Subtitles and transcriptions

Subtitles and transcriptions are available for selected materials for purpose of helping users understand the contents of the educational sessions.

Uncertain words have been indicated with ?? before and after the part. Parts that could not be understood at all have been indicated as [Audio Not Clear].

Every effort has been made to faithfully reproduce the audio of the sessions as recorded. However, no responsibility is accepted for mistakes or omissions. ESO does not endorse any opinions expressed in the presentations.

Prostate cancer digital evaluation

Dr Tolkach: So, dear colleagues, my name is Yuri Tolkach from the Institute of Pathology at the University Hospital of Colon. And I'm happy today to speak about digital evaluation of prostate cancer slides in digital pathology. So, of course, as you probably now know from previous presentations, digital pathology is kind of ongoing transformation in our departments. So, we've worked now without microscope. Also, in Cologne, we started several years ago to proceed some cases digitally, so you scan the slides and then you can analyze them on the monitor. But of course, this opens the possibility for different types of image analysis. And actually, what I'm going to talk about, prostate cancer, digital evaluation is all about image analysis. So, you have kind of algorithms that gonna do some small jobs for you to proceed with digital diagnosis. And maybe, I'll start with a typical prostate biopsy case. You know, we receive many cases with which are prostate biopsies and also larger resections. Prostatectomy specimens, and all of them require a lot of attention from pathologists. For example, in our department, the biopsy cases can be really large. So, some parts of them are from, for example, MRI fusion biopsies, or targeted biopsies. And there will be also some kind of systematically taken cores. And at the end, we have up to 25-cores, plus different levels of single cores. But at the end, this process is very iterative, very cyclical. So, you have to screen a single core to say if there is a tumor or not and to provide a Gleason score if there is a tumor and to measure the Length. And of course, the automatization will be very helpful for pathologist. As one case can take up to 30-40 minutes and even one hour. And all of the larger resections, prostatectomy specimens, which are processed in pathology departments and completely embedded are very large. So, it can also take a lot of time to proceed one case. For example, at the right side you can see on the part of typical case, there is another half of this case with many slides and you see my pen marks where you have to see if there is a tumor in single slide, to measure it, to reconstruct it at the end of the case and provide basic metrics and Gleason grading. So, it can be a very complicated work, kind of complicated reconstructions from single slides. And of course, kind of automatization of this work is something which very important for digital pathology. And for example, this slide, you can see kind of overview of typical, and emerging tools for prostate cancer detection from different commercial companies. And as you can see, the principle is very same. So, this kind of tool will detect prostate cancer, or just kind of tell you where is the area with highest probability of being prostate cancer. And will, provide some basic metrics like tumor square and to also Gleason's grading. So, you can also screen, for example, these homepages of different companies to see also the results of validation studies provided. I will go a little bit more into details of another tool in this list, which we have developed and which we are clinically validating. So, actually, this story started with scientific project, which we published in the last year and we trained or kind of created a classification model, which uses neuronal network and aspect bone, and analyze digital slides of prostate, detects tumor and makes Gleason grading. We have validated it using two external independent cohorts of patients with large number of patients. And it showed a really high accuracy metrics for tumor detection and very impressive results regarding Gleason grading. But you know, when you do a

scientific project, it's you mostly don't see how it would look like for pathologists because it's just kind of Python code. It's not really clinical grade instrument, which you can use in daily routine. So, of course, you have to think how will the pathologist will interact with this instrument, how it's going to enter into typical routine of genital-urinary pathologists and there, there was a progress in the last year because we started a corporation with one company, Indica labs, from United States, and the engineers of Indica labs have implemented these algorithms as real product or real tool, which can be used by pathologists in the daily routine. So, we extended annotations, we reviewed our algorithm once more time. And basically, this tool is about detection prostate cancer. So, it detects different classes, for example, invasive carcinoma. Also, intraductal carcinoma, some other classes and it performs Gleason grading. But the actual system is a little bit more complex because you have firstly to detect tissue. So, you have a kind of tissue detection model, the tissue should be analyzed and there is also one other step, quality control step, which is extremely important. And there is one publication which we recently published about importance of quality control because every artifact or just heterogeneity of staining and cutting can produce misclassifications, so we have to control for quality of your slide before you go to the next step of cancer detection and further to proceed to Gleason grading. So that's an image from the system, how the tumor detection, or basically firstly tissue detection quality control can look like. So, you can see here a green line. This is the detected tissue and the slide, and also red, red and blue regions, which stay for different quality. For example, in this slide, you can see there is a focus deterioration, which is something you will always see in pathology slides, and you always have to control for such regions. Because for example, no one will be able, also, human pathologists and also AI tool will be able to provide reliable predictions if there is a tumor or not in this region. So, this focus in consistencies can be really very focal. So, for example, this slide demonstrates that there were focused problems only in smaller regions, but these regions include epithelial cells. So, basically such focus problems can lead to misclassification of these epithelia. They are completely benign, but they can be classified as tumor and you have to be very rigorous with quality control of slides. For example, this slide demonstrates also detection of scratches. And so, quality control is extremely important. There is a video on these slides and you can basically see how the tool works. So, actually, this is a kind of digital pathology system, where you can just analyze your biopsy course, just have a look at the tissue in all possible details. And of course, there is some outputs of the model or the tool detecting tumor. So, you can just highlight the areas where a tumor with high probability was detected. Of course, the last word is from you as pathologists, but at least you have the regions and you can receive also metrics, kind of square of the tumor and always can agree or disagree with a tool regarding the classification of this region. For this also another region is with highly differentiated prostate cancer here in this core. But anyway, it's really helpful to have such kinds of maps before you start analyzing the case. So, I would like to present the results of the validation study. So, to date, we have included six different centers into validation study and also 13 pathologists as grading experts. But for now, I'd like to show the results for two large retrospective cores stemming from one center in Germany, in one center in Austria. And the overall number of biopsy cohorts analyzed from these two centers was more than 2,500. These were consecutive full routine cases. We did not do any pre-selection and just took the consecutive cases from these departments. And as you can see, one cohort was also scanned with two different scanners, Leica scanner. And at the next slide, you can see actually how heterogeneous our digital pathology material can be. For example, this image is from our center in Cologne. And this one is from Wiener Neustadt in Austria, and you can see that cutting quality on also, staining quality can be really very different among centers. And it's a large force of heterogeneity, which should be addressed during analysis. Also, when you scan the same slide with two different scanners, you see that there could be dramatical differences in color quality or intercepted color quality. And there is also another source of heterogeneity of these slides. So, in this particular part of our validation study, we decided to not to do any stain normalization or any domain adaptation methods, just to have a native output, to have basic idea how our algorithm works. Also, in case of this prominent heterogeneity in the slides. And the next slide shows that we have some important metrics for our AI tool on the per-call basis or analysis of single course. And you can see here, a GT stays for ground truth. So, for analysis results from human pathologists sometimes with immunohistochemistry and

you see also here the analysis provided by AI tool. And one important result is actually that you can reach really high negative predictive values for an AI tool. So, it's of course extremely important when such a tool says there is no tumor in a core, then it should be really no tumor over here. So, it's very important. But otherwise, you can see, for example, sensitivity and specificity is a little bit fluctuating among cohorts. And also, the same cores can cause different metrics a little bit different. But as I mentioned above, we did not do any stain normalization and actually it's a case for addressing this heterogeneity to improve these parameters. Other important metrics is of course analysis time. So, I think no one will would like to wait one or two hours till the tool is ready with analysis of slides. it's really important how quick that this tool works. And for example, for our two cohorts, we have a kind of one minute per slide analysis time or two minute per slide analysis time, course from WNS cohort are little bit particular with regard to diameter, of course, but if you have several GPU instances or several kinds of graphic cards analysis, the analysis is done in 5 to 10 minutes. So, it's actually a very competitive time to start with for a tool and tool will not lead to any kind of lag in diagnostic process. Actually, another important metric we received is a kind of cores where the tumor was not detected by pathologist. So, if the tumor was missed by pathologists, and it's really important metric because having a tool which works in the background or in the foreground before you analyze the cases is an important instrument for quality control of our work. And for detecting some mostly very small and inconspicuous tumor regions, which were not detected by us as pathologists. And of course, it's very important. So, for example, this slide, or this video demonstrates a tumor focus, which was not detected by a pathologist, but was detected by AI tool. So, you can see a really very small inconspicuous glands, surrounded by inflammation infiltrate. And of course, if you analyze this core on the 5 X or even 10 X magnification, you can just miss this very inconspicuous region. So, also when you have a kind of AI tool, we always think about accuracy and from, presentation of other companies, and also from validation study, you always receive kind of accuracy percent. For example, kind of accuracy, 98%. But you maybe want... it would be 100% for such a tool, but I think it's for now not possible and no single tool among those I showed in the beginning provides a 100% accuracy. So, there will be always some kind of false misclassifications. But when we analyze this misclassification, we saw that actually that these false positive alerts of false positive regions are very important for us as pathologists. So, for example, this is one example of a small gland. It's not a tumor, kind of atrophy, partial atrophy, but also, was also highlighted by the AI tool as suspicious or high-probability of being a tumor. Also, some other kind of atrophy and also kind of mimickers of tumor were highlighted by the AI tool. And I think it's actually, when you get only, only real tumor highlighted and don't get such kinds of false positive alerts from such mimicker regions, you don't have a confidence during interaction with such tools. So, I think that all pathologists who work with this tool today are really feeling themselves more confident when they get also the regions highlighted, which they, or which we are considering suspicious and not only clear tumor regions. So, false positive can be really important for getting more sense of confidence during the work. I would like to show several slides with regard to Gleason grading, now we have a large team of 13 experts working with our tool, but these results are for 5 boards certified pathologists. And we compare them to AI tool. And these cases, these 176 scores came from 57 cases from the first cohort showed earlier from WNS cohort. And these are just consecutive cases with different Gleason grades, in this single course. So, here you can see the results of agreement analysis for single pathologists and for AI tool. And when you compare Average Kappas to become metric for agreement between graders, you can see that actually there is no statistically significant difference between agreement of AI tool and average agreement to all other graders between AI tool and single pathologist. And actually, when you see a comparison or Kappas between single pathologists, you can sometimes see some outliers with higher Kappa levels. And when we analyzed further these outliers, it's mostly from pathologies, which used to work together a lot or have common school of grading. So, but actually when you compare the pathologist from different departments, also, a very experienced pathologists, you can see that actually these Kappa levels or agreement levels are actually pretty same for AI tool as for single pathologists. And also, we performed kind of sub-analysis with consensus grading. So, you can see here the number of cases among this course where three or five pathologists agree about Gleason grading. We have three, four or five pathologists agree about

Gleason grading and they are all pathologists provided the same Gleason grading. And you can see how progressively agreement levels of AI tool are increasing in cases where majority of pathologists provide the same grading and actually Kappa considers also some kind of agreement by chance and then actual agreement. So really, the same Gleason grading is even more, for these cases where consensus grading is present. I think the most important message we got from the study we published in the last year is actually that agreement is of course important, but we all know a very subjective nature of Gleason grading for very subjective nature of estimating a tumor architecture, but actually, Gleason grading is a prognostic indicator. And when you see it on the Kaplan-Meier curves, we see a biochemical recurrence as endpoint from our publication. You can see that actually their statistical significance is very similar for human graders and for AI tool. And maybe, if you even see how these curves for different grade groups are located, maybe the certification of AI tool is even better at stratification of these Gleason grade groups. So, I think we have to more investigate this prognostic impact of Gleason grading provided by AI to not only concentrate on agreement. Thank you very much for your attention. And I'm very happy to work with this international group of experts on this study and I hope to provide updated results on validations to our study this year. I am happy to hear your questions.