15th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2019)

New York, New York, USA
20-25 July 2019

Volume 1 of 2

Editors:

Petra Perner

ISBN: 978-1-5108-9516-4
# Table of Content

Arabic Rule-based Named Entity Recognition System Using GATE  
*Hatem M. ElSherif, Khaled Mohammad Alomari, Ahmad Qasim AlHamad and Khaled Shaalan* ............................................................... 1

An Unsupervised Statistical Moving Shadow Detection Method for Video Analysis  
*Hang Shi and Chengjun Liu* ................................................................. 16

Social Activism Analysis: An Application of Machine Learning in the World Values Survey  
*Francielle M. Nascimento, Dante A. C. Barone, and Henrique Carlos de Castro* ................................................................. 28

Deep Dilated Convolutional Nets for the Automatic Segmentation of Retinal Vessels  
*Ali Hatamizadeh, Hamid Hosseini, Zhengyuan Liu, Steven D. Schwartz, and Demetri Terzopoulos* ................................................................. 39

Clonal Selection Algorithm Applied to Object Recognition in Mobile Robots  
*Jose Guillermo Guarnizo and Luis Fernando Nino* ................................................................. 49

Customized use of Expectation Maximization for Glioma Identication from Brain MRIs  
*Nidhi Gupta, Pushpraj Bhatele, and Pritee Khanna* ................................................................. 63

A Novel Phonetic Algorithm for Predicting Chinese Names using Chinese PinYin  
*Hua Zhao and Fairouz Kamareddine* ................................................................. 78

Detecting Customer Churn Signals for Telecommunication Industry Through Analyzing Phone Call Transcripts with Recurrent Neural Networks  
*Junmei Zhong, William Li* ................................................................. 93

A comparison of linear and machine learning models for the simulation of soil moisture  
*Milan Cisty, Frantisek Cyprich and Veronika Soldanova* ................................................................. 104

Vehicle Classification in Video Using Deep Learning  
*Mohammad O. Faruque, Hadi Ghahremannezhad, and Chengjun Liu* ................................................................. 117

A Model for Diagnosis of Alzheimer’s Disease Using Haralick Texture Features and Meta Feature Selection  
*Iago Richard Rodrigues Silva, Marilia Nayara Clemente de Almeida Lima, Wellington Pinheiro dos Santos, and Roberta Andrade de Araújo Fagundes* ................................................................. 132
Mining Data Stream to Detect Behavior Change in a Real-Time Strategy Game  
Eldane Vieira, Rita Maria Silva Julia, and Elaine Ribeiro de Faria .......................... 146

Packet2Vec: Utilizing Word2Vec for Feature Extraction in Packet Data  
Eric L. Goodman, Chase Zimmerman, and Corey Hudson ........................................ 161

Handbags Classification Model via Deep Learning  
Dieinison J. F. Braga, Leodecio Braz S.S., Criston Pereira de Souza, and 
Ticiana L. Coelho da Silva ...................................................................................... 176

Long Short-Term Memory-based Multi-Period Price Prediction for Portfolio Management  
Zhengyong Jiang and Frans Coenen ......................................................................... 187

Fraud Detection Using Explainable Machine Learning Algorithms  
Luciano C M Andrade, André C P L F Carvalho ...................................................... 201

A Fast AEMST Algorithm for High-Dimensional Datasets by Removing Redundant Distance Computations  
Xiaochun Wang, Xia Li Wang, Xuan Xiong Lin ....................................................... 216

Server Failure Prediction Framework for IT Infrastructure Supporting Critical Applications  
Mythili Krishnan and Madhan Kumar Srinivasan ....................................................... 229

A New Minimum Spanning Tree Algorithm Constructed in a Backward Fashion  
Aozhong Wang, Xiaochun Wang, Xia Li Wang ......................................................... 243

Analysis of Linear and Non-Linear Classifiers in Imbalanced Data to Predict Diabetes Induced Complications  
Tahsinur Rahman, Aniqa Zaida Khanom, Sheikh Mastura Farzana, Sharovar Md. Shahriar Khan and Dr. Md. Ashraful Alam ...................................................... 255

FF-SVM: New FireFly-based Gene Selection Algorithm for Microarray Cancer Classification  
Nada Almugren and Hala Alshamlan ......................................................................... 270

Fast K-medoids Clustering Algorithm Using Triangle Inequality  
Xiaochun Wang, Xia Li Wang, Xuan Xiong Lin ......................................................... 282

Design and Implementation of an Intrusion Detection System using Deep Neural Network  
Hyun-chul Chang and Sungbum Park ...................................................................... 296
Community detection in networks with fuzzy boundaries
Pradumn Kumar Pandey, T. Ramalingeswara Rao and Satyajeet Maharana ... 304

Precise Feature Selection and Case Study of Intrusion Detection in an Industrial Control System (ICS) Environment?
Terry Guo, Animesh Dahal, and Ambareen Siraj ............................... 319

Entropy-based Approach for Parameter-free Attribute Clustering
Adison Khomprasert, Thanawin Rakthamanon, Kitsana Waiyamai .............. 333

Seq2SQL - Evaluating Different Deep Learning Architectures Using Word Embeddings
Kevin Stower and Dirk Krechel .............................. 343

Sparse Multimodal Classification of EEG Signals from Rapid Serial Visual Presentation of Diagnostic Images
Valentina Sulimova, Sergey Bukhonov, Olga Krasotkina, Vadim Mottl, Annette Sterr, Kevin Wells, David Windridge ........................................... 355

Neural-Attention Multi-Instance Learning for Predicting User Demographics from Highly Noisy Tweets
Qing Chen, Mingxuan Sun, and Jian Zhang ................................. 367

Explainable Artificial Intelligence for Match Analysis in Association Football
Bruno Marques and Dante Barone ........................................... 382

Using Machine Learning Algorithms to Predict the Likelihood of Recurrent Falls in Older Adults
Leeanne Lindsay, Sonya Coleman, Brian Taylor, Dermot Kerr, Anne Moorhead ................................................................. 392

Tree Based Clustering On Large, High Dimensional Datasets?
Lee A. Carraher, Sayantan Dey, and Philip A. Wilsey ............................ 397

Linear complexity algorithms for high dimensional SVM and regression problems with smart sparse regularization
Vadim Mottl, Olga Krasotkina, Valentina Sulimova, Alexey Morozov, Ilya Pugach, Alexander Tatarchuk .................................................. 412

Authors Index
15th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2019)

New York, New York, USA
20-25 July 2019

Volume 2 of 2

Editors:

Petra Perner

ISBN: 978-1-5108-9516-4
# Table of Content

Novel approach for Label Disambiguation via Deep Learning  
*Ticiana L. Coelho da Silva, Natanael da Silva Araújú, Josè Antônio F. de Macedo, David Araújo, Felipe Melo Soares, Paulo A. L. Rego, and Aloisio Vieira Lira Neto* .......................................................... 431

On combining dynamic selection, sampling, and pool generators for credit scoring  
*Leopoldo Melo Junior, Franco Maria Nardini, Chiara Renso, and Jose Antonio Macedo* .......................................................... 443

Comparative Analysis of Features Selection Techniques for Classification in Healthcare  
*Shruti Kaushik, Abhinav Choudhury, Ashutosh Kumar Jatav, Nataraj Dasgupta, Sayee Natarajan, Larry A. Pickett, and Varun Dutt* .......................................................... 458

Ontology-based Neural Network for Energy Optimization in Data Centers  
*Zhan Li, Li Chen, Guan Wng, Feng Zeng, and Yun Zhang* .......................................................... 473

Semantic similarity estimation for domain specific data using BERT and other techniques  
*R. Prashanth* .......................................................... 488

Educational Data Mining: A Hybrid Approach to Predicting Academic Performance of Students  
*Marlta Nayara Clemente de Almeida Lima, Geovanne Oliveira Alves, Wedson Lino Soares, and Roberta Andrade de Araujo Fagundes* .......................................................... 500

On the Vapnik-Chervonenkis Dimension of Binary Tables  
*Roman Sizov and Dan A. Simovici* .......................................................... 515

A Streaming Analytics Language for Processing Cyber Data  
*Eric L. Goodman and Dirk Grunwald* .......................................................... 527

SumOpinions: Automatic Summarization of Reviews about Tourist Places  
*João Holanda Freires Junior, Joseé Antônio de Fernandes Macêdo, and Igo Ramalho Brilhante* .......................................................... 542

Support Vector Machine with Graphical Network Structures in Features  
*Wenging He, Grace Y. Yi, and Li-Pang Chen* .......................................................... 557

Expanding annotated data with informed labels for weak supervision  
*Eura Shin, Sam Berglin, Jacob Furst, and Daniela Raicu* .......................................................... 571
Anomaly Detection in Process Signals within Machine Learning and Data Augmentation Approach*
Sergey Gavrin, Damir Marzagulov, Alexander Zamyatin ........................................... 585

Dynamic Facial Expression Recognition Based on Visual Rhythms and Motion History Images
Jadisha Yarif Ramirez Cornejo and Helio Pedrini ..................................................... 599

Data Reduction Effects on Classification Accuracy and Training Time
Reham M. Alamro and Abdou S. Youssef ................................................................. 614

Data Mining Framework to Derive Measures for Road Safety
Katherina Meißner and Julia Rieck ................................................................. 625

Failure Analysis on Multivariate Time-series Data given Uncertain Labels
Hao Huang, Shinjae Yoo, and Yunwen Xu .......................................................... 640

Feature Evaluation for Project Preferences Representation in GitHub
Thaciana Cerqueira, Leandro Marinho, and Franklin Ramalho ............................... 655

NOV-mRSI: A Novel Algorithm for Minimal Rare Significance Itemsets Mining
NOT Satisfy the Downward Closure Property in Transactional Databases
Huan Phan and Bac Le ................................................................. 669

Deep Convolutional Multi-Stream Network Detection System Applied to Fall Identification in Video Sequences
Sarah A. Carneiro, Gabriel P. da Silva, Guilherme V. Leite, Ricardo Moreno, Silvio Jamil F. Guimarães, and Helio Pedrini ........................................... 681

Edge Detection Using K-means and Run-Based Two-Scan Connected Components labeling adapted for grayscale images
Chen Avni and Maya Herman ................................................................. 696

The Representation of Wave Trend Direction in Stock Price Time Series via RNN
Yi Wei and Vipin Chaudhary ................................................................. 704

Two-Sample Tests for High Dimensional Distributions Using One-Dimensional Random Projections
Yibo Zhao, Jin Cao, and Ran He ................................................................. 718

The Computing Algorithm of Barrier Tree Based on the Basin Hopping Graph in RNA Structure
Zhendong Liu, Gang Li, and Patrick Wang ............................................................. 733

Hyperspectral Remote Temperature Estimation Robust to Atmospheric Variation with Fourier Transform Infrared and Deep Learning
Sungho Kim, Jungho Kim, Jinyong Lee, and Junmo Ahn .................................. 743
Binary Sparse Dynamic Time Warping
   Youngha Hwang and Saul B. Gelfand ............................................. 748

Applications of Statistical and Machine Learning Time-Series Methods for Predicting Internet Usage
   Naveksha Sood, Shruti Kaushik, Aditya Nigam, Sriram Kailasam, Dileep D. and Varun Dutt ................................................................. 760

A “Sentimental Analysis” of Computer-Generated Papers
   Dan Lemberg and Zeev Volkovich .................................................... 775

Lifetime reward optimization for online ads experiment
   Liang Dai and Ram Akella ............................................................ 782

Hierarchical Deep-Fusion Learning Framework for Lung Nodule Classification
   Kazim Sekeroglu, Omer Muhammet Soysal, and Xin Li ........................... 792

Object Detection and Classification in Low Resolution Underwater Imagery Using Deep Neural Network and Motion-Inspired Features
   Ahsan Jalal and Ahmad Salman .................................................... 805

Modeling and Simulation of UEFA Champions League
   Saman Muthukumarana and Ninad Khargonkar .................................. 820

Stochastic Optimization of Plain Convolutional Neural Networks with Simple Methods
   Yahia Saeed Assiri ................................................................. 833

Annotating Biodiversity Spreadsheets through 5W1H based on Machine Learning
   Ivelize R. Bernardo, Helio Pedrini, and André Santanché .......................... 845

Topological Structure of Linear Manifold Clustering
   Artyom Diky and Robert M. Haralick ................................................. 860

Wheat Ear Detection in RGB and Thermal Images Using Deep Neural Networks
   Zeljana Grbović, Marko Panić, Oskar Marko, Sanja Brdar, and Vladimir Crnojević ................................................................. 875

DeepSeed: A Deep Learning Methodology for Automated Soybean Seed Damage Classification
   Luigi Freitas Cruz, Priscila Tiemi Maeda Saito, and Pedro Henrique Bugatti .. 890

Dynamic Segmentation and Retrieval of Daily Life Activities from Multimodal Life-logs
   Rashmi Gupta and Cathal Gurrin .................................................... 901

Authors Index