An Approach to Explainable AI for Digital Pathology

J. M. Montes-Sánchez , L. Muñoz-Saavedra , F. Luna-Perejón , J. Civit-Masot , S. Vicente-Diaz , A. Civit
Robotics and Computer Technology Lab. Reina Mercedes s/n, E.T.S. Ing. Informática, Universidad de Sevilla
Sevilla, Spain. Email: {jmontes, luimunsaa, fralunper, jcivit, satur, civit}@atc.us.es

***Abstract*—Many medical diagnostics are based, at least, in part on medical imaging. The development of machine learning and, in particular Deep Learning (DL) based image processing in the last decade has led to the growth of diagnostic support aids based on these technologies. A problem regarding the adoption of this systems the lack of understandability of their diagnostic suggestions due to their Blackbox nature. Several approaches have been proposed to increase their explainability including evaluation of the internal layer contributions to outputs, network modifications to make these contributions more meaningful and model agnostic explanations. Medical systems are considered the paradigmatic case where understandability is of outmost importance. Digital Pathology (DP) is an especially difficult, but especially interesting case for image based diagnostic support aids. This is due, among other factors, to the fact that DP images are very large and multidimensional with the information not easily available at first sight. It is important to develop tools that let the pathologists apply their available knowledge easily while improving the diagnostic quality and their productivity. The design and evaluation of an interpretable digital pathology diagnosis aid would open the possibility for developing and deploying larger scale systems that would provide pathologists with reliable and trustworthy tools to help them in their daily diagnosis tasks.**

***Keywords*—Digital Pathology; Explainable Artificial Inteligence; Deep learning;**

I. INTRODUCTION

Many medical diagnostics are currently based, at least, in part on imaging technologies. Currently the interpretation of these images is, in most cases done almost directly by medical professionals. The great development of machine learning (ML) based image processing applications in the last decade has significantly increased the research on diagnostic support aids based on these technologies.

Medical image processing will experiment a breakthrough when this type of ML based diagnostic assistance tools became widely available and accepted by the medical community. Clearly one of the problems regarding the adoption of this type of systems is related to the lack of understandability of their diagnostic suggestions due to their Blackbox nature. This aspect is especially relevant in the case of medical diagnostic aids. This fact has been recently highlighted by several recent articles [1]. In a recent presentation[2] Carlos Guestrin, Senior Director of AI and Machine Learning at Apple considers the transition from black box to inclusivity as one of the four challenges of ML systems for the next few years. In all these cases medical systems are considered the paradigmatic case where understandability is of out-most importance. Some authors even consider that GDPR [3] includes the requirement that companies should be able to give users an explanation for decisions that this type of systems produce. However, in the medical case no real automatic decision making is envisioned as the tools are just diagnostic assistance and are never responsible for any final decision.

Several approaches have been proposed to increase the understandability of CNN based image processing ML systems. The three main approaches are:

- Understanding the internal layer results and their contribution to the system global outputs [4].
- Modifying the system architecture to make the internal layer results more meaningful [5].
- Using a “model agnostic” component that provides complementary explanations [6].

Additionally, there is always the possibility of constructing networks that look for the individual characteristics that doctors use to make a diagnosis. This solution, although less elegant from a scientific point of view, could currently make sense when developing a product with a short time to market. Digital Pathology is an especially difficult, but also especially interesting case. This is due to several factors [7]:

- Digital pathology is not just a transformation of the classical microscopic analysis of histological slides to digital visualization, it is an innovation that is changing medical workflows greatly;
- Much information is hidden in high dimensional spaces, not easily accessible at first sight, thus we need AI systems to help the pathologists in accessing and interpreting this data.
- The new workflows should provide ways in which pathologists can easily use their existing knowledge.

Thus, the possibility of designing and evaluating a small scale interpret able digital pathology image diagnosis aid would open the possibility for developing and deploying, in the near future, larger scale systems that would provide pathologists with reliable and trustworthy tools to help them in their daily diagnosis. These systems would also have a significant potential in the education of pathology students.

II. TECHNOLOGY BENCHMARKING

Among the most widely used ML methods in medical image analysis are support vector machines, random forests, and deep learning (DL). Due to its commercial success DL is currently the most popular framework in ML. Most mapping tasks from input images to an output images can be accomplished via DL given a large enough data set of well labelled training and testing examples. In the medical domain, very good results have been achieved for cancer detection with an accuracy that is similar to that achieved by an average pathologist. Some recent examples related to breast cancer include the results of the Chamelyon 16 challenge for identifying metastatic breast cancer. Some of the participants were able to obtain a Receiver Operating Characteristic (ROC) area under the curve above 92% with and error rate that was below 0.52% when used as an assistance tool by a human pathologist [8], [9] . Other important works have used the DL approach to accurately quantify the tumor extent [10] or to review the real impact of diagnostic tool assistance [11]. To our best knowledge all proposed DL approaches for digital image pathology tools are basically black-box models. However, in the medical domain it is necessary to be able to open this models to a glass-box and to make the results transparent and explainable on demand. The main result of our proposed tool and its success criteria would be to provide a causal explanation that provides useful information to the pathologist, e.g. areas in blue are considered a type X tumor because characteristics A and B are present. This would greatly improve the trustworthiness of the tool and its acceptability by the medical professionals.

III. CONCLUSIONS AND FUTURE WORK

The potential of understandable (explainable, visible, glass-box…) DL diagnostic aids in healthcare is huge. A recent article in Cell [12] emphasizes the importance of visible (as opposed to black-box) approaches to ML in biomedicine. Thus, the consortium considers that designing, implementing, testing and evaluating by medical professionals and students a small scale understandable digital pathology diagnostic aid could represent a major scientific and technological break-through in the field of software and integration for medical imaging. The importance of this ideas for the quality of life of citizens is also very significant as if pathology image diagnostic aids are up-taken this would lead to quicker and better diagnostics by currently heavy overloaded pathologists which would led to faster interventions and better medical prognosis. Last, but also of great importance, the project would lead to a better position of the industry that develop these solutions in the field of DL based-digital image pathology diagnostic aids in general and more specifically in the new field of understandable DL diagnostic aids.

REFERENCES

[1] W. Knight, "The dark secret at the heart of al," *Technology Review*, vol. 120, no. 3, pp. 54–61, 2017.

[2] A. Lauterbach, "Artificial intelligence and policy: quo vadis?" *Digital Policy, Regulation and Governance*, 2019.

[3] F. Pesapane, C. Volonté, M. Codari, and F. Sardanelli, "Artificial intelligence as a medical device in radiology: ethical and regulatory issues in europe and the united states," *Insights into imaging*, vol. 9, no. 5, pp. 745–753, 2018.

[4] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 3145–3153.

[5] C.-C. J. Kuo, M. Zhang, S. Li, J. Duan, and Y. Chen, "Interpretable convolutional neural networks via feedforward design," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 346–359, 2019.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[7] A. Holzinger, B. Malle, P. Kieseberg, P. M. Roth, H. Müller, R. Reihs, and K. Zatloukal, "Towards the augmented pathologist: Challenges of explainable-ai in digital pathology," *arXiv preprint arXiv:1712.06657*, 2017.

[8] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.

[9] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[10] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. N. Shih, J. Tomaszewski, F. A. González, and A. Madabhushi, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Scientific reports*, vol. 7, p. 46450, 2017.

[11] D. F. Steiner, R. MacDonald, Y. Liu, P. Truszkowski, J. D. Hipp, C. Gammage, F. Thng, L. Peng, and M. C. Stumpe, "Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer," *The American journal of surgical pathology*, vol. 42, no. 12, p. 1636, 2018.

[12] K. Y. Michael, J. Ma, J. Fisher, J. F. Kreisberg, B. J. Raphael, and T. Ideker, "Visible machine learning for biomedicine," *Cell*, vol. 173, no. 7, pp. 1562–1565, 2018.