

Petra Perner (Ed.)

# **Machine Learning and Data Mining in Pattern Recognition**

15th International Conference on Machine  
Learning and Data Mining, MLDM 2019,  
New York, NY, USA, July 20-25, 2019  
Proceedings  
Volume II

**ibai** Publishing

---

[www.ibai-publishing.org](http://www.ibai-publishing.org)

Volume Editor

Petra Perner  
Institute of Computer Vision and Applied Computer Sciences, IBaI  
PF 30 11 14  
04251 Leipzig  
E-mail: [pperner@ibai-institut.de](mailto:pperner@ibai-institut.de)

ISSN 1864-9734  
ISBN 978-3-942952-63-7



The German National Library listed this publication in the German National Bibliography.  
Detailed bibliographical data can be downloaded from <http://dnb.ddb.de>.

ibai-publishing  
Prof. Dr. Petra Perner  
PF 30 11 38  
04251 Leipzig, Germany  
E-mail: [info@ibai-publishing.org](mailto:info@ibai-publishing.org)  
<http://www.ibai-publishing.org>

**Copyright** © 2019 ibai-publishing  
ISSN 1864-9734  
ISBN 978-3-942952-63-7

All rights reserved.  
Printed in Germany, 2019

# 15th International Conference on Machine Learning and Data Mining, MLDM 2019

July 20-24, 2019, New York, USA  
[www.mldm.de](http://www.mldm.de)

## Chair

Petra Perner  
Institute of Computer Vision and Applied Computer Sciences, IBA  
Germany

## Program Committee

Reneta Barneva	The State University of New York at Fredonia, USA
Michelangelo Ceci	Universtiy of Bari, Italy
Ireneusz Czarnowski	Gdynia Maritime University, Poland
Roberto Corrizo	Universtiy of Bari, Italy
Christoph F. Eick	Universtiy of Houston, USA
Mark J. Embrechts	Rensselaer Polytechnic Institute and CardioMag Imaging, Inc, USA
Ana Fred	Technical University of Lisboa, Portugal
Giorgio Giacinto	University of Cagliari, Italy
Aminata Kane	Concordia University, Canada
Piet Kommers	University of Twente, The Netherlands
Olga Krasotkina	Russian State University, Russia
Dimitris Karras	Chalkis Institute of Technology, Greece
Adam Krzyzak	Concordia University, Canada
Valerio Pascucci	University of Utah, USA
Gianvito Pio	University of Bari, Italy
Francis E.H. Tay	National University of Singapore, Singapore
Turki Turki	King Abdulaziz University, Saudi Arabia
Zeev Volkovich	ORT Braude College of Engineering, Israel
Patrick Wang	Northeastern University, USA

## Preface

The fifteenth event of the International Conference on Machine Learning and Data Mining MLDM was held in New York ([www.mldm.de](http://www.mldm.de)) running under the umbrella of the Worldcongress “The Frontiers in Intelligent Data and Signal Analysis, DSA2019” ([www.worldcongressdsa.com](http://www.worldcongressdsa.com)).

For this edition the Program Committee received 245 submissions. After the peer-review process, we accepted 65 high-quality papers for oral presentation. The topics range from theoretical topics for classification, clustering, pattern mining to specific data mining methods for the different multimedia data types such as image mining, text mining, video mining and web mining. Extended versions of selected papers will appear in the *International Journal Transactions on Machine Learning and Data Mining* ([www.ibai-publishing.org/journal/mldm](http://www.ibai-publishing.org/journal/mldm)).

A tutorial on Data Mining and a tutorial on Case-Based Reasoning were held before the conference that took pleasure to high participation of researchers and practitioners from industry, social and public services.

We like to thank all presenter for your high-quality presentations and the audience for your high-professional questions and inspiring comments. All that has made the conference to a living and dreadful event. The presenters and the audience went home with a full bag of new insights into different topics the research and inspiring ideas for new work and research. Besides that, gave the banquet an excellent opportunity to network among the participants and set up new co-operations.

We like to thank all reviewers for their highly professional work and their effort in reviewing the papers. We also thank members of Institute of Applied Computer Sciences, Leipzig, Germany ([www.ibai-institut.de](http://www.ibai-institut.de)) who handed the conference as secretariat. We appreciate the help and understanding of the editorial staff of ibai-publishing house ([www.ibai-publishing.org](http://www.ibai-publishing.org)) that prepared and published the proceeding books in two volumes.

We invite you to join us in 2020 in New York to the next Worldcongress ([www.worldcongressdsa.com](http://www.worldcongressdsa.com)) “The Frontiers in Intelligent Data and Signal Analysis, DSA2020” that combines under his roof the following three events: International Conferences Machine Learning and Data Mining MLDM, the Industrial Conference on Data Mining ICDM , and the International Conference on Mass Data Analysis of Signals and Images in Artificial Intelligence and Pattern Recognition with Application in with Applications in Medicine, r/g/b Biotechnology, Food Industries and Diagnostics, Biometry and Security, Agriculture, Drug Discover, and System Biology MDA-AI&PR.

July, 2019

Petra Pernert

## Table of Content

Novel approach for Label Disambiguation via Deep Learning <i>Ticiana L. Coelho da Silva, Natanael da Silva Araújo, José Antônio F. de Macedo, David Araújo, Felipe Melo Soares, Paulo A. L. Rego, and Aloisio Vieira Lira Neto</i> .....	431
On combining dynamic selection, sampling, and pool generators for credit scoring <i>Leopoldo Melo Junior, Franco Maria Nardini, Chiara Renso, and Jose Antonio Macedo</i> .....	443
Comparative Analysis of Features Selection Techniques for Classification in Healthcare <i>Shruti Kaushik, Abhinav Choudhury, Ashutosh Kumar Jatav, Nataraj Dasgupta, Sayee Natarajan, Larry A. Pickett, and Varun Dutt</i> .....	458
Ontology-based Neural Network for Energy Optimization in Data Centers <i>Zhan Li, Li Chen, Guan Wng, Feng Zeng, and Yun Zhang</i> .....	473
Semantic similarity estimation for domain specific data using BERT and other techniques <i>R. Prashanth</i> .....	488
Educational Data Mining: A Hybrid Approach to Predicting Academic Performance of Students <i>Marlia Nayara Clemente de Almeida Lima, Geovanne Oliveira Alves, Wedson Lino Soares, and Roberta Andrade de Araujo Fagundes</i> .....	500
On the Vapnik-Chervonenkis Dimension of Binary Tables <i>Roman Sizov and Dan A. Simovici</i> .....	515
A Streaming Analytics Language for Processing Cyber Data <i>Eric L. Goodman and Dirk Grunwald</i> .....	527
SumOpinions: Automatic Summarization of Reviews about Tourist Places <i>João Holanda Freires Junior, José Antônio de Fernandes Macêdo, and Igo Ramalho Brilhante</i> .....	542
Support Vector Machine with Graphical Network Structures in Features <i>Wenqing He, Grace Y. Yi, and Li-Pang Chen</i> .....	557
Expanding annotated data with informed labels for weak supervision <i>Eura Shin, Sam Berglin, Jacob Furst, and Daniela Raicu</i> .....	671

Anomaly Detection in Process Signals within Machine Learning and Data Augmentation Approach*	
<i>Sergey Gavrín, Damir Murzagulov, Alexander Zamyatin</i> .....	585
Dynamic Facial Expression Recognition Based on Visual Rhythms and Motion History Images	
<i>Jadisha Yarif Ramírez Cornejo and Helio Pedrini</i> .....	599
Data Reduction Effects on Classification Accuracy and Training Time	
<i>Reham M. Alamro and Abdou S. Youssef</i> .....	614
Data Mining Framework to Derive Measures for Road Safety	
<i>Katherina Meißner and Julia Rieck</i> .....	625
Failure Analysis on Multivariate Time-series Data given Uncertain Labels	
<i>Hao Huang, Shinjae Yoo, and Yunwen Xu</i> .....	640
Feature Evaluation for Project Preferences Representation in GitHub	
<i>Thaciana Cerqueira, Leandro Marinho, and Franklin Ramalho</i> .....	655
NOV-mRSI: A Novel Algorithm for Minimal Rare Significance Itemsets Mining NOT Satisfy the Downward Closure Property in Transactional Databases	
<i>Huan Phan and Bac Le</i> .....	669
Deep Convolutional Multi-Stream Network Detection System Applied to Fall Identification in Video Sequences	
<i>Sarah A. Carneiro, Gabriel P. da Silva, Guilherme V. Leite, Ricardo Moreno, Silvio Jamil F. Guimarães, and Helio Pedrini</i> .....	681
Edge Detection Using K-means and Run-Based Two-Scan Connected Components labeling adapted for grayscale images	
<i>Chen Avni and Maya Herman</i> .....	696
The Representation of Wave Trend Direction in Stock Price Time Series via RNN	
<i>Yi Wei and Vipin Chaudhary</i> .....	704
Two-Sample Tests for High Dimensional Distributions Using One-Dimensional Random Projections	
<i>Yibo Zhao, Jin Cao, and Ran He</i> .....	718
The Computing Algorithm of Barrier Tree Based on the Basin Hopping Graph in RNA Structure	
<i>Zhendong Liu, Gang Li, and Patrick Wang</i> .....	733
Hyperspectral Remote Temperature Estimation Robust to Atmospheric Variation with Fourier Transform Infrared and Deep Learning	
<i>Sungho Kim, Jungho Kim, Jinyong Lee, and Junmo Ahn</i> .....	743

Binary Sparse Dynamic Time Warping <i>Youngha Hwang and Saul B. Gelfand</i> .....	748
Applications of Statistical and Machine Learning Time-Series Methods for Predicting Internet Usage <i>Naveksha Sood, Shruti Kaushik, Aditya Nigam, Sriram Kailasam, Dileep D. and Varun Dutt</i> .....	760
A “Sentimental Analysis” of Computer-Generated Papers <i>Dan Lemberg and Zeev Volkovich</i> .....	775
Lifetime reward optimization for online ads experiment <i>Liang Dai and Ram Akella</i> .....	782
Hierarchical Deep-Fusion Learning Framework for Lung Nodule Classification <i>Kazim Sekeroglu, Omer Muhammet Soysal, and Xin Li</i> .....	792
Object Detection and Classification in Low Resolution Underwater Imagery Using Deep Neural Network and Motion-Inspired Features <i>Ahsan Jalal and Ahmad Salman</i> .....	805
Modeling and Simulation of UEFA Champions League <i>Saman Muthukumarana and Ninad Khargonkar</i> .....	820
Stochastic Optimization of Plain Convolutional Neural Networks with Simple Methods <i>Yahia Saeed Assiri</i> .....	833
Annotating Biodiversity Spreadsheets through 5W1H based on Machine Learning <i>Ivelize R. Bernardo, Helio Pedrini, and André Santanché</i> .....	845
Topological Structure of Linear Manifold Clustering <i>Artyom Diky and Robert M. Haralick</i> .....	860
Wheat Ear Detection in RGB and Thermal Images Using Deep Neural Networks <i>Zeljana Grbović, Marko Panić, Oskar Marko, Sanja Brdar, and Vladimir Crnojević</i> .....	875
DeepSeed: A Deep Learning Methodology for Automated Soybean Seed Damage Classification <i>Luigi Freitas Cruz, Priscila Tiemi Maeda Saito, and Pedro Henrique Bugatti</i> ..	890
Dynamic Segmentation and Retrieval of Daily Life Activities from Multimodal Life-logs <i>Rashmi Gupta and Cathal Gurrin</i> .....	901

VIII

Evaluation of ARIMA, LSTM and Hybrid models for Multi-Step ahead Forecasting <i>Neha Gupta, Purushothaman Ethiraj</i> .....	916
Appendix .....	934