



The prospect of artificial intelligence in the differential diagnosis of pancreatic cysts

Marco Montorsi, Giovanni Capretti

Department of General Surgery, IRCCS Humanitas Research Hospital, Humanitas University, Rozzano, Italy

Correspondence to: Prof. Marco Montorsi. Department of General Surgery, Humanitas Research Hospital, Humanitas University, Via Manzoni 56, 20089 Rozzano, MI, Italy. Email: marco.montorsi@humanitas.it; marco.montorsi@hunimed.eu.

Comment on: Kurita Y, Kuwahara T, Hara K, *et al.* Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions. *Sci Rep* 2019;9:6893.

Received: 23 August 2019; Accepted: 16 September 2019; Published: 12 October 2019.

doi: 10.21037/jmai.2019.09.05

View this article at: <http://dx.doi.org/10.21037/jmai.2019.09.05>

The title of the work we are here presenting “*Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions*” immediately strikes the interest of readers, especially of the ones performing pancreatic surgery (1). This article debates a theme of major concern for surgeons, the correct identification of a pancreatic cystic lesion, and a theme of major concern for the medical society and the society in general, the application of artificial intelligence (AI).

To understand the relevance of identifying correctly a cystic lesion it is enough to consider that some cystic lesions have no malignant potential and no prognostic impact on patient's life-span while others, such as main pancreatic duct intraductal mucinous neoplasms (IPMN), may become malignant in the majority of the cases with a prognosis similar to the one of pancreatic cancer (expected 5-year overall survival <10%) (2). For these reasons underestimating a pancreatic cyst is dangerous but also overestimating it can lead to deadly consequences if we consider that operative mortality of the more performed pancreatic surgical procedure, pancreaticoduodenectomy, reaches 7% in national dataset and more than 11% in low-volume centers (3). To add complexity to the problem, it has to be considered that prevalence of these lesions is rapidly growing, due to a larger use of diagnostic tool such as CT scan/MRI and to the progressive aging of population: in fact, prevalence of pancreatic cystic lesions ranges between 10% and 49% in the general population while it can reach 75% in the age range around 80 (4).

The application of AI in medicine is debated, however

it is gaining interest and this can raise clinical, legal and sociological problems. AI is already present in the everyday life of patients and clinicians. According to Cambridge dictionary definition: AI is the study of how to produce machines that have some of the qualities that the human mind has, such as the ability to understand language, recognize pictures, solve problems, and learn. Therefore when a doctor dictate the description of a surgical procedure or a medical record to a working station that write it down he is using a very simple form of AI. When a patient uses an app that helps him managing his insulin therapy or dietary requirement based upon glycemic values or other health data, he is using an AI. The programs of automatic translation or facial recognition, made popular by fictions, are application of AI. These are example of application of AI so widespread in our everyday experience that are commonly accepted and usually not categorized as an AI application.

In medicine, AI has different fields and purposes of application such as automated image screening, clinical decision support, predictive population risk stratification, and patient self-management tools.

The authors of this interesting paper focus their attention on the ability of an AI using deep learning analysis to correctly asses the malignant predictive value of a cystic lesion. This topic had already been explored in two previous papers. One from the same group, “Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas” was published in 2018 (5). In this case the researchers assessed the value of an AI in the evaluation of the malignant potential of different IPMNs utilizing only the images

derived from endoscopic ultrasonography (EUS) of 206 different patients. In this image-based analysis they found a sensitivity, specificity, accuracy and AUROC (Areas under the receiver operating curves) of respectively of 95.7%, 92.6%, 94% and 0.91 revealing an higher accuracy in comparison to human diagnostic ability (56.0%), the simple presence of mural node (68.0%) or a conventional statistical model based on logistic regression (72.0%) (5). The other paper, proposed by researchers of the Mayo Clinic and the University of Central Florida, also aimed stratifying the malignancy risk of IPMN but using, in this case, the image/data obtained through magnetic resonance; 139 cases were analyzed (6). In this paper, AI presented sensitivity and specificity of 75% and 78% respectively, comparable to the application of commonly used radiological criteria [American Gastroenterological Association (AGA) and Fukuoka guidelines]. AUROC was 0.76 (0.70–0.84) for AGA criteria, 0.77 (0.70–0.85) for Fukuoka, and 0.78 (0.71–0.85) for the AI deep learning protocol. Also this comparison confirms a similar performance between human evaluation through renowned guide lines and AI assessment. However, the authors stressed the fact that AI analysis required only 1.82 second per patient, an unparalleled result for a human radiologist (6).

In this second paper the Nagoya group expands the field of evaluation to all the type of pancreatic cystic lesions focusing mainly on the analysis of the cystic fluid. This could be another interesting field of application of AI because so far, no single marker dosed in pancreatic cyst showed a high sensibility and specificity, with CEA showing a 63% of sensibility and specificity (1).

The primary objective is the identification of those cystic lesions harboring malignant features or presenting with high grade dysplasia. Nevertheless, in this specific case the authors made a little, but, in our opinion, fundamental step further. As proposed in the previous article, they included in the input for the deep learning algorithm clinical data, such as patient's gender, cytological findings, site of the cyst, its characteristics and connection with the main pancreatic duct, partially mimicking a human clinical reasoning. The results are extremely interesting with an AI sensitivity, specificity, accuracy and AUROC (Areas under the receiver operating curves) of respectively of 95.7%, 91.9%, 92.9% and 0.96. AI performed better than CEA intra-cystic level alone and cytology accuracy which was respectively 60.9% and 47.8%. When assessing the malignancy risk rate of the different cystic lesions, AI was able to obtain an impressive ROC curve. No previous method has ever shown values so high representing a

new hope for all physicians dealing with this tricky pathology.

Unfortunately, at the same time, this study presents some severe methodological weak points, partially disclosed by the authors. The sample size is small (85 cases with only 23 malignant lesions) and it is a monocentric database. This became a major issue because the researchers utilized an internal validation with multiple resampling (learning and validation cohort) of the same population. This type of approach, that has been used also in the other articles assessing the value of AI on this topic, has been already criticized in previous clinical studies dealing with prognostic score and risk stratification for the problem of overfitting and low reproducibility of the result in external cohorts.

Moreover, the lower the number of the cases the higher is the risk of utilizing an internal resampling. The risk is that the AI “learned” very well how these 85 cystic lesions would behave but we don't know how well it could predict the behavior of other cysts. Also the primary end point, diagnosis of malignancy or high grade dysplasia of the cystic lesion, was not confirmed by its gold standard, histology. 26 cases were not resected and a 1-year follow-up is not comparable with pathological reports. We actually don't know the exact behavior and progression timing of high grade dysplasia in cystic lesion of the pancreas. In their previous work the authors identified as a weak point the selection bias generated by the inclusion of only resected patients. Including, among all the tested, only patients that underwent endoscopic needle aspiration of a cystic lesion but were not resected is also a selection bias. The study design harbors a clear risk of merging two completely different populations (resected and not resected) maintaining a selection bias and losing the possibility of measuring the primary outcome by the gold standard.

Not surprisingly, AI performed better than cytology or CEA intracystic values. These two terms of comparison are known to have weaknesses and in clinical practices they are only part of the elements utilized to formulate the diagnostic hypothesis (7). At the same time, this AI deep learning algorithm presents unexpected high values of accuracy and a AUROC never encountered in other papers dealing with the evaluation of pancreatic cyst malignant potential.

Part of these spectacular results could be ascribed to the increased spectrum of input, including biochemical, radiological and clinical findings, given to AI, establishing the superiority of an extended “clinical” evaluation over a single test. Nevertheless this article demonstrates the ability of AI to perform a sort of clinical reasoning, considering different aspect of a disease presentation, with astonishingly

good results. Further studies are required to confirm these results, and an external validation is needed but, for sure, a clear point has been stated.

Obviously, after the perfect stratification of the malignancy risk of a lesion, still the decision to resect or not should be made, so far, requiring a human intelligence. Nonetheless, this AI estimation could be a major help for clinicians. A future in which doctors will just take the therapeutic decision on data extracted automatically by the clinical records and elaborated by AI is not unrealistic or too distant. This would grant a reduction of the time required to clinicians for cases' evaluation and it would diminish the possibility of missing relevant details.

AI is being progressively applied in medicine and it is expected to be more and more utilized in the next years. A scenario in which AI carries no role at all or in which AI will completely substitute doctors and surgeons is so far unrealistic, and in our opinion, not desirable. A gradual implementation of AI in our practice as a useful tool, an interactive virtual assistant or an instrument of recheck could be a more alluring perspective.

The field of AI research was born at a workshop at Dartmouth College in 1956 (8). From that on, the concept of AI has greatly progressed, with enthusiastic and pessimistic phases, modulating its aims and application fields. In recent years AI application in medicine has become a relevant theme also due to the advent of the "big data" era. The amount of available information is growing exponentially and clinicians need help already to manage them, should it be by multidisciplinary approach, guidelines, nomograms or other. Also, the everyday amount of paperwork for clinicians and the weight of bureaucracy are growing without control stealing valuable time from high trained professionals. AI could be the perfect, welcomed, instrument to face these issues. At the same time AI scares doctors that fear to lose the control of the care process of patients they are responsible for and, taking it to the extreme, to be substituted by automatic system. Philosophically talking of "learning" and "neural network" could be extremely impressive even in our technological era. Is the "0" and "1" code of computers so different to the "all or nothing" activation of our neurons? At the end, as all others human inventions, the worth of AI will be probably defined by how we will use it.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Medical Artificial Intelligence*. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jmai.2019.09.05>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Kurita Y, Kuwahara T, Hara K, et al. Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions. *Sci Rep* 2019;9:6893.
2. European Study Group on Cystic Tumours of the Pancreas. European evidence-based guidelines on pancreatic cystic neoplasms. *Gut* 2018;67:789-804.
3. Balzano G, Capretti G, Callea G, et al. Overuse of surgery in patients with pancreatic cancer. A nationwide analysis in Italy. *HPB (Oxford)* 2016;18:470-8.
4. Kromrey ML, Bülow R, Hübner J, et al. Prospective study on the incidence, prevalence and 5-year pancreatic-related mortality of pancreatic cysts in a population-based study. *Gut* 2018;67:138-45.
5. Kuwahara T, Hara K, Mizuno N, et al. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol* 2019;10:1-8.
6. Corral JE, Hussein S, Kandel P, et al. Deep Learning to Classify Intraductal Papillary Mucinous Neoplasms Using

- Magnetic Resonance Imaging. *Pancreas* 2019;48:805-10.
7. Lévy P, Rebours V. The Role of Endoscopic Ultrasound in the Diagnosis of Cystic Lesions of the Pancreas. *Visc Med* 2018;34:192-6.
 8. Crevier D. AI: the tumultuous history of the search for artificial intelligence. New York: Basic Books, 1993.

doi: 10.21037/jmai.2019.09.05

Cite this article as: Montorsi M, Capretti G. The prospect of artificial intelligence in the differential diagnosis of pancreatic cysts. *J Med Artif Intell* 2019;2:21.