Where radiologists see the added value of machine learning

SECTRA

Knowledge and passion

Using a prostate cancer case example, a survey of radiologists reveals more opportunities than threats

AI and machine learning are set to disrupt the practice of radiology. How and to what extent, we do not yet fully know. However, rather than stop training radiologists and start replacing them as a few suggest, we and many others would prefer to support radiologists in their daily work with the technology of machine learning, leading to increased efficiency and quality of care. But what does this combination of radiologists and machine learning look like? Who is in the driver's seat and when?

Using a case of prostate cancer as a clinical example, we asked radiologists in three of Sectra's key markets (Scandinavia, Benelux, and the US) for their perspectives on machine learning in radiology. What do radiologists think are the right tasks for machine learning applications? Where on a scale between supportive workflow-related tasks and making diagnostic decisions do radiologists see a value in machine learning applications? And can the results from machine learning algorithms be trusted?

The radiologists participating in the survey were told a story describing a future workflow scenario focused on reading multiparametric-MRI (mp-MRI) of the prostate. Machine learning is used throughout the story to support the radiologists in their work. The radiologists were presented with several statements concerning the use of machine learning, and their responses to these statements are presented in this report.

Sectra comments

Daniel Forsberg

Senior Research Scientist, Sectra

Daniel Forsberg holds a Ph.D. in medical informatics from Linköping University. He has conducted post-doctoral research both at Linköping University and, as a postdoctoral fellow, at Case Western Reserve University in Cleveland, Ohio.

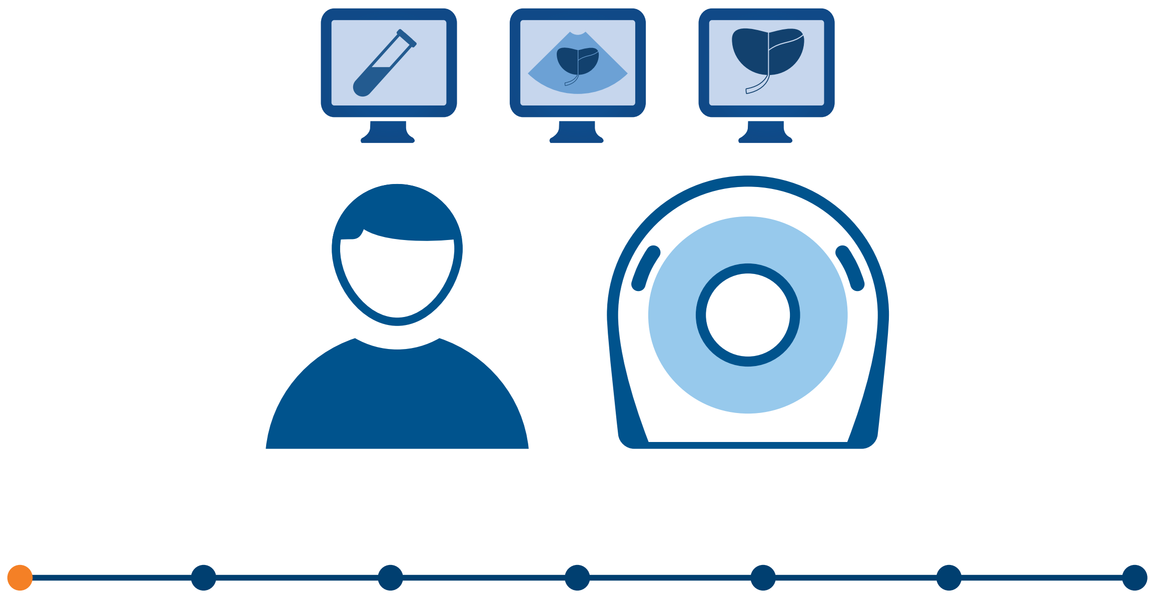
As a research scientist at Sectra, Daniel conducts imaging informatics research in close collaboration with both academic and clinical collaborators. His research currently focuses on population imaging, radiomics, oncology, machine learning and artificial intelligence.

Daniel Forsberg on machine learning

"Today there's so much buzz and hype surrounding machine learning and artificial intelligence. Though many believe that great possibilities lay ahead, few actually know how tomorrow will play out. In all of this, I do my best as a researcher in medical imaging informatics to ensure that machine learning algorithms provide tangible value today in image diagnostics while at the same time exploring the unseen possibilities of tomorrow."

For more inspiration on machine learning,Interact with Daniel on LinkedIn

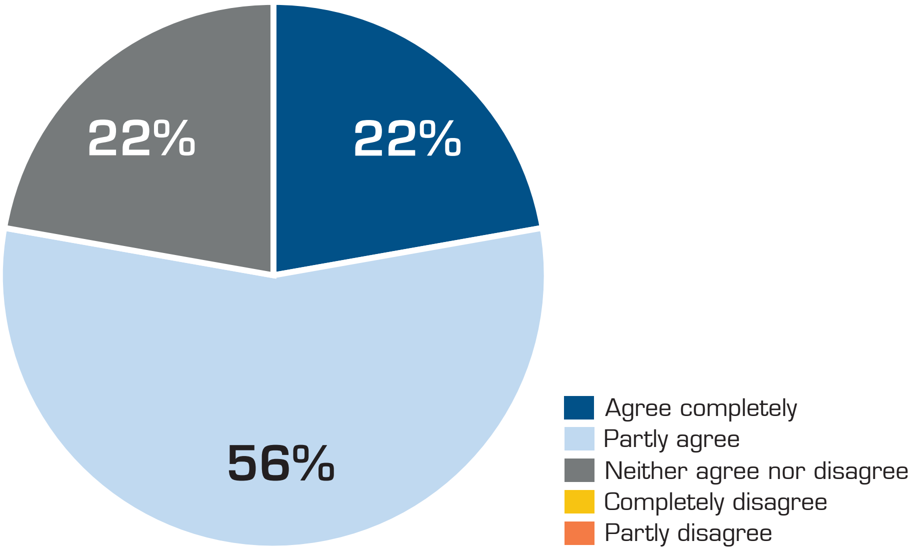
https://www.linkedin.com/in/daniel-forsberg-07189221/



Meet Steve

Steve is a 40-year-old man who for some time has had difficulties urinating. At his annual check-up, his family physician orders a PSA test, which comes back showing elevated PSA levels. Steve is therefore referred to a urologist for a TRUS and a digital examination a few weeks later. The examination shows findings of an enlarged prostate gland and an mp-MRI is scheduled.

On the day of the scheduled mp-MRI, Steve arrives at the local imaging center and performs the ordered mp-MRI. As the images are imported to the PACS, the same images are sent together with clinical information to a cloud-based advanced imaging platform for automatic analysis using various machine learning algorithms. The results from the online processing are returned to the PACS, before any radiologist has even opened the examination.



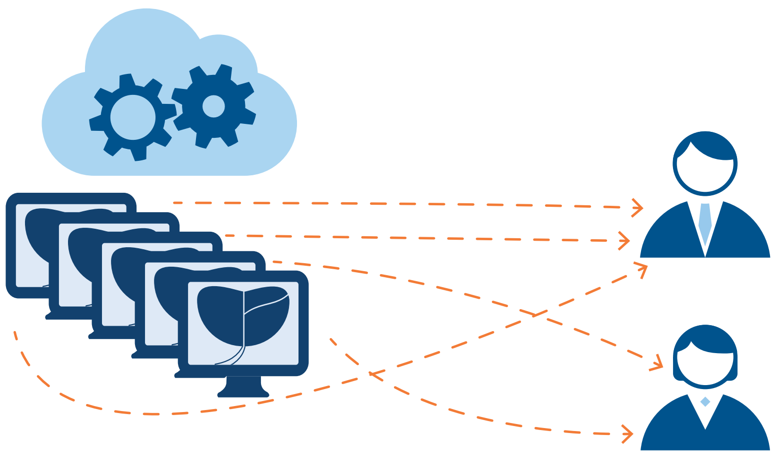
Statement 1: I would feel confident sending identifiable image data and clinical information via an encrypted connection to an online service for automatic processing.

A majority of the respondents would feel partly confident sending data to an online service for machine learning-based automated image processing. None of the respondents disagreed to any extent. Taking comments from the respondents into consideration, the perceived security of cloud-based services may be the reason for not agreeing completely with the statement. Some asked for more information on how the data was encrypted and where the cloud was located (where would images be sent to).

Sectra comments

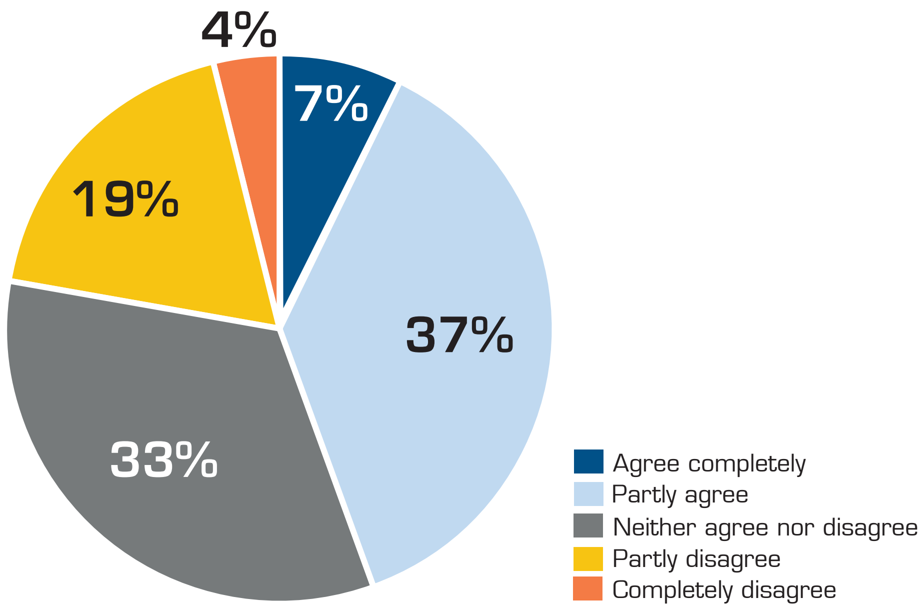
Interestingly, the responses show that there is still some hesitation towards using cloud-based solutions for patient data, even though the US, Benelux and the Scandinavian markets allow this from a legal perspective. As such, it is not surprising that one of the first companies to deliver an FDA-approved deep learning product in radiology, focuses a lot of their communications on how they can securely handle the protected health information as data is transferred from hospitals to their cloud-based computational platform.

It is important to note here that cloud computing will inevitably become an important component in the application and deployment of machine learning on a large scale. To understand why, we must first acknowledge that today AI is narrow, meaning that machine learning algorithms are trained to perform well on one task and one task only. Hence, radiology departments who are interested in making use of machine learning for a wide range of radiology tasks will need multiple algorithms/applications. Now, for a single large academic medical center it might be feasible to extend their datacenter to handle the computational need of modern deep learning applications (e.g. computational nodes with access to GPU hardware for improved computational performance) and to have a few vendors to install their AI applications on the premises, but for the vast majority of hospitals, cloud computing will be a necessity to gain access to the most recent deep learning applications. This will be the case not only from a perspective of scalability and access to adequate hardware, but also from a deployment and usability perspective. Radiologists and other AI end users will prefer access to a large offering of various AI applications through a single common interface instead of being forced to switch between a multitude of different desktop applications.



System-based decision on which cases to read and when

Two radiologists are logged into the PACS covering MR body imaging. Dr Goldstein is a subspecialized radiologist in body imaging with a long experience in reading mp-MRI of the prostate. Dr Williams, on the other hand, has just finished her fellowship in body imaging and joined the radiology group a month ago. The automatic processing has detected some lesions, but is uncertain regarding the degree of malignancy. Because of this and the earlier detected clinical findings, the system decides that both Dr Williams and Dr Goldstein need to review Steve's examination.



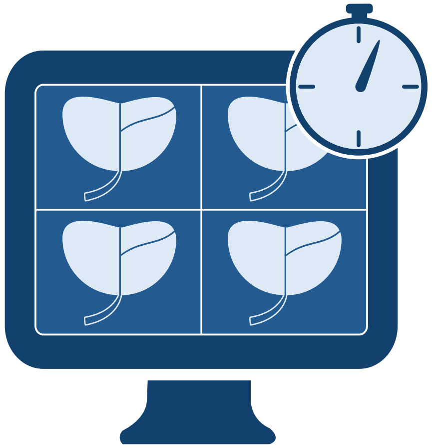
Statement 2: I would trust an autonomous system to make the decisions for me in terms of what to read and when based upon findings from the automatic processing of the image data.

44% of the respondents either agree completely or to some extent. There is, however, a large spread in the responses to this statement. 4% disagree completely, and 19% disagree to some extent. Only 7% agree completely.

Sectra comments

It can be noted that case distribution or triage are two tasks that are frequently promoted as relevant and possible applications for machine learning. For example, it is often suggested that instead of reading the weekend's production of chest CRs in chronological order, why not instead start with those that have been flagged to contain suspicious findings to ensure that patients in need of care are given the appropriate treatment as soon as possible. Considering this, it can appear counterintuitive that quite a few of the respondents hold a negative stance.

However, in this case, it is important to realize that the statement has two aspects. First, the triage aspect, i.e. letting a system perform a prioritization of available unread examinations. Second, the aspect of removing the radiologist's ability to freely choose what to read next. My interpretation of this statement is that radiologists would welcome a system that provides a recommended order of reading unread examinations based upon findings detected by the system, but that the radiologists still prefer full control in terms of which examination to read next. As such, this underlines the importance of working together with radiologists to ensure that developed and deployed machine learning applications solve a problem that makes sense to the end users and that it is integrated into their everyday workflow in a way that supports their work rather than limits it.

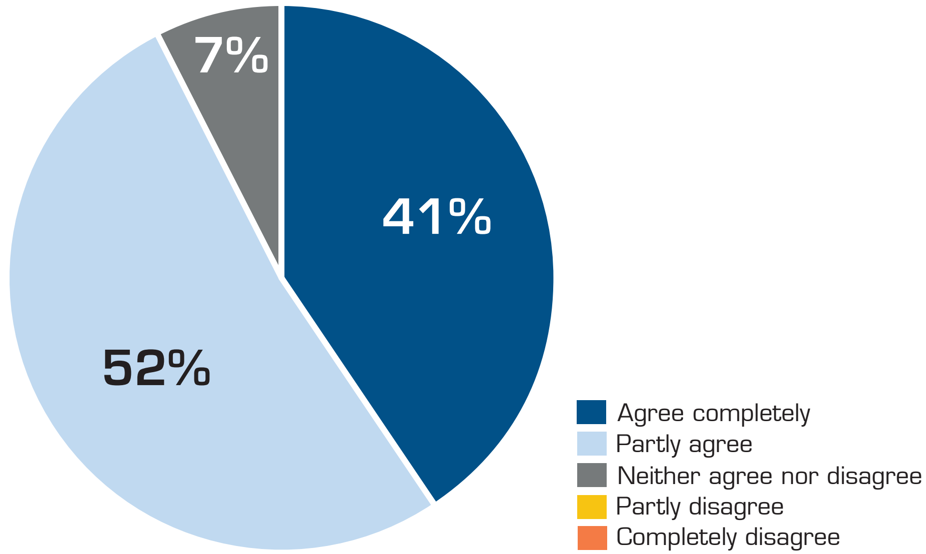


Non-diagnostic workflow support that improves over time

As Dr Williams finishes her current case and clicks on the next case, Steve's examination is automatically hung on her monitors. Dr Williams notices that the hanging almost matches her individual preferences, which differ to the standard hanging protocols. She feels that the performance of the automatic hangings is better than her first week at work and she feels confident that it will soon have the matching accuracy of the automatic hangings provided to Dr Goldstein by the system.

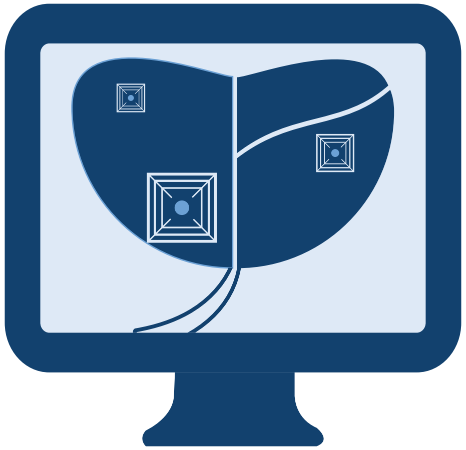
Sectra comments

That machine learning-based applications sometimes make mistakes but that they can improve over time are two aspects of machine learning that are important to convey to potential users to avoid misguided expectations. From a usability perspective and for the developer of a machine learning-based application, it becomes important to design for failure (the failure of a prediction), so that it is easy for the user to spot a mistake and to correct it.



Statement 3: I would be willing to work with a system that over time improves its ability to provide non-diagnostic workflow support, for instance learning hanging protocols.

It is clear that the respondents feel much more comfortable with non-diagnostic support from machine learning applications than with the previous statement. 93% of the respondents agree completely or to some extent. The rest of the respondents, 7%, neither agree nor disagree.

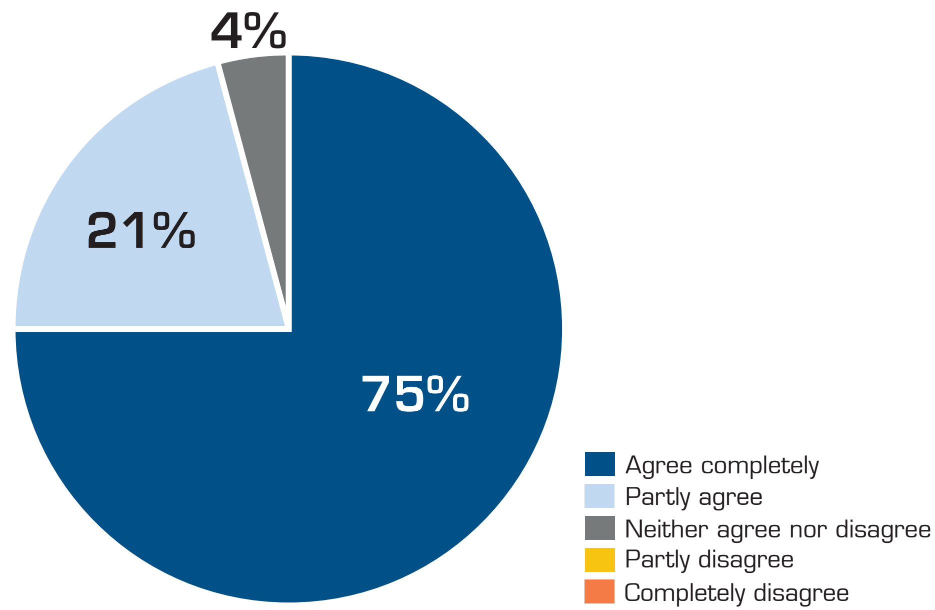


Automatic delineation

Dr Williams starts her review by quickly skimming the clinical summary provided from the EMR. Among the highlighted items, Dr Williams notices that both Steve's father and uncle had prostate cancer, something which is not mentioned by either the urologist or the family physician in their reports. After reading the clinical summary, Dr Williams starts scrolling through the images. As the prostate has been automatically segmented, the different zones of the prostate are highlighted in the various sequences as she browses the images.

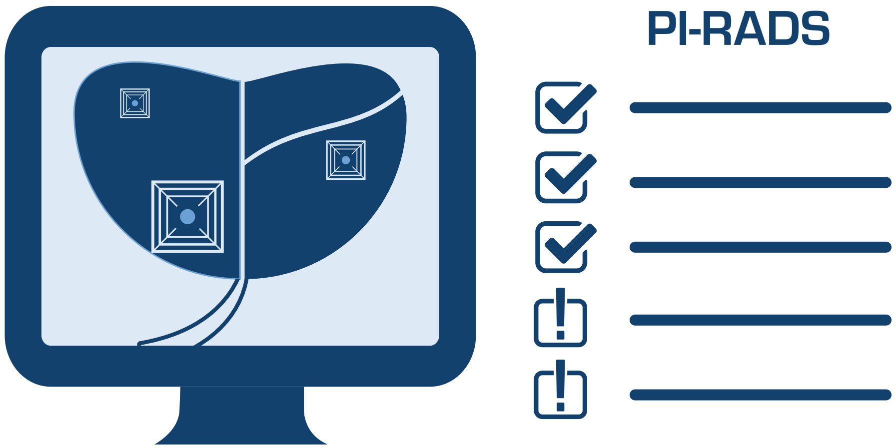
Sectra comments

Automatically annotating organs or parts of the anatomy will, of course, be helpful, not only for prostate mp-MRI, but also for a number of other anatomies and modalities. It can be assumed to be particularly helpful for young radiologists or residents or when there are structured reporting guides to follow. An already demonstrated example of this is spine labeling as showcased by Sectra at RSNA 2016. Segmentation of organs and potentially also lesions has the added benefit of an enabler of quantitative imaging. With segmented target regions, making various size measurements and extracting other quantitative metrics have become straightforward.



Statement 4: Automatic delineation of the prostate and its zones would be valuable for navigation, diagnostic and/or reporting purposes.

75% of the respondents agree completely and another 21% to some extent. This is the statement that has the largest proportion of agreement among respondents. None of the respondents disagree to any extent.



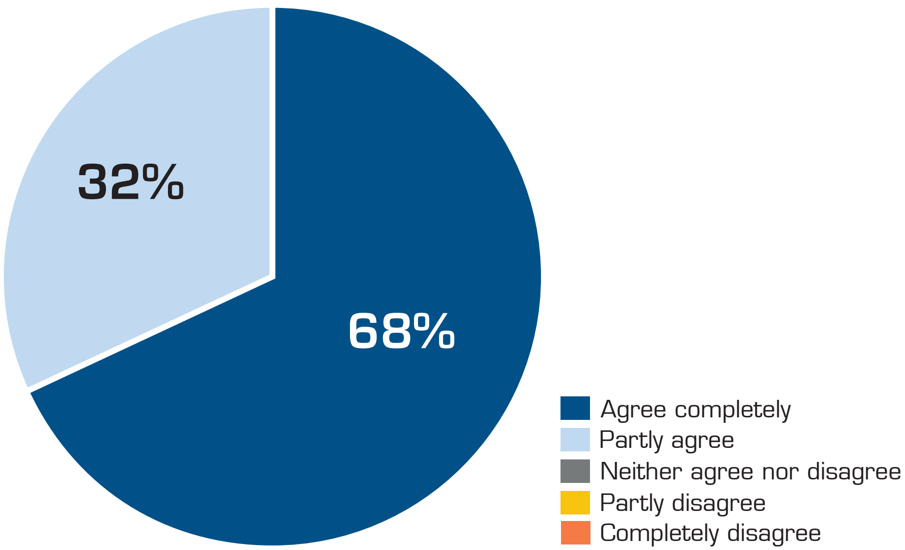
Automatic characterization and scoring of lesions

Dr Williams marks suspicious regions and verifies her findings against the automatically provided detections. Two out of three match the automatically provided detections. The automatic processing has also found a lesion in the transition zone, which she did not mark. However, she rejects the suggestions by the system and proceeds with the three lesions she has marked. Together with the system, she reviews the suggested PI-RADS scores for each lesion. She notes that her own scores differ somewhat from the system-provided scores, but realizes that the reason is mainly due to her somewhat more conservative size measurements.

The final score of the three lesions indicates a potential malignant cancer.

Statement 5: Automatic characterization and scoring of lesions according to internationally accepted criteria would be valuable to me.

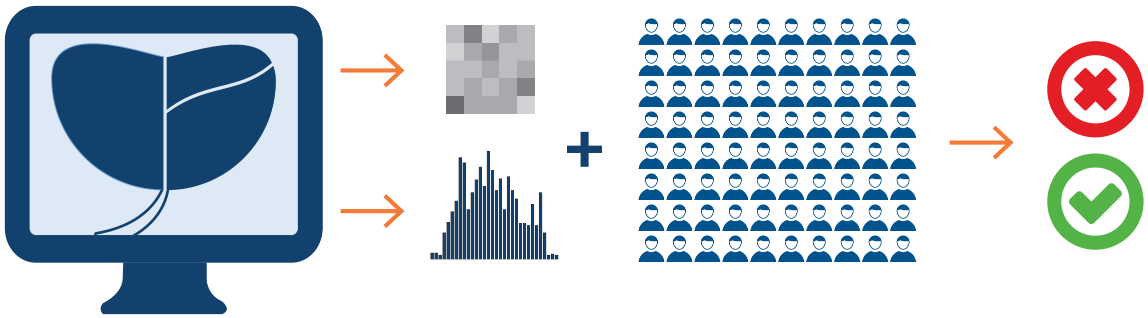
All the respondents agree with this statement. 68% of the respondents agree completely and 32% to some extent.



Sectra comments

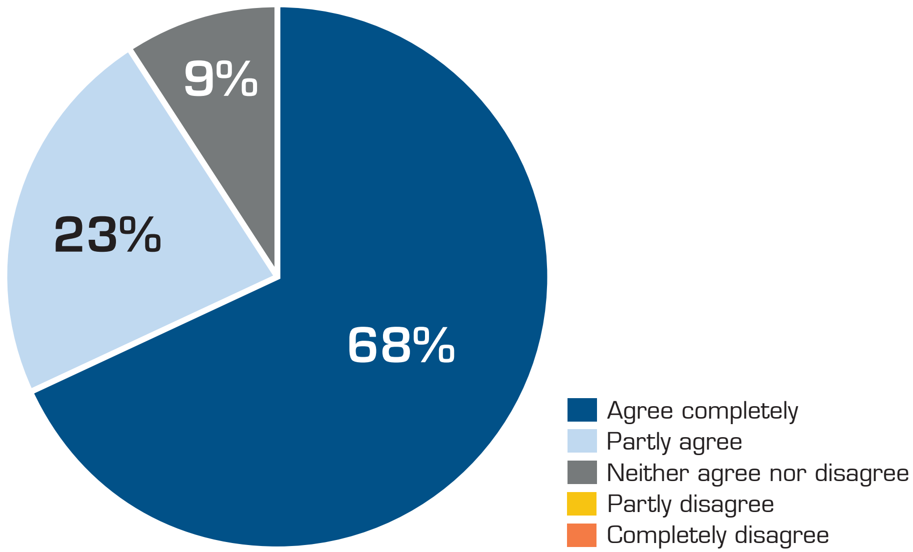
As the whole field of radiology is moving towards more and more quantitative imaging, it is relatively straightforward to envision a future where more and more procedure types will be characterized and scored according to a set of nationally or internationally accepted criteria (think BI-RADS, Lung-RADS and PI-RADS). Hence, the more support available for this, the better.

For this example, it becomes important to consider how automatically generated results are presented and utilized: for detection (that is for supportive purposes with the radiologists in the driver's seat), for diagnosis (automatic diagnostic generation) or something in-between. Most of the time, this is primarily dependent on the user interface and how the results are presented to the end user. Are the results presented in such a way that they provide support to radiologists to make their own final assessment or are they presented in such a way that they claim to be the final result, leaving little room for radiologists to interact with them. Sectra has opted for the former path for our machine learning applications. For example, in Ki67 cell counting for digital pathology, where the results from the machine learning algorithm are presented in such a way that it is easy for pathologists to adjust the results and make the final call, leaving them in full control of the end result.



Radiomics and population imaging

The PI-RADS includes some criteria related to the appearance of lesions. However, in typical images, there are so many more imaging features that can be computed (intensity, shape, margin, texture). These features can be used to create what is called a radiomics signature, which has the potential to differentiate between tumor phenotypes or provide a more fine-grained scoring of lesions. For example, the characteristics of Steve's lesions can be compared with the historical patient population and corresponding clinical outcomes. A more detailed assessment can be provided, potentially omitting the need for a subsequent biopsy to grade the lesions.

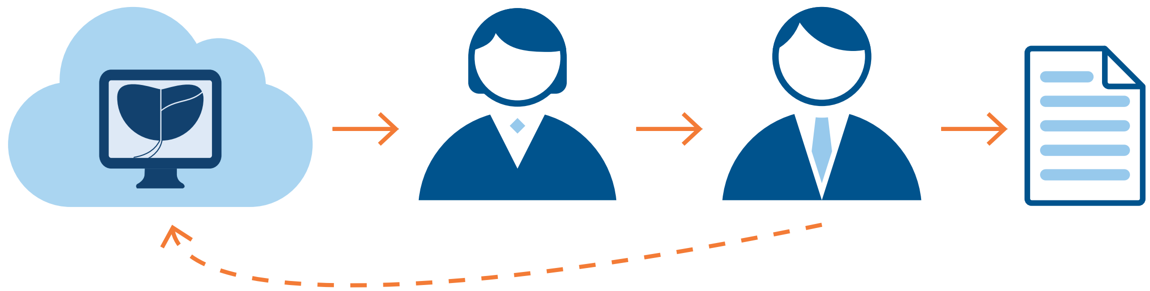


Statement 6: Decision support based upon radiomics and population imaging is the way of the future.

A large majority of the respondents agrees that radiomics and population imaging is the way of the future. 68% agree completely and 23% to some extent.

Sectra comments

Decision support is without a doubt an important task for machine learning applications, and radiomics and population-based imaging are interesting examples of where it could be used. Quantitative imaging, population imaging and consequently radiomics have over the past few years showed very promising results and it is good to see that this has not gone unnoticed by the radiologists themselves. It will be interesting to see how deep learning in the form of convolutional neural networks and "traditional" radiomics will work together or whether deep learning will become prevalent here as well.



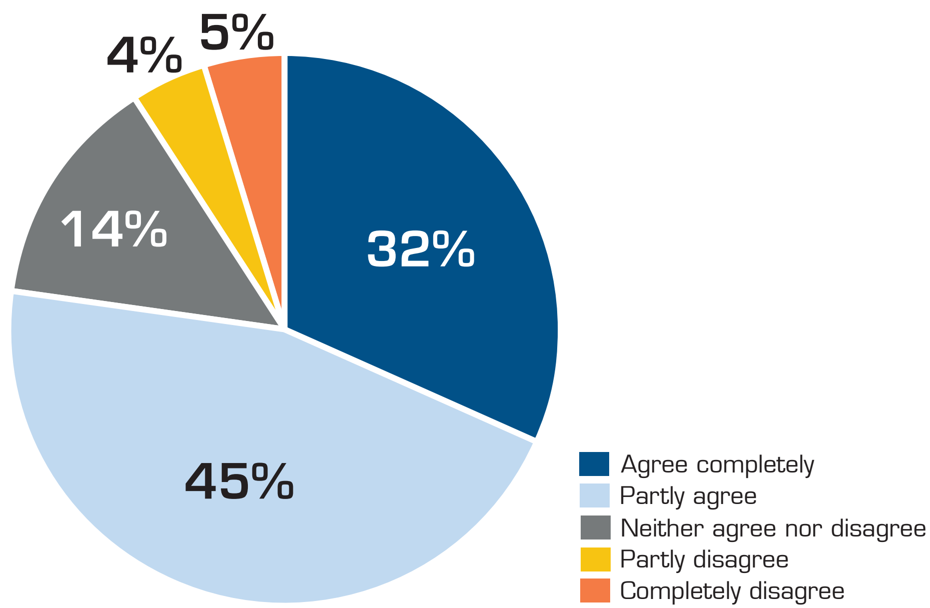
A learning, and therefore changing, system

Dr Willliams finishes her review and signs her report. A few minutes later, the examination is uploaded to Dr Goldstein's workstation and a similar review process commences. The difference is that Dr Williams' findings are also available in the system. Dr Goldstein discards one of Dr Williams lesion detections and he also discards the additional lesions suggested by the automatic processing. He notes to himself that the number of false positive detections from the system is decreasing, but that the system is still struggling somewhat with findings in the transition zone.

As Drs Goldstein and Williams have finished their work with Steve's case, all lesion detections and feedback are sent to the online processing service. Once a month the system is retrained based upon the detections from the previous month and feedback from the radiologists.

Sectra comments

It is understandable that more radiologists are positive towards changes in performance for non-diagnostic tasks than for diagnostics tasks. Again, it is a fact that machine learning algorithms can learn and improve with more data. Many users expect this, especially if the machine learning algorithm includes some features that appear to provide feedback to the application. However, it will be important how new versions of an algorithm are deployed. It is probably better to deploy new algorithm versions at the same time as the overall system is updated rather than just doing it from one day to another when nothing else changes. This would prevent radiologists developing a distrust toward the machine learning applications that randomly seem to change their performance.



Statement 7: I would feel confident to work with a system that once every month changes its performance and behavior (mostly for the better).

This time, when asked about working with a system that changes its performance with algorithms that have the potential to significantly affect the clinical decisions made by the radiologists, we still see that most respondents are positive but that some are negative. 77% either agree completely or to some degree. 5% completely disagree.

Book a meeting

Book a meeting at RSNA and discuss the topic live:

sectra.com/rsna

Join the discussion

What do you think the ideal combination of radiologists and machine learning look like? Let's continue the discussion online:

sectra.com/medical

twitter.com/SectraNews

linkedin.com/company/Sectra

Further reading

• Forsberg D, Sjöblom E, Sunshine JL. Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data. Journal of digital imaging. 2017 Jan 12:1-7.

• Völzke H, Schmidt CO, Hegenscheid K, Kühn JP, Bamberg F, Lieb W, Kroemer HK, Hosten N, Puls R. Population imaging as valuable tool for personalized medicine. Clinical Pharmacology & Therapeutics. 2012 Oct 1;92(4):422-4.

• Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications. 2014 Jun 3;5.

• Abramson RG, Burton KR, John-Paul JY, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Lenchik L, Subramaniam RM. Methods and challenges in quantitative imaging biomarker development. Academic radiology. 2015 Jan 31;22(1):25-32.

• Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2015 Nov 18;278(2):563-77.

• Beam AL, Kohane IS. Translating artificial intelligence into clinical care. JAMA. 2016 Dec 13;316(22):2368-9.

Sectra AB • info.medical@sectra.se • sectra.com/medical

SECTRA

Knowledge and passion