IAEA
International Atomic Energy Agency

Artificial Intelligence in Medical Physics

Roles, Responsibilities, Education and Training of Clinically Qualified Medical Physicists

Endorsed by the American Association of Physicists in Medicine

VIENNA, 2023

TRAINING COURSE SERIES 83

ARTIFICIAL INTELLIGENCE IN MEDICAL PHYSICS

The following States are Members of the International Atomic Energy Agency:

AFGHANISTAN
ALBANIA
ALGERIA
ANGOLA
ANTIGUA AND BARBUDA
ARGENTINA
ARMENIA
AUSTRALIA
AUSTRIA
AZERBAIJAN
BAHAMAS
BAHRAIN
BANGLADESH
BARBADOS
BELARUS
BELGIUM
BELIZE
BENIN
BOLIVIA, PLURINATIONAL STATE OF
BOSNIA AND HERZEGOVINA
BOTSWANA
BRAZIL
BRUNEI DARUSSALAM
BULGARIA
BURKINA FASO
BURUNDI
CABO VERDE
CAMBODIA
CAMEROON
CANADA
CENTRAL AFRICAN REPUBLIC
CHAD
CHILE
CHINA
COLOMBIA
COMOROS
CONGO
COSTA RICA
CÔTE D'IVOIRE
CROATIA
CUBA
CYPRUS
CZECH REPUBLIC
DEMOCRATIC REPUBLIC OF THE CONGO
DENMARK
DJIBOUTI
DOMINICA
DOMINICAN REPUBLIC
ECUADOR
EGYPT
EL SALVADOR
ERITREA
ESTONIA
ESWATINI
ETHIOPIA
FIJI
FINLAND
FRANCE
GABON
GAMBIA
GEORGIA
GERMANY
GHANA
GREECE
GRENADA
GUATEMALA
GUYANA
HAITI
HOLY SEE
HONDURAS
HUNGARY
ICELAND
INDIA
INDONESIA
IRAN, ISLAMIC REPUBLIC OF
IRAQ
IRELAND
ISRAEL
ITALY
JAMAICA
JAPAN
JORDAN
KAZAKHSTAN
KENYA
KOREA, REPUBLIC OF
KUWAIT
KYRGYZSTAN
LAO PEOPLE'S DEMOCRATIC REPUBLIC
LATVIA
LEBANON
LESOTHO
LIBERIA
LIBYA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MADAGASCAR
MALAWI
MALAYSIA
MALI
MALTA
MARSHALL ISLANDS
MAURITANIA
MAURITIUS
MEXICO
MONACO
MONGOLIA
MONTENEGRO
MOROCCO
MOZAMBIQUE
MYANMAR
NAMIBIA
NEPAL
NETHERLANDS
NEW ZEALAND
NICARAGUA
NIGER
NIGERIA
NORTH MACEDONIA
NORWAY
OMAN
PAKISTAN
PALAU
PANAMA
PAPUA NEW GUINEA
PARAGUAY
PERU
PHILIPPINES
POLAND
PORTUGAL
QATAR
REPUBLIC OF MOLDOVA
ROMANIA
RUSSIAN FEDERATION
RWANDA
SAINT KITTS AND NEVIS
SAINT LUCIA
SAINT VINCENT AND THE GRENADINES
SAMOA
SAN MARINO
SAUDI ARABIA
SENEGAL
SERBIA
SEYCHELLES
SIERRA LEONE
SINGAPORE
SLOVAKIA
SLOVENIA
SOUTH AFRICA
SPAIN
SRI LANKA
SUDAN
SWEDEN
SWITZERLAND
SYRIAN ARAB REPUBLIC
TAJIKISTAN
THAILAND
TOGO
TONGA
TRINIDAD AND TOBAGO
TUNISIA
TÜRKİYE
TURKMENISTAN
UGANDA
UKRAINE
UNITED ARAB EMIRATES
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
UNITED REPUBLIC OF TANZANIA
UNITED STATES OF AMERICA
URUGUAY
UZBEKISTAN
VANUATU
VENEZUELA, BOLIVARIAN REPUBLIC OF
VIET NAM
YEMEN
ZAMBIA
ZIMBABWE

The Agency's Statute was approved on 23 October 1956 by the Conference on the Statute of the IAEA held at United Nations Headquarters, New York; it entered into force on 29 July 1957. The Headquarters of the Agency are situated in Vienna. Its principal objective is "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world''.

TRAINING COURSE SERIES No. 83

ARTIFICIAL INTELLIGENCE IN MEDICAL PHYSICS

ROLES, RESPONSIBILITIES, EDUCATION AND TRAINING OF CLINICALLY QUALIFIED MEDICAL PHYSICISTS

ENDORSED BY THE AMERICAN
ASSOCIATION OF PHYSICISTS IN MEDICINE

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA, 2023

COPYRIGHT NOTICE

All IAEA scientific and technical publications are protected by the terms of the Universal Copyright Convention as adopted in 1952 (Berne) and as revised in 1972 (Paris). The copyright has since been extended by the World Intellectual Property Organization (Geneva) to include electronic and virtual intellectual property. Permission to use whole or parts of texts contained in IAEA publications in printed or electronic form must be obtained and is usually subject to royalty agreements. Proposals for non-commercial reproductions and translations are welcomed and considered on a case-by-case basis. Enquiries should be addressed to the IAEA Publishing Section at:

Marketing and Sales Unit, Publishing Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
fax: +43 1 26007 22529
tel.: +43 1 2600 22417
email: sales.publications@iaea.org
www.iaea.org/publications

For further information on this publication, please contact:

Dosimetry and Medical Radiation Physics Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
Email: Official.Mail@iaea.org

ARTIFICIAL INTELLIGENCE IN MEDICAL PHYSICS
IAEA, VIENNA, 2023
IAEA-TCS-83
ISSN 1018–5518

© IAEA, 2023

Printed by the IAEA in Austria
November 2023

FOREWORD

The deployment of technologies based on artificial intelligence (AI) in the medical use of radiation (i.e. radiation oncology, diagnostic imaging and nuclear medicine) and medical physics is expected to grow over the next 5–10 years. Clinically qualified medical physicists are expected to play an important role in ensuring safe and effective clinical implementation of AI based tools.

The use of AI in radiation and medical physics was a topic of discussion at the Technical Meeting on Artificial Intelligence for Nuclear Technology and Applications hosted by the IAEA in 2021. Experts in working groups dealing with human health applications identified opportunities and discussed potential challenges of AI. The experts identified a need for guidelines addressing the use of AI in the field of medical physics.

The 19th Biennial Meeting of the SSDL Scientific Committee (SSC-19) for the Evaluation of and Recommendations on the Dosimetry Programme and the IAEA/WHO SSDL Network identified the need for a new curriculum to update the knowledge and academic education of medical radiation physicists in the areas of data science, regression analysis, statistical learning and deep learning.

To address these needs, in 2021 the IAEA organized a consultancy meeting to develop an IAEA Training Course Series publication to supplement IAEA Training Course Series No. 56 (Rev. 1), Postgraduate Medical Physics Academic Programmes. The present publication contextualizes the roles and responsibilities of clinically qualified medical physicists in the framework of the application of AI in radiation medicine. Continuing professional development activities are also suggested.

The American Association of Physicists in Medicine has endorsed this publication. The IAEA officers responsible for this publication were M. Carrara and E. Titovich of the Division of Human Health.

EDITORIAL NOTE

This publication has been prepared from the original material as submitted by the contributors and has not been edited by the editorial staff of the IAEA. The views expressed remain the responsibility of the contributors and do not necessarily reflect those of the IAEA or the governments of its Member States.

Guidance and recommendations provided here in relation to identified good practices represent expert opinion but are not made on the basis of a consensus of all Member States.

Neither the IAEA nor its Member States assume any responsibility for consequences which may arise from the use of this publication. This publication does not address questions of responsibility, legal or otherwise, for acts or omissions on the part of any person.

The use of particular designations of countries or territories does not imply any judgement by the publisher, the IAEA, as to the legal status of such countries or territories, of their authorities and institutions or of the delimitation of their boundaries.

The mention of names of specific companies or products (whether or not indicated as registered) does not imply any intention to infringe proprietary rights, nor should it be construed as an endorsement or recommendation on the part of the IAEA.

The authors are responsible for having obtained the necessary permission for the IAEA to reproduce, translate or use material from sources already protected by copyrights.

The IAEA has no responsibility for the persistence or accuracy of URLs for external or third party Internet web sites referred to in this publication and does not guarantee that any content on such web sites is, or will remain, accurate or appropriate.

CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

Artificial intelligence (AI) is a term covering a wide spectrum of technologies aiming at giving machines or computers the ability to perform human-like cognitive functions such as learning, problem-solving and decision making [1].

The term AI was first coined to define a scientific field during an academic event at Dartmouth College in 1956 [2]. After the first wave of interest in AI in the 50s and 60s, periods of slowing down called "AI winters" were noted and, these were followed by new waves of interest. The use of AI in healthcare began in the 1980s and 1990s when the availability of personal computers became more widespread, and this coincided with the development of new popular machine learning (ML) algorithms such as Bayesian networks and support vector machines [3]. Specific deep learning (DL) models called Convolutional Neural Networks (CNNs) were first used in medical imaging in the early 1990's [4].

At the same time, the digitization of medical data made it possible to analyse patient records using computer algorithms. The advent of Graphic Processing Units (GPUs) and cloud computers made it easier to access the large computational power needed to apply ML to large datasets. In the 2010s, the remarkable advances in DL approaches and the concept of radiomics, where features describing tumour phenotypes in imaging were analysed to characterize lesions or predict patient outcomes using ML, contributed to the widespread investigation on AI to imaging [5, 6]. Since 2012 results achieved by AlexNet, in the annual ImageNet large scale challenge for image classification, the use of CNNs increased dramatically for image recognition tasks [7].

In the last few years, there has been an unprecedented interest in the use of AI in healthcare [8], and it is commonly recognised that AI will transform healthcare processes [9]. Numerous applications are currently being studied in different health service domains, and this interest is expected to grow even more in the near future. The diversity of data used for the development of AI-based models, known as the *health data ecosystem*, is also expanding dramatically [10]. The health data ecosystem includes personal data from "standard" health sources, for instance data related to anamneses and medical records, genomics and/or imaging, as well as information from "expanded" sources, such as personal information related to environment, behaviour, lifestyle and/or socioeconomic background [11]. Diverse challenges persist at all stages of AI implementation [12], however the use of AI-based systems in the clinic is expected to grow over the next 5-10 years.

A major part of clinical AI applications will focus on facilitating and enhancing healthcare practice [9]. Healthcare professionals will likely be *supported* by AI with a variety of processes pertaining to different "levels of intelligence". In medical uses of radiation, it is generally believed that professionals will not be *replaced* by AI. "Will AI replace radiologists?" is the wrong question, however it is commonly considered that radiologists using AI might in the future replace (or perform better than) radiologists who do not use it [13, 14]. The workload of professionals may be relieved through AI tools by automating repetitive steps of entire work processes (e.g., image segmentation, treatment planning, quality assurance). Work processes would in these cases be supported by AI, however validation/supervision by the responsible professional would be essential. AI tools might also be used to support the professional responsible for taking clinical actions or decisions. In this context, AI may not only complement clinical practice, but augment it by providing efficiencies and insight beyond human capability

[12]. For instance, AI tools might be used to help in the optimization of imaging procedures, by facilitating image reconstruction.

Even though AI is a very popular clinical research and discussion topic, there is a general lack of deployed AI-based tools in use in the clinic. In fact, the implementation of AI in healthcare presents many challenges, with potential risks of unintended consequences in the clinical, technical, safety, ethical, privacy and regulatory domains [11, 15-18]. If these challenges are not addressed, they could present a barrier to more universal clinical applications of AI [19]. The real risks of unintended and negative consequences, the lack of diversity, standardization and harmonization of data elements used to develop and train the AI-based models, the challenges in their clinical implementation, validation and quality assurance, the legal and ethical issues, and the lack of specific education and training of the involved health professionals, are all possible reasons for the caution around wider implementation of AI-based applications. Compared to statistics and other more traditional approaches in data analysis, another challenge related to the implementation of AI is that many AI-based tools may be perceived as "black boxes"; it may be difficult to understand why or how the algorithm takes its decisions, due to the complexity of the algorithm which often uses a large number of inputs and internal variables. Moreover, powerful AI-based tools are data intensive; they need large datasets for accurate modelling and there is generally no established method to determine the sample size needed for training and validation. This implies that as much data as possible are collected and analysed, rather than just a random or statistically relevant sample.

Prior to the clinical use of any AI-based tool, its safety, effectiveness, appropriateness and efficiency have to be ensured. For the output provided by AI-based tools to be interpretable to clinical decision-makers, it needs to be unbiased and vendor-neutral, as well as ethically and legally compliant with certain principles and value standards [20]. A core team of health professionals is therefore needed to lead the process of AI selection and safe implementation into clinical practice. As a physical scientist in a clinical environment, the clinically qualified medical physicist (CQMP) is considered a key professional that needs to be involved in the process [21].

1.2. OBJECTIVE

The roles and responsibilities of CQMPs are published in the IAEA Human Health Series No. 25 [22] can be contextualized in the framework of the application of AI in radiation medicine. Given their "ability to synthesize the knowledge from a diverse set of disciplines" [21], CQMPs must gain the appropriate technical knowledge to address AI with adequate understanding and expertise. Therefore, there is a substantial need to narrow the gap in education and training of current and future CQMPs, who will likely be involved with the clinical implementation of AI. For new generations of CQMP, a supplement to current academic and clinical training programmes is required [23-27], since the standard education of the medical physicist does not usually have sufficient elements built into it that deal with the basics of AI. In addition, for experienced CQMPs, continuing professional development (CPD) programmes are needed that enhance professional competence in the field of AI.

1.3. SCOPE

This publication frames the roles and responsibilities of CQMPs in the field of implementation and utilization of AI in the medical uses of radiation. The publication provides definitions of AI and a few other terms frequently used in the context of AI (e.g., machine learning (ML),

deep learning (DL), big data). Principles, examples, and associated risks of AI-based tools are also discussed.

This publication supplements IAEA guidance on postgraduate medical physics academic programmes. In addition to the core modules, an elective topic "Advanced Statistical Methods" is listed in the IAEA Training Course Series No. 56 (Rev. 1) [28]. An outline of the content of this topic is provided. Furthermore, the principles and needs for continuing education of CQMPs in the field of AI are discussed and CPD course contents are proposed.

1.4. STRUCTURE

Since many incomplete, inaccurate, or even contradicting definitions of AI can be found in literature, Section 2 provides a definition of AI for this publication, and other related technical terms.

Section 3 provides a list of representative examples of clinical processes that are or will in the near future be using AI. To raise awareness about the risks associated with the potential introduction of AI-based tools in the clinic, some of the potential dangers and concerns are also discussed. A brief insight into some of the existing recommendations, guidelines, and checklists to ensure rigorous and reproducible scientific research and clinical studies in AI, are also provided.

CQMPs are key healthcare professionals that need to be considered for selection, safe implementation and use of AI in the clinic. Their roles and responsibilities, which are described in IAEA HHR-25 [22], are contextualized in Section 4 in the framework of the application of AI in radiation medicine.

Section 5 provides guidance on the competencies needed by current and future CQMPs. A detailed outline of an elective module for postgraduate medical physics academic programmes, which might be offered in countries with relevant specialist expertise and resources, is provided. Furthermore, CPD education and training activities dealing with AI in medical physics are discussed.

2. DEFINITIONS OF TERMS RELATED TO ARTIFICIAL INTELLIGENCE

Professionals need to share the same understanding of the main technical terms associated with AI. The definitions of AI, ML, DL, supervised learning, unsupervised learning, reinforcement learning, and big data are given in this section. Figure 1 illustrates the relationship between AI, ML and DL.

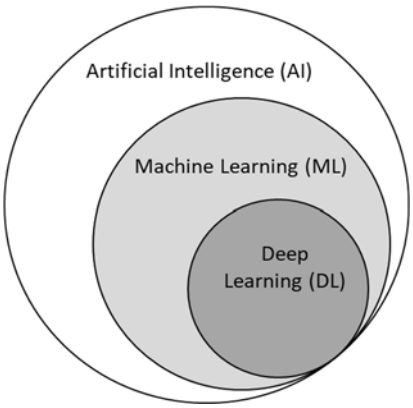

FIG. 1. Relationship between artificial intelligence, machine learning and deep learning.

2.1. ARTIFICIAL INTELLIGENCE (AI)

Artificial intelligence is an array of technologies that enable a machine or computer agent to perform tasks that usually use human intelligence, such as sensing, comprehending, learning, decision-making, and acting. Broadly speaking, AI encompasses computer-aided diagnosis (CAD), radiomics, ML (including DL), computer vision, expert systems and natural language processing. In the last decade, new breakthroughs in DL have greatly accelerated AI applications and enabled the leverage of large datasets and existing experience to render workflows more efficient or automate a task such as prediction, detection, classification, semantic transcriptions, image reconstruction and processing, or sensorimotor control[1].

2.2. MACHINE LEARNING (ML)

ML is a subfield of AI that centres on using computer algorithms to draw inferences and find patterns in data. The algorithms imitate the ways that a human learns by extracting essential features from data and then makes decisions based on inference. For example, a purely physics-based analysis that identifies different tissues in an image would not be considered a ML system as there is no learning involved and the process is based purely on pre-specified rules (e.g. simple thresholding technique on the voxel intensity). If a computer algorithm or other approach is used to learn a mapping from images to tissues from labelled examples of pairs (image patch, tissue characteristics), that would be an example of a ML system. At a high level, ML systems consist of four major elements:

(a) a defined objective or loss function that is a measure of the performance of the system (for example, how to select a feature and how well the predicted tissue labels match the actual labels),

(b) a model class that describes the set of functions that can be used to map from inputs to outputs,

(c) an algorithm used to find the model within the model class that has the best performance according to the objective, and

(d) the independent data used to compute the predictions and the loss (including validation sets).

Three major subfields exist in ML: supervised learning, unsupervised learning, and reinforcement learning. Another specific area of ML is DL.

2.2.1. Deep learning (DL)

The process of DL uses neural networks as its set of functions from inputs to outputs. There are many architectures, or function classes, within DL that are tailored for different types of data – encoders for representation learning, variational encoders (VAEs) and generative adversarial networks (GANs) for generative learning, convolutional networks and vision transformer (ViT) networks for image processing, transformer networks for text, etc. Modern DL models often have millions of parameters. In many contexts, such as image segmentation and classification [29, 30], biomedical data processing [31], and image reconstruction [32], DL models have achieved state-of-the art performance. They tend to require large datasets ("data hungry") to train robust neural networks that generalize well on new data. Overfitting might be an issue when datasets are not sufficiently large, or training and validation processes have not been adequately performed (see Section 3.2).

2.2.2. Supervised and unsupervised learning

Supervised learning is a subfield of ML where the data are provided as examples of (input, output) pairs, and the goal is to learn a function to predict outputs when given inputs, e.g. image patches as input and tissue types as output in the context of segmentation. For lung disease detection, for example, one could have chest X ray or computed tomography (CT) images as input and labelled disease classes as output.

Unsupervised learning is another subfield of ML where the data are provided without any ground truth label, as examples of simply (input), and the goal is to learn some lower-dimensional structure (to characterize the patterns) in the data [31]. For example, one could have image patches and want to cluster them to better understand the types of patches that exist in the data. Sometimes, unsupervised learning can also be used to identify ways to classify and represent the data that can then be used for downstream predictions. For example, the clustering might turn out to be a good way to make predictions about tissue types.

Semi-supervised learning is a hybrid of the above two techniques and proceeds with a small amount of labelled data and a large number of unlabelled examples. It is a special instance of learning with weak supervision [33]. Semi-supervised learning may also be referred to as either transductive or inductive learning and is very useful as it reduces the need for labelled data.

2.2.3. Reinforcement learning

Reinforcement learning is a subfield of ML where data are provided as histories for some sequence (input, action, reward, input, action, reward,...) [34]. Here, the input, or state, is the observation that the ML system receives, action is the decision that it makes, and the reward is the score it gets for its performance. The goal is to learn a mapping (known as a "policy" in computer science) from inputs to actions that maximizes the sum of (potentially weighted) rewards.

2.3. BIG DATA

Big data refers to datasets that are too large or complex to be dealt with using standard data analysis tools. It is typically defined as data that is large in volume (many examples), variety (high heterogeneity), and velocity (rate at which inputs arrive). For example, a research group at NIH has made a dataset with more than 30,000 chest X ray [35], each with up to 14 different thoracic pathology labels. Another example of big data is genomic sequence data, which can include expression data of about twenty thousand of genes and thousands of cells. There is often an implicit assumption that big data are data from more complex sources (e.g. newer instruments, multiple modalities) that require algorithms beyond classical statistical techniques to perform reliable analyses.

According to WHO guidance on ethics and governance of AI for health [11],

> "Over the past two decades, the data that qualify as health data have expanded dramatically. They now include massive quantities of personal data about individuals from many sources, including genomic data, radiological images, medical records and non-health data converted into health data. The various types of data, collectively known as "biomedical big data", form a health data ecosystem that includes data from standard sources (e.g., health services, public health, research) and further sources (environmental, lifestyle, socioeconomic, behavioural and social). Thus, there are many more sources of health data, entities that wish to make use of such data, and commercial and non-commercial applications."

Even if these data exist, they are usually sparse and reside in segregated databases. Methodologies to organize these data in ontologies [36] and federated data systems are important efforts, still in an early stage, that are essential to allow broad AI-based healthcare optimization. Apart from data *quantity*, data *quality* plays also an important role in the successful development of effective AI systems. In this regard, data curation is a fundamental process, since it enables AI models to be developed with relevant, harmonized and error-free quality data.

3. RADIATION MEDICINE PROCESSES BASED ON ARTIFICIAL INTELLIGENCE AND RELATED RISKS

3.1. REPRESENTATIVE EXAMPLES OF AI-BASED RADIATION MEDICINE PROCESSES

Numerous AI applications have been studied in different health service domains. Since 2015, the number of publications dealing with artificial intelligence and/or DL in medicine has grown exponentially [37, 38]. Vendors have started to develop and commercialize some AI-based tools for specific applications in imaging and radiation therapy.

Despite this increase in popularity and availability of AI, clinical implementation of deployed AI-tools is not widely implemented. However, since this is a rapidly evolving field, a wider availability and adoption of AI-based tools to support many processes is expected in the near future. According to predictions based on the hype curve, around 20% of clinical practices may adopt DL models within the next few years [39]. Lists of clinical processes and related applications that are or may soon be based on AI (or include AI components) are given in Table 1 and Table 2.

Table 1 provides examples of processes and related applications in the medical imaging domain, whereas Table 2 provides examples and related applications in the radiation oncology domain. Adaptive radiotherapy and radiomics are not mentioned among the applications given here, since they are considered to be combinations of several of the AI-based applications that are listed separately in the tables.

The tables are not meant to be exhaustive, but rather provide representative examples of processes. For many of the listed applications “there is still a long way to develop accurate, robust, and clinically impactful DL tools to ultimately bring the potential of DL from bench to bed side and to eventually benefit patient care” [37].

TABLE 1. EXAMPLES OF PROCESSES THAT COULD BE BASED ON AI: PART 1 – MEDICAL IMAGING

Process	Applications
Imaging workflow optimization	Prioritization of cases; selection of imaging modality and personalized protocols.
Patient dose estimation in radiological imaging	Tracking of relevant dose metrics; optimization of the imaging procedures by comparing delivered with reference doses.
Image acquisition and reconstruction	CT image acquisition parameter selection; optimal image reconstruction technique for real-time MRI; patient-specific image reconstruction to improve image quality at lower radiation dose; artifact reduction; fully automated data-driven respiratory signal extraction; synthetic image generation; virtual contrast-enhanced imaging; iterative image reconstruction.
Image registration and fusion	Mono-modal and multi-modal image registration; deformable registration; 2D-3D image registration.
Disease identification and characterisation	CAD (detection/diagnosis) of breast cancer, lung cancer, prostate cancer, coronary artery disease, COVID-19 and other pulmonary diseases.
Risk assessment	Breast cancer risk assessment, breast density estimation, imaging-based risk models
Disease monitoring and response assessment	Monitoring of chronic disease, prediction of response to therapy, assessment of risk of recurrence

Note: CT computed tomography; MRI: magnetic resonance imaging; 2D: 2-dimensional; 3D: 3-dimensional; CAD: computer-aided diagnosis.

TABLE 2. EXAMPLES OF PROCESSES THAT COULD BE BASED ON AI: PART 2 - RADIATION ONCOLOGY

Process	Applications
Treatment decision support	Personalized treatment approach (e.g. proton vs photon); pre-treatment patient risk stratification; pre-treatment prediction of tumour response and RT toxicity; individualised RT dose prescription.
Target localization and segmentation	Automated tumour detection (e.g. brain, lung, cervix); automated CTV segmentation considering patient-specific microscopic tumour spread, automated resection cavity delineation (e.g. brain, breast).
OAR volume segmentation	Automated OAR volume segmentation for many sites (e.g., head and neck, breast, pelvis).
Dose prediction and automated planning	Decision support tools for IMRT/VMAT planning; anatomy-based optimal dose prediction; planning process automation to improve efficiency and plan quality; multicriteria treatment plan optimization.
IGRT and motion management	Imaging-based pre-treatment 2D or 3D target localization; automated fiducials or anatomical structure recognition and alignment; real-time tumour tracking; real-time fiducial-based patient motion monitoring; markerless tracking; motion management in MRIgRT; patient motion prediction.
Treatment plan QA	Patient-specific quality assurance (e.g. prediction of passing rates in IMRT/VMAT pre-treatment patient-specific QA).
Equipment QA	Machine-specific QA; performance monitoring over time; prediction of faults to schedule maintenance and QA, additional testing procedures and maintenance.

Note: RT radiotherapy; CTV: clinical target volume; OAR: organ at risk; IMRT: intensity modulated radiotherapy; VMAT: volumetric modulated arc therapy; IGRT: image guided radiotherapy; MRIgRT: magnetic resonance imaging guided radiotherapy; QA: quality assurance.

3.2. RISKS ASSOCIATED WITH THE CLINICAL APPLICATION OF AI-BASED TOOLS

AI is sometimes promoted by industry and researchers as a *panacea* that is going to solve, or, to a large extent mitigate most clinical problems or make complex technologies more accessible. Many processes in the medical use of radiation will likely be based on AI in the near future (see Section 3.1), however broad clinical experience with AI and evidence of its benefit is currently lacking. The inappropriate clinical use of AI-based tools or inappropriately configured AI-based models can severely jeopardize the possible clinical benefits and can clearly result in increased safety and health-related risks for patients.

Together with benefits in healthcare, the application of AI in medicine brings new challenges and risks. It is of paramount importance that the health care professionals are adequately educated and trained and are aware of the risks. If risks are not carefully considered, quality of health service and patient safety may be compromised. Safety issues in the clinical context can be caused by an incorrect or improper use of AI-based tools and/or using AI-based models that were incorrectly or inappropriately trained. This can have consequences on different timescales (short, medium and long term) [40]. Legal and ethical aspects should also be of concern. Some issues related to the use of AI-based tools are mentioned below.

The output of AI-based models is strictly dependent on the quality and quantity of the data that have been collected, annotated and used to train them. *Inconsistent, class-imbalanced, or biased data* might affect the AI output and result in AI models with poor performance. A recent systematic review including 62 studies using ML to detect Covid-19 on chest radiographs and CT scans showed that none of the identified models were of potential clinical use due to methodological flaws and/or underlying biases [41]. In general, a biased dataset might underrepresent the patient population in terms of age, gender, ethnicity, socioeconomic status or geographical origin for instance, or assume a different imaging or therapeutic approach, e.g. the type of technologies used for imaging and radiotherapy and this can affect the performance of an AI model in many ways. It usually results in a biased AI response, such as poorer accuracy and increased risk of harmful effects for the groups that are underrepresented [42]. An improper use of AI might sustain or even amplify existing social inequalities [43] and escalate existing health system inequities [44]. Some AI-based classifiers were shown to consistently underdiagnose under-served patient populations [44]. Gender imbalance in medical imaging datasets was shown to produce biased classifiers for CAD applications, for instance [45]. An improper use of AI might be perceived to impact on human rights [46].

If not adequately harmonized, multi-site medical imaging datasets might show confounding biases [47], which in turn might influence the results of ML algorithms [48]. It has been shown recently that some AI approaches developed to automatically diagnose Covid-19 from chest radiographs were better able to recognize the hospital and country of origin of the radiograph rather than identify disease-related characteristics. This is because some datasets had confounding factors, such as radiopaque markers, pixel size, and overall pixel values that may have inadvertently correlated with the presence of the disease [49]. This type of error, where AI improves performance using a quicker but wrong learning strategy, is called *short-cut learning*.

Dataset shift occurs when the data used for developing an AI model have a different distribution than the data encountered by the AI model in clinical deployment. A dataset shift might happen when patterns of clinical practice evolve over time, due to the introduction of new treatment approaches and technologies, or if patient population characteristics are gradually changing. It

also occurs if a biased training set was used, or the AI-based tool is used in a population different than the one with which it was trained. Dataset shifts may result in increased unreliability of the AI system over time, since conventional ML models usually adapt poorly to a significant change in operational data or to an unanticipated patient context [40]. As stated by Subbaswamy and Saria [50], "beyond contributing to poor performance, failing to account for shifts can also lead to dangerous decisions in practice: the system can fail to diagnose severely ill patients or recommend harmful treatments." An overview of categories and related examples of dataset shifts, as well as strategies for their recognition and mitigation, are given in [51].

AI models are also prone to *overfitting*. Overfitting is an event in which models perform very well on the training data, to the point that they are more sensitive to noise in the data than to their underlying patterns. This is the case when the numbers of features per subject is not significantly lower than the number of subjects in the dataset. As a result, the model is poorly *generalizable*, meaning that it will perform poorly on new, unseen datasets [52]. A recent study showed variable generalization performance of a DL model to detect pneumonia in chest radiographs. Test data, which were acquired in the same hospital where the training data came from, overestimated the performance of the same model when applied to real-world data [53]. Ways to prevent overfitting are multifold, including pre-registration of the analysis protocol, selection of smaller sets of features, controls on the complexity of the model, and using rigorous internal and external validation procedures.

ML training results might suffer from other sources of uncertainties e.g pitfalls in the search of optimal solutions due to the presence of local minima in the multi-dimensional function to be minimized.

As described in Challen et al [40], further short term issues with AI applications include their potential *insensitivity to the real-world impact of a false output* with respect to the clinical context of use, their *black-box* nature and the lack of confidence in the output accuracy (lack of the degree of trustworthiness of an output). Other safety issues might inadvertently arise from the clinician's cognitive bias of *automation complacency*, the phenomenon where the clinician gives more weight to the AI system than it deserves [54]. Clinical decision support systems based on AI might for instance improperly reassure novice physicians and transmit a false sense of security [55]. Instead of conducting a critical and extensive evaluation, inexperienced observers might be primed by these systems and rather search only for confirmatory information [56].

3.3. GUIDELINES FOR DEVELOPING AND REPORTING AI-BASED CLINICAL STUDIES AND SCIENTIFIC RESEARCH

Clinical and scientific studies dealing with AI are also prone to the risks mentioned in Section 3.2. Among the high and rapidly increasing number of recent publications, some showed methodological shortcomings, poor transparency, and poor reproducibility [57]. Incomplete or inconsistent reporting as well as lack of sufficient rigour were also in some cases a matter of concern.

To address these issues, several organizations and peer-reviewed international journals recently developed or are currently developing recommendations, guidelines or checklists to improve the quality of AI-based studies, and to ensure rigorous and reproducible scientific research, publications and clinical studies [58]. A list of some of the publications is given in Table 3.

These publications support algorithm designers, researchers, repository managers, manuscript writers and reviewers, journal editors and model users to achieve best practice in their respective field of competence [59]. Such documents might be useful also to CQMPs, since they may cover many of these roles in the context of radiation medicine clinical studies and research (see Section 4). Other guidelines are being developed, for instance the AI specific version of the STARD checklist (STARD-AI protocol [60]) and the AI-specific extension of the TRIPOD statement and the PROBAST tool for diagnostic and prognostic prediction model studies based on AI (TRIPOD-AI and PROBAST-AI [61]).

With respect to publications on AI in the Medical Physics Journal, specific guidelines including a checklist for AI applications in medical physics (CLAMP) were recently published to ensure rigorous and reproducible scientific research [62].

In addition, multiple societies are collaborating to create large, open datasets of images to provide representative, diverse imaging data to AI researchers to help mitigate biases in datasets and AI algorithms.

TABLE 3. PUBLICATIONS PROVIDING RECOMMENDATIONS, GUIDELINES OR CHECKLISTS TO SUPPORT AI-BASED CLINICAL STUDIES AND SCIENTIFIC RESEARCH

Publication title	Description	Publication year
Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers–From the Radiology Editorial Board [63]	Nine key considerations to evaluate AI research are listed and discussed	2020
Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers [64]	A checklist of 42 items is provided and the items are individually discussed	2020
Minimum information about clinical artificial intelligence modelling: the MI-CLAIM checklist [59]	Six parts of the MI-CLAIM model are outlined and a checklist is provided	2020
Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension [65]	An extension of the SPIRIT checklist is given, with 15 items to be addressed in case of clinical trial protocols for interventions involving AI	2020

Publication title	Description	Publication year
Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension [66]	An extension of the CONSORT checklist is given, with 14 new checklist items added to be additionally addressed in case clinical trial reports for interventions involving AI	2020
The TRUE Checklist for Identifying Impactful Artificial Intelligence-Based Findings in Nuclear Medicine: Is It True? Is It Reproducible? Is It Useful? Is It Explainable? [67]	A simple checklist comprising 4 basic questions is presented to assess the likelihood that a manuscript dealing with AI will be impactful	2021
AI in Medical Physics Guidelines for publication [62]	The CLAMP checklist reflecting the expectations for manuscripts dealing with AI applications that are submitted to the Medical Physics Journal	2021
Nuclear Medicine and Artificial Intelligence: Best Practices for Evaluation (the RELAINCE Guidelines) [68]	Framework to objectively evaluate AI models in nuclear medicine	2022
Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review [69]	Guidelines are given for the development, evaluation and implementation of AI-based prediction models, based on a six-phase structured approach	2022
Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI [70, 71]	A checklist with 17 AI-specific and 10 generic reporting items is presented. Several existing and upcoming reporting guidelines are listed	2022

4. ROLES AND RESPONSIBILITIES OF CLINICALLY QUALIFIED MEDICAL PHYSICISTS IN AI-BASED CLINICAL APPLICATIONS

Medical domain knowledge is typically required to introduce, apply and maintain AI applications effectively and safely in clinics. In clinical settings, the CQMP is the professional in the health care team that can bridge complex technological tools, such as AI applications, with the clinical domain needs. According to the IAEA Human Health Series No. 25 [22], one of the main roles and responsibilities of CQMPs common to all medical physics specialties is the technical supervision of equipment operation and maintenance:

> "CQMPs are responsible for establishing acceptance and commissioning procedures for diagnostic, therapeutic and measurement equipment." They "supervise their implementation, performing quality control and calibration measurements to ensure the safe and optimal performance of the equipment, and authorizing clinical use."

In line with this general statement, it is important to contextualize the roles and responsibilities of CQMPs with respect to implementation, risk mitigation and utilization of AI in radiation medicine. In cases of inappropriate use of AI-based products, CQMPs are in principle accountable for the physical and technical aspects related to local validation, acceptance, maintenance, safety and QA of these tools. Six main areas of responsibility involving CQMPs and the clinical use of AI across all MP specialities can be identified, namely:

(a) Development of technical specifications of the equipment;

(b) equipment acceptance and commissioning;

(c) optimization of the physical aspects of medical procedures;

(d) quality management of the physical, technical and safety aspects;

(e) education and training of other health care professionals, and

(f) scientific research and development.

The sub-sections below are organized according to these main areas of responsibility. To each area of responsibility, a point-by-point description of the potential AI-based processes that involve CQMPs is given. The team of professionals typically consists of (but may not be limited to) clinicians, CQMPs, and medical radiation technologists in radiotherapy, radiology or nuclear medicine.

4.1. TECHNICAL SPECIFICATION

According to IAEA HHR-25 [22],

> "CQMPs have a leading role in preparing equipment specifications according to the needs of the [medical radiological] facility, and they participate in the tender evaluation and purchase recommendation of the equipment. They analyse the functional requirements for clinical use, and specify the necessary conditions for integration, compatibility and connectivity to existing infrastructure of the equipment to be purchased."

Applying these principles, CQMPs are involved in the following aspects of preparing for AI-based tools:

(a) Preparation of the functional specifications of the AI-based tools to be purchased.

(b) Identification of potential vendors/products for the planned clinical application.

(c) Collection of the information for the final product specification that is used for procurement. During this process, it may become evident that the initial functional requirements need to be modified or that more relevant features or products need to be included. The research will preferentially start with several potentially suitable products, allowing cross-product learning from systems that are basically performing/supporting the same clinical task. Typically, the CQMP will:

    (i) Review the key system documentation from vendors and available literature to learn as much as possible about the systems, identify the characteristics of the embedded AI model/engine, including performance on the intended task, safety features, generalizability, interpretability and option/need for training with institutional data, and applied hardware. Part of the information may be found in the overview specification sheets provided by the vendors, but generally more extensive documentation is needed. A detailed understanding of the underlying context, training and validation datasets, and the mathematical models as well as metrics that were used to quantify system performance, is required for meaningful interpretation of the information.

    (ii) Perform a systematic review to assess the experience of existing users with similar AI-based tools.

    (iii) Get a test version of the most relevant products in house, and systematically perform end-to-end tests, usability tests etc, with participation of the full clinical team. Special attention is needed for performance tests on local data.

    (iv) Explore possibilities for embedding the AI-based tool in the existing IT infrastructure.

(d) The CQMP contributes to the development of a functional and technical system specification to be used for procurement, based on the final functional requirements and technical/functional information gathered in the pre-purchase phase.

(e) The CQMP participates in tender evaluation and scoring of offers, providing input especially with regards to the physical, technical and safety aspects.

(f) Together with the clinical team and management, the CQMP verifies that the offers are aligned with national and international guidelines and legislation (e.g., [15, 72]).

(g) The CQMP participates in final negotiations with vendors where input on the physical, technical and safety aspects are discussed. Part of the final purchase requisition is a detailed validation and acceptance protocol, covering all performance, safety and operational aspects, and based on appropriate, agreed metrics.

(h) Agreement is also needed on vendor support for training on local data (if applicable), vendor involvement in maintenance (e.g. related to model updates) and on training of team members for safe, effective and efficient clinical use.

4.2. ACCEPTANCE AND COMMISSIONING

According to IAEA HHR-25 [22],

> "Following the installation of new equipment, CQMPs are responsible for specifying the basic standards to be applied for its acceptance and subsequent commissioning. They ensure that all units and systems function according to their technical specifications and provide advice on any deviation of equipment performance from acceptable criteria". CQMPs also have the "responsibility for the verification of the computer systems and algorithms associated with the new equipment, and for ensuring that they are adequate for safe and effective clinical use."

Adapting these principles to the introduction of AI-based tools in the clinic, CQMPs will be involved with the following technical aspects related to acceptance and commissioning.

(a) The CQMP ensures that acceptance and commissioning of the AI system is in agreement with national and international guidelines and legislation (e.g., [15, 72]).

(b) If there are IT experts available in the department, the CQMP coordinates discussions between these experts and the vendor on the installation of the AI product, which should be compatible to the existing IT infrastructure.

(c) After installation by the vendor, the CQMP performs the agreed acceptance tests and demonstrates compliance in the specifications as described in the contract with the vendor. Other professional belonging to the clinical team might also be involved, depending on the clinical purpose of the procured AI system. In case of discrepancies, analyses and discussions with the vendor may yield possible solutions. In case of persisting deviations, the CQMP will discuss possible clinical consequences with the clinical team. The CQMP supports the management in case of re-negotiations with the vendor, especially if technological performance or safety is at stake.

(d) The CQMP leads the final performance evaluation of the system run with standardized validation datasets and/or local data. Training of the system may also be based on local data. In both cases, the CQMP is responsible for collecting appropriate data, in quantity and quality and for data curation, when needed. In case of training with local data, final validation can usually be performed with a separate local dataset that was not used for the training process. Performance is compared with criteria mentioned in the acceptance protocol and the performance claimed by the vendor or other users (if applicable and known).

(e) In case of system or model updates, the CQMP will lead a new acceptance/commissioning round. Often this will include repeating the tests used for initial acceptance testing.

4.3. OPTIMIZATION OF THE PHYSICAL ASPECTS OF DIAGNOSTIC AND THERAPEUTIC PROCEDURES

As for any practice involving "medical radiological equipment and software that could influence the delivery of medical exposure", it is recommended that practices using AI-based technologies adhere to the International Basic Safety Standards [73] in addition to compliance with national radiation protection regulations.

(a) If applicable, the CQMP will collaborate with the clinician in ensuring that protection and safety is optimized for processes involving AI-based tools.

(b) Examination efficacy as well as image quality and perception may be influenced when using AI-based tools. If applicable, the CQMP will assist the radiation medicine practitioner in their evaluation.

4.4. QUALITY MANAGEMENT OF THE PHYSICAL, TECHNICAL AND SAFETY ASPECTS

As for any routine use of medical equipment, if AI-based tools are used in medical uses of radiation, the CQMP "participates as a team member in designing and implementing a quality management programme" [22], with a specific responsibility for the technical and physical aspects of the quality assurance programme. A list of responsibilities involving a CQMP is provided below. This list is not exhaustive.

(a) The CQMP leads the formal risk assessment procedure that is performed prior to first clinical use. Risk assessments are used to identify possible safety and performance risks in the clinical use of the AI-based tool and to define action levels and processes that can mitigate the risks if they occur. The outcome of the risk assessment is taken into consideration in the development of protocols for clinical use and system QA.

(b) Together with the entire clinical team, the CQMP prepares a protocol with guidelines for clinical use of the AI-based tool. This protocol defines the overall workflow, working instructions and the precise roles and responsibilities of all the team members. It is supposed to ensure that the system is used correctly and safely, and that performance as obtained in the acceptance phase is obtained in daily clinical practice, or even improved in case of e.g., manual adjustments of AI output. A clear description of the following is needed:

    (i) The specific application of the AI tool (e.g., imaging modality/ies, tumour site(s)).

    (ii) The required input(s) in the AI tool

    (iii) The expected output of the AI tool and the way to interpret the AI results

(c) Patient cohorts, or imaging or treatment protocols might change over time. CQMPs need to be aware of these changes (ideally alerted by system capable of monitoring changes and of assessing the impact of these changes on the performance of AI tool) and act upon them with the other team members to assure continued effective and safe use of the AI tool.

(d) The CQMP takes the lead in the establishment and implementation of a QA programme for the AI system, which is essential for quality and safety in its use, e.g. develop a detailed description of the tests that need to be performed on a regular basis. The programme ensures consistent output of the AI system over time, for instance by running the same standardized set of test images and comparing the obtained output against the baseline, obtained during acceptance/commissioning.

(e) In case of system or model updates, the CQMP is responsible for the quality and safety aspects of the adoption of the updated model, including acceptance and commissioning

efforts (section 4.2), adaptation of the QA protocol (if required) and training of the clinical team.

(f) All safety and performance issues that potentially occur during clinical use are to be reported to the CQMP, who will carry out further analyses. The CQMP also leads the reporting and discussions with the vendor.

4.5. EDUCATION AND TRAINING OF OTHER HEALTH PROFESSIONALS

Clinical AI tools are not yet widespread in the medicinal use of radiation, owing to their high complexity and rapidly evolvement. Not all team members involved in the clinical application of AI will necessarily have enough theoretical background or skills for its safe and effective use. Theoretical background therefore also needs to be complemented with practical training. The CQMP participates in continuing professional development, including the development of additional competencies that would be needed for their roles in ensuring safe and effective application of the AI tool. In turn, the CQMP would support the team by:

(a) participating in the assessment of the educational and training needs for other team members.

(b) contributing to fulfilling the educational and training needs, e.g., by providing lectures or identifying educational training opportunities elsewhere. If the training is provided by the vendor, the CQMP reviews and critically evaluates the contents and programme, and suggest modifications if deemed necessary for safe, effective and efficient utilization of AI systems.

(c) assessing the background and skills of IT experts that have a role in the installation and/or maintenance and QA of the AI tool, as relevant to the deployment and utilization of AI tools.

(d) providing mentoring and supervision for residents and students in the team on the theory and skills needed for the AI system to be safe, effective, and efficient in its clinical use.

4.6. SCIENTIFIC RESEARCH AND DEVELOPMENT

Clinical use of AI is highly promising, but in many circumstances it is still limited in evaluation or restricted in scope. For these reasons, substantial research potential exists for all clinical team members, including the CQMP.

(a) CQMPs can perform scientific research and development on the technical aspects of AI tools and their application.

(b) Together with the research team, the CQMP ensures that the relevant reporting guidelines pertaining to the research methodology are taken into consideration (see section 3.3).

(c) The CQMP can investigate adoption of new AI-based procedures, evaluate the clinical performance of existing systems or focus the scientific research to enhance the performance, e.g., by better training, or smarter or wider use. The CQMP can also be involved in the design and development of methodologies for effective maintenance and QA of clinical AI systems.

5. ACQUIRING THE COMPETENCIES IN AI

Current and future CQMPs are expected to acquire the ability to efficiently and safely commission, apply, assure quality, and maintain AI frameworks that are related to medical physics. They are expected to be able to fulfil the responsibilities presented in Section 4. As such it is critical that they acquire the adequate base knowledge, outlined below in this section. Two pathways have been identified to address the knowledge gap of current and future CQMPs: one is specific to postgraduate medical physics academic programmes for students (Section 5.1), and the other is continuing professional development opportunities for CQMPs (Section 5.2).

5.1. POSTGRADUATE MEDICAL PHYSICS ACADEMIC PROGRAMMES

The academic module contained within this programme is aimed at preparing future CQMPs to understand the principles of AI-based frameworks, their inputs and outputs, and the basic approaches that are used.

In addition to the core modules defined in the IAEA Training Course Series No. 56 (Rev. 1) [28], "Advanced Statistical Methods" is listed as an elective topic and may be offered where relevant specialist expertise and resources exist.

It is recommended that the students have prior knowledge of:

fundamental statistics, including central limit theorem; main probability functions and applications; statistical inference testing and assessment of statistical significance; Type I and type II statistical errors; statistical power and sample size calculations; analysis of variance (ANOVA).

basic computational programming, including management of DICOM files and relevant data extraction, and scripting for data and image analysis and their graphical presentation (e.g. Python, R, Octave)

5.1.1. Outline of Advanced Statistical Methods

This elective module is aimed at preparing students in medical physics with background knowledge for safe, effective, and efficient use of AI, and it is expected that the contents are presented at a level that is appropriate for the students to achieve adequate knowledge and skills. The introduction of this module should not be done at the expense of the number, quality and length of the core modules listed in the IAEA Training Course Series No. 56 (Rev. 1) [28]. There might be therefore implications regarding the overall duration of the academic programme.

The outline of the content for this module is the following:

(a) Introduction
    (i) Historical background pertaining to the development of AI
    (ii) The role of the Medical Physicist
(b) Logistic regression for predictive modelling
    (i) Definition and interpretation of the logistic and logit function
    (ii) Problem, data, model, fit, and evaluation for logistic regression

(c) Regression, classification and decision boundary definition
    (i) Difference between regression and classification
    (ii) Converting problem between problem types
    (iii) Interpretation of each model's output
(d) Receiver operating characteristic (ROC) analyses
    (i) Basic concept and interpretation
    (ii) True positive, true negative, false positive, false negative
    (iii) Type I and Type II errors
    (iv) Sensitivity and Specificity
    (v) Area under the curve (AUC)
(e) Co-variance, correlation, regression, R^2
    (i) Definitions and interpretations
    (ii) Analysis
(f) Machine learning categories
    (i) Supervised learning
    (ii) Unsupervised learning
    (iii) Reinforcement learning
    (iv) Hybrid learning categories
        – Semi-supervised learning
        – Self-supervised learning
(g) Machine learning models and data analytics tools
    (i) Linear and logistic regression
    (ii) Neural networks
    (iii) Dimensionality reduction (e.g. Principal Component Analysis)
    (iv) Support vector machine
    (v) Decision trees and random forests
    (vi) Gradient boosting
    (vii) K-means cluster analysis
    (viii) Metrics of evaluation (e.g. confusion matrix)
(h) Training and validation of machine learning models
    (i) Mathematics of training (e.g. loss function, backpropagation, optimization)
    (ii) Data augmentation
    (iii) Model selection and regularization
        – Forward and backward stepwise predictor selection
        – Ridge regression

- LASSO

(iv) Training methods

(v) Hyperparameter optimization

(vi) Required sample size

(vii) Bias-variance trade-off

(viii) Overfitting

(ix) Handling co-linearity of predictors (Variance inflation factor, VIF)

(x) Model validation

- Cross-validation (K-fold, leave-one-out)
- Bootstrap
- Generalisability
- External validation
- Calibration

(i) Deep learning

(i) Deep learning and neural networks

(ii) Convolutional neural networks

(iii) Recurrent neural networks

- Long short-term memory (LSTM) networks

(iv) Transformer networks

- Transformer networks for text
- Vision and Swin transformer networks for image processing

(v) Generative Adversarial Networks (GANs) (e.g. synthetic images)

(vi) Transfer Learning

- Domain Adaptation

(vii) Augmentation

(viii) Metrics of evaluation (e.g. DICE)

(j) Data management

(i) Data collection including data retrieval

(ii) Data quality assessment (sample size, imbalance)

(iii) Data curation

Anonymization and de-identification

Labelling and segmentation

Harmonization

Standardization

Robustness

(k) Pitfalls and bottlenecks of AI in medicine
    (i) Risk analyses
    (ii) Model drift and data drift
(l) Regulatory issues
(m) Overview of AI-based clinical applications
(n) AI-related ethical issues

5.1.2. Practical sessions

Practical sessions are essential to supplement student learning and develop skills as part of the education programme. Due to the similarities between AI models, a single practical session can cover several topics and academic modules. Examples of practical sessions and the knowledge to which they apply are given below.

(a) Training of one classification and one regression model
    (i) Choice of proper model and tuning strategy.
    (ii) Impact of the size of the database (overfitting and generalization).
    (iii) Testing and validation of an AI-based model.
    (iv) Regularization of an AI-based model.
    (v) Establish the quality of the model.

(b) Training of one DL model dealing with image segmentation
    (i) Understand influence of the input data on the model performance.
    (ii) Testing and validation of an AI-based model.
    (iii) Regularization of an AI-based model.
    (iv) Impact of different architectures.
    (v) Establish the quality of the model.
    (vi) Transfer learning.

(c) Assessing DL based segmentation in radiotherapy applications
    (i) Apply a DL segmentation model to segment organs-at-risk and/or tumour tissue in a set of CT images.
    (ii) Evaluate how well the DL-based segmentation corresponds to clinical segmentations.
    (iii) Evaluate how segmentation uncertainty is indicated by the DL model.
    (iv) Create a treatment plan using both DL-based and clinical segmentations and evaluate plan quality differences.

(d) Assessing AI in radiological detection applications (e.g., lung nodule detection and classification, chest X-ray classification)

(i) Apply a DL model to detect and classify potential diseases in a new set of medical images (e.g., X-ray, CT and mammography).

(ii) Assess accuracy of disease detection by the model and evaluate model accuracy in detecting diseases that are less common.

(iii) Assess accuracy of disease localisation. If the model does not offer localisation directly, evaluate visual explanation techniques that explain its decisions

(iv) Evaluate how the model indicates uncertainty in its decision.

(v) Evaluate if the presence of false positive and false negative results affects the further course of treatment.

(e) Assessing image reconstruction from low-dose imaging

(i) Apply a DL model to reconstruct "diagnostic-quality" images from low-dose images, with known ground truth imaging available. Such datasets may need to be constructed, if not publicly available.

(ii) Assess accuracy of the reconstructed images. Evaluate if there are areas in the image, for example specific tissues, where the reconstructed image consistently deviates from the original.

(iii) Assess accuracy of the reconstructed image in the presence of artifacts or artifact sources (e.g., metallic objects in CT). Analyse the effects on reconstructed images of artifacts presence.

(iv) Discuss about clinical implications of the reconstructions and try to understand if clinical decisions would potentially be affected by using reconstructed images (e.g., evaluate if small nodules would be missed or introduced by reconstructing a "diagnostic-quality" image; evaluate if radiotherapy treatment plan quality based on such reconstructed images would be degraded).

Other training activities, similar to the ones listed above and addressing the same or other academic topics, may be considered.

5.1.3. Knowledge resources

A list of resources that can be used to develop the programme is proposed below. Some are available electronically. There are freely downloadable machine learning frameworks to use (e.g. TensorFlow, PyTorch, scikit-learn, mlr3, etc.).

(a) Core resources

(i) JAMES, G., WITTEN, D., HASTIE, T., TIBSHIRANI, R., An Introduction to Statistical Learning. Springer, New York (2013).

(ii) HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J.H., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2009).

(iii) FLACH, P., Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press (2012).

(iv) GOODFELLOW, I., BENGIO, Y., COURVILLE, A., Deep learning, MIT Press (2016).

(v) BISHOP, C.M., NASRABADI, N.M., Pattern recognition and machine learning, Springer, New York (2006).

(vi) BISHOP, C. M., Neural networks for pattern recognition. Oxford University Press (1995).

(b) Additional resources

(i) BURKOV, A., The hundred-page machine learning book. Quebec City, QC, Canada: Andriy Burkov (2019).

(ii) PATTERSON, J., GIBSON, A., Deep learning: A practitioner's approach. O'Reilly Media, Inc (2017).

(iii) LAPAN, M., Deep reinforcement learning hands-on. Packt publishing (2020).

(iv) THEOBALD, O., Machine Learning For Absolute Beginners: A Plain English Introduction. Scatterplot press (2017).

(v) SEGARAN, T., Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, Inc (2007).

(vi) MURPHY, K.P., Machine Learning: A Probabilistic Perspective. MIT Press (2012).

(vii) MÜLLER, A.C., GUIDO, S., Introduction to Machine Learning with Python: a guide for data scientists. O'Reilly Media, Inc (2016).

(viii) RASCHKA, S., Python machine learning. Packt Publishing Ltd (2015).

(ix) REED, R., MARKS II, R.J., Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks. MIT Press (1999).

(x) STEVENS, E., ANTIGA, L., VIEHMANN, T., Deep Learning with PyTorch. Manning Publications (2020).

(xi) CHOLLET, F., Deep Learning with Python. Manning Publications (2021).

(xii) TRASK, A.W., Grokking deep learning. Manning Publications (2019).

(xiii) GÉRON, A., Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc (2019).

(xiv) SHUKLA, N., FRICK LAS, K., Machine learning with TensorFlow. Manning Publications (2018).

5.2. MAINTAINING AND UPDATING COMPETENCIES OF CLINICALLY QUALIFIED MEDICAL PHYSICISTS

According to IAEA Training Course Series No. 71 [74],

> "CQMPs need to continuously update their knowledge and competencies, adapting to the evolution in science and technology. Therefore, CQMPs are expected to not only maintain the standard of practice required, but actively seek to preserve and update their knowledge and competencies".

Given the expected increase of clinical processes that will be based on AI (see section 3.1), the related risks (see section 0), and the roles and responsibilities of CQMPs in the clinical application of AI (see section 4), the need for CPD in the field is advised. Depending on the professional experience with clinical AI, different continuing professional development activities are proposed. Table 4 gives a list of possible CPD course(s) and their contents according to the prior experience of the CQMP. The principles of these activities are discussed in the next subsections.

TABLE 4. CONTINUING PROFESSIONAL DEVELOPMENT ACTIVITIES DEALING WITH AI FOR CQMP

Course content	Type of course	Target CQMPs
Insight into current AI-based tools in MP	Theoretical sessions	As needed
Theoretical principles of AI from the user's perspective	Theoretical sessions; optional practical sessions	Essential for CQMP involved with any AI application in any MP speciality
General approaches and insight on training, validation, application and QA of data in AI	Theoretical sessions; desirable practical sessions	Recommended for CQMP involved with any type of AI-based application in one specific MP speciality
Best practices in AI-based tools applied to a specific clinical task	Theoretical and practical sessions	Recommended for CQMP involved with one specific AI-based application in one specific MP speciality
Theoretical principles of AI from the developer perspective	Theoretical and extensive practical sessions	Optional for CQMPs involved in the research and development of clinical AI models

5.2.1. Education on AI-based tools in MP

Insight into current AI-based tools in medical physics would provide CQMPs a general perspective on the development and use of AI technologies in the field. The courses are meant to not only cover the AI-based tools that are currently used in medical physics, but also provide some insight into what is likely to become an integral part of clinical practice in the future. The courses can provide an overview of the subject areas that CQMPs could be involved in, such as image acquisition and standardization of imaging data, detection of lesions, image reconstruction, synthetic imaging, dose assessment, optimization, segmentation, treatment planning, quality assurance, treatment delivery, follow-up, etc. It is expected that the landscape of AI-tools used in medical physics will evolve over the years as responsibilities and paradigms of the CQMP also evolve, and the courses may need to be updated frequently to maintain relevance. As these courses would not be focused on specific AI-tools in the clinic, these would target CQMPs that seek to participate in future implementation and usage of such technologies.

5.2.2. Education on theoretical principles of AI from the user's perspective

To ensure safe and effective use of AI in the clinic, it is essential that the CQMP has an elementary theoretical background in AI before starting with clinical projects. An insight into the theoretical principles of the main topics listed in Section 5.1.1 is part of the programme, without too many technical details and without approaching the topics from the developer's perspective. This avoids suboptimal and slow clinical introduction as well as the mismanagement of AI models as inscrutable "black boxes". It is recommended that the CQMP does not approach clinical AI applications without an understanding of their principles, and the characteristics of the needed inputs and output(s). The courses can be designed for the CQMP to effectively use the AI technology deployed in the clinic. The CQMP would be able to not only utilize the technology, but also be able to interpret results, identify when the model performance degrades or gives incorrect predictions, and troubleshoot issues (from the user perspective) if they arise.

5.2.3. Education providing a general insight into clinical applications of AI

These courses are designed and focused to provide the CQMPs with insight on existing commonly agreed methods for the training, validation, and clinical application of the AI technologies. The courses can also discuss the potential challenges, hurdles and barriers when integrating, commissioning, validating and deploying the model for clinical use. This would give the CQMP a general pathway to clinically apply the framework in question and facilitate troubleshooting.

In addition, the CQMPs can be educated in the areas of risk analysis, QA, maintenance and update of such technologies, following the deployment. Many data-driven models may experience performance decay as the data distribution slowly shifts due to several aspects such as patient demographic drift, medical technological improvements, clinical practice, protocol improvements, etc.

5.2.4. Education on best practices in AI-based tools applied to a specific clinical task

These courses are designed for a specific subject area, such as diagnosis, image reconstruction, dose assessment, optimization, segmentation, treatment planning, quality assurance, treatment delivery, follow-up, etc., and discuss criteria for equipment purchase/selection and specific methods for AI technology development and application areas for clinical tasks. The course therefore may include examples of specific clinical tasks to allow for the CQMP to grasp a

better understanding of the processes involved. Practical sessions for example clinical tasks are recommended to provide the CQMP with hands-on experience.

5.2.5. Education on the theoretical principles of AI from the developer perspective

Courses on mathematical models, and the theoretical principles and concepts of AI from the developer's perspective are available online. It is practical and effective to learn from these types of courses. Extensive practical sessions that involve coding and AI model development are usually included. These courses are particularly indicated for CQMPs involved in the research and development of clinical AI models. As AI is constantly and rapidly evolving , the content of the courses is expected to follow this trend and to better reflect the major technology developments. Listed below are some examples of courses that are currently available:

(a) Machine Learning Specialization, created by Stanford Online and DeepLearning.AI (offered on Coursera at https://www.coursera.org/specializations/machine-learning-introduction);

(b) Deep Learning Specialization, created by DeepLearning.AI (offered on Coursera at https://www.coursera.org/specializations/deep-learning);

(c) Machine Learning Crash Course (offered by Google AI at https://developers.google.com/machine-learning/crash-course)

(d) Deep Learning A-Z: Hands-On Artificial Neural Networks (offered on Udemy at https://www.udemy.com/course/deeplearning/)

6. CONCLUSION

CQMPs are health professionals that are responsible for the selection, safe implementation and use of AI-based tools applied to medical uses of radiation. Their roles and responsibilities in this framework are described in this publication. To be able to manage AI-based tools with the adequate expertise, CQMPs need specific competencies and CPD is needed. This publication provides a detailed outline of the Advanced Statistical Methods elective module for postgraduate medical physics academic programmes, which might be offered where relevant specialist expertise and resources exist. Furthermore, the principles and needs for CPD in the field of AI are discussed and different CPD activities are proposed.

REFERENCES

[1] XING, L., GIGER, M.L., MIN, J.K., Artificial Intelligence in Medicine: Technical Basis and Clinical Applications, Academic Press, London, UK (2020).

[2] MCCARTHY, J., MINSKY, M.L., ROCHESTER, N., SHANNON, C.E., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence - August 31, 1955, AI Magazine. **27** (2006) 12-14.

[3] GIGER, M.L., CHAN, H.P., BOONE, J., Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM, Med Phys. **35** (2008) 5799-5820.

[4] ZHANG, W., et al., Computerized Detection of Clustered Microcalcifications in Digital Mammograms Using a Shift-Invariant Artificial Neural-Network, Med Phys. **21** (1994) 517-524.

[5] AVANZO, M., et al., Machine and deep learning methods for radiomics, Med Phys. **47** (2020) E185-E202.

[6] SAHINER, B., et al., Deep learning in medical imaging and radiation therapy, Med Phys. **46** (2019) e1-e36.

[7] KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G.E., ImageNet Classification with Deep Convolutional Neural Networks, COMMUNICATIONS OF THE ACM. **60** (2017) 84-90.

[8] AGGARWAL, N., et al., Advancing Artificial Intelligence in Health Settings Outside the Hospital and Clinic, NAM Perspect. **2020** (2020) 1-26.

[9] BOHR, A., MEMARZADEH, K., "Chapter 2 - The rise of artificial intelligence in healthcare applications", Artificial Intelligence in Healthcare, Academic Press, London, UK (2020) 25-60.

[10] VAYENA, E., BLASIMME, A., Biomedical Big Data: New Models of Control Over Access, Use and Governance, J Bioethic Inq. **14** (2017) 501-513.

[11] WORLD HEALTH ORGANIZATION, Ethics and governance of artificial intelligence for health, WHO, Geneva (2021).

[12] AUNG, Y.Y.M., WONG, D.C.S., TING, D.S.W., The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare, Br Med Bull. **139** (2021) 4-15.

[13] LANGLOTZ, C.P., Will Artificial Intelligence Replace Radiologists?, Radiol-Artif Intell. **1** (2019)

[14] PESAPANE, F., et al., Myths and facts about artificial intelligence: why machine- and deep-learning will not replace interventional radiologists, Med Oncol. **37** (2020)

[15] U.S. FOOD AND DRUG ADMINISTRATION (FDA), Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, FDA, Silver Spring, Maryland, USA (2021).

[16] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, OECD LEGAL INSTRUMENTS, (2019).

[17] EUROPEAN COMMISSION EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, Liability for Artificial Intelligence and other emerging digital technologies, EUROPEAN COMMISSION (EC), (2019).

[18] MATHENY, M.E., WHICHER, D., THADANEY ISRANI, S., Artificial Intelligence in Health Care: A Report From the National Academy of Medicine, JAMA. **323** (2020) 509-510.

[19] REDDY, S., ALLAN, S., COGHLAN, S., COOPER, P., A governance model for the application of AI in health care, J Am Med Inform Assn. **27** (2020) 491-497.

[20] WIENS, J., et al., Do no harm: a roadmap for responsible machine learning for health care, Nat Med. **25** (2019) 1337-1340.
[21] XING, L., KRUPINSKI, E.A., CAI, J., Artificial intelligence will soon change the landscape of medical physics research and practice, Med Phys. **45** (2018) 1791-1793.
[22] INTERNATIONAL ATOMIC ENERGY AGENCY, Roles and Responsibilities, and Education and Training Requirements for Clinically Qualified Medical Physicists, Human Health Series No. 25, IAEA, Vienna (2013).
[23] ANDERSSON, J., et al., Artificial intelligence and the medical physics profession-A Swedish perspective, Phys Medica. **88** (2021) 218-225.
[24] ZANCA, F., et al., Expanding the medical physicist curricular and professional programme to include Artificial Intelligence, Phys Medica. **83** (2021) 174-183.
[25] XING, L., GOETSCH, S., CAI, J., Artificial intelligence should be part of medical physics graduate program curriculum, Med Phys. **48** (2021) 1457-1460.
[26] NG, K.H., WONG, J.H.D., A clarion call to introduce artificial intelligence (AI) in postgraduate medical physics curriculum, Phys Eng Sci Med. **45** (2022) 1-2.
[27] DIAZ, O., GUIDI, G., IVASHCHENKO, O., COLGAN, N., ZANCA, F., Artificial intelligence in the medical physics community: An international survey, Phys Med. **81** (2021) 141-146.
[28] INTERNATIONAL ATOMIC ENERGY AGENCY, Postgraduate Medical Physics Academic Programmes, Training Course Series No. 56 (Rev. 1), Vienna (2021).
[29] IBRAGIMOV, B., XING, L., Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks, Med Phys. **44** (2017) 547-557.
[30] SEO, H., HUANG, C., BASSENNE, M., XIAO, R., XING, L., Modified U-Net (mU-Net) With Incorporation of Object-Dependent High Level Features for Improved Liver and Liver-Tumor Segmentation in CT Images, IEEE Trans Med Imaging. **39** (2020) 1316-1325.
[31] ISLAM, M., XING, L., A data-driven dimensionality-reduction algorithm for the exploration of patterns in biomedical data, Nature Biomedical Engineering. **5** (2021) 624-635.
[32] SHEN, L., ZHAO, W., XING, L., Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning, Nature Biomedical Engineering. **3** (2019) 880-888.
[33] LI, X., et al., Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation, IEEE Trans Neural Netw Learn Syst. **32** (2021) 523-534.
[34] MNIH, V., et al., Human-level control through deep reinforcement learning, Nature. **518** (2015) 529-533.
[35] WANG, X., et al., ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, (IEEE Conference on Computer Vision and Pattern Recognition CVPR 2017), (2017) 3462–3471.
[36] MAYO, C.S., et al., American Association of Physicists in Medicine Task Group 263: Standardizing Nomenclatures in Radiation Oncology, Int J Radiat Oncol. **100** (2018) 1057-1066.
[37] SHEN, C.Y., et al., An introduction to deep learning in medical physics: advantages, potential, and challenges, Phys Med Biol. **65** (2020) 05TR01.
[38] ZHANG, A., XING, L., ZOU, J., WU, J.C., Shifting machine learning for healthcare from development to deployment and from models to data, Nature Biomedical Engineering. **6** (2022) 1330-1345.
[39] NETHERTON, T.J., CARDENAS, C.E., RHEE, D.J., COURT, L.E., BEADLE, B.M., The Emergence of Artificial Intelligence within Radiation Oncology Treatment Planning, Oncology-Basel. **99** (2021) 124-134.

[40] CHALLEN, R., et al., Artificial intelligence, bias and clinical safety, Bmj Qual Saf. **28** (2019) 231-237.
[41] ROBERTS, M., et al., Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nat Mach Intell. **3** (2021) 199-217.
[42] MALIK, A., et al., Ten simple rules for engaging with artificial intelligence in biomedicine, Plos Comput Biol. **17** (2021) e1008531.
[43] ZOU, J., SCHIEBINGER, L., Design AI so that it's fair, Nature. **559** (2018) 324-326.
[44] SEYYED-KALANTARI, L., ZHANG, H.R., MCDERMOTT, M.B.A., CHEN, I.Y., GHASSEMI, M., Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations, Nat Med. **27** (2021) 2176-2182.
[45] LARRAZABAL, A.J., NIETO, N., PETERSON, V., MILONE, D.H., FERRANTE, E., Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis, P Natl Acad Sci USA. **117** (2020) 12592-12594.
[46] RAMESH, S., A checklist to protect human rights in artificial-intelligence research, Nature. **552** (2017) 334.
[47] WACHINGER, C., RIECKMANN, A., POLSTERL, S., Detect and correct bias in multi-site neuroimaging datasets, Med Image Anal. **67** (2021) 101879.
[48] ZHAO, Q.Y., ADELI, E., POHL, K.M., Training confounder-free deep learning models for medical applications, Nat Commun. **11** (2020) 6010.
[49] DHONT, J., WOLFS, C., VERHAEGEN, F., Automatic coronavirus disease 2019 diagnosis based on chest radiography and deep learning - Success story or dataset bias?, Med Phys. **49** (2022) 978-987.
[50] SUBBASWAMY, A., SARIA, S., From development to deployment: dataset shift, causality, and shift-stable models in health AI, Biostatistics. **21** (2020) 345-352.
[51] FINLAYSON, S.G., et al., The Clinician and Dataset Shift in Artificial Intelligence, New Engl J Med. **385** (2021) 283-286.
[52] CUI, S.N., TSENG, H.H., PAKELA, J., TEN HAKEN, R.K., EL NAQA, I., Introduction to machine and deep learning for medical physicists, Med Phys. **47** (2020) E127-E147.
[53] ZECH, J.R., et al., Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study, Plos Med. **15** (2018) e1002683.
[54] PARASURAMAN, R., MANZEY, D.H., Complacency and Bias in Human Use of Automation: An Attentional Integration, Hum Factors. **52** (2010) 381-410.
[55] KOHLI, A., JHA, S., Why CAD Failed in Mammography, J Am Coll Radiol. **15** (2018) 535-537.
[56] GAUBE, S., et al., Do as AI say: susceptibility in deployment of clinical decision-aids, Npj Digit Med. **4** (2021) 31.
[57] COLLINS, G.S., MOONS, K.G.M., Reporting of artificial intelligence prediction models, Lancet. **393** (2019) 1577-1579.
[58] CROSSNOHERE, N.L., ELSAID, M., PASKETT, J., BOSE-BRILL, S., BRIDGES, J.F.P., Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks, J Med Internet Res. **24** (2022)
[59] NORGEOT, B., et al., Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist, Nat Med. **26** (2020) 1320-1324.
[60] SOUNDERAJAH, V., et al., Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol, Bmj Open. **11** (2021)
[61] COLLINS, G.S., et al., Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence, Bmj Open. **11** (2021)

[62] EL NAQA, I., et al., AI in Medical Physics Guidelines for publication, Med Phys. **48** (2021) 4711-4714.
[63] BLUEMKE, D.A., et al., Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers - From the Radiology Editorial Board, Radiology. **294** (2020) 487-489.
[64] MONGAN, J., MOY, L., KAHN, C.E., JR., Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers, Radiol Artif Intell. **2** (2020) e200029.
[65] RIVERA, S.C., et al., Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension, Nat Med. **26** (2020) 1351-1363.
[66] LIU, X.X., et al., Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension, Nat Med. **26** (2020) 1364-1374.
[67] BUVAT, I., ORLHAC, F., The TRUE Checklist for Identifying Impactful Artificial Intelligence-Based Findings in Nuclear Medicine: Is It True? Is It Reproducible? Is It Useful? Is It Explainable?, J Nucl Med. **62** (2021) 752-754.
[68] JHA, A.K., et al., Nuclear Medicine and Artificial Intelligence: Best Practices for Evaluation (the RELAINCE Guidelines), J Nucl Med. **63** (2022) 1288-1299.
[69] DE HOND, A.A.H., et al., Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review, Npj Digit Med. **5** (2022) 2.
[70] VASEY, B., et al., DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence, Nat Med. **27** (2021) 186-187.
[71] VASEY, B., et al., Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI, Nat Med. **28** (2022) 924-933.
[72] EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, (2017).
[73] EUROPEAN COMMISSION, FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, INTERNATIONAL ATOMIC ENERGY AGENCY, INTERNATIONAL LABOUR ORGANIZATION, OECD NUCLEAR ENERGY AGENCY, PAN AMERICAN HEALTH ORGANIZATION, UNITED NATIONS ENVIRONMENT PROGRAMME, WORLD HEALTH ORGANIZATION, Radiation protection and safety of radiation sources: International basic safety standards, IAEA Safety Standards Series No. GSR Part 3, IAEA, Vienna (2014).
[74] INTERNATIONAL ATOMIC ENERGY AGENCY, Guidelines for the Certification of Clinically Qualified Medical Physicists, Training Course Series No. 71, Vienna (2021).

ABBREVIATIONS

AI	Artificial Intelligence
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CPD	Continuing Professional Development
CQMP	Clinically Qualified Medical Physicist
DL	Deep Learning
IGRT	Image Guided Radiotherapy
IMRT	Intensity Modulated Radiotherapy
ML	Machine Learning
MP	Medical Physics
MRIgRT	Magnetic Resonance Imaging guided Radiotherapy
OAR	Organ at Risk
QA	Quality Assurance
RT	Radiotherapy
VMAT	Volumetric Modulated Arc Therapy

CONTRIBUTORS TO DRAFTING AND REVIEW

Avanzo, M.	Centro di Riferimento Oncologico di Aviano, Aviano, Italy
Carrara, M.	International Atomic Energy Agency
Ciraj-Bjelac, O.	International Atomic Energy Agency
Doshi-Velez, F.	Harvard University, Cambridge (MA), USA
Gershan, V.	International Atomic Energy Agency
Heijmen, B.	Erasmus Medical Center and Erasmus University, Rotterdam, The Netherlands
Jiang, S.	University of Texas Southwestern Medical Center, Dallas (TX), USA
Nguyen, D.	University of Texas Southwestern Medical Center, Dallas (TX), USA
Seuntjens, J.	McGill University Health Centre, Montreal, Canada
Titovich, E.	International Atomic Energy Agency
Xing, L.	Stanford University, Stanford (CA), USA
van der Merwe, D.	International Atomic Energy Agency
Zwanenburg, A.	National Center for Tumour Diseases (NCT/UCC) Dresden, and OncoRay - National Center for Radiation Research in Oncology, Dresden, Germany

The IAEA wish to acknowledge Andre Dekker (The Netherlands), Maryellen L. Giger (USA), Benjamin Haibe-Kains (Belgium) and Wei Zhao (China) for their valuable comments and suggestions.

Consultants Meetings

Vienna, 19, 23, 24, 26 August and 2 September 2021

IAEA
International Atomic Energy Agency

No. 26

ORDERING LOCALLY

IAEA priced publications may be purchased from the sources listed below or from major local booksellers.

Orders for unpriced publications should be made directly to the IAEA. The contact details are given at the end of this list.

NORTH AMERICA

Bernan / Rowman & Littlefield

15250 NBN Way, Blue Ridge Summit, PA 17214, USA
Telephone: +1 800 462 6420 • Fax: +1 800 338 4550

Email: orders@rowman.com • Web site: www.rowman.com/bernan

REST OF WORLD

Please contact your preferred local supplier, or our lead distributor:

Eurospan Group

Gray's Inn House
127 Clerkenwell Road
London EC1R 5DB
United Kingdom

Trade orders and enquiries:

Telephone: +44 (0)176 760 4972 • Fax: +44 (0)176 760 1640
Email: eurospan@turpin-distribution.com

Individual orders:

www.eurospanbookstore.com/iaea

For further information:

Telephone: +44 (0)207 240 0856 • Fax: +44 (0)207 379 0609
Email: info@eurospangroup.com • Web site: www.eurospangroup.com

Orders for both priced and unpriced publications may be addressed directly to:

Marketing and Sales Unit
International Atomic Energy Agency
Vienna International Centre, PO Box 100, 1400 Vienna, Austria
Telephone: +43 1 2600 22529 or 22530 • Fax: +43 1 26007 22529
Email: sales.publications@iaea.org • Web site: www.iaea.org/publications

23-03231E

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA