



Expectations of Artificial Intelligence for Pathology

Peter Regitnig¹(✉) , Heimo Müller^{1,2}, and Andreas Holzinger² 

¹ Diagnostic and Research Institute of Pathology, Medical University of Graz,
Neue Stiftingtalstraße 6, 8036 Graz, Austria
peter.regitnig@medunigraz.at

² Institute for Medical Informatics, Statistics and Documentation, Medical
University of Graz, Auenbruggerplatz 2/V, 8036 Graz, Austria

Abstract. Within the last ten years, essential steps have been made to bring artificial intelligence (AI) successfully into the field of pathology. However, most pathologists are still far away from using AI in daily pathology practice. If one leaves the pathology annihilation model, this paper focuses on tasks, which could be solved, and which could be done better by AI, or image-based algorithms, compared to a human expert. In particular, this paper focuses on the needs and demands of surgical pathologists; examples include: Finding small tumour deposits within lymph nodes, detection and grading of cancer, quantification of positive tumour cells in immunohistochemistry, pre-check of Papanicolaou-stained gynaecological cytology in cervical cancer screening, text feature extraction, text interpretation for tumour-coding error prevention and AI in the next-generation virtual autopsy. However, in order to make substantial progress in both fields it is important to intensify the cooperation between medical AI experts and pathologists.

Keywords: Artificial Intelligence · Medical AI · Machine learning · Digital pathology · Practical implementations of AI

1 Introduction and Motivation

1.1 What Is the Difference Between Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL)?

A very short answer is $DL \subset ML \subset AI$. For the sake of simplicity, we will use the term AI exclusively in this paper because of the common usage these days. We refer with this term to data-driven machine learning algorithms and, of course, deep learning approaches. The latter represent a currently particularly popular and successful family of learning algorithms [1]. To give the reader a good introduction, we will nevertheless briefly explain these terms.

Artificial Intelligence (AI) deals with the automation of intelligent behaviour with the long-term goal of reaching general intelligence (“strong AI”). This is

the reason why AI cannot be clearly defined because a precise definition of intelligence is lacking [22]. AI had always a close connection to cognitive science and is indeed a very old scientific field going back to the early work of Alan Turing [54]. After extremely high expectations to the classical logic-based AI between 1950 and 1980 the success was relatively low, resulting in a so-called “AI-winter”. From 2010 on, AI regained much interest due to practical success of statistical machine learning and especially the family of deep learning algorithms [23].

Machine Learning (ML) deals with the design, development and evaluation of algorithms that can learn from data, to gain knowledge from experience and improve their learning behaviour over time. Most of all it allows to infer unknowns and to make predictions to support decision making. One challenge is to discover relevant structural and/or temporal patterns (“knowledge”) in data, which is often hidden in arbitrarily high dimensional spaces, thus not accessible to a human expert [21]. Although ML is well defined, demonstrates much practical success, and is seen as the fastest growing technical field in computer science - ML is often summarized under the vague term AI. This may be understandable if we look at the challenges, which include sensemaking, context understanding, and decision making under uncertainty. We have to be fair to say that ML is the technical field that deals with the design, development and evaluation of algorithms, while AI covers also philosophical, social, ethical and juridical aspects, encompasses the underlying scientific theories of human learning vs. machine learning, and also explainability [20] and causability [24] which includes trust, fairness and ethical responsible decision making. Consequently, “Medical AI” is quite an acceptable term for the application of AI to study and pursue the effective use of ML for scientific problem solving and decision making motivated by efforts to safe and improve human health.

Deep Learning (DL) is one methodological family of ML based on, among others, artificial neural networks, deep belief networks, recurrent neural networks, etc. To give a precise example of a feed forward ANN: the multilayer perceptron [56], which is a very simple mathematical function mapping a set of input data to output data. The concept behind such ML models is representation learning, i.e. introducing other representations that are expressed in terms of simpler representations. Despite of the simplicity one huge shortcoming of these approaches is that due to their size they quickly become non-traceable, i.e. it becomes difficult to retrace and to understand why a certain result has been achieved. This makes the field of explainable AI a necessity [14].

AI, ML, DL, or however we call it, may help to solve certain problems, particularly in areas where humans have limited capacities (e.g. in high dimensional spaces, large numbers, big data, etc.). However, we must acknowledge that the problem-solving capacity of the human mind is still unbeaten in certain aspects (e.g. in the lower dimensions, when having little data, complex problems, etc. [42]). Consequently, it is sometimes desired to keep the human-in-the-loop [19], but it is indispensable to keep the doctor-in-control. For all these reasons integrated solutions are urgently needed in the future [25].

1.2 Is Pathology an Uncomplicated Medical Speciality?

In 2015 Levenson et al. [29] published a paper about pigeons, which differentiated between benign breast samples and malignant breast samples.

The excitement in popular media and also within the pathology community was enormous. Newsflashes about this paper were as follows: “Pigeons spot cancer as well as human experts”, “Pigeons to replace pathologists in diagnosing benign from malignant tumours”, and the most aggressive within these headlines came from NBC News: “Bird brain? Pigeons make good pathologists, study finds”.

Of course, such headlines make a pathologist frustrated and lead to increased aversions to all innovations in this field, despite that this was also critically discussed; for example [15] acknowledged that algorithms fall short of matching some basic human vision capabilities, i.e. when humans look at images they can interpret scenes and predict within seconds what is likely to happen after the picture is taken away (“object dynamic prediction”).

By looking at the pictures used in the study by Levenson et al. (2015), one feature is quite apparent, namely that benign samples were less cellular and showed less blue areas compared to the malignant samples. If one takes this single, but very typical characteristic as a diagnostic feature, many breast samples would be misdiagnosed within a few days, but at least for a pigeon, it makes a reasonable success rate. Nevertheless, this work highlighted a phenomenon that even small neuronal networks (the “pigeon brain”) have enough cognitive potential to differentiate between pictures which show less blue and which show bluer, the latter is often coupled to malignancy due to its higher cellularity. Compared to AI generally and neuronal networks particularly, this has something in common: we are also not able to ask a pigeon *why* a certain characteristic it recognised to differentiate between benign and malignant and so we are not able to recognise in most neuronal network architectures, i.e. which features were the most weighted. Although there are already some methods of explainable AI [46] to overcome these problems, in many clinical situations, AI is still kind of black box – lacking interpretability by the human expert. Current pathology text-books contain hundreds of pages per organ and around 10 to 20 typical morphological features per entity. Moreover, if all the requirements and challenges a pathologist faces day by day [26], it is evident that in the near future not a single neuronal network will be able to recognise all the details which clinicians need for further therapy and to show contextual understanding. On specific questions, however, support by AI can be expected, but for the next years it is rather sure that pathologists will be neither replaced by pigeons nor by AI. That does not mean that the workflows might be optimized and that integrated solutions appropriately designed to support effective human-AI interaction will be beneficial to reduce time-pressure, enhance quality, and enable interpretability which would also bring huge benefits in the education and training of pathologists, which are urgently needed.

2 Glossary

Pathology: Discipline in medicine, which diagnoses diseases from tissue, which is obtained by biopsy or surgical procedure. Pathology uses many biomedical techniques, and the two mainly used are microscopy and immunohistochemistry.

Immunohistochemistry: is a biomedical technique to microscopically visualize highly specific certain proteins to determine tissue-typical protein expression for differential diagnosis or treatment decisions.

Breast Cancer: A malignant neoplasm or tumour evolves from the breast gland tissue and which is capable of spreading in local lymph nodes, i.e. lymph node metastasis or via blood vessels in distant organs, i.e. distant metastasis.

Malignant Melanoma: A malignant neoplasm evolves from pigment cells of skin or mucosa and which can spread to lymph nodes or in distant organs.

Lymph Node Metastases: Deposits of malignant tumour cells in regional lymph nodes. Lymph node metastases are crucial prognostic information, because it is often coupled to a systemic tumour disease and indicating later appearing distant metastases.

Sentinel Lymph Node: A specific lymph node, which is the first in a region of a nearby malignant tumour. Sentinel lymph nodes indicate whether other regional lymph nodes might contain metastases, too. Instead of full lymph-node dissection nowadays only the Sentinel lymph node is surgically removed to diminish adverse effects, like lymphoedema.

TNM System: A system describing the extent of a malignant tumour. T stands for local tumour extend, N for lymph node metastases, M for distant metastases. Each cancer patient is staged for the tumour extend at the time of diagnosis to provide prognostic information, which influences the further treatment.

Tumour Grading: A system describing the maturation or similarity of a malignant tumour compared to its normal-tissue counterpart. Grading is a morphological evaluation of architecture of the tumour and its cytological cell features. It ranges from Grade 1 (i.e. well-differentiated) to Grade 3 (i.e. poorly differentiated). Grading also gives important information for prognosis and thus further treatment decision.

Gleason Grading: Tumour grading in prostate cancer facilitates a particular form of Grading, mainly focusing on the architecture of neoplastic glands. This pattern recognition is 5-tired and based on synoptic drawings. For example, Gleason grade 1 is defined as uniform, well circled, evenly distributed glands; grade 4 contains cribriform, i.e. sieve-like structures and grade 5 has solid or single cells unevenly distributed.

Papanicolaou Test: Epithelial cells from the cervix uteri are stained with the Papanicolaou technique and microscopically evaluated for atypical cells. The Pap-test is the most efficient cervical cancer preventive test over the last decades. Nowadays, it is replaced or preceded by the more sensitive HPV test, which focuses on the underlying HPV infection.

3 State-of-the-Art

3.1 The Position of AI in Pathology Today?

Within the last ten years, significant steps have been made to make AI relevant in the field of pathology. Nevertheless, we are currently far away from using AI in daily pathology practice. Scanning and storage of whole slides became quite affordable, but the applications for using AI are still in the field of basic research and highly diverse. During the last decade, machine learning got applicable on whole slide images, and thus AI is also called the third revolution in pathology, after the invention of immunohistochemistry and next-generation sequencing [5,32,38,45].

When one compares the development of AI in pathology to the development of self-driving cars, we see a very long time passing by between the first cars which had adaptive cruise controls in the 1950ies; and even nowadays an autopilot is not standard, as we would have expected some years ago. One reason for that is the fatality-risk of an error caused by the autonomous car; although errors by *human* drivers result in enormous death rates. For example in the US there were in 2014 over 6.1 million reported collisions, 94% attributed to driver error [41]. Consequently, autonomous vehicles have a huge potential to dramatically reduce the contribution of driver error and negligence as the cause of vehicle collisions.

The same is true for a diagnostic error in pathology, where misdiagnosis leads to under- or overtreatment. Even for the non-medical community, it might be evident that chemotherapy for a non-cancer patient is fatal. Therefore in both systems, certainty and reliability are the essential features to be reached and to be tested.

Who really thinks that driving a car is as complex as pathology has no idea how complicated biological tissues can be. The learning curve of a pathologist in training looks like this: To become a good pathologist, it takes about 5 to 6 years to study general medicine, after that 2 to 6 years of residency (depending on national regulations) and after that, another 4 to 10 years to become a good pathologist with a broad knowledge about all possible diseases and differential diagnosis. Becoming an expert in a field, e.g. gynecopathology or neuropathology, takes even more years. General pathologists knowing “everything” do unfortunately not exist anymore compared to 20–30 years ago, because medical knowledge increased so dramatically and the organ-specialisation went in parallel with specialisation in other medical fields. Coming back to the comparison with automatic car-driving: Acquiring a driving license takes a few weeks, probably this comparison of learning times makes it more evident that interpreting morphology in biological tissues is a huge task.

Medical AI will also have great potential to support successfully, but will be similarly or even more dependent on the solution of ethical and legal problems [47].

3.2 Where Could Pathologists Need Support by AI?

Beside the limitations mentioned above, some applications are useful due to their highly specific scope. AI will support the daily work of pathologists within the next few years. Here are some of these summarised as good practice examples.

Finding Small Tumour Deposits Within Lymph Nodes. For cancer therapy, it is crucial to recognise lymph node metastases (see Fig. 1). Some of these are big enough to be seen by naked eye or on radiological imaging, but some require a thorough work-up of the tissue by step sectioning in 200 μm distance or less, resulting in 15 to 20 levels on glass slides for a single lymph node, the so-called sentinel-lymph-node evaluation. This procedure is most often required in breast cancer or malignant melanoma [13,36], but it is also used in other tumour entities nowadays [6].

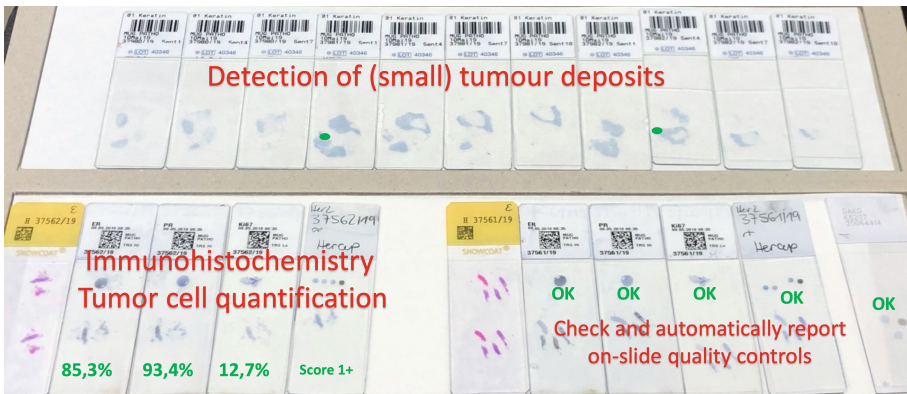


Fig. 1. Glass slides of three pathology cases of breast cancer samples. Upper row: Sentinel lymph node sections immunohistochemically stained with anti-keratin. Lower row: Haematoxylin and Eosin (yellow labels) and immunohistochemically stained (white labels) sections of breast cancer biopsies illustrating potentially supportive quantification of immuno-stained tumour cells, on-slide controls and detection of tumour deposits in lymph nodes. (Color figure online)

This evaluation is laborious, and error-prone due to highly repetitive images seen and a low number of recognisable events, or better said a low number of small tumour deposits. It is a good example where AI can bring pathology to the next level in daily routine work. Applications for metastasis detection in lymph nodes were published in recent years already [2,30].

The main task for AI in this field should be the detection of (small) tumour cells aggregates as a screening method before the stained slides of a case will be seen by a human pathologist. However, the final decision about a metastasis or an artefact should stay in the hands of humans, unless real good trained image recognition software will be able to do the same task equal or better with very high specificity. A high number of cases is needed to train and further to prove the software in a validation study. The first level to reach is a high sensitivity of detection, which can reduce the work-load of pathologists.

Detection and Grading of Cancer. One of the main tasks in pathology is the detection of cancer in tissues removed by biopsy or surgery. The first big question is whether a patient has a tumour or not. In radiological imaging, inflammatory changes often mimic neoplastic growth and thus have to be diagnosed microscopically by pathologists in a certain way.

Within neoplasms, pathologists differentiate between benign, premalignant and malignant and do a further refinement of the tumours, called tumour typing. For malignant tumours, this is done according to the WHO tumour typing [8]. Furthermore, all malignant tumour types are categorized according to the likelihood of their normal tissue counterpart. This is the so-called Grading, where mostly architectural and cytological details are considered. Grading ranges from well-differentiated (G1) to poorly differentiated (G3) and in some instances undifferentiated (G4) exists, too. For each neoplasm, a specific definition of the tumour grades exists. In most tumour types, the grading definitions and microscopical appearance cannot be transferred from one to the other. Tumour-typing and Grading are two of the hallmarks for tumour prognostication and do have direct influence to further treatment decisions. For example, well-differentiated (G1) malignant tumours do have a favourable prognosis and therefore need less often chemotherapy treatment compared to poorly differentiated (G3) tumours.

Thus, AI has a full limitation when trying to generalize in these fields. However, in recent years, good examples of tumour detection algorithms, particularly using deep learning approaches, and its specific Gleason grading [10,11] have been achieved in prostate cancer [3,28,35,43,57].

In breast cancer, counting of mitotic figures is essential besides tubule formation and pleomorphism of nuclei for grading. However, mitotic counting is time-consuming and error-prone. Recently, a method to automatically detect mitotic figures in breast cancer tissue sections based on convolutional neural networks was published [51]. These are good examples that machine learning for a particular type of tumour can be reached successfully in a distinct question, but compared to the number of tumour types, it is still a tiny application. Cancers occur in most types of cells; compared with the 300 or so different types of cells in the human body, we can recognize 200 different types of human cancers [31] and countless variants. It must be recognized that each tumour is highly individual and has distinct biological and thus also morphological features.

Quantification of Positive Tumour Cells in Immunohistochemistry.

Another example of support in pathology will be the exact quantification of positive tumour cells in immunohistochemistry. In daily pathology practice, counting is too time-consuming and seen in the light of relevance not feasible. Therefore, real counting is most times replaced by estimation in a semi-quantitative manner or counted in a fraction of cells. It has been shown that quantification is also possible by relative simple morphometry [12,53], but the limitation is to differentiate tissue types, like malignant and normal epithelial structures or differentiation between invasive and non-invasive tumour cells and surrounding stromal tissue. Machine learning can achieve such a tissue-type differentiation in advance of the final quantification. However, one may never forget, that even such technically sophisticated procedure might only produce pseudo-precision because most tumours show an intra-tumour heterogeneity, which is typical for biological structures.

Checking of On-Slide Quality Controls. Regarding immunohistochemistry, another application of artificial intelligence could be checking of on-slide quality controls. To follow the accuracy of immunohistochemistry additional tissue pieces or cultured cell-lines are placed and stained on the same glass slide as the patient's tissue and are visually checked in the microscope for true positive-, negative-staining and staining intensity in the expected cell types. Such measurements over several runs and several days could be achieved by image similarity measurements, as already shown for histopathology [16].

Pre-check of Papanicolaou-Stained Gynaecological Cytology in Cervical Cancer Screening.

An exciting application for machine learning for microscopy could be a pre-check of Papanicolaou stained gynaecological cytology. The Pap-test is used to detect atypical epithelial cells, which are a precursor of cervical cancer (see Fig. 2). The cells are obtained from the cervix uteri in a non-invasive way. After that directly placed on a glass slide or in a liquid for further processing and staining. Specialised trained biomedical technicians and doctors search for atypical cells, an event occurring in around 1–3 % of cases and seen only in a small fraction of all cells on the glass-slide. On each glass slide, around 20.000 epithelial cells have to be inspected and in a positive case around 10–1000 with a wide variation can be found. Such numbers make it evident that the detection rate is only reaching 50–70%. Based on classical morphometry image analysis, several companies worked already on this topic [4,7,9,27].

Since a few years, machine learning algorithms have been implemented in this field, e.g. DeepPap where the authors implemented a deep convolutional network [58], or the work of [49], where the authors applied a Mask Regional Convolutional Neural Network (Mask R-CNN) to cervical cancer screening using pap smear histological slides. A current review [55] indicated that there are still weaknesses in the available techniques, resulting in low accuracy of classification in some classes of cells and that a shortcoming is that most of the existing algorithms work either on single or on multiple cervical smear images.

However, not only screening for atypical cells could be an application, but also the first level of assessment could be done by a AI system, namely to count how many cells are within a PAP smear and if there are endocervical cells are present or not. All these features are important for interpreting the adequacy of a Pap smear, which underlines the repetitiveness and certainty of a given cell-sample. In the second level of assessment, algorithms could highlight atypical cells among thousands of normal cells within one Pap smear.

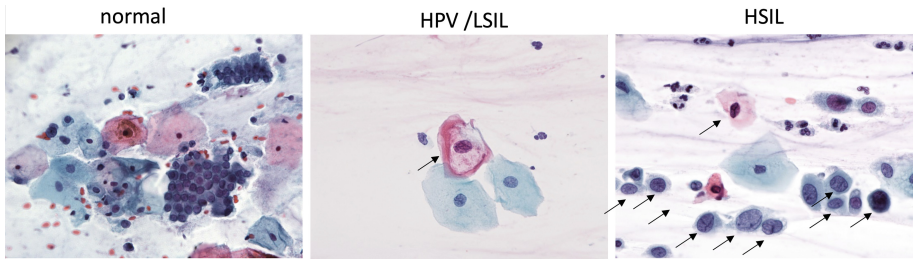


Fig. 2. Papanicolaou-stained cells from the cervix uteri in different appearances: left: different types of normal cells (squamous and glandular epithelial cells and some erythrocytes); middle: one atypical cell of HPV infection (cell with a halo in the cytoplasm, Koilocyte), which is named low grade squamous intraepithelial lesion (LSIL); right shows several squamous cells with high grade atypia, which is named high grade squamous intraepithelial lesion (HSIL). LSIL and HSIL are precursor lesions of invasive cervical cancer.

Text Feature Extraction. A completely different field of medical AI is not connected to images but is connected to pathology-reports, which pathologists typically tend to make like an essay [37, 44, 50]. There are manifold useful applications here, e.g. AI can help to structure or extract specific text parts from routine pathology reports for further scientific purposes. Especially all the reports could be better used for any scientific purpose if it would be easier to search for different disease entities. This could increase the value of millions of biospecimens which are currently stored in pathology archives or biobanks [33, 34].

Text Interpretation and Coding Error Prevention. Another possibility for text-interpreting software could be the preventive error correction. Nowadays errors can occur in classifications during a hard-working day, for example within the TNM system or the tumour grading system. Software running in the background of laboratory information systems might be able to highlight an error of misconduct instantly. For example, instead of well-differentiated, invasive breast carcinoma (NST), G2 it should immediately highlight that “well differentiated” and G2 is incompatible. As a training set thousands, if not millions of already written reports could be taken to learn what is correct and what is outside of this reasonable conduct.

Medical AI in Next-Generation Autopsy. An autopsy is a long and historical field of pathology tasks. Also, in this field, new imaging techniques can be observed. CT or MRI imaging is already used in virtual autopsies and by using such a technique, image analysis comes into the field again. For example, machine learning provides new opportunities for investigating minimal tumour burden and therapy resistance in cancer patients and can lead to an enhanced or augmented autopsy [39,40].

4 Open Problems

In conclusion some specific problems hindering the progress of digital pathology and machine learning can be summarized:

User interfaces are far away from being ergonomic, thus usability engineering [18] is for future Human-AI more necessary than ever. A wise way would be the integration of digital pathology and AI applications together within the laboratory information systems (LIS). On the other hand, a decoupling of such systems is desirable in order to make legal aspects manageable, such as liability via medical device laws. This requires standardized interfaces and thus enables a modular extension of existing workflows via “AI-Apps”. It would also be necessary that measurements would be directly transferred back to the LIS. Such advantages in ergonomics would make the acceptance enormously high and enable trustworthy AI integration.

Nowadays, several proprietary image formats exist among scanner vendors. A single image format like DICOM [17,52,59], which is in development for pathology whole slide images, is necessary for the widespread use and also for consultation purposes between different institutions and standardised exchange of data with neuronal network algorithms. Also, software licensing has to be rethought by vendors. It will be impossible for pathology institutes to purchase several software systems for a wide range of diagnostic reasons and different tissue-specific applications. A way out of this could be a pay-per-case model. A technical issue which might be solved within the next years is the large image file sizes, but technical advances could solve compression and also storage issues. By the way, simple greyscale images, still provide much information for pathologists due to structural preservation.

Finally, the most important and probably obvious problem is that there is too little interaction between AI scientists on the one side and pathologists on the other side. Imagine an AI scientist developing a self-driving car without knowing how to drive a car, which is pretty impossible. Currently, we are in the situation that AI scientists sometimes produce fascinating morphology-interpretation software, but they typically lack knowledge about morphology or the clinical value about image and text interpretation. On the other hand, pathologists, although often very much interested in this topic, lack time, due to shortage of personal and increasing specialisation of organ-specific pathology with sometimes overwhelming daily work-load.

5 Future Outlook

Nowadays, we face an innovative and exciting technique, which will surely find its way in the pathology departments. Some of the advantages are set in the framing of whole slide imaging and digital pathology in a strict sense: By scanning all the thousands of slides in a pathology lab, there will be not so much lost or broken slides anymore. It is expected that there is less time wasting for organisation of cases by bringing together all slides. A tremendous advantage in digital pathology is quick access to archived slides of one patient for comparison of older slides to the actual slides and also to get quicker access to second opinions worldwide [48].

Residential teaching can facilitate typical cases, which will be stored and collected digitally and could be shared worldwide. Measurements of tumour sizes or distances to resection margins will be more objective compared to classical microscopy.

Furthermore, new AI applications will support and improve the quantification, especially in immunohistochemistry, and it might prevent pathologists from typical text-errors. Moreover, the hope exists that machine learning will bring a lot of new insights in tumour prognostication because certain morphological features could be highlighted by unsupervised machine learning only, which humans would hardly recognise at all.

In the author's view, the intensified cooperation between AI scientists and pathologists would bring enormous progress, and we hope that AI will support the next generation of pathologists like computers support pilots in a plane today. Due to the enormous complexity of biology and pathology of tissues, pathologists will not be replaced by AI in the near future, however, such thoughts increase fear and might interfere with further progress.

Acknowledgements. The authors declare that there are no conflicts of interests and the work does not raise any ethical issues. Parts of this work has been funded by the Austrian Science Fund (FWF), Project: P-32554 "A reference model of explainable Artificial Intelligence for the Medical Domain", and by the European Union's Horizon 2020 research and innovation program under grant agreements No 824087 "EOSC-Life" and No 826078 "Feature Cloud". We thank the anonymous reviewers for their critical but helpful comments.

References

1. Arvaniti, E., et al.: Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Sci. Rep.* **8**(1), 1–11 (2018). <https://doi.org/10.1038/s41598-018-30535-1>
2. Bandi, P., et al.: From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE Trans. Med. Imaging* **38**(2), 550–560 (2018). <https://doi.org/10.1109/TMI.2018.2867350>
3. Bulten, W., et al.: Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* (2020). [https://doi.org/10.1016/S1470-2045\(19\)30739-9](https://doi.org/10.1016/S1470-2045(19)30739-9)

4. Cengel, K.A., et al.: Effectiveness of the surepath liquid-based pap test in automated screening and in detection of hsl. *Diagn. Cytopathol.* **29**(5), 250–255 (2003). <https://doi.org/10.1002/dc.10373>
5. Chang, H.Y., et al.: Artificial intelligence in pathology. *J. Pathol. Transl. Med.* **53**(1), 1–12 (2019). <https://doi.org/10.4132/jptm.2018.12.16>
6. Cibula, D., McCluggage, W.G.: Sentinel lymph node (SLN) concept in cervical cancer: current limitations and unanswered questions. *Gynecol. Oncol.* **152**(1), 202–207 (2019). <https://doi.org/10.1016/j.ygyno.2018.10.007>
7. Duggan, M.A., Brasher, P.: Paired comparison of manual and automated pap test screening using the papnet system. *Diagn. Cytopathol.* **17**(4), 248–254 (1997)
8. Ellis, I., et al.: The 2019 who classification of tumours of the breast. *Histopathology* (2020). <https://doi.org/10.1111/his.14091>
9. Elsheikh, T.M., Austin, R.M., Chhieng, D.F., Miller, F.S., Moriarty, A.T., Renshaw, A.A.: American society of cytopathology workload recommendations for automated pap test screening: developed by the productivity and quality assurance in the era of automated screening task force. *Diagn. Cytopathol.* **41**(2), 174–178 (2013). <https://doi.org/10.1002/dc.22817>
10. Epstein, J.I., Amin, M.B., Reuter, V.E., Humphrey, P.A.: The 2014 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am. J. Surg. Pathol.* **40**(2), 244–252 (2017). <https://doi.org/10.1097/PAS.0000000000000530>
11. Epstein, J.I., Amin, M.B., Reuter, V.E., Humphrey, P.A.: Contemporary gleason grading of prostatic carcinoma: an update with discussion on practical issues to implement the 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *Am. J. Surg. Pathol.* **41**(4), e1–e7 (2017). <https://doi.org/10.1097/pas.0000000000000820>
12. Fulawka, L., Halon, A.: Proliferation index evaluation in breast cancer using imagej and immunoratio applications. *Anticancer Res.* **36**(8), 3965–3972 (2016)
13. Garcia-Etienne, C.A., et al.: Management of the axilla in patients with breast cancer and positive sentinel lymph node biopsy: an evidence-based update in a european breast center. *Eur. J. Surg. Oncol.* **46**(1), 15–23 (2020). <https://doi.org/10.1016/j.ejso.2019.08.013>
14. Goebel, R., et al.: Explainable AI: the New 42? In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) *CD-MAKE 2018*. LNCS, vol. 11015, pp. 295–303. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99740-7_21
15. Granter, S.R., Beck, A.H., Papke, D.J.: Alphago, deep learning, and the future of the human microscopist. *Arch. Pathol. Lab. Med.* **141**(5), 619–621 (2017). <https://doi.org/10.5858/arpa.2016-0471-ED>
16. Hegde, N., et al.: Similar image search for histopathology: smily. *NPJ Digit. Med.* **2**(1), 1–9 (2019). <https://doi.org/10.1038/s41746-019-0131-z>
17. Herrmann, M.D., et al.: Implementing the DICOM standard for digital pathology. *Journal of pathology informatics* **9**, 37 (2018)
18. Holzinger, A.: Usability engineering methods for software developers. *Commun. ACM* **48**(1), 71–74 (2005). <https://doi.org/10.1145/1039539.1039541>
19. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform.* **3**(2), 119–131 (2016). <https://doi.org/10.1007/s40708-016-0042-6>
20. Holzinger, A.: From machine learning to explainable AI. In: *2018 World Symposium on Digital Intelligence for Systems and Machines (IEEE DISA)*, pp. 55–66. IEEE (2018). <https://doi.org/10.1109/DISA.2018.8490530>

21. Holzinger, A.: Introduction to machine learning and knowledge extraction (make). *Mach. Learn. Knowl. Extr.* **1**(1), 1–20 (2019). <https://doi.org/10.3390/make1010001>
22. Holzinger, A., Kickmeier-Rust, M., Müller, H.: KANDINSKY patterns as IQ-test for machine learning. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2019. LNCS, vol. 11713, pp. 1–14. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29726-8_1
23. Holzinger, A., Kieseberg, P., Weippl, E., Tjoa, A.M.: Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2018. LNCS, vol. 11015, pp. 1–8. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99740-7_1
24. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H.: Causability and explainability of artificial intelligence in medicine. *WIREs Data Min. Knowl.* **9**(4), e1312 (2019). <https://doi.org/10.1002/widm.1312>
25. Holzinger, A., et al.: Machine learning and knowledge extraction in digital pathology needs an integrative approach. In: Holzinger, A., Goebel, R., Ferri, M., Palade, V. (eds.) *Towards Integrative Machine Learning and Knowledge Extraction*. LNCS (LNAI), vol. 10344, pp. 13–50. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-69775-8_2
26. Kargl, M., Regitnig, P., Mueller, H., Holzinger, A.: TOWARDS A BETTER UNDERSTANDING OF THE WORKFLOWS: MODELING PATHOLOGY PROCESSES IN VIEW OF FUTURE AI INTEGRATION. In: Springer LNCS, vol. 12090, p. 16. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50402-1_7
27. Keyhani-Rofagha, S., Palma, T., O’Toole, R.V.: Automated screening for quality control using PAPNET: a study of 638 negative Pap smears. *Diagn. Cytopathol.* **14**(4), 316–320 (1996)
28. Kott, O., et al.: Development of a deep learning algorithm for the histopathologic diagnosis and gleason grading of prostate cancer biopsies: a pilot study. *Eur. Urol. Focus* (2020). <https://doi.org/10.1016/j.euf.2019.11.003>
29. Levenson, R.M., Krupinski, E.A., Navarro, V.M., Wasserman, E.A.: Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS ONE* **10**(11), e0141357 (2015). <https://doi.org/10.1371/journal.pone.0141357>
30. Liu, Y., et al.: Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch. Pathol. Lab. Med.* **143**(7), 859–868 (2019). <https://doi.org/10.5858/arpa.2018-0147-OA>
31. Lodish, H., Berk, A., Zipursky, S., Matsudaira, P., Baltimore, D., Darnell, J.: Tumor cells and the onset of cancer. In: *Molecular Cell Biology*, 4th ed. Freeman, New York (2000)
32. Madabhushi, A., Lee, G.: Image analysis and machine learning in digital pathology: challenges and opportunities. *Med. Image Anal.* **33**, 170–175 (2016). <https://doi.org/10.1016/j.media.2016.06.037>
33. Müller, H., Dagher, G., Loibner, M., Stumptner, C., Kungl, P., Zatloukal, K.: Biobanks for life sciences and personalized medicine: importance of standardization, biosafety, biosecurity, and data management. *Curr. Opin. Biotechnol.* **65**, 45–51 (2020)
34. Müller, H., et al.: State-of-the-art and future challenges in the integration of biobank catalogues. In: Holzinger, A., Röcker, C., Ziefle, M. (eds.) *Smart Health*. LNCS, vol. 8700, pp. 261–273. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16226-3_11

35. Nagpal, K., et al.: Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. *Nat. Digit. Med. (NPJ)* **2**(1), 1–10 (2019). <https://doi.org/10.1038/s41746-019-0112-2>
36. Nakamura, Y.: The role and necessity of sentinel lymph node biopsy for invasive melanoma. *Front. Med.* **6**(231), 1–7 (2019). <https://doi.org/10.3389/fmed.2019.00231>
37. Napolitano, G., Marshall, A., Hamilton, P., Gavin, A.T.: Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction. *Artif. Intell. Med.* **70**, 77–83 (2016). <https://doi.org/10.1016/j.artmed.2016.06.001>
38. Niazi, M.K., Parwani, A.V., Gurcan, M.N.: Digital pathology and artificial intelligence. *Lancet Oncol.* **20**(5), e253–e261 (2019). [https://doi.org/10.1016/S1470-2045\(19\)30154-8](https://doi.org/10.1016/S1470-2045(19)30154-8)
39. O’Sullivan, S., Holzinger, A., Wichmann, D., Saldiva, P.H.N., Sajid, M.I., Zatloukal, K.: Virtual autopsy: machine learning and AI provide new opportunities for investigating minimal tumor burden and therapy resistance by cancer patients. *Autops. Case Rep.* **8**(1), e2018003 (2018). <https://doi.org/10.4322/acr.2018.003>
40. O’Sullivan, S., Holzinger, A., Zatloukal, K., Saldiva, P., Sajid, M.I., Dominic, W.: Machine learning enhanced virtual autopsy. *Autops. Case Rep.* **7**(4), 3–7 (2017). <https://doi.org/10.4322/acr.2017.037>
41. Paden, B., Cap, M., Yong, S.Z., Yershov, D., Frazzoli, E.: A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Trans. Intell. Veh.* **1**(1), 33–55 (2016). <https://doi.org/10.1109/TIV.2016.2578706>
42. Pohn, B., Kargl, M., Reihls, R., Holzinger, A., Zatloukal, K., Müller, H.: Towards a deeper understanding of how a pathologist makes a diagnosis: Visualization of the diagnostic process in histopathology. In: *IEEE Symposium on Computers and Communications (ISCC)*. IEEE (2019). <https://doi.org/10.1109/ISCC47284.2019.8969598>
43. Poojitha, U.P., Sharma, S.L.: Hybrid unified deep learning network for highly precise gleason grading of prostate cancer. In: *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 899–903. IEEE (2019). <https://doi.org/10.1109/EMBC.2019.8856912>
44. Reihls, R., Pohn, B., Zatloukal, K., Holzinger, A., Müller, H.: NLP for the generation of training data sets for ontology-guided weakly-supervised machine learning in digital pathology. In: *2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1072–1076. IEEE (2019)
45. Salto-Tellez, M., Maxwell, P., Hamilton, P.: Artificial intelligence—the third revolution in pathology. *Histopathology* **74**(3), 372–376 (2019). <https://doi.org/10.1111/his.13760>
46. Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (eds.): *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. LNCS (LNAI), vol. 11700. Springer, Cham (2019). <https://doi.org/10.1007/978-3-030-28954-6>
47. Schneeberger, D., Stoeger, K., Holzinger, A.: The European legal framework for medical AI. In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Fourth IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Proceedings*, p. in print. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-29726-8>
48. Smeulders, A., Van Ginneken, A.: An analysis of pathology knowledge and decision making for the development of artificial intelligence-based consulting systems. *Anal. Quant. Cytol. Histol.* **11**(3), 154–165 (1989)

49. Sompawong, N., et al.: Automated pap smear cervical cancer screening using deep learning. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE (2019). <https://doi.org/10.1109/EMBC.2019.8856369>
50. Tang, R., et al.: Machine learning to parse breast pathology reports in Chinese. *Breast Cancer Res. Treat.* **169**(2), 243–250 (2018). <https://doi.org/10.1007/s10549-018-4668-3>
51. Tellez, D., et al.: Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. *IEEE Trans. Med. Imaging* **37**(9), 2126–2136 (2018). <https://doi.org/10.1109/TMI.2018.2820199>
52. Tuominen, V.J., Isola, J.: Linking whole-slide microscope images with DICOM by using JPEG2000 interactive protocol. *J. Digit. Imaging* **23**(4), 454–462 (2010). <https://doi.org/10.1007/s10278-009-9200-1>
53. Tuominen, V.J., Ruotoistenmäki, S., Viitanen, A., Jumppanen, M., Isola, J.: Immunoratio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res.* **12**(4), 1–12 (2010). <https://doi.org/10.1186/bcr2615>
54. Turing, A.M.: Computing machinery and intelligence. *Mind* **59**(236), 433–460 (1950). <https://doi.org/10.1093/mind/LIX.236.433>
55. William, W., Ware, A., Basaza-Ejiri, A.H., Obungoloch, J.: A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Comput. Methods Programs Biomed.* **164**, 15–22 (2018). <https://doi.org/10.1016/j.cmpb.2018.05.034>
56. Yan, H., Jiang, Y., Zheng, J., Peng, C., Li, Q.: A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Syst. Appl.* **30**(2), 272–281 (2006). <https://doi.org/10.1016/j.eswa.2005.07.022>
57. Yang, Q., et al.: Epithelium segmentation and automated gleason grading of prostate cancer via deep learning in label-free multiphoton microscopic images. *J. Biophotonics* **13**(2), e201900203 (2019). <https://doi.org/10.1002/jbio.201900203>
58. Zhang, L., Lu, L., Nogues, I., Summers, R.M., Liu, S., Yao, J.: Deeppap: deep convolutional networks for cervical cell classification. *IEEE J. Biomed. Health Inf.* **21**(6), 1633–1643 (2017). <https://doi.org/10.1109/JBHI.2017.2705583>
59. Zwoenitzer, R., Kalinski, T., Hofmann, H., Roessner, A., Bernarding, J.: Digital pathology: DICOM-conform draft, testbed, and first results. *Comput. Methods Programs Biomed.* **87**(3), 181–188 (2007). <https://doi.org/10.1016/j.cmpb.2007.05.010>