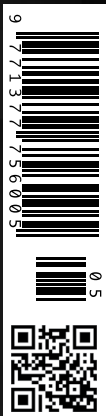




# Labs and Drugs



**440. Paul Timmers**  
Health Data Innovation Perspectives

**445. Renato Cuocolo**  
Radiomics: Recent Trends and  
Assessing Research Quality

**456. Ekaterina Kldiashvili**  
eHealth for Morphology Laboratory  
Practice

**462. Maria C. Carrillo**  
Working Towards a World Without  
Alzheimer's and All Other Dementia

**468. Samna Ghani**  
Why is the Pharmaceutical Industry Not  
Developing New Antibiotics?

**472. Henrique Martins**  
The EU Health Data Centre: A New Total  
Virtual Organisation

# Radiomics: Recent Trends and Assessing Research Quality

◆ Author: Renato Cuocolo | Department of Advanced Biomedical Sciences | University of Naples 'Federico II' | Naples, Italy

Dr Renato Cuocolo, radiologist and research fellow at the University of Naples 'Federico II', recently spoke at the 2021 European Society of Medical Imaging Informatics (EuSoMII) Annual Meeting about the challenges in assessing research quality in radiomics. Given radiomics' transformative potential for medical imaging, HealthManagement.org met with Dr Cuocolo to discuss the recent trends and challenges facing radiomics. Topics ranged from artificial intelligence (AI) integration into the radiological workflow, the appropriateness of specific machine learning algorithms, and assessing research quality.



## Key Points

- Although AI can be applied to facilitate clinical workflow, challenging, high-concept aims drive radiomics research.
- AI can excel in prioritising patients to deal with heavy clinical demand and help with image review and interpretation.
- Radiologist-AI interaction should be seamless but not be based on blind adoption. Radiologist-AI trust can be built using easily verifiable outputs in the initial implementations.
- Despite the growing focus on deep learning, any correctly-applied machine learning algorithm can work well. Simpler models should be preferred if the performance is substantially equivalent.
- If the theory behind a radiomics investigation is sound, then performance should be reproducible under a variety of conditions.
- Most commercially available AI solutions do not have peer-reviewed data backing their performance claims.

## What Needs Now Facing Radiology Can AI Address?

This is a challenging question. The potential for what we all aspire is to have radiomics and machine learning open new possibilities and give us new avenues to bring value to healthcare through radiology; to allow us to obtain information that currently is unavailable from the images, or are not easily obtainable, or require high levels of expertise.

In practice, in the short and medium-term, a feasible goal is to lean on radiomics and machine learning to help us improve the quality of life and speed up the repetitive and less interesting tasks. Consequently, radiologists can be more fully dedicated to the more challenging and interesting aspects of clinical practice.

For example, automated lesion size measurements, segmentation, with less focus on their characterisation; the last topic is

still too challenging for widespread clinical adoption of predictive modelling.

## Can AI Help Tasks That Are Inaccessible, Hard, and Tedious?

Yes. For example, there are multiple sclerosis lesion load comparisons over time or oncological patients staging or follow-up exams. These are tasks that already have some software tools available. Machine learning can certainly improve on those that are available, and this is already a reality.

In the long-term, with the development of the field, one would hope that we could use these tools to obtain additional information compared to what we currently can: for example, the genomic or phenotypical profiling of diseases, which we currently mostly cannot do. This is more interesting from a

research perspective right now because it's the furthest away from a clinical practice point of view. On the other hand, what is more interesting from a clinical practice point of view is these repetitive, boring, and time-consuming tasks that are not challenging for radiologists. Those are the ones that are less interesting from a research point-of-view, and maybe there's less incentive on publishing on those topics because they're less glamorous. One has less opportunity to have high visibility with those efforts.

### Rather than the Tedium of the Workflow, What Is Driving the Innovation?

No, it's not driving the innovation, but I think it's where radiomics can find an easier application in the short and medium-term. What is driving the research are more high concept rewards, but those are more challenging to implement. I think

but not too much. Our resources were focused. Regarding radiomics, I think they could not help speed up the reporting of these.

But machine learning in this setting could be useful in areas not tied to image analysis because machine learning also has some models and approaches to improve patient prioritisation and management of triaging and waiting lists.

Machine learning could have a role in addressing the increased demand for radiology due to COVID-19 or other future reasons where we would like to provide more exams. That space would require the digitalisation of healthcare databases providing information about the patients to correctly select which patients should have easier or earlier access to the exams.

That's a delicate and challenging topic, but it's a space where machine learning could help. It's something that's already

---

## If there's something real behind that experiment, then it should come up independently and from more groups – because there's something there that we're all looking at

---

those are where the attention is focused, but those applications are still very far in the future in a credible manner.

There is a disconnect between where the research is focused, where the funding is going, and where I think radiomics can make a short and medium-term clinical impact in the next five years or ten years. When you're modelling for genotypical aspects or similar outcomes, it's very challenging to reproduce the results across the board and have a product that is implementable everywhere in the world because the settings are incredibly different. Even when you develop a good product, maybe an institution will change their scanner three years in the future? Then you may have to start over pretty much. On the other hand, there are simpler tasks, like lesion segmentation, that are easier to verify from the radiologist's point of view because you can see and check the output in real-time. That's easier to implement, but it's less interesting. It's less glamorous from the research, academic, and funding point of view. It's more challenging to obtain an interest in research in that field. So, I think there is a disconnect between what can be done right now and what we would like radiomics and machine learning to do in the future.

### How Have Radiology Departments Handled Increased Demands Due to COVID-19?

Yes, there was a high increase in chest x-rays and chest CTs in my department, but unfortunately, there was also a decrease in many other areas. The overall amount of activity increased

done in other areas where machine learning has already been applied. They're less critical than healthcare, but there is good experience in this kind of work in other fields.

### Is One of AI's Best Applications Prioritising Patients?

Patient realisation and prioritisation of the exams will help manage the resources when the demand is higher than the resources. Machine learning can help in correctly distributing the resources to allow people who should access healthcare, radiology in this case, so that they won't be left out because there's too much demand.

### How Can AI Help Improve Image Review and Interpretation?

Yes, it can help. To help clinical practice and imaging interpretation review, AI solutions should be integrated with the current software we already use for image reporting, such as PACS systems visualisation and reporting systems. Some solutions work together with PACS vendors and provide good integration with modules within the viewing system. In some cases, they are automatically filling up some of the parts of the reporting. This could be ideal because this software usually has a very practical application like lesion detection, measurements of lung nodules, brain aneurysms, or other findings. Volumes of brain haemorrhage are easy for radiologists to double-check.

We have its output integrated within our clinical workflow, and it's easy for us to see that the algorithm in these applications is working as intended. So, we can easily trust the output of the model. Focusing on these kinds of tasks eases the introduction of these tools in practice because it's easier for radiologists to trust something they can verify immediately.

When you have outputs that refer to information that has to be obtained after surgery, down the line, or prognosis after

self-driving cars. It's something that has been promised for many years. Even now, there's lots of software, but they always require a driver who has their hands on the wheel. Even when the car is in perfect conditions on the highway with low traffic, supervision is always required. No one would ever suggest using the car without any kind of supervision. The same thing is applicable in healthcare and radiology. It's probably equally dangerous because it's always a life-and-death situation. In

## What is driving the research are more high concept rewards, but those are more challenging to implement

ten years, it's very challenging to trust the output by someone who doesn't know how the system works or hasn't worked on developing it. When it is outputted and integrated into a report, they have to sign and take responsibility for it.

### Should AI Systems Seamlessly Integrate into the Workflow and Not Be a 'Black Box'?

I think interaction should be as seamless as possible. It can be done by not using external software as much as possible and not using a dedicated workstation as much as possible.

It should not be left confined to niche areas or specific experts only. It should be made as easily accessible as possible, so the interaction requires as little action from the radiologist as possible. It should be just an overview of what the output is. And use outputs that are easily verifiable by the radiologist without complex technical knowledge. That would be useful.

### Should Radiologists See 'Behind the Curtain' and See How the Model Works?

Yeah, you have to see how the model is working. This is challenging because it's impossible in most cases, or the output needs to be easy to understand.

For example, I can see if the software detects a nodule and measures it. I can see what the measurement is and where the nodule is located. So then, it's easy for me to verify that it acted correctly, even if I don't know how it detected the nodule and how it performed the measurement. I see it's correct, and then I can trust it. That trust can build to introduce more complete tasks where we can start trusting it a little more, delegate, and step back.

This is a challenging balance because you don't want to get to the point where you let the algorithm work completely unsupervised. We want to trust, but not too much. We want someone to check what's going on. It's similar to what's happening in

both cases, you can have a car accident or misdiagnose a lesion or not see a lesion and its secret features.

### What Information Should the AI Provide? What Are Useful Features?

Suppose one wants to dig into how the software works internally. In that case, this should be made as available as possible - for example, seeing feature distribution, seeing how the model is built. If the model uses specific features, it could provide some information on how these features have been distributed within the lesion and, maybe, on the training database where it was used. It should give some insight into how it arrived at its conclusion. For deep learning, you can have activation maps to see where the model's image detection was focused. If one wants to have some information, it should be available because there can be some doubts about the output.

But the front-end for the general user should be as simple as possible, so that information can be accessible but not be mandatory to look at it. It can get too complex for the general user. To become something that we use routinely, it should not get into this level of detail for every exam. Otherwise, it becomes a hindrance instead of perfection.

The ideal implementation depends on what we're talking about. For example, for prognosis, probably just having a probability and an outcome is useful, so we know the progression of disease in five years or something like that. But it would be pretty extraneous to what we usually report right now in radiology. It would not be easy to integrate this information within what we are used to having in our final exam reports. That requires a little bit of work once these technologies are widespread.

### Which Machine Learning Algorithms Lend Themselves Well to Radiomics?

Pretty much you can use any algorithm with radiomics, even

if there is always a challenge tied to the number of patients or lesions or instances available for the training of the model. The main issue is that radiomics usually produce by definition hundreds or even thousands of features for each case. It's known that in machine learning, like in statistics, one cannot use the whole data set because the amount of noise is excessive.

So long as there is a correct pipeline before implementing the machine learning model, there's a good feature reduction. This can include good feature stability, univariate analysis, multivariate analysis, dimensionality reduction with the prin-

When comparing various models, I can say all of them can be useful. There may be cases where deep learning is indicated even if the amount of data that we usually have in radiology is not comparable to what is available for deep learning in other tasks.

Deep learning models have reached prominence in other fields where data sets consist of millions of entries, while in radiology and medicine, we have tens or hundreds of patients. When we have hundreds of patients, we are already happy because we have a rich dataset for our field. But, if you compare those numbers with what is available, for example,

---

## Patient realisation and prioritisation of the exams will help manage the resources when the demand is higher than the resources

---

cipal component analysis, or even more complex algorithms. These could be considered machine learning algorithms but unsupervised ones. Then much any kind of model can be used. From a methodological point of view, if we can obtain a similar performance with a simpler model, it would always be preferable to start out using the simplest model available: even a logistic regression or a linear regression, and then build up from there. Simpler models should always be preferred when possible because the simpler model is easier to understand and to verify that it's working correctly.

As we increase the complexity of the model with ensemble approaches, as with random forests, which are still very understandable, or support vector machines, the complexity increases to the point that deep learning becomes can go to support vector machines that can get fairly complex with learning pretty much a black box. Interpretability becomes limited. You usually can improve performance, but you pay the price in terms of interpretability. So different models should be investigated, but we should select the simplest one for the final implementation, giving the results we wish. This leads to finding the best balance between accuracy and explainability. This is a real advantage of simple models as compared to deep learning.

Today, there is a tendency to go directly to deep learning for any kind of issue. This happens not only in healthcare and radiology but in research in general. There is hype for deep learning because it's more complex and it requires higher computing. It looks more interesting. In the beginning phases of research, there is a tendency to overshoot and go directly to deep learning rather than starting with simpler models, which would probably be more correct from a methodological point of view and even from a practical implementation view.

in image-net or in other datasets, it's pretty much a drop in the ocean.

To summarise, all models can be useful if selected for the right task. One should start simple and move to complexity only, if necessary, after experimentation, and not start with deep learning because that's what the trend is right now in research.

### When Is 'Deep Learning' Appropriate to Use?

Deep learning by design uses a large number of parameters. That's already an issue when the number of data from which those parameters are derived is small. It holds the risks of bias and unreliable results. You can also use deep learning on features that have been extracted by hand or by manual analysis of the image.

The use of deep learning has to be justified from prior experience. Or, one should also use a simpler model for comparison and to prove the added value of a neural network. Even when this has been done in other fields, deep learning was not always the best solution. Random forests or even logistic regressions in many tasks and other fields are still competitive. Only when the amount of data becomes overwhelming (and this has to be demonstrated experimentally), deep learning has the upper hand unequivocally.

In radiology, we have not yet reached saturation level with simpler models, so that deep learning is required to improve what you're currently doing. I think the results that are reported right now in many cases are still obtainable with simpler methods. More understandable results are easier to present and propose to those not directly involved in the field. One can then build upon those. Once large enough data sets are available, then deep learning could probably become viable for more complex tasks that are not yet doable right now.



### Does the Algorithm Selection Depend on the Imaging Modality, the Organ Tissue, or the Disease?

Those factors can influence the selection of the model but mostly in terms of the availability of data. Because in some modalities, like X-rays, it's easier to collect very large databases, and usually, there should be less variation. For other modalities, like ultrasound, image characteristics can vary greatly even within a site. Because each operator uses different settings and this changes the way that the images are acquired. This can introduce biases that are not visible to the human eye but become relevant when analysing the images quantitatively.

on machine learning papers. In both cases, we have to limit our pulling of accuracies to AUC data because the raw data of the test stress was not available. There is a widespread issue of not presenting the entirety of the obtainable results. That's the main issue.

Usually, researchers tend to stay more general and provide the AUC as a general accuracy metric, but then they don't always test more prospectively. This applies not only to a prospective study but to even an experiment of clinical implementation with a specific cut-off and providing, for example, a specific confusion future metric with true positive, false positives response. This would be more informative. From a clinical

---

## You don't want to get to the point where you let the algorithm work completely unsupervised. We want to trust, but not too much

---

In general, I don't think there is a direct correlation between a specific image modality or organ kind of lesion and a preferred machine learning algorithm. I think the choice of the algorithm depends more on the task that we have in mind because if we are talking about lesion detection, then an algorithm that works on the images directly. This type of algorithm depends, not much on the organ or modality, but more on the aim and the kind of data set we have to work with.

### What Challenges Do You Face in Comparing the Performance of Different Algorithms?

There is no preferred metric, even if some specific metrics are more commonly used for some tasks. For example, in segmentation, the dice score is the same as the F1 score used in classification, and so on.

One of the challenges is that researchers often expect to report just the area under the receiver operating characteristic curve (AUC-ROC) or one metric used as the reference, especially those not with a medical background. Usually coming from a more technical background, they're used to tuning the machine learning pipeline to focus on a metric that becomes the reference used for tuning the model, its hyperparameters, and the whole pipeline.

This translates to a tendency to focus on a single metric and then report only that metric within their paper. In medicine, we are used to having more metrics available and even the tools to obtain additional metrics reported in the paper. This information is necessary metrics reported in the paper. Suppose one wants to obtain additional information or even allow format analysis and other types of studies that aggregate data differently; this information is necessary to perform those analyses. In my experience, we did perform two meta-analyses

point of view, specific metrics gain different values based on the problem we discuss. If it's a screening program, we could accept more false positives if it means we are not missing significant lesions. Providing only the AUC gives us no information on that side of thing, so although we may know that the accuracy is good, we don't know the practical distribution of the patients. We might prefer a lower accuracy with a better negative predictive. But I wouldn't focus on expecting a specific metric from each paper. I think it's better to ask for as much information as possible because that's the only way to go forward and have reliable results and build trusted systems. As long as we're only providing one metric, it can always give the impression of being cherry-picked and selective reporting, which only feeds the doubts that some people have towards these techniques. In my experience, we did two meta-analyses on machine learning applications. In both cases, we have to limit our assessment to AUC data because the raw data of the test stress was not available. There is a widespread issue of not presenting the entirety of the obtainable results. That's the main issue.

From a clinical point of view, specific metrics gain value based on the problem we discuss. If it's a screening program, we could accept more false positives if it means we are not missing significant lesions. Providing only the AUC gives us no information on that side of thing, so although we may know that the accuracy is good, we don't know the practical distribution of the patients. We might prefer a lower accuracy with a better negative predictive. But I wouldn't focus on expecting a specific metric from each paper.

I think it's better to ask for as much information as possible because that's the only way to go forward and have reliable results and build trusted systems. As long as we're only

providing one metric, it can always give the impression of being cherry-picked and selective reporting, which only feeds the doubts that some people have towards these techniques.

### Should the Best Metric to Use in Comparing Algorithms Depend on Its Intended Function?

Even if it is the best metric, it's always a limited amount of information. One should always ask for as much information as possible; all the possible metrics that can be reasonably obtained without going overboard.

I don't mean that everyone who presents a single metric does so malevolently. As stated previously, this is especially understandable when researchers don't have a clinical background. You usually have to select one metric during valida-

expertise from behind the research or the availability of sufficient information to assess that point from a reader point of view if the paper has already been published.

But in any case, if it becomes standard practice to expect a thorough reporting of the results in these kinds of papers, the issue will resolve naturally over time. When that information becomes available, we can perform meta-analyses as we do in other fields using classical statistics. We have come to expect this degree of information from clinical trials, not using machine learning. It's not reasonable to not apply the same standards that we have always expected from the other fields and not apply them to machine learning. It's not as if because it's machine learning, we don't have to expect the same degree of information in the end result.

## There is hype for deep learning because it's more complex and it requires higher computing. It looks more glamorous

tion that becomes the reference metric during the development process. There is a tendency for machine learning developers, engineers, and researchers to focus exclusively on that metric. But that metric alone, at the end of the process when one wants to hypothesise the clinic applicability of the result of the resulting model, does not give the full picture. Having the full confusion matrix, which is all the basic obtainable metrics, gives us a better picture and helps us understand if some problems were not obvious to the researchers. For example, because they didn't have the required clinical background or they overlooked it. It can happen.

In general, the solution is for the journals, the readers, the reviewers, to require that all the reasonably obtainable metrics are produced to allow a complete evaluation of the actual result fully.

### How Do You Evaluate Other People's Research When That Info Is Absent?

Well, if I'm a reviewer, usually, I ask for the confusion matrix as a requirement for the assessment of the paper. If I'm a reader, as I said, we did perform two meta-analyses. And in those cases, we had no other choice but to focus exclusively on the AUC values because that was the only metric reported consistently.

This is not ideal. For example, we already know that magnetic resonance imaging has a high negative predictive value in prostate cancer. If I'm developing a model for detecting lesions, I would be interested in a model with a high positive predictive value because then that complements better what we're already able to do as radiologists. But that requires some

### To Facilitate Comparisons Across Studies, Should Researchers Present All Their Data Within Reason?

There will always be a limitation in machine learning because unless the model itself is available for implementation, with details on the pre-processing pipeline of the data, you will never be able to reproduce the result completely.

From the psychology reproducibility crisis, one of the concepts that have emerged is that reproducibility should not be limited to the reproduction of the experiment in and of itself. So taking the pipeline, taking the code in the case of machine learning, having the data set, clicking, and having the same result is useful, but it's of limited interest.

The idea is that if the concept behind the study is sound, if the idea at the basis of a prediction or a predicted model of a classification model or regression model is sound, one should obtain within a certain degree similar results even approaching the problem slightly differently. If the information is there for the exam type, for that lesion type (for example, if you're talking of oncologic patients as one of the most common applications), even if I'm not using the same method, if the theory is good behind this experiment, I should still obtain similar results because the information has to be there. Otherwise, if I am just modelling some random noise in my data set that's not present in your data set or another group's data set, then I would never be able to reproduce. If I give you my data and my model, you will be able to replicate my results. But those results may still not be true or not supported by a real theory behind the experiment. So we should present all the information to assess what's being produced by the model. Reproducing the specific experiment is only interesting up to a certain point. We should also aim to develop a more general understanding

of what we're looking at in the images; what those patterns mean. If there is a pattern that is informative in that lesion, then it should be informative regardless (within certain limits) of how I am looking at it, detecting it, or classifying it. That signal should be there.

In any case, we could have a more optimal solution that gains a little bit better accuracy or a less optimal solution that's less accurate. But if the information is there, it should still be evident even if we slightly diverge on the methods we're using.

So it's not the specific experience. It's more what's behind the experiment. If there's something real behind that experiment, then it should come up independently and from more groups – because there's something there that we're all looking at.

on the specific reasons why a specific model works for one outcome or not. There should be a greater effort in building up a good theory behind some of the applications of machine learning - why it works for a specific game that we have in mind.

(To explain) There is a large amount of data on a specific outcome, such as prostate imaging, breast imaging, and neuro-oncological imaging. Some fields already have a large number of studies that have been published. But they're always very small and narrow in their overview. We should start having some works that try to aggregate this data and look at the bigger picture. And try to develop a larger theory within each of these areas of why radiomics works or doesn't work for something. This is very challenging, but in the long-term, if we want

---

## If there's something real behind that experiment, then it should come up independently and from more groups – because there's something there that we're all looking at

---

### How Can This Strategy Address the Robustness and Replicability Crisis in the Literature?

From a more immediate point of view, we should raise the standards of what we expect from machine learning research in radiology. This process is already beginning because checklists have been developed by the editors and journals that are more specific to machine learning research than more general research checklists—these aid in ensuring that the correct amount of information is present in the paper. The includes the accuracy metrics that we talked about.

Also, there has been growing interest from various research groups, including my own, in using external tools to assess the quality of studies that have already been published. And the results of those efforts are usually not satisfying currently. The quality is generally found to be always very low across the board, independently of the application. There is a problem there. There is a small trend in improvement over the years, and we have to build upon that to obtain greater improvement.

In the short term, we have to continue raising the publication standards, especially on the more prestigious journals with the resources to implement more strict peer-reviewing. And maybe involve a technical editor for the more methodological aspects that may not be known to a clinical reviewer, that are usually involved in this process. Then from a more general point of view, we should develop the theory behind radiomics and machine learning.

For now, usually, research goes in this manner: You have an idea. You build a data set. And then you try, if you're able, to predict whatever you want to predict based on the idea that you had in the region. But only a few groups have tried to work

to make radiomics a robust field, it should have some theory and some understanding of how it works in a more general sense, not only because it works practically and empirically it stops there.

Something similar happened for functional brain imaging and brain connectivity. And there have been other areas in radiology where initial results then brought building up a more robust theory for what's going on in the brain. It is possible to take a more practical aspect and quantitative results and experimental science. To build upon that to obtain a more theoretical understanding of what is happening biologically. I think that's what we should aspire to as machine learning researchers. It might not be possible, but we should try at least.

### What New Directions Will Radiomics Take Within the Next Five Years?

In the next years, I think there will remain a high interest in radiomics for challenging tasks that radiologists cannot currently achieve: for example, genomic profiling, currently big profiling operations, and the prediction of outcomes at ten years. Research that's already going on right now will continue. I hope that there will be greater attention to the more practical side of things and more easily obtainable results that are clinically implementable and would allow for a real application of these tools in practice. Building trust between the radiologists and the tool, and the patient and the tool, will enable us to develop the necessary regulation and legal frameworks. Having simpler tools that are more easily verifiable will open the door for all the rest.



I hope that this realisation will become widespread. Not from academia, but from the companies? Working in this area, there is already a greater understanding of how to move forward. For example, even from improving image quality and speeding up image acquisition in MRI or lowering the dose, those are applications of deep learning to which radiologists are less aware. Companies are investing much in things that are practical and visible. Verifying the image quality is still diagnostic, and the information that we can get from those images is still useful. There is a tendency to go in this way from a commercial view, which I hope will drive the rest of the field. The problems it will solve in the near future in the next five years will be more practical: like speeding up the acquisition of the burden of repetitive and ring tasks.

### Do You Think That the Fear That AI May Replace Radiologists Is Justified?

Not really, because radiology is fairly complex and fortunately too complex for now to be substituted by an automated tool. If we're talking about if AI ever got to the point where it can substitute for radiologists, then there will be other issues to address; it will be able to substitute many other workplaces before radiology. There will probably be a reorganisation of society as a whole before that. In the field of medicine, other specialties are more immediately in danger from AI. For example, pathologists and other specialties that analyse images or also have these kinds of tasks. In that case, it's probably easier to develop tools that obtain similar results because it's more straightforward, and there's more homogeneity in the workflow. So I don't think that the fear is justified, even in the future.

As I said before, we are seeing what's happening even with self-driving cars. It's been ten years that self-driving cars are coming in the next five years. The hope and the expectations with AI are always too high compared to what it can do practically. Med students might not have sufficient knowledge both in radiology and in AI to correctly assess the situation. Until there is a shortage of radiologists, I would not worry much about it.

### Will Demand for Radiologists Decrease Because AI Will Increase Their Efficiency?

No, I don't think so. Radiology is also becoming more and more active on the interventional side of things, so there's a whole side of radiology that's completely not interested in this problem. The current proposed applications of AI completely ignore the more practical side of things.

I think there is a greater chance that maybe teleradiology and other technologies might reduce or redistribute the work in radiology before AI can impact. Because of the tasks that AI can do, I expect a very limited impact on most of the work radiologists do in clinical practice in small centres. Most of the work that's done right now is aimed at higher levels of

care and niche cases. Or even at increasing the number of exams that can be performed, which increases the demand for radiologists. Because if we can speed up MRI imaging from 40-minute exams to 10-minute exams or 5-minute exams, then instead of acquiring 20 MRI exams in one morning, we could acquire 100 MRI exams in one morning. Then we would probably need more people to report on those exams. I think it's very difficult to make predictions at this scale.

### How Do These Algorithms Become Commercialised?

Well, that's challenging. You need solid computer science people and software engineers. Around the AI model, you have to develop a whole software infrastructure that allows for data management. Because you input raw data and feed it to the model after the correct prognosis, you have to implement the whole pipeline, developed in the research setting, the user interface, all the user experience aspects, and integrate it with the current solutions. The challenge is that it requires the involvement of many other people from different fields.

If you have an idea and a product, and you're able to trademark it and register it, you can go to a company and then use their expertise. For example, several medical scanners and technology vendors are already buying up smaller companies or are working together with researchers to develop their own different solutions.

In actuality, there is already a large amount of software that's commercially available for radiology. Recently, there has been even a repository with an accompanying paper published in *European Radiology*, including solutions already having either FDA approval and or European CE marking for medical use. So there is a large amount of software.

It's challenging for a research group alone. Probably it will never reach that point without either expanding in a start-up company and building up the necessary infrastructure or working together with a larger company that already has the necessary know-how.

Regarding the concordance between the research and the commercial aspects, this review highlighted how out of 100 commercially available solutions, most did not have any research supporting their performance. So, when they come to propose a product, most have no research. Of those with research (36%), only half of the research was vendor-independent, not directly authored or sponsored by the software vendor. While it's true that software is commercially available, it's probably not true that there is sufficient peer-reviewed evidence to support their implementation. We should see the actual quality of that research – if it's reliable and reproducible, and all the things we have discussed in our previous questions and answers.

There are commercially available solutions. Companies have come to my institution to propose some of these. I think it's still too soon to implement them. Maybe some of the vendor solutions to speed up image acquisition timing are already

useable. For the rest, I would not invest any money in these solutions at this time as, most often, we will be early adopters. In any technology, it's not always a good position to be in because the early adopters also end up being beta testers. They end up paying for the privilege of using something that's optimally ready. I would still wait a little bit more. If I had to spend money at my department and was head of that department, I would not invest in any AI products right now. There are still probably more viable expenses before spending that in that area for now. Maybe they'll be more mature in four or five years and have more evidence to support their use. For now, I think it should still remain mainly in the research field.

### **Don't the EMA or FDA Need Data to Approve the AI Solutions?**

Most that have their approvals have it for technical feasibility, not based on clinical impact. They might have studies demonstrating that the results are reproducible and robust. It's not that they don't have any evidence. It may not be published, it's not openly available, and it does not undergo the classical external review process. They might have internal evidence that they may have produced to the legislative bodies. They can be used clinically, but most of them have no proven clinical impact.

Considering the United States, there's also a whole other discussion. In the last months, the first solutions have obtained the ability to be reimbursed by insurances. This is less an issue in Europe, because in Europe, usually, the final payer is, in large part, the state, at least, in Italy. There's always public coverage of most of the expenses. One of the questions is, who pays for AI? If it's valuable in Italy, the hospital pays for it.

In the end, usually, it's the national health care system. In the United States, reimbursement is not so easy. It's a challenge for these companies trying to get recognition from insurance companies and be reimbursable. Translating the research to the technical practice, and especially on the commercial side of things, is a whole other world. It's very challenging.

I don't know if there are any implementations of 'upkeep over time' and guarantees that if the data distribution changes at your institution, the vendor takes care of this. Who takes responsibility if the model stops working? Who covers the costs, for example, of retraining a model on updated data? It never ends if you want to go into that side of things.

I am a believer in this technology. I think these technologies do work and can work and should be implemented in radiology in the future. It's just that probably, right now, we are going a little bit too fast. This can be counterproductive in the long term because we are riding the hype wave right now. If we proceed too fast and don't work as expected, we would have a backlash. It would be a long-term negative outcome because these technologies have solid bases and can be implemented correctly.

I do not wish to give the impression that I'm negative at all. I work mainly in this research area. It would be hypocritical of me to say I don't believe in it. I believe they work, but we should be very careful how we implement and develop this kind of research.

### **Conflict of Interest**

Dr Cuocolo has reported no conflicts of interest.

Watch the full interview [here](#). ■