The impact of AI-derived software in a high-throughput breast screening center

The number of papers published on the use of Artificial Intelligence (AI) in radiology is growing exponentially, with the field of breast imaging being highly active in the implementation of AI approaches, which are no longer restricted to exploratory R&D and evaluation studies but are now being increasingly implemented in clinical routine.
The Danube City Diagnostic Center in Vienna, Austria is particularly active in breast imaging and is an example of a clinic which is keen to benefit from the latest technology when relevant clinically. They have recently implemented the Transpara AI software package from the Dutch company Screenpoint Medical and are using the software in day-to-day routine breast screening.
To find out more, we had a wide-ranging conversation with Dr. Friedrich Vorbeck, radiologist and vice-director of the clinic

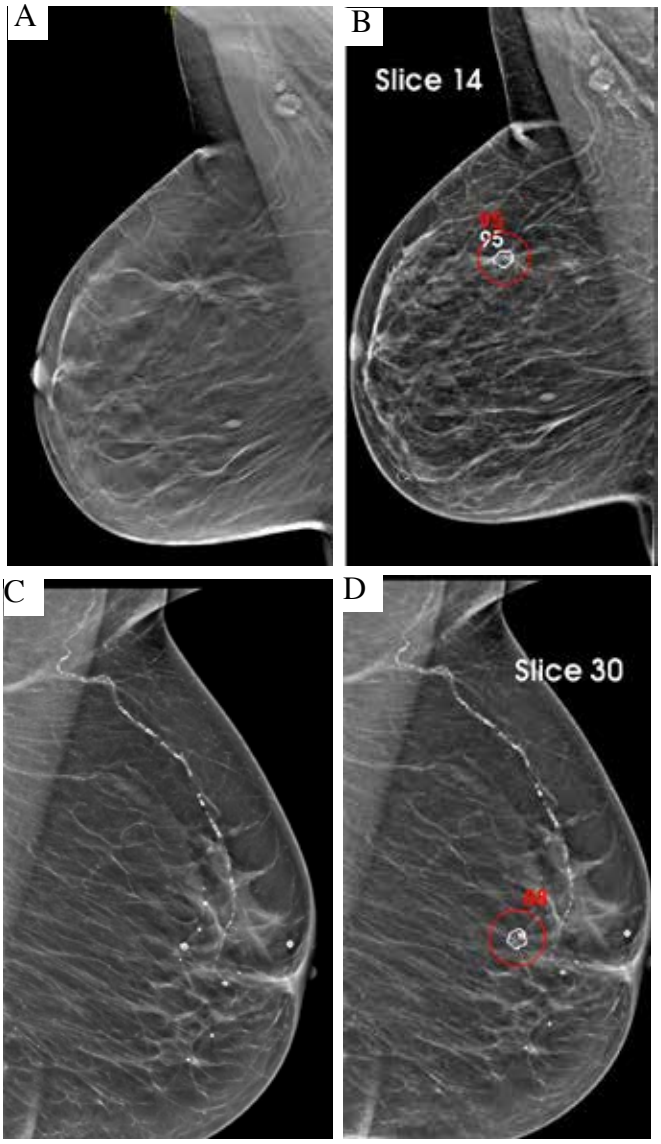
Dr. Friedrich Vorbeck is radiologist and vice-director of the Danube City Diagnostic Center in Vienna, Austria

email:
f.vorbeck@radiologie-wien.at

Q To start with, please describe your practice at Danube Imaging Center and in particular your breast screening activities

Yes, of course but before we get into the details, let me first pick up on the very valid comment in your introduction about AI software no longer being the sole preserve of academic/research centers. So let me start by putting into some sort of context the role that our center,and other private practices plays in breast screening. It is true that for example when you attend the mammography sessions in congresses such as ECR or in specialized breast imaging congresses such as EUSOBI, etc., the majority of presentations are given by specialized speakers from academic/research centers. Although such talks are always very interesting in terms for example of the statistical analysis of multi-centric and multi-national trials or the preliminary evaluation of possible new imaging modalities, the hard reality is that right now the vast majority of imaging in breast screening is carried out by private practices like ours or in primary care centers. Any difficult or special cases can then be referred out of primary care to specialized hospitals but certainly here in Austria, more than 80% of breast screening examination are carried out in centers such as ours. The result is that hospital physicians, despite their undoubted other qualities, are in fact not the best equipped to comment about routine screening mammography.
This is not just a theoretical observation, but in fact has practical consequences when, for example we want to talk about the real impact of AI-developed algorithms in breast screening and what precisely it can bring to breast radiologists. To give an example, if I refer a difficult patient to a specialized center, there the physician will use his specific expertise and experience to determine for example the precise BI-RADS category of the patient including most likely taking a biopsy which will ultimately determine the therapeutic approach to be followed. The specialized physician will probably — and understandably — say that AI does not actually contribute much in such a particular case and may even conclude that in fact there is no real need for AI.
Of course I am talking here of the situation with which I am most familiar, namely Austria, where, as I said, the vast majority of screening cases occur outside of tertiary care hospitals. I know that this organizational structure also occurs in other European countries, such as Germany. However in some countries such as The Netherlands, screening mammography also occurs in tertiary care centers.

Anyway, to come back to our center in particular, each year we see in total about 7000 women per year for breast imaging. The majority of these examinations take place as part of screening programs but there are approximately 1000 – 1500 patients that we see who present for diagnostic imaging or are being examined for symptomatic reasons.
For these 7000 women, we routinely



Images A & C above are from original breast tomosynthesis examinations; images B & D show the output of the analysis by the Transpara software, which highlights any lesion detected. The software also displays a lesion score on a scale of 1 - 100 indicating the confidence of the finding. The slice of the tomosynthesis examination in which the lesion was detected is also shown.
Biopsy subsequently confirmed the above lesions as being carcinoma.
The Danube City Diagnostic Center in Vienna, Austria, use the absence of such software - flagged lesions to indicate to the woman being screened that there is a very low probability of her having a problem and that there is no immediate reason for further investigation. In any case, her images will always be subsequently read by two radiologists, without the radiologists needing to carry out a full face-to-face examination, thus saving a considerable amount of radiologists' time.

acquire two views for each breast, so we ended up with a total of 28000 breast views per annum. Breast cancer is detected in about 70 women so our cancer detection rate (CDR) is approximately 1%. The cancer detection rate is of course an important parameter and has been used to compare the performance of screening centers, but this has to be done with caution. The CDR can be influenced by many factors, such as the characteristics of the women being examined, e.g. their age profile, or the category of the density of their breast (which varies with age). In Austria we have geographical regions where the age profile of the women being screened can be much younger than in other areas, so it is to be expected that the CDR can vary significantly from area to area. To get back to the CDR in our center, as I said we detect about 70 cancers per year, and since — as is normal — we take two views of each breast, that means that we have 140 views containing a lesion. To put that into context, we have to look at a total of 28000 images to detect 140 with cancer, which is a good example of the "needle in a haystack" analogy often used to illustrate screening statistics.

The search for the needle in the haystack gets even worse when we bring digital breast tomosynthesis (DBT) into the picture. There are of course many compelling reasons for the use of breast tomosynthesis, not the least of which is that DBT significantly increases the sensitivity of mammography, particularly in women with dense breasts. As you know the principle of DBT is to acquire several images at different angles around each breast. In our system (The Mammomat Revelation from Siemens Healthineers) we take 50 - 100 images per breast To simplify the calculation, let's assume we "only" acquire 50 tomosynthesis images per breast. We also acquire a synthetic 2D image prepared from the tomosynthesis slices to mimic a classical 2D mammography image. But that doesn't absolve us of the medical and legal need to read every slice. This means that we no longer end up reading 28000 images per year if we use mammography alone, but close to 700 000 per year with tomo. All this to detect 140 images with cancerous lesions (actually slightly more carcinomas than the average of 140, since as I mentioned above tomosynthesis is more sensitive than mammography).

In the context of the subject of our conversation, namely the overall usefulness of AI in breast imaging, you can see where I'm going with these figures. It is the overwhelming screening statistics of the ratio of images read to cancers detected that highlight the need for AI- based algorithms to help the breast screening radiologist. The need for AI support in the relatively few symptomatic cases or in the establishment of BI-RADS category is nowhere near as pressing.

Q *So how exactly do you use the Transpara software in your center ?*

The basic rationale for the use of the Transpara AI software system from Screenpoint in our center is to alleviate the workload involved in the reading of normal images, so that the radiologist can focus more on the relatively few cases which are difficult or where carcinomas are suspected.
In Austria the breast screening system has traditionally been quite intense, with a preselection of patients, and a large proportion of the women also undergoing an ultrasound examination. The reason for the supplemental ultrasound is simply to increase the chance of detecting suspicious lesions that may have been

missed in mammography, but of course the disadvantage is that it is enormously time-consuming, user-dependent, not to mention the fact that it has been shown that ultrasound gives rise to a large number of BI-RADS 3 cases, which is not desirable in a screening context. With any increase in sensitivity compensated by a decrease in specificity, right now the whole question of the true value of supplemental breast ultrasound is still a matter of debate.

In our clinic, there isn't a question of whether a woman should receive mammography or tomosynthesis — all the women being screened get a tomosynthesis examination. As I said, the tomosynthesis system we use is the Siemens Healthineers Mammomat. I know that there have been some debates on the significance of the slight technical differences, for example in the different angle of sweep between tomosynthesis units from various manufactures, e.g. Hologic, GE Healthcare, etc. I think this argument is another example where the pros and cons of a subject are investigated in detail in academic reference hospitals but which in my opinion are not all that important in the context of a screening environment. We are very satisfied with the Siemens Healthineers Mammomat system, but I suppose Hologic or GE systems may also be suitable.

But let's get back to the workflow.

The tomosynthesis examination is carried out by the technician who then initiates the image analysis by the Transpara Software. This only takes three – four minutes during which time the woman gets dressed and waits outside in the waiting room.

The technician then receives the results of the Transpara AI analysis which is presented in the form of a Transpara score ranging from a value of 0 – 10, with 10 indicating that there is a high likelihood of a tumor being present. This is a non-linear scoring system so a score of 5 doesn't mean that there is half the chance of a tumor as a score of 10. In fact all scores under 8 have a very low (and exponentially decreasing) risk.

In addition the technician will visually assess the breast density according to the RCR system 1,2,3 or 4, with 1 being low density and 4 very high density.

If the Transpara score is less than 8 and the breast density is 1 or 2, the technician can directly inform the woman that the probability of having a cancer is very low and that she can return home, with the results to follow by email. We are very happy with the performance of the Transpara system in this role.

Q *And how does the woman react to the fact that it was basically an AI system that gave her the all-clear?*

Firstly the women are happy that they can go home. They are not concerned that the decision was made on the basis of an AI system, since they know that anyway the images will later be read by human radiologists (actually two human radiologists since we use a double reader system).

Q *So apart from giving the woman an early indication of the likely results of the screening how does the AI system actually help you, given that two radiologists will anyway read the examination?*

"...there is also the added benefit of AI in that it can significantly increase the confidence of the radiologist by identifying abnormalities that are otherwise very easy to miss..."

You are right — the final diagnosis/report is always made by double reading radiologists but for us in screening mode the advantage of AI is enormous simply because of the amount of time it saves us. We don't use AI as a final clinical decision-maker but as a prioritization tool, basically by flagging up the risk of the presence of lesions. In practice, this means that in all low risk cases, the radiologist looks only at the images and does not get involved in other tasks such as having a discussion with the woman, visually examining the breast, eventually carrying out an ultrasound, which together is hugely time-consuming for the radiologist. I estimate that I would spend a couple of minutes only reading the images as compared to much more than ten minutes for a full examination with the woman. This advantage is very real, although the general requirements / regulations governing screening may be unique to Austria.

I should emphasise that all this only applies to screening cases. In contrast, in all symptomatic or diagnostic cases, the radiologist is fully involved in the examination, taking whatever additional time that is needed.

However in addition to the primary advantage of saving time in screening, there is also the added benefit of AI in that it can significantly increase the confidence of the radiologist by identifying abnormalities that are otherwise very easy to miss. We have already had several examples of this, e.g. in picking up carcinoma lesions <4mm (subsequently verified by biopsy) which were located on only three or four tomo slices, and were practically invisible on synthetic 2D, even when on retrospective viewing we knew where to look.

The Transpara AI software also gives a confidence rating for the lesions it has identified, typically by circling in red the suspected lesion. Psychology studies have shown that there is a danger that such marking may have the unintended consequence of the radiologist not spending time on other non-marked areas. This would be a real problem if the AI software couldn't be trusted to also identify other suspect lesions. In our experience with Transpara this is not the case. On the contrary, the issue could be that the software is *too* sensitive, and flags up areas which in fact are not breast cancer. An example of this is arterial calcification in arteriosclerosis where the software often identifies this as a malignant calcification. For me this is not a problem, since a quick look enables me as radiologist to identify the arterial calcification. I'd much rather have this than the AI not picking up the calcification at all.

So to summarize, our appreciation of the Transpara software is solely based on its usefulness in screening, rather than in diagnostic imaging. For example if a woman presents with a palpable mass of say 3 or 4 cm diameter, there is no point in applying AI to the images of such an easily diagnosable case. In fact the AI would probably not even detect such a lesion, simply because it hasn't been trained for that. This can be the source of unwarranted dismissal of AI technology,

for example in academic hospitals where the hasty conclusion sometimes is that if software can't even detect a 4 cm lesion then there is no point in evaluating it further.

Q *so what is the reaction of your radiologists to all this? Is AI perceived as a threat or as a much-needed help ?*

The answer to that is simple. In total in our center we have four radiologists and if I were to ask them whether we should drop the use of AI, the answer is a resounding "No". This reaction is totally based on the implications regarding workload reduction. I have heard that in some academic centers, particularly in the United States, the legal aspects of clinical responsibility has put off some radiologists from using AI. As you saw from our workflow, this is not an issue with us.

The hard reality in the field of breast screening in Austria at least is that we have annually more and more images to read and the reimbursement/financing doesn't follow this growth curve. So we have really no alternative but to increase our efficiency, and our experience so far with Transpara is such that AI is becoming indispensable in this respect.

Q *Clearly the workflow you established in your center whereby AI plays a key role in deciding the initial prioritization of cases has produced efficiency gains, but the final clinical decision still remains firmly the responsibility of the radiologist. Is a logical future next step, at least for screening patients, that the software alone determines the clinical decision, without the involvement of radiologists ?*

I don't think that this will happen. The legal responsibility will always remain with the clinician. It's like blood analysis where the hematologist is responsible for interpreting the results — it's never the hematology measurement system, however reliable it may be, that determines the diagnosis. In radiology the legal responsibility is ultimately with the radiologist — but as part of that responsibility the radiologist of course must in turn be aware of, and able to evaluate, the performance characteristics of any software that may contribute to his/her clinical decision.

Q *Going back to the development of AI algorithms, generally the more images the algorithm is trained on, the more reliable it is. Do you get regular updates of the Transpara software as the software is trained on more images?*

The question of updates is important. Of course a better-trained algorithm is always welcome, but on the other hand there are always issues of regulatory approval, so understandably the AI producing companies prefer not to issue many, relatively minor updates but restrict themselves to a few major updates.

Q *It sounds as though you are completely satisfied with Transpara. Are there no weaknesses at all?*

Yes there are a few. For example one inconvenience we've come up against is that comparing prior examinations can be a bit cumbersome and tricky — although this may in fact more be an issue of PACS integration. To their credit I have to say that Screenpoint are wide open and happy to receive feed-back from centers like ours and consider the points in their future development plans.

"...it is all the more important to ease the current work-load on radiologists, so I am firmly in favour of using all suitable technology, such as AI, to help us..."

Q *How do you see the future development of breast screening? There has been some debate recently about whether breast screening should be population-based or risk-based.*

That's almost an ideological question where cost-efficiency is key and where questions such as the whether the healthcare budget is better spent on one form of screening compared to another form. But to put that into perspective from the budget point of view, the reality is that the medical/hospital/pharmaceutical costs of the last few weeks of life a woman dying from breast cancer is huge compared. My personal opinion is that we should screen as many women as possible for the simple reason that it saves lives and also reduces the huge end-of-life costs.

Q *You said earlier that the radiologists in your center could not envisage dropping AI software. But what about radiologists without practical experience of AI? For example could a perceived threat of AI to the role of radiology be the reason for a shortage of radiologists, or shortage of students choosing radiology as a speciality?*

No I don't think so. I think that the shortage of radiologists is just one part of a more general problem which is that — in Austria at least — there are too few medical students in general. This is not because there are too few applicants but rather because of purely political decisions to limit the budget and so the number of students. I don't have the exact figures for the number of newly qualified medical graduates opting for radiology in Austria, but it is certainly not only radiology where there are shortages. I'm a member of the Vienna Council of Physicians and I see for example that general practice is chronically short of new entrants into the profession. When I studied medicine in the 1980s, general practice had a good reputation, with GPs being highly respected in the community and also being financially well paid for their work. Unfortunately that has now changed completely, so GPs are less respected, less well paid and on top of that have an ever increasing work load. No wonder that there are fewer entrants into general practice. As I said I don't have the exact figures for today but radiology in Austria is much better appreciated.

That doesn't mean that everything is rosy in radiology — it's not that easy to recruit good radiologists — but it's even worse in sectors such as orthopedics or pediatrics. For me this means that it is all the more important to ease the current work-load on radiologists, so I am firmly in favour of using all suitable technology, such as AI, to help us, so long of course that it is appropriately validated.

As I hope you have gathered from our conversation, the implementation of AI in our breast screening service has already led to significant reductions in our radiology workload, with no compromise on quality.