Computing and Informatics, Vol. 41, 2022, 1–11, doi: 10.31577/cai_2022_1_1

GUEST EDITORIAL: SPECIAL ISSUE ON DATA SCIENCE APPROACHES AND APPLICATIONS

Nafaa Jabeur

Computer Science Department
German University of Technology in Oman
Muscat, Sultanate of Oman
e-mail: nafaa.jabeur@gutech.edu.om

Ansar Yasar

Transportation Research Institute
Hasselt University
Hasselt, Belgium
e-mail: ansar.yasar@uhasselt.be

Yehya Mohamad

Fraunhofer Institute for Applied Information Technology
Bonn, Germany
e-mail: yehya.mohamad@fit.fraunhofer.de

Michele Melchiori

Dipartimento di Ingegneria dell'Informazione
University of Brescia
Brescia, Italy
e-mail: michele.melchiori@unibs.it

This special issue is based on best papers selected from the 1st International Conference on Applied Data Science and Intelligence (ADSI 2021), the 1st International Conference on Recent Theories and Applications in Transportation and Mobility

(RTATM 2021), and the 1st International Conference on Informatics Revolution for Smarter Healthcare (IRSH 2021). The conferences, which were planned to take place on 14–15 October 2021 in Prague, Czech Republic, were held online on the same dates due to COVID restrictions. They attracted a large number of scientific papers. Based on a thorough peer review process as well as on the evaluations of the Guest Editors, about 20 % of the papers were selected for their significant contributions to the field of data science via innovative approaches and applications. We present in what follows a short abstract of each paper.

Paper 1: Using Machine Learning for Intrusion Detection Systems, by Quang-Vinh Dang [1]

The author analyzes the machine learning models that have been proposed in the recent years. He proposes new techniques to improve the performance of these models. In order to demonstrate the performance of the proposed techniques, he presents experimental results using real-world datasets. The results show an improvement of predictive accuracy of selected machine learning models.

Paper 2: Review of Smart Contracts for Cloud-Based Manufacturing, by Ahmad Hassan Afridi [2]

The paper conducts a scoping review of smart contracts for cloud manufacturing. It studies the best practices to start manufacturing activities supported by blockchain smart contracts. It explores some recent case-studies and concepts related to data extraction from digital libraries and online repositories. The paper answers four research questions:

1. What are some key architectural approaches is blockchain uses for manufacturing?
2. What kinds of smart contracts are available for cloud-based manufacturing?
3. What are some dApps available for cloud-based manufacturing?
4. What are some research issues/software development related to blockchain application to cloud-based manufacturing?

Paper 3: Topic Extraction in Social Networks, by Chaima Messaoudi, Zahia Guessoum and Lotfi BenRomdhane [3]

The authors address two main challenges; namely extracting data from Twitter and extracting topics from user tweets. In contrast with approaches based on data crawling and natural language processing techniques, they use K-means clustering and Latent Dirichlet Allocation to extract the prevalent topics from this dataset. The proposed solution is generic and can, thus, be reused by scientists to annotate any text collection.

Paper 4: Semantic Segmentation of Text Using Deep Learning, by Tiziano Lattisi, Davide Farina and Marco Ronchetti [4]

Given a text, can we segment it into semantically coherent sections in an automatic way? Can we detect the semantic boundaries, if we know how many

they are? Can we determine how many semantically distinct sections are in the text? To answer these questions, the authors use the Bidirectional Encoder Representation from Transformer (BERT) to analyze the text. They evaluate a function, called local incoherence, to show maxima at the points where a semantic boundary is detected. The results suggest that the proposed approach can be successfully applied.

Paper 5: Reducing the Effect of Imbalance in Text Classification Using SVD and GloVe with Ensemble and Deep Learning, by Tajbia Hossain, Humaira Zahin Mauni and Raqeebir Rab [5]

The authors address the topic of text classification. They particularly focus on the problem of text data imbalance. They propose several techniques and unique approaches to tackle this obstacle. They prepare four datasets of varying degrees of imbalance to conduct their experimentations. They prove that feature extraction techniques singular value decomposition (SVD) and GloVe are key to reducing the effect of imbalance in text classification, especially in ensemble and deep learning. Based on the results obtained, they propose a modified ensemble classifier that can classify imbalanced and balanced data alike.

Paper 6: Fault Diagnosis of Discrete-Event Systems from Abstract Observations, by Gianfranco Lamperti, Marina Zanella and Xiangfu Zhao [6]

Active systems (ASs) are a special class of (asynchronous) discrete event systems (DESs). Diagnosis engines are being used to find out possible faults based on the DES models as well as on a sequence of observations gathered during operation. These engines need to know what is observable in the behavior of the DES and what is not. The notion of observability, which serves this purpose, still needs considerable investigations. The authors present in this paper a generalized notion of observability, where an observation is abstract rather than concrete, since it is associated with a DES behavioral scenario rather than a single component transition. To support the online diagnosis engine, they perform an offline knowledge compilation. They generate a set of data structures, called watchers, which allow for the tracking of abstract observations.

Paper 7: Smart Water Management Using Intelligent Digital Twins, by Slim Zekri, Nafaa Jabeur and Hana Gharrad [7]

Providing and distributing fresh water to large communities is a major global concern. In addition to its scarcity as well as to its wastage, this vital resource is being affected by challenging environmental conditions. New approaches are, therefore, urgently needed for an optimized, fair, and efficient use of fresh water. Within this context, the authors propose a new framework based on multi-agent systems and the Digital Twins paradigm to enable intelligent and autonomous data acquisition, data processing, asset control, service generation, and service delivery. The multi-agent system is responsible of running data analytics mechanisms in order to assess water consumption and generate relevant

feedbacks to users using. It is also responsible of simulating asset operations under specific working constraints for the purpose of failure and/or defect detection.

Paper 8: Using Machine Learning Techniques to Support the Emergency Department, by Roeland van Delft and Renata M. de Carvalho [8]

This research lays down foundations for a stronger presence of machine learning in the emergency department. Using machine learning to make predictions on a patient's situation can increase patient's health and decrease the waiting time. This paper explores to what extent it is possible to accurately predict Emergency Room (ER) outcome. The predictions will be based on routinely available ER data from a Dutch hospital. The data set used is representative for any Dutch hospital. Prediction performance is compared between ML predictors. This research found that for more than half of the adult patients, the algorithm can very accurately predict hospitalization, with similar results for children and during the COVID-19. Moreover, several plans are introduced to substantially improve the ER process, for example by quickly reviewing patients selected by the algorithms.

Paper 9: Is Transfer Learning Helpful for Neural Combinatorial Optimization Applied to Vehicle Routing Problems? by Ali Yaddaden, Sébastien Harispe and Michel Vasquez [9]

Recently, combinatorial optimization problems have aroused a great deal of interest in Machine Learning, and leading to interesting advances in Neural Combinatorial Optimization (NCO): The study of data-driven solvers for NP-Hard problems based on neural networks. This paper studies the benefit of Transfer Learning for NCO by evaluating how model training can be improved taking advantage of knowledge learned while solving similar tasks. The authors focus, in particular, on two famous routing problems: The Traveling Salesman Problem (TSP) and the Vehicle Routing Problem (VRP). They study the effect of applying Transfer Learning from a model trained to solve TSP while training a model learning to solve the Capacitated VRP (CVRP).

Paper 10: Practice-Centered Approach to Design Cooperative Healthcare Information Systems: Data, Architectural and Organizational Challenges, by Rahma Marref, Khuloud Abou Amsha and Myriam Lewkowicz [10]

The healthcare sector is a collaborative environment that requires a joint action for delivering care. Health professionals who work in different organizations or settings must assimilate a massive amount of data generated during the patient care journey. Electronic healthcare records offer a starting point for supporting cooperation among healthcare professionals by saving and sharing traces of the patient's medical acts. The authors claim that these records merely store and share data, which disregards how health professionals use this data to understand the patients' situations and make decisions. They argue that focusing on the cooperative practices of managing patients gives design-

ers new insights to design future healthcare information systems supporting the cooperation.

Paper 11: Maturity Model for IoT Adoption in Hospitals, by Faruk Hasić, Bram Beirens and Estefanía Serral [11]

Hospitals are facing a wide variety of challenges in terms of quality and efficiency of healthcare. IoT is a technology used by organisations to increase efficiency and quality by recording measurements for historic analysis. Unfortunately, the benefits provided by a successful IoT adoption are currently out of reach for many hospitals, mainly due to the lack of appropriate maturity models. The authors develop, in this paper, an IoT maturity model that is tailored to the healthcare industry with an emphasis on Belgian hospitals. The proposed model is benchmarked against the scientific literature and industry expert opinions.

Paper 12: How to Overcome the Lack of Health Record Data and Privacy Obstacles in Initial Phases of Medical Data Analysis Projects, by Yehya Mohamad, Alexander Gabber, Sonja Heidenblut, Daniel Zenz, Anam Siddiqi and Henrike Gappa [12]

The authors present related work on medical data formats and data generation of medical health records. They present an approach to generate synthetic electronic health records (HER) that are readily available; suited to research; and free of legal, privacy, security and intellectual property restrictions to be used in home care research projects. They adapt and use Synthea™, an open-source software framework that simulates the lifespans of synthetic patients to generate synthetic EHRs. They also present the use-case of homecare from the capturing of patients' requirements, translating the requirements into a data model, feeding the data model into Synthea™ framework, to using EHRs to build an initial machine learning data model for homecare.

Paper 13: Brain Tumor Detection Using Selective Search and Pulse-Coupled Neural Network Feature Extraction, by Brad Niepceron, Filippo Grassia and Ahmed Nait Sidi Moh [13]

The identification of tumorous tissues in the brain based on Magnetic Resonance Images (MRI) analysis is a challenging and time consuming task that highly depends on radiologists' expertise. As prompt diagnosis of tumors can often be inherent to the patient's survival, it is crucial to decrease the amount of time spent on the manual analysis of MRI while increasing the accuracy of the detection process. To tackle these issues, the authors use the Selective Search (SS) algorithm coupled with a simplified Pulse-Coupled Neural Network (PCNN) for visual feature extraction and detection validation.

Paper 14: Automating the Dataset Generation and Annotation for a Deep Learning Based Robot Trajectory Adjustment Application for Welding Processes in the Automotive Industry, by Mohamed Slim Werda, Theodor Al Saify, Khalid Kouiss and Jaafar Gaber [14]

Industrial companies are more and more interested in the use of artificial intelligence (AI) in the control and the monitoring of their processes. They try to take advantage of the power of this technology in order to increase the level of automation and to build smarter machines with new capabilities of self-adaptation and self-control. Especially, the automotive industry, with their high requirements in productivity and diversity management, are eager to adapt AI concepts to their processes. However, the training of Deep Learning (DL) models requires an important effort of data preparation, to provide a dataset of all possible configurations. Considering the fact that automotive industry deals with a huge number of references and that it needs to modify very often and very quickly their products, it is very difficult, if not impossible, to gather a sufficient datasets for each produced reference and to have the time to train the DL models in the plants with the traditional methods. This paper presents an innovative methodology to prepare the dataset by creating virtual images instead of collecting real ones and then automatically annotating them. It demonstrates that this method will reduce significantly the efforts and the time of the preparation of the dataset. It also presents how this method was deployed for the quality control of welding operations in the automotive industry.

Paper 15: Home Health Care Scheduling Problem Under Uncertainty: Robust Optimization Approaches, by Adnen El-Amraoui, Slim Harbi and Ahmed Nait Sidi Moh [15]

The paper deals with home health care services (HHCS). It aims to formulate a model that addresses the homecare scheduling problem while taking into account the behaviors of patients, travel time uncertainty, as well as the behaviors of the medical team and the social actors.

Paper 16: Proposal of Critical Success Factors for eHealth Services Deployment, by Luis E. Mendoza, Lornel Rivas and Cristhian Ganvini [16]

The paper proposes a set of critical success factors (CSFs) for the implementation of eHealth. These factors allow for the identification of gaps and the proposal of alternatives for the optimization of healthcare services. They also facilitate the planning of eHealth projects and related activities. The paper investigates the requirements for an appropriate adoption of the CSFs based on the application context and the expected services.

Paper 17: Boosted Performance, Quick Response, and Better QoS Using IoT Plus, by Omar Mohammad, Zahra Yousef, Mirna Atieh and Kassem Ahmad [17]

The authors focus on some problems related to the IoT realm. They aim to lower data transmission time, reduce routing processes, decrease internet usages, increase response speed, deliver important and sensitive data first, improve the quality of services, and enhance the overall performance of the network. They present a new model in the application layer to classify IoT applications according to their valued data. They also propose an architecture

for fog computing and use the cell operator as the main fog center to store data.

Paper 18: Portability of Interfaces in Healthcare EAI Environments, by Severin Linecker and Wolfram Wöß [18]

Enterprise Application Integration (EAI) and HL7 (Health Level Seven) messaging are well established technologies in healthcare environments. Due to the age and longevity of HL7 standards (especially HL7 V2.x) and their widespread use, many interfaces outlive the middleware on which they run and must be ported to new systems. This often requires the entire code of the interface to be rewritten, which is associated with great effort and costs. The paper shows a generic EAI framework based on configuration and dependency injection for implementing reusable interfaces upfront and the results when applied to a real production EAI environment of an Austrian healthcare provider.

Paper 19: Agent-Based Approach for Connected Vehicles and Smart Road Signs Collaboration, by Mayssa Hamdani, Nabil Sahli, Nafaa Jabeur and Nadhira Khezami [19]

Road traffic is drastically increasing in big cities around the world. In order to enable a flexible management of this traffic, Intelligent Transportation System (ITS) solutions are relying on emergent ubiquitous, mobile, and communication technologies, particularly to intelligently deal with the limited capacities of the existing road infrastructures. While intelligence is left to the autonomous and connected vehicles as well as to the ITS, the road infrastructure has been mostly playing a passive role (as a source of data). Road signage, in particular, are in best cases dynamic but do not play an active role in monitoring traffic and incidents. The paper proposes to build Smart Road Signs (SRS) that can collaborate with Connected Vehicles in order to monitor traffic and warn drivers about any incident or danger. These SRSs are meant to operate autonomously in order to detect road traffic problems, share appropriate information with vehicles in the vicinity, and display relevant messages based on the ongoing contextual situation. To meet these goals, the authors rely on multi-agent systems to design SRSs as proactive components in the ITS landscape.

REFERENCES

[1] Dang, Q. V.: Using Machine Learning for Intrusion Detection Systems. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 12–33, doi: 10.31577/cai_2022_1_12.

[2] Afridi, A. H.: Review of Smart Contracts for Cloud-Based Manufacturing. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 34–55, doi: 10.31577/cai_2022_1_34.

[3] Messaoudi, C.—Guessoum, Z.—BenRomdhane, L.: Topic Extraction in Social Networks. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 56–77, doi: 10.31577/cai_2022_1_56.

[4] Lattisi, T.—Farina, D.—Ronchetti, M.: Semantic Segmentation of Text Using Deep Learning. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 78–97, doi: 10.31577/cai_2022_1_78.

[5] Hossain, T.—Mauni, H. Z.—Rab, R.: Reducing the Effect of Imbalance in Text Classification Using SVD and GloVe with Ensemble and Deep Learning. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 98–115, doi: 10.31577/cai_2022_1_98.

[6] Lamperti, G.—Zanella, M.—Zhao, X.: Fault Diagnosis of Discrete-Event Systems from Abstract Observations. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 116–134, doi: 10.31577/cai_2022_1_116.

[7] Zekri, S.—Jabeur, N.—Gharrad, H.: Smart Water Management Using Intelligent Digital Twins. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 135–153, doi: 10.31577/cai_2022_1_135.

[8] van Delft, R. A. J. J.—de Carvalho, R. M.: Using Machine Learning Techniques to Support the Emergency Department. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 154–171, doi: 10.31577/cai_2022_1_154.

[9] Yaddaden, A.—Harispe, S.—Vasquez, M.: Is Transfer Learning Helpful for Neural Combinatorial Optimization Applied to Vehicle Routing Problems? Computing and Informatics, Vol. 41, 2022, No. 1, pp. 172–190, doi: 10.31577/cai_2022_1_172.

[10] Marref, R.—Amsha, K. A.—Lewkowicz, M.: Practice-Centered Approach to Design Cooperative Healthcare Information Systems: Data, Architectural and Organizational Challenges. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 191–212, doi: 10.31577/cai_2022_1_191.

[11] Hasić, F.—Beirens, B.—Serral, E.: Maturity Model for IoT Adoption in Hospitals. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 213–232, doi: 10.31577/cai_2022_1_213.

[12] Mohamad, Y.—Gabber, A.—Heidenblut, S.—Zenz, D.—Siddiqi, A.—Gappa, H.: How to Overcome Lack of Health Record Data and Privacy Obstacles in Initial Phases of Medical Data Analysis Projects. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 233–252, doi: 10.31577/cai_2022_1_233.

[13] Niepceron, B.—Grassia, F.—Nait Sidi Moh, A.: Brain Tumor Detection Using Selective Search and Pulse-Coupled Neural Network Feature Extraction. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 253–270, doi: 10.31577/cai_2022_1_253.

[14] Slim Werda, M.—Al Saify, T.—Kouiss, K.—Gaber, J.: Automating the Dataset Generation and Annotation for a Deep Learning Based Robot Trajectory Adjustment Application for Welding Processes in the Automotive Industry. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 271–287, doi: 10.31577/cai_2022_1_271.

[15] El-Amraoui, A.—Harbi, S.—Nait Sidi Moh, A.: Home Health Care Scheduling Problem Under Uncertainty: Robust Optimization Approaches. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 288–308, doi: 10.31577/cai_2022_1_288.

[16] Mendoza, L. E.—Rivas, L.—Ganvini, C.: Proposal of Critical Success Factors for eHealth Services Deployment. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 309–329, doi: 10.31577/cai_2022_1_309.

[17] Mohammad, O.—Yousef, Z.—Atieh, M.—Ahmad, K.: Boosted Performance, Quick Response, and Better QoS Using IoT Plus. Computing and Informatics, Vol. 41,

2022, No. 1, pp. 330–356, doi: 10.31577/cai_2022_1_330.

[18] Linecker, S.—Wöss, W.: Portability of Interfaces in Healthcare EAI Environments. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 357–375, doi: 10.31577/cai_2022_1_357.

[19] Hamdani, M.—Sahli, N.—Jabeur, N.—Khezami, N.: Agent-Based Approach for Connected Vehicles and Smart Road Signs Collaboration. Computing and Informatics, Vol. 41, 2022, No. 1, pp. 376–396, doi: 10.31577/cai_2022_1_376.

Nafaa Jabeur is Associate Professor and Director of Research at the German University of Technology in Oman (GUtech). He received his Ph.D. and M.Sc. degrees in computer science from the Laval University, Quebec, Canada in 2006 and 2001, respectively. He received his engineer degree in computer engineering in 1999 in Morocco. He has more than 19 years of experience in the industrial and academic sectors. He worked in several countries, including Tunisia, Morocco, Canada, Belgium, and the Sultanate of Oman. In the industrial sector, he worked as software engineer developer, project manager, business developer, and CEO of an IT company. In the academic sector, he worked in several universities, as Assistant/Associate Professor, Head of Department, and Director of Research. He has participated in several R & D and consultancy projects, edited 2 books, and authored more than 80 research papers in prestigious conferences and high ranked journals. His main research interests include smart cities, transportation, IoT, blockchain, artificial intelligence, drones, network security, and augmented reality.

Ansar Yasar is Professor at the Transportation Research Institute (IMOB), Hasselt University Belgium, one of the top European institutes focusing on the broader themes of road safety and transport management using new modes of mobility. He is Senior Member of IEEE, Chartered Engineer UK, Fellow BCS and EUR ING (European Engineer). At IMOB, he worked on the European FP7 project DATA SIM (2011–2014), European ERA-NET Smart-PT (2014–2017), H2020 Track & Know (2018–2020) and H2020 iSCAPE (2016–2019) projects. He is currently responsible for the H2020 Search & Rescue project (2020–2023) with a consortium of several international partners. He is responsible for the Intelligent Transport Systems (ITS) and Intelligent Topics in Transportation courses for the masters of transportation sciences students and related research around this topic at UHasselt. Furthermore, he is currently Head of the Business Development Unit at the institute focusing on several applications including Safe Routes to Schools & Safe-Trucking system that resulted in creation of successful spin-off companies. His broader research interests include smart cities and communities, connected and intelligent mobility, drones system management, road safety using smart transport solutions and mobility management. As a part of his basic research, he has expertise in modelling

pedestrian flow using various IoT-based techniques and simulations tools such as VisWalk. He has supervised several Master and Ph.D. students in the topics of transport and pedestrian modelling using infused ITS techniques. He received his Ph.D. in engineering from Katholieke Universiteit Leuven, Belgium, his M.Sc. degree in computer science and engineering from Linkoping University, Sweden and his B.Sc. degree in software engineering from Foundation University Islamabad, Pakistan. He has authored more than 200 research articles in renowned international journals, conferences and workshops. As one of his accomplishments, he has co-edited a book entitled "Data Science and Simulation in Transportation Research" published by IGI Global in December 2013. Furthermore, he has been involved in organization of many international peer-reviewed conferences, summer schools and other scientific events. He is also a technical expert to evaluate project proposals submitted to the European R & D – EUREKA and COST frameworks.

Yehya Mohamad holds his Ph.D. in computer science from the Technical University of Aachen (Germany). His Ph.D. works focused on vital sign detection in real time from sensors and wearables. He holds a Masters in computer science from the Technische Universität Berlin. He is a senior researcher, lecturer, and digital health expert at the Fraunhofer Institute for Applied Information Technology FIT (Germany) since 2001. His research interests include digital health, compliance, usability and accessibility of information systems, standards, user profiling, internet technologies, and web of things. He was involved in several funded European projects and published about hundred research papers in prestigious conferences and journals. He also organized several successful international research events.

Michele Melchiori is Associate Professor at the Department of Information Engineering of the Università degli Studi di Brescia, Italy, where he has been doing research and teaching activities since 1998. He earned his Ph.D. in information engineering from the University of Brescia in 2001 and, previously, a Master degree in computer science from the University of Milan, Italy. His main research interests are in the area of models and tools for the design of information systems, and specifically on semantic resources discovery, semantic-based recommendation systems for web services and mash-ups, semantic interoperability, ontologies for interoperability, semantic web service discovery in distributed environments, semantic/social technologies, and integration of semi-structured data. On these topics, he has worked on in the context of various funded competitive projects. International project: INTEROP (network of excellence, INTEROPerability of enterprise applications and software). National projects: VISPO (Virtual District Internet-based Service Platform, research activity with strategic character, co-funded by the Italian Ministry of Education, University and Research (MIUR); MAIS (Multichannel Adaptive Information

Systems, MIUR FIRB project); ESTEEM (Emergent Semantics and cooperaTion in multi-knowledgE EnvironMents – Advanced methods and tools for semantic cooperation in Web virtual communities, MIUR FIRB project). He has been scientific coordinator of operational unit in a project with government funding (FEARS projects), Title of the project: "Portal and Collaborative Tools Design for Traceability of the Supply Chain and Openness of Food-Farming Products." Previous national projects: D2I (From Data to Information, MIUR scientific research programs with relevant national interest); INTERDATA (Methodologies and Technologies for Data and Process Management on Internet and Intranets, co-funded by MIUR). He published a large number of papers in highly ranked journals and conferences.