Big and Deep Hype and Hope: On the Special Issue for *Deep Learning and Big Data in Healthcare*

José Luis Rojo-Álvarez [1,2,3]

1 Departamento de Teoría de la Señal y Comunicaciones, Sistemas Telemáticos y Computación, Universidad Rey Juan Carlos, 28943 Fuenlabrada, Spain; joseluis.rojo@urjc.es; Tel.: +34-91-488-8744
2 Center for Computational Simulation, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Spain
3 Persei Vivarium, Inc., 28013 Madrid, Spain

Deep Learning networks are revolutionizing both the academic and the industrial scenarios of information and communication technologies. Their theoretical maturity and the coexistence of large datasets with computational media is making this technology available to a wide community of makers and users, and recent evolution has been remarkable in techniques such as deep belief networks, Boltzmann machines, auto encoders, or recurrent networks. In a different yet often closely related arena, the analysis of large amounts of data from the Electronic Health Recording, the Hospital Information Systems, and other medical data sources, success cases on companies, and new products have made possible new tools for estimation of in-hospital stay duration, chronic patient identification, and policies to reduce readmissions by preventing illness progression. Large and small companies have paid attention to this new era, in which machine learning and statistical analysis need to be revisited if they want to provide suitable algorithms, especially in healthcare scenarios, where patient data more than ever is becoming the key to improving patient healthcare.

Healthcare is now an open field for advantageous use of deep learning and big data advancements, and challenges are open in order to provide systems that can be accurate enough to be useful to the clinician and the patient in the health itinerary. Not only are large amounts of data available, but also sensitivity and specificity are to be paid special attention, as well as support systems rationally fitting into the health system.

The goal of this special issue was to put together relevant contributions, condensed into five key cornerstones of deep learning and big data applications in healthcare. On the one hand, the applications have included works with medical images (magnetic resonance, radioscopy and tomography, echography, nuclear medicine), contributions to signal processing (cardiac, neural, long-term monitoring, wellness devices), or data from large forms (primary attention, specialized medicine, clinical practice, electronic health recordings, hospital information systems, interoperability). The feature interpretation remains an open issue in deep learning and big data state-of-the-art, but it takes special relevance in healthcare applications in order to gain confidence in their use both by the healthcare staff and by the patients; so contributions including insights into this hot and open topic have naturally been provided.

The result is that accepted submissions have naturally grouped around a set of topics and techniques that can be considered as actual knowledge clusters in these fields, and they can be summarized as follows.

Medical structured data: Many applications deal with what we could denominate structured data, which includes as sources tabular data from trials, from the electronic health recording itself, or from data forms obtained from the patients. In [1], the failure of dental implants is early predicted using machine learning systems and considering a variety of features compared with previous approaches based on single factors. In [2], patient eligibility in interventional cancer clinical trials is predicted in terms of short statements considered as inclusion or exclusion criteria, and using

pretrained word-embeddings as inputs to deep classifiers. In [3], blood transfusion in the sensitive case of orthopedic procedures is scrutinized in advance to surgery, and different classifiers (such as logistic regression and decision trees) are used to open the way to this new application. In [4], daily questionnaires from mobile applications are shown to suffer from missing data, and they are completed with a principled approach based on long-term–short-term memory networks, with representative case applications given by their use for activity, food, and biometrics in athletes. In [5], adverse drug reactions are analyzed from medical records of patients with Stevens-Johnson Syndrome, using a generalized sequential pattern algorithm for early detection of their onset. In this work, the results provided with clinical information relevant for different medication groups, which can be useful for the involved physicians to stop the administration of suspected drugs in the early stages, are mined.

Medical unstructured data: A different application field can be established in the unstructured data (broadly speaking, free text, images, and others) that can be obtained from Hospital Information Systems and other related sources. In [6], concept detection in medical images is addressed with the help of unsupervised learning applied to the biomedical literature, with the use of bags of visual words, autoencoders, and generative adversarial networks. In [7], the behavioral characteristics of smartphone addiction are scrutinized by automatically collecting and analyzing data through an app. Cognitive bias is present between the self-reports and the automatically collected data, and here again data mining allows to show the relevant terms and items, hence contributing to knowledge discovery in these sources. In [8], a file system structure with distributed topography is developed to support the storage of large amounts of data from wellness recordings in the Internet of Things. In [9], the integration of atomic association rules with classification is improved to adapt large amounts of data, and it is used with advantage in Tibetan medical syndrome, a model of which is built in shorter time, with fewer rules but more understandable, and with higher accuracy than the conventional methods.

Medical signals: Medical signal processing is a key application set in the medical field, as many diagnostics and medical examinations are based on time-changing measured magnitudes. The electromiogram signals represent the basis for many biomedical and clinical tests, as well as for rehabilitation and prosthetic applications. In [10], stacked and sparse autoencoders are used for hand motion classification in terms of electromiogram signals in robotic arms and prosthetic hands. In [11], surface electromiogram signals are used to recognize the gait and to control the lower-limb exoskeleton or prostheses, as well as to analyze with detail the effect of load variation, using for this back propagation neural nets, support vector machines, and nearest neighbor classifiers.

Another emerging field is related with biomechanics, in which a variety of waveforms naturally encode non-trivial information about our health and wellness. In [12], the proper postural habit in childhood is supported by a monitoring system classifying their sitting posture, and convolutional neural networks therein exhibit improved performance with respect to other machine learning schemes. In [13], motional patterns of human postures are identified by introducing the so-called data density functional method, by mapping the sensed time signals into specific physical spaces, and then turning them into discrete states that are detected using clustering algorithms. In [14], a detailed review puts together the topics of biomechanical feedback, motor skills, and wearables, both for sports and arts scenarios. Focusing on real-time biomechanical feedback, which is still a scarce application nowadays, specific areas are identified as useful in this arena, such as 3D-motion capture, anthropometry, biomechanical modelling, sensing technology, and artificial intelligence. As a synthesis, a two-chain body model integrates the inertial measurement units with deep learning technology, and the availability of massive data seems to be the key for this scheme yielding an advantageous alternative to existing and conventional methods.

Cardiac signals are always strongly present in the machine learning literature. In [15], the early detection of lethal arrhythmia ventricular fibrillation is addressed from electrocardiographic signals, with the hard-to-achieve real time requirement. Time-frequency representations are used to yield an image representation of the cardiac bioelectric potentials, thus avoiding the need for previous feature extraction and selection methods before feeding this input into a combination of classifiers.

In [16], classification of heart sound signals from phonocardiographic measurements obtained in digital stethoscopes is addressed by using one normal and five abnormal categories. Features are extracted from their cepstral coefficients and their wavelet transform, and they are fed to a deep neural network, exhibiting competitive accuracy.

Medical images: Deep learning techniques have yielded a number of advantages when used with multimedia images. It is not surprising that their use for medical images is being increasingly explored and gaining in sophistication and use specialization. In [17], the early detection of atherosclerotic plaque rupture in sudden coronary deaths is addressed in terms of thin cup fibroatheroma or vulnerable plaque detection from virtual histology intravascular ultrasound images, with geometric and texture features, and using different classifiers for this purpose including back-propagation neural networks, nearest neighbors, and support vector machines). In [18], an automatic system for fluorescence intensity classification is designed to support the autoimmune diagnostics with human epithelial cells from indirect immuno-fluorescence indirect images, using convolutional neural networks and support vector machines. In [19], skull stripping in brain magnetic resonance imaging is designed as an alternative to the time-consuming manual segmentation task by using a 3-D U-net for automatic segmentation of the full brain with satisfactory results. In [20], early detection of polyps are a convenient target to prevent colorectal cancer. Aiming to reduce false negatives during manual inspection and support screening tests, convolutional neural networks are trained in this work to mark the regions, and features are reinforced with transfer learning and fine tuning among databases.

Technical shared elements: Some technical considerations have a transverse interest for many applications of the techniques targeted in this special issue. The input space and its expression in terms of feature extraction and selection, the data imbalance, and other topics, should devote specific attention and benefit from development of new approaches and by adequate adaptation of existing ones in other fields. In [21], novelty detection is addressed in terms of deep autoencoders with density based clustering by sending the results to a density-based cluster, so that points not involved in any group are considered as novelties. This represents a likely convenient approach to diagnostic formulations from patient data in those situations where labels are not so easy to obtain. In [22], data imbalance is analyzed in terms of its impact on machine learning algorithms, based on the analysis of patient data from electronic health records. The inclusion of multi-labeling tasks in allowing to give several diagnoses for each patient is a natural requirement which degrades the performance of most of the learning machines, and it is alleviated here with the use of regularized ensembles.

In [23], a comprehensive review is stated to put this special issue into context. The reader can find therein simple descriptions of the technical elements of both big data technologies and deep learning techniques. Application fields in which success stories can be told are first summarized therein and are followed by a view to the current state of healthcare scenarios. A specific analysis of the scope and limitations is provided for the literature on electrocardiographic arrhythmia classification using public databases, and a case-study example is provided for beginners to start to play with these concepts. Overall, this special issue has outlined the state in which we find ourselves today, which is hard to summarize, but probably we can think that healthcare applications for these technologies are not only promising, but also necessary. Maybe the scientific exploration is taking longer than in other fields, and we can speculate on the possibility that we need much more reliable algorithms in healthcare than in multimedia applications, for instance; then, the additional effort that is being devoted by the scientific community becomes natural. Nevertheless, the disruption in this field has not yet fully arrived, but it will probably come soon, and then this special issue and this period will become nicely obsolete. The body of work by the community in this field represents an effervescent and necessary exploration, so every big and small work in this setting deserves much respect and encouragement to keep moving forward.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Liu, C.H.; Lin, C.J.; Hu, Y.H.; You, Z.H. Predicting the Failure of Dental Implants Using Supervised Learning Techniques. *Appl. Sci.* **2018**, *8*, 698. [CrossRef]
2. Bustos, A.; Pertusa, A. Learning Eligibility in Cancer Clinical Trials Using Deep Neural Networks. *Appl. Sci.* **2018**, *8*, 1206. [CrossRef]
3. Chang, C.M.; Hung, J.H.; Hu, Y.H.; Lee, P.J.; Shen, C.C. Prediction of Preoperative Blood Preparation for Orthopedic Surgery Patients: A Supervised Learning Approach. *Appl. Sci.* **2018**, *8*, 1559. [CrossRef]
4. Rodenburg, F.J.; Sawada, Y.; Hayashi, N. Improving RNN Performance by Modelling Informative Missingness with Combined Indicators. *Appl. Sci.* **2019**, *9*, 1623. [CrossRef]
5. Ou-Yang, C.; Chou, S.C.; Juan, Y.C.; Wang, H.C. Mining Sequential Patterns of Diseases Contracted and Medications Prescribed before the Development of Stevens-Johnson Syndrome in Taiwan. *Appl. Sci.* **2019**, *9*, 2434. [CrossRef]
6. Pinho, E.; Costa, C. Unsupervised Learning for Concept Detection in Medical Images: A Comparative Analysis. *Appl. Sci.* **2018**, *8*, 1213. [CrossRef]
7. Lee, M.; Han, M.; Pak, J. Analysis of Behavioral Characteristics of Smartphone Addiction Using Data Mining. *Appl. Sci.* **2018**, *8*, 1191. [CrossRef]
8. Ergüzen, A.; Ünver, M. Developing a File System Structure to Solve Healthy Big Data Storage and Archiving Problems Using a Distributed File System. *Appl. Sci.* **2018**, *8*, 913. [CrossRef]
9. Zhu, X.; Zhang, L.; Zhang, Y.; Wang, L.; Wang, S.; Liu, P. Research on Classification of Tibetan Medical Syndrome in Chronic Atrophic Gastritis. *Appl. Sci.* **2019**, *9*, 1664. [CrossRef]
10. Zia ur Rehman, M.; Gilani, S.O.; Waris, A.; Niazi, I.K.; Slabaugh, G.; Farina, D.; Kamavuako, E.N. Stacked Sparse Autoencoders for EMG-Based Classification of Hand Motions: A Comparative Multi Day Analyses between Surface and Intramuscular EMG. *Appl. Sci.* **2018**, *8*, 1126. [CrossRef]
11. Zhang, X.; Sun, S.; Li, C.; Tang, Z. Impact of Load Variation on the Accuracy of Gait Recognition from Surface EMG Signals. *Appl. Sci.* **2018**, *8*, 1462. [CrossRef]
12. Kim, Y.M.; Son, Y.; Kim, W.; Jin, B.; Yun, M.H. Classification of Children's Sitting Postures Using Machine Learning Algorithms. *Appl. Sci.* **2018**, *8*, 1280. [CrossRef]
13. Huang, S.J.; Wu, C.J.; Chen, C.C. Pattern Recognition of Human Postures Using the Data Density Functional Method. *Appl. Sci.* **2018**, *8*, 1615. [CrossRef]
14. Zhang, X.; Shan, G.; Wang, Y.; Wan, B.; Li, H. Wearables, Biomechanical Feedback, and Human Motor-Skills' Learning and Optimization. *Appl. Sci.* **2019**, *9*, 226. [CrossRef]
15. Mjahad, A.; Rosado-Muñoz, A.; Guerrero-Martínez, J.F.; Bataller-Mompeán, M.; Francés-Villora, J.V.; Dutta, M.K. Detection of Ventricular Fibrillation Using the Image from Time-Frequency Representation and Combined Classifiers without Feature Extraction. *Appl. Sci.* **2018**, *8*, 2057. [CrossRef]
16. Yaseen.; Son, G.Y.; Kwon, S. Classification of Heart Sound Signal Using Multiple Features. *Appl. Sci.* **2018**, *8*, 2344. [CrossRef]
17. Rezaei, Z.; Selamat, A.; Taki, A.; Mohd Rahim, M.S.; Abdul Kadir, M.R.; Penhaker, M.; Krejcar, O.; Kuca, K.; Herrera-Viedma, E.; Fujita, H. Thin Cap Fibroatheroma Detection in Virtual Histology Images Using Geometric and Texture Features. *Appl. Sci.* **2018**, *8*, 1632. [CrossRef]
18. Cascio, D.; Taormina, V.; Raso, G. Deep Convolutional Neural Network for HEp-2 Fluorescence Intensity Classification. *Appl. Sci.* **2019**, *9*, 408. [CrossRef]
19. Hwang, H.; Rehman, H.Z.U.; Lee, S. 3D U-Net for Skull Stripping in Brain MRI. *Appl. Sci.* **2019**, *9*, 569. [CrossRef]
20. Sornapudi, S.; Meng, F.; Yi, S. Region-Based Automated Localization of Colonoscopy and Wireless Capsule Endoscopy Polyps. *Appl. Sci.* **2019**, *9*, 2404. [CrossRef]
21. Amarbayasgalan, T.; Jargalsaikhan, B.; Ryu, K.H. Unsupervised Novelty Detection Using Deep Autoencoders with Density Based Clustering. *Appl. Sci.* **2018**, *8*, 1468. [CrossRef]

22. Han, H.; Huang, M.; Zhang, Y.; Liu, J. Decision Support System for Medical Diagnosis Utilizing Imbalanced Clinical Data. *Appl. Sci.* **2018**, *8*, 1597. [CrossRef]

23. Bote-Curiel, L.; Muñoz-Romero, S.; Gerrero-Curieses, A.; Rojo-Álvarez, J.L. Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners. *Appl. Sci.* **2019**, *9*, 2331. [CrossRef]