

"I'm doing twice as much CT reporting": The impact of Veye Lung Nodules at Salisbury NHS

PAGE 2

Deep learning applications in radiology: a deep dive on classification

PAGE 3

Who is the stricter regulatory parent? Clinical decision support software in the EU and the US

PAGE 4

Ten tips for running real-world research into the clinical utility of radiology AI

PAGE 8

# The MedAI Times

## "I'm doing twice as much CT reporting": The impact of Veye Lung Nodules at Salisbury NHS

Author: Catalina Barzescu, Aidence

News from the UK's National Health Service (NHS) has been highly concerning as of late. Patients are waiting for months for critical diagnostic tests, often medical images. At the same time, the radiology workforce is significantly understaffed and under immense pressure to clear the backlog.

Dr Katharine Johnson has been practising radiology within the NHS for over 20 years. She thinks radiologists are currently severely time-constrained.



PAGE 2

"AI will make you a faster, more accurate reporter. But it's up to us radiologists to ensure that the AI we dopt is fit for purpose, safe, and worth the cost."

PAGE 2

Dr. Katharine Johnson

However long you think it will take, double it, and then double it again!

Get the right advice early on.

Understand the decision authority to proceed.

Ten tips for running real-world research into the clinical utility of radiology AI

Author: David King, Aidence

PAGE 6

## Who is the stricter regulatory parent? Clinical decision support software in the EU and the US

PAGE 4

Author: Ricardo Roovers, Aidence

Medical technology is rapidly evolving, challenging regulators to maintain oversight while also encouraging innovation. The EU and the US don't always take the same path to achieve this balance – like parents who may have different views on how to best nurture their children.

### Veye Lung Nodules

Your one-click AI solution for pulmonary nodule management



PAGE 6

### Deep learning applications in radiology: A deep dive on classification

Author: Valerio Fortunati PhD, Quantib

PAGE 3

# "I'm doing twice as much CT reporting": The impact of Veye Lung Nodules at Salisbury NHS

**Author:**  
Catalina Barzescu, Aidence

## The challenges

News from the UK's National Health Service has been highly concerning. Patients are waiting for months for critical diagnostic tests, often medical images. At the same time, the radiology workforce is significantly understaffed and under immense pressure to clear the backlog.

Dr Katharine Johnson has been practising radiology within the NHS for over 20 years. She thinks radiologists are currently severely time-constrained:

**"There's far more imaging than there's ever going to be radiologists. It's not possible to keep up. There is significantly less time to clinically assess patients such that imaging is often the primary diagnostic tool."**

Dr Johnson has also noticed that her hospital's current lung nodule follow-up has been challenged in its post-pandemic recovery. The ideal follow-up for patients considers the British Thoracic Society (BTS) guidelines, which recommend using

volumetry rather than diameter to analyse and report on the growth of a lung nodule. However, performing the measurements relies on an imperfect method:

**"Using an axial measurement and putting on callipers manually is inherently inaccurate. I've always used volumes for lung nodules analysis, but calculating volumes is laborious and time-consuming."**

She is, thus, advocating for a better way to manage lung nodules on chest CT scans.

## The solution

In 2020, Katharine Johnson helped set up a lung cancer screening service at the University Hospital Southampton NHS Foundation Trust. The initiative is part of NHS England's Targeted Lung Health Checks, a programme that has so far caught over 1,000 cancers at an early stage.

Most of the areas involved in the checks use artificial intelligence (AI)

for lung nodule management, and Aidence's Veye Lung Nodules is the preferred solution. Veye Lung Nodules automatically detects, measures, classifies, and tracks the growth of pulmonary nodules, then delivers its analysis into the original study.

Dr Johnson has been working with Veye for over three years. She got accustomed to the AI tool when reporting on screening CTs at Southampton. In January 2022, she also started using it at Salisbury hospital for all high-resolution diagnostic chest exams.

**"We use Veye Lung Nodules routinely on plain chest CTs. When I sit down to look at these scans, they already have nodules highlighted, and volumes applied. If there's a relevant previous scan, they already have the volume doubling time (VDT)."**

The adoption of Veye Lung Nodules at Salisbury is supported by the Cancer Alliance and has recently been extended for another year.



(FIG. 1) EXAMPLE OF A REPORT GENERATED BY VEYE LUNG NODULES

## The results

Dr Johnson has greatly benefited from the use of the AI solution in her daily work:

**"I have reduced my chest CT reporting time by about 50%. I'm doing twice as much reporting. I think it's brilliant."**

The potential to report on CTs significantly faster with Veye Lung Nodules is supported by clinical research. In a study published in the European Journal of Radiology Open, two radiologists achieved an average 40% reduction in reading time when reporting pulmonary nodules and reached a higher level of agreement.

Veye's automated lung nodule volumetry is particularly useful, says Dr Johnson:

**"The volumetry is fantastic! It saves me from manually performing the analysis. With Veye, we're also likely to reduce human error and increase consistency between different readers."**

**"A less subjective and more quantifiable report is more reliable. Ultimately, we have a much more robust nodule follow-up in half the time".**

CONTINUES BELOW, LEFT ▶

## ▶ CONTINUES BELOW, LEFT

She also shares a clinical example of Veye's impact on patient care:

**"A man presented with severe interstitial lung fibrosis. Veye Lung Nodules highlighted an area in the background of the lung disease, which I suspect I may have overlooked. It turned out to be early-stage lung cancer."**

**"In this case, Veye has helped us arrive at the diagnosis robustly. For the patient, it means curable disease with a small operation. An enormous impact!"**

Dr Johnson acknowledges that Veye Lung Nodules is not a perfect tool; it occasionally returns false positives and does not detect larger masses. IT integration is demanding at times. Good service is essential, and Aidence has been able to provide it:

**"The Aidence team has been incredibly responsive. If there's a problem, they've gone above and beyond to sort it out."**

## "Up to us"

Dr Katharine Johnson admits she was initially "a bit cynical" about the use of AI in radiology. She has definitely come around and encourages her colleagues to be the ones driving AI adoption:

**"AI will make you a faster, more accurate reporter. But it's up to us radiologists to ensure that the AI we adopt is fit for purpose, safe, and worth the cost."**

What does the future hold? Autonomous AI dismissing scans with no actionable findings might be the most promising use case in medical imaging:

**"It would be great if a computer could dismiss 'normal' so I can focus on the cases that need my attention."**

We are not there yet, yet testimonials like Dr Johnson's make all the work to achieve it more than worth it.

## Background

About Dr Katharine Johnson, BSc.  
MBBS MRCP FRCP

- \* Consultant General Radiologist since 1999;
- \* Special interest in thoracic imaging;
- \* Lead for thoracic imaging and lung multidisciplinary team meetings at Salisbury NHS Foundation Trust;
- \* Helped set up the lung cancer screening programme (Targeted Lung Health Check) at the University Hospital Southampton NHS Foundation Trust.

About Salisbury NHS Foundation Trust

- \* Award-winning trust;
- \* Delivers a broad range of clinical care to approximately 270,000 people in Wiltshire, Dorset and Hampshire.



# Deep learning applications in radiology: A deep dive on classification

**Author:**  
Valerio Fortunati PhD, Quantib

Classification methods divide the input they are handed into groups or "classes". For example, an algorithm which assesses a mammogram and provides an answer to the question whether a patient has breast cancer (yes or no), is a classification algorithm. Additionally, any algorithm that outputs a certain discreet scoring, such as BI-RADS, is also a classification algorithm. Deep learning (DL) is extremely suited to

automate the solving of these classification problems. But which DL networks work best for these tasks? And what are the main characteristics of these networks?

In "old-fashioned" networks, features are designed manually. This means that the developer needs to sit down, dive in and figure them out.

## Why use deep learning for classification in radiology?

Neural networks are very suited for classification. This has to do with the way the algorithm extracts features.

Deep learning networks, on the other hand, ease this task by figuring out, based on the data, which features are important. The only thing you have to do is feed the data to the network (including the ground truths) and let the training procedure 'teach' the network (i.e. by updating its weights) how to best perform the task at hand. (Fig. 2)

## Which deep learning networks are the best for classification in radiology and how do they work?

### AlexNet: the starting point

AlexNet is a deep convolutional neural network (CNN) that was created in 2012 and has a surprisingly good performance in classifying images into 1000 classes. AlexNet combined a whole lot of new techniques to the training of a CNN that became pretty much standard in DL. For example, applying dropout, using multiple GPUs to cut back on training time, applying data augmentation and using a specific activation function during training which also dramatically decreased time needed to create the algorithm.

There is research that applied AlexNet applied to medical images with the objective of classifying skin

lesions into seven different classes. The strong performance metrics show the suitability of such an algorithm for classification tasks.

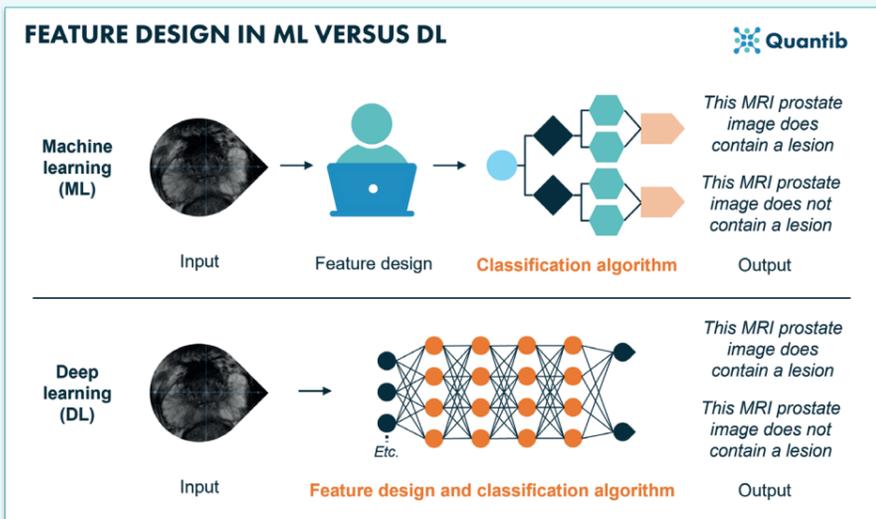
### VGGs

The VGG neural networks appeared with the main goal of increasing the depth, i.e. the amount of layers of the neural network. However, because of its depth, the VGG takes longer to output results. What makes VGGs different is that they use very small kernels, that is taking information from just a few voxels, that they then push from one layer to the next. (Fig. 3)

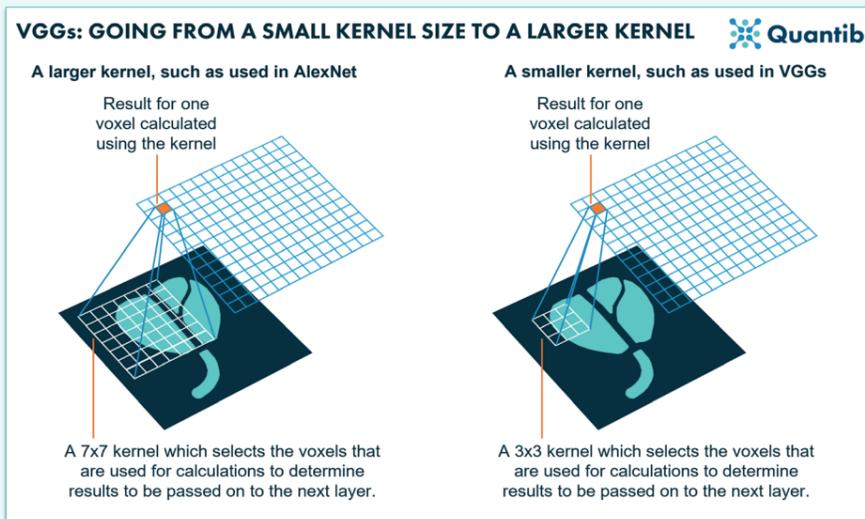
As an example, a recent study used a VGG model to review chest-X-rays (CXR) to detect the presence of COVID-19 suspect regions. The performance results suggest its suitability for this classification tasks.

CONTINUES BELOW, LEFT ▶

(FIG. 2) MACHINE LEARNING REQUIRES MANUAL FEATURE DESIGNING BEFORE THE TRAINING OF THE ALGORITHM CAN START. DEEP LEARNING COMBINES FEATURE EXTRACTION AND THE TRAINING OF THE ALGORITHM USING A NEURAL NETWORK.



(FIG. 3) ALEXNET USES RELATIVELY LARGE KERNELS WHILE VGGs USE RELATIVELY SMALL KERNELS<sup>4,5</sup>

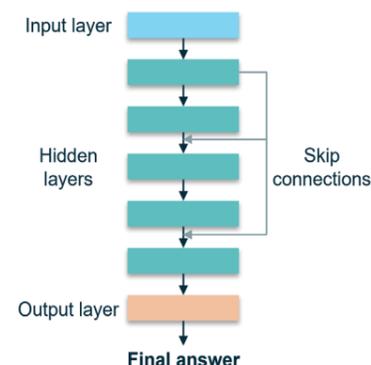


▶ CONTINUES

## The ResNet family

Currently, the most used algorithms for classification are part of the ResNet family. We stick to the design of CNNs, but with the added improvement of "identity shortcut connections" or "skip connection" between the layers. These connections take information from a specific layer in the deep neural network and feed that directly into the final result or to a layer further down the network. This enables each layer of the network to learn extra information and they simplify and accelerate the training. (Fig. 4)

### RESNET: INTRODUCTION OF SKIP CONNECTIONS



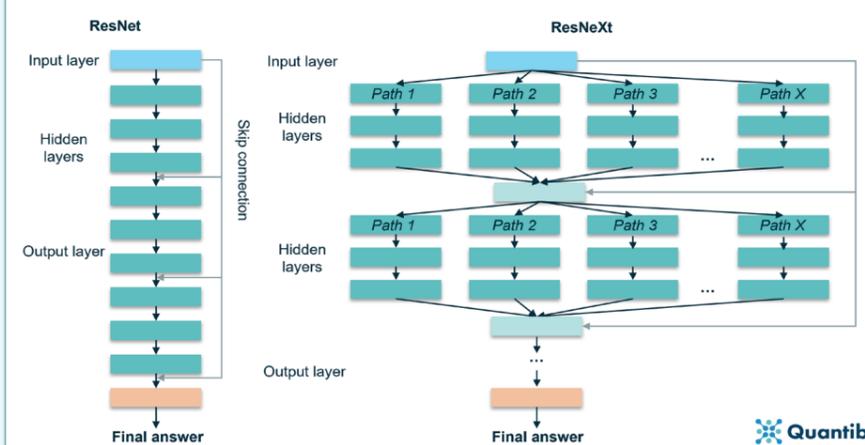
(FIG. 4) A RESNET USES SKIP CONNECTIONS WHICH AVOIDS LOSING TOO MUCH OF THE ORIGINAL INPUT INFORMATION IN THE NETWORK WHICH WOULD MAKE THE ALGORITHM STRONGLY DIVERGE.

ResNext is one of the newer versions of high performing CNNs. A ResNext actually combines multiple paths of stacks of hidden layers. Research shows that enlarging the network in this way actually leads to better performance than increasing the number of layers in a network. (Fig. 5)

CONTINUES ON PAGE 4 ▶

(FIG. 5) A RESNEXT FEEDS THE INPUT INTO MULTIPLE PATHS. EACH PATH IS TRAINED IN A SLIGHTLY DIFFERENT WAY GIVING THE ALGORITHM AN OPPORTUNITY TO LEARN MORE AND THEREFORE PERFORM BETTER.

### RESNEXT: MULTIPLE PATHS OF THE SAME RESNET BLOCK



► CONTINUES FROM PAGE 3

Researchers trained a ResNet to predict the probability of breast cancer in patients based on DCE-MRI. Performance checks on both an internal test set as well as on a retrospective study show good performance.

### Promising new techniques that are still to be applied to medical imaging

#### Vision Transformers

Vision transformers (ViTs) are fundamentally different to the deep neural networks discussed in the examples above. They do not use convolutional layers, they use transformer layers. A transformer layer inputs patches, which are sets of voxels in a square. The layer compares each patch to all the other patches in the image, thus checking the relative importance of each set of voxels versus all the other same-sized sets of voxels throughout the complete image. This means that the network could learn relevant relationships between parts of the image at any scale. Hence, it can simultaneously tell you something about structures that cover a large part of the image, but also, for example, reveal a correlation between small lesions that can be found in multiple locations. Additionally, you are not constrained by using the

same kernel for the whole image which allows for more flexibility as well.

#### ConvNext

The ConvNext network is a combination of a ResNet and a ViT with the aim of implementing the latest techniques of ViT to the CNN architecture.

### Conclusion

A wide range of methods has been developed with the purpose of performing the task of classification and research on how to optimize artificial intelligence methods for medical image processing is moving forward with grand strides. DNNs, such as AlexNet, VGGs and those of the ResNet family show vast results when applying them to diagnostic images.

Other promising, more recent developed algorithms, including ViT and ConvNext methods, have been successfully developed in several fields. Although they are still too premature to apply to medical images, there is great promise in these types of modern algorithms. Upon availability of more computing power and bigger, good quality datasets. ◀

### Curious to learn more AI in radiology?

Read our Ultimate Guide to AI in radiology

Scan the QR code



or go to <https://www.quantib.com/the-ultimate-guide-to-ai-in-radiology>

#### Find sources on:

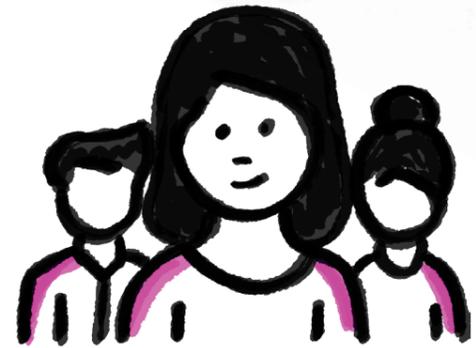
<https://www.quantib.com/blog/medical-image-segmentation-in-radiology-using-deep-learning>

Originally published by Quantib.

Read more articles on <https://www.quantib.com/blog>



aidance



## We are Aidance

Driven by the purpose of giving lung cancer patients a fighting chance.

Part of RadNet's AI Division

# Who is the stricter regulatory parent? Clinical decision support software in the EU and the US

**Author:**  
Ricardo Roovers, Aidance

Medical technology is rapidly evolving, challenging regulators to maintain oversight while also encouraging innovation. The EU and the US don't always take the same path to achieve this balance – like parents who may have different views on how to best nurture their children.

The US Food and Drug Administration (FDA) recently published a new set of guidance documents regarding Software as a Medical Device (SaMD) and its classification. In this article, I take an imaginary example of a software product and run it through the applicable regulatory governance in these new guidelines.

By comparing FDA's approach to the EU's, I highlight some significant differences, aiming to open a debate around which framework is stricter and will better safeguard medical applications.

### Understanding the terms

My focus is solely on the guidelines regarding Clinical Decision Support Software (CDS), Medical Device Data Systems (MDDS) and Device Software Functions and Mobile Medical Applications. Some examples of products that fit these terms are:

**CDS:** Software that analyses patient-specific measurements to identify patients potentially experiencing a time-critical disease.

**MDDS:** Software that collects output from a medical device and transmits the information to a central patient data repository.

**Device Software Functions and Mobile Medical Applications:** Software that is able to control a connected medical device.

As a disclaimer, I will *not* touch upon Computer-Assisted Software systems such as CADe, CADx or CADt devices, for which separate guidances and definitions apply.

### Imaginary Report

Let's consider a fictive medical data reporting tool, with the following intended use:

**"ImaginaryReport" functions as a report for physicians to fill in and store medical information about patients. ImaginaryReport also incorporates information transferred from connected software medical devices into this report.**

**Based on this information and on clinical guidelines, ImaginaryReport calculates a certain metric (routinely used in clinical practice) for diagnosis. Based on this metric, the medical**

**information, and another clinical guideline, ImaginaryReport provides the physician with a recommendation for a diagnostic procedure for the patient. ImaginaryReport is not intended to replace the clinical judgement of the physician.**

The ImaginaryReport, thus, has three functions:

1. Transfer, storage and display of medical information from physicians and connected medical devices.
2. Calculating a metric routinely used in clinical practice for diagnosis, based on clinical guidelines.
3. Providing a recommendation for a diagnostic procedure, based on existing clinical guidelines.

I will analyse each of these functions against the existing legislation and guidance, to determine whether these

functions qualify as a medical device within the US and the EU, and, if so, what their respective classification would be.

### Function 1: Transfer and storage of information from physicians and connected medical devices

#### The EU approach

Article 2 of the EU Medical Devices Regulation 2017/745 (MDR) contains this definition:

**Medical device means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes:**

**– Diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease,  
– Diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,**

CONTINUES ON NEXT PAGE ►

– Investigation, replacement or modification of the anatomy or of a physiological or pathological process or state,  
– Providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations.

Our focus is on software, thus, Non-Device-MDDS. FDA clarifies that Non-Device-MDDS is not considered a medical device and not subject to FDA's regulations.

Function 1 of ImaginaryReport is solely intended to transfer, store, and display medical device data. So, applying the FDA MDDS Guidance, **Function 1 would also not be considered a medical device in the US.**

**Comparing the EU and the US**

Regarding the first function of our imaginary medical device, the EU and the US take a similar regulatory approach. (Fig. 6)

**Function 2: Calculating a metric routinely used in practice for diagnosis, based on clinical guidelines**

**The EU approach**

Rule 11 of Annex VIII of the MDR clarifies that:

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa.

Additionally, Section 3 of the MDCG 2019-11 does not restrict this function from being regarded as a medical device. Annex 1 of the MDCG 2019-11 explicitly mentions that:

Decision Support Software are qualified as

medical devices. These are computer-based tools which combine general medical information databases and algorithms with patient-specific data. They are intended to provide healthcare professionals and/or users with recommendations for diagnosis, prognosis, monitoring and treatment of individual patients.

Since **Function 2** aims to provide the physician with additional diagnostic information, calculated by the device itself, it fits the definition of a medical device under the MDR.

**The US approach**

The second guidance updated by the FDA is the Policy for Device Software Functions and Mobile Medical Applications, explaining its oversight over:

1. Software functions that are an extension of one or more medical devices by connecting to such device(s) for purposes of controlling the device(s) or analysing medical device data;

2. Software functions (typically, mobile apps) that transform the mobile

platform into a regulated medical device by using attachments, display screens, or sensors or by including functionalities similar to those of currently regulated medical devices;

3. Software functions that become a regulated medical device by:  
a. performing patient-specific analysis and providing specific output(s) or directive(s) to health care professionals for use in the diagnosis, treatment, mitigation, cure, or prevention of a disease or condition, or;  
b. performing patient-specific analysis and providing patient-specific diagnosis or treatment recommendations to patients, caregivers, or other users who are not health care professionals.

It further sets out criteria for software functions that are not subject to regulatory oversight. These may meet the definition of a medical device. However, FDA intends to exercise **enforcement discretion**, i.e. not enforce the regulations for them, because they pose a low risk.

These functions either:

- Provide or facilitate supplemental clinical care, by helping patients manage their health in their daily environment;
- Are specifically marketed to help patients communicate with health-care professionals by capturing an image for patients to convey to their doctors about potential medical conditions;
- Perform simple calculations routinely used in clinical practice. (Note: the FDA does not clarify the thresholds for "simple calculations" and "routine use in clinical practice", leaving this point open to a certain level of interpretation.)

Function 2 of ImaginaryReport meets the third point of performing simple calculations. Subsequently, **Function 2 may be considered a medical device for which the FDA does not intend to enforce regulations.**

**Comparing the EU and the US**

Here, there is a clear distinction between the governance of the software in the US and the EU. The EU is taking a more cautious approach compared to the US. With the MDR and Rule 11 in place, many of these software devices will be classified at least as class IIa medical devices. (Fig. 7)

**Function 3: Providing a recommendation for a diagnostic procedure, based on clinical guidelines**

**The EU approach**

The same interpretation as for Function 2 applies to Function 3.

(FIG. 6) EU VS US, FUNCTION 1



(FIG. 7) EU VS US, FUNCTION 2



**Function 3 would also render the product a medical device**, of at least class IIa.

**The US approach**

The third guidance recently released by the FDA is the guidance on Clinical Decision Support Software. CDS is a broad concept which encompasses various types of products and functions, such as:

- Computerised alerts and reminders for providers and patients;
- Clinical guideline automation;
- focused patient data reports and summaries;
- diagnostic support; and
- contextually relevant reference information.

This type of software can either be a regulated medical device or can be considered a (non-medical) device which is not regulated by the FDA. CDS not considered to be medical devices are those which meet the four criteria set out in the guidance, as derived from Section 520(o)(1) of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 360j):

1. Not intended to acquire, process, or analyse a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system;

2. Intended for the purpose of displaying, analysing, or printing medical information about a patient or other medical information (such as peer-reviewed clinical studies and clinical practice guide lines);

3. Intended for the purpose of supporting or providing recommendations to a health care professional about

prevention, diagnosis, or treatment of a disease or condition;

4. Intended for the purpose of enabling such health care professional to independently review the basis for such recommendations that such software presents so that it is not the intent that such health care professional rely primarily on any of such recommendations to make a clinical

diagnosis or treatment decision regarding an individual patient.

ImaginaryReport's function 3 would meet all four criteria:

- It is not intended to acquire, process, or analyse a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system;
- It displays medical information, namely the outcome of a clinical practice guideline;
- It provides a recommendation to a healthcare professional about the diagnosis;

CONTINUES ON NEXT PAGE ▶

► CONTINUES FROM PAGE 5

- It is based on a clinical guideline, and the physician can always consult the clinical guideline to verify or determine the recommendation. This is also communicated to the physician within ImaginaryReport.

As a result, **Function 3 would not be considered a medical device.**

### Comparing the EU and the US

As in the previous section, the FDA is less concerned with these devices than the EU, making them easier and faster to bring onto the US market than the EU. (Fig. 7)

### Different turns

Where does this analysis leave our imaginary product? **ImaginaryReport would not be considered a medical device in the US. However, in the EU, it would be at least a class IIa medical device.**

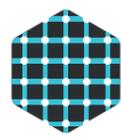
Which benefits patients more? Does the lack of regulatory control in the US also mean that there is an inherent safety risk for patients?

As a medical device manufacturer, we argue it is important to be subjected to appropriate documentation of

requirements, verification and validation testing, and risk management assessments, irrespective of regulatory needs. Even though these types of devices apply known algorithms and medical protocols, there can always be risks, e.g. incorrect interpretations of clinical classification scores, interface usability aspects, or risks associated with implementing the protocols into the software.

The new FDA documents provide a clear framework for lower-risk classes of medical device software, which was not always straightforward to understand. Yet, as noted, the FDA insufficiently defines what "simple" calculations are or when a guideline is "routinely used" in clinical practice. In the EU, the MDCG guidance also leaves room for interpretation since the specific example of implementing a clinical guideline or known risk scores is not addressed.

The two regulatory authorities have taken different turns regulating a product such as ImaginaryReport. FDA's leniency could stimulate innovation in the field of CDS. But, the EU's stricter approach could set CDS to higher standards. It remains to be seen which parenting style fits the development of CDS over the coming years. ◀



aidance

## Veye Lung Nodules

Your one-click AI solution for pulmonary nodule management

Learn more at [www.aidance.com](http://www.aidance.com)



# Ten tips for running real-world research into the clinical utility of radiology AI

**Author:**  
David King, Aidance

A longtime user of our AI solution, Veye Lung Nodules, once said he "feels naked without it". This indicates its clinical utility – bringing reassurance, confidence, or comfort to the radiologist – and we were happy to hear it! But much work must be done to translate such statements into peer-reviewed evidence of Veye's benefits in clinical practice.

At Aidance, we are building an "evidence dossier" for our AI medical devices. In a previous article, my colleague Maurits explained the framework we use to do so, outlining our existing studies and the work in progress. One of the latter is INPACT, a joint evaluation programme looking into the effect of AI on radiology decision-making in lung cancer care.

INPACT is made possible with funding from the UK's National Health Service (NHS) through the AI in Health and Care Award. Should it demonstrate a positive impact on healthcare outcomes and cost-effec-

tiveness, it will support the national adoption of our solution. It may improve trust and encourage the implementation of proven tools that can make a lasting difference in people's lives.

Setting up INPACT has been a unique, complex undertaking and an excellent opportunity for us as a vendor of emerging technology to learn and derive best practices, particularly within the NHS context. In this article, we're sharing our top tips, hoping to support further research into AI in radiology towards the benefit of physicians and patients.

### INPACT basics

The INPACT study is performed in the radiology departments of six hospitals in the UK. Radiologists first analyse chest CTs unaided by Veye Lung Nodules, then provide a consecutive reading with access to Veye's results. A radiology expert at each centre independently evaluates all

cases where one or more nodules were identified by either the radiologist or Veye and compares the radiologists' performance and confidence level. The sample consists of up to 750 unique chest CT scans per hospital, adding up to approximately 4,500 cases over six months.

We're working on this study with the University of Edinburgh and clinical consultancy Hardian Health. The university is leading the technology-specific evaluation team (TSET), an independent academic body appointed by NHS England to oversee the programme and provide objective reporting of results. Hardian Health backs up the research for the health economics workstreams and also works closely with the six sites.

We've had strong support from the AI Award team and the academic TSET, which has been invaluable in helping to navigate challenges along the way.

So, what have we learned?

### Our top tips (Fig. 8)

#### 1. However long you think the project will take, double it, and then double it again!

Study design, research governance, registration and approvals, technology installation, ethics applications, participant recruitment, training, and study setup all have significant lead times. Given their interdependencies, they will have knock-on effects on the calendar.

So, build in realistic timelines from the start and validate these with the experts before committing to a "final end date".

#### 2. Get the right advice early on.

Real-world research, by definition, means testing a research hypothesis in a live clinical setting, where it will likely have a direct impact on the patient. Thus, it can be fraught with complications, such as blurring the distinction between research and service evaluation.

Framing the research questions and study design are non-trivial tasks. There are plenty of sources of guidance that you could, and should draw from. Here is a list of the UK advisory services:

- The Health Research Authority
- The NHS AI Lab
- The AHSN Network
- The NIHR Clinical Research Network
- The NHS Innovation Service
- The multi-agency advisory service (MAAS) for AI and data-driven technologies

In our case, talking to AI specialists at the UK Health Research Agency helped us frame the problem and finalise our study design.

#### 3. Keep an open mind on research questions and design.

In medical imaging research, reader studies are often the go-to design for testing a particular research question. Nonetheless, they remain artificial constructs. Consequently, their utility in real-world research is perhaps limited. So, you may want to take a more creative approach to designing a study protocol.

The critical question is: how can we balance the capture of data in a real-world clinical context with minimising the burden on the already stretched clinical workforce? A randomised controlled trial may be the gold standard for measuring outcomes and impact. Yet, it is not always desirable, practical, or affordable.

CONTINUES ON NEXT PAGE ►

► CONTINUES FROM PAGE 6

There is no one-size-fits-all approach. It's best to play around with ideas and assess the strengths and weaknesses of different designs. Many advisory bodies (see point #2) can also provide valuable insights.

#### 4. Map out all the parties that need to be engaged.

Moving from identifying an evidence gap to concluding and publishing a research study is a long journey, measured in years, not months. Many stakeholders will get involved along the way. This includes a mix of regulation, advisory, support and execution, with authority to allow or block a project from moving to the next stage.

Even within a single institution, like an NHS hospital, there will be multiple people to engage – research coordination, technology teams, finance, information governance, clinical and procurement, to name but a few. Much as we'd like a single point of contact in cases like this, that's not the reality, so wide-ranging engagement is absolutely essential.

Start your project by sketching a mind map of everyone that needs to be involved. And then ask them all who else needs to participate, and how. Repeat this exercise until the mind map stops growing.

#### 5. Involve frontline clinicians from the start.

We have been fortunate to have a good plan for our real-world research agreed upon early on with our TSET and the NHS AI Award team. But, in the build-up to starting our study and its execution, we are constantly learning new things. These are minor tweaks and improvements to, for instance, our inclusion criteria or choice of data categories.

The majority of these insights come from the clinical teams involved in the study. Although they represent issues the study team might have been able to anticipate, the clinical end users of our technology are much better placed to challenge us and spot such things up front.

Thus, if you want to collect real-world clinical evidence, make sure you have real-world clinicians advising your study team. In our example, a dedicated clinical advisor works with our clinical investigators, but many other models could work equally well.

#### 6. Keep it simple and focus on your primary research goals.

As our study moved from design to execution, it was not uncommon to ask ourselves questions like "[x] thing is interesting, maybe we should explore [y]?" or "Would [a] be able to tell us anything more about [b] if we do [c]?". Be warned – that way leads to scope creep! It requires discipline to stay focused on the primary aims of the research, and, fortunately, our TSET has maintained a laser focus on these aims.

Before commencing a real-world study, do one last sense check of the primary research questions and the study's aims. Are these truly the questions you want answers to? If so, stick to them, but don't discard other ideas you might have as you



(FIG. 8) TEN TIPS FOR RUNNING REAL-WORLD RESEARCH INTO THE CLINICAL UTILITY OF RADIOLOGY AI

progress. Collect them somewhere for consideration in future research.

#### 7. Understand the decision authority to proceed.

The hardest part of our research journey (so far, at least) has been getting started. We had sketched out some initial timelines which we thought were realistic (see point #1). Still, we quickly realised that getting the green light to proceed involved navigating several stakeholders (point #4), which all took time.

A further complication was that our project was sponsored by a Scottish body, with different governance arrangements from the English partner sites. At times, even our TSET (with its vast institutional experience of running clinical research) was unclear on who had authority over which decision. It doesn't help that many of the decisions required to initiate clinical research are interdependent. A won't sign B until C has signed D, which depends on A approving E.

The takeaway: document carefully, in the form of a decision tree, who has decision authority over your project, and how that authority is exercised. Note the lead time, prerequisites, and any interdependencies. Doing this upfront will take time, but your project will run much more smoothly in the long term.

#### 8. Incentivise the research effort.

To run a real-world clinical study, you need real-world clinical people. However, NHS clinicians are under immense pressure from multiple systemic challenges. And, in almost every aspect of care, they go above and beyond.

Therefore, one of our early principles

was not to take their participation in our research for granted. We recognise the time and commitment required to support a study like ours. In some cases, this might simply be a recognition of their efforts in any resultant research publications. But principally, we want to ensure the financial incentive is appropriate—institutionally and personally—for the extra effort required to support the research.

If your research involves the use of clinician time (which it invariably will), discuss with them what a fair system of remuneration and recognition might look like and build into your study process.

#### 9. Think about the meaning of "real world."

This might sound abstract, but it is, nonetheless, key to understanding the value of any research. You could have the best and most performant AI algorithms in the world and prove, without question, that clinicians using them will be consistently more accurate or work faster. But that doesn't answer the pivotal question: "So what?"

What does that mean in practical terms? How many patients will benefit, and how? What does it do for costs and efficiency? Is the solution easy to use in practice, and will it actually be used?

That is why it is so important to think of clinical utility and cost-effectiveness in the context of real-world clinical care. We discarded many of our early research ideas because they told us nothing about what would happen in real life.

You may not be able to reliably test a solution in every real-world dimen-

sion. So, be clear about which dimensions are important for your evidence gaps and design around those. These may be efficiency, clinical decision-making, ease of use, potential bias, etc.

#### 10. Don't plan to do everything in sequential steps.

Our final advice is one of practicality more than anything else. As an AI Award-funded programme, we are expected to wrap up our project in line with the funding award agreement. However, as timelines have drifted to the right (point #1 again), there is less and less time at the end of the project for data collation, results analysis, interpretation, and reporting.

Our solution is to bring forward as much work as possible from these final stages of the project to minimise the time required for the last steps. That means prepping our analytics work with dummy data, pre-writing elements of our research papers that are not dependent on the actual results, and generally ensuring that we can still hit our deadline for deliverables without compromising the study's integrity.

It's an old cliché, but there is merit in "starting with the end in mind". If that means drafting a mock-up scientific paper before you even start your study, that's likely to be a good course of action. It will expose any weaknesses in your methodology or data strategy and ensure everyone is sighted on what you need to achieve your aims. Don't wait until the end of the project to start thinking about what it all means.

#### Never forget the patient!

This is our "bonus" eleventh tip.

When conducting scientific research, it is easy to get lost in theory, governance, and statistics. So it behoves researchers to continually focus on those who matter most: the patients and citizens at the receiving end of all medical technology innovations.

It is best practice to involve patient and public representation throughout a research programme, from conception and study design to execution and analysis. In the NHS, there are plenty of resources to help create effective patient-focused research. We recommend the NIHR Learning for Involvement website as a great place to start.

Active engagement with patients and the public helps us never lose sight of our driving mission: to give all lung cancer patients a fighting chance.

#### Thanks are due

All the learnings we derived from setting up a research programme into AI within the NHS will help us to grow as a company and streamline our following projects. They may also allow other technologies facing a similar conundrum to, for instance, understand how and where to get started.

Preliminary results of INPACT will be made available to the NHS AI Award team in March 2023, and we will publish our findings thereafter. We want to thank all the organisations involved, especially the radiologists in the study sites. They are helping us build the body of evidence that will realise patient benefits across the whole NHS. ◀