Strahlenschutzkommission
Geschäftsstelle der
Strahlenschutzkommission
Postfach 12 06 29
D-53048 Bonn
http://www.ssk.de

Application of artificial intelligence in image reconstruction and processing in radiology and nuclear medicine

Recommendation by the German Commission on Radiological Protection

Adopted at the 339th meeting of the German Commission on Radiological Protection on 27/28 October 2025

The German original of this English translation was published in 2025 by the Federal Ministry for the Environment, Climate Action, Nature Conservation and Nuclear Safety under the title:

Anwendung von künstlicher Intelligenz bei der Bildrekonstruktion und -verarbeitung in Radiologie und Nuklearmedizin

Empfehlung der Strahlenschutzkommission

This translation is for informational purposes only, and is not a substitute for the official statement. The original version of the statement, published on www.ssk.de, is the only definitive and official version.

Preface

Artificial intelligence (AI) is concerned with the development of machines designed to perform tasks that previously required human intelligence. An important sub-area of AI is machine learning (ML), which enables machines to recognise patterns in data sets. Deep learning (DL) refers to machine learning based on large amounts of data.

In medicine, more and more artificial intelligence methods are being developed in various fields of application and some of them are already being used. The use of AI-based methods also plays an important role in medical imaging.

In addition to support in image evaluation and diagnosis, many approaches are pursued to perform image reconstruction of three-dimensional medical image data sets and/or image processing for better visualisation, e.g. by noise reduction using AI-based methods. This is also done for image data sets whose initial data were obtained using ionising radiation.

Optimisation of the medical application of ionising radiation is required in terms of radiation protection. Therefore, AI-based approaches are used to obtain better diagnostic information from the images at the same patient exposure, or to reduce patient exposure while maintaining the same diagnostic significance of X-ray imaging or nuclear medical imaging.

Thus, if possible in the near future, it is necessary to exploit the possibilities offered by AI-based methods, in particular those using machine learning. On the other hand, it is important to assess the extent to which potential problems associated with the use of machine learning are relevant. This applies in particular to radiation protection aspects. For example, when using AI-assisted methods for reconstruction or noise reduction, there is a possibility that information relevant to the diagnosis may be lost or distorted, or that false information may be added. This may not be recognised in the images to be diagnosed and, in many cases, cannot even be reproduced on the basis of the stored data. In this case, ionising radiation would then be applied to humans without fulfilling the medical justification. This must be assessed in the context of radiation protection law (in particular § 83 StrlSchG (StrlSchG 2017)).

Against this background, the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV), in a letter dated 7 October 2020, asked the German Commission on Radiological Protection (SSK) to consider the use of artificial intelligence (AI) methods in image reconstruction in radiology and nuclear medicine, with regard to the ionising radiation used, from the point of view of radiation protection.

The present advisory opinion answers the questions posed in the advisory mandate with a particular focus on computed tomography, imaging during interventions, positron emission tomography (PET), single photon emission computed tomography (SPECT), cone beam computed tomography and digital tomosynthesis. The automated evaluation of image data is not part of the consultation mandate.

To prepare this recommendation, the SSK established a working group in 2021, consisting of the following members:

Prof. Dr. Christoph Hoeschen (University of Magdeburg), Chairperson of the working group

Prof. Dr. Stefan Delorme (German Cancer Research Center)

Prof. Dr. Jörg van den Hoff (Technical University of Dresden)

Prof. Dr. Marc Kachelrieß (German Cancer Research Center)

Prof. Dr. Andreas Maier (University of Erlangen)

Prof. Dr. Georg Schramm (University of Leuven, Belgium)

Dr. Sarina Thomas (University of Oslo, Norway)

Prof. Dr. Michael Uder (University Hospital Erlangen)

The SSK would like to thank the BfS for providing comprehensive technical information and support.

Bonn, December 2025

Prof. Dr. Christoph Hoeschen	*Prof. Dr. Ursula Nestle*
Chairperson of the working group "AI in Medicine"	Chairperson of the German Commission on Radiological Protection

CONTENTS

1 Introduction

1.1 Introduction

Artificial intelligence (AI) is a field of computer science that deals with the development of machines capable of performing tasks that previously required human intelligence. An important sub-area of AI is machine learning (ML), which enables machines to recognise patterns in data sets. Deep learning (DL) refers to machine learning based on large amounts of data, i.e. DL is a part of machine learning, which in turn is a part of AI. This corresponds to the definition of the European Commission in Article 3 (1) of the EU Artificial Intelligence Act (EU 2024). There, an "AI system" is defined as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".

The basic process for creating an AI-based procedure is as follows:

First, a problem to be solved by the AI-based procedure must be defined. A so-called model is developed or selected. The models are based on neural networks, i.e. they imitate a network of neurons at different levels that are interconnected. The chosen model describes the structure of data processing. Depending on their complexity, these models have a varying number of parameters, i.e. internal variables that enable the model to be adapted to the given task. These parameters are learned. To learn the parameters, the model is trained, e.g. by comparing a desired result with an input signal or by recognising certain patterns in large amounts of data. The result of the training is then validated. At the end of the creation of the AI-based procedure, the model is tested with the parameters resulting from the training and validation.

In medicine, more and more artificial intelligence methods are being developed in various fields of application and some of them are already being used. The use of AI-based procedures also plays an important role in medical imaging. Medical imaging methods comprise a number of different techniques: microscopic imaging mainly for pathological, ophthalmological or dermatological questions, ultrasound imaging, projection imaging using X-rays or creation of scintigrams with the help of a planar gamma camera, tomographic methods such as X-ray computed tomography (CT), single-photon emission computed tomography (SPECT) and positron emission tomography (PET), magnetic resonance imaging (MRI) or optical coherence tomography, etc. In addition to supporting image evaluation and evaluation, many approaches are pursued to perform image reconstruction of three-dimensional medical image data sets generated using ionising radiation and/or image processing for better visualisation, e.g. by means of noise reduction using AI-based methods.

Optimising the medical application of ionising radiation is essential in terms of radiation protection. Therefore, AI-based approaches are used to obtain more recognizable diagnostic information from the images at the same patient exposure, or to reduce patient exposure while maintaining the same diagnostic significance of X-ray imaging or nuclear medical imaging. Thus, if possible in the near future, the use of AI-based methods, especially those using machine learning, should be pursued. On the other hand, it is necessary to assess the relevance of possible problems associated with the use of machine learning. This applies in particular in the sense of radiation protection. For example, when using AI-assisted methods for reconstruction or noise reduction, there is a possibility that information relevant to the diagnosis may be lost or distorted, or that false information may be added. This may not be recognised in the images to be diagnosed and, in many cases, cannot even be reproduced using the stored data. In this case, ionising radiation would then be applied to humans without fulfilling the medically justifiable purpose. This must be assessed in the context of radiation protection law (in particular § 83 StrlSchG (StrlSchG 2017)).

1.2 Explanation of terms

Based on the general definition of artificial intelligence, further terms necessary for understanding the text are explained below:

In algorithmic modelling in the context of machine learning, a basic distinction is made between white-box, grey-box and black-box models. The choice of method depends on the type of learning problem, i.e. the task to be solved, and the available data:

- White-box models are not learning-based methods, as the structure of the model and the associated parameters are known, for example from physics. These methods are often used in imaging because they are transparent and make decision-making comprehensible. Essentially, all classical methods of tomographic image reconstruction, from filtered back projection to iterative reconstruction, can be classified as white-box models.
- Black-box models, on the other hand, are methods in which the internal structure of the model and the associated parameters are not known or not sufficiently known. Here, the models are trained using suitable training data. These models are used in imaging, including in the medical field, provided that they have delivered powerful results in other areas such as pattern recognition or data processing and can be easily applied to imaging. Examples of black-box models are U-Net reconstructions and AUTOMAP (see section 3.1.2).
- Grey-box models are a mixture of white-box and black-box models. In these methods, the structure of the model is known, but not all parameters. Grey-box models can benefit from prior knowledge to simplify the learning problem, improve the outcome, and reduce the number of parameters required. An example of a grey-box model is the so-called Known Operator Learning, in which part of the model is already known and the model only needs to learn the remaining parameters. Methods such as variational networks, FBP-Net, or other hybrid methods (Maier et al. 2022a) thus allow classical theory to be combined with machine learning.

In this recommendation, the focus of the considerations regarding AI is on black-box and grey-box models, as these approaches have achieved remarkable results in the literature and are currently being discussed in particular. In addition to the black-box and grey-box models, white-box models are also important AI methods. However, they should be considered separately, as the classical analysis and quality assurance procedures can essentially be adopted for this purpose. This is not possible with black-box models, which is why they are the focus of this recommendation. Many black-box models are based on so-called convolutional neural networks (CNNs), which are trained using input data. These convolution-based networks involve numerous non-linear filter operations that are applied to the image data or their intermediate results. In contrast to so-called fully connected networks, in which each voxel of an intermediate result is linked to each voxel of the following intermediate result, only closely neighbouring voxels are linked in CNNs. This means that a) the neighbourhood relationships of the image content are preserved and b) the number of unknowns to be determined by the training is several orders of magnitude smaller.

Like a natural neural network, an artificial neural network consists of many processing layers. Each layer processes the data non-linearly and passes it on to the next layer. After typically a few dozen to hundreds of layers, the neural network outputs the processed data, image, or image stack. Specifically, in image processing neural networks, a convolution[1] is usually performed in the layers, i.e. the image is filtered with a 3×3 convolution kernel (or several) and the result of the convolution operation for each pixel is sent through a non-linear scalar function (e.g. by setting negative values to zero). Each convolution kernel has, for example, 9 unknown coefficients for two-dimensional kernels and 27 for three-dimensional kernels, per dimension of a so-called feature. Due to the numerous convolution kernels per layer and the numerous layers, the number of coefficients to be determined during training typically amounts to several million to several billion. The coefficients are determined during training, which requires the large amounts of training data.

Description/clarification of different data classes

The basic process of developing an AI-based method requires various data sets, also known as data classes:

– Training data describe the input data that are used to optimise the neural network parameters by means of learning.
– Validation data describe the data used during development to assess the quality of the network. If the network is not yet sufficiently good, its parameters (network layout, number of training epochs, learning rate, etc.) are changed and retrained with the training data. Thus, the validation data are indirectly included in the training process, namely to optimise higher-level model parameters.
– The test data are an independent data set used to test the developed AI method after validation. They shall be as representative as possible of the relevant population in the clinical application of the method and are either acquired externally or taken from the total data used for development.

Once the development has been completed, the network and its parameters finalised and evaluated, the AI-based method should be evaluated on another (second) test data set, also known as the verification data set. This verification data set should be treated completely separately from the development and should cover the overall variability of the input data of the final application, i.e. multicentric (obtained from different centres) and representative. In the case of medical data, particular attention should be paid to influencing factors such as gender, weight, origin, pathology and age, provided that the manufacturer or developer of the procedure has access to these data.

[1] The convolution of two functions is a mathematical operation in which the value of the first function f is replaced by a function value that is obtained as the integral of the value of the first function multiplied by a shifted and mirrored second function g over all shifts. This means that the superposition of the functions is considered and f is essentially "smeared" by g. The mathematical formulation for two functions f and g is $(f * g)(x) = \int_{-\infty}^{+\infty} f(x') \cdot g(x - x') \mathrm{d}x'$. Due to the way it works, convolution can be used very effectively to represent filter functions, for example. Convolution can also be performed in multiple dimensions and used as a discrete operation, as is the case with the aforementioned CNNs, for example.

1.3 Situation regarding the use of AI in medical imaging

The use of electronic data processing is particularly important in all areas of medicine, which have always dealt with large amounts of data. This is especially true when doctors are required to pay particular attention and there is a risk that important findings may be overlooked. This therefore affects all disciplines dealing with images, be it radiological or endoscopic images or photographs, e.g. of the skin or microscopic tissue sections, which can easily be converted into digital format.

For example, the examination of microscopic tissue sections requires a high degree of attention, e.g. when examining a biopsy of a lesion suspected of being a tumour, the detection of atypical cells in blood or lymphatic vessels is evidence of the malignancy and invasiveness of the process. Such invasive foci can be very small, so the requirement for attention during the examination is high.

The situation in radiology is similar. For example, the exclusion of pulmonary nodules in a computed tomography scan of the thorax requires high attention, as all the vessels (pulmonary arteries and veins) cut in the individual sectional image usually have an oval shape, and their elongated shape only becomes apparent when "scrolling" through the image stack. Examiners find the task both strenuous and tedious, so that the risk of overlooked findings should not be underestimated. The situation is similar in whole-body examinations when screening bones for circumscribed abnormalities.

Whole-body MRI examinations are somewhat different, as they contain an almost unmanageable wealth of information and are prone to artefacts. In addition to the sheer task of coping with the volume of information, the examiner must also use their intellectual abilities to distinguish genuine findings from artefacts.

A number of digital tools are already in use: The simplest of these merely change the image presentation to make the examiner's task easier. This includes, for example, the thick-slab MIP (MIP: maximum intensity projection) used when searching for nodules in lung CT images. In this process, MIPs are generated from approximately 1 cm thick stacks of thin-slice CT images of the lungs (with a slice thickness of 1 mm). In these MIPs, vascular sections are elongated while nodules still appear round – without any loss of resolution, which makes it much easier to find even small nodules. Other algorithms use elastic deformation to convert the representation of the ribs into a top view, in which the ribs are stretched and arranged in parallel, which makes it much easier and faster to find osteolytic or osteosclerotic lesions.

More advanced image processing approaches involve automatically detecting pathological findings, e.g. nodules in the lungs. Some of these solutions are already integrated into the scanner software, while others can be purchased as an option or as separate stand-alone software. Many of these solutions also include tools for segmenting the lesions found, e.g. for volume measurements as required to assess the therapeutic response of tumours based on specified criteria such as ‘RECIST’ (Response Evaluation Criteria In Solid Tumors, (Therasse et al. 2006)). All of these solutions require a high level of interaction with the person carrying out the evaluation, who must either identify and annotate the findings themselves after the images have been processed accordingly, or correct or reject an annotation suggested by the software. The required expertise of evaluating the images is correspondingly high. Most of these approaches are based on algorithmic approaches without the use of deep learning; initial approaches to the use of deep learning-based methods are being pursued.

Approaches are being scientifically tested with the help of artificial intelligence and so-called radiomics methods are used to derive correlations with the diagnosis of lesion type (e.g. benign or malignant lesion) from local image properties of findings (such as signal intensity, texture or contour) (Lambin et al. 2012). Correlations with biological properties of findings (e.g. probability of treatment response) or general disease prognosis are also investigated. Without exception, these methods are currently undergoing scientific testing and are not available as medical devices or parts thereof. The user is also free to assess, modify or reject the results of such processes, as the entire image information is available to them at any time.

These approaches to supporting diagnosis and evaluating given image data for clinical or scientific purposes must be distinguished from changes to the image information made by the AI method itself. The latter may affect whether the original information in the processed images is still preserved in the same way. In addition, if the original data are severely affected by noise or other artefacts, the person making the diagnosis has no way of verifying the accuracy of the images generated. Such methods are currently under development and, depending on the application, are either closer to clinical trials or further away. More detailed descriptions can be found in the scientific rationale.

The use of AI methods in the processing of data from imaging examinations – both in the image and raw data space – described in the last paragraph can be utilised in a variety of ways: For example, AI methods can aim to reduce radiation exposure without compromising image quality (prioritisation of radiation reduction) or to generate better images at a given radiation exposure (prioritisation of image quality). An AI-based method, which in the case of image processing is typically realised by a neural network, learns what to do from training data in preparation for the actual processing of data from a specific examination. In the examples mentioned, where more noisy images need to be converted into less noisy images, this can be done in the simplest case by presenting a neural network with a large number of data pairs of highly noisy and low-noise patient images (or image stacks). The highly noisy images or image stacks can be generated, for example, by adding realistic noise to the less noisy images or image stacks. The neural network is thus trained to denoise the more noisy images so that they become as similar as possible to the less noisy images.

An AI method can only make an estimate to supplement the missing information. This information then comes from the distribution of training data from a larger number of other examinations and not from the patient being examined. This often involves paired training data (supervised learning, (Mohri et al. 2012)), so that, for example, in the case of noise reduction, there is one CT volume with a lot of noise and one corresponding CT volume with little noise. This means that there is a possibility that information displayed in the image as a result of an attempt to improve image quality may not originate from the current measurement. Here, too, AI methods have only been able to learn the probability distribution of the training data and can therefore only transfer the information contained therein to the image of the actual patient.

Probably the most visible and typical field of application of AI in imaging, which in some cases is already at an advanced stage of clinical testing, is the processing of images with the aim of reducing noise while maintaining image sharpness. This corresponds to an increase in the contrast or signal-to-noise ratio. Such a method can also achieve a reduction in dose if the examination protocols are adapted accordingly, e.g. by reducing the exposure time or the tube current on the CT. These AI noise reduction methods are not only used in computed tomography, but also in projection imaging, such as X-ray radiography or fluoroscopy, and in nuclear medicine procedures.

Another field of application of AI methods in imaging with ionising radiation is the improvement of image quality in X-ray images, including fluoroscopy and mammography. In contrast to tomographic images, in which the grey values displayed are also quantitatively significant and therefore should not be changed by the AI methods if possible, in projection images it is usually only the qualitative image impression that is important. The grey value of a pixel in the projection image is a superimposition of the attenuation values of structures lying one behind the other in the beam path, and the grey value of the pixel does not allow any quantitative conclusions to be drawn about the conditions in the individual structures. Non-AI-based changes to quantitative physical information, such as non-linear contrast enhancements or contrast reductions, including those that depend on the image content and those that are irreversible and/or reduce information, have been used successfully for a long time and are a clinical standard, provided that the information relevant to the diagnosis is demonstrably preserved. Simple concrete examples of this, whose mode of operation is well understood and comprehensible, are gamma curves or bias filters. There is no obligation under radiation protection legislation to archive the unaltered raw data, and therefore they are not typically stored. However, the image data used for reporting should always be archived without any loss of quality, as recommended by the SSK in 2016 (SSK 2016). With the help of AI methods, it is now possible to achieve significantly more sophisticated effects, such as selectively highlighting or suppressing bone structures in X-ray images, identifying vessels (virtual DSA) (Duan et al. 2024) or compensating for patient motion in fluoroscopic images. However, the functioning of AI methods is not easy to comprehend, let alone understand, especially when large amounts of data are involved. This applies, for example, to foundation models[2], which influence results based on large image and text databases that link patient findings with radiological images. In any case, care must be taken to ensure that the use of AI does not remove any relevant information from the X-ray images. This applies to all medical imaging methods, and, due to radiation protection regulations, particularly to those that use ionising radiation.

AI methods can be applied not only to images but also to raw data in order to reduce artefacts or compensate for inadequacies in imaging systems (Magonov et al. 2024). These processes may not be visible to users and run without their knowledge.

1.4 Advisory Mandate

The Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (now the Federal Ministry for the Environment, Climate Action, Nature Conservation and Nuclear Safety, BMUKN) formulated an advisory request dated 7 October 2020, asking the SSK to investigate the problems associated with the use of AI in medicine and to propose ways of solving them. This advisory request focuses on the following questions:

– Which AI applications can be useful for image optimisation (e.g. scatter reduction, dose reduction, artifact reduction), and what are the risks of using AI in medical imaging? Can the limitations and potentially adverse effects of an AI method be made clear to the user, and can criteria be derived for assessing the suitability of the methods used?

– What is the current framework for the use of these methods, and what conditions must be in place in the future to ensure that the use of ionising radiation is safe and that the actual diagnostic features are shown with sufficient reliability?

  When considering the relevant requirements for AI algorithms and their quality assurance, the training data, the requirements for the validation and testing of the algorithms, and the testing of AI-supported image reconstruction methods accompanying clinical application should be taken into account.

---

[2] A foundation model is a large machine learning model that has been pre-trained on a broad database and, thanks to its versatility, serves as the basis for a variety of specific applications.

- What contribution can international standardisation and medical device law for imaging systems provide to the quality assurance of AI methods with regard to approval?
- AI methods can also be used in situations with low noise levels or in high-contrast areas to achieve improvements in image quality. The methods to be used may therefore differ depending on the issue at hand and the imaging situation (low- and high-contrast images). Can requirements for dealing with different available AI methods take this circumstance into account, and how could this be done?
- Can requirements be developed for the methodology of an AI algorithm in image optimisation with regard to the intended field of application, the usable devices and acquisition parameters, the required test data sets and validation and, on the other hand, for the establishment of a test system with which the information-preserving application of procedures can be verified? If this is possible, can these requirements be used to specify what needs to be tested and documented, and how, in order to ensure that the framework conditions are complied with under all circumstances? This also includes how the test procedures themselves can be tested and how this can be used to ensure that the intended range of application of the AI procedures can actually be verified.
- The use of AI-based image reconstruction methods also influences metrological checks, which are carried out, for example, as part of acceptance tests or to determine image quality and dose efficiency. The application of AI methods to phantoms in such tests could lead to distortions in the characterisation of parameters such as image quality and dose efficiency. The requirements for the validation and testing of the algorithms should also take this into account and, if necessary, propose alternative test procedures.

2 Recommendations and statement

For the use of AI-based methods for image reconstruction or image processing in medical imaging methods based on the use of ionising radiation, the SSK recommends that

1. in order to rule out image processing-related misdiagnoses that would call into question the benefit of the examination and thus the justifying indication (see section 4.1.1), when using AI methods for image reconstruction or further processing of patient-related image data, it needs to be ensured that all structures relevant for the diagnosis are retained with at least essentially unchanged sensitivity and specificity in terms of diagnostic reliability compared to non-AI-supported methods and that no artificial structures are generated. Similarly, when using computed tomography image data for treatment planning in radiotherapy, the Hounsfield units[3] must not be significantly altered;
2. in the case of a dose reduction in combination with the use of AI methods to generate the images to be evaluated, all structures relevant to the diagnosis should be identifiable with comparable or higher sensitivity and specificity than would be the case without dose reduction and AI use, in order to satisfy the justifying indication for the examination (see section 4.1.2);
3. in order to demonstrate compliance with the first two recommendations, human-observer studies should be carried out, unless scientifically validated results are already available which confirm the equivalent significance of other (metrological) methods. In the latter case, it must be ensured that the methods used are selected in such a way that they do not favour the AI methods being evaluated and that the absence or creation of subtle structures is sufficiently taken into account (see sections 4.1.2 and 4.2.5).

A prerequisite for the quality of data processing to be trusted is the transparency of the processes (as described in section 7.5). Therefore, the SSK continues to recommend that

4. the entire process of creating and using AI-based methods for image reconstruction or image processing should be quality assured. In particular, generative networks should always produce a unique output data set from an input data set that is not influenced by chance.
5. a competent body should be designated to verify the performance of an algorithm for the specified use cases on independent, sufficiently characterised test data prior to its clinical use (see sections 4.2.5, 5.3, 7.2, 7.3, 7.5). This competent body may either be an existing body into which additional competencies are integrated, or it may be newly created. The points to be examined by the competent body should include the following aspects, in particular for the approval and assessment of AI-based reconstruction and post-processing algorithms in radiology and nuclear medicine:
   – Validity of the metrics and methods used by the developers to assess the performance of the AI algorithm for specific clinical applications.

---

[3] The Hounsfield unit describes the attenuation of X-rays (assumed to be monoenergetic) by tissue in a computer tomography image. The Hounsfield scale was proposed by Godfrey Hounsfield. The attenuation of X-rays by tissue is described by its attenuation coefficient μ, and the respective value is related to the attenuation coefficient of water. The CT value or Hounsfield unit is then calculated as follows:

$$\mathrm{CT-Wert}\ [\mu_{\mathrm{Gewebe}}] = \frac{\mu_{\mathrm{Gewebe}} - \mu_{\mathrm{Wasser}}}{\mu_{\mathrm{Wasser}} - \mu_{\mathrm{Luft}}} * 1000\ \mathrm{HU}$$

- Performance of the AI algorithm to be assessed for the specified application based on the disclosed subset of test data and metrics provided by the developers.
- External and independent assessment of the performance of the AI algorithm under evaluation for the specified application based on independent representative test data that were **not** available to the developers.
- Appropriate involvement of experienced clinical and imaging experts in the assessment of the performance of an AI algorithm in accordance with recommendations 1 and 2, if there is reason to believe that an assessment of performance cannot be guaranteed by mathematical and technical metrics.
- Scope and representativeness of practically relevant input data for evaluating the generalisability of AI algorithms, i.e. for which applications, devices and examination parameters and which patient cohorts an algorithm can be used.
- Existence and plausibility of risk assessment and hazard assessment for potential failure scenarios and misuse of the AI algorithm. This refers to a potentially possible malfunction of the AI method.

Appropriate binding regulations should be put in place.

The SSK recommends the following for the implementation:

6. the competent body should be given access to verification data that are independent of the training data in order to verify the quality of the system used, similar to the phantoms used in conventional methods (see sections 5.3, 7.5).
7. the independent body should receive, for the analysis it is to carry out, a sufficient amount of multicentre and representative data from various devices and manufacturers, including the necessary metadata and annotations, i.e. the markings or classifications of image data. The image data must be suitable for verifying the quality of the algorithms. The data must not be published, to ensure that the verification data are independent and did not influence the training of the algorithms (see sections 4.2.2 and 7.5).
8. in order to assess the data quality with regard to the usability for the development of AI-based methods, independent clinical experts should also be involved in the evaluation process (see sections 4.3, 5.1).
9. Standards should be defined and reviewed (see sections 4.3, 5.1) to ensure the quality of the annotations (see Recommendation 7).

For the development, introduction and clinical implementation of AI-based methods in accordance with recommendations 1 and 2, the SSK recommends:

10. Binding requirements should be imposed on the manufacturers and developers of an AI procedure as a necessary prerequisite for verifying compliance with the first two recommendations:
    - For development, training, validation and test data should be randomly sampled from a common population – if this is not the case, reasons should be given. Manufacturers should specify the use case (see also sections 4.2.5, 5.1, 6 and 7.5).
    - Comprehensive documentation of the training, validation and test data used with regard to the patient cohort, the radiopharmaceutical used, the dose or at least the activity applied, or the exposure to X-rays, as well as the acquisition and reconstruction parameters, should be provided. This should be done on condition

that data protection is guaranteed and provided that the manufacturer has access to the aforementioned examination-relevant data. If data augmentation methods were used, these should also be described (see sections 4.1.4, 4.2.5, 5.1, 6 and 7.5).

- A representative part of the training, validation and test data should be disclosed, at least to a competent body (see sections 4.2.5, 5.1, 6 and 7.5).
- The criteria used to select the training, validation and test data should be explained, along with why this approach ensures that out-of-distribution cases will not occur to a relevant extent (see also sections 4.2.5, 5.1, 6 and 7.5).
- It should be explained whether and, if so, what measures are taken by the manufacturer to identify out-of-distribution cases (see section 7.5).
- Manufacturers should explain at which point in the data pre-processing, image reconstruction or image processing AI algorithms are used and what purpose they serve. A specification of the input and output data should be provided and it should be stated whether generative networks are used and whether pre-trained networks are used. If this is the case, it should also be documented and disclosed what the networks were pre-trained with and how this was done (e.g. supervised, semi-supervised, unsupervised, transfer learning) (see sections 4.2.1, 4.2.5, 5.1 and 7.2).
- Manufacturers should provide a risk assessment of what a malfunction of the AI, e.g. in the case of an out-of-distribution patient, could mean for the patient. It should be explained whether a malfunction leads to clearly visible artefacts and is therefore easily recognisable by users, or whether it may lead to subtle changes in small details that are difficult to detect and could easily be misinterpreted as a missing or newly developed pathology. The manufacturer should explain how it arrived at its assessments and thus present a risk analysis (see sections 5.2, 5.3, 7.3, 7.4).
- Manufacturers shall explain for each AI procedure how it modifies its input data (see section 4.2.1).

11. The manufacturer shall define the field of application of the algorithm based on the verification data. They shall also clearly identify any limitations that could have an adverse effect on the patient, e.g. a significant underrepresentation of certain characteristics such as age groups or, where applicable, ethnic groups in the training data (see section 7.5). If necessary, restrictions on the areas or classes of application, e.g. a restriction to image data from groups of people or for certain diseases, should be specified.

The SSK also recommends the following general aspects for the safe introduction and long-term safe use of AI-based procedures in accordance with recommendations 1 and 2:

12. Where possible, established conventional methods should continue to be provided alongside AI-based methods so that users can decide for themselves which method to use and, in case of doubt, can refer to a reference (see sections 4.3 and 5.3).
13. Further research should be promoted and funded to develop test procedures that enable the characterisation and quality assurance, including the potential limitations and difficulties of medical image reconstruction and processing when using AI methods (see sections 4.2.5, 5.3 and 10).

14. International coordination on approval, testing and quality assurance methods for the use of AI methods for image reconstruction and further processing of medical image data should be promoted (see sections 4.2, 5.2 and 8).
15. When acquiring data for the development of algorithms based on artificial intelligence, a data agreement between clinical and industrial partners should stipulate that a sufficiently large portion of the data be made available for independent, external verification (see section 4.2.2). These data could then potentially be made available to the competent body (see recommendation 6).

Statement

In the context of image reconstruction and further processing using AI methods for medical image data generated using ionising radiation, the SSK considers it helpful to consider the possibility of standards and its use in order to support the recommendations outlined above and make them easier to implement. The SSK therefore takes the following position:

– The SSK considers standards, or at least standardisation of the requirements for training and validation data, to be possible and useful for quality assurance. The same applies to procedures for the development, testing and quality assurance of AI-based applications (see chapters 8 and 10). The SSK does not consider standards of AI-based methods for reconstruction or image processing itself to be useful, as the individual algorithms function differently, classical reconstruction methods have not been standardised either, and the SSK believes that standards of individual methods are hardly possible and does not promise any additional security. In addition, AI-based methods are developing so rapidly that the standardisation process would be too slow.

Scientific rationale

3 Application of AI methods in medicine

3.1 Imaging diagnostics in radiology and nuclear medicine as well as interventional imaging

3.1.1 Noise reduction in medical imaging

Medical diagnostics makes extensive use of image information for diagnosis. When information is to be obtained from inside the body, imaging usually involves either ultrasound-based methods, magnetic resonance imaging, X-ray-based methods or nuclear medicine procedures. Methods that provide three-dimensional information are being used more and more frequently.

In all of these procedures, image noise is a key factor in the image quality that can be achieved and the associated diagnostic quality. In the above-mentioned imaging methods, noise correlates either inversely with radiation exposure (X-ray-based and nuclear medicine procedures) or with measurement time (nuclear medicine, ultrasound and magnetic resonance imaging). The vast majority of X-ray-based examinations are projection images. Radiation exposure is comparatively low, so manufacturers are currently less focused on promoting the use of noise-reducing image processing techniques or publicising them than for tomographic methods. The image quality of medical tomographic imaging modalities in radiology (CT, MRI) and nuclear medicine (SPECT, PET) is affected by image noise to varying degrees, depending on the specific modality. In tomographic procedures based on the use of ionising radiation (CT, SPECT, PET), there is a direct or indirect relationship between image noise and the radiation dose associated with the examination. In CT, the latter is determined by the time integral of the tube current for the acquisition in question.

In nuclear medicine procedures, on the other hand, radiation exposure depends solely on the radiopharmaceutical used and its activity, but not on the selected acquisition time. The activity to be applied also depends on the physical and technical characteristics of the imaging system. Image noise is determined here by the parameters of activity and acquisition time. In contrast to CT, noise can therefore be reduced within certain limits by longer acquisition times without changing the activity of the radiopharmaceutical applied. However, since there are both practical (e.g. throughput, patient compliance, patient movement) and physical (e.g. half-life of the nuclide used) limitations for acquisition parameters, there is in principle a causal relationship for nuclear medicine methods as well between a reduction in radiation exposure through a reduction in the injected activity and the associated undesirable increase in image noise or deterioration in the signal-to-noise ratio.

The often desired reduction in radiation exposure leads to increased image noise unless countermeasures are taken. If the aim is to improve image quality while maintaining the same level of radiation exposure, this can often be achieved by using noise reduction techniques.

An essential requirement for noise reduction methods is the preservation of spatial resolution so that the images do not become blurred. Noise reduction must therefore be carried out in a way that preserves edges and is thus, by definition, not translation-invariant. Classical noise reduction methods often consist of one or more components that detect or evaluate edges and components that smooth the images in such a way that the edges are not smoothed or, at least, are only smoothed parallel to the edge. In most cases, the methods are also non-linear: regions with high noise are smoothed more than areas with low noise. In many cases, noise reduction is performed iteratively.

Neural networks, some of which are already used in practice for noise reduction, have the same objective. However, due to the complexity of the networks, the above-mentioned separation of edge detection and noise reduction does not exist or cannot be recognised – at least not without considerable effort.

The networks are often trained by subsequently adding noise to existing data (e.g. clinical images or volumes) using an algorithm in order to simulate lower radiation exposure. These low-dose data (input) and the associated normal-dose data (label) then serve as a data pair for training the network. Alternatively, noise reduction networks are also trained unsupervised, i.e. with a data pool of low-dose data and another data pool of normal-dose data, which do not have to come from the same patients. It is also conceivable to use purely synthetically generated data for training.

In CT, AI-based noise reduction methods have been offered by three of the four major CT manufacturers for some time (AiCE from Canon, TrueFidelity from GE, Precise Image from Philips). However, only sparse information is available about the algorithms. Ultimately, all three methods are probably supervised training methods. The input consists of images to which noise has subsequently been added and that have been reconstructed using filtered back projection (FBP). Canon and GE appear to use iteratively reconstructed images of the normal dose data as labels, whereas Philips uses FBP reconstructions of the normal dose data. Another thing these methods have in common is that they cannot be applied to images that have already been reconstructed but are linked to the actual FBP reconstruction so that they can be marketed as image reconstruction methods. The fourth manufacturer, Siemens, does not currently officially offer any AI-based noise reduction (or "reconstruction methods").

In CBCT (interventional C-arm systems, dental cone beam CT (DVT) systems, imaging CBCTs of radiotherapy equipment) and radiography (fluoroscopy, mammography, digital subtraction angiography), there are currently no known applications of AI methods for noise reduction installed in medical devices. In general, however, this is to be expected in the medium term, especially for dynamic imaging systems or CBCT systems. In this context, the approaches to noise reduction may well be similar to those used in CT, especially with regard to the related methods. In scientific publications, too, the majority of publications on noise reduction methods using AI methods relate to CT (Balogh und Janos Kis 2022, Pashazadeh und Hoeschen 2023).

In nuclear medicine, there is also a high demand for image noise reduction, which is generally high compared to CT for methodological reasons. The current standard approach is to filter iterative maximum likelihood expectation maximisation (MLEM) image reconstructions, which are stopped after a few iterations, using spatially invariant low-pass filters (e.g. Gaussian filters) that selectively suppress high-frequency noise. However, this filtering always results in a noticeable loss of spatial resolution, which particularly adversely affects the detection and quantification of structures close to the inherent resolution limit of the tomograph. There are also other methods that use locally adaptive edge-preserving filtering techniques. However, these are not yet routinely used for various reasons (variable quality of results, excessive dependence on various selectable parameters for adjusting the filter properties, etc.). Alternatively, the prior knowledge that the object under investigation is homogeneous even in regions and therefore cannot be heavily noisy can be used directly within the iterative reconstruction for noise suppression. This method is referred to as maximum a posteriori reconstruction (MAP reconstruction) with a "smoothing prior" from "prior knowledge" (Qi 2006). In this context, the latter is a "penalty function" that rejects noisy reconstructions.

Deep learning-based methods using deep neural networks, such as CNNs, now offer an alternative option to locally adaptive filtering of tomographic data sets, some of which are already available in everyday clinical practice. These methods enable a relevant improvement in noise reduction with less loss of resolution than is the case with today's common *white-box* methods. In this context – as in the field of image restoration in general (in addition to noise reduction in the narrower sense, this includes contrast enhancement, resolution recovery, motion correction, etc.) – a special network architecture has proven to be particularly popular and successful: the so-called U-Net (which belongs to the so-called encoder-decoder networks) proposed in 2015 (Ronneberger et al. 2015). This architecture was originally developed for segmentation tasks, but is now successfully used in a wide range of image restoration tasks.

The most widely used approach for so-called network training, i.e. the optimisation process for determining suitable values for the usually extremely large number of free parameters of the network (e.g. weights of learned convolution kernels), is *supervised learning*. In the context of noise reduction, the network is trained to restore images with a significantly improved signal-to-noise ratio from suitable training data from noisy images. As far as nuclear medical imaging is concerned, the training uses, where available, data sets that have been generated with the clinically standard activities currently in use and the longest possible acquisition times, so that a high-quality reference is available. These reference data sets are often referred to as *ground truth*. High-quality, low-noise images are particularly helpful for training denoising networks in PET. Data from *"total-body PET"* systems are now also available for obtaining such image data. These have only been on the market since 2020, are characterised by extremely high sensitivity and can therefore deliver low-noise, high-quality images with shorter acquisition times or lower activities.

From raw nuclear medicine data, which are usually recorded in list mode format (list of photon detection events), virtual raw data sets can be simulated by stochastically extracting a subset of data (*Poisson thinning*), which are comparable to images with shortened measurement times or reduced applied activity or a scanner with lower sensitivity (shorter axial field of view, i.e. extension of the tomograph in the head-to-toe direction). The reconstruction of the virtual partial data and the reconstruction of the original complete raw data make it possible to generate pairs of images consisting of a simulated input image with increased noise level and the original *ground truth* image retrospectively (stochastically correct). These image pairs can then be used to optimise and validate the network parameters.

A key requirement for training neural networks is the availability of a considerable number of validated input-output image pairs, which must cover the range of inter-individual variability in image properties that occurs in practice. In the context of nuclear medicine, these include different physical and biological tracer properties, non-standardised applied activities and acquisition times and, as a result, different signal-to-noise ratios, (patho)physiological and anatomical differences between different patients, and differences in the image quality of different scanners. Adequate coverage of this range is important in order to obtain networks that deliver valid and robust results in everyday clinical practice. Furthermore, it is essential to know the range of training data used in order to be able to draw conclusions about the suitability of a given network for a given question, such as: Can a given, trained network be used for images generated with a different radiotracer (or device) that could not yet be included in the training data?

Since the low-noise reference images required for *supervised* training are not available or generatable for all fields of application, other approaches to generating reference images pursue complementary strategies. These include, for example, (a) the use of Monte Carlo-simulated data, where the noise-free reference image is known, (b) more complex reconstructions (MAP) or post-processing of clinical images with edge-preserving filters or structural prior knowledge.

In addition, there are a variety of training methods that aim to reduce the amount of training data required. These include *transfer learning* and *unsupervised* training methods. In *transfer learning*, networks are pre-trained on a specific data set from one domain (e.g. a large data set of frequently used radiotracers). Subsequently, some of the network parameters are further optimised (parameter fine-tuning) using an additional small data set from another domain (e.g. a new radiotracer). *Unsupervised learning* methods do not even require the presence of reference images. These include the *"deep image priors"* method and "conditional generative networks". However, the latter are more difficult to train than *supervised* or *semi-supervised learning* methods and are definitely inferior when good reference images are available.

Compared to conventional methods, DL approaches are characterised by very simple implementability and high flexibility and are often superior to established conventional methods in visual assessment or avoid the need for frequent user intervention, thereby minimising interobserver variability and time requirements. On the other hand, as with all CNN applications, there is the *black-box* nature of the ultimately operational algorithm, which essentially eludes mathematical analysis. Accordingly, in the context of noise reduction, the essential questions of verifiability and generalisability of the results achieved in training and validation cohorts to the patient data to be processed prospectively in clinical use must also be considered essential and not conclusively resolved.

There is a risk that DL-based noise reduction methods will be used uncritically, under the misperception that they allow the exposure times and injected activities in nuclear medicine examinations or the radiation exposure in radiological examinations to be reduced without any loss of quality or information. There are only a few multicentre studies on this topic that are relevant to clinical questions.

3.1.2 Image reconstruction

In medical imaging, image reconstruction refers to the calculation of a medical image or image stack (volume) from the measured raw data. In general, this is a function that maps from the raw data domain to the image domain. The raw data domain depends on the modality. For example, MR imaging is measured in Fourier space, CT imaging in projection space (X-ray transformation), and PET imaging involves line integrals from all possible directions. Image reconstruction therefore mediates between two different domains and is thus a solution to an inverse problem. Algorithmic methods that only process an image or a stack of images and generate a modified image or a modified stack of images are usually referred to as image processing or image restoration, but not image reconstruction. The processing of already reconstructed images is also referred to as post-processing. This section deals with image reconstruction methods based on AI methods. Post-processing methods are covered in the sections on noise reduction and super-resolution. For marketing reasons, some CT manufacturers link (AI-free) image reconstruction with AI-based image post-processing and then incorrectly refer to AI-based reconstruction.

In addition to their use for noise reduction on already reconstructed images, ML methods can also be integrated directly into the tomographic image reconstruction process, which calculates multidimensional images from measured "raw data" (in this context, "projection data"). The aim of this direct integration of ML methods is to improve the resulting image quality compared to conventional reconstruction methods in terms of various parameters such as noise, resolution or the occurrence of various image artefacts.

Various scientific comparative studies (Muckley et al. 2021, Sidky and Pan 2022) have shown that methods that integrate ML directly into the (iterative) reconstruction process can be superior to post-reconstruction ML methods. However, it should be noted that training ML reconstruction methods is often very time-consuming and the implementation and application of these methods is more complex compared to post-reconstruction ML methods, which hinders the widespread use of the former.

The approaches to integrating ML into medical image reconstruction published in recent years can be broadly categorised as follows:

1. In *"end-to-end learning"*, a network is trained to transform recorded projection data directly into images without using existing knowledge about the physics of data acquisition or the mathematics of tomographic methods. Examples of this are "AUTOMAP" (Zhu et al. 2018) or "DeepPET" (Häggström et al. 2019). *Compared* to other methods, however, *end-to-end learning* has the disadvantage that very large networks and very large amounts of data are required for training. Furthermore, it has been shown that these approaches often suffer from falsely added image structures and are not superior to other less complex ML methods.
2. In *physics-driven learning*, conventional *image updates* based on the physics and statistics of the data acquisition process are combined with learned regularisation steps in the form of trainable networks. Examples of this are the so-called *"unrolled variational networks" (*Hammernik et al. 2018, Mehranian et al. 2021), "learned primal-dual" reconstruction (Adler and Oktem 2018), "plug-and-play" methods (Ebner and Haltmeier 2024) and the learned filtered backprojection (Würfl et al. 2016).
3. "Conditioned generative models" aim to first train a generative model that is designed to learn the diversity of possible high-quality images. After network training, the generation process is conditioned with the measured projection data in order to obtain a high-quality image that is simultaneously "compatible" with the projection data (Chung et al. 2022, Singh et al. 2023).
4. In *"post-reconstruction learning"*, networks are trained to improve the quality of a "conventionally" reconstructed image after reconstruction. Conceptually, this method is similar to the noise reduction described in section 3.1.1, but also includes resolution enhancement and artefact reduction (Berker et al. 2018).
5. In *"learning the physics of the data acquisition process (learning of the forward model)"*, networks are trained to learn certain physical processes that are "classically" difficult to model in a data-driven manner. Examples of this include the modelling of scattered radiation in CT (Maier et al. 2018, Maier et al. 2019b) and PET reconstruction (Schramm et al. 2021).

The above approaches differ in terms of the complexity of the required networks, generalisability, and the amount of training data required. It can be shown mathematically that methods that make greater use of known knowledge about the physics and statistics of the data acquisition process and the mathematics of tomography (e.g. with known forward operators) generalise better and require less training data (Maier et al. 2019a).

"Bayesian learning" is another promising ML reconstruction approach that aims to generate not only the reconstructed image but also a confidence map representing the regional uncertainty of the ML reconstruction model.

As with all networks, so-called "adversarial attacks" – attacks in which input data are deliberately manipulated by exploiting known vulnerabilities in order to intentionally mislead AI models in an already trained network – are a problem. However, prior knowledge can help to secure the network in as many parts as possible so that it can be analysed using classical methods (Gottschling 2022, Huang et al. 2018).

ML reconstruction methods in radiology

The idea of using prior operators can also be applied to networks inspired by classical filtered backprojection to better approximate limited angle geometries that typically cannot be solved by classical analytical inversion models. Interestingly, errors in the discretisation or initialisation of the filter steps are intrinsically corrected by the learning process. The method is also compatible with other DL approaches that are capable of learning additional data-driven regularisation but, in principle, perform classical iterative reconstruction.

In the field of CBCT imaging, e.g. for interventional imaging, approaches similar to those used in DL-based reconstruction are generally found. Research focuses in particular on image reconstruction with limited projection angles, metal artefact reduction and the calculation of suitable trajectories. Optimal planning results in a reduced or modified number of projections for reconstruction, which can have a positive or negative effect on both radiation exposure and the impact of confounding factors.

ML reconstruction methods in nuclear medicine

A key requirement for nuclear medicine cross-sectional imaging is to ensure the quantitative accuracy of the image data, as nuclear medicine functional imaging is based in principle on the derivation of quantitative parameters from the tomographic image data. This means that the reconstruction method under consideration, including the above-mentioned scatter radiation and absorption corrections, must not only deliver distortion-free images with unaltered local contrast, but also ensure the correct global scaling of the data. With regard to the use of ML-based methods, this increases the need for suitable quality control instruments. Currently, most ML reconstruction methods in nuclear medicine focus on improving the image quality of *low-count* images. In addition, there are also ML approaches to reduce artefacts in PET images with limited angular geometry (Chin et al. 2022).

3.1.3 Resolution enhancement (superresolution)

Similar concepts to those used for noise reduction can also be used to improve the spatial resolution of images. These super-resolution solutions are primarily found in the processing of photographic images or videos. However, there is one application from the manufacturer Canon in the field of CT imaging: Canon's PIQE (Precise IQ Engine) is a neural network that aims to increase the spatial resolution of CT data beyond the resolution limit of the respective CT device (Canon Medical Systems n.d.). The intended area of application is CT cardiac imaging. With the Aquilion Precision, Canon has launched a CT device that offers approximately twice the spatial resolution of all other conventional CT systems except for the new CT from Siemens with a photon-counting detector. The high-resolution patient images from the Precision now serve as training labels. The training input is a smoothed version of the high-resolution images.

However, it should be noted that such methods do not actually improve the resolution of the imaging system, but merely make edges appear sharper. This is particularly important in CT in order to suppress blooming artefacts. This refers to the effect whereby calcifications in vessels appear larger than they actually are due to the narrow grey scale windowing typically used to view the vessels. This leads to radiologists tending to underestimate the lumen of the vessel at the site of calcification. With the help of super-resolution AI techniques, this effect would not occur and the lumen would be correctly assessed.

What such methods cannot achieve is an actual increase in spatial resolution: two closely adjacent lesions that appear as a single lesion in the CT system (without AI) due to the limited resolution remain visible as a single lesion even after the application of resolution-enhancing AI methods.

3.1.4 Artifact reduction in CT and X-ray diagnostics

3.1.4.1 Missing data

Some CT artefacts result from missing data. For example, metal artefacts arise when certain object components attenuate the X-ray beam so strongly that no usable information can be measured in the metal shadow. The projections therefore have one or more holes, i.e. areas of missing data that must be closed using an inpainting procedure, for example.

Another example is truncation artefacts. These occur when the object is larger than the CT detector. Truncation artefacts occur particularly in CBCT, as today's flat detectors are usually only about 40 cm in size, which means that the field of view diameter is approximately 20 cm due to the beam geometry, applying the beam set. To reduce the artefacts, the detector must be virtually enlarged and filled with meaningful data, i.e. data must be added around the actual field of view.

There are numerous suggestions in the literature for reducing radiation exposure by taking fewer projections per rotation while maintaining the same dose per projection. Even though it would be more obvious and sensible to simply reduce the dose per projection instead of the number of projections, the resulting artefacts can be reduced either by an inpainting method that supplements the missing projections or data, or by a special reconstruction method.

Ultimately, for the above-mentioned use cases in which data or data ranges are missing, damaged or faulty and therefore unusable, neural networks can be used to fill in these data gaps. The solutions described in the literature either do this directly in the raw data space, i.e. in the projection data, or they correct the corresponding artefacts in the image space and use this corrected volume for a forward projection. The data missing from the measurement data are then taken from the forward-projected data and copied into the gaps in the original data. This ensures that only the missing areas are supplemented in the original data. Correction algorithms that merely calculate a more visually appealing volume from the artefact-laden volume are considered unreliable, as they also alter parts of the good raw data.

Examples of such methods are described (Byl et al. 2021) in the field of metal artefact correction, (Fonseca et al. 2021, Kabelac et al. 2025, Ketola et al. 2021) for the detruncation of CT data using neural networks, and (Han et al. 2017, Kofler et al. 2020) and (Ronneberger et al. 2015) in the field of sparse-view projections.

3.1.4.2 Erroneous data

Another part of CT artefacts comes from erroneous data, such as incorrectly calibrated detectors, detector properties that vary over time, scattered radiation or beam hardening.

CT ring artefacts, for example, are caused by detector pixels whose response behaviour during measurement differs from that during calibration. This can be due to temperature fluctuations or the exposure history of the pixel or its neighbours. The correction can be performed traditionally or using neural networks, either in the raw data space or in the image space, or a combination of both. An example of such a ring artefact correction method using neural networks is described in (Trapp et al. 2022).

Scattered radiation artifacts are caused by scattered radiation that is generated in the object or in the pre-filters and reaches the detector from a direction that does not correspond to that of the primary radiation. Scattered radiation artefacts also occur when anti-scatter grids are used. In X-ray images, scatter artefacts lead to a loss of contrast, which, as described above, can be corrected by image processing measures. A physically correct correction is not necessary in this case. In order to correct scatter radiation artefacts in CT, however, the scatter radiation component must be estimated as accurately as possible and subtracted from the raw data. The gold standard for this estimation is the use of Monte Carlo scatter calculations, which, however, require computing times of hours or days and also require the presence of a 3D model of the patient, such as a CT volume. They cannot therefore be used to correct individual projection images (e.g. X-ray projection images or fluoroscopy images). In the field of AI, deep scatter estimation (DSE) is the standard method for estimating scattered radiation (Maier et al. 2018, Maier et al. 2019b). DSE does not require a 3D model and can therefore also be used for projection images. In addition, the runtime is very short at one to two milliseconds per projection. Based on the ideas of DSE, there are numerous follow-up publications in the scientific literature (Laurent et al. 2023, Xiang et al. 2020).

3.1.4.3 Motion correction and compensation in X-ray diagnostics

Patient movements during the examination lead to artefacts and loss of resolution in the reconstructed tomographic images, the nature and extent of which vary depending on the modality. DL methods are typically used to determine displacement vector fields (DVFs), which describe time-dependent three-dimensional, arbitrary and involuntary movement and deformation (organ mobility, respiratory and cardiac cycles). In principle, motion correction methods can be divided into two categories: those that modify the projection data fed into the image reconstruction in advance to compensate for patient motion, and those that operate on the motion-compromised reconstructed tomographic image data. In radiology, the term motion correction is reserved for post-reconstruction methods that essentially improve the visual impression of the image. Where possible, the projection data can be corrected during image reconstruction. This is often referred to as motion compensation. However, this conceptual distinction is unusual in the field of nuclear medicine, where both cases are referred to as motion correction.

In CT images, patient movements can lead to motion artefacts in both projection images and reconstructed CT volumes. Numerous algorithmic methods for reducing or correcting such artefacts can be found in the literature. In the field of AI, these methods are often cosmetic in nature and thus produce a visually appealing result, the accuracy of which, however, is difficult to verify. Where possible, motion compensation methods should be used. In contrast to motion correction, these methods estimate the actual motion of a voxel and then incorporate it into the calculation of the compensated images, for example, by moving the respective voxel during image reconstruction in accordance with the patient, so that the contributions are always added to the correct voxel during back projection. In motion compensation, the task of the neural network is usually to estimate motion, i.e. to generate a motion vector or a motion vector field (Maier et al. 2021, Maier et al. 2025). Here, too, incorrect estimates may occur. In contrast to pure correction, however, plausibility checks can be performed, e.g. with regard to the speed of movement.

3.1.5 Motion correction in nuclear medicine

In nuclear medicine, too, the influence of unavoidable voluntary and involuntary patient movement during acquisition is a relevant aspect for the resulting image quality. This is particularly relevant for nuclear medicine imaging due to the much longer acquisition times compared to computed tomography. Both cyclic and stochastic movement (restlessness) lead to a loss of resolution. The most significant problem in practice is respiratory cycle-correlated movement during whole-body examinations of oncology patients, which typically has amplitudes of 2 cm. In addition to sensor-based stroboscopic measurement protocols ("respiratory gating"), purely data-driven approaches are becoming increasingly important. In this context in particular, DL-based methods are attractive for similar reasons as in noise suppression and resolution recovery in image reconstruction and are in some cases very similar in design, since the problem of motion correction formally requires deconvolution of the image affected by motion blur.

3.1.6 Further methods for reducing radiation exposure in diagnostics

Methods for reducing radiation exposure in diagnostics refer to all methods that aim to reduce radiation exposure during diagnostic imaging procedures. In addition to the reconstruction methods already mentioned, which aim to reduce radiation by reducing the number of projections or the number of photons required per projection while maintaining image quality, there are approaches that already use AI methods during projection acquisition. These include collimation optimisation and automated patient positioning. Systems based on AI methods are already available, particularly for 2D X-ray diagnostics. One example is the "YSIO X.pree" method used by Siemens (Siemens Healthineers AG n.d.-a), which determines the position of the thorax using a 3D camera and AI methods and adjusts the collimation based on this. Omega Medical Imaging offers automatic calculation of the collimator setting based on target region determination (Omega Medical Imaging 2023). In the field of CT imaging, various manufacturers (Siemens FAST Integrated Workflow™ (Nikolaus 2018), Philips CT Precise Suite™ (Philips GmbH 2021), GE Revolution™ Maxima (GE Healthcare n.d.)) offer automatic patient positioning to move the target structure to the isocentre. A camera system with depth information is used for 3D patient reconstruction, from which the target structure within the patient is derived for positioning. However, more detailed information on the manufacturers' AI methods cannot be found.

3.1.7 Segmentation

Recent developments in deep learning have brought great advantages to image segmentation, in which the contours of organs or anatomical structures are precisely determined. CNNs have dominated this field, and several approaches have been developed using CNNs. Examples include Deeporgan, brain MR segmentation with CNNs, and fully convolutional multi-energy 3D U-Nets. In particular, the so-called U-Net (Ronneberger et al. 2015) has established itself as the state of the art in this field. Investigations into a wide variety of problems have shown that the original version from 2015 with automatic parameter setting is generally superior for certain problems. This self-configuring version of U-Net is known by the name nn-U-Net (Isensee et al. 2021), introduced by the authors.

Approaches to grey-box models have been around since 2004 and have experimented, for example, with the fusion of neural networks and active contour models. However, these approaches were still component-based (i.e., the neural network was trained for only part of the task and the contour model was optimised separately) and did not use deep networks or end-to-end training (i.e., for the entire task). It seems promising to rethink traditional segmentation approaches and merge them with deep learning based on end-to-end training. This makes it possible to significantly reduce the number of parameters while still obtaining stable results. However, the hybrid methods do not yet demonstrate the performance of black-box models.

Overall, segmentation techniques are important because they are fundamental to the contouring of organs and subsequent dose estimates, as well as to radiation therapy planning. At first glance, there does not appear to be any relevance to radiation protection in the sense of the advisory request described above, but due to their importance for radiation therapy methods, a more detailed examination will be necessary at a later date.

3.1.8 Classification

Based on segmented images or the detection of specific structures or structure clusters, classification methods are being developed for various applications in medical imaging, and corresponding models are already being used in commercially available analysis products. This applies in particular to breast imaging techniques, especially projection mammography and breast tomosynthesis. As already mentioned, image areas or entire images are divided into classes based on various characteristics. These classes range from assignment to specific organs, to tissue at risk or suspicious tissue areas, to the definition of tumour tissue or diseased tissue as a whole. In mammographic imaging, particularly in screening, classification is carried out in suspicious areas with regard to possible tumour diseases and potentially normal tissue, but there are also classifications for other areas of the body where other classifications play a role, for example for the detection of perfusion disorders or emphysema. The methods used today are primarily based on the use of large data sets with known diseases and corresponding image data sets. CNN approaches, often U-Net-based methods or extensions, but also almost all other DL-based approaches, are used for such classification strategies. Since these applications generally do not alter the image information in today's systems, but are used to indicate potentially diagnostically relevant information, this type of classification does not initially involve any radiation protection aspects. However, work is being done on classification systems in which decisions about the further course of treatment, including the use of ionising radiation, are made on the basis of the classification. One example is the classification of healthy tissue, tumour tissue and destroyed tumour tissue in minimally invasive tumour therapy (Mahmoodian et al. 2022). Since in this application, for example, decisions about further ablation are made using X-ray-based imaging, aspects relevant to radiation protection could arise. However, this issue is not specified in the advisory request. Furthermore, since such methods are still under development, this aspect will not be examined further here. However, it should not be forgotten for future considerations.

3.1.9 Radiomics

The process known as "radiomics" and all similar approaches go one step further than simple classification. Radiomics is based on the assumption that the information obtained from three-dimensional or four-dimensional data sets in image data is so diverse and unmanageable that DL methods are better able to detect subtle image information and assign diseases than a purely human evaluation of large amounts of data. In the development of radiomics methods, various parameters and characteristics of image data are therefore determined in general terms and tested on the basis of large case numbers to determine whether there are connections between such quantifiable characteristics or parameters (generally referred to as features) and certain diseases. The difference to the methods described previously in the area of classification is essentially that areas are no longer classified and assumptions are made about the cause of the classification, but rather that the pattern that can lead to recognition consists of a lot of information that is not directly accessible, such as the shape of structures, neighbourhood analyses, structural gradients and many others (Lambin et al. 2012). The same considerations apply to the radiation protection-related evaluation of radiomics methods as to the classification approaches.

3.1.10 Dose calculation

The calculation of patient-specific dose distributions in X-ray imaging and radiotherapy is based on the spatial distribution $\mu(x, y, z)$ of the attenuation coefficients of the patient's tissues. The best method currently available for calculating dose distributions is Monte Carlo simulation of all particle histories within the patient. However, such Monte Carlo calculations are computationally and therefore time-consuming, so that they often have to be replaced by approximation methods, such as kernel-based dose calculation. In connection with CT, dose distributions are currently calculated exclusively offline and only for scientific purposes. Widely used and sometimes routinely recorded dose values, such as the CTDI value, the dose length product (DLP), the size-specific dose estimates (SSDE) or the effective dose calculated from the DLP using the k-factor are not patient-specific, but refer to dose values of a phantom followed by a few simple conversions.

However, in order to optimise scan protocols, it may be useful to determine the dose distribution in the patient in advance. For example, in CT there is what is known as tube current modulation (TCM). If the attenuation value distribution of the patient and thus the location of the organs were roughly known before the scan, and if a dose distribution could then be calculated in a very short time, tube current modulation could be performed with significantly reduced radiation exposure and thus reduced radiation risk for patients (Baader and Kachelrieß 2025, Klein et al. 2022). The prerequisites for this can be achieved with the help of AI methods, e.g. by calculating a CT volume from the topogram (i.e. the first scan performed as a projection image for setting the scan area) followed by a dose calculation in real time. It is therefore foreseeable that AI methods will be used in the future that have a direct influence on the scan execution and thus on radiation exposure.

For the dose calculation itself, AI can be used as follows (Maier et al. 2022b): Based on the patient's CT volume and the scan geometry, a first-order dose distribution is calculated, i.e. a dose distribution as it would appear without scattered radiation. This calculation corresponds to a back projection of the radiation intensity into the volume and can be performed in a matter of nanoseconds. Based on the CT volume and the first-order dose volume, a neural network trained with Monte Carlo calculations, for example, then calculates the dose distribution including all higher orders.

3.2 Radiotherapy

The use of AI in radiotherapy, except for image reconstruction and noise reduction for diagnostic applications such as radiation therapy planning and staging, is not covered by this recommendation. It should be noted here that, in addition to the requirements that apply to images created for diagnostic purposes, it is also particularly important for image reconstruction and noise reduction using AI in CT imaging for radiation therapy planning that the Hounsfield units in the reconstructed images are not significantly altered by the AI methods. For the sake of completeness, however, other potentially relevant areas of application for AI in radiotherapy will be briefly discussed.

3.2.1 Pseudo-CT

In radiotherapy, the term pseudo-CT refers to the creation of an artificial CT image based on a real MR image. This procedure is used if no CT image is available for the patient or if CT scans should not be repeated too frequently for radiation protection reasons. Pseudo-CT can thus serve as an aid for planning or optimisation without exposing the patient to radiation. With the help of registered MR-CT image pairs, a neural network can be trained to generate corresponding Hounsfield unit maps based on the MR images. There are already a few examples in commercial use: The Siemens system, Synthetic CT™ (Siemens Healthineers AG n.d.-b), combines a U-Net for tissue segmentation and a cGAN (conditional Generative Adversarial Network) for CT generation. This network consists of a generator that produces images and a competing discriminator that evaluates the realism of the image. Spectronic Medical AB (integrated into the systems of Canon and GE Healthcare (Spectronic Medical AB n.d.)) estimates a local affine transfer function based on several 3D CNNs, according to the manufacturer. MRCAT™ Brain from Philips (Philips Austria GmbH n.d.) was developed for applications in the head and neck area and generates a synthetic CT image; however, the underlying method has not been disclosed. In addition to classical U-Net-based approaches, recent literature also increasingly features methods based on so-called denoising diffusion models, another category of generative models that gradually add noise to a structured image in order to generate new data sets and thus learn to denoise data sets.

3.2.2 Dose calculation

Similar to what is described in section 3.1.10 in the case of CT, dose calculation can also be performed in radiotherapy (Martins et al. 2023). Based on the attenuation value distribution of the patient, i.e. the CT volume, an approximate, computationally efficient dose calculation method is first used to obtain an approximate dose distribution. A neural network then converts this into a realistic dose distribution. Other approaches only consider the density distribution of the patient (i.e. a CT image data set or a pseudo-CT image data set) and the information about the therapy beam, i.e. the radiation therapy plan, and a neural network calculates the dose distribution from this. An initial calculation of an approximate dose distribution is not required.

3.2.3 Decision support systems

As already described in sections 3.1.8 on classification and 3.1.9 on radiomics, image-based approaches are used in conjunction with DL methods to extract specific information from the image data sets that is suitable for supporting the diagnosis by providing clues. These methods are also used to generate suggestions for therapy optimisation. This also applies to radiotherapy procedures, whether minimally invasive or brachytherapy procedures, or for the (online) optimisation of teletherapy methods – for example, by changing radiation therapy plans or by determining new radiation therapy parameters for better compliance with treatment plans. The systems generally suggest diagnostic options or treatment options and/or changes to the treatment regimen, thereby helping the physician to make treatment decisions. In addition, radiation therapy plans are already being created automatically. Here, too, virtually all common DL-based approaches are used. These approaches are playing an increasingly important role, particularly in adaptive radiotherapy, due to the growing possibilities of online imaging and the algorithms available. Due to the rapid development in this field, it was decided not to list the algorithms and products currently available on the market. The decision support systems used in radiotherapy applications are relevant from a radiation protection perspective, as they can alter radio-oncological treatment. However, the final decision rests with the specialist physician, so there is currently no clear direct aspect under radiation protection law. A reassessment will be necessary as soon as, on the basis of the proposed changes to treatment regimens, decisions have to be made so quickly that a genuine reassessment by the physician is no longer possible, which is conceivable in the future in the interests of better patient management.

4 Framework conditions

4.1 Legal framework conditions

When considering the use of artificial intelligence in radiation-based medicine, and in particular for the use of noise reduction and reconstruction techniques for X-ray imaging, four aspects are of particular importance from a legal perspective. Firstly, the use of radiation on humans is only permissible if it is associated with an expected benefit. Further details on this are discussed in section 4.1.1 below on the "justifying indication". Secondly, there is a requirement to optimise patient exposure in the medical use of ionising radiation. The implications of this requirement are discussed in section 4.1.2 "Optimisation requirement". Thirdly, it must be taken into account that liability issues arise if an incorrect clinical decision is made on the basis of added or suppressed image information. Fourthly, it is important to consider the legal implications for the development of models and training in the application of artificial intelligence arising from the data protection requirements of the GDPR (EU 2016), as patient data must be used in this context. A corresponding assessment can be found in section 4.1.4.

In addition to the legal framework, ethical aspects of the use of artificial intelligence for medical imaging may also play a role. This naturally concerns the use of patient data for training the models that will later be used for image processing and is therefore related to data protection requirements. In addition, questions must be considered that are related to the equal availability of optimised methods and thus also to the potentially reduced risk associated with the use of ionising radiation. AI-based methods that allow for dose reduction in imaging may not be available everywhere or for all patients because they are expensive, can only be used safely on certain devices, or because the patients belong to so-called out-of-distribution cases.

Conversely, possible misdiagnoses and the responsibility for them are associated not only with legal aspects, but also with ethical ones. However, these aspects play a subordinate role in the consideration of issues relevant to radiation protection. This also applies to decision support systems, which are not the subject of this recommendation anyway and were only considered in section 3.2.3 for the sake of completeness. Their ethical dimension should not differ from previous methods as long as the final clinical decision is made by the treating physician together with the patient.

4.1.1 Framework conditions based on the justifying indication

The use of ionising radiation on humans in the context of medical imaging diagnostics is only permitted on the basis of a justifying indication. This is given if the potential risks of radiation use are outweighed by its expected benefits (Section 83(3) StrlSchG, (StrlSchG 2017)). To this end, the specialist physician who establishes the justifying indication must be able to assess the significance and possible limitations of the respective examination. This is possible in imaging procedures without the use of machine learning because there is sufficient relevant experience with them and the mathematical methods used to calculate the diagnostic images from the raw data are defined and known.

If, on the other hand, the generation of images from raw data is based on machine learning methods and the "solution path" is neither predictable nor disclosed retrospectively, then any assessment of the significance and limitations is subject to uncertainty. Especially when the raw data are subject to a high level of noise and the images reconstructed from them using conventional methods can only be interpreted with limitations, it cannot be ruled out that the images generated using AI methods no longer contain diagnostically important details or that they contain details that do not actually exist.

For establishing the justifying indication, this may mean that when AI methods are used for image reconstruction from raw data with a high level of noise, the assessment of a possible benefit is subject to uncertainty unless measures are taken to ensure that the image information has not been unintentionally distorted.

Together with a note on the use of image data for radiation therapy planning in chapter 3.2, this results in recommendation 1 of the SSK in chapter 2:

For the use of AI-based methods for image reconstruction or image processing in medical imaging procedures based on the use of ionising radiation, the SSK recommends that

- when using AI methods for image reconstruction or further processing of patient-related image data, care should be taken to ensure that all structures relevant to the diagnosis are retained with essentially unchanged sensitivity and specificity in terms of diagnostic reliability compared to non-AI-supported methods, and that no artificial structures are generated in order to rule out image processing-related misdiagnoses that would call into question the benefit of the examination and thus the justifying indication. When using computed tomography image data for radiation therapy planning in radiotherapy, the Hounsfield units must not be significantly altered.

4.1.2 Framework conditions based on the optimisation requirement

Although current radiation protection legislation does not specify any exposure limits for patients when ionising radiation is used for diagnostic or therapeutic purposes, it does stipulate an optimisation requirement (Section 83(5) StrlSchG, (StrlSchG 2017)). In particular, this means that diagnostic imaging methods must be optimised in such a way that exposure to ionising radiation is reduced as far as possible while maintaining diagnostic accuracy. This reduction in radiation exposure is limited by physical and technical constraints. However, AI-based methods, especially DL-based methods, could be suitable for truly exploiting these limits. In this sense, if this cannot be guaranteed by other methods, the use of DL-based methods may even be indicated. However, it must be ensured that

- the physical limits are not exceeded (i.e. sufficient patient- and situation-related information is available to be able to represent real information) and
- no relevant information is lost during optimisation or additional information that is not available is added, because otherwise the diagnostic significance will not be preserved.

This gives rise to the following two recommendations 2 and 3 from the SSK:

- When reducing the dose in combination with the use of AI methods to generate the images to be evaluated, it should be possible to identify all structures relevant to the evaluation with comparable sensitivity and specificity as it would be the case without dose reduction and AI use, in order to satisfy the justifying indication for the examination.
- Human-observer studies should be conducted to verify compliance with this recommendation, unless scientifically validated results are already available that confirm the equivalent significance of metrological methods. Care must be taken to ensure that the metrological methods are selected in such a way that AI methods are not given preference in the evaluation and that the absence or generation of subtle structures is adequately taken into account in the examination (see section 4.2.5).

4.1.3 Legal issues arising from potentially caused incorrect treatment

In addition to the actual radiation protection issues that may arise due to incorrect additional or suppressed information in the context of the use of ionising radiation on humans without the expected benefit of a reliable diagnosis (see sections 4.1.1 and 4.1.2), other legal issues will certainly also be considered relevant. This applies in particular to the possibility that incorrect information or the incorrect presentation of information – which can be very difficult to prove – may result in incorrect treatment or therapy decisions. This can have far-reaching consequences and, for example, lead to a tumour or other serious illness not being detected or being assessed as less serious, resulting in necessary treatment not being provided. However, the opposite may also be the case, whereby an illness is considered more serious than it actually is due to the application of AI, or is only presumed to exist, resulting in unnecessary treatment. In both cases, the question of who is responsible for the therapy decision or treatment strategy becomes relevant. However, these questions are not relevant in terms of radiation protection law and are therefore not discussed here; it should be noted, however, that they must be taken into account when introducing or using the methods.

The extent to which special medical device law aspects need to be considered in AI-based image reconstruction methods is currently difficult to predict. In general, a procedure that is subject to the Medical Devices Act should do what it is designed to do – in this case, produce images that accurately reflect reality. In the case of "black-box" methods, whose mode of operation cannot be monitored and documented, this can be called into question unless sufficient empirical, experimental or clinical application data prove that the medical device fulfils its stated purpose. However, the use of methods that represent a black box for the user and even for the manufacturer as part of image generation is a novelty. This is not yet taken into account at all by current medical device law.

4.1.4 Data protection aspects

AI models that are intended to generate diagnostically usable images from raw data with a high noise content must be trained with data sets from examinations of patients who have undergone the examinations for clinical reasons. These data are transmitted for the development of the AI methods, possibly with supplementary data, e.g. clinical data. The recipients are predominantly third parties, i.e. the manufacturers whose devices are operated in the institution and who intend to develop corresponding AI methods.

Complete anonymisation of the data sets is practically impossible, as a pseudonym must be used to record which data sets can be assigned to the same person. Similarly, relevant clinical information such as the type of pathology depicted and general information such as age, gender and, where applicable, ethnicity must be identifiable. In principle, it is never possible to completely rule out the possibility of tracing data from imaging examinations back to the individual, e.g. based on facial or body contours, skeletal features or changes due to previous medical interventions, even if data that enable direct identification, such as the name, have been removed. The storage and transfer of these data are therefore subject to the requirements of data protection law. The informed consent of the persons concerned must also be obtained, either directly in relation to the research purpose or at least in connection with the treatment and examination contract with the institution.

4.2 Technical framework conditions

This section explains the technical framework for the use of AI for image reconstruction and its potential impact on dose reduction. To this end, it explains the training of deep neural networks (DNN), the backpropagation algorithm and variants such as supervised learning and unsupervised learning (including autoencoders). Difficulties arise in obtaining a sufficient number of training examples, which must be collected in compliance with data protection regulations, and in providing sufficient computing capacity. There is also a lack of effective and objective measurement methods for assessing the quality of AI reconstruction methods.

4.2.1 Training of deep neural networks and backpropagation algorithm

AI methods have the potential to improve image reconstruction in medical applications while reducing radiation exposure for patients. This technology is based on deep neural networks

(DNNs), which consist of many layers of artificial neurons. These networks are trained using algorithms such as backpropagation to recognise complex patterns in the data and make predictions or reconstructions based on them.

The backpropagation algorithm is used in artificial neural networks to adjust the weights of the neurons during the training process. It calculates the gradient of the network's loss function by repeatedly applying the chain rule to the individual layers of the network. This gradient is then used to optimise the weights of the network through gradient descent in order to minimise the loss function, i.e. the parameters of the model are adjusted iteratively until there are almost no changes. By adjusting the weights during the training process, the network can learn to recognise patterns in the input data and make accurate predictions, and even learn complex problems such as image reconstruction.

There are different training approaches for DNNs, such as supervised learning and unsupervised learning. In supervised learning, both input data and the correct output data (labels) are presented to the DNNs in order to adjust the network. In contrast, unsupervised learning does not use labels and instead attempts to model the structure or distribution of the input data directly. Autoencoders are an example of unsupervised learning, in which a DNN attempts to compress the input data efficiently and then reconstruct it.

4.2.2 Data collection and data protection in Europe

A key problem in the application of AI in image reconstruction is the need to collect many thousands of training examples in order to train the neural network effectively. In Europe, collecting these data is challenging because strict data protection laws restrict the use of patient data. It is also important to cover a sufficient variety of pathologies, patients of different ethnic origins and imaging protocols across different clinical standard procedures to ensure a comprehensive and reliable AI solution. Large industrial providers around the world are already accumulating large amounts of such data, although the quality and origin of the data cannot always be clarified beyond doubt. Among other things, the realisation of the European Health Data Space could be used to make suitable data available on a large scale.

4.2.3 Computing capacity for training deep networks

The hardware requirements for developing DNNs for tomographic imaging are substantial. This is due to the fact that the iterative optimisation process is inherently computationally intensive and operates on a large amount of three-dimensional training data. In terms of hardware, high-performance graphics cards (GPUs) have become the norm for this task. Even with this dedicated hardware, training the network can take several days of computing time in typical applications. Furthermore, since the design and optimisation of a new network architecture is usually an empirical or heuristic trial-and-error process that requires repeating the training of the network after each modification and then comparing the results of different design variants, the availability of the most powerful hardware possible is of crucial importance for the efficient development of new DL approaches.

A particular problem in the implementation of AI solutions, especially for image reconstruction and dose reduction, is the computing capacity required for training DNNs. Such computing power is often available at large companies and global players, while universities, clinics and public institutions may not have sufficient access to such resources. For smaller studies and highly specialised procedures, the standard equipment of research institutes is typically sufficient. However, for processing large amounts of data, such as those required for foundation models, the computing equipment may be inadequate. In particular, traditional large-scale computing systems in Germany rely on classic central processing units (CPUs), while a large number of ML algorithms require general-purpose graphical processing units (GPUs), which are not yet sufficiently available in all German high-performance computing centres. This poses a significant hurdle to the reproducibility and dissemination of these methods. Furthermore, the leading manufacturer of GPU computing hardware (TSMC for Nvidia) is located in Taiwan, which could become a geopolitical risk.

In contrast, the hardware requirements for the use of trained networks are less critical, and standard modern workstations are usually sufficient for imaging diagnostics.

4.2.4 Software

On the software side, a small number of comprehensive, freely available libraries have emerged over the last decade, which dominate the field to a large extent and are used as "building blocks". The following development platforms are particularly noteworthy here:

- Tensorflow (Google),
- PyTorch (Meta),
- MONAI (Nvidia),

which are currently used in the vast majority of applications of DL methods (including medical imaging). This has made it comparatively easy to apply standard techniques to specific tasks. At the same time, however, the degree of complexity of the calculations involved is such that users are rarely able to develop a deeper understanding of the behaviour of trained networks. This is particularly the case when predictions about network behaviour are required when applied to data that lie at the edge of or outside the manifold of the training data.

4.2.5 Quality management and measurement methods

Another challenge in the field of AI-supported image reconstruction and, in part, other AI-based image processing algorithms is quality management. Positive results are often found in the literature, while negative study results are published less frequently. There is currently no recognised measurement method that can reliably measure the quality of AI reconstruction methods in particular. Traditional quality assurance tools such as phantoms are susceptible to manipulation when included in the training data.

In order to use AI methods safely for dose reduction and to avoid hallucinations (i.e. image content that does not exist in reality but is generated by the AI-based method), effective and objective measurement methods for assessing quality are essential. Depending on the extent of AI use and the number of free parameters, different approaches can be considered. It is important to develop and establish such methods in order to comply with the radiation protection principles of justification and optimisation and to promote confidence in the use of AI in medical imaging. Ultimately, it is essential that the methods used can be shown to be equivalent to existing methods or procedures in terms of diagnostic reliability. For this reason, so-called human-observer studies should be regarded as a reference. These studies can be used to evaluate the sensitivity and specificity of a procedure for a specific diagnostic question. However, these studies are very costly, and their significance depends on the number and variance of the cases considered, the number and suitability of the persons reading the images, and the complexity and difficulty of the question under consideration. Sufficient statistical support must therefore be ensured for the intended comparisons in all cases. This must be documented. If measurement techniques are to be used to demonstrate the comparability of AI-based methods with standard methods, it must be ensured that these methods would show equivalent results to human-observer studies. This must be proven in comparative studies conducted using both approaches –

i.e. human-observer studies and measurement studies – and showing comparable results using representative examples. In any case, examples with representative differences in diagnostic reliability in human-observer studies and those without such differences must be included. Appropriate methods for AI-based applications for image reconstruction and noise reduction still need to be developed and evaluated.

The SSK therefore recommends that

*cor*respond*ing research funding should be made available as a key step towards the safe application of AI-based methods for these areas of medical imaging, especially when using ionising radiation.*

An essential aspect for assessing the performance of an AI method is to apply the method to independent, well-characterised test data that have not been used previously in the development and initial evaluation of the method ("unseen data"). Such a database should therefore not be freely accessible, but reserved for use by a competent body in order to rule out the aforementioned problem of possible manipulative "optimisation" of the method by integrating the said data into a re-training of the method under consideration.

Accordingly, in recommendation 5 in chapter 2, the SSK recommends that

- a competent body should be designated to verify the performance of the algorithm for the specified use cases on independent, sufficiently characterised test data. This competent body can either be an existing body into which additional competencies are integrated, or it can be newly created.
- Appropriate legal regulations or sub-legal provisions should be put in place.

The individual tasks of this body are set out in the explanations in the various chapters and are summarised in recommendation 5. The following recommendations describe the possible implementations, which are also based on the various explanations in the following chapters.

4.3 User-specific framework conditions

Users have certain requirements for AI-based systems, which are listed below:

Speed: The processes in radiological and nuclear medicine facilities are tightly scheduled and interlinked, so that significant delays in the provision of images calculated using AI methods are likely to be unacceptable. At the same time, it is also possible in principle to speed up procedures and imaging, which can certainly be an advantage in everyday clinical practice.

An *improvement in image quality* should be recognisable based on visual impression alone, not *solely* on numerical parameters such as the contrast-to-noise ratio on clinical or phantom images, or a modulation transfer function. Depending on the application, an improvement in image quality can mean that images calculated using AI methods are of equivalent quality to those produced conventionally, despite significantly reduced exposure or applied activity, or that the images are noticeably better while maintaining the same dose. The evaluation of image quality should not be limited to a mere subjective impression of the image, which is undoubtedly influenced by preferences and viewing habits: image noise, sharpness or the clarity of pathological findings and important anatomical image features are examples of possible, objectifiable criteria. From the perspective of radiation protection and the clinical user, diagnostic significance is the decisive criterion.

The *anatomical correctness* of the images is essential. It goes without saying that existing structures must not be suppressed and non-existent structures must not be generated, but the requirements go far beyond this. The sharp or blurred boundaries of organs and lesions, the smooth or irregular texture of edges, such as the surface of organs, are important diagnostic features that must not be distorted under any circumstances. This also includes blurring due to the partial volume effect of interfaces that run diagonally through the slice and provide important information about the anatomical conditions. On devices already in clinical use, for example, the effect has been observed that the liver surface is incorrectly reproduced in the image as having small bumps, which would be an indication of cirrhosis of the liver. Another observed artificial false finding can occur at the boundary between the liver and the filled gallbladder, which runs through the slice at a flat angle. Normally, the boundary would be correctly displayed as blurred between two different grey values due to the partial volume effect, but a transition with a discrete grey value was calculated, which mimicked a thickening of the gallbladder wall, e.g. due to inflammation. The uncritical removal of diagnostically important artefacts or acceptance of side effects during image smoothing can lead to misdiagnoses.

The *availability* of *conventionally calculated images* as a reference is indispensable, at least until AI-based image reconstruction has been evaluated to such an extent that anatomically correct representation can be trusted. Until then, the regular comparison of images reconstructed using AI and conventional methods should be part of the implementation process.

For the same reason, *AI-based images* must be *labelled* accordingly in the overlays, as must medical findings created using these images.

The *integrity of the images, also with regard to hardware and software updates*, should ensure that findings do not appear to have been added or disappeared, or that other image characteristics have changed, solely as a result of reconstruction or image processing in general. Of course, images also undergo changes when switching to a new generation of devices, but these changes are predictable and explainable given knowledge of the physical and mathematical processes involved. However, when changes are made to AI-based reconstruction or image processing, the necessary transparency is not necessarily provided. Similarly, a minimum level of *comparability between images from different manufacturers* must be ensured.

It must be ensured that the physician fulfils his or her responsibility when using ionising radiation and that the diagnosis can be made on the basis of the available images. This means that the physician must be able to be sure that the information displayed is genuine and that no information has been lost or added.

In practice, however, it has been observed that the use of AI methods creates the risk of a loss of skills and routine among medical staff. In addition, time pressure and resource shortages in clinical care can mean that the results of AI methods are not actually checked in detail by medical staff. Measures must therefore be taken to prevent this development. However, aspects of practical implementation must be considered, in particular the potential loss of skills and routine, also against the background of clinical time pressure and resource shortages.

In particular, the data used for modelling or training DL-based methods must meet certain quality requirements. This does not mean that only ideal images with the best image quality are required, but rather that the conditions under which the images were taken are known and meet certain specifications and, in this sense, must be of high quality. This makes it possible to show how accurately the methods can work, what deviations and uncertainties in image reconstruction or image representation are to be expected, and how this depends on patient parameters. Knowledge of this accuracy and the potential deviations is just as important for the physician as knowledge of the procedure itself in order to be able to assess how confident he or she can be in their diagnosis and/or therapy decision. It must also be verifiable and traceable how this may change with more data sets available. Support in assessing the procedure and its effects can be provided, for example, by an MPE. The MPE in radiology and nuclear medicine is responsible for the quality assurance of radiological and nuclear medicine facilities and the selection of devices and equipment, which includes AI methods.

5 Potential problems and limitations

5.1 Garbage in, garbage out

In machine learning, the quality of the training data is crucial to the performance of the trained model. If the data are of poor quality, the models will only deliver suboptimal results. This problem is often referred to as "garbage in, garbage out", which means that if poor data are entered into a model, poor results will come out.

Quality assurance of training data should therefore be a high priority to ensure that the data are error-free, consistent and representative. This includes, for example, checking for incorrect or incomplete data, removing outliers and balancing classes to ensure that the model is not overtrained on one class. The number of training examples must also be appropriate in relation to the number of training parameters obtained from them.

Overall, it is important to be aware that the quality of the training data directly influences the performance of a model. It is therefore crucial to carefully check and ensure that the data are of high quality in order to achieve accurate and reliable results. Using high-quality data does not mean that only ideal images with the best image quality must be used, but rather that the conditions under which the images were taken are known and meet certain specifications as indicated above.

As described in the introduction (chapter 1.2), the data are divided into training, validation and test data. It is important that this is done randomly. If this were not the case, the training, validation and test data might have different statistical distributions. For example, it would be questionable to use only data from obese patients for training and validation and then only data from patients with a low BMI for testing. In order to assess whether the method has been trained with meaningful data, it is important to comprehensively document the data used. This documentation includes a description of the patient cohort, the body region examined (e.g. head, neck, thorax, abdomen, pelvis) and the imaging parameters. It must also be known whether data from test objects were included in the training exclusively or additionally. The description should also include the criteria used to select the data and a justification for this selection. Statistical parameters (minimum, median, maximum, etc.) relating to patient data such as age, height and weight are also helpful for assessing data quality, as is information on whether the statistical distribution of the data corresponds to the distribution of the population as a whole or whether at least all relevant groups are represented sufficiently frequently. This information can be used to draw conclusions about the suitability of an AI procedure for a specific patient. For example, if only patients weighing up to 80 kg were included in the training and the specific patient weighed 130 kg, the application of the AI method could potentially deliver suboptimal results. It is also important to know whether data augmentation methods have been used. These are methods that modify the initial data in order to multiply the amount of training data.

Ideally, the data should be disclosed so that they can be assessed by an independent body. Since the disclosure of all training data is contrary to the commercial interests of the manufacturers, an alternative would be to disclose the data only to a competent testing or certification body (competent body), or to disclose only a small subset of the training data.

5.2 Recognising problems

The fundamental problem with all ML-based interventions in the image generation process, but especially with DL-based interventions, is the question of whether the results generated can be validated in individual cases. This can be traced back to the fundamental question of the trained network's ability to generalise. It must be investigated whether data not used in training ("unseen data") can be processed correctly or whether unrecognised specific characteristics of the training data cause the network to incorrectly supplement them in the application to new data, i.e. generate artefacts in the images. Empirically, this question can best be answered by regularly comparing conventionally generated images with those that have undergone DL-based processing.

Significant possible artefacts are

1. False positive lesions: Conversion of a noise artefact by the network into a lesion perceived as such by the reporting physician.
2. False negative findings: Removal of low-contrast structures interpreted as noise, which could otherwise have been identified by the reporting physician.
3. Signal amplification in the context of resolution recovery. This problem is not specific to DL approaches, but also affects them. For nuclear medicine, this is particularly problematic in view of the quantitative determination of tracer accumulation that is frequently performed here, for example to evaluate the response to therapy in tumour diseases.
4. (Local) systematic bias in image intensity, which influences absolute quantification.

Currently, all manufacturers of CT, SPECT and PET devices, as well as manufacturers specialising in software, are pushing products onto the market that are advertised as "AI-supported" or "AI-assisted". The exact tasks performed by the AI modules are not specified in detail. It is particularly problematic that commercial development will probably lead to a situation where users will not be able to deactivate the corresponding interventions in the image generation process and will no longer be able to easily assess the extent, quality and validity of the image modifications made in individual cases. In the foreseeable future, the problem of verifiability can therefore only be resolved by large clinical validation studies.

5.3 Fundamental problems with the verifiability of AI methods

The testing of DL-based methods presents a fundamental problem: the established conventional methods for assessing image quality and integrity in radiological and nuclear medical imaging are based on measurements with dedicated phantoms of simple geometry that undergo the respective imaging and image processing process. Quantitative parameters for assessing image quality can then be derived from the resulting image data sets using suitable physical evaluation methods. Furthermore, such standardised measurements with invariant phantoms provide the basis for a valid observer-based quantitative assessment of image quality.

A fundamental difficulty is that this approach can only be applied to the verification of AI-based methods with major limitations and reservations, since physical phantoms are always outside the intended scope of application of the method, which was trained with (and for application to) patient data. In this respect, the results obtained with the phantom would not allow any reliable conclusions to be drawn about the expected quality in the application case (neither in a positive nor in a negative sense). Suitable phantom geometries, which reproduce the anatomical conditions as faithfully as possible and yet cannot be recognised 1:1 by the algorithm, are conceivable, but would massively increase the effort involved in every respect. It would also be ideal if the test method could evaluate the actual image acquisition process as well as the DL-based image processing individually.

An alternative strategy, which would first have to be implemented, is the use of complex software phantoms that adequately represent the conditions in real patient data. In any case, however, the problem would be that the use of invariant standardised phantom geometries would in principle offer the developer of the AI method the opportunity to train their method specifically to achieve good results in corresponding phantom-based quality control measurements, which would of course devalue or render them worthless. Variable phantom geometries known only to the tester could help to mitigate this problem, but would further increase the methodological effort. As long as the complex phantom geometries are purely software phantoms without physical representation, testing the entire system as described above is not possible.

One option currently available is to base the verification of AI methods on the use of real patient data. The obvious conceptual difficulty here is that the inherent interindividual variability of these data and the inability to freely select structure and contrast ratios as required for measurement purposes, as is the case with phantoms, leads to new methodological problems in the quantitative and qualitative evaluation of the resulting images. Collections of well-characterised image data sets from examinations of patients that can be subjected to AI-supported processing and allow a well-defined comparison between expected and actual values are therefore likely to be significant for objective evaluation. These collections must not be known to or accessible by the developers of AI systems. In addition, DL-based image processing and reconstruction methods generally produce a very good visual impression, which does not necessarily correlate with diagnostic sensitivity and specificity.

Another difficulty is that standard test methods, particularly in terms of statistical evaluations or the determination of deviations, can only be meaningful to a limited extent, as these methods often do not sufficiently evaluate deviations in subtle structures because greater weight is placed on deviations with higher contrasts or large areas. By definition, these can often be better represented using DL-based methods than with other conventional methods.

In order to enable the objective evaluation of DL-based methods and to discover potential weaknesses in AI algorithms, it is desirable to grant independent approval authorities' access to a representative portion of the verification data.

Ideally, these data should reflect the entire range of planned clinical applications. This approach ensures greater transparency and credibility, as the performance of the system is independently tested and validated.

It is clear that the evaluation of imaging techniques that use AI-based image processing methods poses major challenges and that there is therefore an increased need for research in order to establish the same level of quality assurance for new methods as for existing ones.

6 Quality-assured input data

As described in section 5.1, it is of central importance that the input data for training and evaluating a network are quality assured. This applies regardless of the specific issue under consideration. In principle, all relevant aspects of data generation, including pre-processing, must be taken into account when documenting the data. The quality of the input data used for training and evaluating a neural network is crucial for its objective performance, whereby its correct assessment in turn depends on the quality of the independent test data used. As a side note, it should be mentioned here that the input data must of course, also meet the same quality standards when using AI-based processing. This chapter presents the key aspects of quality assessment in the context of the methods of tomographic image reconstruction, noise reduction and motion correction covered in this recommendation. These are:

– How many independent data sets are used?

  A large number of separate data sets (see below) is a necessary (but not sufficient) prerequisite for the training data as a whole to be considered representative of the range to be covered later in clinical application.

– How well is the typical interindividual (patho)physiological variability of regional signal distribution (global image dynamics, local contrast, texture, etc.) covered?

  This aspect plays a crucial role, as insufficient coverage means that the network is trained to reproduce specifics of the data used for training (which are not representative), which can lead to suboptimal performance or artefacts when applied to other "unexpected" data. For example, care must be taken to consider the extent to which aspects of the given prevalence should be included in the selection of data sets.

– How well is the relevant inter-instrument variability and the variability inherent in the performance of an examination (different tomographs with different performance parameters such as intrinsic resolutions and sensitivities, different acquisition and reconstruction parameters or radiotracer activities, different noise levels, etc.) covered in the event that the methods are to be used flexibly? This means that the more system-independent and application-independent an AI method is to be used, the more different data sets are required for training.

– Is there a qualified ground truth definition by experienced reporting physicians or imaging experts (if *supervised learning* approaches are used)?

No generally applicable specifications can be made with regard to the minimum number of separate training data sets required. For some of the procedures and applications, it can be assumed that the threshold is more likely to be in the range of 10-100 3D image volumes than 1,000 or more. For other methods, for example in the context of screening with low-dose CT and correspondingly low prevalence, a higher number of data sets is likely to be required, depending on the task to be solved by the network.

In order to ensure the use of quality-assured input data, the question or task must first be clearly formulated. The relevant parameters for the task must be identified and specified with regard to diagnostic procedures, imaging parameters and meaningful quality criteria. In the case of *supervised learning* approaches, the creation of the ground truth must be described. The individual aspects must be clearly documented. The documentation must be carried out for all data used for model generation, testing and validation in such a way that the quality of the data can be presented completely at any time. This includes information on the status of AI-based software, patient-related data, acquisition parameters, device parameters, proven quality assurance of the device, software versions, etc.

Standard operating procedures (SOPs) can help to significantly facilitate the quality assurance of the input data. However, even when using SOPs, appropriate documentation must be ensured.

7 Possible test methods for quality assurance

7.1 Variable phantoms

As described in particular in section 5.3, one of the central problems in quality assurance for AI-based image processing methods is that testing, verification and quality assurance are considered very difficult, as an ML-based method may be able to recognise the test phantoms based on the raw or image data provided and thus display images of any desired quality. This is particularly important given that most of the test phantoms used today for quality assurance contain relatively simple structures and are therefore easy to recognise using only a small amount of data.

One approach to testing, examining and subsequently performing quality control on AI-based image processing methods is therefore to create basic phantoms containing complex structures with changing, variable inserts, so that the phantoms to be examined differ only slightly, but in a measurable and relevant way, and the overall structure of the phantom is complex. It is important, of course, that the contents and arrangement of the phantoms and their components are either known precisely from the outset or can be recorded retrospectively by means of a high-dose, high-resolution CT or PET measurement in order to obtain a ground truth and perform genuine quality assurance.

In order to design the complex basic structure, it is advisable to examine image data sets of areas of interest with often critical image quality, save them as files with predefined information and then, for example, produce 3D prints from them. They should contain areas into which variable inserts can then be introduced, e.g. structures for resolution determination, contrast evaluation and noise measurement. The phantoms must be designed in such a way that the system identifies the phantom images as belonging to the overall task and neither develops special algorithms for processing the phantom data nor impairs the processing of patient data. The latter could happen if the system learns to optimally reconstruct or denoise images of the phantoms and this does not meet the same requirements as for patient images. This poses particular challenges for nuclear medical imaging, because typical radionuclide distributions must be simulated and corresponding resolution and contrast test inserts filled with radioactive material must be able to be introduced, while still allowing for the principle of variable phantoms.

7.2 Data specifications – data testing

An integral part of validating trained networks is a careful analysis of the results obtained in independent (external) test data. Such test data help to quantify the scope in which a given network can deliver valid results (as well as the proportion of manifest errors). Ideally, such test runs should be repeated regularly with independent test cohorts and use metrics that are closely related to relevant clinical tasks. In this context, it also seems desirable that well-characterised but otherwise inaccessible test cohorts be made available by the designated independent competent body in order to be able to objectively verify the performance of the method in question as claimed by the manufacturer. In this context, it may be advisable to seek the appropriate involvement of experienced imaging and clinical experts if an assessment of performance cannot be guaranteed by mathematical and technical metrics alone. Furthermore, it would be desirable if findings and annotations for the characterisation of the test and validation data sets were not only created by a single person, but rather reflected a consensus among several experts.

It would also be desirable to have the option of deactivating the network's intervention in the image generation process in order to enable a "before/after" comparison on a case-by-case basis or as part of regular quality checks, as this would enable experienced users to assess the extent of the network's influence on the resulting image data and to make experience-based decisions as to when results should be considered untrustworthy or invalid. However, it can be assumed that, to an increasing extent, the deep integration of DL-based methods into the image generation process by manufacturers will make it difficult or impossible to deactivate or bypass DL-based processing steps.

7.3 Statistical analyses

Statistical analyses are essential for quality assurance in order to objectively evaluate and optimise the results of AI methods and to formulate hypotheses about their usability and robustness. When comparing images calculated using AI methods with reference images, quantitative, image-based metrics can be used, but as explained in section 5.3, these are only of limited significance. Clinical experts can assess the quality of the image impression, which is usually divided into scales. In addition, other relevant characteristics such as radiation exposure or execution time can be quantified. Generative AI methods in particular carry the potential risk of hallucinations that can add or remove information, as described in section 4.2. A distinction should be made between whether, depending on the input data, significant anatomically incorrect changes occur in the image in the event of an error, or whether there are anatomically correct but nevertheless erroneous changes that are more difficult to identify, such as interstitial lung changes, which could potentially influence the diagnosis. Identification and quantification of these image errors by experts is useful for performing a risk analysis.

These considerations lead to recommendation 4 of the SSK in chapter 2:

The SSK recommends that

– the entire process of creating and using AI-based methods for image reconstruction or image processing should be quality assured. In particular, generative networks should always produce a unique output data set from an input data set that is not influenced by chance.

In the development of AI methods, the metrics described can be determined for the entire verification data and serve as a basis for statistical analyses. By calculating parameters such as mean errors, quantiles and standard deviations, probabilities can be calculated under which the AI methods deliver a desired result. In addition, significance tests such as p-tests, t-tests or Wilcoxon tests (Siegel 2001) or Bayesian methods can also be used to determine the superiority of one method over another AI or classical method. The choice of method depends on the distribution of the data. Furthermore, a multivariate analysis can provide information on whether the results depend on certain influencing factors or an underrepresentation of groups within the data. This enables a differentiated evaluation in order to identify and minimise distortions.

In patient care, such analyses can be carried out using representative sample collections or in the context of clinical studies. Depending on the application, this requires that reference data can be acquired simultaneously with reasonable effort.

Another key aspect of quality assurance is the performance of statistical analyses during routine patient care. The aim of such analyses is to continuously monitor the performance of the AI models used and to ensure that they deliver consistent and valid results even under real clinical conditions. Various metrics can be used for this purpose, such as the sensitivity, specificity, accuracy or F1 score of diagnoses (Lang 2023), which are calculated on the basis of validated test data in routine patient care. Clinical parameters such as quality of life or treatment success can also be taken into account. These metrics should be defined with regard to clinical application, regularly reviewed and compared with the results achieved in the model development process in order to identify any deviations at an early stage.

A central component of these analyses is the monitoring of model performance over different time periods. Changes in the underlying data, such as the use of new devices, modified protocols or different patient groups, can influence the performance of the model (so-called "data drift"). This can occur, for example, due to changes in the patient population in terms of age, gender, weight or other factors. It is therefore essential to continuously analyse the distribution of the input data and the distribution of the model results.

In addition, methods for detecting "out-of-distribution" data can be used to ensure that the model is not applied to data that lie outside its originally intended scope of application. Ideally, such cases can be detected automatically and either reported to the user or stored for further analysis and retraining of the model.

In order to translate the results of statistical analyses into practice, a feedback mechanism should be implemented. This could enable clinical staff to report conspicuous or potentially erroneous results to the development teams. In addition, a central database would be helpful in which statistical quality parameters and annotated and validated error cases are collected for use in the continuous improvement of AI models. Furthermore, it would be advisable to repeatedly perform diagnostics using classical approaches (e.g. higher doses) in order to continuously collect evidence for the improved treatment based on AI algorithms.

In summary, statistical analyses are an indispensable tool for quality assurance, as they enable the performance and reliability of AI-based image processing methods to be monitored in practical use, deviations to be identified at an early stage and continuous improvement to be ensured.

7.4 Virtual phantoms and quality assurance methods

Instead of using physical phantoms, which are costly to produce and whose variability is limited at least by the number of possible quality assurance measurements, it can be highly advantageous to generate virtual data and have them processed by AI-based reconstruction or post-processing methods. However, this only helps with quality assurance of the image processing methods, not with quality assurance of the image-generating systems or the interaction between the two components. Preferably, such virtual data are based on patient data, as this is the only way to generate samples that correspond to the actual expected statistical distribution of the data. It is important to have a sufficiently large amount of clinical data. In practice, the amount of patient data can be multiplied by means of so-called data augmentation, i.e. a modification of the data set. For example, an existing CT data set can be manipulated by scaling, shifting, deforming and changing the grey scale, and thus fed into the training as a new data set. Further examples of the generation of additional data sets are:

- Existing image data can be virtually subjected to additional noise – "noise injection" – in order to test AI-based noise reduction methods.
- Existing image data can be artificially blurred to test AI-based super-resolution algorithms.
- Simulations of the scattered radiation component can be performed and this component increased in order to simulate contrast changes; if necessary, the volume of the patient can be virtually adjusted for this purpose.

However, such quality assurance methods require suitable interfaces to the imaging system or its software, which would have to be provided by the manufacturer. Furthermore, these data generally do not provide any means of checking the overall system, but only the AI-based image processing methods. It is also necessary that the data augmentation itself is carried out according to quality-assured principles and that it is ensured that neither nonsensical data are generated nor that the data are not too similar in terms of learning.

7.5 Disclosure of the training, validation and test data distribution

The disclosure of essential characteristics of the training, validation and test data used is a central component of quality assurance. Only through such disclosure is it possible for regulatory authorities or independent bodies to assess the suitability of the underlying data for the intended areas of application of the AI algorithm.

Careful documentation of the data helps to identify possible biases in the data. It should be ensured that the training, validation and test data reflect the entire spectrum of clinical reality. In addition, the documentation of the data must be designed in such a way that data protection regulations are complied with.

In order to enable a sound assessment of the data quality and its suitability for the respective area of application, the following aspects of the training, validation and test data used should be documented:

- How many patients were used for training?
- What was the age, gender, weight and other characteristics of the patients associated with the data?
- Which acquisition and reconstruction protocol was used? Which augmentation techniques were used?
- Was training also carried out on quality assurance phantoms? If so, why?
- Does training on phantoms influence the results of clinical data and vice versa?
- How is it ensured that the training distribution corresponds to the expected patient distribution?
- How was the training conducted (supervised, unsupervised, etc.)?
- How were the necessary data pairs generated (measurement, simulation, etc.)?
- What measures were taken to detect and intercept out-of-distribution (OOD) cases or, if possible and necessary, to integrate them into the model?
- Is there ongoing training of the model used based on new clinical image data? How are changes documented in this case?

Based on this information, special real or virtual phantoms could then be generated for quality assurance purposes in order to apply AI methods to them and carry out quality control. Furthermore, such documentation is necessary to allow a competent body to assess the validity of the metrics and methods used by the developers with regard to the appropriate performance of the AI algorithm for specific clinical applications.

Out-of-distribution cases (OOD cases) refer to input data that lie outside the distribution on which the AI algorithm was trained. These could be, for example, extremely obese patients or patients with rare anomalies that were underrepresented or not included at all in the training data. Detecting such OOD cases is important because AI models are optimised for data from their training distribution. For data that are not part of this distribution, unpredictable results cannot be ruled out, which can pose a risk in clinical practice. OOD detection (e.g. through an analysis of acquisition parameters) is desirable and, for some models, even necessary in order to minimise unpredictable results and ensure the safety and reliability of clinical application.

8 Inclusion in regulations

The recommendations of the SSK on the use of AI methods for image reconstruction and processing of medical images should be incorporated into the relevant regulations to ensure their strict implementation. In particular, the Medical Devices Implementation Act (MPDG 2020), the Medical Devices Operator Ordinance (MPBetreibV 2025) and the Radiation Protection Ordinance (StrlSchV 2018) and subordinate regulations should be addressed here. International regulations must also be observed, of course.

The overriding priority is to ensure that

- no information relevant to the diagnosis is lost in medical images and no image content that could be misinterpreted is regenerated,
- all information necessary to ensure these requirements is available,
- the use of AI procedures is clearly described.

At the same time, however, no unnecessary hurdles should be created for the use of AI methods for optimised imaging techniques while ensuring quality and, in particular, radiation protection. Therefore, when implementing the recommendations in the regulations, it must be clarified:

- How can or must the relevant aspects be incorporated into regulations?
- In which regulations is this best and most effectively possible?

One way to establish regulations for procedures is to standardise them. For classical image reconstruction methods (filtered backprojection or iterative methods) and image processing such as noise reduction (various filters or non-linear methods), it should be noted that these methods have not been and are not currently subject to standardisation.

The SSK considers the standardisation of AI-based procedures for reconstruction or image processing to be even more difficult, as the individual algorithms have different modes of operation, some of which are not comprehensible or mathematically describable, as was made clear in the previous chapters. Furthermore, the standardisation of individual procedures seems hardly possible. Due to the potential for incorrect application in non-evaluated areas of use, the SSK does not believe that standardising individual methods would offer any additional safety.

As an alternative to standardising AI-based methods themselves, the development of suitable standard operating procedures (SOPs) is more relevant for the application of such methods in terms of patient safety and radiation protection. This facilitates the implementation of the recommendations in chapter 2 regarding the development, verification and use of AI-based procedures. However, the SSK considers the standardisation of requirements for training and validation data, or procedures for the development, testing and quality assurance of methods, to be possible and, where appropriate, sensible.

9 Guidance for the development of AI-supported reconstruction and post-processing algorithms in radiology and nuclear medicine

In order for manufacturers to develop and market AI-based methods that are consistent with the above recommendations, the SSK provides the following guidance on how this can be achieved:

1. The task of the AI-based reconstruction or post-processing algorithm must be clearly specified
   - Positive example: "The aim of the AI reconstruction algorithm developed is:
     - to reconstruct diagnostic images from scans with a reduced radiation dose (e.g. 25% reduction) compared to the previous standard, which are not inferior to standard reconstruction in terms of diagnostic quality.
     - to remove/reduce image artefacts in an organ/specific region/entire image due to respiratory movement.
     - to improve the spatial resolution of the reconstructed images."
   - Negative example: "The aim of the AI reconstruction algorithm developed is:
     - to improve image quality (problem: no specification of what better image quality means)
     - to denoise the images (problem: no indication of radiation dose/noise levels in the input and reference data)."
2. A strict separation between training, validation and test data must be maintained. The latter must not have any direct (optimisation of network weights) or indirect (e.g. optimisation of training or network hyperparameters) influence on the development of the AI algorithm, which is submitted to the competent body for approval.
   - Positive example: "From the available data, 80% of the data were randomly selected for training (optimisation of network parameters), 10% for validation (optimisation of network and training hyperparameters) and 10% for final testing."
   - Negative example: "From the available data, 80% of the data were randomly selected for training and 20% for testing." Problem: It remains unclear whether the test data had an influence on the choice of hyperparameters.
3. A detailed characterisation of the training, validation and test data, as well as any data augmentation techniques used, must be developed.
   - Positive example: "Tabular listing of relevant demographic data (age, gender, height, weight, etc.), the prevalence of pathologies and examination indications, the prevalence of relevant artefacts, the radiotracers used and the acquisition time (if relevant: nuclear medicine). The following techniques were used for data augmentation: ..."
4. Detailed documentation of the device and protocol parameters used to record the raw data must be created:
   - Positive example: "Tabular specification of scanner model, acquisition protocol, reconstruction parameters and algorithm."
5. The "class" of the AI approach used must be specified.

   - Positive example: "The network developed is an 'end-to-end' network / 'unrolled variational network' / 'conditioned diffusion model' / 'denoising network' that acts in the image space after reconstruction."
   - Negative example: "The developed network is a neural network." Problem: Such a statement is too unspecific.
6. The higher-level network class used and the order of magnitude of the optimised network weights should be specified.
   - Positive example: "The developed network is a convolutional neural network (CNN) with U-Net architecture/transformer network – fully connected network/FCN with approx. 10^2, 10^3, 10^6 or 10^9 trainable parameters."
   - Negative example: "The AI algorithm developed is a neural network." Problem: Here, too, the information is too vague.
7. The cost function and optimisation algorithm used in the training process to optimise the network weights should be specified.
   - Positive example: "A supervised or unsupervised approach and the mean square error were used as the cost function for network training. The ADAM optimiser with an initial step size of 0.001 was selected for optimising the network parameters."
8. The metrics and methods used to demonstrate the performance of the developed algorithm for specific applications based on the selected test data must be specified.
   - Positive example: "The performance of the network was evaluated based on the sensitivity and specificity of the detection of the following lesions/structures/pathologies: ..."
   - Negative example: "The network was compared with a reference method." Problem: It is not specified how the comparison was carried out.
9. A small but representative portion of the test data, which enables external and independent verification of the performance of the AI algorithm, should be disclosed, at least to an approving or assessing body.
   - Reason: Transparent and comprehensible testing promotes the confidence of users and regulatory authorities in the safety and effectiveness of the AI algorithm.

10 Final considerations

AI-based methods are becoming increasingly important in many areas. Accordingly, there are also many approaches to developing AI-based methods for image reconstruction and noise reduction in medical image data (see chapter 3). In principle, such approaches may be suitable for better utilisation of the recorded or measured (raw) image data. If the data can be used more efficiently, this can enable obtaining better images in diagnostic terms at the same radiation exposure for the patient, or to achieve a reduction in radiation exposure for images of equal diagnostic suitability. Despite the wide range of approaches, products have so far only been used to a very limited extent in routine clinical practice.

Due to the algorithmic processing of image data in AI-based methods, which in most cases is not traceable, and the original information, which often can no longer be validated, as well as the potential possibility of suppression or artificial insertion of image content, diagnostic equivalence to standard methods must be ensured in all cases. This is also a key requirement from the perspective of radiation protection legislation (see chapters 4 and 8).

These aspects give rise to the first three basic recommendations in chapter 2.

In order to develop methods and use them safely in clinical practice – also in terms of radiation protection and patient safety – various requirements must be met. In addition to documenting which models are used and how they are trained, it is particularly important to select data that are suitable for training, testing and validation purposes (see chapters 6 and 7). Standards for the requirements of the various data sets could possibly be defined with the help of international standardisation, for example. Overall, international coordination regarding the framework conditions for the use of AI-based image reconstruction and noise reduction methods in medical images is certainly advisable (see chapter 8).

In addition, suitable methods must be used and, in many cases, developed in the first place in order to evaluate AI-based methods for medical imaging and also for the entire imaging process, including data acquisition, and to check them for quality assurance purposes. In this area in particular, there is a great need for research from a radiation protection perspective (see chapters 4, 5 and 7).

In addition to the verification of systems, the correct application of procedures also plays a central role in quality assurance in terms of radiation protection. In this context, the understanding of the methods by medical users and standard operating procedures are of central importance (see chapters 4 and 8). However, the SSK considers the standardisation of individual methods as such to be less useful.

Developments in other areas of AI-assisted image processing in medicine should be closely monitored in order to examine the implications for radiation protection-specific applications.

In summary, the use of AI-based methods offers many opportunities to optimise image-guided patient care in terms of diagnostic reliability and reduced radiation exposure, provided that the recommendations and statements set out in this SSK recommendation are implemented. To facilitate this in practice, recommendations and guidance on the development, testing and verification of AI-based methods have been provided for developers, manufacturers and competent bodies (see chapters 2 and 9).

11 References

Adler and Oktem 2018	Adler J, Oktem O. Learned Primal-Dual Reconstruction. IEEE Trans Med Imaging. 2018 Jun;37(6):1322-32, doi: 10.1109/TMI.2018.2799231
Baader and Kachelrieß 2025	Baader E, Kachelrieß M. Risk-minimising tube current and tube voltage modulation for CT: A simulation study. Med Phys. 2025 Aug;52(8):e18047, doi: 10.1002/mp.18047
Balogh and Janos Kis 2022	Balogh ZA, Janos Kis B. Comparison of CT noise reduction performances with deep learning-based, conventional, and combined denoising algorithms. Med Eng Phys. 2022 Nov;109:103897, doi: 10.1016/j.medengphy.2022.103897, Epub 20220924
Berker et al. 2018	Berker Y, Maier J, Kachelrieß M. Deep Scatter Estimation in PET: Fast Scatter Correction Using a Convolutional Neural Network. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), 10-17 Nov. 2018. p. 1-5
Byl et al. 2021	Byl A, Klein L, Sawall S, Heinze S, Schlemmer HP, Kachelrieß M. Photon-counting normalised metal artefact reduction (NMAR) in diagnostic CT. Med Phys. 2021 Jul;48(7):3572-82, doi: 10.1002/mp.14931, Epub 20210528
Canon Medical Systems n.d.	Canon Medical Systems. Precise IQ Engine (PIQE): A New Concept in Clarity and Confidence in Cardiac Imaging. https://global.medical.canon/publication/ct/V38_CTEU220164, last accessed on 12 November 2025
Chin et al. 2022	Chin M, Chinn G, Innes D, Levin CS. Deep Learning-Based Limited Angle Tomography for a 1-Millimetre Resolution Dual-Panel Clinical PET System. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 5-12 Nov. 2022. p. 1-3
Chung et al. 2022	Chung H, Kim J, McCann MT, Klasky ML, Ye CJ. Diffusion Posterior Sampling for General Noisy Inverse Problems. 2022, doi: 10.48550/arXiv.2209.14687
Duan et al. 2024	Duan L, Eulig E, Knaup M, Adamus R, Lell M, Kachelriess M. Training of a deep learning based digital subtraction angiography method using synthetic data. Med Phys. 2024 Jul;51(7):4793-810, doi: 10.1002/mp.16973, Epub 20240214
Ebner and Haltmeier 2024	Ebner A, Haltmeier M. Plug-and-Play Image Reconstruction Is a Convergent Regularisation Method. IEEE Trans Image Process. 2024;33:1476-86, doi: 10.1109/TIP.2024.3361218, Epub 20240221
EU 2016	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Official Journal L 119 of 4 May 2016, p. 1, corr. Official Journal L 314 of 22 November 2016, p. 72, Official Journal L 127 of 23 May 2018, p. 2 (GDPR)

EU 2024	European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144, and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Regulation on artificial intelligence). http://data.europa.eu/eli/reg/2024/1689/oj, last accessed on 05.11.2025
Fonseca et al. 2021	Fonseca GP, Baer-Beck M, Fournie E, Hofmann C, Rinaldi I, Ollers MC, van Elmpt WJC, Verhaegen F. Evaluation of novel AI-based extended field-of-view CT reconstructions. Med Phys. 2021 Jul;48(7):3583-94, doi: 10.1002/mp.14937, Epub 20210531
GE Healthcare n.d.	GE Healthcare. Revolution Maxima. https://www.gehealthcare.de/products/computed-tomography/revolution-family/revolution-maxima, last accessed on 12 November 2025
Gottschling 2022	Gottschling N. On existence, stability, accuracy and learning of approximate decoders for ill-posed inverse problems. doi: 10.17863/CAM.91560
Häggström et al. 2019	Häggström I, Schmidtlein CR, Campanella G, Fuchs TJ. DeepPET: A deep encoder-decoder network for directly solving the PET image reconstruction inverse problem. Med Image Anal. 2019 May;54:253-62, doi: 10.1016/j.media.2019.03.013, Epub 20190330
Hammernik et al. 2018	Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, Knoll F. Learning a variational network for reconstruction of accelerated MRI data. Magn Reson Med. 2018 Jun;79(6):3055-71, doi: 10.1002/mrm.26977, Epub 20171108
Han et al. 2017	Han Y, Gu J, Ye JC. Deep learning interior tomography for region-of-interest reconstruction. arXiv. 2017, doi: 10.48550/arXiv.1712.10248
Huang et al. 2018	Huang Y, W\ rfl T, Breininger K, Liu L, Lauritsch Gn, Maier A. Some Investigations on Robustness of Deep Learning in Limited Angle Tomography. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018: Springer-Verlag. p. 145–53
Isensee et al. 2021	Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021 Feb;18(2):203-11, doi: 10.1038/s41592-020-01008-z, Epub 20201207
Kabelac et al. 2025	Kabelac A, Eulig E, Maier J, Hammermann M, Knaup M, Kachelrieß M. Latent space reconstruction for missing data problems in CT. Med Phys. 2025 Jul;52(7):e17910, doi: 10.1002/mp.17910, Epub 20250604
Ketola et al. 2021	Ketola JHJ, Heino H, Juntunen MAK, Nieminen MT, Siltanen S, Inkinen SI. Generative adversarial networks improve interior computed tomography angiography reconstruction. Biomed Phys Eng Express. 2021 Oct 29;7(6), doi: 10.1088/2057-1976/ac31cb, Epub 20211029

Klein et al. 2022	Klein L, Liu C, Steidel J, Enzmann L, Knaup M, Sawall S, Maier A, Lell M, Maier J, Kachelrieß M. Patient-specific radiation risk-based tube current modulation for diagnostic CT. Med Phys. 2022 Jul;49(7):4391-403, doi: 10.1002/mp.15673, Epub 20220429
Kofler et al. 2020	Kofler A, Haltmeier M, Schaeffter T, Kachelriess M, Dewey M, Wald C, Kolbitsch C. Neural networks-based regularisation for large-scale medical image reconstruction. Phys Med Biol. 2020 Jul 6;65(13):135003, doi: 10.1088/1361-6560/ab990e, Epub 20200706
Lambin et al. 2012	Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012 Mar;48(4):441-6, doi: 10.1016/j.ejca.2011.11.036, Epub 20120116
Lang 2023	Lang N. F1 score – explained simply! 9 August 2023 / Statistics. https://databasecamp.de/statistik/f1-score, last accessed on 12 November 2025
Laurent et al. 2023	Laurent B, Bousse A, Merlin T, Nekolla S, Visvikis D. PET scatter estimation using deep learning U-Net architecture. Phys Med Biol. 10 March 2023;68(6), doi: 10.1088/1361-6560/ac9a97, Epub 20230310
Magonov et al. 2024	Magonov J, Maier J, Erath J, Sunnegardh J, Fournie E, Stierstorfer K, Kachelriess M. Reducing windmill artefacts in clinical spiral CT using a deep learning-based projection raw data upsampling: Method and robustness evaluation. Med Phys. 2024 Mar;51(3):1597-616, doi: 10.1002/mp.16938, Epub 20240116
Mahmoodian et al. 2022	Mahmoodian N, Thadesar M, Sadeghi M, Georgiades M, Pech M, Hoeschen C. Liver tumour segmentation using deep learning method: RESLU-NET. Physica Medica. 2022 12/01;104:S132, doi: 10.1016/S1120-1797(22)02430-9
Maier et al. 2022a	Maier A, Köstler H, Heisig M, Krauss P, Yang SH. Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future. Progress in Biomedical Engineering. 2022a;4(2), doi: 10.1088/2516-1091/ac5b13
Maier et al. 2019a	Maier AK, Syben C, Stimpel B, Würfl T, Hoffmann M, Schebesch F, Fu W, Mill L, Kling L, Christiansen S. Learning with Known Operators reduces Maximum Training Error Bounds. Nat Mach Intell. 2019a Aug;1(8):373-80, doi: 10.1038/s42256-019-0077-5, Epub 20190809
Maier et al. 2018	Maier J, Sawall S, Knaup M, Kachelrieß M. Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-Ray CT Using a Deep Convolutional Neural Network. J Nondestruct Eval. 2018 2018/07/10;37(3):57, doi: 10.1007/s10921-018-0507-z
Maier et al. 2019b	Maier J, Eulig E, Vöth T, Knaup M, Kuntz J, Sawall S, Kachelrieß M. Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. Med Phys. 2019b Jan;46(1):238-49, doi: 10.1002/mp.13274, Epub 20181126

Maier et al. 2021	Maier J, Lebedev S, Erath J, Eulig E, Sawall S, Fournié E, Stierstorfer K, Lell M, Kachelrieß M. Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac CT. Med Phys. 2021 Jul;48(7):3559-71, doi: 10.1002/mp.14927, Epub 20210526
Maier et al. 2022b	Maier J, Klein L, Eulig E, Sawall S, Kachelrieß M. Real-time estimation of patient-specific dose distributions for medical CT using deep dose estimation. Med Phys. 2022b Apr;49(4):2259-69, doi: 10.1002/mp.15488, Epub 20220214
Maier et al. 2025	Maier J, Sawall S, Arheit M, Paysan P, Kachelrieß M. Deep learning-based cone-beam CT motion compensation with single-view temporal resolution. Med Phys. 2025 Jul;52(7):e17911, doi: 10.1002/mp.17911, Epub 20250604
Martins et al. 2023	Martins JC, Maier J, Gianoli C, Neppl S, Dedes G, Alhazmi A, Veloza S, Reiner M, Belka C, Kachelrieß M, Parodi K. Towards real-time EPID-based 3D in vivo dosimetry for IMRT with Deep Neural Networks: A feasibility study. Phys Med. 2023 Oct;114:103148, doi: 10.1016/j.ejmp.2023.103148, Epub 20231004
Mehranian et al. 2021	Mehranian A, Wollenweber SD, Walker MD, Bradley KM, Fielding PA, Su K-H, Johnsen R, Kotasidis F, Jansen FP, McGowan DR. Image enhancement of whole-body oncology [18F]-FDG PET scans using deep neural networks to reduce noise. Eur J Nucl Med Mol Imaging. 2021;49(2):539-49, doi: 10.1007/s00259-021-05478-x
Mohri et al. 2012	Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning. The MIT Press, 2012, ISBN 9780262018258
MPBetreibV 2025	Medical Devices Operator Ordinance of 14 February 2025 (Federal Law Gazette 2025 I No. 38), last amended by Article 1 of the Ordinance of 31 October 2025 (Federal Law Gazette 2025 I No. 263)
MPDG 2020	Medical Devices Implementation Act of 28 April 2020 (Federal Law Gazette I p. 960), last amended by Article 3 of the Act of 23 October 2024 (Federal Law Gazette 2024 I No. 324)
Muckley et al. 2021	Muckley MJ, Riemenschneider B, Radmanesh A, Kim S, Jeong G, Ko J, Jun Y, Shin H, Hwang D, Mostapha M, Arberet S, Nickel D, Ramzi Z, Ciuciu P, Starck JL, Teuwen J, Karkalousos D, Zhang C, Sriram A, Huang Z, Yakubova N, Lui YW, Knoll F. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. IEEE Trans Med Imaging. 2021 Sep;40(9):2306-17, doi: 10.1109/TMI.2021.3075856, Epub 20210831
Nikolaus 2018	Nikolaus K. Improved CT scans – thanks to artificial intelligence. Siemens AG. https://www.siemens.com/de/de/unternehmen/stories/forschung-technologien/kuenstliche-intelligenz/verbesserte-ct-scans-dank-kuenstlicher-intelligenz.html, last accessed on 12 November 2025
Omega Medical Imaging 2023	Omega Medical Imaging. Innovations in Interventional Imaging. https://www.omegamedicalimaging.com/innovations-in-interventional-imaging/, last accessed on 12 November 2025

Pashazadeh and Hoeschen 2023 Pashazadeh A, Hoeschen C. Possibilities of artificial intelligence in radiation protection. Radiology. 2023 Jul;63(7):530-8, doi: 10.1007/s00117-023-01167-y, Epub 20230622

Philips Austria GmbH n.d. Philips Austria GmbH. MRCAT Brain. Clinical MR-RT application. https://www.philips.at/healthcare/product/HCNMRF320/mrcat-brain-klinische-mr-rt-anwendung, last accessed on 12 November 2025

Philips GmbH 2021 Philips GmbH. Philips Incisive CT Precise Suite. KMA-Online, Georg Thieme Verlag KG. https://www.kma-online.de/unternehmen/philips/produkte/detail/philips-incisive-ct-precise-suite-82, last accessed on 12 November 2025

Qi 2006 Qi J. Comparison of Lesion Detection and Quantification in MAP Reconstruction with Gaussian and Non-Gaussian Priors. Int J Biomed Imaging. 2006;2006:87567, doi: 10.1155/IJBI/2006/87567, Epub 20060629

Ronneberger et al. 2015 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015, doi: 10.48550/arXiv.1505.04597

Schramm et al. 2021 Schramm G, Rigie D, Vahle T, Rezaei A, Van Laere K, Shepherd T, Nuyts J, Boada F. Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network. Neuroimage. 1 January 2021;224:117399, doi: 10.1016/j.neuroimage.2020.117399, Epub 20200921

Sidky and Pan 2022 Sidky EY, Pan X. Report on the AAPM deep-learning sparse-view CT grand challenge. Med Phys. 2022 Aug;49(8):4935-43, doi: 10.1002/mp.15489, Epub 20220209

Siegel 2001 Siegel S. Nonparametric statistical methods. Klotz, 2001, ISBN 9783880741027

Siemens Healthineers AG n.d.-a Siemens Healthineers AG, (Siemens Healthineers AG). YSIO X.pree. Intelligence for excellence. https://www.siemens-healthineers.com/de/radiography/digital-x-ray/ysio-xpree, last accessed on 12 November 2025

Siemens Healthineers AG n.d.-b Siemens Healthineers AG. MR-only Radiotherapy Planning. Powered by syngo.via RT Image Suite. https://www.siemens-healthineers.com/magnetic-resonance-imaging/clinical-specialities/synthetic-ct, last accessed on 12 November 2025

Singh et al. 2023 Singh IRD, Denker A, Jin B, Thielemans K, Arridge S. Investigating Intensity Normalisation for PET Reconstruction with Supervised Deep Learning. 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD), 4-11 Nov. 2023

Spectronic Medical AB n.d. Spectronic Medical AB. MRI only radiotherapy planning. https://medical.spectronic.se/page-2/index.html, last accessed on 12 November 2025

SSK 2016 Radiation Protection Commission. Dose documentation and archiving of digital image and examination data in radiology and nuclear medicine, adopted at the 282nd meeting of the Radiation Protection Commission on 23/24 June 2016. urn:nbn:de:101:1-201703087447. Announcement in BAnz AT 20 March 2017 B3

StrlSchG 2017	Act on Protection against the Harmful Effects of Ionising Radiation (Radiation Protection Act – StrlSchG) of 27 June 2017 (Federal Law Gazette I p. 1966), last amended by the announcement of 3 January 2022 (Federal Law Gazette I p. 15)
StrlSchV 2018	Ordinance on Protection against the Harmful Effects of Ionising Radiation (Radiation Protection Ordinance – StrlSchV) of 29 November 2018. Federal Law Gazette I p. 2034, 2036, amended
Therasse et al. 2006	Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: a review of validation studies on tumour assessment. Eur J Cancer. 2006 May;42(8):1031-9, doi: 10.1016/j.ejca.2006.01.026, Epub 20060417
Trapp et al. 2022	Trapp P, Amato C, Sawall S, Kachelrieß M, Vöth T. DeepRAR: a CNN-based approach for CT and CBCT ring artifact reduction. SPIE Medical Imaging, 2022: SPIE. p. 337-44
Würfl et al. 2016	Würfl T, Ghesu FC, Christlein V, Maier A Eds. Medical Image Computing and Computer-Assisted Intervention - MICCAI. 2016 Cham
Xiang et al. 2020	Xiang H, Lim H, Fessler JA, Dewaraja YK. A deep neural network for fast and accurate scatter estimation in quantitative SPECT/CT under challenging scatter conditions. Eur J Nucl Med Mol Imaging. 2020 Dec;47(13):2956-67, doi: 10.1007/s00259-020-04840-9, Epub 20200515
Zhu et al. 2018	Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 21 March 2018;555(7697):487-92, doi: 10.1038/nature25988

Definitions

Adversarial attacks	Targeted manipulations of input data that may be barely visible to humans but cause an AI model to produce extremely incorrect results.
Annotation	The manual or semi-automatic marking of image data (e.g. outline of a lesion, noise structures or heavily noisy images, etc.) so that it can be used for training AI methods.
Artificial intelligence (AI)	A field of computer science that deals with the development of machines that exhibit human-like intelligence.
Autoencoder	A neural network that converts data into a compressed representation and reconstructs it from there. Useful, for example, for noise reduction or feature extraction.
Backpropagation algorithm	The standard method for training neural networks, in which errors between the network output and the training label are propagated backwards through the network to adjust the weights.
Bayesian learning	A method that explicitly takes probabilities and uncertainties into account in the model, thus also enabling statements to be made about the reliability of predictions.
Black-box model	A model whose internal structure and parameters are not fully known or comprehensible. Its mode of operation can only be deduced from the input-output relationships. These may be secret functions or, as in the case of neural networks, functions with an extremely large number of parameters (e.g. millions to billions of parameters) which, although not secret, are impossible to understand due to their sheer number.
Convolutional neural network (CNN)	A special type of neural network that uses convolution operations. Particularly suitable for image data, as neighbourhood relationships are preserved. 2D and 3D CNNs often play an important role in image reconstruction.
Data augmentation	Methods for artificially expanding data sets (e.g. by rotating, mirroring or adding noise to images) in order to increase the robustness of AI processes. Augmentation can increase the amount of training data many times over.
Deep learning (DL)	Deep learning refers to machine learning (see there) using large amounts of data.
Deep scatter estimation	An AI method for extremely fast and accurate estimation of scattered radiation in CT and PET data, based on deep neural networks. The gold standard for scattered radiation calculation would be the Monte Carlo method or the Boltzmann transport method. However, both are very computationally intensive and therefore require significantly longer computing times.

End-to-end learning	An approach in which a neural network converts raw data directly into a result (e.g. an image) without explicitly involving physical or mathematical models. The opposite of end-to-end learning would be a modular or component-based approach in which the entire processing process is not mapped by a single network but is broken down into several successive steps that are trained or optimised individually.
Forward projection	The computational simulation of projections from an image or volume (image stack) based on the physical properties of the imaging technique. The forward projection of a CT volume produces a projection that looks like an X-ray image.
Foundation model	A foundation model is a large machine learning model that has been pre-trained on a broad database and, thanks to its versatility, serves as the basis for a variety of specific applications.
Generative networks	Neural networks that are trained to generate new data that is as similar as possible to the training data. In imaging, for example, they can generate realistic-looking but artificially generated images or image segments. Generative networks almost always have a stochastic component – i.e. a random number or random vector as input. As a result, such generative networks with stochastic components do not always deliver the same deterministic result, but can generate many different, yet "realistic-looking" variants from the same data distribution.
Gradient descent	An optimisation method that iteratively changes parameters in the direction of the steepest descent of an error function in order to minimise it. Backpropagation (see above) is the technique used to calculate the gradients.
Grey-box model	A hybrid model in which the structure is known, but not all parameters. Prior knowledge is used to reduce the number of parameters to be learned. See also black-box model.
Ground truth	Reference values or images that are assumed to be "true" and are used for comparison with the results of AI methods, i.e. for training. In this case, the ground truth serves as a label (see below) for training.
Hyperparameters	Hyperparameters refer to parameters of an AI model that are used to control the training algorithm. Their values must be defined before the model is trained.
Inpainting method	The inpainting method refers to adding destroyed, damaged or missing image sections or projection data to produce a revised image that is as free of artefacts as possible.

Label	The reference information associated with an input data point that serves as a target in supervised learning. Labels can be, for example, class affiliations (tumour present/absent), image markings (segmentations) or reference images (e.g. those with very little noise or few artefacts). They are typically created through annotation.
Layer	Processing layer in the neural network (see there)
Machine learning (ML)	Machine learning (ML) is an important subfield of AI (see there) that enables machines to learn from experience, recognise patterns in data sets and thus automate tasks that would normally require human intelligence.
Maximum a posteriori reconstruction (MAP)	A reconstruction method that incorporates prior knowledge ("priors") into the calculation in order to achieve more robust results with noisy or incomplete data.
Monte Carlo simulation	A method that uses random numbers to simulate physical processes. It is often used in medical physics, for example, to calculate dose or scattered radiation distributions.
Neural network	Any number of interconnected neurons; in artificial neural networks, digital neurons whose tasks and connections are inspired by biological neurons. A neural network consists of many processing layers. Each layer processes the data non-linearly and passes it on to the next layer. After typically several dozen to hundreds of layers, the neural network outputs the processed data, the processed image or the processed image stack.
Out-of-distribution	Refers to data that differs significantly from the training data. Models often deliver uncertain or incorrect results in this case. Out-of-distribution cases can be reduced, but not completely eliminated, by using diverse and representative training data and appropriate testing strategies. There are also mechanisms for detecting and displaying such uncertainties.
Physics-driven learning	A method that integrates prior physical knowledge (e.g. image formation processes) into the learning process in order to increase the robustness and plausibility of the results.
Poisson thinning	A method for stochastically generating correct subsets from complete measurement data that correspond to a reduced count rate, for example in the case of shorter measurement times or lower doses.
Population	The totality of all objects or persons about which statements are to be made in a study. Samples for training, validation and testing should be taken from this at random.
Post-reconstruction learning	AI methods applied to already reconstructed images to improve their quality (e.g. noise reduction, artefact reduction).

Prior	Prior knowledge or an assumption that is incorporated into a model before data analysis, e.g. smoothness or homogeneity of certain structures.
Radiomics	A mostly AI-based method in which local image properties of findings, such as signal intensity, texture or contour, are used to derive correlations with the diagnosis of lesion type (e.g. benign or malignant lesion).
Regularisation	A general approach to stabilising models and avoiding overfitting. Additional knowledge or a side condition is incorporated into the optimisation, e.g. smoothness. Regularisation replaces missing information with prior knowledge, thereby stabilising the solution selection.
Regularisation step	A measure in training that prevents overfitting by reducing or smoothing model complexity.
Semi-supervised learning	A hybrid form in which both annotated and unannotated data are used for training.
Supervised learning	Supervised learning is a learning method in which input data with known target values (labels, see there) are trained. The network learns to understand the assignment of input to output.
Total-body PET	A PET scanner with a large axial field of view (often 1 m to 2 m) that can capture a large anatomical area or even the entire body at once, enabling particularly high sensitivity.
Transfer learning	An approach in which a pre-trained model is adapted to a new, similar task in order to require less training data and computing time.
U-Net	A special architecture of CNNs that processes input data in an encoder-decoder structure, with the encoder and decoder layers connected by direct connections, known as skip connections. The encoder processes the input data and generates a vector that contains the relevant information in a compressed form. The decoder uses this information, can filter it and then generate an output from the data. Originally developed for segmentation, it is now also used in many image enhancement tasks.
Unsupervised learning	Unsupervised learning is a learning method that does not require labels (see there). The network independently recognises structures or patterns in the data.
White-box model	A model with completely known structures and parameters, often physically sound and easily comprehensible.
Wilcoxon test	A non-parametric statistical test for comparing two related samples without assuming a normal distribution.

List of abbreviations

AI	Artificial intelligence
AUTOMAP	Automated transform by manifold Approximation (by Zhu et al. 2018)
BMUKN	Federal Ministry for the Environment, Climate Protection, Nature Conservation and Nuclear Safety, formerly BMUV: Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection
CBCT	Cone beam computed tomography
cGAN	Conditional generative adversarial network
CNN	Convolutional neural network
CPU	Central processing unit
CT	Computed tomography
CTDI	Computed tomography dose index
DL	Deep learning
DLP	Dose-length product
DNN	Deep neural network
DSA	Digital subtraction angiography
DSE	Deep scatter estimation
DVF	Displacement vector fields
DVT	Digital volume tomography
FBP	Filtered back projection
GDPR	General Data Protection Regulation
GPGPU	General-purpose computation on graphics processing units
GPU	Graphics processing unit
HU	Hounsfield unit
MAP	Maximum a posteriori reconstruction
MIP	Maximum intensity projection
ML	Machine learning
MLEM	Determination of the maximum conditional probability of occurrence for a found result (maximum likelihood expectation maximisation)
MRI	Magnetic resonance imaging
OOD	Out-of-distribution
PET	Positron emission tomography

RECIST	Response Evaluation Criteria In Solid Tumors
SOP	Standard operating procedure
SPECT	Single-photon emission computed tomography
SSDE	Size-specific dose estimation
SSK	German Commission on Radiological Protection
StrlSchG	Radiation Protection Act
StrlSchV	Radiation Protection Ordinance
TCM	Tube current modulation
U-Net	CNN for fast and precise image segmentation
Voxel	3D image element (volume picture element)