Editorial foreword

The present special issue is the outcome of an open call for papers under the broad theme of "Applications of Data Science: blending system design and engineering, advanced analytics and large-scale experimentation". This action is a further dissemination regarding the organization of the 5th International Conference on Real Time Intelligent Systems (RTIS'2020), which was held in Biarritz, France between 30th June and 3rd July 2020.

Data science has emerged as a topic encompassing machine learning and data mining, statistics, big data management and discovery science. Thus, this special issue focuses on applications, end-to-end systems and experience reporting rather than proposal of novel algorithms and model development methods. The emphasis is shifted towards the combination of: (i) advanced algorithmic solutions for data analytics, (ii) implementation and engineering techniques so that the resulted products are applied to address specific real-world problems, and (iii) experience reporting preferably through large-scale evaluation of working prototypes. The described systems should be either made publicly available or described in adequate detail so that third-parties can re-implement them with reasonable effort.

Fourteen papers were submitted, originated form several countries around the globe: Algeria, China, France, Germany, India, Oman, Pakistan, Saudi Arabia, South Korea, Tunisia, Turkey, UK, and USA. Each paper passed a rigorous review process by three independent reviewers. Finally, the undersigned guest editors selected six papers for publication. A brief presentation of these six papers follows.

The first paper is entitled "Bio-signal data sharing security through watermarking: A technical survey" and is co-authored by N. Sharma, A. Anand and A.K. Singh (NIT Patna, Bihar, India). Sensitive medical information is easily transmitted over the networks due to smart ICT-based healthcare systems. However, the crime of stealing healthcare data is increasing resulting in great financial loss. Researchers are developing various data hiding techniques for smart healthcare applications. In this paper, the authors first introduce various aspects of data hiding along with major properties, generic embedding and extraction process, as well as recent applications. This survey provides a comprehensive view on data hiding techniques, and their new trends for solving new challenges in real-world applications. In the sequel, the authors survey the various notable bio-signal based data hiding techniques. The summary of some notable techniques in terms of their objective, type of data hiding, methodology and database used, performance metrics, important features, and

limitations are also presented. Finally, the authors discuss research directions and promising areas for future research.

The title of the second paper is "Context-aware recommender system using trust network" and the authors are El Yebdri Zeyneb, Benslimane Sidi Mohammed, Lahfa Fedoua, Barhamgi Mahmoud and Benslimane Djamal (Abou Bekr University of Tlemcen, École Supérieure en Informatique at Sidi Bel Abbes and Claude Bernad Lyon1 University). Context-Aware Recommender Systems (CARS) improve traditional Recommender Systems (RS) in a wide array of domains and applications. However, CARS suffer from several inherent issues such as data sparsity and cold start. Incorporating trust into recommender systems can handle these issues. Trust-aware recommender systems (TARS) use information from social networks such as trust statements, which prove another valuable information source. This paper exploits the advantages of these two systems by incorporating both trust and context information. The authors propose a hybrid approach: Trust based Context aware Post Filtering Approach (TCPoFA). Their approach utilizes the relative average difference among the context on output of trust aware collaborative filtering by incorporating explicit and implicit trust information. They also use a confidence concept to remove non-confident users from the trust network before generating prediction. The experiments show that the proposed approach improves the standard RS on real world dataset.

The third paper is entitled "Popularity vs quality: Analyzing and predicting the success of highly rated crowdfunded projects on Amazon" and is co-authored by Vishal Sharma, Kyumin Lee and Curtis Dyreson (Utah State University and Worcester Polytechnic Institute). The popular online crowdfunding platforms provide a stage for innovators worldwide to bring ideas to reality. Despite the popularity and success of many projects on the platforms, it is yet to be determined whether successful projects always produce high quality products. The quality of crowdfunded products in the market has not been statistically and scientifically evaluated. To this end, the authors (i) compare crowdfunded products with traditional products in terms of their ratings on Amazon, (ii) analyze negative reviews of crowdfunded products, (iii) analyze characteristics of the successful projects and unsuccessful projects, and (iv) build machine learning models at three different stages, to predict high or low star ratings for a crowdfunded product. Their experiments show that crowdfunded products received lower ratings than traditional products. Their ensemble model effectively identifies which product will receive high star-ratings from customers on Amazon.

The title of the fourth paper is "Analysing environmental impact of large-scale events in public spaces with cross-domain multimodal data fusion" and is co-authored by Suparna De (University of Winchster), Wei Wang (Xi'an Jiaotong Liverpool University), Yuchao Zhou (University of Surrey), Charith Perera (Cardiff University), Klaus Moessner (Chemnitz University of Technology) and Mansour Naser Alraja (Dhofar University). In this study, the authors demonstrate how to quantify environmental implications of largescale events and traffic in public spaces, and identify specific city regions that are impacted. They develop an innovative data fusion framework that synthesizes the state-of-the-art techniques in extracting pollution episodes and detecting events from citizen-contributed, city-specific

messages on social media platforms. They further design a fusion pipeline for this cross-domain, multimodal data, which assesses the spatio-temporal impact of the extracted events on pollution levels within a city. Results of the analytics have great potential to benefit citizens and in particular, city authorities, who strive to optimize resources for better urban planning and traffic management.

The fifth paper is entitled "Designing and implementing a big data benchmark in a financial context application to a cash management use case" and is co-authored by Lilia Sfaxi and Mohamed Mehdi Ben Aissa (University of Carthage). This paper details the steps followed to benchmark a cash management platform of an investment bank using a generic benchmarking solution called BABEL. The authors highlight the modular design of BABEL, and present an evaluation methodology and best practices for its application on real world systems. The performance results for the cash management use case enables to define the right tradeoffs in terms of consistency and availability, in a way that respects the service level agreements defined by the clients. On the other hand, the authors show that the overhead caused by BABEL's integration with the platform at runtime is very negligible.

The title of the sixth paper is "A parallel text clustering method using Spark and hashing" and is co-authored by Mohamed Aymen Ben HajKacem, Chiheb-Eddine Ben N'cir and Nadia Essoussi (University of Tunis and University of Jeddah). The increasing growth of available textual data from web, social networks and open platforms have challenged the task clustering textual data. It becomes important to design scalable clustering method able to effectively organize huge amount of textual data into topics. In this context, the authors propose a new parallel text clustering method based on Spark framework and hashing. The proposed method deals simultaneously with the issue of clustering huge amount of documents and the issue of high dimensionality of textual data by respectively integrating the divide and conquer approach and implementing a new document hashing strategy. These two facts have shown an important improvement of scalability and a good approximation of clustering quality results. Experiments have shown the effectiveness of the proposed method compared to existing ones in terms of running time and clustering accuracy

We own gratitude to the referees who kindly participated in the effort and helped in delivering a special issue with interesting papers, and hopefully with papers that will demonstrate significant impact in the future. In particular, we would like to thank the following academicians:

Hong-Ning Dai, Macau University of Science and Technology, Macau
Laurent D'Orazio, University Rennes, France
Feiran Huang, Jinan University, China
Andreas Kosmatopoulos, CERTH, Greece
Georgia Kougka, Aristotle University of Thessaloniki, Greece
Dominique Laurent, University Cergy Pontoise, France
Elio Mansour, Université de Pau et des Pays de l'Adour, France
Mohamed Mousa, Suez Canal University, Egypt
Yongrui Qin, University of Huddersfield, UK
Amjad Rattrout, Arab American University, Palestine
Salma Sassi, University of Jendouba, Tunisia

M. Tanveer, Indian Institute of Technology Indore, India
Joe Tekli, Lebanese American University, Lebanon
Sang Won Yoon, Binghamton University, USA
Nadia Yacoubi-Ayadi, Institut Supérieur de Gestion de Tunis, Tunisia
Ruofei Zhang, Microsoft Corp, USA

Finally, thanks are due to the Editor-in-Chief of Computing journal, Professor Schahram Dustdar, for giving us the opportunity to run this project.
Guest Editors
Richard Chbeir (Université de Pau et des Pays de l'Adour, France)
Anastasios Gounaris (Aristotle University of Thessaloniki, Greece)
Yannis Manolopoulos (Open University of Cyprus, Cyprus)
Jolanta Mizera-Pietraszko (Military University of Land Forces, Poland)