Bade

Deep Learning on the Wing: A Survey of Resource-Efficient Object Detection and SLAM for Autonomous UAVs 243
Rayan Abri, Sara Abri and Salih Cetin

SIGMA: A Lightweight Multi-Perspective Statistical-Signature Framework for Concept Drift Detection in Business Process Event Logs 249
Lerina Aversano, Felice Franchini and Debora Montano

Probabilistic Variational Archetypal Learning on the Simplex for Interpretable Medical Image Analysis 258
Ali Yassin, Narjes Soudin, Abbass Rammal and Mageda A. A. Sharafeddin

Plywood Veneer Quality Classification Using CNNs with Squeeze-and-Excitation Blocks 270
Cristian Gotardo, Andreia Marini, Marlon Marcon and André Roberto Ortoncelli

SHORT PAPERS

Evaluating England's Plastic Bag Reduction Strategies, 2015-2025 285
Erdem Sarialioglu

Selection Bias Correction in Retail Intelligence 291
Spandan Ghose Chowdhury

Towards Fine-Grained Emotion Detection in Hebrew: Dataset Creation and Evaluation 299
Natalia Vanetik, Shilat Haya Yosefi and Keturah Shlomo

Identification of Relevant Association Rules from Biased Target Value Distributions of Empirical Studies 307
Friedemann Schwenkreis

Towards Realistic Privacy-Preserving Stress Detection for Nurses Using Smartwatch Data 317
Cedric Aebi, Albin Grataloup, Souhir Ben Souissi and Christoph Golz

Behavioral Profiling of Active E-Commerce Customers: Integrating Post-Purchase Interaction Signals into Comparative Clustering Models 325
Serena Tomakyan, Arda Raştak and Nilay İşeri

DeepSupp: Attention-Driven Correlation Pattern Analysis for Dynamic Time Series Support and Resistance Levels Identification 331
Boris Kriuk, Logic Ng and Zarif Al Hossain

Enhancing Visual Perception in Foggy Conditions via Multiclass Fog Density Modeling 338
Mohamad Mofeed Chaar and Galia Weidl

Spatially Constrained Clustering for Analyzing Spatio-Temporal Dynamics of Automobile Insurance Risk 346
Shengkun Xie, Samy El Mouttalibi, Jin Zhang and Clare Chua-Chow

Hybrid Approach Based on Supervised Learning and Synthetic Data Generation for University Student Dropout Risk Prediction 354
Laura Caicedo, Juan José Muñoz and Nestor Diaz

Regime Detection and Forecasting of Financial Indicators in Electric Transmission Sector Companies Using Hidden Markov Models 363
Giovanni Vallim Maniezzo, Elder Oroski and Luis Ledo Mota Melo Júnior

Evaluating Nearest-Neighbor Variants for Time Series Classification with Manhattan-Based Measures 371
Zoltán Gellér, Vladimir Kurbalija and Mirjana Ivanović

A Unified Semantic Representation Framework for Multi-Source Heterogeneous Data Fusing Protocol Syntax Constraints and Cross-Source Alignment: A Case Study of in-Vehicle Intrusion Detection 379
Yixuan Liu, Cong Yang, Haibo Cheng, Kaimin Zhuo, Peng Liu and Yixuan Wang

Sequence-Constrained Data Augmentation: A First- and Second- Order Markov Transition Matrix Approach 387
Hanyu Li, Yuanbo Kong and Meng Wang

Quantum Computed Moments Drive Orthogonal Neodymium Selectivity in Lanmodulin Dimer Pockets through Hamiltonian Distance Learning 395
Don Roosan, Mazharul Karim, Rubayat Khan and Md Rahatul Ashakin

Quantum Pharmacogenomics for Polypharmacy Risk Minimization in Elderly Care 403
Don Roosan, Mazharul Karim, Rubayat Khan and Inyene Essien-Aleksi

Quantum Kernel Support Vector Machine for Quantum Dot State Recognition 411
Md Rahatul Ashakin, Rubayat Khan, Avik Mahata, Mazharul Karim, Nuruzzaman Sojib and Don Roosan

From Prediction to Decision: A Counterfactual Machine Reasoning Framework for ESG Analysis 419
Tanzina Nizam and Yeon Kyupil

Scored Rule Sets for Interpretable Multiclass Classification 426
Robin Nunkesser

Enhancing Wildfire Prevention: An Operational Pipeline Integrating WRF Mesoscale Modelling and Automated CFFDRS Calculation 434
Afonso Oliveira, Eva Maia and Isabel Praça

A Multi-Mode Online Learning Framework for Drift-Aware Maritime Fuel Consumption Prediction 442
Nadeem Iftikhar and Finn Ebertsen Nordbjerg

Integrating Causal Inference with Graph Neural Networks for Alzheimer's Disease Analysis 450
Pranay Kumar Peddi and Dhrubajyoti Ghosh

Uncertainty Quantification for Annual Budgets: Bootstrap Methods on Short Monthly Time Series 458
Emmanuel Zarpas and Martín Gómez-Abejón

Advancing m6Am Site Prediction Through Deep Learning on Diverse mRNA Sequence Landscapes: The m6Am-DLcat Approach 466
Alfredo Cuzzocrea, Tasmin Karim, Md. Shazzad Hossain Shaon, Md. Fahim Sultan and Mst Shapna Akter

A Two-Stage Clustering Framework for Profiling E-Commerce Search Queries 475
Melis Öztürk Umut, Enis Teper and Mustafa Keskin

Hierarchical Structure-Based Influence Maximization 480
Longyun Wang, Qing Zhang, Min Wu, Huijun Zheng, Wenchuan Yang and Xin Lu

Outlier Detection in Low-Cost Air Quality Sensor Networks: Distinguishing Sensor Errors from Environmental Events 488
Victoria Beltran, Daniel Fernandez-Garcia and Jose Santa

A Statistical Analysis of the State-of-the-Art Transformer-Based Remaining Useful Life Models on C-MAPSS FD001 500
Gabriel Vinicius Boin Freitas and Yuri Kaszubowski Lopes

Neural Network-Based Group Sparsity for Nonlinear Feature Selection in Multi-Output Problems 508
Mamadou Kanouté and Florence Forbes

Age- and Sex-Specific Normative Profiles of Reactive Strength Index in Football Athletes: A GAMLSS Approach 517
P. D. Chougale and U. Ananthakumar

Agreement beyond Chance in LLM-Based Hazard Severity Classification for Railway Functional Safety 524
Padma Iyenghar

Ontology-Driven Semantic ETL for Executable Knowledge Engineering 536
Antony Seabra, Daniel Schwabe and Sergio Lifschitz

AgentSearch: Learning Efficient Agentic Workflows via Deliver Tree Search 544
Dinesh Chowdary Attota and Ying Xie

Developing an ML Model for Orthophoto Segmentation Using Automatically Generated Labels 555
Kevin Kocon and Michel Krämer

Data First, AI Second: A Four-Lenses Conceptual Architecture for Enabling AI to Deliver Precision Impact 563
Christopher J. Garasi and Aaron J. Moreno

CITRUS: Screening the Past, Ranking the Future 571
Sayak Chakrabarty, Souradip Pal and Imon Banerjee

DeepSeekMath Meets Order Book: Group-Aware Policy Optimization for High-Frequency Directional Trading 579
Sayak Chakrabarty and Souradip Pal

Intent-Aware Predictive Orchestration for Softwarized Networks Using LLM-Assisted Policy Translation 587
Pınar Ersoy and Mustafa Erşahin

Progressive Multi-Objective Optimization for Improved t-SNE Embeddings 594
Samir Brahim Belhaouari, Skander Bensegueni, Lyes Fennour and Dhia El Hak Amani

Chunking Is Not the Bottleneck: Why RAG Evaluation Must Look beyond Segmentation Strategy 605
Aaditya Chauhan and Nivas Hegde

Detection of Anxiety and Depression from Social Media Text Using Natural Language Processing 613
Muhammad Azhar, Adeen Amjad, Bilal Hussain, Chan Kit Ling, Taimoor Niaz and Muhammad Hamza Akbar

Text-to-SQL with Large Language Models: Challenges Revisited and New Dimensions 621
Luca Sala, Giovanni Sullutrone and Sonia Bergamaschi

Temporal Stability in Student Dropout Prediction Using Weekly Learning Analytics 629
Tiago Franco, Paulo Alves, José Rufino, Maria de Fátima Pacheco and Nuno Ribeiro

Individual-Level Probabilistic Model Integrating Rule Representation and Logistic Curves 637
Akari Oda, Yoichi Seki and Kaoru Shimada

CRAFT: Clustered Regression for Adaptive Filtering of Training Data 647
Parthasarathi Panda, Asheswari Swain and Subhrakanta Panda

AI-Powered Early Detection of Child Malnutrition Using Dimensionless Anthropometric Ratios and Deep Learning 655
Vaishnu Kanna and Subhrakanta Panda

BLEO: A Binary Language Education Optimization for Feature Selection 661
Beatriz Souza Sé Barros, Guilherme N. Marques, Douglas Rodrigues and João Paulo Papa

Data-Driven Multi-Horizon Taxi Demand Forecasting Using Transformer-Based Temporal Modeling 669
Magesh Ram Rajakumar, Christine Markarian and Shadi Atalla

Automated & Real-Time Privacy Quantification for Microservices Architectures 677
Catarina Silva, Bernardo Falé, João Paulo Barraca and Paulo Salvador

Exceptional Itemsets with Statistical Background Discovered by Evolutionary Computation 685
Kei Arai and Kaoru Shimada

Time Series Forecasting of Sports Ticket Sales with Venue-Independent Fan Segments 694
Shun Takagi, Haru Terashima, Naoki Tamura, Kazuyuki Shoji, Kenta Urano, Takuro Yonezawa and Nobuo Kawaguchi

Corporate Reputation Exploration with Knowledge Graphs on Bipolar Extremes for Massive Online Consumer Review Datasets 704
R. E. Loke and T. Schouten

SwitchEmbed: Representation Learning for Arabic-English Code-Switched Text 712
Mariam Rizkallah, Amani Ghonim, Ahmed Sherif and Caroline Sabty

Benchmarking YOLOv8 and YOLOv11 for Meteor Echo Detection in BRAMS Radio Spectrograms 720
Diogo Ramalho Fernandes, Stijn Calders, Olivier Parisot, Mickaël Stefas, Hervé Lamy and Katrien Kolenberg

Evaluating and Comparing Optimized CNN and Transformer Architectures for EEG-Based Depression Detection 728
Khaoula Ansari, Ibtissam Abnane and Ali Idri

Toward Reliable Automated Radiography: Improving Chest X-Ray Diagnosis through Multimodal Fusion and Label Cleaning 736
Barbod Yadali Jamalouei, Pourya Jafari and Rayan Abri

Identifying Brand Advocates: A Neural Network Approach to Social Media Analysis 744
R. E. Loke and O. Ran

Media-Derived Epidemic Intelligence for Avian Influenza: A Multi-Granularity Evaluation of LLM-Augmented Event-Based Surveillance 757
Bibek Paudyal, Karine Deschinkel, David Laiymani and Christophe Guyeux

Beyond Accuracy: Comparative Explainability Analysis of Random Forest and XGBoost Using SHAP and LIME 766
Danilo Medeiros Eler, Leonardo Menosse Ribeiro, Thales Vinicius de Brito Ue and Wilson Estecio Marcílio Junior

Agentic GraphRAG and Deterministic Schema Reconciliation for High-Compliance Domains: An LLMOps and FinOps Approach 774
Marcelo Massashi Simonae, André Roberto Ortoncelli and Marlon Marcon

SENTINEL: Semantic Validation of Financial News Clustering with LLM-Based Evaluation 782
Braeden Mefford, Andrew Berg, Houston Miles and Mia Y. Wang

Evaluating Retrieval-Augmented Generation on Social Bias Benchmarks across Small Language Models 790
Manuel Amir Freer Valdez and Arghir-Nicolae Moldovan

A Comprehensive Evaluation of Pre-Trained Language Models for Irony Detection in Tweets 798
Mayara C. Marinho and Vinicius R. P. Borges

DATABASES & DATA MANAGEMENT

FULL PAPERS

Scalable Data Management for Smart Campus Digital Twin Applications 811
Luís Carlos Casanova Afonso, Jo ao Rafael Almeida and José Luís Oliveira

Read Fast, Write Carefully: Empirical Evidence on the Performance Trade-Offs of Denormalization in PostgreSQL under TPC-H Workloads 823
Luís Mendonça, Pedro Martins, Filipe Madeira and Maryam Abbasi

Racial Inequality in Brazilian Computing Education: An Analysis of Black Students Using Census Data 835
Francisco Carlos Monteiro Souza, Marlon Marcon, André Roberto Ortoncelli, Rodolfo Adamshuk Silva and Alinne Cristinne Corrêa Souza

SHORT PAPERS

Modeling Higher-Order Relationships in the Context of Big Data: Methods, Applications, and Prospects 849
Huijun Zheng, Qing Zhang, Longyun Wang, Min Wu, Wenchuan Yang and Xin Lu

Research on Core Technology Identification Methods Based on High-Order Dependency Metrics 857
Qing Zhang, Siyu Lai, Huijun Zheng, Longyun Wang, Wenchuan Yang, Min Wu and Xin Lu

Finding the Sweet Spot: Query Cost and Load Distribution for Spatio-Temporal Queries in Peer-to-Peer Systems 866
Zhan Ye, Laurent Yeh and Iulian Sandu Popa

MLBRS: A Multi-Layer Behavioural Risk Scoring Framework for Insider Threat Detection 874
V. L. Kartheek, Aayush Shah, Rishav Jain, R. Gururaj and Subhrakanta Panda

DATA ENGINEERING, INFRASTRUCTURE AND BUSINESS APPLICATIONS

FULL PAPERS

Data-Centric Lifecycle Integrity Validation for IEC 61508 Using Executable Documentation 887
Padma Iyenghar, Christopher Zimmerman and Claudio Gregorio

Decision Support System for the Use of Robotic and Intelligent Process Automation in Small and Medium-Sized Enterprises 899
Ann-Kathrin Holaj and Sascha Alpers

Beyond Context Windows: Data Transformations at Scale with LLMs 910
Abhigith Abraham, Fariz Rahman, Fadil Rahman, Monika Janchevska and Tanaya Amar

SHORT PAPERS

Graph-Based Named Entity Management System with Ontology-Supported Matching Capabilities 925
Andrea Leoni, Andrea Molinari and Simone Sandri

Architectural Alternatives for Data Distribution in Autonomous Driving: From Real-Time Performance to Semantic Interoperability 933
Mario Martín-Pérez and Marta Zorrilla

Closing the Loop: How Fleet IoT and Explainable AI Can Target Severe-Crash Risk on U.S. Roadways 941
Suryakant Kaushik

A Holistic Framework for Assessing and Representing Vehicle Measurement (Time-Series) Data Quality via a Single Quality Score 949
Sven Dominka and Henrik Smith

AUTHOR INDEX 957