Bioinformatics and Medicine in the Era of Deep Learning

Davide Bacciu[1,*], Paulo J.G. Lisboa[2], José D. Martín[3],
Ruxandra Stoean[4], Alfredo Vellido[5,†]

1- Università di Pisa - Italy

2- Liverpool John Moores University - UK

3- Universitat de València - Spain

4- University of Craiova - Romania

5- Universidad Politécnica de Cataluña - Spain

Abstract. Many of the current scientific advances in the life sciences have their origin in the intensive use of data for knowledge discovery. In no area this is so clear as in bioinformatics, led by technological breakthroughs in data acquisition technologies. It has been argued that bioinformatics could quickly become the field of research generating the largest data repositories, beating other data-intensive areas such as high-energy physics or astroinformatics. Over the last decade, deep learning has become a disruptive advance in machine learning, giving new live to the long-standing connectionist paradigm in artificial intelligence. Deep learning methods are ideally suited to large-scale data and, therefore, they should be ideally suited to knowledge discovery in bioinformatics and biomedicine at large. In this brief paper, we review key aspects of the application of deep learning in bioinformatics and medicine, drawing from the themes covered by the contributions to an ESANN 2018 special session devoted to this topic.

1   Introduction

Deep Learning (DL) [1] has become an increasingly popular Machine Learning (ML) approach in the last decade, as shown in Fig. 1. One of the main reasons for its success stems from its internal representation in the form of high-level features, allowing the modelling of difficult problems, and a smart initialization of some other deep structures. Moreover, staging the difficult task of efficient feature selection by using multiple layers has been crucial in solving extremely difficult problems of image classification [2, 3, 4], or Natural Language Processing (NLP) [5, 6, 7] by means of Convolutional Neural Networks (CNNs) and Deep Recurrent Neural Networks (RNN), respectively. The success achieved in such complicated problems has generated a great interest not only in the academic community but also in industry, with many private companies involved in the development of commercial products based on DL [8, 9].

Although CNNs and recurrent networks have already produced significant advances in biomedical imaging and biomedical signal processing [8], the impact
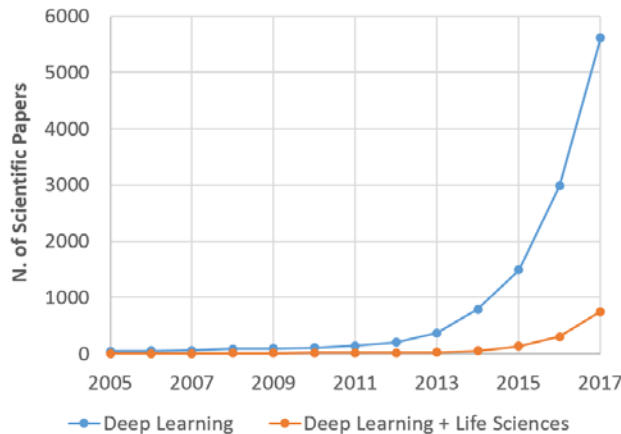

Fig. 1: Evolution on the number of papers published on DL topics with respect to those on DL in BioInformatics (Source: query on abstracts keywords on Scopus data; date: February 1st, 2018)

.

of DL in Bioinformatics is still limited. This is possibly related to some of the challenges that many bioinformatics data sets represent, such as insufficient and unbalanced data, or challenges related to the area of application itself, such as the lack a straightforward interpretation of deep models. This special session aims to bring together some of the most recent advances in DL as applied to bioinformatics and (bio)medicine.

The contents of this tutorial are outlined as follows: Section 2 deals with the problem of structured biomedical data. Section 3 reviews the contribution of DL to image processing in medicine. Section 4 provides a glimpse of a key issue when dealing with DL approaches in the clinical field, namely, intepretability. Finally, Section 5 gives some insights on future trends and challenges derived from the papers submitted to the session.

2   Deep Learning of structured biomedical data

The term structured data denotes a whole selection of data types encoding different forms of relational information, ranging from sequences to more complex and general classes of graphs. Structured data can be seen as representing compound information made of atomic entities, i.e. the labels of the nodes/vertices of the structures, linked by relationships, denoted by the edges between the nodes (possibly complemented by edge labels). As such, structured data place emphasis on evaluating and interpreting information "in context" rather than following the classical i.i.d. assumption, where the context is defined by the set of nodes an atomic piece of information is linked to.

Structured data arise naturally as the result of many biological, physical and

chemical processes. Sequences are the simplest form of structured data and are a natural representation, for instance, for nucleotide compounds (e.g. DNA, RNA) and for physiological signals (e.g. ECG, EEG, MEG data). Trees, instead, are widely used to represent cell phylogeny and clonal diversity, while more general classes of graphs find wide application to represent proteomic and metabolomic interactions, protein structures as well as chemical compounds.

The field of ML for structured data processing has a long consolidated history dating back to the sixties for the first approach to sequential data, namely the Hidden Markov Model (HMM) [10]. Later, in the nineties, the scientific community started tackling more complex structures, with works proposing recursive approaches for acyclic graph processing [11, 12], later followed by a number of approaches tackling more general classes of graphs with cycles [13, 14].

The first and most popular applications of DL to biomedicine concern the processing of sequential data. CNN models have found wide application to genomics, in particular for the prediction of protein binding sites such as in the DeepBind [15] and DeepSEA [16] methods, where it has also been proposed as a method to visualize the effect of wild-type mutation on binding site prediction, contributing to the interpretability of the learned model. Despite their limitation in treating only fixed length subsequences, CNNs have quickly become the reference DL model for genomic studies: a recent review of the main CNN approaches in the field can be found in [17]. RNNs, such as the Long Short Term Memory (LSTM), have found less application in genomics despite their ability in modeling variable length sequences, mostly due to a widespread belief of the biomedical community, which considers them difficult to train [18]. An attempt to bring the LSTM model into the biomedical community is put forward in [19], where an approach to learn interpretable features from an LSTM trained on real-world clinical time-series is proposed. Along the same line, [20] proposed the use of a LSTM to classify 128 diagnoses from multivariate clinical time series collected in an intensive care unit (ICU). Use of more parsimonious RNN models is discussed in [21] for scoring stress levels from heart-rate information. On the unsupervised learning side, stacked denoising autoencoders have been proposed to learn informative encodings on noisy ECG time series [22]. The ubiquitous Generative Adversarial Networks (GANs) have very recently been used [23] for the generation of molecules encoded as character sequences.

Several papers of this ESANN 2018 special session deal with DL approaches for biomedical sequences. In [24], it is proposed a framework integrating deep autoencoders with kernel methods to learn effective encodings of multivariate clinical time series in presence of missing information. A deep RNN approach is discussed in [25] to diagnose Parkinson's disease from spiral drawing tests. CNNs are used in [26] to classify sleep stages from EEG recordings, using a fixed window approach splitting time series into subsequences of 30 seconds of electrical activity.

The long wave of the DL revolution is now approaching and rediscovering the processing of more complex and expressive forms of structured data, in particular as regards cheminformatics and drug discovery. Again, CNNs are among

the most popular approaches here: [27] proposed the first attempt to process molecular data within a CNN using circular fingerprint encodings imported from the chemistry literature. A proper generalization of the concept of convolution from images to graphs of varying dimensionality has been proposed in [28], with application to the characterization of chemical compounds. Graph convolutions are used in [29] together with an attentional model based on LSTM to predict molecule properties with a one-shot learning approach. Very recently, a DL model, based on a similar encoding to the one in [14], has been proposed in [30] for the processing of large scale networks of proteomics and metabolomics interactions for predicting new drug-disease associations in drug repurposing. The special session contributes to this flourishing topic with a paper [31] discussing a deep graph kernel for disease gene prioritization in bioinformatics.

3 Deep Learning for medical image processing

The popularity of DL has exponentially risen especially due to its capability to process images independently from human intervention. This includes inherent robustness to variations in position, rotation, scale, perspective and occlusion. These traits have consequently appeared as particularly valuable in the medical sector. The amount of image data available for analysis keeps increasing with the modernization and constant use of imaging devices. Time-saving decision support in this area was achieved by ML techniques before the arrival of DL, but with a different supplementary human cost, i.e. that of highlighting the regions of interest in the images, of handcrafting the relevant features for the diagnosis and of labelling each image. This has proved to be a turning point for DL, as the accurate provider of both image processing and image interpretation. Lacking the need for expert handcrafted features by automatically learning the optimal attributes from the available images [32] and benefiting from large amounts of available data, DL has therefore successfully entered the realm of medical imaging. Its applications range from landmark detection and tissue segmentation to diagnosis and prognosis [33].

The main DL approach to image analysis is the CNN. Its special convolutional layers are able to detect the relevant features gradually, from those low-level to the high-level structures, by inspecting small portions of the training images. Following the independent feature discovery, the image is labelled according to the given task. CNN modelling may target images from histopathology [34, 35], CT scans [36, 37], or MRI [38], to name a few. Stacked autoencoders can also be employed for automatic feature extraction, such as in [39] for MRI imaging and in [40] for CT and ultrasound images.

The success of DL for medical imaging can be appreciated in the constant presence of the topic at conferences specialized in biomedical computing and in journals dealing with medical image processing, as well as in the advent of companies using these methods in the area (e.g. Enlitic, Arterys, or Lunit [41]).

Shifting now from the brighter side of the DL potential for image mining in medicine, a big problem that the paradigm faces in this real-world complex field

of application, as compared to its use for general images, must be acknowledged. The actual scarcity of labelled medical images makes it harder for the approach to perform well, leading to overfitting and hard parametrization. Current solutions include data geometric augmentation, transfer learning and fine-tuning [33, 42] from general image data sets such as ImageNet or preferably [43] from those emerging in the medical domain (e.g. on Kaggle), as well as GANs [44].

Several papers in the current special session cover some of the key aspects of the application of DL for image processing in medicine and public health. As the field of application is concerned, two of them target medical imaging tasks from CT and HP and one deals with recognising pollen from microscopic images for allergy and asthma prevention in medicine. As for the DL architectures involved, two CNN (of which one is a GAN) and one autoencoder emulator are employed to solve the problems. In [45], the authors propose an expansion of an initially low number of available 2D lung CT scans through a GAN methodology. In [46], a NMF approach for learning a reduced feature representation in a fashion similar to an autoencoder is described for the processing of histopathological colorectal cancer slides. Finally, from the public health perspective, a CNN for pollen recognition from a collection of microscopic images is employed in [47].

4 Interpretable DL models in biomedicine and bioinformatics

Over the last decades, data have become central to the life sciences. Biomedicine and, perhaps more expressively, bioinformatics are perfect examples of that. The introduction of computerized and networked digital systems and the impressive advances in non-invasive data acquition technologies are putting data at the heart of these disciplines.

Data are hoped to become the key to the discovery of new knowledge at all physiological scales, opening the doors to hitherto unaccessible medical advances. Such transformation from data to knowledge is a natural goal for ML. To be accepted in biomedicine and bioinformatics, and very especially in medical practice, ML-based approaches must be trusted. One of the main challenges ML faces to achieve such desired trustworthiness is that of becoming explainable and interpretable [48, 49].

The relevance of this challenge is heightened by a pressing societal issue: the implementation of the new European Union directive for General Data Protection Regulation (GDPR). It is to be enforced throughout 2018 with minimum variation between European countries. Its Article 22, concerning "Automated individual decision-making, including profiling", mandates a *right to explanation* of all decisions made by automated or artificially intelligent algorithmic systems [50]. The GDPR directive makes model interpretability a core concern in biomedical decison making; it directly involves ML and should particularly concern those aiming to see DL being used in medical practice, beyond basic research.

Another of the reasons for explainability and interpretability to have become,

of late, contentious and hotly debated issues in ML is, precisely, the new life given to connectionism in the form of DL. In general, DL approaches are extreme cases of *black box* models. Being such a success story for ML, DL models' lack of interpretability has become a pressing concern in the area, reflected by recent literature. A concern of no easy solution, given the difficulty of making these often extremely complex systems become somehow transparent. In [51], an adversarial training scheme was recently proposed, where model neurons "are endowed with human-interpretable concepts" and interpretable representations can trace outcomes back to influential neurons, providing an explanation of how models make their predictions. Also recently, interpretability criteria based on analysis of deep networks in the information plane [52] were described.

Lack of transparency has indeed been argued to be one of the main barriers to the acceptance and adoption in medicine of ML methods in general and DL methods in particular [43]. This view is shared by Che and co-workers; in [53], they propose gradient boosting decision trees to extract interpretable knowledge from a trained deep network. In related work [54], deep models are regularized so that their class-probability predictions can be modeled with minimum loss by decision trees with few nodes. Interpretation from these trees is far more tractable and intuitive than from the original models.

An example of how to deal with interpretability in DL can be found in the ESANN 2018 special session covered in this tutorial. In [55], an extension of *interpretable mimic learning* that teaches reasonably simple and interpretable models so as to mimic predictions of complex DL models without missing in the performance is introduced. It focuses on problems of ordinal classification and illustrates the capabilities of the model with a problem of ordinal response to cancer treatment.

5   Some future trends and challenges

The application of DL methods to problems in biomedicine and bioinformatics is a many-faceted problem. At this point in time and in a brief tutorial such as this, it would be impossible to provide a comprehesive view of the future trends in the area and the various important challenges faced by such applications in real-world scenarios.

Some future trends, though, have been outlined for biomedicine in [56] and for bioinformatics in [8]. For biomedicine, DNNs are quoted to be of potential benefit to areas as varied as semantic linking, biomarker development, drug discovery (structural analysis and hypothesis formulation through DNN abstract learned representations analysis), clinical recommendations and transcriptomic data analysis . For bioinformatics, investigating proper ways to encode raw and multi-modal data forms, instead of human-processed features, and learn suitable features from those multi-modal or raw forms is seen as a future challenge for DNNs.

Note that the previous comments do not cover medical applications in clinical settings and healthcare. In these, and as reported in the previous sections,

explainability and interpretability are major challenges for DL. Despite extensive research into the interpretation of neural networks in clinical settings dating back several decades [57] and extending to more recent work [58], we are only now weaving the first strands of methods capable of *translating* the complex inner workings of deep architectures. This means that this is possibly both a future strong research trend and far-from-trivial challenge. The latter is exemplified by a recent study by Google Brain researchers [59], who show the surprising result of the lack of sensitivity of local model explanations to deep neural network (DNN) parameter values. The study concludes that "the architecture of DNN is a strong prior on the input, and with random initialization, is able to capture low-level input [image] structure".

As mentioned, another likely future trend and definite challenge for DL is the coherent integration of multi-modal data [60]. A case in multi-omics data integration is presented in one of the studies in the ESANN 2018 special session covered by this tutorial [61]. It delivers a novel *super-layered* neural network architecture named cross-modal neural network. Interestingly, it is meant to perform well in scenarios with a limited number of training examples available (something not uncommon in multi-modal biomedical problems).

A final glimpse of the future is provided in [62], where a reinforcement learning brain-dynamics-model-free approach to control response properties of *biological* neural networks is described. Beyond this proof-of-concept, such approach could have an enormous impact on the study of degenerative brain diseases.

References

[1] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[2] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. 2012.

[3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[4] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In C. Cortes *et al*, editor, *Advances in Neural Information Processing Systems 28*.

[5] A. Graves, A.-R. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649. IEEE, 2013.

[6] I. Sutskever, O. Vinyals, and V. Le Quoc. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 3104–3112. MIT Press, 2014.

[7] R. Nallapati, B. Xiang, and B. Zhou. Sequence-to-sequence RNNs for text summarization. *CoRR*, abs/1602.06023, 2016.

[8] S. Min, B. Lee, and S. Yoon. Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18(5):851–869, 2017.

[9] K. Murnane. Thirteen companies that use deep learning to produce actionable results, 2016. Online: https://www.forbes.com/sites/kevinmurnane; accessed 20 February 2018.

[10] L.E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 12 1966.

[11] P. Frasconi, M. Gori, and A. Sperduti. A general framework for adaptive processing of data structures. *IEEE Transactions on Neural Networks*, 9(5):768–786, Sep 1998.

[12] D. Bacciu, A. Micheli, and A. Sperduti. Compositional generative mapping for tree-structured data - part I: Bottom-up probabilistic modeling of trees. *IEEE Transactions on Neural Networks and Learning Systems*, 23(12):1987–2002, 2012.

[13] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, Jan 2009.

[14] A. Micheli. Neural network for graphs: A contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3):498–511, March 2009.

[15] B. Alipanahi, A. Delong, M.T. Weirauch, and B.J. Frey. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8):831–838, July 2015.

[16] J. Zhou and O.G. Troyanskaya. Predicting effects of noncoding variants with deep learning based sequence model. *Nature Methods*, 12(10):931–934, August 2015.

[17] W. Jones, K. Alasoo, D. Fishman, and L. Parts. Computational biology: deep learning. *Emerging Topics in Life Sciences*, 1(3):257–274, 2017.

[18] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle. Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 2016.

[19] Z. Che, S. Purushotham, R.G. Khemani, and Y. Liu. Distilling knowledge from deep networks with applications to healthcare domain. *NIPS Workshop on Machine Learning for Healthcare (NIPS-MLHC)*, abs/1512.03542, 2015.

[20] Z.C. Lipton, D.C. Kale, C. Elkan, and R. Wetzell. Learning to diagnose with LSTM recurrent neural networks. *International Conference on Learning Representations (ICLR 2016)*, 2016.

[21] D. Bacciu, M. Colombo, D. Morelli, and D. Plans. Randomized neural networks for preference learning with physiological data. *Neurocomputing*, 2018.

[22] M.M. Al Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R.R. Yager. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, 345:340–354, 2016.

[23] G. Lima Guimaraes, B. Sánchez-Lengeling, P.L. Cunha Farias, and A. Aspuru-Guzik. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. *CoRR*, abs/1705.10843, 2017.

[24] F.M. Bianchi, K.O. Mikalsen, and R. Jenssen. Learning compressed representations of blood samples time series with missing data. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[25] C. Gallicchio, A. Micheli, and L. Pedrelli. Deep echo state networks for diagnosis of parkinson's disease. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[26] I. Fernández-Varela, D. Athanasakis, S. Parsons, E. Hernández-Pereira, and V. Moret-Bonillo. Sleep staging with deep learning: a convolutional model. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[27] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R.P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, pages 2224–2232. MIT Press, 2015.

[28] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In M.F. Balcan and K.Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2014–2023, New York, New York, USA, 20–22 Jun 2016. PMLR.

[29] H. Altae-Tran, B. Ramsundar, A.S. Pappu, and V. Pande. Low data drug discovery with one-shot learning. *ACS Central Science*, 3(4):283–293, 2017.

[30] W.L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. *CoRR*, abs/1706.02216, 2017.

[31] T. Van Dinh, N. Navarin, and A. Sperduti. Deep: decomposition feature enhancement procedure for graphs. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[32] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, and C.I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[33] D. Shen, G. Wu, and Suk H.-I. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19(1):221–248, 2017.

[34] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab. AggNet: Deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1313–1321, 2016.

[35] S. Postavaru, R. Stoean, C. Stoean, and G. Joya Caparros. Adaptation of deep convolutional neural networks for cancer grading from histopathological images. In I. Rojas, G. Joya, and A. Catala, editors, *Advances in Computational Intelligence*, pages 38–49, Cham, 2017. Springer International Publishing.

[36] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5):1207–16, 2016.

[37] H.R. Roth, C.T. Lee, H.C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R.M. Summers. Anatomy-specific classification of medical images using deep convolutional nets. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 101–104, April 2015.

[38] S. Pereira, A. Pinto, V. Alves, and C. A. Silva. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.

[39] G. Wu, M. Kim, Q. Wang, B.C. Munsell, and D. Shen. Scalable high-performance image registration framework by unsupervised deep feature representations learning. *IEEE Transactions on Biomedical Engineering*, 63(7):1505–1516, July 2016.

[40] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Scientific Reports*, 6, 2016.

[41] A.-R. Ali. Deep learning applications in medical imaging. TechEmergence, Accessed February 2018.

[42] H. Greenspan, B. van Ginneken, and R.M. Summers. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35:1153–1159, 2016.

[43] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Pérez, B. Lo, and G.Z. Yang. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, (21):4, 2017.

[44] J.M. Wolterink, T. Leiner, M.A. Viergever, and I. Išgum. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging*, 36(12):2536–2545, Dec 2017.

[45] U. Javaid and J.A. Lee. Capturing variabilities from computed tomography images with generative adversarial networks. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[46] M. Atencia and R. Stoean. Non-negative matrix factorization for medical imaging. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[47] N. Khanzhina. Pollen grain recognition using convolutional neural network. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[48] A. Vellido, J.D. Martín-Guerrero, and P.J.G. Lisboa. Making machine learning models interpretable. In *Proceedings of the 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012)*, pages 163–172.

[49] V. Van Belle and P.J.G. Lisboa. Research directions in interpretable machine learning models. In *Proceedings of the 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, pages 533–541.

[50] B. Goodman and S. Flaxman. European Union regulations on algorithmic decision making and a "right to explanation". *AI Magazine*, 3(38), 2017.

[51] Y. Dong, H. Su, J. Zhu, and F. Bao. Towards interpretable deep neural networks by leveraging adversarial examples. Technical report, arXiv preprint, arXiv:1708.05493, 2017.

[52] R. Schwartz-Ziv and N. Tishby. Opening the black box of deep neural networks via information. Technical report, arXiv preprint, arXiv:1703.00810v3, 2017.

[53] Z. Che, S. Purushotham, R. Khemani, and Y. Liu. Distilling knowledge from deep networks with applications to healthcare domain. Technical report, arXiv preprint, arXiv:1512.03542, 2015.

[54] M. Wu, M. Hughes, S. Parbhoo, and F. Doshi-Velez. Beyond sparsity: Tree-based regularization of deep models for interpretability. In *In: Neural Information Processing Systems (NIPS) Conference. Transparent and Interpretable Machine Learning in Safety Critical Environments (TIML) Workshop*, 2017.

[55] J.P. Amorim, I. Domingues, P.H. Abreu, and J. Santos. Interpreting deep learning models for ordinal problems. In *In: 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine learning (ESANN)*, 2018.

[56] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov. Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, 13, 2016.

[57] S.H. Yang, R.N. Van Gelder, and J.S. Pepose. Neural network computer program to determine photorefractive keratectomy nomograms. *Journal of Cataract & Refractive Surgery*, 24(7):917–924, 1998.

[58] I. Sturm, S. Lapuschkin, W. Samek, and K.R. Muller. Interpretable deep neural networks for single-trial EEG classification. *Journal of Neuroscience Methods*, 274:141–145, 2016.

[59] J. Adebayo, J. Gilmer, I. Goodfellow, and B. Kim. Local explanation methods for deep neural networks lack sensitivity to parameter values. In *Sixth International Conference on Learning Representations (ICLR 2018), workshop track*.

[60] G. Bhanot, M. Biehl, T. Villmann, and D. Zühlke. Biomedical data analysis in translational research: Integration of expert knowledge and interpretable models. In *Proceedings of the 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2017)*, pages 177–186.

[61] I. Bica, P. Velickovic, H. Xiao, and P. Li. Multi-omics data integration using cross-modal neural networks. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.

[62] J. Wülfing, S.S. Kumar, J. Boedecker, M. Riedmiller, and U. Egert. Controlling biological neural networks with deep reinforcement learning. In *Proceedings of the 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018)*.