

Manuel Graña
José Manuel López-Guede
Oier Etxaniz
Álvaro Herrero
Héctor Quintián
Emilio Corchado *Editors*

International Joint Conference SOCO'16-CISIS'16- ICEUTE'16

San Sebastián, Spain, October
19th–21st, 2016 Proceedings

Advances in Intelligent Systems and Computing

Volume 527

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagrais, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlm@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

Manuel Graña · José Manuel López-Guede
Oier Etxaniz · Álvaro Herrero
Héctor Quintián · Emilio Corchado
Editors

International Joint Conference SOCO'16- CISIS'16-ICEUTE'16

San Sebastián, Spain, October 19th–21st, 2016
Proceedings

Editors

Manuel Graña
Computational Intelligence Group
University of the Basque Country
(UPV/EHU)
San Sebastian
Spain

José Manuel López-Guede
University of the Basque Country
Vitoria
Spain

Oier Etxaniz
Computational Intelligence Group
University of the Basque Country
(UPV/EHU)
San Sebastian
Spain

Álvaro Herrero
Department of Civil Engineering
University of Burgos
Burgos
Spain

Héctor Quintián
University of Salamanca
Salamanca
Spain

Emilio Corchado
University of Salamanca
Salamanca
Spain

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-3-319-47363-5

ISBN 978-3-319-47364-2 (eBook)

DOI 10.1007/978-3-319-47364-2

Library of Congress Control Number: 2016953011

© Springer International Publishing AG 2017, corrected publication 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume of *Advances in Intelligent and Soft Computing* contains the accepted papers presented at SOCO 2016, CISIS 2016 and ICEUTE 2016, all conferences held in the beautiful and historic city of San Sebastián (Spain), in October 2016.

Soft computing represents a collection or a set of computational techniques in machine learning, computer science and some engineering disciplines, which investigate, simulate, and analyze very complex issues and phenomena.

After a thorough peer-review process, the 11th SOCO 2016 International Program Committee selected 45 papers which are published in these conference proceedings, and this represents an acceptance rate of 45%. In this relevant edition, a special emphasis was laid on the organization of special sessions. Two special sessions were organized related to relevant topics such as Optimization, Modeling and Control Systems by Soft Computing, and Soft Computing Methods in Manufacturing and Management Systems.

The aim of the 9th CISIS 2016 conference is to offer a meeting opportunity for academic and industry-related researchers belonging to the various, vast communities of computational intelligence, information security, and data mining. The need for intelligent, flexible behaviour by large, complex systems, especially in mission-critical domains, is intended to be the catalyst and the aggregation stimulus for the overall event.

After a thorough peer-review process, the CISIS 2016 International Program Committee selected 20 papers which are published in these conference proceedings achieving an acceptance rate of 40%.

In the case of 7th ICEUTE 2016, the International Program Committee selected 14 papers, which are published in these conference proceedings.

The selection of papers was extremely rigorous in order to maintain the high quality of the conference and we would like to thank the members of the program committees for their hard work in the reviewing process. This is a crucial process to the creation of a high-standard conference and the SOCO, CISIS and ICEUTE conferences would not exist without their help.

SOCO'16, CISIS'16, and ICEUTE'16 enjoyed outstanding keynote speeches by distinguished guest speakers: Prof. Ajith Abraham—MIR Labs (USA), Prof. Michal Wozniak—Wroclaw University of Technology (Poland), Carlos Toro—Vicomtech (Spain), and Sebastian Rios University of Chile (Chile).

SOCO'16 has teamed up with the *Journal of Applied Logic* (ELSEVIER) for a suite of special issue including selected papers from SOCO'16.

For this CISIS'16 special edition, as a follow-up of the conference, we anticipate further publication of selected papers in a special issue of the prestigious *Logic Journal of the IGPL* Published by Oxford Journals.

Particular thanks go as well to the conference main sponsors, COESI, IEEE Systems, Man and Cybernetics-Spanish Chapter, AEPIA, The International Federation for

Computational Logic, Python San Sebastián Society, University of Basque Country, who jointly contributed in an active and constructive manner to the success of this initiative. We want also to extend our warm gratitude to all the special sessions chairs for their continuing support to the SOCO, CISIS and ICEUTE series of conferences.

We would like to thank all the special session organizers, contributing authors, as well as the members of the program committees and the local organizing committee for their hard and highly valuable work. Their work has helped to contribute to the success of the SOCO 2016, CISIS 2016 and ICEUTE 2016 events.

October 2016

The editors

Manuel Graña
José Manuel López-Guede
Oier Etxaniz
Álvaro Herrero
Héctor Quintián
Emilio Corchado

SOCO 2016

Organization

General Chairs

Manuel Graña University of Basque Country, Spain
Emilio Corchado University of Salamanca, Spain

International Advisory Committee

Ashraf Saad Armstrong Atlantic State University, USA
Amy Neustein Linguistic Technology Systems, USA
Ajith Abraham Machine Intelligence Research Labs -MIR Labs,
 Europe
Jon G. Hall The Open University, UK
Paulo Novais Universidade do Minho, Portugal
Amparo Alonso Betanzos President Spanish Association for Artificial Intelligence
 (AEPIA), Spain
Michael Gabbay Kings College London, UK
Aditya Ghose University of Wollongong, Australia
Saeid Nahavandi Deakin University, Australia
Henri Pierrevall LIMOS UMR CNRS 6158 IFMA, France

Program Committee Chairs

Emilio Corchado University of Salamanca, Spain
Manuel Graña University of Basque Country, Spain
Álvaro Herrero University of Burgos, Spain

Program Committee

Ajith Abraham Machine Intelligence Research Labs (MIR Labs), USA
Cesar Analide University of Minho, Portugal
Ángel Arroyo University of Burgos, Spain
Mehmet Emin Aydin University of the West of England, UK
Antonio Bahamonde University of Oviedo at Gijón, Spain
Marius Balas Aurel Vlaicu University of Arad, Romania
Anna Bartkowiak University of Wrocław, Poland
Bruno Baruque University of Burgos, Spain
Rosa Basagoiti Mondragon University, Spain
Anna Burduk Wrocław University of Technology, Poland
Robert Burduk Wrocław University of Technology, Poland

Humberto Bustince	Public University of Navarre, Spain
Jose Luis Calvo-Rolle	University of A Coruña, Spain
Andre Carvalho	University of Saõ Paulo, Brazil
José Luis Casteleiro-Roca	University of Coruña, Spain
M. Chadli	University of Picardie Jules Verne, France
Camelia Chira	Technical University of Cluj-Napoca, Romania
Sung-Bae Cho	Yonsei University, Korea, Republic of
Emilio Corchado	University of Salamanca, Spain
Jose Alfredo Ferreira Costa	Universidade Federal do Rio Grande do Norte, Brazil
José Valente De Oliveira	Universidade do Algarve, Portugal
Jorge Díez	University of Oviedo, Spain
Enrique Domínguez	University of Malaga, Spain
Abraham Duarte	University Rey Juan Carlos, Spain
Richard Duro	University of Coruna, Spain
Wilfried Elmenreich	Alpen-Adria-Universität Klagenfurt, Austria
Florentina-Magda Enescu	University of Pitesti, Romania
Florentino Fdez-Riverola	University of Vigo, Spain
Leocadio G. Casado	University of Almeria, Spain
José Gámez	University of Castilla-La Mancha, Spain
Poonthalir Ganesan	PSG College of Technology, India
Esteban García-Cuesta	Universidad Europea de Madrid, Spain
Jorge García-Gutiérrez	University of Seville, Spain
Jose García-Rodríguez	University of Alicante, Spain
George Georgoulas	TEI of Epiruw, Greece
Juan Gomez Romero	University of Granada, Spain
Manuel Graña	University of País Vasco, Spain
David Griol	Universidad Carlos III de Madrid, Spain
Maciej Grzenda	Warsaw University of Technology, Poland
Álvaro Herrero	University of Burgos, Spain
Cesar Hervas	University of Cordoba, Spain
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Wei-Chiang Hong	Nanjing Tech University, Taiwan
Valeriu-Manuel Ionescu	University of Pitesti, Romania
Krzysztof Kalinowski	Silesian University of Technology, Poland
Fanny Klett	German Workforce ADL Partnership Laboratory, Germany
Przemyslaw Korytkowski	West Pomeranian University of Technology in Szczecin, Poland
Damian Krenczyk	Silesian University of Technology, Poland
Mitiche Lahcene	University of Djelfa, Algeria
Carlos Laorden	University of Deusto, Spain
José Manuel López Guede	University of País Vasco, Spain
Martin Macas	Czech Technical University in Prague, Czech Republic
Eleni Mangina	University College Dublin, Ireland
Francesco Marcelloni	University of Pisa, Italy
María Martínez Ballesteros	University of Seville, Spain
Francisco Martínez-Álvarez	University Pablo de Olavide, Spain

Oliviu Matei	North University of Baia Mare, Romania
Manuel Mejia-Lavalle	Cenidet, Mexico
José M. Molina	University Carlos III de Madrid, Spain
María N. Moreno García	University of Salamanca, Spain
Mohamed Mostafa Fouad	Arab Academy for Science, Technology, and Maritime Transport, Egypt
Dimitris Mourtzis	University of Patras, Greece
Luís Nunes	Instituto Universitário de Lisboa, Portugal
Enrique Onieva	University of Deusto, Spain
Marcin Paprzycki	IBS PAN and WSM, Poland
Daniela Perdukova	Technical Univerzity of Kosice, Slovakia
Carlos Pereira	ISEC, Portugal
Henri Pierreval	LIMOS-IFMA, France
Stefano Pizzuti	Energy New technologies and sustainable Economic development Agency (ENEA), Italy
Jiri Pospichal	University of SS. Cyril and Methodius, Slovakia
Dilip Pratihari	Indian Institute of Technology Kharagpur, India
Héctor Quintián	University of Salamanca, Spain
Luis Paulo Reis	University of Minho, Portugal
Dragan Simic	University of Novi Sad, Serbia
Georgios Ch. Sirakoulis	Democritus University of Thrace, Greece
Aureli Soria-Frisch	Starlab Barcelona, S.L., Spain
Rui Sousa	University of Minho, Portugal
Alicia Troncoso	Universidad Pablo de Olavide, Spain
Zita Vale	GECAD - ISEP/IPP, Portugal
M ^a Belén Vaquerizo	University of Burgos, Spain
Sebastián Ventura	University of Cordoba, Spain
José Ramón Villar	University of Oviedo, Spain
Eva Volna	Univerzity of Ostrava, Czech Republic
Michał Wozniak	Wroclaw University of Technology, Poland
Daniela Zaharie	West University of Timisoara, Romania
Cui Zhihua	Taiyuan University of Science and Technology, China
Urko Zurutuza	Mondragon University, Spain

Special Sessions

Optimization, Modeling and Control Systems by Soft Computing

Program Committee

Héctor Alaiz Moreton	University of Leon, Spain
José Luis Calvo Rolle	University of A Coruña, Spain
Jose Luis Casteleiro Roca	University of A Coruña, Spain
Luis Alfonso Fernández Serantes	FH-Joanneum University of Applied Sciences, Austria
Eloy Irigoyen	University of País Vasco, Spain

Esteban Jove Pérez	University of Salamanca, Spain
Maria Del Carmen Meizoso López	University of A Coruña, Spain
Andrés José Piñón Pazos	University of A Coruña, Spain
Matilde Santos	Complutense University of Madrid, Spain

Soft Computing Methods in Manufacturing and Management Systems

Program Committee

Maria Baron-Puda	University of Bielsko-Biala, Poland
Grzegorz Bocewicz	Koszalin University of Technology, Poland
Anna Burduk	Wroclaw University of Technology, Poland
Grzegorz Cwikla	Silesian University of Technology, Poland
Reggie Davidrajuh	University of Stavanger, Norway
Paul Eric Dossou	Icam, France
Laszlo Dudas	University of Miskolc, Hungary
Franjo Jović	Josip Juraj Strossmayer University of Osijek, Croatia
Krzysztof Kalinowski	Silesian University of Technology, Poland
Damian Krenczyk	Silesian University of Technology, Poland
Ivan Kuric	University of Žilina, Slovakia
Hongze Ma	KONE, Finland
Dariusz Plinta	University of Bielsko-Biala, Poland
Sebastian Saniuk	University of Zielona Gora, Poland
Bożena Skołod	Silesian University of Technology, Poland
Remus Zagan	Constanta Maritime University, Romania
Marcin Zemczak	University of Bielsko-Biala, Poland

Organising Committee

Manuel Graña	University of Basque Country, Spain
Oier Etxaniz	University of Basque Country, Spain
José Manuel López-Guede	University of Basque Country, Spain
Emilio Corchado	University of Salamanca, Spain
Álvaro Herrero	University of Burgos, Spain
Héctor Quintián	University of Salamanca, Spain



Enrique De La Hoz	University of Alcalá, Spain
Raúl Durán	University of Alcalá, Spain
Jesús Díaz-Verdejo	University of Granada, Spain
José Ferrer	Universitat de les Illes Balears, Spain
Amparo Fuster-Sabater	Institute of Applied Physics (CSIC), Spain
Leocadio G. Casado	University of Almeria, Spain
Debasis Giri	Haldia Institute of Technology, India
Ana I. González-Tablas	University Carlos III of Madrid, Spain
Manuel Grana	University of Basque Country, Spain
Juan Pedro Hecht	University of Buenos Aires, Argentina
Luis Hernandez Encinas	Institute of Physical and Information Technologies (ITEFI), Spain
José Luis Imana	Complutense University of Madrid, Spain
Gustavo Isaza	University of Caldas, Colombia
Tomasz Kajdanowicz	Wroclaw University of Technology, Poland
Wojciech Kmiecik	Wroclaw University of Technology, Poland
Zuzana Kominkova Oplatkova	Tomas Bata University in Zlin, Czech Republic
Bartosz Krawczyk	Wroclaw University Of Technology, Poland
Carlos Laorden	University of Deusto, Spain
Wenjian Luo	University of Science and Technology of China, China
José M. Molina	University Carlos III of Madrid, Spain
Guillermo Morales-Luna	CINVESTAV-IPN, Mexico
Enrique Onieva	University of Deusto, Spain
José A. Onieva	University of Malaga, Spain
Gianluca Papaleo	CNR-IEIIT, Italy
Alberto Peinado	University of Malaga, Spain
Carlos Pereira	ISEC, Portugal
Araceli Queiruga Dios	University of Salamanca, Spain
Héctor Quintián	University of Salamanca, Spain
Gerardo Rodriguez Sanchez	University of Salamanca, Spain
Luis Enrique Sanchez Crespo	University of the Armed Forces (ESPE-L), Ecuador
Gerald Schaefer	Loughborough University, UK
Hugo Scolnik	University of Buenos Aires, Argentina
Roman Senkerik	Tomas Bata University in Zlín, Czech Republic
Sorin Stratulat	Université de Lorraine, France
Javier Sánchez-Monedero	University Loyola Andalucía, Spain
Juan Tena	University of Valladolid, Spain
Antonio J. Tomeu-Hardasmal	University of Cádiz, Spain
Fernando Tricas	University of Zaragoza, Spain
Eva Volna	University of Ostrava, Czech Republic
Antonio Zamora Gómez	University of Alicante, Spain
Rodolfo Zunino	University of Genoa, Italy
Urko Zurutuza	Mondragon University, Spain

Special Sessions

Security in Wireless Networks: Mathematical Algorithms and Models

Program Committee

Raúl Durán Díaz	University of Alcalá, Spain
Luis Hernández Encinas	Institute of Physical and Information Technologies (ITEFI), Spain
Ángel Martín Del Rey	University of Salamanca, Spain
Agustín Martín Muñoz	Information Security Institute (CSIC), Spain
Alberto Peinado Domínguez	University of Malaga, Spain
Araceli Queiruga Dios	University of Salamanca, Spain
Gerardo Rodríguez Sánchez	University of Salamanca, Spain

Organising Committee

Manuel Graña	University of Basque Country, Spain
Oier Etxaniz	University of Basque Country, Spain
José Manuel López-Guede	University of Basque Country, Spain
Emilio Corchado	University of Salamanca, Spain
Álvaro Herrero	University of Burgos, Spain
Héctor Quintián	University of Salamanca, Spain

International Conference on



*Computational Intelligence in
Security for Information Systems*

ICEUTE 2016

Organization

General Chairs

Manuel Graña University of Basque Country, Spain
Emilio Corchado University of Salamanca, Spain

Program Committee Chairs

Emilio Corchado University of Salamanca, Spain
Manuel Graña University of Basque Country, Spain
Álvaro Herrero University of Burgos, Spain
José López Guede University of Basque Country, Spain

Program Committee

Estibaliz Apiñaniz- University of Basque Country, Spain
 Fernandez de Larrinoa
Karmele Artano-Perez University of Basque Country, Spain
Ruperta Delgado-Tercero University of Basque Country, Spain
Manuel Graña University of Basque Country, Spain
José Manuel López Guede University of Basque Country, Spain
Pilar Martínez-Blanco University of Basque Country, Spain
Amaia Mesanza-Moraza University of Basque Country, Spain
Paulo Moura Oliveira UTAD University, Portugal
Paulo Novais University of Minho, Portugal
Iñaki Ochoade Eribe University of Basque Country, Spain
Eduardo Solteiro Pires UTAD University, Portugal
Inmaculada Tazo University of Basque Country, Spain

Organising Committee

Manuel Graña University of Basque Country, Spain
Oier Etxaniz University of Basque Country, Spain
José Manuel López-Guede University of Basque Country, Spain
Emilio Corchado University of Salamanca, Spain
Álvaro Herrero University of Burgos, Spain
Héctor Quintián University of Salamanca, Spain



Contents

SOCO 2016: Classification

Predicting 30-Day Emergency Readmission Risk	3
<i>Arkaitz Artetxe, Andoni Beristain, Manuel Graña, and Ariadna Besga</i>	
Use of Support Vector Machines and Neural Networks to Assess Boar Sperm Viability	13
<i>Lidia Sánchez, Héctor Quintian, Javier Alfonso-Cendón, Hilde Pérez, and Emilio Corchado</i>	
Learning Fuzzy Models with a SAX-based Partitioning for Simulated Seizure Recognition.	20
<i>Paula Vergara, José Ramón Villar, Enrique de la Cal, Manuel Menéndez, and Javier Sedano</i>	
Real Prediction of Elder People Abnormal Situations at Home	31
<i>Aitor Moreno-Fernandez-de-Leceta, Jose Manuel Lopez-Guede, Manuel Graña, and Juan Carlos Cantera</i>	

SOCO 2016: Machine Learning

Assisting the Diagnosis of Neurodegenerative Disorders Using Principal Component Analysis and TensorFlow	43
<i>Fermín Segovia, Marcelo García-Pérez, Juan Manuel Górriz, Javier Ramírez, and Francisco Jesús Martínez-Murcia</i>	
Cyclone Performance Prediction Using Linear Regression Techniques	53
<i>Marina Corral Bobadilla, Roberto Fernandez Martinez, Rubén Lostado Lorza, Fátima Somovilla Gomez, and Eliseo P. Vergara Gonzalez</i>	
Time Analysis of Air Pollution in a Spanish Region Through k -means	63
<i>Ángel Arroyo, Verónica Tricio, Álvaro Herrero, and Emilio Corchado</i>	
Using Non-invasive Wearables for Detecting Emotions with Intelligent Agents	73
<i>Jaime Andres Rincon, Angelo Costa, Paulo Novais, Vicente Julian, and Carlos Carrascosa</i>	
Impulse Noise Detection in OFDM Communication System Using Machine Learning Ensemble Algorithms	85
<i>Ali N. Hasan and Thokozani Shongwe</i>	

SOCO 2016: Soft Computing Applications

A Hybrid Method for Optimizing Shopping Lists Oriented to Retail Store Costumers 95
Santiago Porras and Bruno Baruque

Estimation of Daily Global Horizontal Irradiation Using Extreme Gradient Boosting Machines 105
Ruben Urraca, Javier Antonanzas, Fernando Antonanzas-Torres, and Francisco Javier Martinez-de-Pison

The Control of the Output Power Gas Temperature at the Heat Exchanger . . . 114
Martin Pieš, Blanka Filipová, and Pavel Nevřiva

Industrial Cyber-Physical Systems in Textile Engineering. 126
Juan Bullón Pérez, Angélica González Arrieta, Ascensión Hernández Encinas, and Araceli Queiruga-Dios

Optimal Scheduling of Joint Wind-Thermal Systems 136
Rui Laia, Hugo M.I. Pousinho, Rui Melício, and Victor M.F. Mendes

ANN Based Model of PV Modules 147
Jose Manuel Lopez-Guede, Jose Antonio Ramos-Hernanz, Manuel Graña, and Valeriu Ionescu

SCADA Network System for the Monitoring and Control of an Electrical Installation Supplied by a Hydro-Generator 156
Florentina-Magda Enescu, Cosmin Ştirbu, and Valeriu Ionescu

CO₂ and Idling Emission Estimation for Vehicle Routing Problem with Mid Way Halts 167
Ganesan Poonthalir, R. Nadarajan, and S. Geetha

Agent-Based Spatial Dynamic Modeling of Opinion Propagation Exploring Delaying Conditions to Achieve Homogeneity 177
Leire Ozaeta and Manuel Graña

SOCO 2016: Genetic Algorithms

Coevolutionary Workflow Scheduling in a Dynamic Cloud Environment 189
Denis Nasonov, Mikhail Melnik, and Anton Radice

Searching Parsimonious Solutions with GA-PARSIMONY and XGBoost in High-Dimensional Databases 201
Francisco Javier Martinez-de-Pison, Esteban Fraile-Garcia, Javier Ferreiro-Cabello, Rubén Gonzalez, and Alpha Pernia

A K-means Based Genetic Algorithm for Data Clustering 211
Clara Pizzuti and Nicola Procopio

Improvement in the Process of Designing a New Artificial Human
 Intervertebral Lumbar Disc Combining Soft Computing Techniques
 and the Finite Element Method 223
*Rubén Lostado Lorza, Fátima Somovilla Gomez,
 Roberto Fernandez Martinez, Ruben Escribano Garcia,
 and Marina Corral Bobadilla*

SOCO 2016: Image and Video Analysis

Object Recognition by Machine Vision System of Inspection Line 235
Ondrej Petrtyl and Pavel Brandstetter

Pixel Features for Self-organizing Map Based Detection of Foreground
 Objects in Dynamic Environments 247
*Miguel A. Molina-Cabello, Ezequiel López-Rubio,
 Rafael Marcos Luque-Baena, Enrique Domínguez,
 and Esteban J. Palomo*

Reliable Workspace Monitoring in Safe Human-Robot Environment 256
*Amine Abou Moughlbay, Héctor Herrero, Raquel Pacheco,
 Jose Luis Outón, and Damien Sallé*

Forecasting Store Foot Traffic Using Facial Recognition, Time Series
 and Support Vector Machines 267
*Paulo Cortez, Luís Miguel Matos, Pedro José Pereira, Nuno Santos,
 and Duarte Duque*

**SOCO 2016: Special Session on Optimization, Modeling
 and Control Systems by Soft Computing**

Using GPUs to Speed up a Tomographic Reconstructor Based on
 Machine Learning 279
*Carlos González-Gutiérrez, Jesús Daniel Santos-Rodríguez,
 Ramón Ángel Fernández Díaz, Jose Luis Calvo Rolle,
 Nieves Roqueñi Gutiérrez, and Francisco Javier de Cos Juez*

An Intelligent Model for Bispectral Index (BIS) in Patients Undergoing
 General Anesthesia 290
*José Luis Casteleiro-Roca, Juan Albino Méndez Pérez,
 José Antonio Rebozo-Morales, Francisco Javier de Cos Juez,
 Francisco Javier Pérez-Castelo, and José Luis Calvo-Rolle*

Detection of Stress Level and Phases by Advanced Physiological Signal Processing Based on Fuzzy Logic	301
<i>Unai Zalabarría, Eloy Irigoyen, Raquel Martínez, and Asier Salazar-Ramírez</i>	
Reinforcement Learning for Hand Grasp with Surface Multi-field Neuroprostheses	313
<i>Eukene Imatz-Ojanguren, Eloy Irigoyen, and Thierry Keller</i>	
Fuzzy Candlesticks Forecasting Using Pattern Recognition for Stock Markets	323
<i>Rodrigo Naranjo and Matilde Santos</i>	
Analysing Concentrating Photovoltaics Technology Through the Use of Emerging Pattern Mining	334
<i>A.M. García-Vico, J. Montes, J. Aguilera, C.J. Carmona, and M.J. del Jesus</i>	
Mobile Wireless System for Outdoor Air Quality Monitoring	345
<i>Anton Koval and Eloy Irigoyen</i>	
SOCO 2016: Special Session on Soft Computing Methods in Manufacturing and Management Systems	
ANN-Based Hybrid Algorithm Supporting Composition Control of Casting Slip in Manufacture of Ceramic Insulators	357
<i>Arkadiusz Kowalski and Maria Rosienkiewicz</i>	
Genetic Algorithm Adoption to Transport Task Optimization	366
<i>Anna Burduk and Kamil Musiał</i>	
Detecting Existence of Cycles in Petri Nets: An Algorithm that Computes Non-redundant (Nonzero) Parts of Sparse Adjacency Matrix.	376
<i>Reggie Davidrajuh</i>	
An Instance Generator for the Multi-Objective 3D Packing Problem	386
<i>Yanira González, Gara Miranda, and Coromoto León</i>	
Solving Repetitive Production Planning Problems. An Approach Based on Activity-oriented Petri Nets	397
<i>Bożena Skołud, Damian Krenczyk, and Reggie Davidrajuh</i>	
The Concept of Ant Colony Algorithm for Scheduling of Flexible Manufacturing Systems	408
<i>Krzysztof Kalinowski and Bożena Skołud</i>	

Multistage Sequencing System for Complex Mixed-Model Assembly Problems	416
<i>Marcin Zembczak and Bożena Skołud</i>	
Robustness of Schedules Obtained Using the Tabu Search Algorithm Based on the Average Slack Method	426
<i>Iwona Paprocka, Aleksander Gwiazda, and Magdalena Bączkiewicz</i>	
Heterogeneous Fleet Vehicle Routing and Scheduling Subject to Mesh-Like Route Layout Constraints	435
<i>Grzegorz Bocewicz, Zbigniew Banaszak, and Damian Krenczyk</i>	
Application of the Hybrid - Multi Objective Immune Algorithm for Obtaining the Robustness of Schedules	445
<i>Iwona Paprocka, Aleksander Gwiazda, and Magdalena Bączkiewicz</i>	
Outperforming Genetic Algorithm with a Brute Force Approach Based on Activity-Oriented Petri Nets	454
<i>Reggie Davidrajuh</i>	
Integration of Manufacturing Functions for SME. Holonic-Based Approach . . .	464
<i>Bożena Skołud, Damian Krenczyk, Krzysztof Kalinowski, Grzegorz Ówikła, and Cezary Grabowik</i>	
CISIS 2016: Applications of Intelligent Methods for Security	
Feel Me Flow: A Review of Control-Flow Integrity Methods for User and Kernel Space	477
<i>Irene Díez-Franco and Igor Santos</i>	
A Secure Mobile Platform for Intelligent Transportation Systems	487
<i>Alexandra Rivero-García, Iván Santos-González, and Pino Caballero-Gil</i>	
Learning Deep Wavelet Networks for Recognition System of Arabic Words . . .	498
<i>Amira Bouallégué, Salima Hassairi, Ridha Ejbali, and Mourad Zaied</i>	
Intrusion Detection with Neural Networks Based on Knowledge Extraction by Decision Tree.	508
<i>César Guevara, Matilde Santos, and Victoria López</i>	
Using Spritz as a Password-Based Key Derivation Function	518
<i>Rafael Álvarez and Antonio Zamora</i>	
A Multiresolution Approach for Blind Watermarking of 3D Meshes Using Spiral Scanning Method	526
<i>Ikbel Sayahi, Akram Elkefi, and Chokri Ben Amar</i>	

Data Is Flowing in the Wind: A Review of Data-Flow Integrity Methods to Overcome Non-Control-Data Attacks. 536
Irene Díez-Franco and Igor Santos

CISIS 2016: Infrastructure and Network Security

Towards a Secure Two-Stage Supply Chain Network: A Transportation-Cost Approach. 547
Camelia-M. Pinteá, Anisoara Calinescu, Petrica C. Pop, and Cosmin Sabo

The HTTP Content Segmentation Method Combined with AdaBoost Classifier for Web-Layer Anomaly Detection System. 555
Rafał Kozik and Michał Choraś

Cluster Forests Based Fuzzy C-Means for Data Clustering 564
Abdelkarim Ben Ayed, Mohamed Ben Halima, and Adel M. Alimi

Neural Visualization of Android Malware Families 574
Alejandro González, Álvaro Herrero, and Emilio Corchado

Time Series Data Mining for Network Service Dependency Analysis. 584
Mona Lange and Ralf Möller

Security Analysis of a New Bit-Level Permutation Image Encryption Algorithm 595
Adrian-Viorel Diaconu, Valeriu Ionescu, and Jose Manuel Lopez-Guede

Characterization of Android Malware Families by a Reduced Set of Static Features 607
Javier Sedano, Camelia Chira, Silvia González, Álvaro Herrero, Emilio Corchado, and José Ramón Villar

CISIS 2016: Security in Wireless Networks: Mathematical Algorithms and Models

A Comparison of Computer-Based Technologies Suitable for Cryptographic Attacks 621
Victor Gayoso Martínez, Luis Hernández Encinas, Agustín Martín Muñoz, Óscar Martínez-Graullera, and Javier Villazón-Terrazas

Cryptanalysis of a Key Authentication Scheme Based on the Chinese Remainder Theorem and Discrete Logarithms. 631
Alberto Peinado

A SCIRS Model for Malware Propagation in Wireless Networks 638
*Angel Martín del Rey, José Diamantino Hernández Guillén,
and Gerardo Rodríguez Sánchez*

Malware Propagation Models in Wireless Sensor Networks: A Review 648
*Araceli Queiruga-Dios, Ascensión Hernández Encinas,
Jesus Martín-Vaquero, and Luis Hernández Encinas*

A Study on the Performance of Secure Elliptic Curves for
Cryptographic Purposes 658
*Raúl Durán Díaz, Victor Gayoso Martínez, Luis Hernández Encinas,
and Agustín Martín Muñoz*

A SEIS Model for Propagation of Random Jamming Attacks in Wireless
Sensor Networks 668
Miguel López, Alberto Peinado, and Andrés Ortiz

ICEUTE 2016

Educational Big Data Mining: How to Enhance Virtual
Learning Environments 681
Pietro Ducange, Riccardo Pecori, Luigi Sarti, and Massimo Vecchio

Application of the PBL Methodology at the B.Sc. in Industrial Electronics
and Automation Engineering 691
*Isidro Calvo, Jeronimo Quesada, Itziar Cabanes,
and Oscar Barambones*

Coordination and Cooperative Learning in Engineering Studies. 701
*Karmele Artano-Pérez, Aitor Bastarrika-Izagirre,
Ruperta Delgado-Tercero, Pilar Martínez-Blanco,
and Amaia Mesanza-Moraza*

Welcome Program for First Year Students at the Faculty of Engineering
of Vitoria-Gasteiz. Soft Skills 711
*Estibaliz Apiñaniz-Fernandez de Larrinoa, Javier Sancho-Saiz,
Amaia Mesanza-Moraza, Ruperta Delgado-Tercero, I. Tazo-Herrán,
J.A. Ramos-Hernanz, J.I. Ochoa de Eribe-Vázquez, J.M. Lopez-Guede,
E. Zulueta-Guerrero, and J. Díaz de Argandoña-González*

Revisiting the Simulated Annealing Algorithm from a Teaching Perspective . . . 718
*Paulo B. de Moura Oliveira, Eduardo J. Solteiro Pires,
and Paulo Novais*

Minecraft as a Tool in the Teaching-Learning Process of the Fundamental Elements of Circulation in Architecture	728
<i>Maria Do Carmo López Méndez, Angélica González Arrieta, Marián Queiruga Dios, Ascensión Hernández Encinas, and Araceli Queiruga-Dios</i>	
Skills Development of Professional Ethics in Engineering Degrees in the European Higher Education Area	736
<i>Lidia Sanchez, Javier Alfonso-Cendón, Hilde Pérez, Héctor Quintián, and Emilio Corchado</i>	
Expert System for Evaluating Teachers in E-Learning Systems	741
<i>Bogdan Walek and Radim Farana</i>	
The Quadrotor Workshop in Science Week. Spread of Technical and Scientific Applications in Society	751
<i>Julian Estevez</i>	
Erasmus Innovative European Studies	758
<i>Jose Manuel Lopez-Guede, Erol Kurt, Necmi Altin, Manuel Graña, and Valeriu Ionescu</i>	
Study of Huffman Coding Performance in Linux and Windows 10 IoT for Different Frameworks	766
<i>Alexandru-Cătălin Petrini and Valeriu-Manuel Ionescu</i>	
Virtualization Laboratory for Computer Networks at Undergraduate Level . . .	776
<i>Valeriu Manuel Ionescu and Alexandru-Cătălin Petrini</i>	
Using the Phone's Light Sensor to Detect the TV Video Stream	785
<i>Valeriu Manuel Ionescu, Cosmin Stirbu, and Florentina Magda Enescu</i>	
Comparing Google Cloud and Microsoft Azure Platforms for Undergraduate Laboratory Use.	795
<i>Valeriu Manuel Ionescu and Jose Manuel Lopez-Guede</i>	
Correction to: Optimal Scheduling of Joint Wind-Thermal Systems	C1
<i>Rui Laia, Hugo M. I. Pousinho, Rui Melicio, and Victor M. F. Mendes</i>	
Author Index	803

SOCO 2016: Classification

Predicting 30-Day Emergency Readmission Risk

Arkaitz Artetxe^{1,2(✉)}, Andoni Beristain^{1,2}, Manuel Graña²,
and Ariadna Besga³

¹ Vicomtech-IK4 Research Centre, Mikeletegi Pasealekua 57,
20009 San Sebastian, Spain

{aartetxe, aberistain}@vicomtech.org

² Computation Intelligence Group, Basque University (UPV/EHU),

P. Manuel Lardizabal 1, 20018 San Sebastian, Spain

³ Department of Internal Medicine,
Hospital Universitario de Alava, Vitoria, Spain

Abstract. *Objective:* Predicting Emergency Department (ED) readmissions is of great importance since it helps identifying patients requiring further post-discharge attention as well as reducing healthcare costs. It is becoming standard procedure to evaluate the risk of ED readmission within 30 days after discharge. *Methods.* Our dataset is stratified into four groups according to the Kaiser Permanente Risk Stratification Model. We deal with imbalanced data using different approaches for resampling. Feature selection is also addressed by a wrapper method which evaluates feature set importance by the performance of various classifiers trained on them. *Results.* We trained a model for each scenario and subpopulation, namely case management (CM), heart failure (HF), chronic obstructive pulmonary disease (COPD) and diabetes mellitus (DM). Using the full dataset we found that the best sensitivity is achieved by SVM using over-sampling methods (40.62 % sensitivity, 78.71 % specificity and 71.94 accuracy). *Conclusions.* Imbalance correction techniques allow to achieve better sensitivity performance, however the dataset has not enough positive cases, hindering the achievement of better prediction ability. The arbitrary definition of a threshold-based discretization for measurements which are inherently is an important drawback for the exploitation of the data, therefore a regression approach is considered as future work.

Keywords: Readmission risk · Imbalanced datasets · SVM · Classification

1 Introduction

The number of people aged over 65 is projected to grow from an estimated 524 million in 2010 to nearly 1.5 billion in 2050 worldwide [1]. This trend has a direct impact on the sustainability of health systems, in maintaining both public policies and the required budgets.

This growing population group represents an unprecedented challenge for health-care systems. In developed countries, older adults already account for 12 to 21 % of all ED visits and it is estimated that this will increase by around 34 % by 2030 [14].

Older patients have increasingly complex medical conditions in terms of their number of morbidities and other conditions, such as the number of medications they use, existence of geriatric syndromes, their degree of physical or mental disability, and the interplay of social factors influencing their condition [9]. Recent studies have shown that adults above 75 years of age have the highest rates of ED readmission, and the longest stays, demanding around 50 % more ancillary tests [15]. Notwithstanding the intense use of resources, these patients often leave the ED unsatisfied, and with poorer clinical outcomes, and higher rates of misdiagnosis and medication errors [16] compared to younger patients. Additionally, once they are discharged from the hospital, they have a high risk of adverse outcomes, such as functional worsening, ED readmission, hospitalization, death and institutionalization [17].

In this paper we present our recent work on ED readmission risk prediction. We utilize historic patient information, including demographic data, clinical characteristics or drug treatment information among others. Our work focuses on high risk patients (two higher strata) according to the Kaiser Permanente Risk Stratification Model [11]. This includes patients with prominence of specific organ disease (heart failure, chronic obstructive pulmonary disease and diabetes mellitus) and patients with high multi-morbidity. Predictive models are built for each of the stratified groups using different classifiers such as Support Vector Machine (SVM) and Random Forest. In order to deal with class imbalance and high dimensional feature space, different filtering techniques have been proposed during experimental approach.

The main contributions of this work are:

- We extend the work by Besga et al. [2] applying well-known machine learning techniques such as class balancing and feature selection in order to obtain better sensitivity.
- We compare two well established supervised classification algorithms, Random Forests and SVM, and analyze their performance in different scenarios.
- We make use of a wrapper feature selection method that maximizes the prediction ability while minimizes models' complexity.

The paper is organized as follows. In Sect. 2 we present some related works on predictive modelling for readmission risk estimation. In Sect. 3 we present the dataset as well as the methodological approach followed in order to build our models. Next, we describe the evaluation methodology and the experimental results. In Sect. 5 we discuss the conclusions and future work.

2 Related Work

Readmission risk modelling is a research topic that has been extensively studied in recent years. The main objective is usually to reduce readmission costs by identifying those patients with higher risk of coming back soon. Patients with higher risk can be followed-up after discharge, checking their health status by means of interventions such as phone calls, home visits or online monitoring, which are resource intensive. Predictive systems generally try to model the probability of unplanned readmission (or death) of a patient within a given time period.

In a recent work, Kansagara et al. [9] presented a systematic review of risk prediction models for hospital readmission. Many of the analyzed models target certain subpopulation with specific conditions or diseases such as Acute Myocardial Infarction (AMI) or heart failure (HF) while others embrace general population.

One of the most popular models that focus on general populations is LACE [3]. The LACE index is based on a model that predicts the risk of death or urgent readmission (within 30 days) after leaving the hospital. The algorithm used to build the model is commonly used in the literature (logistic regression analysis) and, according to the published results, the model has a high discriminative ability. The model uses information of 48 variables collected from 4812 patients from several Canadian hospitals.

A variant called LACE + [4] is an extension of the previous model that makes use of variables drawn from administrative data.

A similar approach is followed by Health Quality Ontario (HQP) with their system called HARP (Hospital Admission Risk Prediction) [10]. The system aims to determine the risk of patients in short and long term future hospitalizations. HARP defines two periods of 30 days and 15 months for which the model infers the probability of hospitalization, relying on several variables. From an initial set of variables of 4 different categories (demographic, feature community, disease and condition and meetings with the hospital system) the system identifies two sets of variables, a complex and a simpler one, with the most predictive variables. Using these sets of variables and a dataset containing approximately 382,000 episodes, two models for one month and 15 months are implemented. The models were developed using multivariate regression analysis. According to the committee of experts involved in the development of HARP, the most important metric was the sensitivity (i.e. the ability to detect hospitalizations). Regarding this metric, claimed results suggest that both simple and complex models achieve high sensitivity, although the complex model gets better results. The authors of this work suggest that the simple model could be a good substitute when certain hospitalization data is not available (e.g. to perform stratification outside the hospital).

A recent work by Yu et al. [5] presents an institution-specific readmission risk prediction framework. The idea beneath this approach is that most of the readmission prediction models have not sufficient accuracy due to differences between the patient characteristics of different hospitals. In this work an experimental study is performed, where a classification method (SVM) is applied as well as regression (Cox) analysis.

In [2] Besga et al. analyzed patients who attended Emergency Department of the Araba university Hospital (AUH) during June 2014. We exploit this dataset improving their results with further experiments.

3 Materials and Methods

The dataset, presented by Besga et al. in [2], is composed of 360 patients divided into four groups, namely: case management (CM), patients with chronic obstructive pulmonary disease (COPD), heart failure (HF) and Diabetes Mellitus (DM). For each patient a set of 97 variables were collected, divided into four main groups: (i) Sociodemographic

data and baseline status, (ii) Personal history, (iii) Reasons for consultation/Diagnoses made at ED and (iv) Regular medications and other treatments. Dataset contains missing values.

In order to build our model following a binary classification approach, the target variable was set to *readmitted/not readmitted*. Those patients returning to ED within 30 days after being discharged are considered readmitted (value = 1), otherwise are seen as not readmitted (value = 0).

It is noteworthy that one patient returning the first day and another returning the 30th are both considered as *readmitted*. On the other hand, a patient returning the 31th day is considered as *not readmitted*, while in practice underwent a readmission. We believe that having the number of days passed before readmission would have been much more meaningful for identification and would have permitted even identifying a more accurate prediction, including the predicted time for readmission.

All the tests were conducted using 10-fold cross-validation. The evaluation metrics that we have used are: sensitivity, specificity and accuracy. In order to avoid any random number generation bias, we have conducted 10 independent executions with different random generating seeds and averaged the results obtained.

Table 1. Distribution of variables by category

Variable	No. (%) of variables n = 96
Sociodemographic and baseline status	4 (4.2)
Personal history	43 (44.8)
Reasons for consultation	16 (16.7)
Regular medications	33 (34.3)

According to the data shown in Table 1 our dataset has a high dimensional feature space. In this scenario we have carried out some feature selection techniques. The goal is to find a feature subset that would reduce the complexity of the model, so that it would be easier to interpret by physicians, while improving the prediction performance and reducing overfitting.

We are going to use the following approaches: filter methods and wrapper methods. Filter algorithms are general preprocessing algorithms that do not assume the use of a specific classification method. Wrapper algorithms, in the other hand, “wrap” the feature selection around a specific classifier and select a subset of features based on the classifier’s accuracy using cross-validation [18]. Wrapper methods evaluate subsets of variables, that is, unlike filter methods, do not compute the worth of a single feature but the whole subset of features.

- **Filter method:** We have used Correlation-based Feature Selection (CBFS) method since it evaluates the usefulness of individual features for predicting the class along with the level of inter-correlation among them [19]. In this work we have used the implementation provided by Weka [8].

- **Wrapper method:** We have selected SVM as the specific classification algorithm and Area Under the Curve (AUC) as evaluation measure. Since an exhaustive search is impractical due to space dimensionality, we used heuristics, following a greedy stepwise approach. In this work we have used the implementation provided by Weka.

3.1 Support Vector Machine

Support vector machines (SVM) are supervised learning models which have been widely used in bioinformatics research and many other fields since their introduction in 1995 [7]. It is often defined as a non-probabilistic binary linear classifier, as it assigns new cases into one of two possible classes. In the readmission prediction problem, the model would predict whether a new case (the patient) will be readmitted within 30 days.

This algorithm is based on the idea that input vectors are non-linearly mapped into a very high dimensional space. In this new feature space it constructs a hyperplane which separates instances of both classes. Since there exist many decision hyperplanes that might classify the data, SVM tries to find the maximum-margin hyperplane, i.e. the one that represents the largest separation (margin) between the two classes.

In this work we have used the libSVM¹ implementation of the algorithm, which is the common implementation used for experimentation, and can be easily integrated to weka [8] using a wrapper. We have used a radial basis kernel function: $\exp(-\gamma*|u-v|^2)$ where $\gamma = 1/\text{num_features}$ and $C = 1$.

3.2 Random Forest

Random Forest [6] is a classifier consisting of multiple decision trees trained using randomly selected feature subspaces. This method builds multiple decision trees at training phase. In order to predict the class of a new instance, it is put down to each of these trees. Each tree gives a prediction (votes) and the class having most votes over all the trees of the forest will be selected. The algorithm uses the bagging method, i.e. each tree is trained using a random subset (with replacement) of the original dataset. In addition, each split uses a random subset of features.

One of the advantages of random forests is that generally they generalize better than decision trees, which tend to overfitting and naturally perform some feature selection. They can also be run on large datasets and can handle thousands of attributes without attribute deletion. In this work we have used Weka's implementation of the algorithm.

¹ <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

4 Results

In this section we analyze the prediction performance of different models on the emergency department short-time readmission dataset presented in [2]. As shown in Table 2 we have considered besides the original four subpopulations a fifth dataset that encompasses all of them.

Table 2. Comparative information about the subpopulations of the dataset

	Readmission within 30 days, no. (%) of patients		
	Overall no. of patients	No	Yes
	$n = 360$	n = 296 (82.2)	n = 64 (17.7)
Case management	94 (26.1)	73 (77.7)	21 (22.3)
Heart failure	70 (19.4)	62 (88.6)	8 (11.4)
Chronic obstructive pulmonary disease	80 (22.2)	64 (80)	16 (20)
Diabetes mellitus	116 (32.2)	97 (83.6)	19 (16.4)

4.1 Class Balancing

In readmission prediction analysis like in any other supervised classification problem, imbalanced class distribution leads to important performance evaluation issues and problems to achieve desired results. The underlying problem with imbalanced datasets is that classification algorithms are often biased towards the majority class and hence, there is a higher misclassification rate of the minority class instances (which are usually the most interesting ones from the practical point of view) [13].

As shown in Table 3, class imbalance is causing an accuracy paradox. If we just

Table 3. Confusion matrix of SVM on the diabetes mellitus dataset

		Predicted	
		Readmitted	Not readmitted
Actual	Readmitted	97	0
	Not readmitted	19	0

look at the accuracy of the model we get an 83.62 % although SVM just behaves as suing only the greater *a priori* probability to make the classification decision.

Resampling. There are several methods that can be used in order to tackle the class imbalance problem. Building a more balanced dataset is one of the most intuitive approaches. In our experiment we have used under-sampling as a preliminary approach and continued with an over-sampling using synthetic samples.