Artificial intelligence for analysing chest CT images

Medtech innovation briefing

Published: 5 January 2021

www.nice.org.uk/guidance/mib243

Summary

- The **technologies** described in this briefing are artificial intelligence (AI) technologies for chest CT. They are used for assisting with triaging, reporting, and identifying abnormalities.

- The **innovative aspects** are that the software helps radiologists and radiographers detect abnormalities in chest CT images.

- The intended **place in therapy** would be to support radiologists and radiographers when reviewing chest CT images in secondary care for people who have been referred for chest CT.

- The **main points from the evidence** summarised in this briefing are from 2 retrospective studies. The best quality evidence came from 1 UK study showing that Veye Chest (Aidence) performed similarly to chest radiologists for lung nodule segmentation growth assessment. The studies were limited in quality and no studies were published in full.

- **Key uncertainties** around the evidence for CT AI are that more published evidence is needed to show how the technologies perform compared with radiologists, and how they impact clinical management and outcomes in CT reporting. Generalisability may also be limited if the AI algorithm is trained on a non-local population. The technology will not be used instead of a radiologist but as a tool to aid faster diagnosis. Further research is needed to understand the risks of AI automation.

- The **cost** of the technologies varies significantly. icolung is offered pro bono as part of the icovid.ai initiative. Veye Chest costs between £5 and £7.50 per output (depending on selected features and volume of scans). Veolity (MeVis) offers a perpetual licence for £44,000 with no per scan costs. The 2019/20 national tariff for a CT scan of 1 area is £69 for people aged over 18 and £73 for people aged between 6 and 18. The cost of reporting is £20 for all ages. The use of AI technologies has the potential to reduce resource use by helping reduce the workload of staff reading CT images.

The technology

The chest imaging artificial intelligence (AI) technologies in this briefing are standalone software platforms that use machine or deep learning algorithms to analyse or interpret radiology images. Some technologies allow images to be transferred from the hospital to the software platform, which is hosted in an NHS accredited secure data centre. The software analyses the chest DICOM (digital imaging and communications in medicine) image using proprietary algorithms. The image analysis may be sent directly back to the hospital to be viewed with hospital systems such as picture archiving and communication system (PACS) and some radiology information systems using protocols such as DICOM and HL7. Some technologies may also allow uploading and viewing of images and analysis using a web interface.

Version updates and periodic maintenance activities are needed for these technologies. This can be done remotely.

The technology may help identify images as normal or abnormal, highlight suspected abnormalities and provide results as heat maps or clinically relevant labels. It may also provide support for prioritising CTs for specialist review. The AI analyses are intended to be used with radiology images to support radiologist review and decision to improve diagnostic accuracy. Turnaround time may be decreased for time sensitive conditions such as pneumothorax or catheter malposition. They are not intended to be used as medical advice.

The following technologies are post-processing image analysis software for chest CT. Other, similar technologies may be available but are not included in this briefing (for example if they were not identified, or the company chose not to participate).

- **Veye Chest (Aidence)** – used for automatic detection, classification, measurement and growth assessment of solid and sub-solid pulmonary nodules. Can be used on low-dose or standard-dose, and non-contrast or post-contrast scans with a maximum axial slice thickness of 3 mm.

- **icolung (icometrix)** – used for automatic detection, segmentation and measurement of lung abnormalities in 5 lung lobes in non-contrast scans. Has a maximum axial slice thickness of 5 mm (about 1 mm thickness is recommended). The company states that version 0.6.0 onwards is expected to provide support for COVID-19 diagnosis by including the probability of an image being from a person with COVID-19.

- **Veolity (MeVis)** – used for reading chest CTs including automatic detection, segmentation and measurement of pulmonary nodules. Can be used on low-dose or standard-dose, and non-contrast or post-contrast scans. The algorithm automatically compares the current scan with previous scans and assesses change in nodule size.

Innovations

AI for analysing chest CT may help increase diagnostic accuracy and reduce time to diagnosis by providing additional information for radiologists. The technology automatically reads medical images and identifies abnormalities.

Current care pathway

Depending on the intended population, the technology potentially applies to a range of NICE guidance including:

- The section on diagnosis and staging in NICE's guideline on lung cancer advises to offer people with known or suspected lung cancer a contrast-enhanced chest CT scan to further the diagnosis and stage the disease. It states to include the liver, adrenals and lower neck in the scan.

- The section on diagnosis in NICE's guideline on metastatic malignant disease of unknown primary origin in adults states to offer a CT scan of the chest, abdomen and pelvis.

- NICE's guideline on venous thromboembolic diseases advises to offer patients an immediate CT pulmonary angiogram if pulmonary embolism is suspected along with a likely 2-level Wells score.

- NICE's guideline on diagnosing chronic obstructive pulmonary disease (X-ray images or CT scans).

- NICE's guideline on idiopathic pulmonary fibrosis in adults (CT and X-ray).

- NICE's guideline on tuberculosis recommends that local hospitals, Clinical Commissioning Groups and local multidisciplinary teams should consider developing a pathway for people with imaging highly suspected of tuberculosis. X-ray imaging or CT may be used in children and young people with suspected pulmonary tuberculosis, or in a number of groups with suspected extrapulmonary tuberculosis.

- Also see the NICE topic page on respiratory conditions.

Population, setting and intended user

Radiologists (and also radiology specialist registrars and reporting radiographers) in secondary care review and interpret images from people referred for radiological imaging because of suspected abnormalities in the chest. This allows them to make diagnoses and inform planning of patient management. AI for chest CT is intended to support this process by providing an additional source of automatic analysis.

Training may be needed for radiologists to learn how to use the software and the reports it produces.

Costs

Technology costs

- Veye Chest (Aidence), price per output ranges from £5 to £7.50, depending on features selected and volumes of scans. A yearly fee of between £4,000 and £9,000 for cloud server hosting, monitoring and support. IT integration, training and deployment is a one-off cost of £8,500.

- icolung (icometrix) is offered pro bono as part of the icovid.ai initiative. The costs associated with consumables, maintenance or training are included free of charge during the endemic phase of the COVID pandemic, and at least for 6 months. Costs may apply after this phase.

- Veolity (MeVis) offers a one-off perpetual software licence for £44,000 with no per scan costs. This includes first year maintenance (yearly maintenance cost is then £8,800). Initial installation, testing and training costs are £9,000. Yearly support costs are £6,000.

Costs of standard care

Chest CT images are interpreted by radiologists as standard practice in the NHS. The 2019/20 national tariff for a CT scan of 1 area is £69 for people aged over 18 and £73 for people aged between 6 and 18. The cost of reporting is £20 for all ages.

Resource consequences

No published evidence on the resource consequences of AI for chest CT was found.

The costs of adopting chest CT reporting services vary among the technologies included in this review. The cost associated with the technologies generally consists of integration costs, fixed cost per scan processed and yearly maintenance costs. Chest CT is used in the NHS to help diagnosis in a number of clinical pathways. The unit cost of the technologies depends on the total CT throughput.

Minimal changes are needed in facilities or infrastructure as long as software companies comply with NHS communication standards and there is a robust IT infrastructure in the implementing organisation. Existing IT infrastructure and software may vary across NHS

organisations, and unforeseen issues may arise because of implementing novel software.

Regulatory information

Veye Chest (Aidence) is CE marked as class IIb under the new medical device regulations (October 2020).

icolung (icometrix) is a CE-marked class I medical device (May 2020).

Veolity (MeVis) is a CE-marked class IIa medical device (May 2014).

The following manufacturer field safety notices or medical device alerts for this technology have been identified.

Veye Chest (Aidence): it was noted that there was a bug in the software that was introduced with the release of Veye Chest 2.6.0. Aidence released a fix to address the bug. Medicines and Healthcare products Regulatory Agency reference: 2019/011/019/291/017.

Equality considerations

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others.

People aged 65 or over are more likely to have cancer. Age is a protected characteristic under the Equality Act 2010.

Clinical and technical evidence

A literature search was carried out for this briefing in accordance with NICE's interim process and methods statement. This briefing includes the most relevant or best available published evidence relating to the clinical effectiveness of the technology. Further information about how the evidence for this briefing was selected is available on request by contacting mibs@nice.org.uk.

Published evidence

This briefing summarises 1 study published as a preprint (icolung, icometrix) and 1 UK study into Veye Chest (Aidence) published as 2 conference posters.

The icolung study included automatic segmentation of lung lesions in the chest CT scans of 17 confirmed or suspected people with COVID-19 from Europe, South America and a public dataset. The Veye Chest study included 349 chest CT examinations in 324 people and assessed accuracy of lung nodule segmentation and growth tracking. Three ongoing studies into Veolity (MeVis) are described later in the briefing.

The clinical evidence and its strengths and limitations are summarised in the overall assessment of the evidence.

Overall assessment of the evidence

Two retrospective studies into artificial intelligence (AI) for chest CT are summarised. Neither was published in full. One relatively small study of 17 people with suspected COVID-19 compared the performance of AI algorithms in icolung with chest radiologists to identify abnormal scans and potentially significant lesions in COVID-19. The study included a number of non-UK datasets, therefore the generalisability of the findings to the NHS is unclear. Another study into Veye Chest compared the performance of AI software with chest radiologists for segmentation and growth assessment of lung nodules in adults (aged 50 to 74, who currently smoke or have a smoking history, or are reported to have radiological evidence of pulmonary emphysema). This study was published as 2 conference poster presentations, which limits the amount and quality of information available. The study was UK-based, which may help generalisability to the NHS.

There is limited evidence into AI for chest CT. More studies of the impact on clinical management and outcomes would help provide evidence to support clinical adoption. Ideally, studies would be UK-based and prospective in design.

Tilborghs et al. (2020)

Study size, design and location

A retrospective study into 17 people with suspected COVID-19. Images were from

European and South American centres and a public dataset.

Intervention and comparator

The segmentation accuracy of 12 algorithms (incorporated into icolung) was compared with radiologists. The segmentation accuracy of the AI algorithm was calculated as the dice coefficient (statistical method used to judge the similarity of 2 samples). COVID-19 status was confirmed by laboratory testing.

Key outcomes

Dice scores for lung segmentation, binary lesion segmentation and multiclass lesion segmentation were 0.982, 0.724 and 0.469, respectively. The AI algorithm performed binary lesion segmentation (identifying abnormalities) with an average volume error that was better than visual assessment by human readers. The algorithm performed least accurately in multiclass lesion segmentation (including identifying consolidation and ground glass opacity, which the authors note are important lesions in COVID-19).

Strengths and limitations

The study is very small (n=17) and published as a preprint (rather than as a formally peer reviewed paper). There is no information on the size or composition of the training datasets. A multicentre dataset was used and it is unclear how this generalises to an NHS setting. The authors note that a version of the software in this study is available as icolung in the US and Europe. It is unclear how the software in this study differs from icolung.

Murchison et al. (2019a and 2019b)

Study size, design and location

A retrospective study of 337 chest CT scans from 314 people (173 women, 164 men) with a total of 470 pulmonary nodules included (Murchison et al. 2019a and Murchison et al. 2019b). Images were from 1 UK regional healthcare database. Inclusion criteria were people aged between 50 and 74, who currently smoke or have a history of smoking, or are reported to have radiological evidence of pulmonary emphysema.

Intervention and comparator

The segmentation accuracy of Veye Chest was compared with 3 experienced chest radiologists. The segmentation accuracy of readers was calculated as the dice coefficient between each radiologist's segmentation and the segmentations of the others and subsequently averaged (inter-reader dice coefficient).

When looking at nodules visible on sequential scans, nodule registration from the AI was scored as either a true positive-pair if the detected registration was included in the nodule registration reference standard, or as a false positive-pair. The mean discrepancy between growth percentages determined by radiologists and AI alone was calculated.

Key outcomes

The software was able to successfully segment 95% of the total 428 nodules between 3 mm and 30 mm. The performance of the AI software for segmenting pulmonary nodules on chest CT was comparable with that of experienced thoracic radiologists.

The mean growth percentage of lung nodule pairs was similar between readers and by standalone AI.

Strengths and limitations

This study was done in a UK setting. The Fleischner Society's definition for pulmonary nodules was broadly used during this study. Training data were from people aged 50 to 74 in a registry of people who smoke. It is unclear how these data will apply to other patient populations.

The study is presented as a conference poster presentation. The study population is relatively small.

Recent and ongoing studies

Five recent and ongoing studies involving Veolity (MeVis) were identified in the development of this briefing. These included the following 3 studies that are registered with a clinical trials database:

- International Lung Screen Trial. ClinicalTrials.gov identifier: NCT02871856. Status: active, not recruiting. Indication: people who may be at increased risk of lung cancer because of age and smoking history. Study completion date: December 2023.

- Yorkshire Lung Screening Trial. ISRCTN42704678. Status: Enrolling by invitation. Indication: people who may be at increased risk of lung cancer because of age and smoking history. Study completion date: July 2024.

- The SUMMIT Study: a cancer screening study. ClinicalTrials.gov identifier: NCT03934866. Status: enrolling by invitation. Indication: people who may be at increased risk of lung cancer because of smoking history. Study completion date: August 2030.

Expert comments

Comments on this technology were invited from clinical experts working in the field and relevant patient organisations. The comments received are individual opinions and do not represent NICE's view.

All experts were familiar with artificial intelligence (AI) for imaging, and 3 had used AI technologies for imaging before. Two experts were currently using AI technology – 1 to assess bone age in children and 1 in lung cancer screening. None had been involved in research and development of AI technologies. None of the experts believed that clinical AI software was in routine use in the NHS (except in some centres running the national lung cancer screening programme). The technology is relatively new and having ongoing validation and development.

Level of innovation

Three experts stated that AI technology for chest imaging was a novel concept in the NHS. One expert noted that the technology could introduce a paradigm shift in UK radiology practice. Two experts noted that the technology could improve patient safety by reducing the risk of abnormalities being missed.

Two experts were aware of other competing technologies and highlighted variability in the field. An expert highlighted that many CT scanner vendors provide similar AI technologies embedded on their software package. One expert noted that comparing the different technologies was challenging as the concept of 'deep learning' is broad and a detailed

description of technologies is needed for a comprehensive comparison. Performance of the technologies may also vary, with the results of nodule detection and measurement differing significantly across different AI software. Another expert highlighted that there is a broad range of AI technologies emerging in all radiological disciplines for both diagnostic and therapeutic purposes.

Potential patient impact

All 4 experts noted that the technology could improve diagnostic accuracy in image interpretation. One expert felt that AI should be mandatory for any CT imaging of the thorax to minimise the risk of missing early-stage lung cancer or lung metastases. For example, implementing automated lung nodule detection tools before a radiologist interprets the scans may reduce human error and so significantly reduce the number of delayed diagnoses of lung cancer or metastases. One expert suggested that using AI software may increase the abnormality detection rate, potentially through increased sensitivity, but at the cost of reduced specificity. Three experts suggested that the technology may help improve triaging or speed up time to diagnosis. However, 2 experts explained there may be limitations to this. One noted that from experience the time spent by radiologists to report each case may increase, because results from the AI analysis need to be taken into account in addition to the time taken to do routine reporting. Another expert cited evidence from a study showing that algorithms designed to help prioritise scans based on perceived abnormality may produce a trade-off, with other people waiting longer for their results (Annarumma et al. 2019). One expert noted that AI technology may decrease interobserver and intraobserver variability in diagnostic work. All experts felt that the technology has potential to change the current pathway or clinical outcomes by improving interpretation accuracy and reporting times. Earlier diagnosis may result in less invasive treatment or allow curative treatment (which may not be possible with a late diagnosis). However, 1 expert noted that identification of clinically insignificant disease may unnecessarily increase patient anxiety and demand on the CT scanning service. One expert suggested that although tuberculosis assessment AI is unlikely to add significant value to UK radiology practice, assessment of other pathologies such as pneumothorax, pleural effusion, and nodules would useful.

Potential system impact

Experts thought that improved accuracy and reporting speed of interpreting radiographs and radiology reporting would be the main system benefits of the technology. One expert

noted that productivity may improve in settings such as lung cancer screening if the technology could provide an initial reading, avoiding the need for 2 reading radiologists. However, there is no evidence for this. Another expert advised that over-diagnosis was a risk with adopting AI for CT, estimating that around 70% of people referred for a chest CT present with a lung nodule, and only a very small minority of these nodules have clinical implication. One expert highlighted that radiologists and radiographers can feel guilt from missed diagnoses, and potential improvements in accuracy may help reduce these effects.

Expert opinions varied about the cost impact of AI technologies compared with standard care. One noted that installation costs would be significant and included software hosting and support, integration with existing software, and security. Another highlighted the need for additional specialised staff, and that there may be an increase in the time radiologists and radiographers spend on each case reported. Two experts thought that costs would increase (for example because of increased need for scanners, radiography staff, radiologist, lung nurses and chest clinicians), but that these would be offset by savings in efficiency. Another expert stated that using AI technology could result in cost savings if the cost is offset by earlier disease diagnosis. One noted that improved accuracy would also help offset costs. One expert did not anticipate significant resource gains from these technologies. Finally, 1 expert noted that there were diverse opinions about whether the technology would enhance the function of radiology departments or potentially replace radiologists.

Experts noted the need for a robust IT structure before implementing AI software. At installation, software hosting and bandwidth may need capital investment, and the AI software would need to be effectively integrated into clinical systems (usually picture archiving and communication system [PACS] or electronic health record). Specific training may be needed to understand limitations and scope of the technology function. One expert suggested that training needs may be fairly modest, and depend on product design and radiologist or radiographer experience levels.

Experts highlighted a number of potential safety and regulatory issues. A fundamental safety issue is the validity of training data and its application to the local population. AI software is based on machine learning of large datasets and will therefore be influenced by the characteristics in those populations. One expert described an example of algorithm training from people in China being unlikely to be fully applicable to a typical NHS population for assessing tuberculosis (because prevalence differs between these populations). One expert suggested that there may be data security issues if software companies want to use NHS patient data to develop their products. Another expert

highlighted that there is significant variability of methods used to quantify lung nodules which needs to be accounted for in assessment and guidelines involving AI for volumetric quantification of lung nodules.

One expert described 2 potential cognitive biases that may affect interpretation of AI output. Firstly, in 'automation bias' clinical staff may be overconfident in the results of an automated system, which may bias clinical opinion negatively if the AI output is not accurate. Secondly, the expert noted the 'satisfaction of search' bias when a reader may carry out an incomplete assessment. AI technology may promote this error if the reader incorrectly believes the image has been analysed. In relation to this, another expert added that de-skilling of radiologists (and reporting radiographers) may be a factor after AI implementation.

Two experts mentioned discussions around regulation and legal responsibilities. One noted that the regulatory environment was complex (including the Medicines and Healthcare products Regulatory Agency and Care Quality Commission oversight) and that the current framework may not be adequate to keep up with the sector.

One expert noted that the evidence about AI technology is still limited. Further research into multicentre cohorts would be needed to understand risks associated with implementation in clinical practice.

General comments

One expert provided advice based on their own experiences with AI software. Firstly, when automated analysis is available, the technology may be met with considerable scepticism from some staff, who then request older 'manual' methods. Clinical engagement is vital in developing clinical AI technologies, and this must include training in the software scope and limitations for all clinical staff. Secondly, automated assessment may reveal areas of inadequate clinical practice, for example when a clinician has used an incorrect method for image interpretation. The software may incorrectly be labelled inaccurate. Thirdly, implementation of clinical AI software, will need significant involvement of NHS IT staff, for hosting software, firewall configuration, software integration and troubleshooting. Storage of any metadata associated with the AI image analysis will need to be considered. Ideally these data should be stored long term in NHS systems. Metadata may be lost if stored in vendor's software storage or potentially at the end of contract when transferring from one vendor to another.

Another expert noted that thoracic imaging scans can show numerous changes within the lung, mediastinum, chest wall or visualised portions of the neck and abdomen. They suggested the AI technologies may not be properly designed or tested for analysis of the chest wall, neck and upper abdomen.

Other considerations

All experts thought a large number of people would be eligible for chest AI technology in the NHS. One expert noted that over 5 million CT scans are done every year in the UK, and a large proportion of those would cover the thorax. AI software could be applied to all acute chest CT. For example, nodule volumetry could apply to all cases of lung cancer and lung metastatic disease.

All experts felt that, at least in the short to medium term, this technology would be an addition to current standard care. One expert suggested that in 7 to 10 years, this technology may begin to replace current standard care.

In terms of the practical aspects of the technology, 1 expert indicated that training may be needed to understand and use the technology appropriately. Another highlighted that outsourcing scans to be analysed by AI software may be of concern. One expert noted that the AI technology would need to be seamlessly integrated into software so that the AI interpretation appears in the same report box as the standard X-ray or CT report. The expert noted that technology that needed 'extra mouse clicks' may not be used.

All experts suggested a number of potential barriers to this technology being adopted. Factors included clinical governance and costs, difficulty integrating the technology into clinical pathways, lack of IT capacity for integration (and potential issues of incompatibility between the AI technology and existing IT infrastructure). Clinical acceptance was mentioned as a significant barrier, but one expert noted that this may change over time with wider acceptance of AI technologies in different fields. Another suggested that the lack of robust clinical evidence on the broader impact of the technologies also affects the translation from research to clinical practice.

Two experts were not aware of further evidence for the technology included in this briefing. Other experts noted that there were various developing research projects for AI in thoracic imaging, including tools for quantitative analysis of diffuse interstitial lung, cardiovascular and pleural diseases.

All experts proposed potential research studies to address uncertainties in the evidence base. This included post-marketing studies of effectiveness to independently show the use of these technologies for NHS populations. One expert noted that the evidence base for the technologies in this briefing would benefit from more publications independent from the software companies that had formal peer review. One expert specified that in vivo research would be helpful into the diagnostic accuracy and variability of measurement methods. They suggested further assessments would benefit from being randomised, prospective, blinded, UK-based, controlled and well powered.

All experts felt that NICE guidance into chest AI technologies would be very useful or crucial. One expert suggested the techniques, limitations and scope of these technologies is poorly understood. Another expressed concerns about uncontrolled introduction of AI software. One noted that it may be too early in the life cycle of the technologies to produce a fully informed assessment.

Expert commentators

The following clinicians contributed to this briefing:

- Dr Jim Carmichael, consultant radiologist, Guy's and St Thomas' NHS Foundation Trust. Dr Carmichael has received private income from HCA international, unrelated to any AI technology.

- Dr Nicholas Hollings, consultant radiologist, Royal Cornwall Hospital. Expertise in chest and cardiothoracic radiology. Dr Hollings is the director of T2 Star Ltd, a publishing and radiology consultancy business. He has received a small honorarium for market research into the CT investigation of interstitial lung disease.

- Dr Klaus Irion, consultant chest radiologist, Manchester Royal Infirmary. Dr Irion has approved a research grant for an AI tool for diagnosis of pleural malignancy and is in discussions with Warwick University about joint research on classification of chest radiographs as normal or abnormal.

- Dr John Reynolds, consultant radiologist with expertise in cardiothoracic radiology, Birmingham Heartlands Hospital. Declared no conflicts of interest.

Development of this briefing

This briefing was developed for NICE by the King's Technology Evaluation Centre (KiTEC). NICE's interim process and methods statement sets out the process NICE uses to select topics, and how the briefings are developed, quality-assured and approved for publication.

ISBN: 978-1-4731-3951-0