Year : 2019

Image quality characterisation in computed tomography to assess the relevant diagnostic information

Viry Anaïs

Viry Anaïs, 2019, Image quality characterisation in computed tomography to assess the relevant diagnostic information

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_791599613EE34

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Institut de Radiophysique

Image quality characterisation in computed tomography to assess the relevant diagnostic information

Thèse de doctorat ès Sciences de la vie (PhD)

Présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne
par

Anaïs VIRY

M.Sc. Physique Médicale, Université Joseph Fourier, Grenoble, France
M.Sc. Ingénierie Biomédicale, Université Aix-Marseille II, France

Comité de thèse

Président du comité de thèse : Prof. Fabio MARTINON (UNIL)
Directeur de thèse : Prof. Francis R. VERDUN (CHUV-UNIL)
Co-Directeur de thèse : Dr. Fabio BECCE (CHUV-UNIL)
Expert : Prof. Mika KORTESNIEMI (HUS Medical Imaging Center, Helsinki, Finlande)
Expert : Dr. Françoise PEYRIN (Creatis, Lyon, France)
Expert : Dr Clarisse DROMAIN (CHUV-UNIL)

Lausanne 2019

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Institut de Radiophysique

Image quality characterisation in computed tomography to assess the relevant diagnostic information

Thèse de doctorat ès Sciences de la vie (PhD)

Présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne
par

Anaïs VIRY

M.Sc. Physique Médicale, Université Joseph Fourier, Grenoble, France
M.Sc. Ingénierie Biomédicale, Université Aix-Marseille II, France

Comité de thèse

Président du comité de thèse : Prof. Fabio MARTINON (UNIL)
Directeur de thèse : Prof. Francis R. VERDUN (CHUV-UNIL)
Co-Directeur de thèse : Dr. Fabio BECCE (CHUV-UNIL)
Expert : Prof. Mika KORTESNIEMI (HUS Medical Imaging Center, Helsinki, Finlande)
Expert : Dr. Françoise PEYRIN (Creatis, Lyon, France)
Expert : Dr Clarisse DROMAIN (CHUV-UNIL)

Lausanne 2019

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur Prof. Fabio **Martinon**
Directeur·trice de thèse	Monsieur Prof. Francis **Verdun**
Co-directeur·trice	Monsieur Dr Fabio **Becce**
Expert·e·s	Monsieur Prof. Mika **Kortesniemi**
	Monsieur Prof. Françoise **Peyrin**
	Madame Prof. Clarisse **Dromain**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Anaïs Viry

Master Sciences, technologie, santé* / Diplôme d'ingénieur** Université Joseph Fourier, Grenoble France* / Université d'Aix-Marseille, France**

intitulée

Image quality characterisation in computed tomography to assess the relevant diagnostic information

Lausanne, le 8 mars 2019

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Fabio Martinon

"Success is not final, failure is not fatal: it is the courage to continue that counts".

Winston S. Churchill

Content

Remerciements

Comme cela se perçoit dans la lecture de ce manuscrit, de nombreuses personnes ont contribué à ce travail et ont permis sa réalisation. J'aimerais avant tout les remercier pour leur participation, leur aide et leur soutien tout au long de ces trois années.

Tout d'abord, je tiens à remercier mon directeur de thèse, le Pr. Francis Verdun, pour m'avoir ouvert les portes de la physique en imagerie médicale en Suisse. Il a toujours su me guider dans ce travail de thèse afin de garder une ligne directrice pertinente, tout en me laissant la liberté nécessaire pour expérimenter et tester mes propres idées. Mes remerciements sont également destinés à mon co-directeur de thèse, le Dr Fabio Becce, pour toutes les explications cliniques qu'il a pu me fournir pour enrichir ce travail, et en particulier pour son implication constante dans le projet sur l'imagerie spectrale musculo-squelettique.

Je tiens également à remercier tous les co-auteurs des différents articles qui ont été rédigées pendant cette thèse, leur implication est essentielle pour améliorer la pertinence scientifique du travail. Parmi eux, je remercie particulièrement l'équipe de l'université d'Otago en Nouvelle-Zélande pour m'avoir accueilli au sein de leur hôpital afin de mieux comprendre le fonctionnement d'un scanner spectral, et de m'avoir permis d'acquérir toutes les données nécessaires à ce travail de recherche. Il est important également de souligner l'apport scientifique de mon collègue, le Dr. Damien Racine dans les différents projets de cette thèse. Son implication et nos discussions (parfois longues) pour analyser les résultats nous ont permis d'avancer bien plus rapidement et efficacement dans nos projets communs.

Je remercie mes collègues (trop nombreux pour être tous cités) pour la bonne humeur au sein de l'institut, qui rendent très agréable le travail quotidien. Je voudrais particulièrement remercier le Pr. François Bochud pour m'avoir accueillie à l'institut, ainsi que pour son expertise scientifique très appréciable dans ce travail. Sans oublier mes collègues du groupe d'imagerie médicale, Christel, Julie, Nick, Pascal et Silvano pour la bonne ambiance entre nous et nos nombreux échanges pour répondre aux questions de nos partenaires des différents centres de radiologie romands. Merci également à tous mes collègues doctorants ou ex-doctorants (Alexandre, Josh, Marie, Mathieu, Patrick, Valentin et Siria), qui ont toujours su me soutenir et qui ont largement contribué à rendre ces trois années mémorables, entre travail et activités extérieures !

Je tiens à remercier chaleureusement toute ma famille pour leur soutien et leur aide, en particulier, mes parents et mon frère, tous et toutes entrepreneurs et passionnés par leur travail. Ils m'ont montré, dès mon enfance, la détermination qu'il faut pour mener à bien nos différents projets professionnels et personnels. Je remercie également ma famille suisse pour m'avoir très bien accueilli ici et m'avoir permis de découvrir le pays d'une partie de mes origines. En dernier lieu, mes remerciements s'adressent à Charles, pour sa générosité et son soutien sans faille depuis notre rencontre.

Toutes ces personnes m'ont aidé à relever avec succès les étapes que constitue un travail de thèse. Je suis très heureuse aujourd'hui de vous présenter ce travail qui marque la liaison entre mes (longues) études et le métier de physicienne médicale.

Abstract

Computed tomography (CT) plays a major role in patient care for diagnostic and screening purposes, including therapeutic procedures. Over the last twenty years, the frequency of CT examinations has continuously increased, leading to an increase in the annual effective dose per inhabitant. This is why over the last ten years Public Health Authorities have applied strong pressure to decrease the dose levels of CT examinations. To manage this trend, two radiation principles should be carefully applied: the justification of imaging examinations, and the optimisation of radiological procedures. The principle of justification requires that "any decision that alters the radiation exposure situation should do more good than harm". The optimisation process should ensure that the radiation exposure is kept as low as reasonably achievable to enable a correct diagnosis. In response, manufacturers have proposed various technologies to reduce patient exposure and improve CT performances. In parallel, new generations of CTs, called dual-energy CT (DECT) and spectral CT, are currently focused on developing quantitative imaging to discriminate between different materials of interest based on their effective atomic number and density. In such a context, it is necessary to ensure that radiation exposure is related to the content of useful diagnostic information. Hence, to optimise CT clinical protocols or develop new applications for quantitative imaging, the characterisation of image quality in terms of diagnostic performances is essential. For this purpose, task-based image quality assessment is an interesting framework to ensure that an image fulfils its diagnostic objective. The goal of this PhD thesis was to apply task-based image quality methodologies adapted for pertinent clinical tasks to the various generations of CT (conventional CT and multi-energy CT).

This work focused on two clinical indications: abdominal and musculoskeletal CT imaging. To ensure the choice of pertinent clinical tasks for the two indications, a close collaboration with radiologists was a prerequisite. The work began by assessing a high and low-contrast detection task with mathematical observers. This work confirmed that their use was a good choice to assess a detection task and overcome the difficulties for the image quality assessment with iterative algorithms. This thesis also dealt with DECT devices installed in various radiology departments and a prototype of spectral CT for the discrimination and the quantification of various materials involved in arthropathies. From the current knowledge of task-based image quality assessment for detection tasks with iterative reconstruction techniques, adaptive task-based tools were applied to evaluate these quantitative imaging devices. For DECT, the reliability of announced quantification performances was verified and for spectral CT, the discrimination performances of this new imaging technique were assessed.

The project confirmed that the use of adapted task-based image quality methodologies is essential for assessing the diagnostic performances of the different CT generations. The pertinence of the proposed approaches was analysed to correctly characterise and optimise clinical protocols. Finally, this work demonstrates that the use of a task-based approach makes it possible to establish a bridge between medical physicists and radiologists when dealing with the optimisation of clinical protocols. The challenge that remains is to define the content of the required diagnostic information in image patients for the different clinical tasks.

Keywords : *Computed tomography, diagnostic performances, image quality, model observers, detection, discrimination, quantification*

Résumé

La tomodensitométrie (TDM) est une technique d'imagerie médicale qui joue un rôle majeur dans la prise en charge du patient, tant pour le diagnostic et le dépistage que pour réaliser des procédures thérapeutiques. La fréquence de ces examens augmente depuis 20 ans ; ce qui engendre une augmentation de la dose efficace annuelle moyenne par habitant. Les deux principes de radioprotection, la justification de chaque examen d'imagerie et l'optimisation des protocoles cliniques doivent donc être soigneusement appliqués pour juguler cette tendance. En particulier, l'optimisation des examens cliniques doit permettre d'obtenir une exposition des patients minimale tout en assurant un diagnostic fiable. Les différents constructeurs ont de ce fait proposé des évolutions technologiques majeures pour réduire les doses et améliorer les performances des scanners. En parallèle, ils se concentrent actuellement sur le développement de l'imagerie quantitative, avec l'apparition des scanners à double énergie, ou encore des scanners spectraux. Ces nouvelles techniques permettent d'identifier, de différentier et de quantifier différents matériaux d'intérêts en se basant sur leur numéro atomique effectif et leur densité. Dans ce contexte, il est nécessaire de s'assurer que l'exposition du patient permet toujours d'obtenir une information diagnostique pertinente. Ainsi, que ce soit pour optimiser un protocole clinique ou pour développer de nouvelles applications, la caractérisation de la qualité d'image est essentielle. L'évaluation de la qualité d'image basée sur la tâche clinique est un concept intéressant car il permet d'assurer qu'une image répond à une question clinique précise. Le but de ce travail de thèse est d'appliquer des méthodes d'analyse de la qualité d'image basées sur des tâches cliniques pertinentes pour les différentes générations de scanners existantes ou en développement.

Ce travail concerne deux indications cliniques majeures : l'imagerie abdominale et l'imagerie musculo-squelettique. Afin de choisir des tâches cliniques pertinentes pour chaque application, une collaboration étroite avec les radiologues est prérequis. Dans un premier temps, nous avons évalué la détection de lésions à haut et bas contraste pour différents scanners à l'aide d'observateurs mathématiques. Ce travail a confirmé que leur utilisation était pertinente pour évaluer ce type de tâche clinique. Dans un second temps, nous avons étudié des scanners à double énergie installés dans différents services de radiologie, ainsi qu'un prototype de scanner spectral pour la quantification et la discrimination de différents matériaux. A partir des connaissances actuelles sur l'évaluation d'une tâche de détection, des méthodes d'analyse de la qualité d'image spécifiques aux tâches de quantification et de discrimination ont été développées. Pour les scanners à double énergie, il s'agissait de vérifier les performances diagnostiques annoncés pour la quantification tandis que pour le scanner spectral, il s'agissait d'évaluer pour la première fois les performances diagnostiques de différenciation de matériaux impliqués dans les arthrites microcristallines.

Ce travail confirme que les méthodes d'analyse de la qualité d'image basées sur des tâches cliniques sont essentielles pour évaluer les performances diagnostiques des scanners. La pertinence des approches proposées a été analysé pour permettre une caractérisation et une optimisation des protocoles cliniques. Finalement, ce travail démontre que l'approche proposée permet de créer un lien entre les physiciens médicaux et les radiologues lors de l'optimisation des protocoles cliniques. Le défi est maintenant de déterminer les prérequis diagnostiques pour les différentes tâches cliniques.

Résumé public

Caractérisation de la qualité d'image en tomodensitométrie afin d'obtenir des examens d'imagerie à visée diagnostique.

Le scanner, aussi appelé tomodensitométrie est une technique d'imagerie médicale utilisant des rayons X qui permet d'obtenir une imagerie tridimensionnelle des structures anatomiques d'un patient. Il joue un rôle majeur dans sa prise en charge, pour le diagnostic, le dépistage ou pour guider certaines procédures chirurgicales et thérapeutiques. Le nombre d'équipements et la fréquence des examens augmentent continuellement depuis 20 ans, ceci contribue à augmenter l'exposition de la population aux rayons X. De plus, comme leur utilisation peut comporter certains risques pour le patient, il est nécessaire d'appliquer soigneusement deux principes de radioprotection. Dans un premier temps, la justification de chaque examen permet de s'assurer que le bénéfice pour le patient reste toujours supérieur aux risques engendrés. Dans un second temps, l'optimisation des examens d'imagerie doit permettre de déterminer l'exposition minimale des patients tout en assurant un diagnostic fiable. Pour ce faire, il est essentiel de caractériser la qualité d'image des examens pour différentes tâches diagnostiques, semblable à celles réalisées par le radiologue dans sa pratique quotidienne (détecter, caractériser une lésion, différentier des tissus…). Le but de ce travail de thèse est de déterminer les tâches diagnostiques les plus pertinentes, puis de développer et d'appliquer des méthodes d'analyse de la qualité d'image adaptées aux tâches définies. Ainsi, les performances diagnostiques des scanners actuels et futurs ont été évaluées pour différents protocoles cliniques.

List of figures

List of tables

List of abbreviations

ALARA	As low as reasonable achievable
ASIC	application specific integrated circuit
ATCM	Automatic tube current modulation
AUC	Area under the ROC curve
CHO	Channelized Hotelling observer
CNR	Contrast-to-noise ratio
CPP	Calcium pyrophosphate
CT	Computed tomography
CTDIvol	Volumetric computed tomography dose index
DDoG	Dense difference of gaussian
DECT	Dual energy CT
DFOV	Displayed field of view
DLCT	Dual layer CT
DRL	Diagnostic reference level
DSCT	Dual source CT
EID	Energy integrating detector
ESF	Edge spread function
FBP	Filtered back projection
FDA	Food and Drug Administration
FN	False negative
FP	False positive
FSCT	Fast switch CT
FWMH	Full width maximum height
HA	Calcium hydroxyapatite
HO	Hotelling observer
HU	Hounsfield units
ICRP	International commission of radiation protection
TN	True negative
TNF	True negative fraction
TP	True positive
TPF	True positive fraction
Zeff	Effective atomic number
IR	Iterative reconstruction
LDA	Linear discriminant analysis
LG	Laguerre Gauss
LSF	Linear spread function
MBIR	Model based iterative reconstruction
MRI	Magnetic resonance imaging
MSU	Monosodium urate
MTF	Modulation transfer function
NEMA	National Electrical Manufacturers Association
NPS	Noise power spectrum
NPV	Negative predictive value
NPW	Non prewhitening
OCP	Octacalcium phosphate
PCD	Photon counting detector
PET-CT	Positron emission tomography
PHA	Public health authorities
PMMA	Polyméthacrylate de méthyl
PPV	Positive predictive value
PSF	Point spread function
PW	Prewhitening
RMSE	Root mean square error
ROC	Receiver operating characteristic
ROI	Region of interest
SKE	Signal-known-exactly
SKS	Signal-known-statistically
SNR	Signal-to-noise ratio
SPCCT	Spectral photon counting CT
SPECT-CT	Single Photon Emission CT

1. Introduction

1.1 Framework of patient exposure with medical imaging examinations

Over the last twenty years, the number of medical imaging examinations has considerably increased for diagnostic and therapeutic procedures. In particular, more than 50 000 Computed Tomography (CT) devices were recorded over the world, and the expansion of CT concerns all western countries [1]. In Switzerland, these numbers have doubled over the last twenty years; in 2017, there were 326 CT units for about 8.5 million inhabitants. The extensive use of CT is due to its major role in patient care, whether for diagnostic (alone or coupled with nuclear medicine devices SPECT-CT, PET-CT), for screening (only in a few countries), for its role in minimally invasive procedures (percutaneous abscess drainage, percutaneous biopsy …) or for its role in therapeutic procedures (radiotherapy).

Among all the medical imaging techniques, examinations including X-rays need to be carefully monitored because of their impact on the collective radiation effective dose. From the data collected for the "Medical Radiation Exposure of the European Population" report, pie charts of the European median frequency of X-ray examinations and contribution to the collective effective radiation dose were plotted for the four main categories (plain radiography including dental procedures and mammography, CT, fluoroscopy and interventional radiology) (Figure 1)[2]. The highest contribution to the collective radiation dose was from CT examinations (61%), despite its relative low frequency (7.8 %) compared to other X-ray imaging modalities. This is because of the substantially higher radiation exposure for CT in comparison with plain radiography (typically 0.8 mSv for an adult abdominal radiography and 10 mSv for an adult abdominal CT). In Switzerland, the contribution of CT examinations to the collective radiation dose was equal to 70% for a frequency of 9.6 % in 2013, thus placing Switzerland in the third quartile of the distribution of European countries in terms of frequency of CT procedures. Moreover, the frequency of CT examinations have continuously increased from 2008 to 2013, leading to a 17% increase in the average annual effective dose per inhabitant.[3]

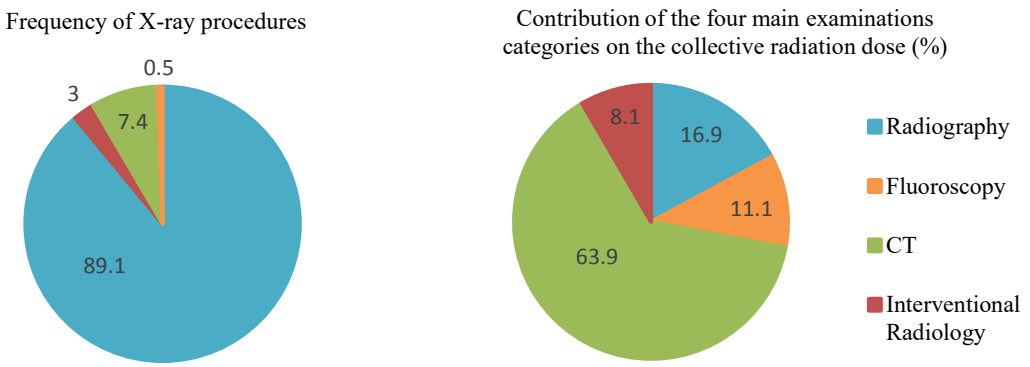


Figure 1 *: Pie charts of the European median frequency of X-ray examinations (left) and the collective radiation exposure (right) were plotted for the four main categories (plain radiography including dental procedures and mammography, CT, fluoroscopy and radiology interventional).*

To manage this trend, two radiation protection principles should be re-enforced: the justification and the optimisation of radiological procedures. In Europe, due to the high contribution on the overall collective dose, the Public Health Authorities (PHA) carefully monitor CT devices. The justification and optimisation process are the main principles of radiation protection. They are the core of ICRP publications 103 and 105, as well as the last European directive (2013/59/Euratom) [4–7]. These two principles are equally important and need to be applied together to avoid unjustified patient exposure.

The justification process should ensure that each prescribed examination is able to answer a clinical question, which help the diagnosis, taking into account the radiation exposure risk. Indeed, the major risk associated with the use of X-ray in CT is the risk of inducing cancers. Even if the benefits of CT examinations may outweigh radiation risks, it is unnecessary to use ionising radiations if other imaging techniques could provide an accurate diagnosis (ultrasound, magnetic resonance imaging, MRI). Various studies have reported a number of unjustified examinations ranging from 7 to 20% depending on the country and clinical situation [8–11]. In this context, referring medical practitioners and radiologists should be involved to ensure the justification process, building on referral guidelines developed to assist them in decision-making for appropriate healthcare [12,13]. Financial incentives may also be considered when analysing the situation.

The optimisation process should ensure that the radiation exposure is kept as low as reasonably achievable (ALARA) while allowing a correct diagnosis. So far, PHA have focussed their actions on dose reductions without taking into account the image information content. In response, manufacturers have proposed various technologies to reduce patient exposure, optimising the fluence of photons on the detectors (with automatic tube current modulation, ATCM), or extracting as much image information as possible (with iterative image reconstruction (IR) algorithms). Another aspect that has to be considered when dealing with the optimisation process is the variation of the practice even for a well-defined indication. To handle this problem, the diagnostic reference levels (DRL) were proposed by the ICRP (International Commission on Radiological Protection) in 1996 as a tool to reduce the range of radiation exposure applied in a country for a given examination[14]. Defined as the third quartile of the distribution of median values of a given dose indicator ($CTDI_{vol}$ or DLP for CT) across numerous centres for an anatomical region, the DRL should initiate actions when the dose local indicator is systematically higher than the national DRL value proposed. However, DRL have two major limitations: (1) they are not related to precise clinical indications, (2) nor with any image quality criterion. This last point is particularly critical since the necessary content of information required to allow a correct diagnosis is strongly related to image quality. Even if some countries have begun to propose updated DRL values linked to general clinical indications [15], an effort should be made to standardise the examination nomenclature and find a consensus on the number of phases for each examination (since DRL are often given for one acquisition phase). Having improved the situation concerning these medical aspects, adapted image quality metrics still need to be developed and applied to link radiation exposure with the necessary diagnostic information contained in images.

With Filtered Back Projection techniques (FBP), the simple observation of images (checking for the high content of image noise or presence of starving artefacts) could alert radiologists about a limited image quality level related to a risk of missing subtle lesions. With IR algorithms, the situation has changed, and the

detection of low image quality level is not obvious anymore. Indeed, these algorithms produce images that have very low image noise levels and where common artefacts are significantly reduced. Hence, the images "look" acceptable but might not contain the necessary information to provide a correct diagnosis. Thus, the risk of missing a pathological lesion increases when decreasing the dose levels without the presence of alerting signs. In such a context, a new paradigm needs to be found to ensure that radiation exposure is related to the content of useful diagnostic information.

Figure 2 proposes a block diagram to explain the ideal different steps of the justification and the optimisation for medical imaging examinations using X-rays, and in particular, CT examinations. It is important to note that the task-based image quality assessment is not yet a routine methodology in the current optimization process. Hence, this proposed chain of the different steps imposes a strong collaboration between referring medical practitioners, radiologists, and medical physicists.

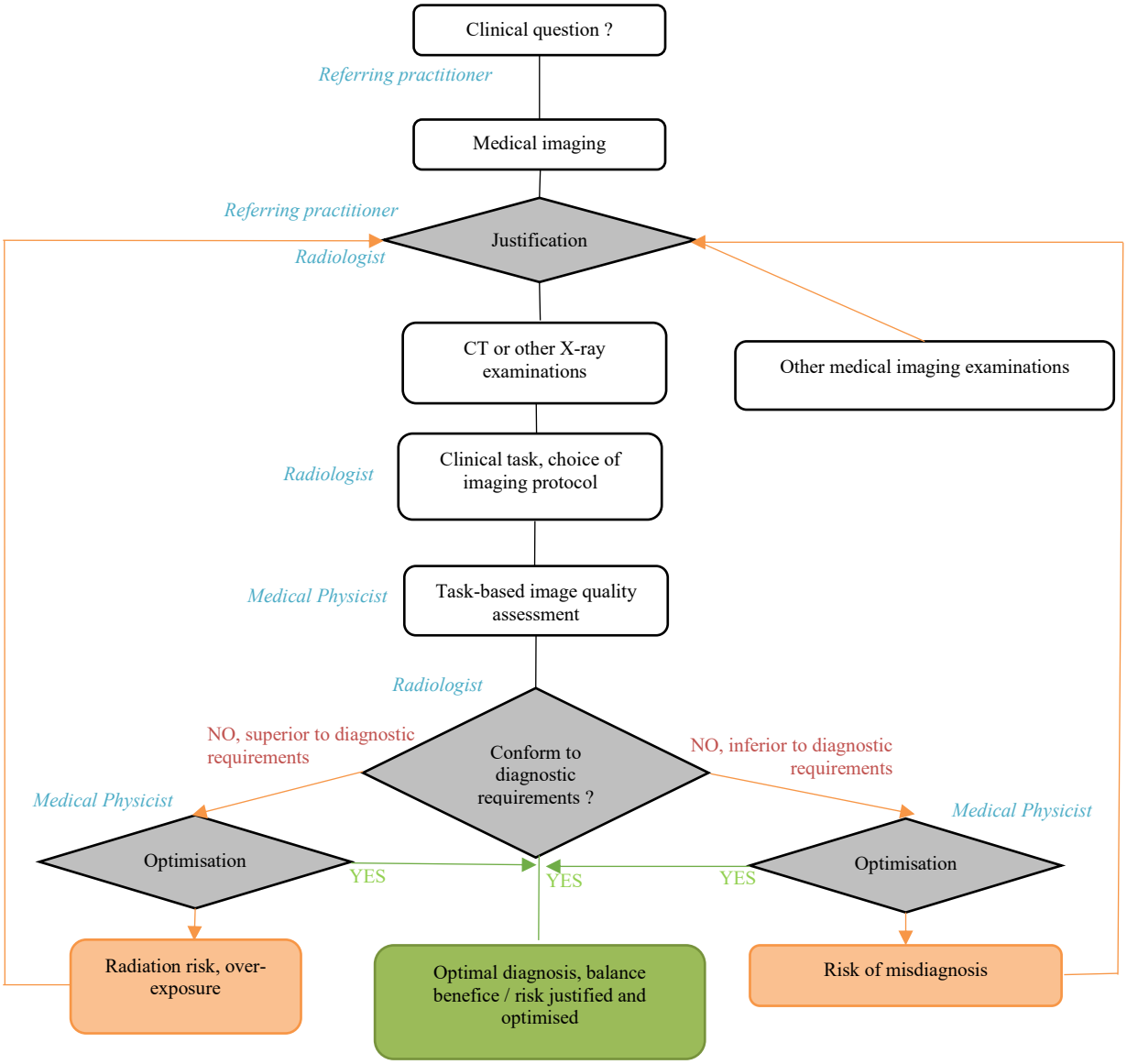


Figure 2 : *Schematic representation of justification and optimisation process for medical imaging examinations. In blue, the responsible parties for each step were indicated.*

In the last Euratom directive, roles and responsibilities of medical physicists working in the field of medical imaging are more explicit [4]. They should, in particular:

- Ensure that each CT device conforms to the local legislation (quality assurance process).
- Verify the reliability of the marketing claims concerning any technological developments.
- Ensure that image quality makes it possible to answer the clinical question with a radiation exposure as low as reasonably achievable (optimisation process).

However, the application of the directive in many European countries remains difficult and medical physicists are still mainly involved in CT quality assurance and have a minor role in the optimisation of clinical processes. In this context, the knowledge of medical physicists is required to adapt and apply clinically relevant image quality metrics.

1.2 **Major advances in conventional CT**

From the first application of the third generation of CT in the 1990's (a combined rotation of X-ray tube and detector), CT have known different axes of development to improve image quality and decrease patient exposure. The first axis of development was focused on detectors to increase signal-to-noise ratio, spatial resolution, z coverage and examination speed. The second axis of development was dedicated to the development of efficient tools to reduce radiation exposure, while trying to keep a constant "image quality level" (ATCM and IR algorithms).

1.2.1 *Automatic Tube Current Modulation*

ATCM have been proposed since 1994 to reduce radiation exposure. This idea is based on the principle that the amount of X-ray output could be adjusted as a function of patient's size and diagnostic requirements. ATCM systems can use the localizer radiograph acquired before the actual CT scan to determine patient attenuation in the three directions (x, y, z) and adjust the tube current as a function of patient attenuation. Hence, it is possible to modulate the current in the z-axis (useful for thoraco-abdominal scans for example) and in the x-y direction (useful to take into account the ellipsoidal shape of a patient's trunk). To automate tube current, manufacturers propose various strategies. The first strategy chosen by GE and Canon maintains a selected image noise level for different patient sizes. The second strategy, chosen by Philips and Siemens, maintains an overall image quality level for different patient sizes, allowing a higher noise level (lower tube current in comparison to the first strategy) for larger patients to avoid over-exposure [16]. The two strategies are summarized in Figure 3. It is interesting to notice that the "image quality level" mentioned in this context is either a simple image quality metric (noise standard deviation) or the use of a reference image that looks adequate, which is difficult to link with any clinical criteria.

Several authors showed a potential of dose reduction from 15 to 53% with ATCM in comparison with constant tube current [17–19]. However, none of the strategies controls whether a minimal amount of diagnostic information is included in images for each patient size. This lack of gold standard leads to the necessity of

optimising clinical protocols that take into account the diagnostic requirements for different patient's body habitus.

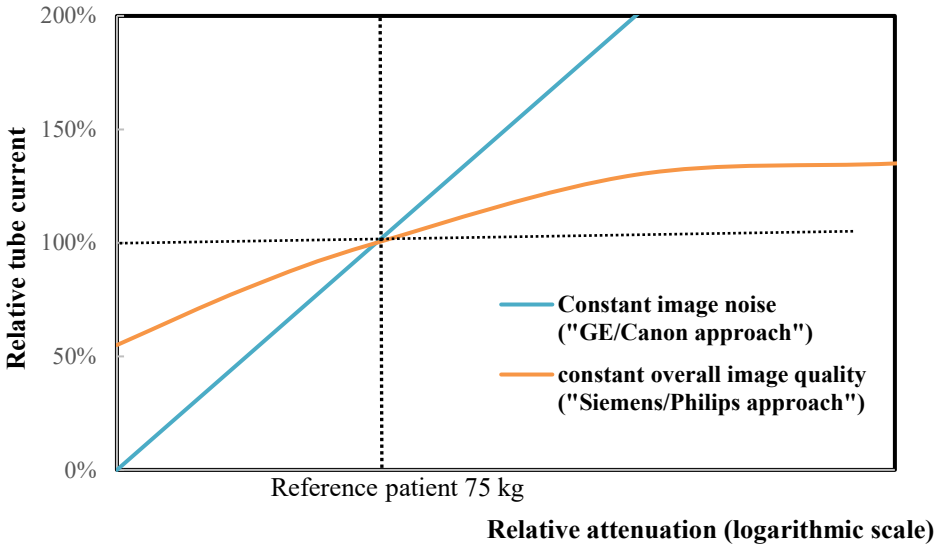


Figure 3 : *Adaptation of the tube current for different patient sizes for the two strategies used in routine. The figure represents only the tendency for the two strategies and is not based on actual measurements.*

1.2.2 *Iterative reconstruction algorithms*

Historically, CT images were reconstructed using Filtered Back Projection, a simple and fast algorithm, which required many pre-processing operations to correct for cone beam geometry, scatter process, and detector's responses. For each projection, data are back-projected and filtered in one single step. However, for heavy patients or low-dose protocols, images contain many artefacts and large image noise levels that might hide important image features. Hence, the concept of IR algorithms were initially introduced in 1970's to generate images with lower noise levels and artefacts [20]. However, this technique was not introduced in the clinical world due to the lack of computational power.

In 2008, manufacturers developed hybrid IR algorithms, so-called Statistical IR, working iteratively in raw-data space and/or image space to reduce noise level. Statistical noise model is included in the reconstruction process. Avoiding multiple projection/back-projection operations between raw-data space and image space, the speed of image reconstruction is preserved. In 2012, a second generation of iterative algorithms, so-called model-based iterative reconstruction (MBIR) was developed. MBIR can operate few projection/back-projection operations and include supplemental models in the calculation process in comparison with statistical IR. The complexity of models integrated in commercially algorithms influences the reconstruction image speed. Thus, manufacturers propose fast versions or more time-consuming versions. In particular, GE proposed a fully iterative reconstruction algorithm, VEO, which takes into account the data statistics, the geometry, X-ray interaction physics and the optic characteristics of the CT in the reconstruction process [21]. Then, in each iteration, VEO works in raw-data space and image space using successive back-projection/projection operations. Due to the high computational demand, the reconstruction time is longer than

hybrid IR or partial version of MBIR (about 30 minutes for each series of 100 images). Therefore, VEO reconstruction can be difficult to manage in a clinical workflow.

In Figure 4, a schematic outlines the basic process of the different generations of IR algorithms. Since each manufacturer proposed a different version, its exact calculation is always proprietary and limited information is available in the scientific literature. However, based on the collected data from various articles, a classification between the three categories (hybrid IR, model-based IR and full IR) are proposed for each commercially IR in Table 1 [22–25].

Various scientific articles have shown that IR algorithms can achieve major noise reduction while maintaining spatial resolution and also reduce artefacts (beam hardening, metal artefacts …) [26,27]. In particular, model-based IR can offer a higher noise reduction in comparison to statistical IR. A potential limitation of IR algorithms is that the noise texture is modified in comparison to FBP algorithm, with a shift toward the low frequencies of the Noise Power Spectrum (NPS) peak.

Considering only the noise reduction, manufacturers have claimed a dose reduction ranging from 20% to 80% for the different commercially available IR algorithms [28]. However, even if IR algorithms can improve quantitative image quality metrics (such as the noise, the contrast-to-noise ratio, or the spatial resolution) in comparison with FBP, it is important to note that no gold standard exists in terms of diagnostic performances for different body habitus and anatomical locations. Moreover, as opposed to standard FBP algorithms, IR algorithms are no longer linear and cannot be considered stationary anymore. Thus, common image quality metrics used by medical physicists have been updated (modulation transfer function, noise power spectrum) or abandoned (contrast-to-noise ratio). [26].

Nowadays, the general use of ATCM and IR algorithms reinforces the need for a good collaboration between medical physicists and radiologists to establish links between standardised image quality metrics and relevant diagnostic information.

CT	**Hybrid IR**		**Model-Based IR**	
	Image space	Image / Raw-data space	Partial MBIR	Full IR
GE	-	ASiR (2008)	ASiR-V (2014)	VEO (2011)
Phillips	-	iDose (2011)	IMR (2013)	-
Siemens	IRIS (2010)	SAFIRE (2011)	ADMIRE (2014)	-
Toshiba / Canon	QDS	AIDR (2012)	FIRST (2016)	-

Table 1 : *Classification of the different commercial IR algorithms. The year of release of each algorithm was indicated in parentheses.*

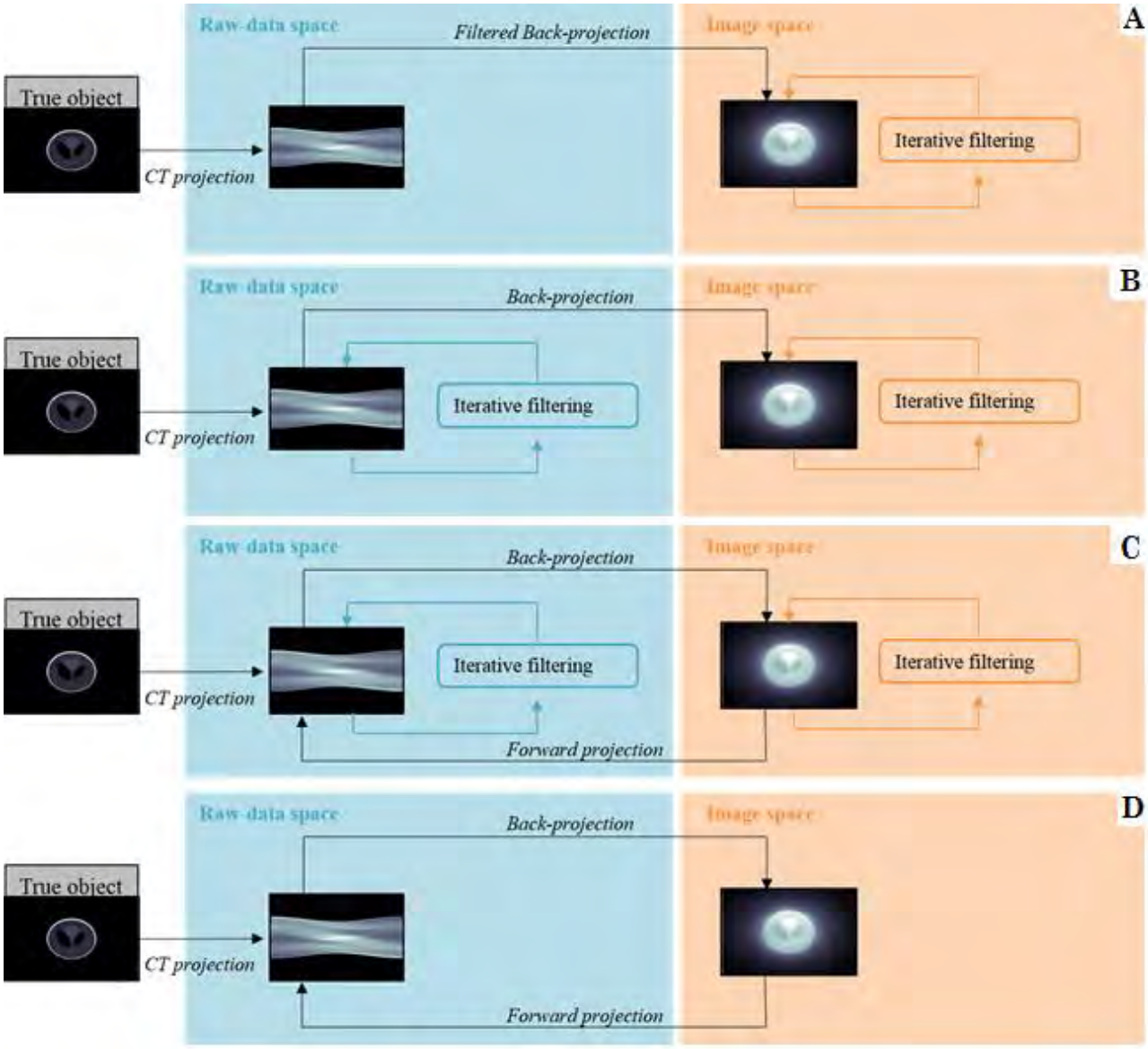


Figure 4 : *Basic process of the different generations of IR algorithms: (A) hybrid IR working in image space, (B) hybrid IR working in raw-data and image space, (C) partial model-based IR, (D) Full IR.*

1.3 **Major advances in the field of quantitative CT imaging**

In parallel with the development of conventional CT (detector improvement and dose reduction techniques), another axis of development was dedicated to material discrimination in CT. In fact, a remaining issue in conventional CT is that two materials (with different elemental compositions) can produce a similar pixel value (similar linear attenuation coefficient). For example, the classification between contrast medium and bone can be very challenging in CT (see Figure 5). Indeed, the linear attenuation coefficient for a given material depends on its effective atomic number (Z_{eff}) number (i.e its elemental composition) and its density at a given photon energy [29]. The concept of spectral imaging or multi-energy imaging was first introduced in 1976 by Godfrey Hounsfield [30,31]. Its principle is based on two physical properties, illustrated in Figure 6:

- X-ray attenuation of low Z_{eff} materials (water, soft tissues) is strongly related to Compton interactions, and is independent of photon energy (in the usual limited CT energy range used).
- X-ray attenuation of high Z_{eff} materials (bone, contrast agent …) is strongly related to photoelectric interactions and is strongly dependent on photon energy.

Hence, being able to image at various photon energies potentially allow material discrimination based on their Z_{eff} number and/or concentration. So far, the clinical applications of dual-energy CT (DECT) are based on these properties (use of the Compton/photoelectric interactions). Nowadays, multi-energy CT are under development aiming at replacing energy integrating detectors (EID) by photon-counting detectors (PCD).

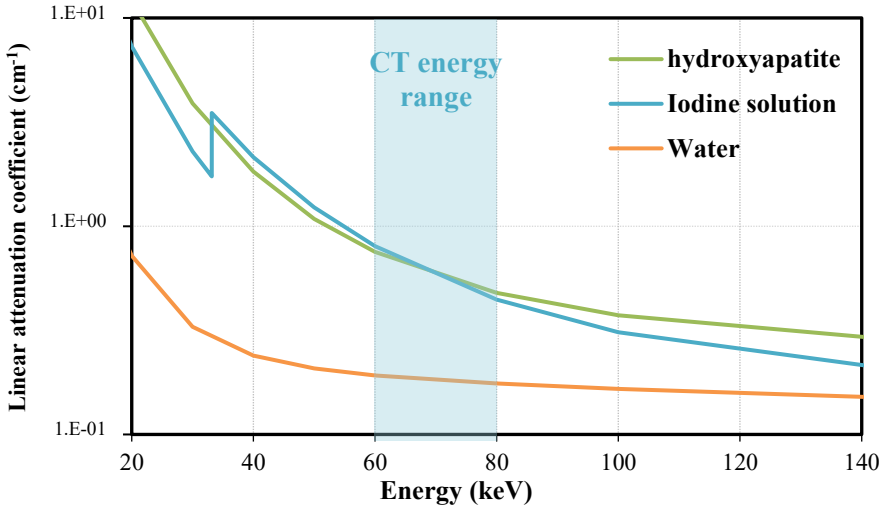


Figure 5 : *Linear attenuation coefficients as a function of photon energy for hydroxyapatite (present in cortical bone, assuming a density of 1.8 g/cm³), for water (assuming a density of 1 g/cm³), and for an iodine solution (a 10% dilution in physiological serum, assuming a density of 1.2 g/cm³). Data computed from the National Institute of Standards database.*

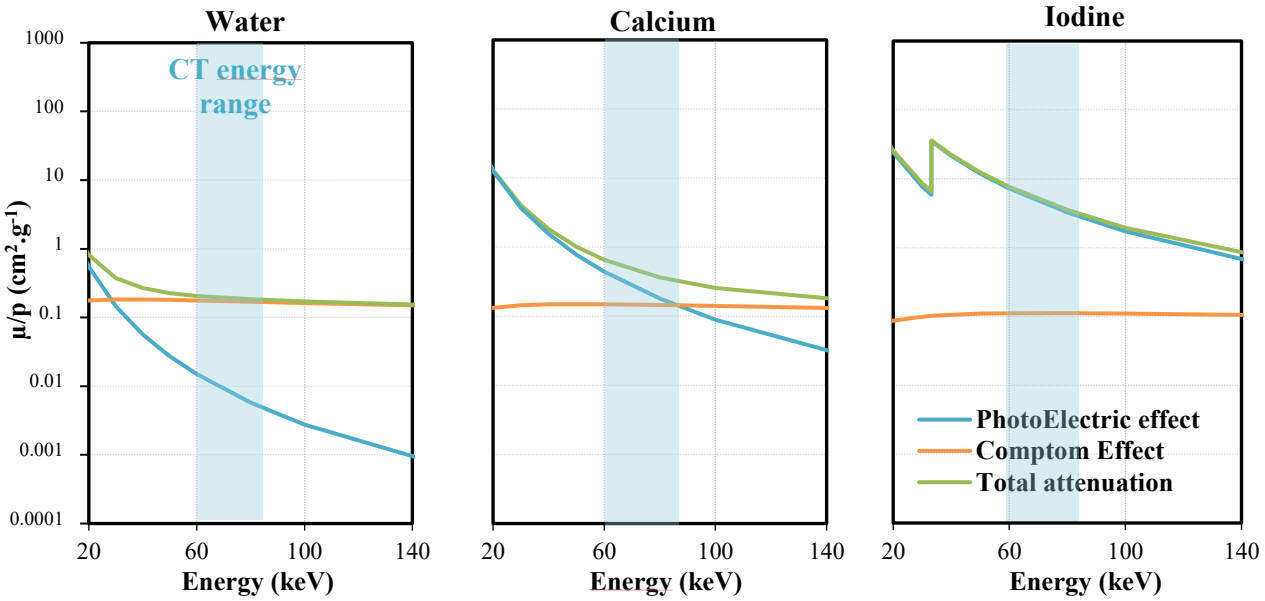


Figure 6 : *Mass attenuation coefficient as a function of energy range for three materials: water, iodine and calcium. Data computed from the National Institute of Standards database.*

1.3.1 *Dual-energy CT*

An ideal DECT should use two monochromatic beams with an identical number of photons and acquire the data at the same time. If new X-ray sources (such as synchrotron facilities) can provide high intensity beams with narrow energy bandwidths, their implementation in the clinical world is not possible. Therefore, DECT systems are still based on the use of polychromatic X-ray sources; sources that are prone to beam hardening (change in effective energy, which is material dependent) and beam energy separation. Manufacturers have proposed various strategies to take advantage of the potential benefit of DECT. They all have their own advantages and limitations. Details can be found in Table 2 [32,33].

DECT techniques	Manufacturer	Principle	Advantages	Limitations
Sequential scans	All	Two consecutive scans (high and low kVp)	Optimise spectral separation using different filtrations	Patient motion between two scans Projection not paired
Dual-Source CT (DSCT)	Siemens	CT with two X-ray tubes (high and low kVp)	ATCM available Optimise spectral separation using different filtrations	Cross-talk of scattered radiation between both tube-detector couples: necessity to apply a scatter-correction algorithm Projection data not paired
Fast-Switch CT (FSCT)	GE	Switch the tube potential between two consecutive views	Projection data paired	ATCM not available Not possible to optimise filtrations to increase the spectral separation Difficult to obtain the same number of photons at detector level
Dual-Layer CT (DLCT)	Phillips	CT with a single tube potential beam and layered detectors	Projection data paired ATCM available DECT results available for each examination	Difficult to obtain the same number of photons at detector level
Split-Filter dual energy CT	Siemens	An X-ray tube with a tin and gold filtration	Projection data paired	Not possible to optimise tube voltage to increase the spectral separation

Table 2: *Advantages and limitations for the different DECT techniques used in clinic*

An important parameter for DECT systems to consider is the separation between the two spectra. Indeed, the larger the spectral separation is, the higher the signal-to-noise ratio will be, leading to a better material discrimination.

Combining the information obtained at 80 and 140 kVp (effective energy of approximately 40 and 70 keV), one can obtain a high-contrast between two materials based on their Z_{eff} and density. In order to illustrate this property, it is common to use a 2D graph to represent CT numbers for various materials at low and high energy (Figure 7) [32]. For low Z_{eff} materials (water, fat, muscle …), CT numbers are constant at 80 kVp and 140 kVp, and points are aligned on the bisecting line of the aforementioned graph. For high Z_{eff} materials, due to the contribution of photoelectric effect, CT numbers at 80 kVp are higher than CT numbers at 140 kVp, points are positioned above the bisecting line. The higher the effective Z, the higher the angle is between the bisecting line and the point corresponding to the material.

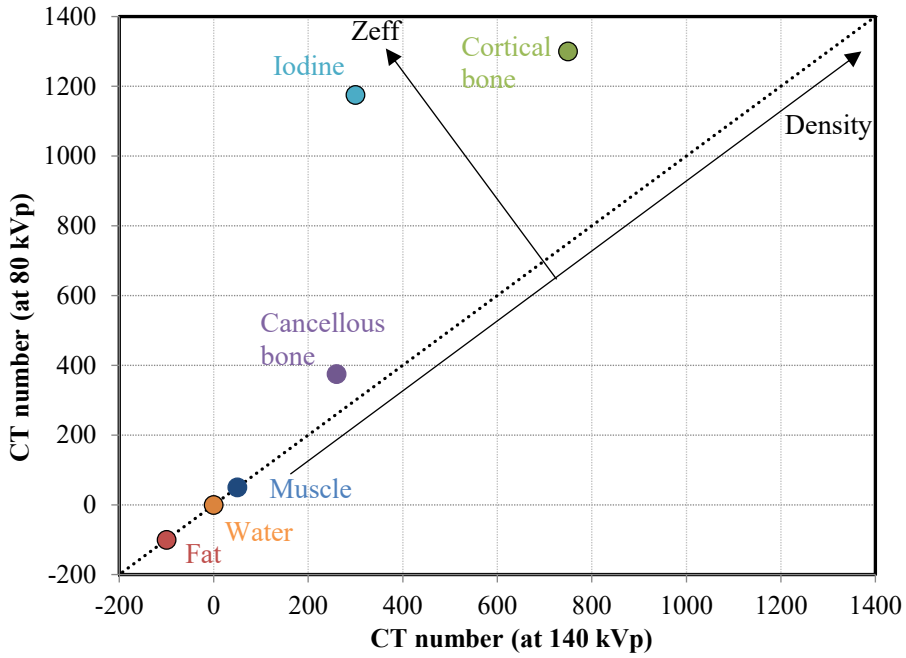


Figure 7 : *CT numbers at 80 and 140 kVp for different materials, measured on patient images scanned with a DECT protocol*

In the simplest image formation model, the attenuation *A* for a specific material depends on the energy and can be expressed using equation 1:

$$A = \int_E w(E) e^{-\int_l \mu(E,l)dl} dE \qquad \text{[Eq 1]}$$

Where w(E) is an energy-dependent factor specific to the system configuration (X-ray tube and detector) and μ, the energy-dependent attenuation coefficient for the material. dl denotes the "line integral" function along the line l described by the incident ray. μ can be expressed as a linear combination with Compton scattering (SC) and photoelectric effect (PH) (equation 2).

$$\mu(E,l) = \mu_{PH}(l,E) + \mu_{SC}(l,E) \qquad \text{[Eq 2]}$$

We can express the previous expression as a combination between energy-dependent factors (f_{PH} and f_{SC}) material-dependent factors (a_{PH} and a_{SC}) for Compton scattering and the photoelectric effect (equation 3).

$$\mu(E,l) \sim f_{PH}(E)a_{PH}(l) + f_{SC}(E)a_{SC}(l) \quad \text{[Eq 3]}$$

As energy-dependent factors f_{PH} and f_{SC} are known, it is necessary to perform the measurements at two energy levels to solve the equation and determine a_{PH} and a_{SC} for various materials (means energies produced at 80 kVp and 140 kVp). Hence, it is possible to reconstruct two maps, one for the contribution of the photoelectric effect and one for the contribution of the Compton scattering effect. This is the first step before reconstructing the various decomposition maps required for the clinical applications. The post-processing operations can be classified into two categories: material-specific maps (virtual non contrast images, iodine map, material differentiation images) based on a material decomposition algorithm and non-material-specific map (Z_{eff} map, virtual mono-energetic images, electron density map) [33]. Post-processing is an important part for DECT techniques and includes many possibilities to qualitatively visualize and quantitatively analyse images. If it is well described, it is still time-consuming to reconstruct images and interpret them (due to the large amount of information available in comparison to the standard single-energy CT).

Initially used between 2006 and 2009 to characterise non-calcified atherosclerotic plaques in CT angiography, today DECT offers various clinical applications. In abdominal imaging, monochromatic images and/or virtual unenhanced images are used to better visualize and characterise liver lesions, pancreatic adenocarcinomas and renal masses. DECT can also be useful for diagnosing urinary calculi, and is able to discriminate between uric acid and non-uric acid calculi [34]. In lung imaging, the DECT ability to quantify the specific tissue content with contrast medium is very useful for the diagnosis of acute or chronic pulmonary embolism [35,36]. In musculoskeletal imaging, DECT applications are based on the characterisation and quantification of various materials, such as calcium, iodine, mono-sodium urate or iron. For example, virtual non-calcium CT can help for the detection of bone marrow lesions. The characterisation of mono-sodium urate in DECT is very useful to confirm the diagnosis of crystal-induced arthropathies, such as gout. Moreover, the quantification of mono-sodium urate could also serve as a tool to monitor the disease. Another major application of DECT is the ability to reduce metal artefacts using high mono-energetic images [37].

For the various clinical applications, different post-processing algorithms are essential to ensure a diagnosis (material decomposition, virtual unenhanced images, monochromatic images …). As for conventional CT, it is necessary to verify the reliability of post-processing algorithms in performing a diagnosis.

1.3.2 *Multi-energy spectral CT*

DECT techniques can provide tissue specific images and various clinical applications have been proposed. However, the different DECT techniques (see Table 2) have two main disadvantages: (1) The

separation between the two spectra is limited, (2) material decomposition algorithms are limited to two basis material [38]. In order to discriminate more than two materials (for example, soft tissues and two contrast media), it is necessary to expand the number of basis functions (to take into account the k-edge discontinuities). In that case, the number of independent measurements must be strictly superior to two [39]. Hence, PCD have been recently developed for medical X-ray applications. PCD improve the major limitations of EID usually installed on CT devices and offer this possibility of discriminating between more than two materials. Indeed, by integrating the energy response, EID logically lose photon energy information and add electronic noise (see Figure 8). PCD can count the number of photons and the pulse height can be relied to the energy of each photon by comparing them with given energy thresholds [40].

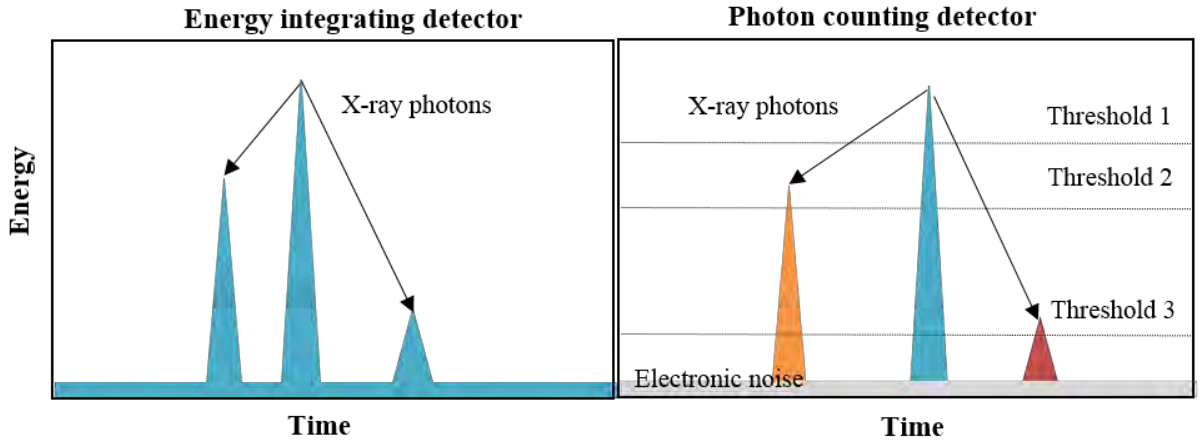


Figure 8 : *Schematic response of an EID and PCD. For EID, the number of photons is indirectly estimated from the area under the blue curve. For PCD, the detector counts the number of photons and pulse height is compared to the given threshold to discriminate the photons' energy. By applying a threshold sufficiently high above the noise the detector records only the events originating from ionising radiation interacting with the sensor.*

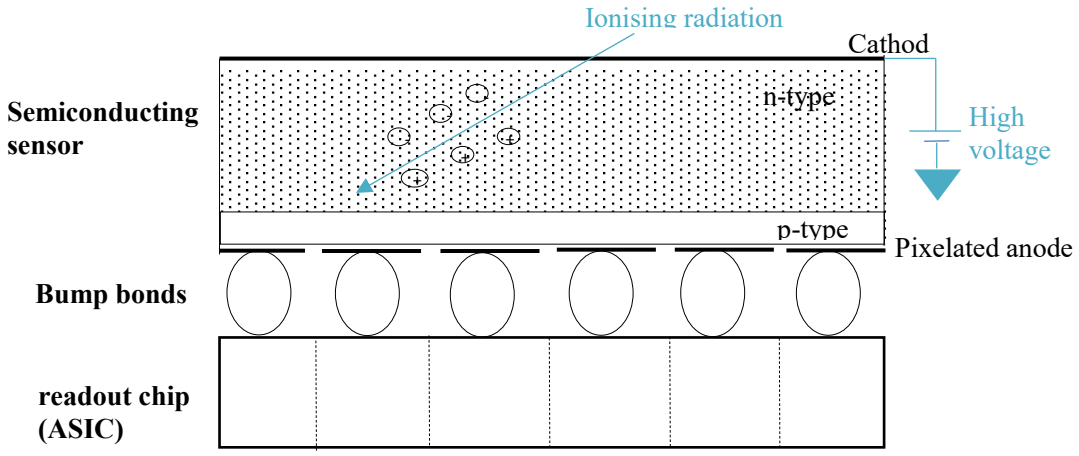


Figure 9: *Simplified 2-dimensional view of the architecture of a photon counting hybrid pixel detector.*

PCD consists of semi-conductor sensors combined in a 2D pixel matrix. Each pixel in the pixelated anode has a corresponding pixel on the application specific integrated circuits (ASIC) readout chip via micro bumps (see Figure 9) [41]. The detector is also called a hybrid pixel detector. One advantage is being able to

choose the semi-conducting material depending on the application. Cadmium-telluride, cadmium-zinc-telluride and silicon could be used for medical imaging applications. The signal received from the sensor is amplified and filtered with a pulse shaper to improve the signal-to-noise ratio. Then, the pulse height of the output signal is compared to implementing adjustable thresholds with a multi-level comparator. Each comparator is followed with a counter. At the output of the readout chip, the energy and the number of photons are stored.

PCD are promising and could offer various clinical applications in medical X-ray imaging with their ability to directly count each incoming photon and classify them according to their energy. For multi-energy CT, detectors should be able to handle high count rates with a high efficiency. X-ray beam shaping filters, composed with a classical bowtie filter and a dynamical part should be used to decrease the count rate and optimise the intensity and energy spectrum. Multiple stack layers have also been proposed to decrease the count rates that each layer need to handle and use the depth interaction to obtain energy information [42]. However, PCD has different limitations that can slow its introduction in clinical routine. Firstly, the charge sharing effect occurs when a photon is counted by two adjacent pixels due to the diffusion of the secondary charge in the semiconductor. Hence, there is a loss of energy information and a distortion of the energy spectrum. This phenomenon can be avoided by using larger pixel sizes (decreasing spatial resolution) or by implementing a charge summing between adjacent pixels (square of 2x2 pixels) to detect coincidence and attribute the charge to a single pixel. Secondly, the pile-up effect occurs when two consecutive photons are counted at the same time, yielding to a loss of counts and an incorrect energy registration. This effect is always present and depends on the count rates and dead time of the detector. This phenomenon can be limited by decreasing the pixel size or by shortening the peaking time (the time for an amplified pulse to reach its maximum). Thirdly, with K-escape X-ray and Compton scattering, X-ray fluorescence or any secondary photons can be registered by the same pixel as two separate events or an adjacent pixel, yielding to an incorrect record of the energy value. The effect can be limited by extending the peaking time to collect all the charges produced by a given photon. Hence, pile up effect, charge sharing, K-escape X-ray and Compton scattering effects could lead to conflicting detector configurations (pixel size, peaking time) [43]. Despite this, in order to correct the energy spectrum distortion, a correction model should be included in the reconstruction process [44,45] . It is interesting to note that energy distortion and the pile up effect have also been present for a long time in nuclear medicine; these both skew the imaging process.

Currently, at least four prototypes of multi-energy spectral CT with a photon-counting detector have been reported in the literature: First, a GE CT system, equipping a LightSpeed model with a first generation PCD showed its first clinical images in 2009 [46] . However, so far, no other publications were found regarding this prototype. The number of energy thresholds of this PCD is limited to two. Second, a Siemens CT system, equipping a dual source CT SOMATOM Definition Flash with an EID and PCD was proposed. The PCD is composed of a cadmium-telluride semiconductor, and the number of energy thresholds is effectively up to four. It is the only prototype able to directly compare an EID and PCD performances during the same acquisition [47]. Third, Philips proposed a small field of view SPCCT system (spectral photon counting CT) equipped with a cadmium-zinc-telluride semiconductor and allowing up to six different energy thresholds [48].

Finally, the MARS CT collaboration (New Zealand) proposed a small-animal SPCCT equipped with a Medipix3RX chip developed by CERN and enabling a discrimination between eight energy bins [49–51]. The two last SPCCT systems have announced the availability of a large bore SPCCT, suitable for human imaging in a near future [52].

The capacity to identify various materials makes spectral CT very promising. So far, the major application is the development of K-edge imaging. Indeed, the use of various materials having a K-edge discontinuity in a relevant energy range (that could be used in spectral imaging for medical applications) can enlarge the application of spectral CT possibilities for diagnostic or therapeutic purpose [53]. Indeed, due to the absorption of lower energy photons, the use of iodine (K edge 33 keV) for K-edge imaging is challenging. Other contrast media or nano-particles could be employed such as barium (K-edge 37.4 keV), gadolinium (K-edge 50 keV), gold (K-edge 80.6 keV) or bismuth (K-edge 90.5keV). However, toxicity issues need to be considered. The range of clinical applications using high Z-materials being wide [54–57], it could play a role for atherosclerotic plaque imaging [58], monitoring biomarkers for cancer [59], or assessing the treatment response for various inflammatory diseases. Spectral imaging can also improve soft tissue contrast (lipid, water …) [60] and material discrimination for musculoskeletal imaging [61]. As for DECT, post-processing algorithms are essential to display a final image. Various possibilities exist: synthetic CT numbers, energy bin images, energy weighted images, material decomposition images, or synthetic mono-energetic images. The choice of the display method depends on the imaging goal, for example: enhancing a specific material; quantifying a concentration or maximizing the contrast for a known or unknown target. As for DECT, it is necessary to link image quality parameters of the different reconstruction approaches to diagnostic performances in order to correctly assess spectral imaging possibilities.

1.4 **Characterisation of image quality in CT: a challenging task**

With the latest technological developments of CT (for dose reduction or quantitative imaging), the characterisation of image quality is essential to ensure several outcomes: conformity to the legislation, reliability of the latest technological developments in terms of diagnostic performances, and optimisation of clinical practice. However, assessing image quality in medical imaging remains a challenging task. In 1991, Fryback and Thornbury proposed a six-level scale, ranging from the technical to the societal efficacy in order to assess the societal cost and benefits of a diagnostic imaging technique [62]. Image quality metrics based on the technical efficacy (spatial resolution, image noise, contrast to noise ratio) are used to ensure that each CT device conforms to manufacturer's claims and legal requirements. In Switzerland, even if conformity to legislation is verified directly by the manufacturers, a validation by the medical physicist is still important to ensure a high safety level of the medical device (and to eliminate the risk of being both "judge and judged"). However, it is important to note that legal requirements for CT are only given for conventional CT. For dual-energy specificities, adapted metrics for quality assurance need to be proposed and introduced in recommendations and national legislations [63].

In order to optimise the clinical protocols, the use of basic image quality metrics are of limited interest because they are not related to any clinical requirement in term of image quality. Historically, an improvement of these metrics was associated to better diagnostic performances. For example, contrast-to-noise ratio could be used, keeping the reconstruction kernel constant, to infer the low-contrast detectability in a white noise background, using the Rose Model[64]. In the past, when iterative reconstruction techniques (IR) were introduced, methods to assess the diagnostic accuracy were not sufficiently developed. This led to an overestimation of the potential dose reduction because the estimation was only based on the improvement of the contrast-to-noise ratio [65]. However, the modification of noise level and texture with IR algorithms implies a significant increase of the contrast-to-noise ratio without necessarily improving the detection of low-contrast structures[66]. The standard image quality metrics being not adequate anymore, it is necessary to try to go a step further on the six-level scale and assess the diagnostic accuracy, initially defined as the measure of the specificity and sensitivity.

Barrett and Myers proposed a framework to quantify if "the images fulfil their diagnostic purpose", called task-based image quality assessment [67]. This methodology can extract the desired information from an image for clinically relevant goals. To apply the methodology, a strong collaboration between the radiologist and the medical physicist should ensure that for a given CT examination, the clinical indication can be associated to a clinical task. In medical imaging, two types of tasks can be distinguished: a classification task or an estimation task [67,68]. The classification task is a distinction between several categories (for example a distinction between normal and abnormal case). Signal detection, pattern recognition, or material discrimination are examples of classification tasks that could be performed in medical imaging. An estimation task is a quantification of a parameter in an image (size, density, shape...). For a particular diagnostic case, it can happen that classification task and estimation task are combined. In the framework of task-based image quality assessment, four key elements should be defined:

- A task associated with the clinical indication: a classification or/and an estimation task.
- The statistical properties of signal and backgrounds.
- The observer: human observers or algorithms can perform the specified task.
- A figure of merit to quantify how the observer can perform the specified task.

The gold standard for assessing diagnostic accuracy are the clinical trials studies. These studies involve several radiologists reading a statistically sufficient number of real cases where ground truth (normal or abnormal case) is known. However, such studies have several limitations: first, the ground truth is generally difficult to obtain (biopsy and anatomic pathology analysis). Second, such studies are time-consuming, non-objective, non-reproducible and involve high financial costs due to the number of cases and readers necessary to obtain statistically significant results. The complexity of new CT devices, which always involve a higher number of parameters, make these studies even more complicated to conduct. To overcome these limitations, alternative studies, which replace one or several elements of the task-based image quality framework by a surrogate, could be performed. For example, using phantom-based cases where the ground truth is known or using mathematical observers instead of human observers could be a solution to easily and speedily assess diagnostic accuracy.

Mathematical observers are computer algorithms useful for task-based image quality assessment. Several investigators have started to use task-based image quality metrics to evaluate a detection task, when willing to optimise their clinical protocols using IR algorithms [69–73]. The Food and Drug Administration (FDA) in the United States even promotes the use of model observers to check the veracity of the dose reduction claims[74], whilst a standardised phantom has been proposed by the National Electrical Manufacturers Association (NEMA) to assess image quality.

Even if many efforts are currently focused toward dose optimisation with IR algorithms, the emergence of dual-energy and spectral techniques imposes the adaptation of image quality methodologies. The multi-energy CT diagnostic performances can be easily associated to two clinical tasks: a classification task to discriminate between various materials (iodine, calcium, uric acid …) and an estimation task with the quantification of the concentration and Z_{eff} for various components in the human body. From the current state of knowledge of task-based image quality assessment for detection tasks with IR techniques, it is necessary to develop adaptive task-based tools to evaluate quantitative imaging CT devices (DECT and SPCCT). For DECT, task-based image quality assessment could verify the reliability of diagnostic performances announced and adapt the post-processing for a specific clinical indication (for example, choice of the "best" virtual mono-energetic images). For SPCCT, we need to anticipate the user's needs in terms of objective characterisation of image quality, in order to correctly evaluate the clinical potential of a new imaging technique.

2. Goal of the PhD thesis

The goal of this PhD thesis was to develop tools and metrics to characterise image quality in CT linked to various clinical questions in order to assess the relevant diagnostic information. The first part will be devoted to applying mathematical observers and analysing the pertinence of the approach to characterise and optimise clinical protocols.

Two situations will be analysed, a high and low-contrast detectability for conventional abdominal CT imaging. With the introduction of quantitative imaging with Dual-Energy CT and Multi-energy spectral CT, new clinical questions are emerging and adaptive tools to assess diagnostic image quality need to be developed. Hence, the second part will be devoted to assessing the quantification of various materials in Dual-Energy CT and the third part will focus on assessing a discrimination task between various crystals involved in arthropathies with a multi-energy spectral CT prototype.

Finally, this work proposes to build a bridge between the medical physicists who are responsible for assessing image quality in CT and the radiologists who are responsible for ensuring the CT diagnostic performances for various clinical questions.

These objectives are summarized in Figure 10.

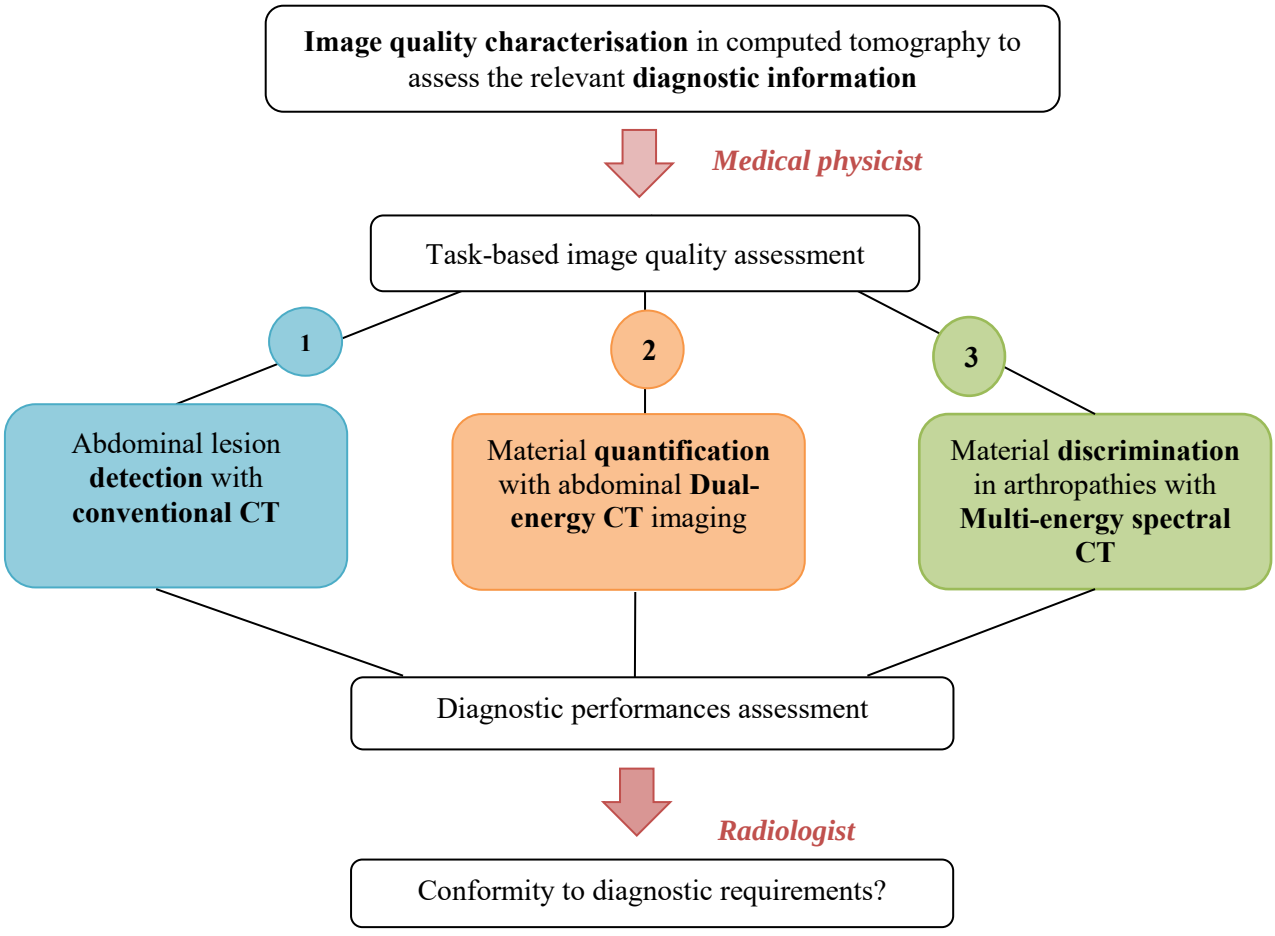


Figure 10: *Schematic representation of the three thesis milestones*

3. Methodologies for image quality assessment in CT

This section will detail the methodologies applied in the thesis for image quality assessment. The first part will be devoted to describing the physical metrics used to assess the technical efficacy of a CT system (first level on the six-level scale). The second part will describe task-based image quality methodologies, developed to assess the diagnostic accuracy of a clinical protocol (second level on the six-level scale). The various methodologies will be applied for the three CT modalities (single-energy CT, dual-energy CT and multi-energy spectral CT).

In medical imaging, image quality can be assessed in the image domain or in its corresponding spatial frequency domain. Working in the image domain is the most intuitive manner because this is the classic representation of medical images. In the frequency domain, the representation is more complex, but only one analysis is necessary to determine the input/output relationship through the CT system for different object sizes.

3.1 Image quality metric for technology efficacy assessment

3.1.1 *Physical metrics in image space*

In CT, three image quality parameters can affect the detection of specific structures: noise, contrast and spatial resolution (image blur). While the detection of large structures roughly depends on the contrast-to-noise ratio, spatial resolution is the critical parameter for the detection of small structures.

Noise: Noise is caused by random variations of Hounsfield Units (HU) voxel from another. This phenomenon has various sources: quantum noise (statistical fluctuation of photon generation and detection), fixed pattern noise (including anatomical noise), and electronic noise (added by detector electronics). Noise can be estimated by computing the standard deviation σ of pixel values in a homogeneous region of interest (ROI) in an image (Equation 4, with x_i the pixel values and $\bar{x}$ the mean pixel values in the ROI).

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{[Eq4]}$$

The principal limitation of standard deviation to estimate noise is that two images could have equal variances but different textures. This phenomenon can appear when comparing various IR algorithms or different kernels. Figure 11 shows two noise images with equal variances. Noise texture is generally assessed in the frequency space using the Noise Power Spectrum (NPS) or more rarely in the image space with the auto-covariance of pixel values.

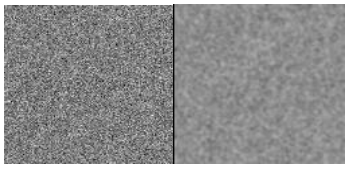


***Figure 11**: Two simulated noise images with equal variances but different textures.*

Contrast: Contrast is defined as the absolute difference between the mean pixel values in a ROI positioned in the object and the mean pixel values in another ROI in the image background (Equation 5). From noise and contrast, it is possible to calculate the contrast-to-noise ratio (CNR) (Equation 4).

$$C = |\overline{x_{ROI1}} - \overline{x_{ROI2}}| \quad \text{[Eq 5]} \qquad CNR = \frac{|\overline{x_{ROI1}} - \overline{x_{ROI2}}|}{\sqrt{\frac{\sigma^2_{ROI1} + \sigma^2_{ROI2}}{2}}} \quad \text{[Eq 6]}$$

Historically, an improvement of the CNR was associated with a better low-contrast detectability. However, this figure of merit is not task-oriented. Figure 12 shows various signals in homogenous backgrounds with equal CNR. The same CNR can result in different low-contrast detectability levels. The difference between the three images is the spatial frequency content of the noise. Thus, appropriate task-based image quality metrics should be used to assess lesion detectability.

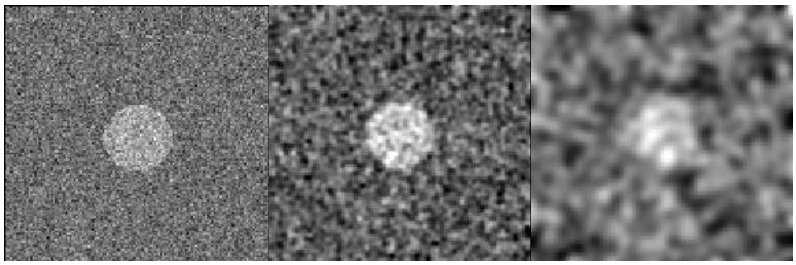


Figure 12: *Three simulated images with equal CNR but different detectability levels.*

Spatial resolution: Spatial resolution is related to the sharpness of the image. Indeed, increasing blur reduces the detectability of small objects. In CT, focal spot size, detector aperture, motion, voxel size (depending on the reconstructed field of view and matrix size), and the filtering type all impact spatial resolution. CT system responses to various input functions are generally described in image space:

- CT response to input Dirac stimuli described the Point Spread Function (PSF)
- CT response to line stimuli described the Line Spread Function (LSF)
- CT response to edge stimuli described the Edge Spread Function (ESF).

The first derivative of the ESF is the LSF. The full width at half maximum (FWMH) of the PSF can be used as a figure of merit to estimate spatial resolution in image space.

3.1.2 *Physical metrics in Fourier Space*

Working in the Fourier space requires the linearity and stationarity of the CT system. While these two conditions are approximately satisfied with the FBP reconstruction algorithm, it is not necessarily the case with IR algorithms. In these conditions, some precautions should be taken to assess image noise (NPS, Noise Power Spectrum) and spatial resolution (MTF, Modulation Transfer Function).

Noise: In Fourier space, noise power is expressed in terms of the NPS. It measures the autocorrelation of pixel values in the spatial frequency domain, providing its amplitude over the entire frequency range. NPS is calculated using equation 7:

$$NPS_{2D}(f_x, f_y) = \frac{\Delta_x \Delta_y}{L_x L_y} \frac{1}{N_{ROI}} \sum_{i=1}^{N_{ROI}} |FT_{2D}\{ROI_i(x,y) - \overline{ROI_i}\}|^2 \qquad \text{[Eq 7]}$$

Where Δ_x, Δ_y are the pixel sizes in millimetres in the x and y direction while Lx and Ly are the lengths in number of pixels of the ROI. N_{ROI} is the number of ROIs used to calculate NPS and $\overline{ROI_i}$ is the mean pixel values of the i^{th} ROI. Since the NPS is a decomposition of noise power (variance) in the spatial frequency range, the integral of the NPS in the two spatial frequency directions, f_x and f_y, is equal to the variance. To simplify the representation, a mean radial 1D NPS can be plotted as a function of the radial frequency f_r, expressed from the spatial frequency in x and y direction f_x and f_y using the equation $f_r = \sqrt{{f_x}^2 + {f_y}^2}$. Figure 13 shows three NPS for a CT acquisition of a homogeneous test object reconstructed with three different algorithms. The reduction of noise level with IR algorithms (ASIR and VEO) in comparison with FBP reconstruction results in a decrease of the NPS amplitude over most of the frequency range. The change in noise texture with IR algorithms results in a shift of the NPS peak toward lower frequencies.

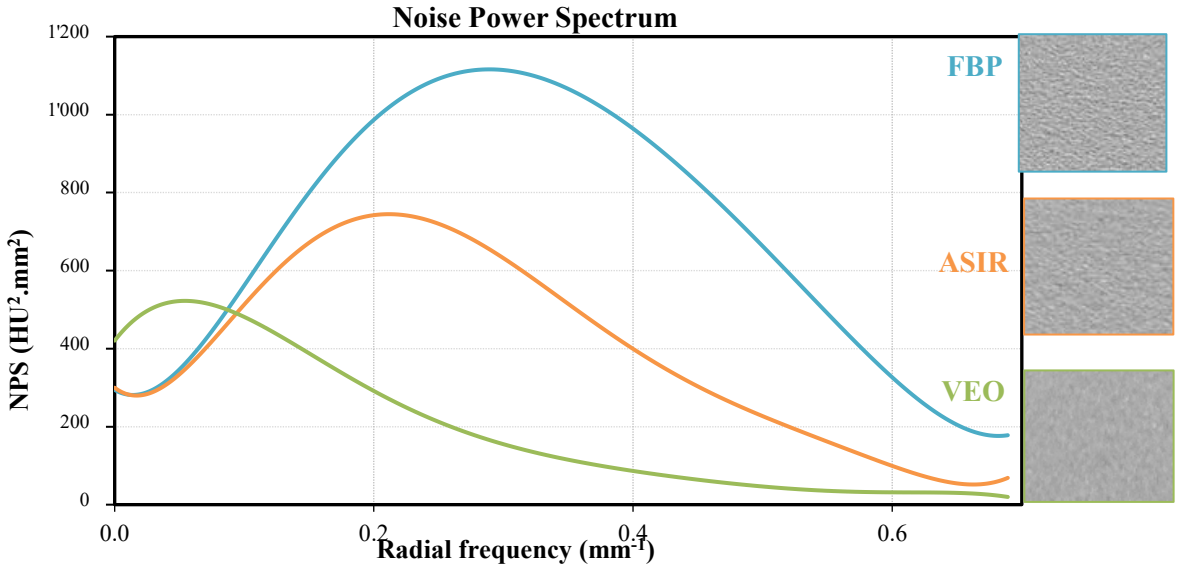


Figure 13: *NPS for three different algorithms: FBP (blue), ASIR (a hybrid IR in orange), VEO (a full IR in green) obtained at the same dose level. Corresponding ROIs were displayed using the same window level. The curves was published in a SPIE proceeding in 2017, more details about CT acquisition and could be found in the paper* [75].

Spatial resolution: The MTF provides an assessment of the contrast transfer as a function of spatial frequency. It is assessed in Fourier space. The MTF describes the capability of a system to transfer the spatial frequency of a signal, from the object to the image. The MTF is the Fourier transform of the LSF and is normalized by its zero-frequency value (Equation 8).

$$MTF_{2D}(f_x, f_y) = \frac{|FT_{2D}\{LSF(x,y)\}|}{|FT_{2D}\{LSF(0,0)\}|} \qquad \text{[Eq 8]}$$

The LSF is obtained from the first derivative of the ESF, the CT response to a high-contrasted edge. When using iterative reconstruction techniques, the imaging system is no longer linear and introduces a non-linear relationship between contrast, noise, and spatial resolution. Therefore, the Target Transfer Function (TTF) was proposed to characterise the spatial resolution specific to a given contrast level. A dedicated custom-made phantom containing cylindrical rods of three materials (Teflon, Polyethylene (PE), and polymethyl methacrylate (PMMA) in a water environment can be used to assess the TTF, as described by Ott [26]. CT images of the test object are shown in Figure 14. As the test object used to calculate the ESF is circular (to avoid artefacts), a mean radial TTF is plotted. Figure 15 shows that the TTF does not depend on contrast (materials) for FBP reconstructions, while it is for IR algorithms.

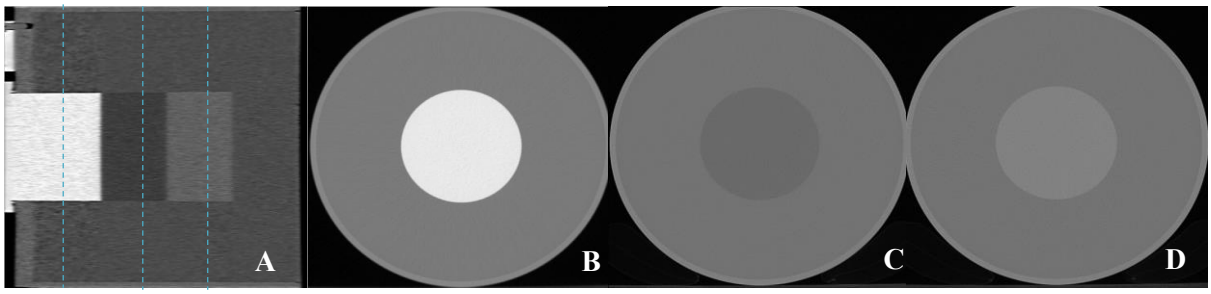


Figure 14 : *CT images of the dedicated custom-made phantom for TTF calculation: (A) Sagittal view, dashed lines represents the slices corresponding to the three axial views presented in B, C and D, (B) axial view with the Teflon cylinder (900 HU), (C) axial view with the Polyethylene cylinder (-80 HU), (D) axial view with the PMMA cylinder (120 HU)*

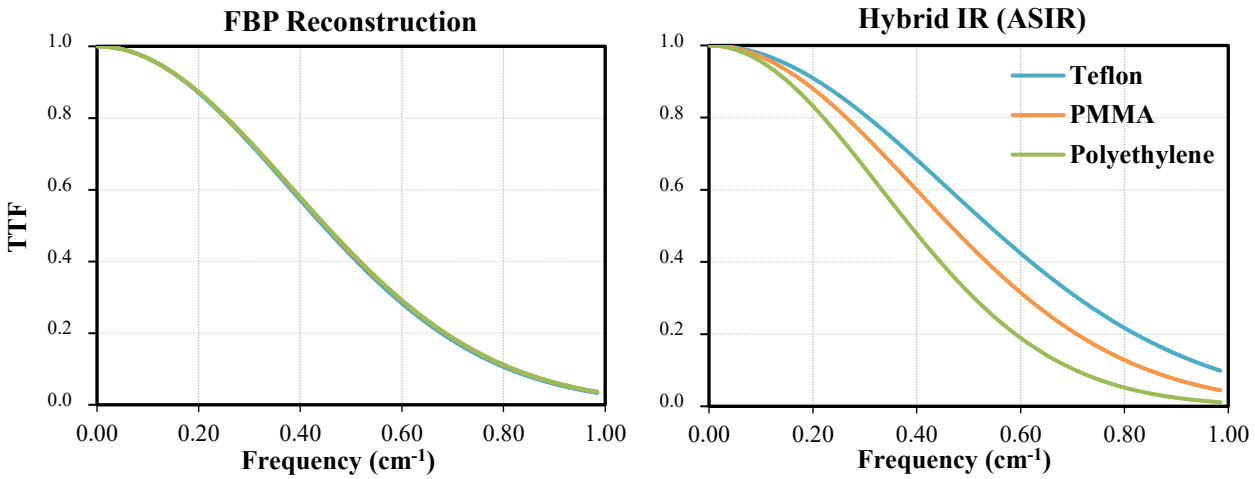


Figure 15 *: TTF for three contrast materials Teflon, PMMA and Polyethylene. Images were acquired with a standard kernel used for abdominal explorations and reconstructed with standard FBP and ASIR 50% at the same dose level*

Physical metrics in image or Fourier spaces are usually used for quality assurance control (acceptance test or annual test) with standardised phantoms. Spatial resolution is assessed in image space with a high-resolution test gauge or in Fourier space with a small spherical Tungsten insert. This is not suitable anymore for IR algorithms. Unfortunately, currently no standardised phantom is proposed for the TTF calculation.

As noise texture and spatial resolution can impact the detectability and become specific to a clinical task (contrast) with IR algorithms, it is necessary to go a step further and use task-based image quality assessment to assess diagnostic performances.

3.2 **Task-based image quality methodologies for diagnostic accuracy assessment**

3.2.1 *Receiver operating characteristics analysis*

The second level of the image quality hierarchical classification is the assessment of diagnostic accuracy. It will measure the performances of an imaging system in a clinical workflow to make a diagnosis. To assess diagnostic accuracy, it is necessary to compare radiologist diagnosis based on medical images with ground truth, the actual states of diseases (normal or abnormal). For many years, diagnostic accuracy was assessed by determining the percentage correct (the number of correct decisions divided by the number of cases). However, this measure does not take disease prevalence into account nor the number of false negative and false positive diagnoses (summarized in Table 3).

State of disease / Diagnosis	Normal	Abnormal
Positive	False Positive (FP)	True Positive (TP)
Negative	True Negative (TN)	False negative (FN)

Table 3: *Four classifications for a diagnostic test*

These two limitations can be overcome by determining the sensitivity (fraction of patients having an abnormal lesion and correctly diagnosed) and the specificity (fraction of patients having a normal lesion and correctly diagnosed). The term "true positive fraction (TPF)" and "true negative fraction (TNF)" are equivalent.

$$sensitivity = \frac{TP}{TP+FN} \qquad specificity = \frac{TN}{FP+TN} \qquad \text{[Eq 9]}$$

However, reporting a single value for sensitivity and specificity is not enough to compare two imaging systems (for example, two reconstruction algorithms) because it represents only a single decision threshold to make a diagnosis. Hence, the Receiver Operating Characteristic (ROC) analysis was introduced to describe the combinations between sensitivity and specificity values that radiologists can obtain using different decision thresholds (see Figure 16A). All the different combinations are usually represented by plotting the *sensitivity* as a function of *1- specificity* (see Figure 16B). This curve is called a ROC curve. The higher the separation between the "normal" and "abnormal" distribution, the better the performances of the CT system. From the characteristics TP, FP, TN, FN, it is possible to compute various performance indexes (Equation 10):

- Accuracy (ratio between the number of correctly diagnosed cases, positively or negatively, and the total number of cases).

- Negative predictive value (NPV, ratio between the number of negatively correct diagnosed cases and the number of negative calls).
- Positive predictive value (PPV, ratio between the number of positively correct diagnosed cases and the number of positive calls).

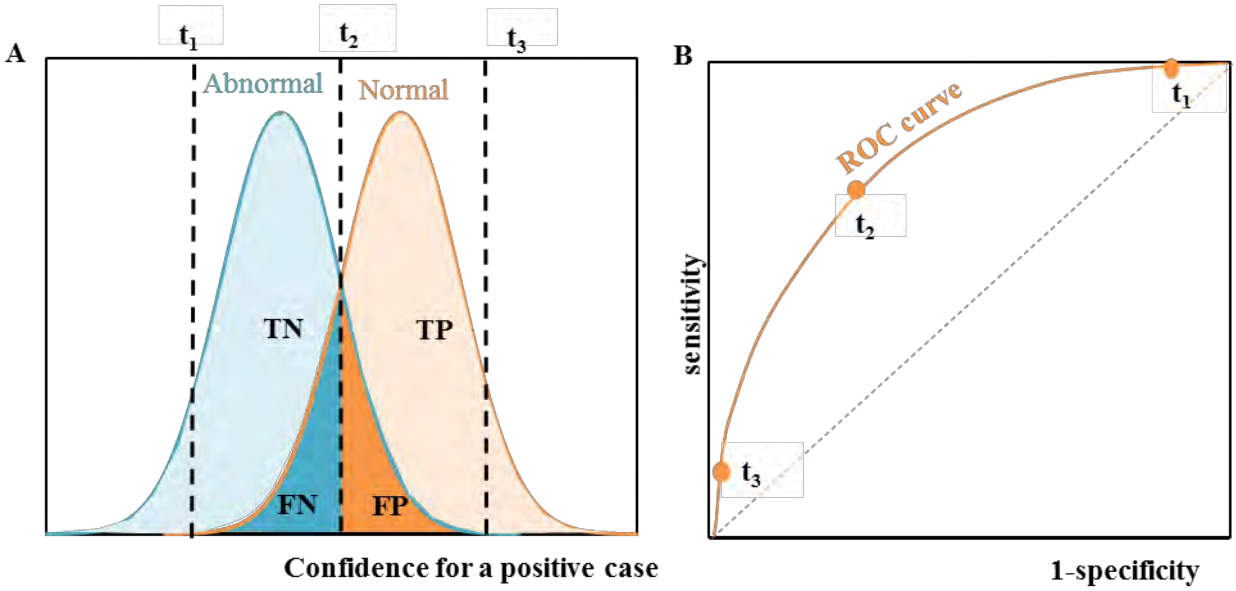


Figure 16: **(A)** *Probability density functions for abnormal (blue) and normal cases (orange). For the decision threshold t_2, the following characteristics: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) were represented in colour (light orange, dark orange, light blue, dark blue, respectively). This decision threshold produces a combination of sensitivity and specificity that is represented in graph B by the point labelled t_2. Similarly, the decision thresholds t_1 and t_3 have a corresponding point in graph B.* ***(B)*** *Graph representing the sensitivity as a function of the specificity for all decision thresholds (solid line).*

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FN} \qquad NPV = \frac{TN}{TN+FN} \qquad PPV = \frac{TP}{TP+FP} \qquad \text{[Eq 10]}$$

From a ROC analysis, two figures of merit are generally computed: the area under the roc curve (AUC) (equation 11) and the signal-to-noise ratio (equation 12). Assuming normal distributions, SNR can be linked with AUC using equation 13. The relation was plotted in Figure 17.

$$AUC = \int_0^1 dFPF\ TPF(FPF) \qquad \text{[Eq 11]}$$

Where FPF is the False negative fraction (defined as 1- TNF) and TPF is the True positive fraction. As TPF and TNF can vary from 0.5 to 1, the AUC varies from 0.5 to 1. (where 0.5 is similar to a hazard decision and 1 to a perfect decision).

$$SNR = \frac{\overline{\lambda_{abnormal}} - \overline{\lambda_{normal}}}{\sqrt{\frac{\sigma^2_{\lambda_{normal}} + \sigma^2_{\lambda_{normal}}}{2}}} \qquad \text{[Eq 12]}$$

Where $\lambda_{abnormal}$ is the confidence for a negative case and λ_{normal} is the confidence for a positive case. With an assumption of equal variance for normal and abnormal cases, SNR is specifically called detectability index. The range of the SNR is from 0 to infinity.

$$AUC = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{SNR}{2}\right) \qquad \text{[Eq 13]}$$

With erf, the error function, defined as $\mathrm{erf}(x) = \int_0^x e^{-t^2} dt$.

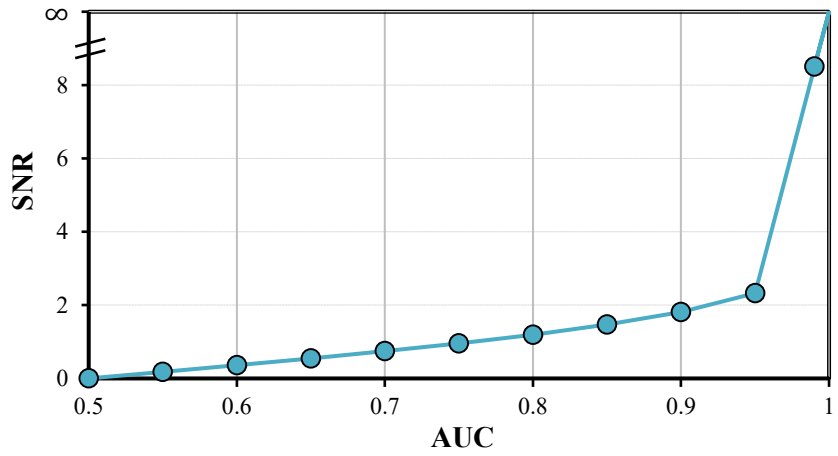


Figure 17: *Relation between SNR and AUC.*

3.2.2 *Ideal mathematical observers (2 class problem)*

Mathematical observers are algorithms useful for task-based image quality assessment in medical imaging. As discussed in the introduction, it is an efficient way to overcome the limitations of clinical studies: time, financial costs, and the large intra- and inter-variability among human observers. Initially introduced by Barrett and Myers for medical imaging tasks, mathematical observers are based on statistical decision theory [67,76]. They are defined for specific tasks (the common task applied is a lesion detection in a noisy background) and conditions (Signal Known Exactly, SKE or Signal Known Statistically SKS).

Let us define g, a vector containing all the pixel values of an image. When a radiologist looks at an image g, he has to make a decision Di (D_1 the disease is absent or D_2 the disease is present) based on prior knowledge and the cost relative to each decision. An ideal observer will make its decision by minimizing the overall average cost $\bar{C}$ of making a decision (equation 14):

$$\bar{C} = C_{22}\,\Pr(D_2|H_2)\,\Pr(H_2) + C_{12}\,\Pr(D_1|H_2)\,\Pr(H_2) + C_{21}\,\Pr(D_2|H_1)\,\Pr(H_1) + C_{11}\Pr(D_1|H_1)\Pr(H_1) \qquad \text{[Eq 14]}$$

With $\Pr(D_i|H_j)$ being the probability of making a decision D_i when the hypothesis H_j is true, $\Pr(H_j)$ the prior probability under hypothesis H_j (the prevalence of the disease) and C_{ij} the cost associated with making a decision D_i when the hypothesis H_j is true. Assuming that the decision is deterministic, this probability can be also defined as:

$$\Pr(D_i|H_j) = \int_{\Gamma_i} d^M g\ \mathrm{p}(g|H_j) \qquad \text{[Eq 15]}$$

$\mathrm{p}(g|H_j)$ is the conditional density function on the data g under hypothesis H_j and Γ_i a subspace containing images g conducting to make a decision D_i.

Assuming that the observer is forced to make a decision, we can write:

$$\int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_2) + \int_{\Gamma_1} d^M g\ \mathrm{p}(g|H_2) = 1 \qquad \text{[Eq 16]}$$

Using equation 15, the overall average cost (equation 14) can be written:

$$\bar{C} = C_{22}\Pr(H_2)\int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_2) + C_{12}\Pr(H_2)\int_{\Gamma_1} d^M g\ \mathrm{p}(g|H_2) + C_{21}\Pr(H_1)\int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_1) + C_{11}\Pr(H_1)\int_{\Gamma_1} d^M g\ \mathrm{p}(g|H_1) \qquad \text{[Eq 17]}$$

Using equation 16, we can write $\bar{C}$ as a function of Γ_1 only:

$$\bar{C} = C_{22}\Pr(H_2)\int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_2) + C_{21}\Pr(H_1)\int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_1) + C_{12}\Pr(H_2)\left(1 - \int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_2)\right) + C_{11}\Pr(H_1)\left(1 - \int_{\Gamma_2} d^M g\ \mathrm{p}(g|H_1)\right)$$

$$\bar{C} = \int_{\Gamma_2} d^M g\ \Pr(H_1)\,(C_{21} - C_{11})\mathrm{p}(g|H_1) - \int_{\Gamma_2} d^M g\ \Pr(H_2)\,(C_{12} - C_{22})\mathrm{p}(g|H_2) + C_{12}\Pr(H_2) + C_{11}\Pr(H_1) \qquad \text{[Eq 18]}$$

Minimizing $\bar{C}$ equals choosing Γ_2 in order to minimize the two integral terms (since the non-integral terms are constants). Assuming that the cost of making a wrong decision C_{12} or C_{21} is superior to the cost of making a right decision C_{22} or C_{11} and probability density functions are positive, Γ_2 is a region for which:

$$\Pr(H_0)\,(C_{12} - C_{22})\mathrm{p}(g|H_2) + \Pr(H_1)\,(C_{21} - C_{11})\mathrm{p}(g|H_1) < 0 \qquad \text{[Eq 19]}$$

The decision rule is to make a decision comparing the quantity $\frac{\mathrm{p}(g|H_2)}{\mathrm{p}(g|H_1)}$ (also called the likelihood ratio Λg) to a criterion $\frac{\Pr(H_1)(C_{21}-C_{11})}{\Pr(H_2)(C_{12}-C_{22})}$ (also called Λc):

$$\frac{\mathrm{p}(g|H_2)}{\mathrm{p}(g|H_1)} \underset{D_1}{\overset{D_2}{\gtrless}} \frac{\Pr(H_1)(C_{21}-C_{11})}{\Pr(H_2)(C_{12}-C_{22})} \qquad \text{[Eq 20]}$$

This inequality can be read "choose H_2 true if the likelihood ratio is higher than $\frac{\Pr(H_1)(C_{21}-C_{11})}{\Pr(H_2)(C_{12}-C_{22})}$, choose H_1 true if the likelihood ratio is lower than $\frac{\Pr(H_1)(C_{21}-C_{11})}{\Pr(H_2)(C_{12}-C_{22})}$.

Particular case: Signal-known-exactly / Background known-exactly (or Gaussian correlated noise)

We can define two hypotheses:

$$H_1: g = s_1 + n \qquad H_2: g = s_2 + n$$

For a correlated Gaussian noise with $\boldsymbol{K_n}$ the noise covariance matrix, the associated probability density follows the multi-variate Gaussian (equation 21).

$$\mathrm{p}(g|H_j) = \frac{1}{\sqrt{2\pi^M \det(\boldsymbol{K_n})}}\, e^{-\frac{1}{2}(g-s_j)^T \boldsymbol{K_n}^{-1}(g-s_j)} \qquad \text{[Eq 21]}$$

The decision rule (equation 20) can be written in this case:

$$\frac{e^{-\frac{1}{2}(g-s_2)^T \boldsymbol{K_n}^{-1}(g-s_2)}}{e^{-\frac{1}{2}(g-s_1)^T \boldsymbol{K_n}^{-1}(g-s_1)}} \begin{matrix} >^{D_2} \\ <_{D_1} \end{matrix} \frac{\Pr(H_1)(C_{21}-C_{11})}{\Pr(H_2)(C_{12}-C_{22})} \qquad \text{[Eq 22]}$$

The likelihood ratio Λg (left part of the inequality) can be simplified:

$$\lambda_g = \ln(\Lambda g) = \frac{1}{2}(g-s_1)^T \boldsymbol{K_n}^{-1}(g-s_1) - \frac{1}{2}(g-s_2)^T \boldsymbol{K_n}^{-1}(g-s_2)$$

$$\lambda_g = \ln(\Lambda g) = \frac{1}{2}[g^T \boldsymbol{K_n}^{-1}(s_2-s_1) + (s_2-s_1)^T \boldsymbol{K_n}^{-1} g] + \frac{1}{2}[{s_1}^T \boldsymbol{K_n}^{-1} s_1 - {s_2}^T \boldsymbol{K_n}^{-1} s_2\ ] \qquad \text{[Eq 23]}$$

with $\Delta s = s_2 - s_1$. Noting that in equation 24 the last term does not depend on g, the decision rule can be simplified:

$${\lambda'}_g = \Delta s^T \boldsymbol{K_n}^{-1} g \begin{matrix} >^{D_1} \\ <_{D_0} \end{matrix} \Lambda c \qquad \text{[Eq 24]}$$

In this case, the ideal observer is linear. We can note that multiplying an image by $\boldsymbol{K_n}^{-1/2}$destroys the correlation of the noise. Hence, it is possible to decompose ${\lambda'}_g$ (equation 24):

$${\lambda'}_g = \left(\boldsymbol{K_n}^{-1/2} \Delta s\right)^T \boldsymbol{K_n}^{-1/2} g$$

The ideal observer destroys the correlation of Δs and images g before computing the likelihood ratio ${\lambda'}_g$.

3.2.3 *Linear observers*

We have seen that to compute the ideal observer, it is necessary to compute the likelihood ratio or its log and compare it with numerous criteria. For Gaussian data, the ideal observer is linear. However, usually for non-Gaussian data, or in case there are other sources of variability (random signal, random background), the ideal observer is no longer linear, and computing it can be difficult. To overcome this limitation, one could use a suboptimal linear observer, which takes the general form:

$${\lambda_k}^{(i)} = w^T {g_k}^{(i)} \qquad \text{[Eq 25]}$$

${\lambda_k}^{(i)}$ is the scalar response of a vector image ${g_k}^{(i)}$ from class k through *w* (in a 2 class problem, k=0 or 1). We can define $\overline{\lambda_k}$ as the mean scalar responses of vector images from class k and σ_k as the standard deviation of the scalar responses for vector images from class k:

$\overline{\lambda_k} = w^T \overline{g_k}$, with $\overline{g_k}$ the mean vector images for class k.

${\sigma_k}^2 = \sum_i (w^T {g_k}^{(i)} - w^T \overline{g_k})^2 = \sum_i w^T ({g_k}^{(i)} - \overline{g_k})({g_k}^{(i)} - \overline{g_k})^T w = w^T \boldsymbol{K_k} w$, with $\boldsymbol{K_k}$ the covariance matrix of class k.

Then try to find the optimal direction of w to maximize the signal to noise ratio defined in equation 12 (the discrimination between the two classes):

$$SNR^2(w) = \frac{2\left(w^T \overline{g_1} - w^T \overline{g_0}\right)^2}{w^T(\boldsymbol{K_0}+\boldsymbol{K_1})w} = \frac{2\, w^T \Delta_g {\Delta_g}^T w}{w^T(\boldsymbol{K_0}+\boldsymbol{K_1})w} \qquad \text{[Eq 26]}$$

where Δ_g is the difference between the mean images in class 1 and class 0, respectively. As we shall demonstrate in Annex 1 using the Fisher linear discriminant analysis, the optimal direction of w that maximizes SNR is given by equation 27:

$$w_{Opt} = (\boldsymbol{K_0} + \boldsymbol{K_1})^{-1}\Delta_g \qquad \text{[Eq 27]}$$

This optimal linear observer is also called the Hotelling observer (equation 28) for a detection task (detection of a signal in a noisy background) and requires the knowledge of the covariance matrix and the mean images:

$$\lambda_{HO} = {\Delta_g}^T \left[\frac{1}{2}\,(\boldsymbol{K_0} + \boldsymbol{K_1})^{-1}\right] g \qquad \text{[Eq 28]}$$

Particular case: Signal-known-exactly

For a detection task when the signal is known-exactly, the only image fluctuation comes from the noise and usually both image classes have the same covariance matrix ($\boldsymbol{K_0}$=$\boldsymbol{K_1}$=$\boldsymbol{K_n}$) and Δ_g, the difference between the mean image in class 1 and class 0, is usually the searched signal s. In this case, the Hotelling observer is the Prewhitening observer (PW) and takes the form:

$$\lambda_{PW} = s^T\, {\boldsymbol{K_n}}^{-1} g \qquad \text{[Eq 29]}$$

The SNR quantifies the performance of this observer, and takes the form:

$$SNR_{PW} = \sqrt{s^T\, {\boldsymbol{K_n}}^{-1} \mathrm{s}} \qquad \text{[Eq 30]}$$

When the noise is a correlated Gaussian noise (equation 24), the PW observer is ideal.

Particular case: Signal-known-exactly / non-correlated noise

For a detection task when the signal is known-exactly and the noise is non-correlated (white noise), the Hotelling observer is the Non-Prewhitening observer (NPW) and takes the form:

$$\lambda_{NPW} = s^T\, g \qquad \text{[Eq 31]}$$

The SNR quantify the performance of this observer, and takes the form:

$$SNR_{NPW} = \frac{[s^T s]^2}{s^T \boldsymbol{K_n}\, \mathrm{s}} \qquad \text{[Eq 32]}$$

Sometimes, NPW could be used in correlated noise. In this case, the NPW is suboptimal in comparison with PW observer. The NPW model is successful to predict human observer performance in white noise but not in a real anatomical background[77].

3.2.4 *Channelized Hotelling Observer*

To compute the Hotelling observer, the mean image for each class and the covariance matrix is required. If it is possible to easily estimate the mean images using a subset of images belonging to each class, estimating the covariance matrix is more complicated. Indeed, to be invertible and non-singular, the number of sample images used to estimate the covariance matrix should be at least equal to the number of pixels in an image. Thus, in medical imaging the Hotelling observer is rarely computed. To overcome this limitation, an alternative is the Channelized Hotelling Observer (CHO) [78,79]. This observer reduces the dimensionality of data using a set of channels and keeps only the frequencies included in images. The ith channelized vector image ${v_k}^{(i)}$ from class k is expressed using equation 33.

$$v_k{}^{(i)} = \boldsymbol{U}^T g_k{}^{(i)} \qquad \text{[Eq 33]}$$

With $g_k{}^{(i)}$ the i[th] vector image from class k containing N components, U is a N x P matrix where columns are the channel profiles (P channels) and v the resulting channelized image vector with P components.
The CHO works directly with channelized images and is expressed using equation 34:

$$\lambda_k{}^{(i)} = \Delta v^T \left[\frac{1}{2}\left(\boldsymbol{K_{v_o}} + \boldsymbol{K_{v_1}}\right)^{-1}\right] v_k{}^{(i)} \qquad \text{[Eq 34]}$$

With $\boldsymbol{K_{v_k}}$, the covariance matrix of channelized images from class k and Δv the difference between the mean channelized image from class 0 and 1. When both image classes have the same covariance matrix ($\boldsymbol{K_{v_0}}$=$\boldsymbol{K_{v_1}}$=$\boldsymbol{K_{v_n}}$), the SNR takes the form:

$$SNR = \sqrt{\Delta_v{}^T \, \boldsymbol{K_{v_n}}^{-1} \Delta_v} \qquad \text{[Eq 35]}$$

As for the HO, the mean channelized image for each class and the covariance matrix of channelized images is required to compute the CHO. The covariance matrix is a P x P matrix, with P being the number of channels. Hence, at least P^2 sample images should be used to estimate the covariance matrix.

The channels could be designed to be efficient (giving the best discrimination between two classes), or more anthropomorphic (reproducing the spatial frequency selectivity of the human visual system).

Efficient channels

The Laguerre Gauss (LG) channels have been proposed by Gallas and Barrett as the most efficient channels to detect a radially symmetric signal with a known location in a background with a correlation structure that has no preferred orientation [79]. The LG channels are the product between Laguerre polynomials and Gaussian functions. The p[th] channel is defined as:

$$u_p(r, a_u) = \frac{\sqrt{2}}{a_u} exp\left(\frac{-\pi r^2}{{a_u}^2}\right) L_p\left(\frac{2\pi r^2}{{a_u}^2}\right) \qquad \text{[Eq 36]}$$

Where a_u is the width of the Gaussian function, chosen relative to the signal size, and Lp is the p[th] Laguerre polynomial (see equation 37). The total number of channels depends of the complexity of the background. The CHO with LG channels is called the "ideal channelized linear observer".

$$L_p(x) = \sum_{k=0}^{n} (-1)^k \binom{p}{k} \frac{x^k}{k!} \qquad \text{[Eq 37]}$$

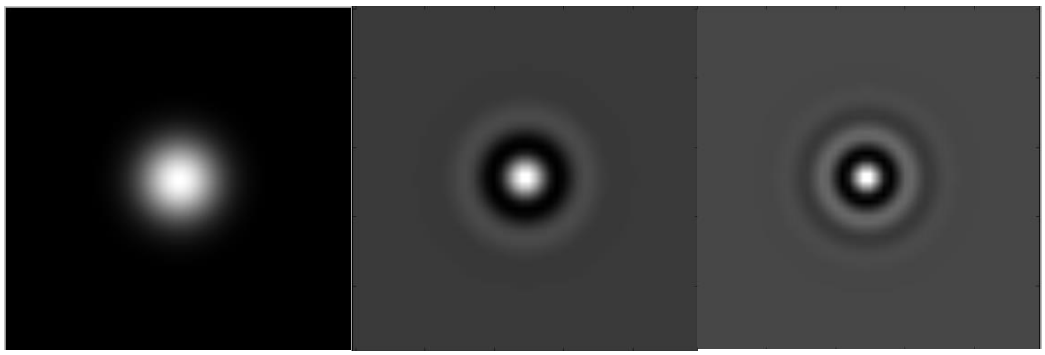


Figure 18 : *50x50 Laguerre Gauss channels with a parameter a_u=10 : (A) channel with p=0, (B) channel with p=4, (C) channel with p=9.*

Anthropomorphic channels

Among anthropomorphic observers, dense difference of Gaussian (DDoG) channels and Gabor channels are the most computed channels in the literature to assess a detection task. The DDoG channels are particularly interesting for a radially symmetric signal. Moreover, they use only a small number of channels in comparison with Gabor channels. Hence, the covariance matrix could also be estimated with a smaller number of sample images.

The DDoG channels are commonly expressed in the spatial frequency domain using equation 38. The radial profile for the p[th] channel is:

$$C_p(\rho) = exp\left[-\frac{1}{2}\left(\frac{\rho}{Q\sigma_p}\right)^2\right] - exp\left[\frac{1}{2}\left(\frac{\rho}{\sigma_p}\right)^2\right] \qquad \text{[Eq 38]}$$

Where ρ is the spatial frequency, p the channel number, Q the bandwidth of the channel, and σ_p the standard deviation for channel p, defined as: $\sigma_p = \sigma_O \alpha^{p-1}$. The following parameters are generally used: $\sigma_O = 0.005$, $\alpha = 1.4$, $Q = 1.67$.[77] These channel profiles were plotted in Figure 19

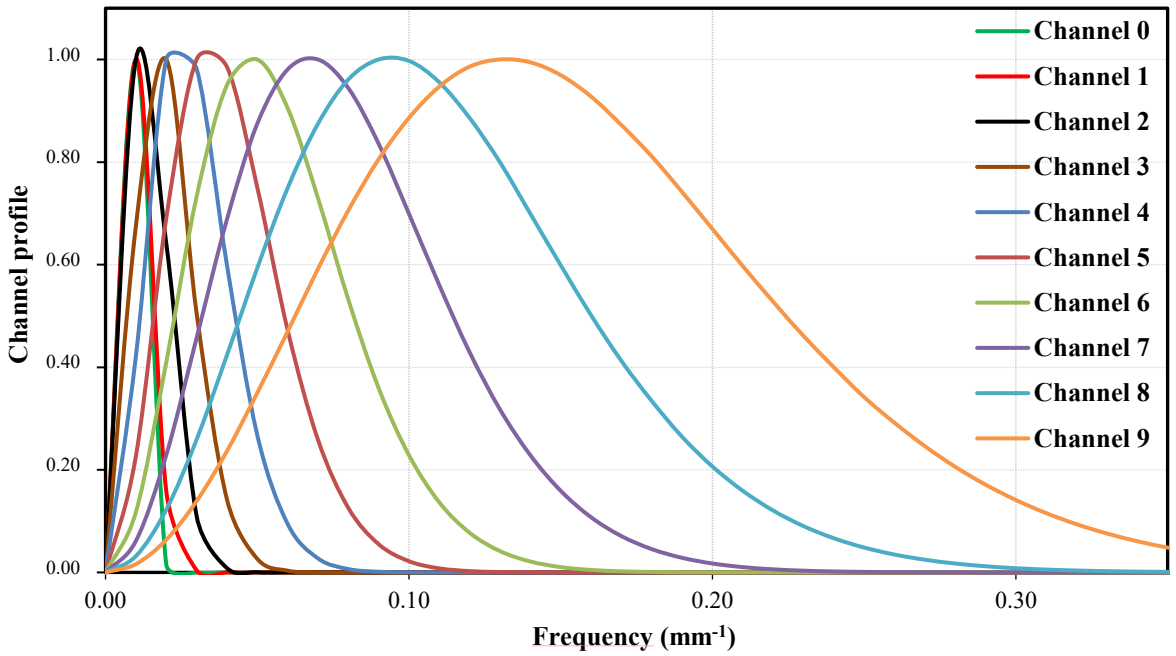


***Figure 19**: 10 DDoG channels with the following parameters $\sigma_O = 0.005$, $\alpha = 1.4$ and $Q = 1.67$*

Contrary to the DDoG channels, Gabor channels are particularly interesting for a non-radially symmetric signal in an oriented noise. The Gabor channels were expressed in the image domain for different spatial frequencies, orientations, and phases using equation 39:

$$V(x,y) = exp\left[-4\ln(2)\left(\frac{x^2+y^2}{{w_s}^2}\right)\right] cos[2\pi f(xcos\Theta + ysin\Theta) + \beta] \qquad \text{[Eq 39]}$$

Where w_s is the channel width, f the spatial frequency, Θ the orientation, and β the phase.

3.2.5 *Model observer in Fourier space*

SNR of mathematical observers (PW or NPW observers) can be computed in the image domain (see equation 30) or in the frequency domain (see equation 39) when the covariance matrix and the background mean are stationary. By analogy, the covariance matrix in the image domain is equivalent to the NPS in the frequency domain. The searched signal s in the image domain is equivalent to the Fourier transform of the theoretical signal multiplied with the TTF and the signal contrast. For a radially symmetric signal, SNR can be written as a continuous function in the frequency-domain using equation 40 and 41 for PW and NPW observers, respectively.

$$SNR_{PW} = \sqrt{2\pi}\Delta HU\sqrt{\int_0^{f_{ny}} \frac{S^2(f)TTF^2(f)}{NPS(f)} f df} \qquad \text{[Eq 40]}$$

$$SNR_{NPW} = \sqrt{2\pi}\Delta HU \frac{\int_0^{f_{ny}} S^2(f)TTF^2(f) f df}{\sqrt{\int_0^{f_{ny}} S^2(F)TTF^2(f)NPS(f) f df}} \qquad \text{[Eq 41]}$$

Where ΔHU is the contrast difference between the signal and the background, f_{ny} the Nyquist frequency, and S(f) is the Fourier transform of the theoretical signal. For an input signal defined with a cylinder of radius R (equation 42), S(f) is expressed using equation 43.

$$f(x,y) = \begin{cases} 1, & \sqrt{x^2+y^2} \leq R \\ 0, & \sqrt{x^2+y^2} > R \end{cases} \qquad \text{[Eq 42]}$$

$$S(f) = \frac{R}{f} J_1(2\pi R f) \qquad \text{[Eq 43]}$$

Where J_1(x) is a Bessel function of the first kind and $f = \sqrt{u^2+v^2}$, with u and v the spatial frequency in x and y direction.

The choice between model observers in the image or frequency domain depends on the goal of the study and the experimental conditions. Model observers in the image domain should be used to directly compare human and model observer performances with the same set of images. However, this requires the use of an appropriate phantom with inserted lesions. Then, the high number of images necessary to compute the HO or CHO implies scanning the phantom many times (especially if there is only one lesion with a given contrast, shape and size). Model observers in the frequency domain could be used when experimental conditions are not fulfilled (phantom, number of scans). However, the system should be linear and stationary to apply the methodology.

4. Achieved results

The principal results that compose the core of the thesis were developed around three milestones:

- A task-based image quality assessment applied for a detection task with conventional CT. This part is associated with two published peer-reviewed papers.
- A task-based image quality assessment applied for a quantification task in DECT. This part is not associated with a peer-reviewed paper.
- A task-based image quality assessment applied for a discrimination task in SPCCT. This part is associated with two under submission peer-reviewed papers. The first paper demonstrates the performance of an SPCCT system for a discrimination task in in-vitro conditions; the second paper proposes to validate the results for ex-vivo conditions.

The following section will summarize each milestone with its associated articles, my contribution, the clinical context of the article, the applied task-based image quality methodology, and the main outcomes. The articles will be included in the subsequent section.

4.1 Task-based image quality assessment for the detection of low and high-contrast lesions with conventional CT

4.1.1 *Comparison of multiple CT scanners for the detection of low and high-contrast lesions*

Contribution

The study deals with an objective comparison of multiple CT scanners for two tasks: the detection of high and low-contrast structures. The study and the writing of the manuscript was performed in collaboration with Damien Racine during his PhD. We equally contributed to the paper, Damien Racine was in charge of the part concerning the detection of low-contrast structures and I was in charge of the part concerning the detection of high-contrast structures.

Clinical context

The quality of a CT examination depends on a wide range of parameters such as acquisition time, temporal resolution, spatial resolution, and energy resolution when dealing with kV optimisation. Thus, the actual determination of the clinical performance of a CT device is quite complex, and the clinical question needs to be clarified in order to set a standard for image quality level. In this context, we proposed a comparison of eight CT devices (an older and a newer one, from four manufacturers) for two abdominal protocols; one used to detect urinary stones and the other used to detect focal liver lesions. We could translate these indications into two simple tasks: the detection of a high-contrast structure and the detection of a low-contrast structure. A task-based methodology was applied to assess the two tasks.

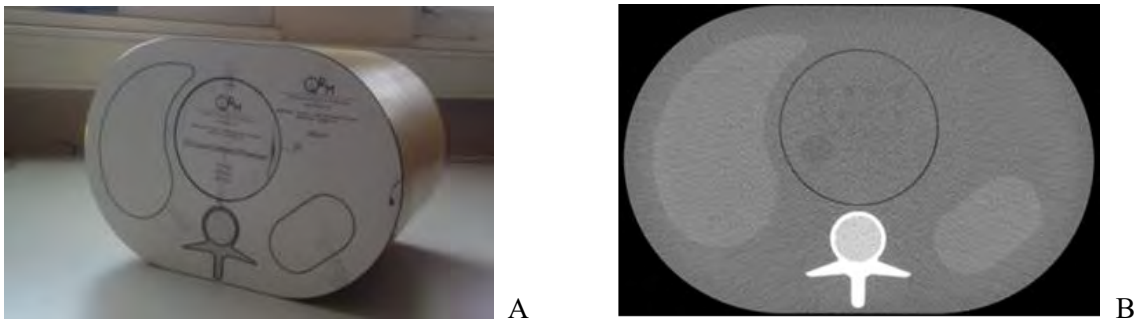
Task-based image quality assessment

We studied 8 CT devices from the four major manufacturers. In order to benchmark the different CT devices, we chose to use an ideal linear observer, which will maximize observer performances. For the two tasks, the four items of the task-based image quality assessment are described in Table 4.

	High contrast detection task	Low contrast detection task
Type of task	Classification task	Classification task
Properties	**Signal**: high-contrast rods of different diameters (1 to 10mm) and contrasts (1000HU, 120HU, -80HU) **Noise**: homogeneous background **Phantom**: no existing phantom for this task.	**Signal**: low-contrast spheres of different diameters (5, 6, 8 mm) and a contrast of 20HU **Noise**: homogeneous background **Phantom:** Abdominal phantom with inserted lesions
Observer	PW observer in Fourier Space	CHO with LG channels (image space)
Figure of merit	SNR	AUC_w (linear combination between AUC results for each size).

Table 4: *Summary of task-based image quality assessment used in the study for the two principal tasks: the detection of high and low-contrast lesions.*

For the detection of high-contrast lesions, no standardised phantom with specific high-contrast structure exists. In this case, a PW model observer in the frequency domain was used (see equation 40). NPS and TTF were assessed with the homemade phantom presented in Figure 13 and the model observer was computed for different simulated high-contrast rods. For the TTF calculation, the Teflon insert was chosen since the contrast is close to the contrast of renal stones calculi. For the detection of low-contrast lesions, we used a specific abdominal phantom containing four low-contrast spheres with a given contrast and size (see Figure 20B).
To reduce the number of images required, the CHO was used instead of the HO observer in the image space. LG channels were used in order to compute an ideal CHO (see equation 36).



A B

Figure 20*: (A) Photo of the abdominal phantom, (B) axial view of the phantom where the 20 HU spheres are located*

Main outcomes

The results were anonymised, the capital letters A, B, C and D refer to the four manufacturers, the lower-case letter "n" refers to a newer CT generation, and "o" to an older CT generation. Figure 21 shows the detectability (SNR) of high-contrast rods with different diameters for an abdominal protocol used for the detection of urinary renal stones. The SNR increased with the lesion size. For each manufacturer, we observed that the newer CT device improved the SNR compared to the older one. Particularly, the largest improvement was

observed for manufacturer C (80% for lesions from 3 to 5 mm) whereas moderate improvements were found for manufacturers A, B and D (26%, 22% and 40%, respectively).

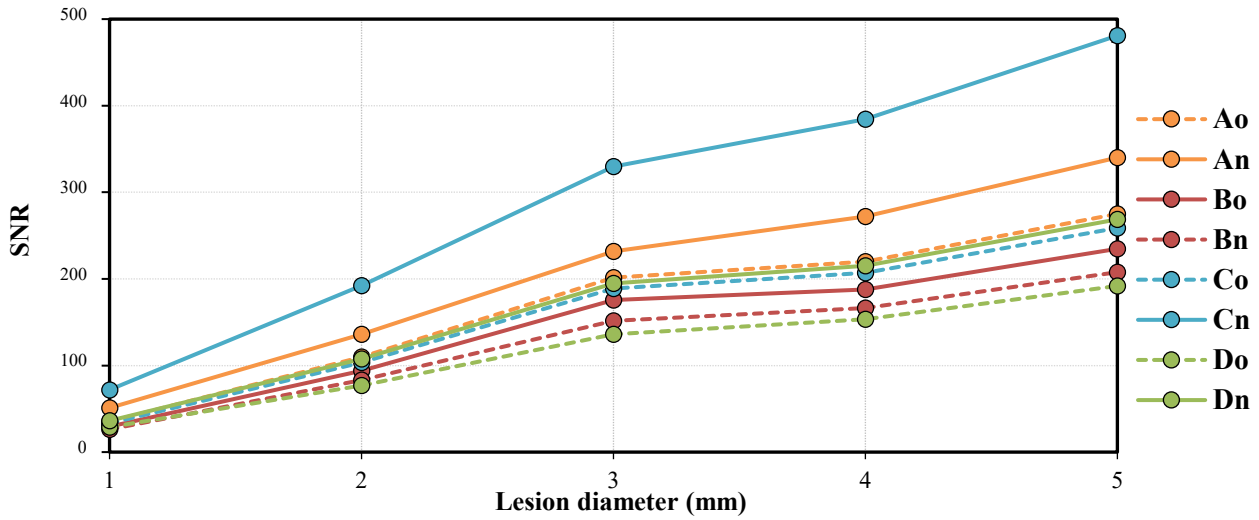


Figure 21: *The detectability (SNR) of high-contrast rods with different diameters (for a nominal contrast of 1000 HU) for an abdominal protocol.*

Figure 22 shows the detectability of low-contrast spheres for an abdominal protocol (detection of focal liver lesions) with different dose levels. The highest dose level 15 mGy corresponds to the DRL used in Switzerland for abdominal protocols (data published in 2008 by the Federal Office of Public Health). For this dose level, there was no major difference among the various scanners; the detectability is trivial with AUC_w values of nearly 1. Reducing the dose level to 10 mGy slightly decreased the detectability for all scanners, except scanner "Do" where a larger reduction can be observed. For the lowest dose level of 5 mGy, we observed that newer scanners improved the detectability compared to older ones for manufacturer A, C and D. The largest improvement was noted for manufacturer D (11%) whereas moderate improvements were found for manufacturers B and C (3 and 7 %, respectively). Surprisingly, for manufacturer B the detectability was higher for the older CT compared to the newer CT.

Finally, these results show that technological improvements between older and newer versions of CT devices have an impact on the high-contrast and low-contrast detectability. We also noticed that the variability of image quality can be important between the various CT systems especially for a low dose level. That is why, it is necessary to independently optimise each CT for the different clinical protocols with adapted clinical tasks. Particularly, the potential for optimisation can be more limited for older CTs compared to newer CTs.

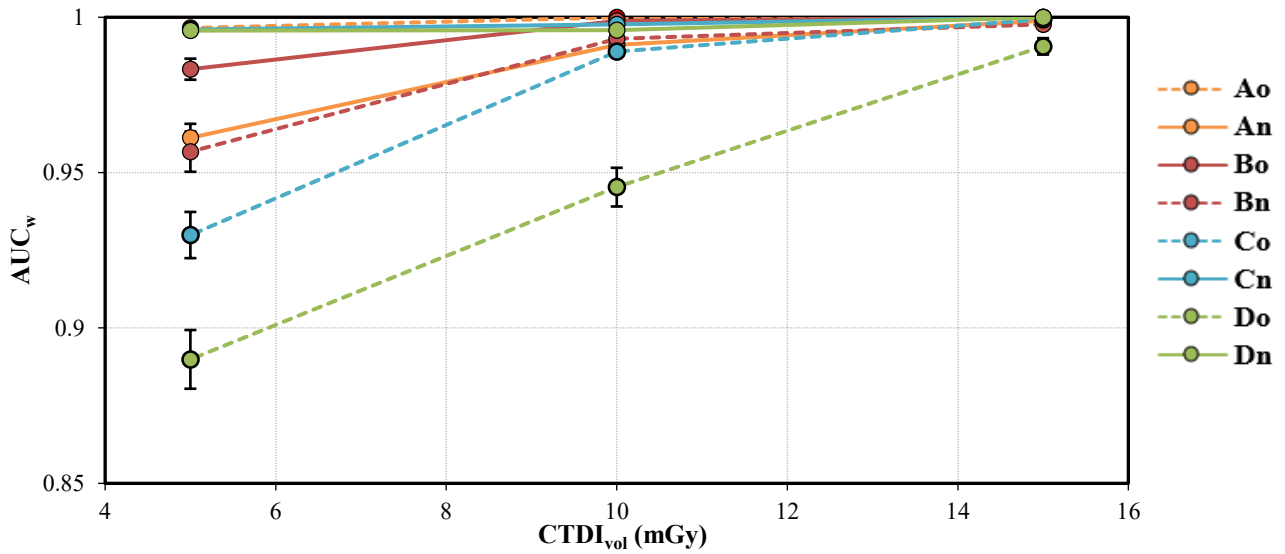


Figure 22: *Low contrast detectability as a function of dose levels for an abdominal protocol*

4.1.2 *Characterisation of five iterative reconstruction techniques for the detection of low-contrast lesions with an abdominal protocol*

Contribution

The measurements were conducted with Christoph Aberle from Basel University Hospital. Christophe Aberle and I independently computed the task-based assessment analysis and finally compared the results.

Clinical context

The optimisation process is an important step, and it could be particularly critical when dealing with the detection of low-contrast lesions in the abdominal region. Indeed, the risks of misdiagnosis could be high if the dose level is drastically decreased for abdominal protocols with IR techniques. In this context, we first proposed characterising the impact of five IR algorithms on low-contrast detectability for various patient body habitus using a clinical abdominal protocol on four scanners.

Materials

Four CT units were studied: Revolution, and Discovery CT750HD (GE HealthCare), as well as SOMATOM Definition Flash and SOMATOM Definition Edge (Siemens HealthCare). We scanned the abdominal phantom, containing hypo-dense lesions of 5, 6 and 8 mm, with two supplemental annuli in order to mimic different body habitus. The clinical abdominal protocol chosen on each CT machine was the porto-venous acquisition phase. The ATCM settings were those used in clinical protocols. Other acquisition parameters were kept as comparable as possible between each manufacturer. Images were reconstructed using FBP and the IR algorithms available on each CT.

Task-based image quality assessment

In order to characterise and optimise a clinical abdominal protocol in terms of low-contrast detectability, we used an anthropomorphic model observer. As for the previous part, to reduce the number of images required

we used a CHO instead of a HO. We chose 10 DDoG channels (anthropomorphic channels) since the lesions present in the phantom have a spherical symmetry. AUC was used as a figure of merit to assess the low-contrast detectability. It is important to note that, even if the studied task was the same as the previous study, material and methods should be adapted. Indeed, the characterisation and optimisation of clinical protocols with only one phantom size is insufficient and requires the use of various phantom size.

Main outcomes

Figure 23 shows the impact of the IR algorithms for the detection of a 5mm sphere as a function of the displayed $CTDI_{vol}$ for each scanner and phantom size. For the small phantom size, all scanners performed very well (AUC nearly 1), and IR algorithms did not improve the low-contrast detectability. For the medium phantom size, AUC values for each CT decreased and only the full IR (VEO) improved the low-contrast detectability compared to FBP (5.8%). For the large phantom size, a similar behaviour for the full IR was noticed (improvement of 8.5% compared to FBP). Hence, the different generations of IR algorithms (hybrid IR, partial MBIR) did not significantly improve the low-contrast detectability, except the full IR for the larger phantom sizes.

Finally, the potential of dose optimisation with the different IR algorithms should be limited when dealing with the detection of low-contrast lesions in abdominal region. As expected, the image quality decreased with phantom size, which requires to conduct optimization studies for different patient body habitus.

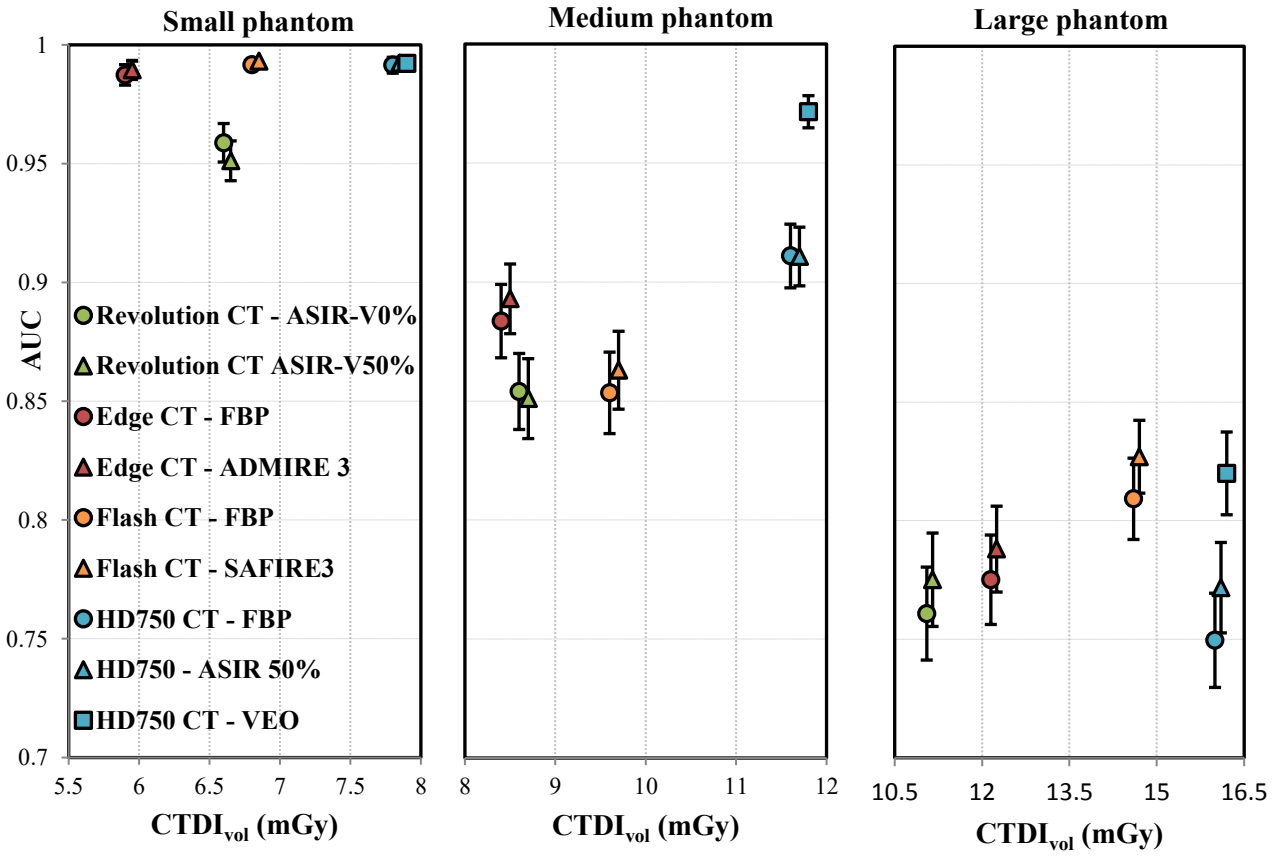


Figure 23: *Impact of the image reconstruction algorithms on low-contrast detectability as a function of the CTDIvol of each CT for a 5-mm diameter sphere and the three phantom sizes.*

4.2 **Image quality assessment applied for a quantification task in dual-energy CT**

4.2.1 *Effective Z map accuracy in material characterisation in DECT*

Contribution

The measurement and the analysis were performed with the help of Remy Petkantchin (master student). The methodology and the results were fully described in this part because the work is not associated with a peer-reviewed article.

Context

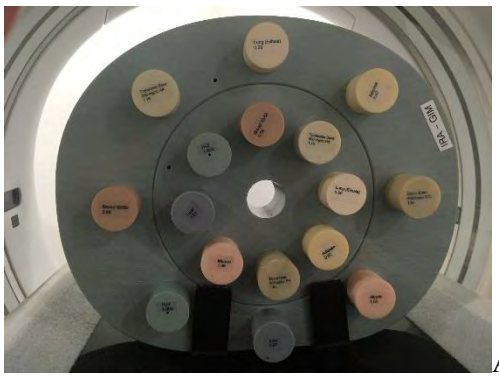
Even if efforts are currently focused on reducing dose level for conventional CT with IR techniques, DECT devices are emerging and the number of clinical applications is increasing. In particular, DECT offers the possibility of discriminating between different materials based on their Z_{eff} and density. Z_{eff} map can be reconstructed in DECT, which is clinically useful for various applications: the correct identification of renal stones with the differentiation between uric acid and non-uric acid calculi [80,81], the composition of carotid plaques to identify vulnerable and stable plaques[82], or the differentiation between uric acid and calcium pyrophosphate crystals in the diagnosis of gout [32]. Hence, the main goal of this study was to assess the accuracy and the precision of the GE HD750 Discovery CT scanner for measuring Z_{eff}, and to compare with Z_{eff} reference values, for various materials with an abdominal protocol, in order to ensure the capacity of diagnosis with DECT images..

Methodology

We scanned the CIRS Electron Density Reference Phantom model 062 with the GE HD750 Discovery DECT scanner (see Figure 24). A routine DECT abdominal protocol was reviewed. The relevant parameters are listed in Table 5. The materials selected for the study were Adipose (fat), Muscle, Trabecular Bone, and Dense Bone because they are both typically present in abdominal region. In addition to those, we also added a 50 mL syringe filled with distilled water.

Acquisition parameters	**Abdominal protocol**
Voltage	GSI (80-140kVp)
Dose	13.33 mGy
Slice thickness/increment	2.5/1.25 mm
Scan Type	Helical
Rotation Time	0.5 s
Pitch	1.375
Reconstruction parameters	
DFOV	35 cm
Kernel	Standard
Algorithm	ASiR 50%

Table 5 : *Acquisition and reconstruction parameters for the abdominal protocol.*


A

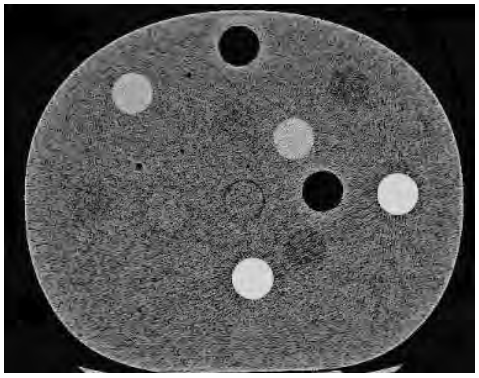

B

Figure 24: (A) *Photo of the CIRS phantom with selected inserts. A syringe filled with distilled water was inserted in the central hole for acquisitions.* (B) Effective *Z map of the CIRS phantom.*

Since, the manufacturer of the CIRS phantom provides the element composition of the various inserts, it was necessary to estimate reference Z_{eff} values for each material. There are many ways to estimate Z_{eff} [83], the more common is the simplistic power law approach [84](equation 44):

$$Z_{eff} = (\sum_i \alpha_i \, Z_i^m)^{\frac{1}{m}} \qquad \text{[Eq 44]}$$

With $\alpha_i = \frac{\frac{P_i Z_i}{A_i}}{\sum \frac{P_j Z_j}{A_j}}$, the fractional number of electrons of the ith element, Z_i the atomic number of the ith element and m, an exponent which could vary as a function of the energy beam considered. For low-energy beams, m=3.1 is generally used. Hence, the following reference values were calculated for the selected materials (see Table 6).

CIRS materials	Reference Zeff (from equation 42)
Adipose	6.37
Distilled water	7.45
Muscle	7.4
Trabecular bone	10.06
Dense bone	12.57

Table 6*: Reference Zeff values for CIRS materials of interest*

For each insert, 20 ROIs of 1.5 cm^2 were extracted along the volume of the insert. The mean and standard deviation of pixel values were calculated for each ROI and insert. To evaluate the accuracy of the Z_{eff} measurements, the deviation from the reference Zeff values was calculated.

Results

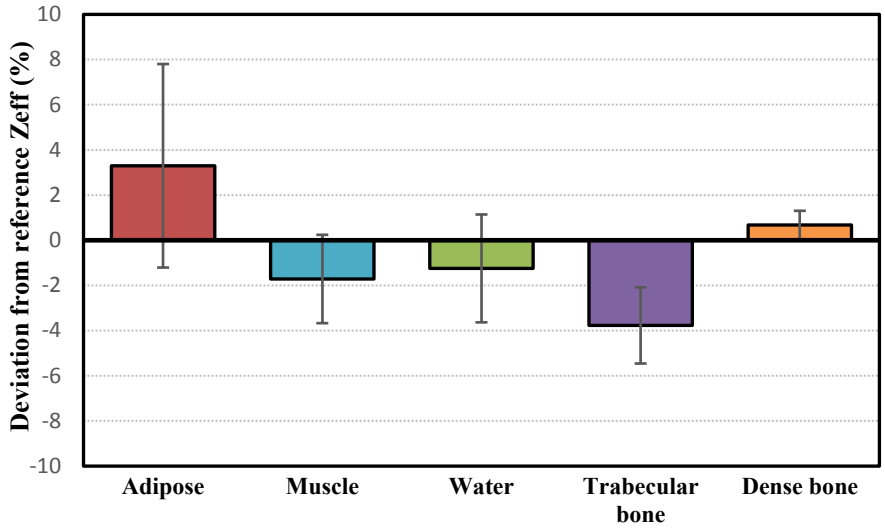


Figure 25: *Deviation from reference Zeff values for the various materials of interest. Error bars represent the 95% confidence interval of the mean deviation*

The deviation of Z_{eff} measurements from reference Z_{eff} values are represented in Figure 25. The measured mean values were close to the tabulated reference values, with a maximum deviation of -3.8% for the trabecular bone. The most accurate measurement was obtained for the distilled water, with a deviation of -1.2 % from the reference value. The uncertainties are higher for low Z_{eff} materials (adipose, muscle, water) than high Z_{eff} materials (dense bone).

This basic study should be performed for annual quality control of DECT systems for different dose levels, kernel, and reconstruction algorithms in order to ensure the accuracy of the quantitative Zeff measurements. To go a step further, specific materials involved in the different pathologies should be used to assess Z_{eff} accuracy for clinical protocols (iodine, iron, calcium, uric acid …). An adapted anthropomorphic phantom should be used to mimic patient attenuation.

4.2.2 *Iodine quantification in DECT*

Contribution

I was the sole contributor for the measurements and the analysis. A CT radiographer prepared the four diluted solutions of iodine. The methodology and the results are fully described in this part because the work is not associated with a peer-reviewed article.

Context

As shown in the previous section, DECT makes it possible to discriminate different materials based on their Z_{eff} values. Often, the material decomposition algorithm is applied to provide iodine-specific maps. In this case, the iodine density in a tissue can be expressed in mg/mL. The quantification of iodine in abdominal

imaging is very useful for furnishing information on the microvascular environment of the tissue. It can help to characterise and quantify lesion vascularity. For example, the detection of iodine in a tissue can characterise malignant renal masses and discriminate from hyper-attenuating renal cysts. Setting an iodine density threshold of 0.5 mg/mL to determine whether contrast enhancement is present in a renal mass [85] will necessitate an accurate and precise calibration of the DECT systems. Hence, the goal of this study was to determine the accuracy of iodine density quantification in an anthropomorphic abdominal phantom for two different DECT systems (DLCT and FSCT).

Methodology

Four diluted solutions of iodine at the following concentrations: 0.5, 1, 2, and 4 mg/mL were prepared by diluting a commercial contrast media solution (Iomeron 400 mg/mL, Bracco) with a NaCl solution at 0.9%. Four 20 mL syringes were filled with the various diluted iodine solutions. A supplemental syringe was filled with a solution of NaCl, and another was filled with distilled water (H2O) to serve as a reference without contrast media. The concentration values were set based on iodine concentration values measured on DECT patients' abdominal images. The concentration uncertainty for the various diluted iodine solutions were estimated at 0.15 mg/mL from the estimated volume uncertainty drawn into a 1mL syringe. The concentration uncertainty of the commercial contrast media solution was neglected.

The anthropomorphic abdominal QRM phantom (25 cm in diameter), containing the six syringes (see Figure 26), was scanned on an IQon DLCT (Philips) and a Discovery HD750 FSCT (GE HealthCare) with the venous phase of a routine multi-phase abdominal protocol. The display FOV were set to 35 cm for the two CT to obtain equivalent pixel sizes. For the DLCT system, two scans were successively acquired: the first one using a fixed dose level of 10 mGy, the second one using the ATCM. For the FSCT, only one acquisition at 10.76 mGy was performed. Indeed, FSCT systems don't offer the possibility to modulate the tube current. Parameters of the abdominal protocol are resumed in Table 7.

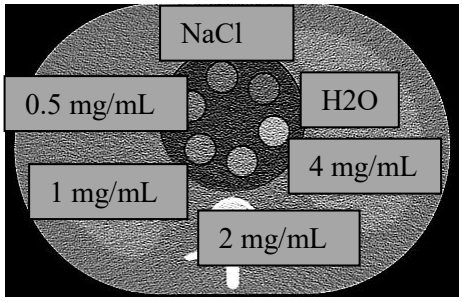


Figure 26: *DLCT acquisition of the abdominal phantom. The concentration of the syringes was indicated.*

An iodine-specific map was post-processed and transferred to a specific workstation, Spectral Diagnostic Suite (Philips) and AdvantageSim (GE Healthcare). For each iodine syringe, the mean and standard deviation values of 15 ROIs along the volume of the syringe were measured. Then, the overall mean and standard deviation was calculated for each syringe. The correlation between the known and the measured concentration values were analysed using a Pearson correlation. The Pearson coefficient was calculated using equation 45, where c^i_{mes} represents the various measured concentrations, c^i_{knowm} the various prepared iodine concentrations (0, 0.5, 1, 2 and 4 mg/mL).

$$r^2 = \frac{\sum_{i=1}^{N}(c_{mes}^{i} - \overline{c_{mes}})(c_{known}^{i} - \overline{c_{known}})}{\sqrt{\sum_{i=1}^{N}(c_{mes}^{i} - \overline{c_{mes}})^2}\sqrt{\sum_{i=1}^{N}(c_{known}^{i} - \overline{c_{known}})^2}} \quad [45]$$

To measure the overall iodine accuracy, the root mean square error (RMSE) was calculated using equation 46:

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N}(c_{mes}^{i} - c_{known}^{i})^2}{N}} \quad [46]$$

The uncertainty for the RMSE was estimated by randomly varying the known concentrations and the measured concentration values in the range of their standard deviations, considering a Gaussian distribution.

Acquisition parameters	**Protocol DLCT**	**Protocol FSCT**
Voltage	120 kVp	140 kV/80 kV (GSI)
Dose	10 mGy (fixed tube current) 4.6 mGy (ATCM)	10.76 mGy (fixed tube current)
Slice thickness/increment	2/1 mm	2.5/1.25 mm
Rotation Time	0.33s	0.7s
Pitch	1.17	1.375
Reconstruction parameters		
DFOV	35 cm	35 cm
Kernel	YB	Standard
Algorithm	iDose 3	ASiR-50%

Table 7: *Acquisition and reconstruction parameters of the abdominal protocol*

Results

Figure 27 shows the correlation between the known and measured concentration values for the different syringes. All measured concentrations showed a very strong agreement with iodine prepared concentrations. (r^2=0.996 for the 10 mGy DLCT acquisition, r^2=0.996 for the 4.6 mGy DLCT acquisition, r^2=0.999 for the FSCT acquisition). The various DECT systems can reveal actual changes in iodine concentration. For concentration values higher than 1 mg/mL, the measured concentration values were systematically lower than the known concentrations.

The RMSE showed a relatively good agreement for the iodine quantification for the three acquisitions (Figure 28). No significant difference for overall iodine quantification was noted between the DLCT and FSCT systems at the same dose level (one-tailed t-test, p=0.59). For the DLCT system, even if there is no significant difference between the two dose levels (one-tailed t-test, p=0.57), the lower the dose level, the higher the uncertainties on iodine quantification accuracy. Hence, a fixed iodine density threshold should not be set to determine whether contrast enhancement is present and characteristic of a malignant lesion. The uncertainty for iodine quantification with DECT should be taken into account.

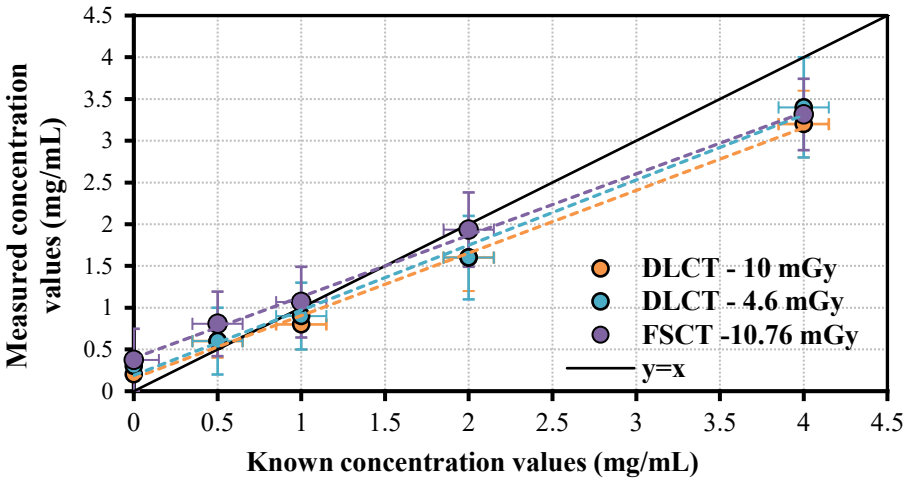


Figure 27: *Correlation between known and measured iodine concentration values*

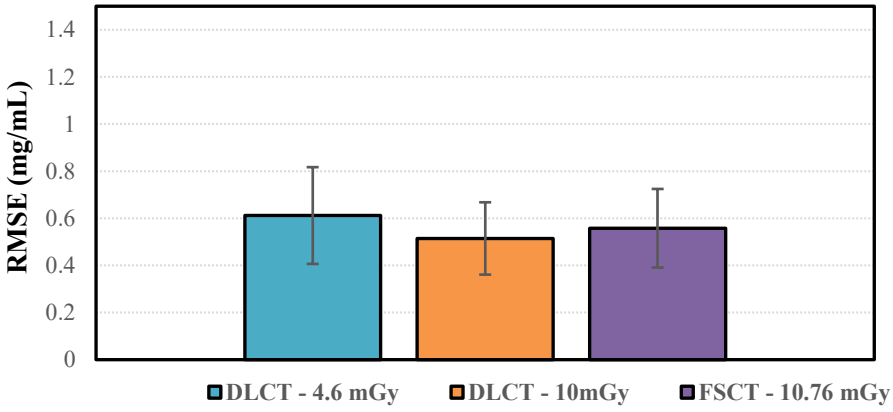


Figure 28: *Overall accuracy for iodine quantification for the 3 DECT acquisitions*

4.3 Task-based image quality assessment applied for a discrimination task in multi-energy spectral photon-counting CT

4.3.1 *In-vitro diagnostic performances for crystal-related arthropathies*

Contribution

The study was completed with the research team of Otago University in New Zealand (MARS collaboration), where a prototype of a small animal SPCCT was designed using a Medipix3RX chip developed by CERN. I did an internship there for one month to learn the functions of this device and carry out the measurements. I was the sole contributor for the analysis, which was developed based on the previous works for single energy CT.

Clinical context

Crystal-related arthropathies are joint disorders characterised by an accumulation of crystals in joints. Different types of crystals can be involved, monosodium urate (MSU), basic calcium phosphate (with hydroxyapatite,

HA and octacalcium phosphate, OCP) and calcium pyrophosphate (CPP). It is important to identify the crystals for diagnosis and management of the disease (by assessing the treatment response)[86,87]. In order to diagnose patients, a biopsy following by a polarizing microscopy are generally prescribed. Among imaging techniques, DECT have shown promising results in distinguishing MSU from other calcium crystals for patients with tophaceous gout (MSU deposition) [88]. However, DECT is still limited for use in early stages of gout [89] and no imaging technique is capable to non-invasively discriminate between calcium crystal types. Hence, the goal of this study was to assess SPCCT performances for the discrimination (1) between MSU and calcium crystals and (2) between HA, OCP and CPP calcium crystals. A secondary goal was to study the influence of the sample's concentration for the HA / CPP pair.

Materials

A homemade phantom composed of various concentrations of crystal samples was scanned with a SPCCT protocol for bone imaging. Crystal samples were made by blending crystal powder in agar. The concentrations were selected based on those found on clinical Dual-Energy CT images of patients affected by the various crystal-related arthropathies. Images were reconstructed using four different energy bins: 20-30 keV, 30-40 keV, 40-50 keV and 50-80 keV, taking advantage of the potential better discrimination between the various crystals for lower energy photons due to the higher contribution of the photoelectric effect (see Figure 29). Energy-resolving MARS CT of the phantom are displayed in Figure 29.

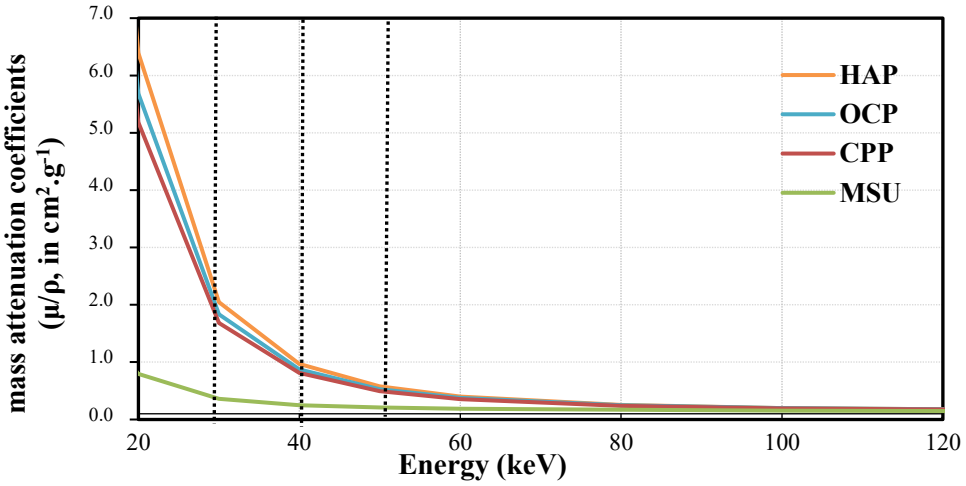


Figure 29: *Theoretical mass attenuation coefficients for the four crystals as a function of photon energy. The dashed lines indicate the separation between the different energy bins. Data computed from the National Institute of Standards database*

Task-based image quality assessment

The task associated to this study is a discrimination between different materials. As opposed to the previous studies where the detection task was performed on images, the discrimination assessment was performed directly on pixel values.

Firstly, a roc analysis was computed to assess the discrimination between two materials for each energy bin. For this purpose, AUC was chosen as a figure of merit, considering that the higher the AUC value, the better

the discrimination. Accuracy, sensitivity, and specificity were computed to assess diagnostic performances for this task. This analysis is included in the associated paper for Radiology (see section 5.3).

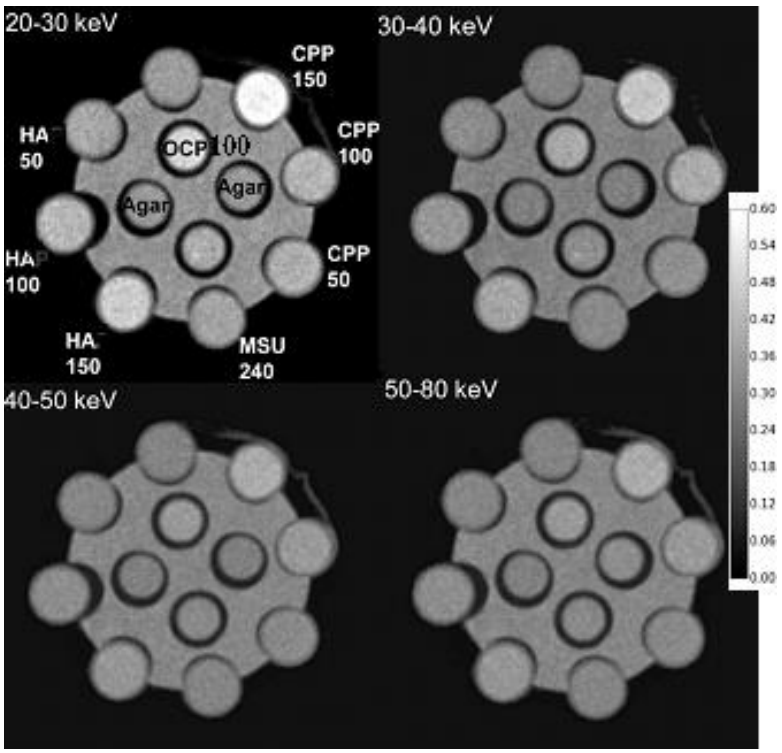


Figure 30: *MARS CT images of the phantom. The quantitative contrast scale corresponds to linear attenuation coefficients. On the top left image, the name and concentration of each crystal suspension in mg/mL were indicated.*

Secondly, a linear discriminant analysis (LDA) was performed to find a linear combination between energy bins which maximize the discrimination between the four materials. This method is generally used in statistics for machine learning or pattern recognition. It is interesting to notice that the Hotelling observer is based on this methodology. In this case, the methodology is applied to a multi-class problem, each class corresponds to a material. This analysis was not included in the Radiology article.

The original data can be combined in a matrix $X = \left[x_k^1 \ldots . x_k^{N_k}\right]$ where x_k^i is a vector containing the four energy bins information for the ith pixel from class k. X is a matrix 4xN where N is the sum of the number of pixels in each class (N_k). In a 4-class problem, we can find a matrix W which projects the data into a lower dimensional space 3xN and improve the class separability. The LDA can be summarized in 3 steps:

- Calculate the separability between the different classes (distance between the mean of each class), defined with the between class scatter matrix S_B.

  $S_B = \sum_{k=1}^{4} N_k(\mu_k - \mu)(\mu_k - \mu)^T$ with $\mu_k = \frac{1}{N_k}\sum_{i=0}^{N_k} x_k^i$ the pixels mean for class k and $\mu = \frac{1}{N}\sum_{i=0}^{N} x^i$ the total pixels mean for all the classes.

- Calculate the scatter of pixel values for each class, defined with the within class scatter matrix: $S_W = \sum_{k=1}^{4}(N_k - 1)K_K$ with K_K the covariance matrix for class k

- Calculate the optimal projection matrix W. We can find w (the column vectors of W) solving the following generalized eigenvalue equation: $\boldsymbol{S_W}^{-1}\boldsymbol{S_B}w = \lambda w$. The solutions are the eigenvectors corresponding to the 3 non-zero eigenvalues. This step is demonstrated in Annex 2 and 3.

Principal outcomes

Figure 31 shows the different ROC curves obtained for the discrimination between MSU and all calcium crystals for the four energy bins. The lower the energy range, the better the discrimination. For the two first energy bins, MSU and calcium crystals can be discriminated with an accuracy higher than 95%.

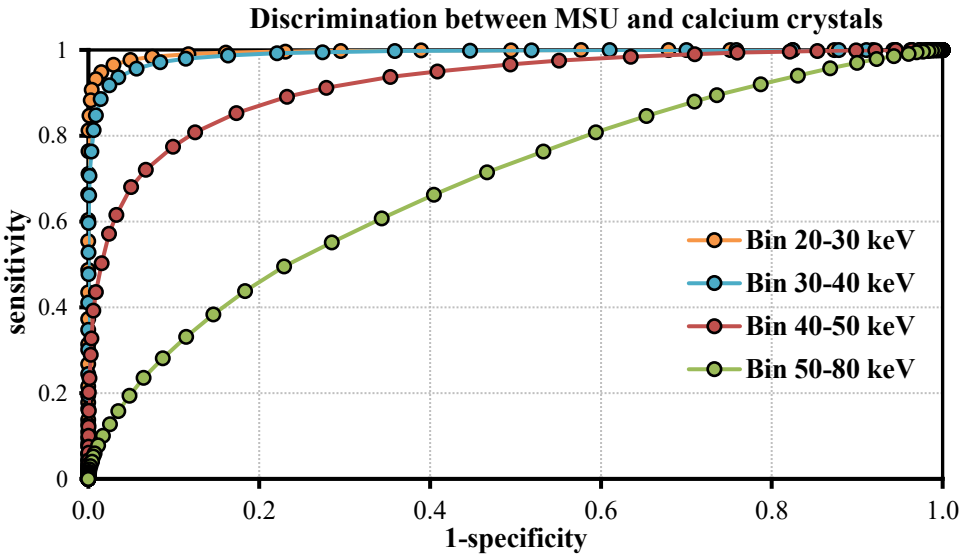


Figure 31: *ROC curves showing the discrimination between MSU and the mix of calcium crystals for the four energy bins.*

Comparing the different calcium crystals, the discrimination was moderate for the HA/CPP pair and for the HA/OCP pair for the first energy bin with an AUC equal to 0.80 and 0.78, respectively (see Figure 32). The accuracy reached 72.2% and 71%, respectively. Then, the discrimination performances decreased when increasing the energy range. For the OCP/CPP pair, the discrimination was poor, with an accuracy near to 55% whatever the energy bin considered.

Finally, Table 8 shows the influence of the sample's concentration for the HA and CPP pair. For an identical concentration between HA and CPP samples, the higher the concentration, the easier the discrimination. For the highest concentration, 150 mg/mL, HA was differentiated from CPP with an accuracy of 76.5%, whereas for the lowest concentration, 50 mg/mL, the discrimination between HA and CPP was more difficult with an accuracy of 62.9%.

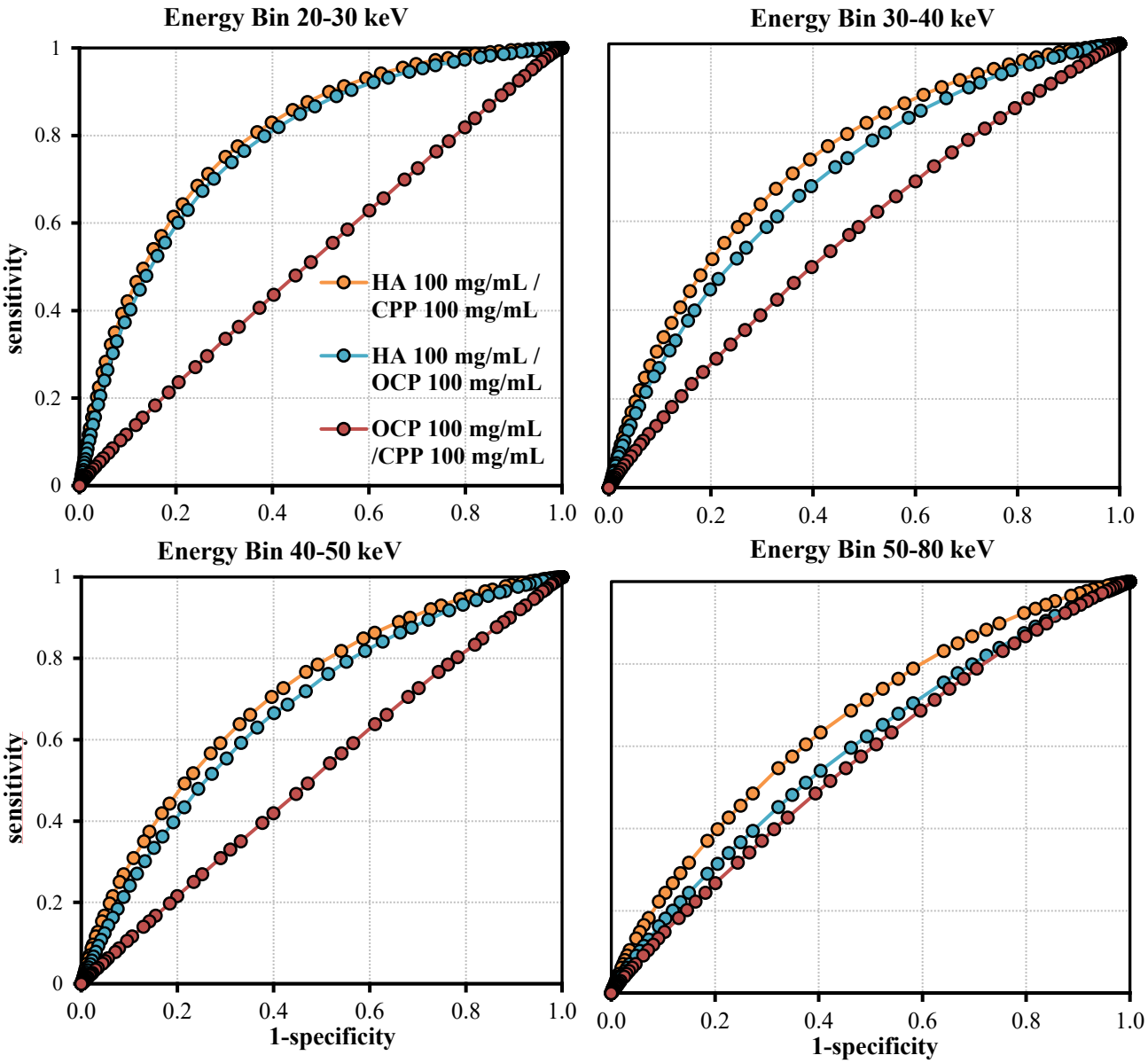


Figure 32 : *ROC curves showing the discrimination between HA and CPP, HA and OCP, OCP and CPP for the different energy bins. The concentration of the different calcium crystal suspensions was equal to 100 mg/mL.*

Accuracy (%)	CPP 50 mg/mL	CPP 100 mg/mL	CPP 150 mg/mL
HA 50 mg/mL	62.9 [62.6-63.2]	96.3 [96.0-96.4]	99.1[99.0-99.2]
HA 100 mg/mL	79.9 [79.4-80.3]	72.20 [71.7-72.6]	91.9 [91.7-92.1]
HA 150 mg/mL	95.7 [95.4 - 95.8]	59.8 [59.4-60.1]	76.5 [76.1-76.9]

Table 8: *Accuracy values for the discrimination between CPP and HA for three concentrations: 50, 100 and 150 mg/mL. 95% confidence intervals were detailed in brackets.*

Supplemental outcomes

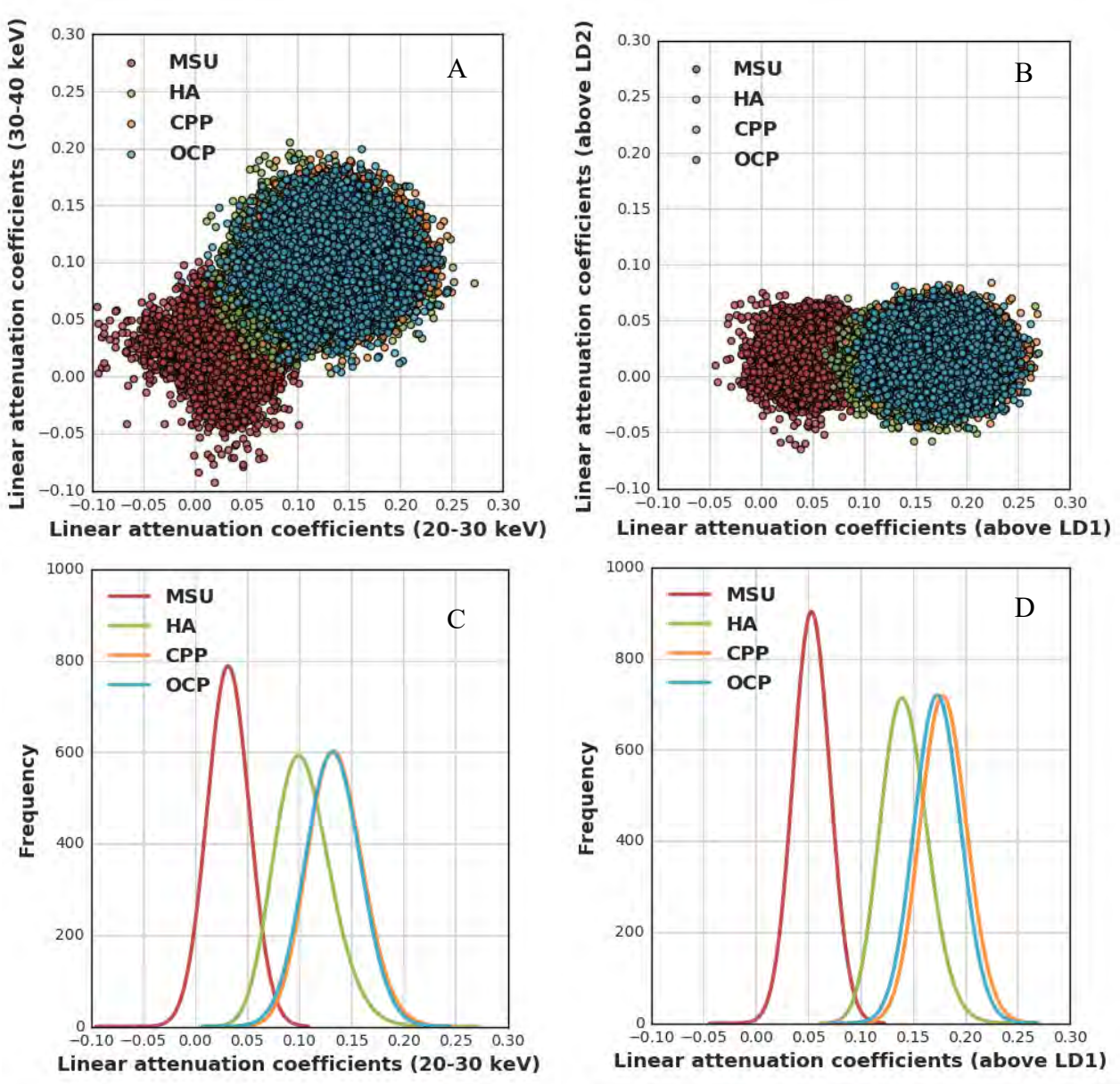


Figure 33:*(A) Scatter plot of the pixel values of the different crystals for the two first energy bins. (B) Scatter plot of the projected pixel values above the new feature subspace, (C) Histogram of pixel values for the first energy bin, (D) Histogram of pixel values for LD1.*

LDA methodology was computed to find the best linear combination between energy bins to maximize the discrimination between the four materials. Figure 33A shows a scatter plot of pixel values for the four materials and the two first energy bins, and Figure 33B shows a scatter plot of projected pixel values for the four materials above the new feature subspace, obtained by applying the LDA methodology.

LD1 is the first linear discriminant and corresponds to the projection of the data onto the eigenvector corresponding to the highest eigenvalue. LD2 is the second linear discriminant and corresponds to the projection of the data onto the eigenvector corresponding to the second highest eigenvalue. Note that the discrimination between materials is better along LD1 compared to LD2.

Hence, it is possible to compute the AUC values between the different pairs of crystals along LD1, and compare to the AUC values obtained using only the first or the second energy bin (see Table 9) . As expected, a better discrimination was obtained using LD1 than the first or second energy bin.

	Energy bin 1	Energy bin 2	LD1
MSU / HA	0.99 [0.98-1.0]	0.98 [0.98-0.99]	1.0 [0.99-1.0]
MSU / CPP	1.0 [0.99-1.0]	1.0 [0.99-1.0]	1.0 [0.99-1.0]
MSU / OCP	1.0 [0.99-1.0]	0.99 [0.99-1.0]	1.0 [0.99-1.0]
HA / OCP	0.78 [0.77-0.78]	0.70 [0.69-0.70]	0.88 [0.87-0.89]
HA / CPP	0.80 [0.79-0.82]	0.73 [0.72-0.74]	0.84 [0.83-0.85]
OCP / CPP	0.52 [0.50-0.54]	0.57 [0.56-0.58]	0.57 [0.56-0.59]

Table 9*: AUC results between the different pairs of crystals using the information from the first energy bin, the second, and the first linear discriminant LD1. 95% confidence intervals were detailed in brackets*

Finally, the in-vitro results suggest that MSU can be accurately discriminated from calcium crystals with SPCCT and could become a promising technique to efficiently diagnose patients with tophaceous gout. The moderate discrimination between calcium crystals, especially between HA and CPP is also promising. SPCCT could be the first imaging technique able to diagnose patients with CPP deposits (pseudo gout). Results should now be confirmed ex-vivo on anatomical pieces and later in vivo.

4.3.2 *Ex-vivo diagnostic performances for crystal-related arthropathies*

Contribution

This article gathers initial ex-vivo results on two anatomical pieces (an excised meniscus and a gouty finger) to determine the discrimination between MSU, HA and CPP. The study was conducted with the research team of Otago University. I was in charge of the preparation, acquisitions, and post-treatment of the excised meniscus and calibration phantoms.

Clinical context

Gout, the most common inflammatory arthritis, is caused by the presence of MSU crystals within the joint. Frequently, clinically indistinguishable with gout, pseudo-gout is characterised by a deposition of CPP crystals. The long-term management of these diseases requires the distinction between the different crystals involved. Diagnosing arthritis requires a biopsy and a visualization of the crystals via polarized microscopy. However, this technique remains operator-dependent and shows moderate sensitivity and specificity for crystals other than MSU [90]. Among imaging techniques, DECT have shown promising results to diagnose gout, by correctly identifying MSU crystals from soft tissues and calcium crystals. However, the performance is still

limited in the early stages of gout. Concerning pseudo-gout, no imaging technique is currently capable of correctly identifying CPP crystals from HA crystals. The goal of this study is to discriminate MSU, HA and CPP crystals in excised human specimens using the Mars SPCCT system.

Methodology

Calibration phantoms, containing known concentrations of MSU, CPP, HA, water, and lipid were scanned with the MARS SPCCT system using the bone imaging protocol. Concentrations of the various samples were indicated in Table 10. Different ROIs were extracted for each insert to measure the mean and standard deviation of the linear attenuation coefficient values. In order to compute the material decomposition algorithm, the calibration curves for each crystal were plotted (linear attenuation coefficient as a function of concentration for the different energy bins). Figure 34 shows a good linearity between the SPCCT response and the concentration for each crystal and each energy bin.

	CPP	HA	MSU
Concentration 1	50 mg/mL	50 mg/mL	74.5 mg/mL
Concentration 2	75 mg/mL	75 mg/mL	137 mg/mL
Concentration 3	100 mg/mL	100 mg/mL	160 mg/mL
Concentration 4	150 mg/mL	150 mg/mL	240 mg/mL

Table 10: *Concentrations of the various samples included in the calibration phantom for the material decomposition algorithm*

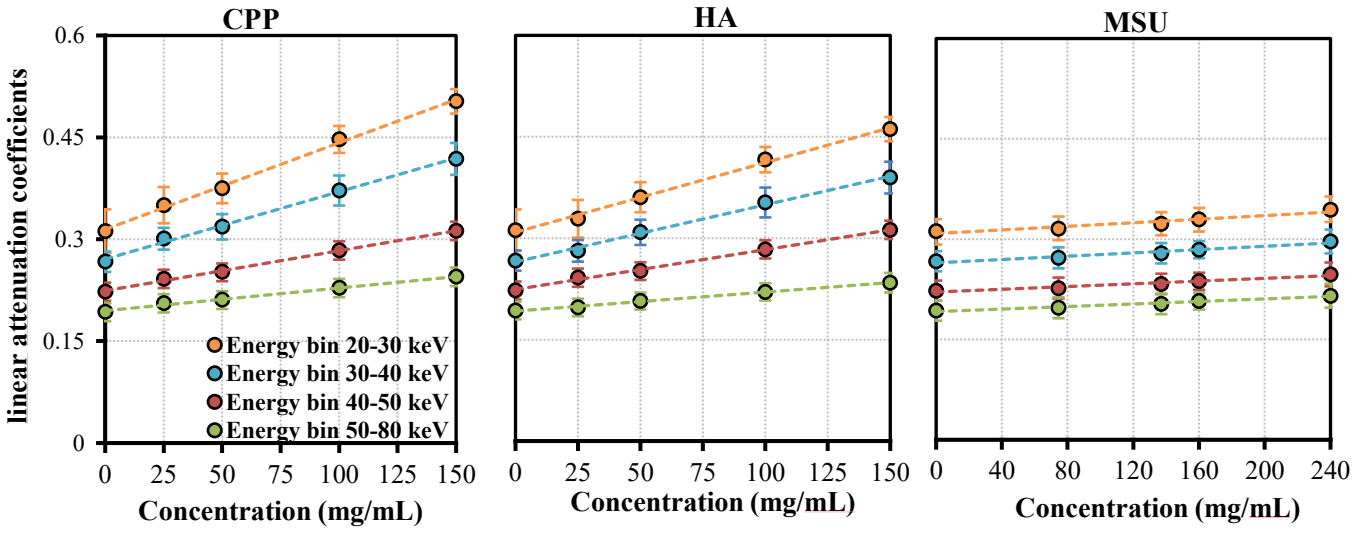


Figure 34: *Calibration curves for the material decomposition algorithm*

Then, two excised specimens, a finger with a gouty tophus and a meniscus, excised at the time of joint replacement surgery were scanned with the MARS SPCCT system using the same bone imaging protocol. Images of specimens were reconstructed with the material decomposition algorithm, previously calibrated with the calibration phantoms. To validate SPCCT results, crystals were scrapped from the two ex-vivo samples

and examined with polarized light microscopy and X-ray diffraction. The gouty finger was also scanned with a DECT.

Principal outcomes

Figure 35 shows the results obtained for the gouty finger. Plain X-ray radiography indicated the edge and soft tissue changes that could be consistent with tophus material. DECT and SPCCT identified MSU crystal deposits. The better spatial resolution of SPCCT compared to DECT enabled a visualisation of finer details. Polarized light microscopy confirmed the presence of MSU crystals in the finger.

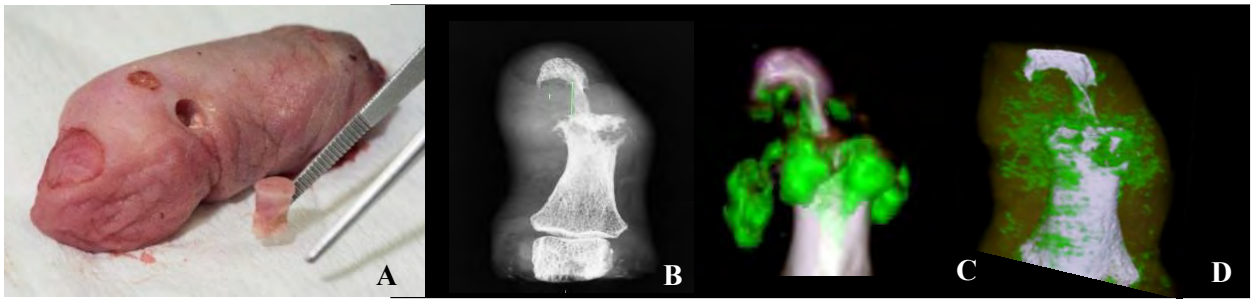


Figure 35: *(A) Photo of the finger with a gouty tophus, (B) Plain X-ray image, (C) DECT image, (D) SPCCT image.*

Figure 36 shows the results obtained for the excised meniscus obtained after total knee replacement surgery. Plain X-ray images revealed a chondrocalcinosis (pseudo-gout). SPCCT images show a predominance of CPP crystals in the calcified region of the meniscus. Polarized light microscopy and X-ray diffraction confirmed the presence of CPP in the meniscus.

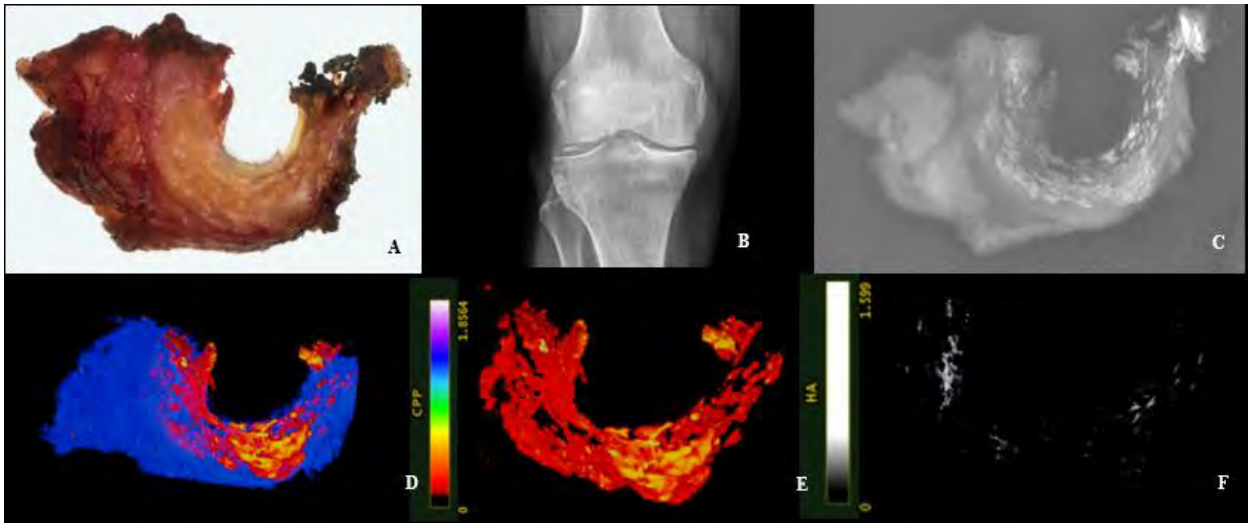


Figure 36: *(A) Photo of the excised meniscus, (B) Plain X-ray of the knee before surgery, (C) Plain X-ray of the meniscus, (D) SPCCT image, (E) SPCCT image showing only CPP deposits, (F) SPCCT image showing only HA deposits.*

Finally, our results confirmed the results obtained in-vitro. SPCCT could be an interesting imaging technique to diagnose crystal-related arthropathies. The results need to be completed with a larger number of human samples before to start the first human clinical trial when the large bore MARS CT will be developed.

5. Scientific articles

5.1 Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners.

Published in Medical Physics (2017)[91]

Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners

Damien Racine[a)]* and Anaïs Viry*
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

Fabio Becce and Sabine Schmidt
Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Rue du Bugnon 46, 1011 Lausanne, Switzerland

Alexandre Ba and François O. Bochud
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

Sue Edyvean
Medical Dosimetry Group, Centre for Radiation Chemicals and Environmental Hazards, Public Health England, Didcot OX11, UK

Alexander Schegerer
Department of Medical and Occupational Radiation Protection, Federal Office for Radiation Protection (BfS), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

Francis R. Verdun
Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland

(Received 30 September 2016; revised 3 February 2017; accepted for publication 7 March 2017; published 13 September 2017)

Purpose: We sought to compare objectively computed tomography (CT) scanner performance for three clinically relevant protocols using a task-based image quality assessment method in order to assess the potential for radiation dose reduction.
Methods: Four CT scanners released between 2003 and 2007 by different manufacturers were compared with four CT scanners released between 2012 and 2014 by the same manufacturers using ideal linear model observers (MO): prewhitening (PW) MO and channelized Hotelling (CHO) MO with Laguerre-Gauss channels for high-contrast spatial resolution and low-contrast detectability (LCD) performance, respectively. High-contrast spatial resolution was assessed using a custom-made phantom that enabled the computation of the target transfer function (TTF) and noise power spectrum (NPS). Low-contrast detectability was assessed using a commercially available anthropomorphic abdominal phantom providing equivalent diameters of 24, 29.6, and 34.6 cm. Three protocols were reviewed: a head (trauma) and an abdominal (urinary stones) protocol were applied to assess high-contrast spatial resolution performance; and another abdominal (focal liver lesions) protocol was applied for LCD. The liver protocol was tested using fixed and modulated tube currents. The PW MO was proposed for assessing high-contrast detectability performance of the various CT scanners.
Results: Compared with older generation CT scanners, three newer systems displayed significant improvements in high-contrast detectability over that of their predecessors. A fourth, newer system had lower performance. The CHO MO was appropriate for assessing LCD performance and revealed that an excellent level of image quality could be obtained with newer scanners at significantly lower dose levels.
Conclusions: This study shows that MO can objectively benchmark CT scanners using a task-based image quality method, thus helping to estimate the potential for further dose reductions offered by the latest systems. Such an approach may be useful for adequately and quantitatively comparing clinically relevant image quality among various scanners. © *2017 The Authors. Medical Physics published by Wiley periodicals, Inc. on behalf of American Association of Physicists in Medicine.* [https://doi.org/10.1002/mp.12224]

Key words: computed tomography, high-contrast spatial resolution, image quality, low-contrast detectability, model observers

1. INTRODUCTION

In most Western countries, the radiation exposure of the population due to computed tomography (CT) examinations has increased steadily for 20 yr.[1] A survey performed in 2006 in the United States showed that the average effective dose due to CT reached 1.5 mSv per capita, per year.[2] The last surveys performed in Switzerland in 2008 and 2013 showed a similar trend, with the average dose per capita from CT increasing from 0.8 mSv to 1.0 mSv within this 5-year period.[3,4]

In this context, the radiation protection requirements in diagnostic radiology (justification of the examination and optimization of the imaging protocol) need to be reinforced. The justification aspect is beyond the scope of this article. The optimization of a CT examination is achieved when image quality enables the clinical question to be answered while keeping patient radiation dose as low as reasonably achievable (ALARA).[5] This goal is, however, difficult to apply in practice. The quality of a CT examination depends on a wide range of parameters such as acquisition time, temporal resolution, and energy resolution when dealing with kV optimization or spectral CT imaging, and other factors. Thus, the actual determination of the clinical performance of a CT scanner is quite complex, and the clinical question needs to be clarified to enable a standard for image quality level to be set.[6] Task-oriented image quality criteria can then be used as surrogates for the assessment of actual clinical image quality.[7–10] They will necessarily be simple in comparison to the clinical situations, but will make it possible, for example, to predict the ability to detect simple structures of high and low contrast within homogeneous backgrounds.[9,10,11] This represents the most basic task that can be considered a surrogate for measuring clinical image quality.

To add complexity, making the task more realistic, one could then not only consider the detection but also the determination of the correct position of the detected structure. Then, the performance with which these tasks are performed could be assessed over more realistic structured backgrounds that mimic the actual anatomy. To go a step further, one could also check if the sizes and contrasts measured on the images correspond to the actual values. This strategy is still far from actual clinical image quality assessment, but may aid in the optimization of clinical protocols.

The aim of this study was to propose a way to objectively compare CT scanner performance using the simplest task-based image quality assessment: detection. This method was used in particular to evaluate the impact of technological developments on the potential for radiation dose reduction. In addition, we also wanted to investigate if major differences in performance existed among different manufacturers in the limited image quality criteria chosen. We compared the outcomes of four CT scanners released by the four major manufacturers from 2003 to 2007 with the outcomes of newer systems introduced from 2012 to 2014, using ideal model observers (MO) on three clinically relevant protocols.

2. MATERIALS AND METHODS

2.A. CT scanners and clinical protocols

This study was conducted using eight different CT scanners: two per major manufacturer including models released between 2003 and 2007 (referred as "older"), and models released between 2012 and 2014 (referred as "newer"). These eight CT scanners were, listed as "older"/"newer": LightSpeed VCT/Revolution CT (GE Healthcare, Milwaukee, WI, USA), Brillance 40/Ingenuity Core 128 (Philips Medical Systems, Best, the Netherlands), Somatom Sensation 64/Force (Siemens Healthcare, Forchheim, Germany), and Activion 16/Aquilion Prime (Toshiba Medical Systems, Tokyo, Japan). Basic characteristics and the image reconstruction methods used for this study are summarized in Table I.

For all CT systems, the displayed weighted computed tomographic dose index ($CTDI_w$) data were verified by measuring the normalized weighted computed tomographic dose index ($nCTDI_w$) using a 32-cm diameter CTDI test object and a 10-cm long CT pencil ionization chamber connected to an electrometer (model 1035-10.3 CTDI chamber and MDH model 1015 electrometer, Radcal, Monrovia, CA, USA), calibrated in RQR9 and RQA9 beams according to IEC 61267 and traceable to the Swiss Federal Office of Metrology.[12] The volume CT dose index ($CTDI_{vol}$) is defined as the $CTDI_w$ divided by the helical pitch factor; the values used in this study were taken directly from the displayed ones.

The image acquisition protocols used to compare the performance of the CT units were proposed by a panel of four senior radiologists working in three different University Hospitals in Switzerland. Among a large number of clinically relevant protocols we focused on three: two requiring a relatively high level of spatial resolution for the detection of high-contrast structures in the head and abdomen, and one requiring a high level of low-contrast resolution in the abdominal region. When dealing with the assessment of low-contrast resolution performance in the abdominal region, two approaches were chosen: one using fixed dose levels (5, 10, and 15 mGy), with the 15 mGy dose level corresponding to the Swiss abdominal CT Diagnostic Reference Level (DRL) using one phantom size; and the other using the tube current modulation option (with the local settings used for the clinical indication of the acquisitions) using three phantom sizes.[13] The details of the acquisition parameters used for each protocol are given in Table I. For technical reasons, the acquisition parameters used were not exactly the same.

2.B. Phantoms and data acquisitions

2.B.1. High-contrast performance

The assessment of high-contrast performance for head and abdominal protocols was made using a dedicated custom-made phantom containing cylindrical rods of different contrast materials (Teflon® or polytetrafluoroethylene [PTFE], polyethylene, and polymethylmethacrylate [PMMA])[14] (Fig. 1). The edge of this internal cylinder, at different z-axis positions, is used as the interface for the high-contrast numerical evaluation. The external diameter of this phantom is 250 mm; the high-contrast internal cylinder diameter is 100 mm.

2.B.2. Low-contrast performance

A modified anthropomorphic abdominal phantom (QRM 401, QRM, Moehrendorf, Germany) (Fig. 2) was used to investigate the low-contrast resolution performance of the CT units. It is made of calibrated tissue-equivalent material.

TABLE I. Acquisition parameters for each clinical protocol.

Manufacturer	GE Healthcare		Philips		Siemens		Toshiba	
CT Scanner	Light speed	Revolution	Brilliance	Ingenuity core	Sensation	Force	Activion	Aquillion
Release year	2005	2014	2006	2011	2003	2012	2007	2012
Rows	64	256	64	256	64	192	16	80
Algorithms	FBP	ASIR-V	FBP	Idose	FBP	Admire	FBP	AIDR 3D
Voltage	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV	120 kV
Acquisition mode	Helical	Helical	Helical	Helical	Helical	Helical	Helical	Helical
Slice Thickness	2.5 mm	2.5 mm	2.5 mm	2.5 mm	3 mm	3 mm	3 mm	3 mm
Head HR: Trauma								
FOV	25 cm							
$CTDI_{vol}$	15 mGy							
Kernel	Bone Plus	Bone Plus	D	D	H70h	Hr64h	FC80	FC30
Pitch	0.984	0.969	1	1	1	1	0.825	0.825
Abdomen HR: Search for urinary stones								
FOV	25 cm							
$CTDI_{vol}$	15 mGy							
Kernel	Standard	Standard	B	B	B31s	Bf40s	FC18	FC18
Pitch	1.375	1.375	1.295	1.295	1.4	1.4	1.475	1.475
Abdomen LCD: Detection of focal liver lesions								
FOV	32 cm							
$CTDI_{vol}$	5-10-15 mGy							
Kernel	Soft	Soft	B	B	B31s	Bf40s	FC18	FC18
Pitch	0.984	0.969	1	1	1	1	0.825	0.825
FOV	32-36-42 cm							
$CTDI_{vol}$	Automatic tube current modulation							
Kernel	Soft	Soft	B	B	B31s	Bf40s	FC18	FC18
Pitch	1.375	1.375	1.295	1.295	1.4	1.4	1.475	1.475

The body of the phantom (equivalent diameter of 24 cm) contains muscle, liver, spleen, and bone (vertebrae) tissue equivalents. A module can be inserted into the phantom body that includes spheres of different sizes: 8, 6, 5, 4, and 3 mm; each size having a contrast 20 HU relative to the background at 120 kV. For practical reasons, only three spheres (5, 6, and 8 mm) were used.[15] Two additional annuli (increasing the phantom's effective diameter to 29.6 cm and 34.6 cm, respectively) were added to simulate a range of body habitus (from an approximate patient weight of 50 kg for the equivalent diameter of 24 cm to 75 kg and 100 kg for the equivalent diameter of 29.6 cm and 34.6 cm).

Ten successive scans of the phantom fixed in place were performed to obtain 40 regions of interest (ROIs) with the spheres, and 100 ROIs without any target. This phantom was scanned using two protocols. First, the small phantom was

FIG. 1. Photo and sagittal view of the TTF phantom filled with water with its different contrast rods: PTFE, polyethylene, and PMMA.

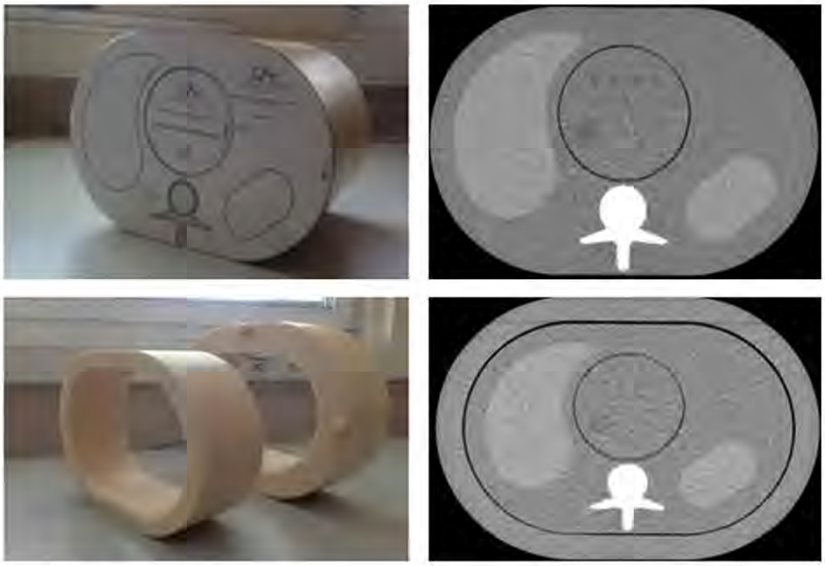


FIG. 2. Pictures of the QRM 3D LCD phantom and additional annuli (left) and example of images (right) provided by the QRM abdominal LCD phantom (equivalent diameters of 24 and 29.6 cm).

scanned at three dose levels to assess the baseline values of the CT scanners. Then, datasets were again acquired, this time with the two additional rings installed, to investigate the effect of body habitus on low-contrast detectability (LCD).

2.C. Image quality metrics

2.C.1. High-contrast performance

Spatial resolution: The parameter usually used to assess spatial resolution when dealing with CT images is the Modulation Transfer Function (MTF). However, iterative reconstruction (IR) algorithms are known to be highly nonlinear and therefore might introduce a dependency of the image contrast over the spatial resolution. Boone[16] and Richard et al.[17] proposed target transfer function (TTF) metrics to overcome this problem by characterizing spatial resolution taking into account contrast properties. MTF and TTF are similar metrics, except that TTF may be applied on different contrast levels. In this study we took a similar approach. Using the rod phantom, TTF was calculated for each rod from the radial mean of the edge spread function (ESF) profiles. The ESF's raw data were fitted and analytically differentiated to provide line spread functions (LSFs). Finally, performing a Fourier Transform on the LSFs gave the TTFs, which were normalized to 1 at the zero frequency. More details about the methods can be found in Ott et al.[14]

Noise power spectrum: The rod phantom also allows the assessment of the noise power spectrum (NPS). ROIs of 100 × 100 pixels, which were located in the center of a homogeneous region from 10 images, were used to calculate the NPS. The 2D NPS was computed using the following equation:

$$NPS_{2D}(f_x,f_y)=\frac{\Delta_x\Delta_y}{L_xL_y}\frac{1}{N_{ROI}}\sum_{i=1}^{N_{ROI}}\left|FT_{2D}\{ROI_i(x,y)-\overline{ROI_i}\}\right|^2 \quad (1)$$

Where Δ_x, Δ_y are the pixel sizes in the x and y dimensions, L_x, L_y are ROI sizes for the two directions ($L_x = L_y = 100$ pixels), N_{ROI} is the number of ROIs ($N_{ROI} = 10$), and $\overline{ROI_i}$ is the mean pixel value for the ith ROI. The 2D NPS was then radially averaged to provide the 1D NPS_{1D} according to the methodology presented in ICRU 87.[18]

Prewhitening model observer: To perform a task-based image quality assessment of high-contrast structures, the detectability index (d′) of different diameters structures having a nominal contrast of 1080 HU at 120 kVp (PTFE/water), 120 HU (PMMA/water), and −80 HU (polyethylene/water) was computed using the prewhitening mathematical model observer, PW,[19] and to reduce inconsistencies due to the use of iterative reconstructions, the MTF function, that should be used in that model, was replaced by the TTF function (see Eq. 2)[14].

$$d' = \sqrt{2\pi}|\Delta HU|\sqrt{\int_0^{f_{Ny}}\frac{S^2(f)TTF^2(f)f}{NPS_{1D}(f)f}df} \quad (2)$$

where f is the frequency, f_{Ny} is the Nyquist frequency of the image, $|\Delta HU|$ is the absolute contrast difference between the signal and the background and S(f) is the Fourier transform of the input signal, $S(f) = \frac{R}{f}J_1(2\pi Rf)$ with J_1, a Bessel function of the first kind. In our study, the rod phantom was used to provide the estimation of TTF and NPS, which are needed for PW MO, but not for the direct measurement of the small size disks' high-contrast detectability. To overcome

this limitation, we simulated a virtual disk with a radius R varying from 0.5 mm to 2.5 mm.

It is of note that scatter reduces not only image contrast but also the amplitude of the TTF. It was decided to take into account the scatter effect by using the measured contrasts rather than the nominal ones. Thus, TTFs were fitted to avoid the effect of scatter (spatial resolution drop in the low frequency range) as presented in reference.[14]

The uncertainty of the PW outcome was assessed by varying randomly the contrast and TTF values in the range of their standard deviations, considering a Gaussian distribution, measured on 30 images. The NPS parameter was not considered due to the fact that its uncertainty is negligible compared to that of the contrast and TTF parameters.

2.C.2. Low-contrast performance

Channelized Hotelling observer: LCD was evaluated in the image domain using a channelized Hotelling observer (CHO), with Laguerre-Gauss (LG) channels. This model is an estimation of the Hotelling observer, which itself is the ideal linear MO. The use of LG channels is appropriate in this case because they are known to maximize task performance.[20,21] The computation was made up to the tenth order of the LG polynomials and for one orientation only (due to the circular symmetry of the structure to be detected), resulting in a total of 10 channels. The pth channel (u_p) is obtained by multiplying the Laguerre polynomial at the pth order by a Gaussian function:

$$\boldsymbol{u}_p(r|a_u) = \frac{\sqrt{2}}{a_u}\exp\left(\frac{-\pi r^2}{a_u^2}\right)L_p\left(\frac{2\pi r^2}{a_u^2}\right) \tag{3}$$

where L_p is a Laguerre polynomial, r is a two-dimensional spatial coordinate, and a_u is the width of the Gaussian function (taken to be = 9 in the present study).

Laguerre polynomials are defined by:

$$L_p(x) = \sum_{k=0}^{p}(-1)^k\binom{p}{k}\frac{x^k}{k!} \tag{4}$$

The image is passed through the 10 LG channels. The channel output is a scalar $\mathbf{v}_i$ obtained by the dot product between the channel u_p and the image g:

$$v_i = U^T g \tag{5}$$

Where U represents the matrix of the channels, each column is one of the 10 channels:

$$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \boldsymbol{u}_{10}] \tag{6}$$

The CHO is then computed from the template w_{LG}:

$$w_{LG} = \left(K_{v/n}\right)^{-1}(\langle v_s\rangle - \langle v_n\rangle) \tag{7}$$

where $(K_{v/n})$ is the covariance matrix calculated from 100 signal absent images as perceived through the channels (channelized images).$\langle v_s\rangle$ represents the mean of 40 channelized signal images and $\langle v_n\rangle$ the mean of 100 channelized absent signal images.

The decision variable λ_{LG} of the CHO model is obtained by combining the template w_{LG} and the channelized image v_i:

$$\lambda_{i,LG} = w_{LG}^T v_i \tag{8}$$

In the end, the MO was tested with the same set of images as with the training set although this could overestimate its performance.[22] For each category (lesion and phantom size as well as dose levels), a receiver operating characteristic (ROC) curve was calculated with 50 threshold levels.[23] To summarize the information, an area under the ROC curve (AUC) was calculated using the trapezoidal method. The average and standard deviation of the model observers were estimated by performing a bootstrap method.[24] In practice, 500 ROC experiments were performed for each category.

For each dose level and phantom size, we used an image quality metric called "AUC_w" which combines the AUCs for lesions of different sizes. This metric is computed thus:

$$\mathrm{AUC_w} = \frac{\sum_{i\in(8;6;5)}\frac{\mathrm{AUC_{lesion(i)}}}{i}}{\sum_{i\in(8;6;5)}\frac{\mathrm{AUC_{lesion(i,max)}}}{i}} \tag{9}$$

where i represents the lesion sizes: 8, 6, or 5 mm, $AUC_{lesion(i)}$ represents the outcome of model observer for each lesion size. With such a definition, $AUC_{lesion(I,max)}$ corresponds to the value of this metric when the performance is maximal for each lesion size ($AUC_{lesion(I,max)} = 1.0$).

3. RESULTS

To ensure the impartiality of this work, the results are reported in an anonymous manner consistently throughout the manuscript. A capital letter (A, B, C, and D) was assigned to each manufacturer and the lower case letter "a" and "b" was added for respectively "newer" and "older" CT units. Differences between the displayed and measured $CTDI_{vol}$ were within 15%, in conformity with Swiss legal requirements (limit of ± 20%).

3.A. Image quality for high-contrast structures

3.A.1. High-contrast detection for the head protocol

For the detection of high-contrast structures, a d′ was calculated for different contrast values for each CT using a head protocol. As expected, the d′ increased with the diameter and the nominal contrast of structures to be detected (Figs. 3(a)–3(c).

Comparison of performance of new and old scanner models from each manufacturer: Figure 3(a) shows that for manufacturers A, C, and D, there was a noticeable improvement of the detectability when switching from the older CT scanners to newer ones while a slight reduction was observed for manufacturer B. The largest improvement was observed for manufacturer C (283% for lesions from 3–5 mm), whereas moderate improvements were found for manufacturers A and D (18% and 37%, respectively).

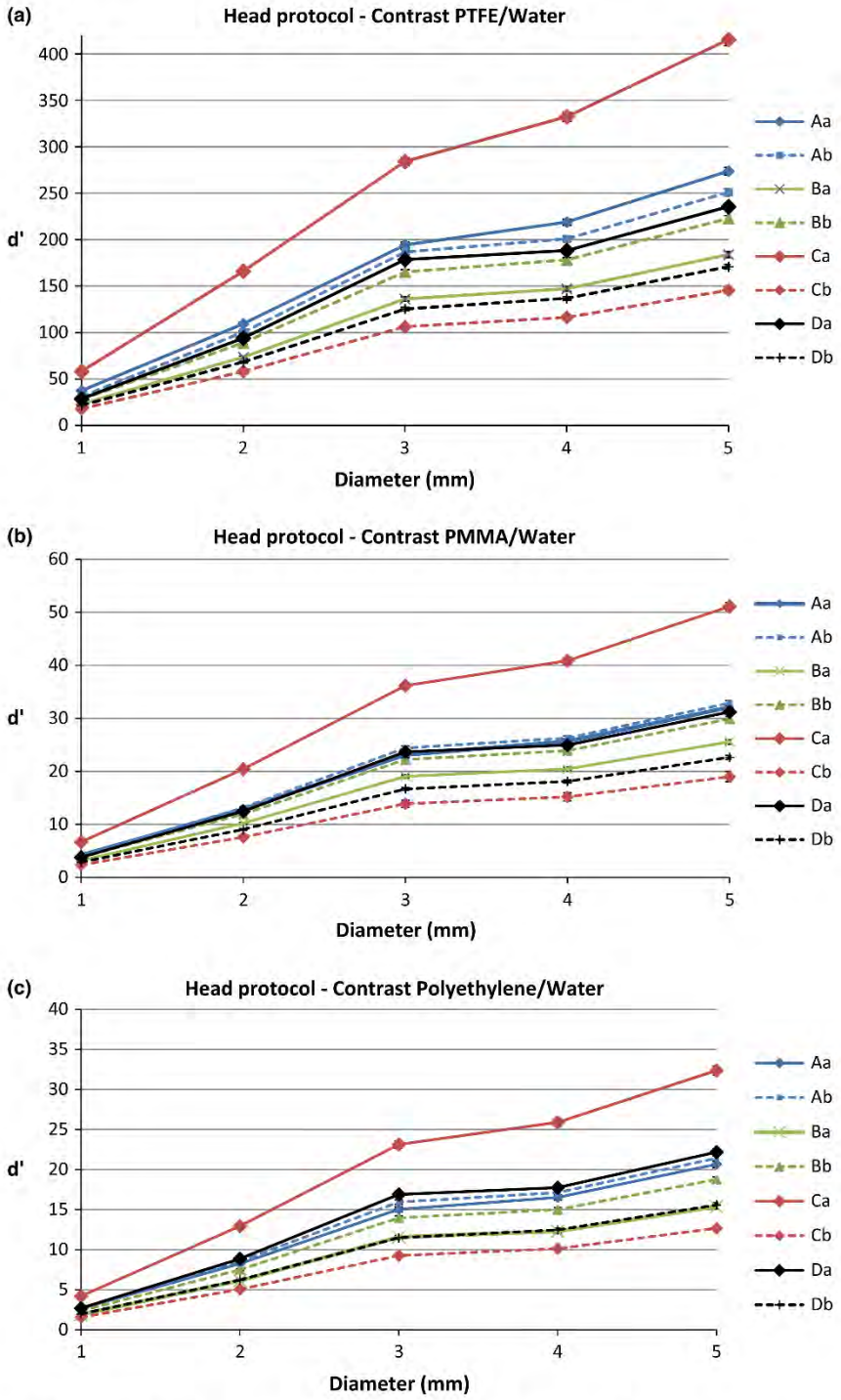


FIG. 3. (a) Detectability index (d′) calculated with the PW model for a nominal contrast of 1080 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs. (b) Detectability index (d′) calculated with the PW model for a nominal contrast of 120 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs. (c) Detectability index (d′) calculated with the PW model for a nominal contrast of −80 HU. The horizontal axis represents the diameter of structures. Dotted lines represent older CTs and solid lines represent newer CTs.

In Figs. 3(b) and 3(c), similar behavior was observed for manufacturers B, C, and D. For manufacturer A, no major difference appeared between the older and newer CT units.

Differences between manufacturers (newer CT models): For the three contrast levels tested with the newer CT models presented in Figs. 3(a)–3(c), the d′ reached the highest value

for manufacturer C, and manufacturers A and D provided better results than manufacturer B.

High-contrast detection for the abdomen protocol: The same methodology was applied to assess the detectability of high-contrast structures for the abdominal protocol.

Comparison of performance of new and old scanner models from each manufacturer: In Fig. 4(a), for the highest contrast level, detectability improved when switching from older to newer CTs for manufacturers A, C, and D. A major improvement was noted for manufacturer C (86%) and a moderate improvement was noted for manufacturer A and D (23% and 40%, respectively). For manufacturer B, a trend similar to the one identified in the head protocol was observed (Fig. 3(a)).

In Figs. 4(b) and 4(c), similar behaviors were observed for manufacturers B, C, and D. For manufacturer A, smaller differences appeared between the older and newer CT units for materials PMMA and polyethylene than for PTFE.

Differences among manufacturers (newer CT models): For each contrast level and each structure's diameter, the results of the comparison of newer CT models was very similar to the results for the head protocol. Manufacturer C reached the highest performance. Manufacturers A and D provided better results than manufacturer B.

3.B. Image quality for low-contrast detectability

3.B.1. Abdomen low-contrast detection — $CTDI_{vol}$ variation

Imaging the small abdomen phantom with a $CTDI_{vol}$ of 15 mGy (Fig. 5) showed no major differences among the various scanners. Reducing the $CTDI_{vol}$ to 10 mGy, the image quality metrics slightly decreased for all scanners (AUC_w going from 1.0 to 0.985), with a larger reduction observed for scanner "Db" (AUC_w going from 1.0 to 0.945). These variations are statistically significant as an uncertainty of 0.003 was set for these measurements ($P < 0.05$). At the lowest $CTDI_{vol}$, we investigated (5 mGy), all newer scanners provided better results than the older ones except for scanner "Aa".

To investigate the robustness of the method used, the measurements were repeated five times on the same scanner "Da" using the small abdomen phantom with a $CTDI_{vol}$ of 5 mGy (Fig. 5) under "positioning uncertainties," and demonstrate that comparable results could be obtained when repositioning the phantom several times.

To investigate how the method would vary when characterizing various scanners of the same type, the methodology was applied on five different "Da" scanners (Fig. 5), and represented as "CT machine uncertainties." Comparable results were found with different machines of the same type.

3.B.2. Abdomen low-contrast detection – Phantom size variation

Using automatic tube current modulation and the small abdominal phantom, Fig. 6 shows that it is possible to reach a similar level of image quality for all scanners (differences within 5%). However, this high level of image quality is obtained at noticeable different $CTDI_{vol}$ values (almost 300%).

When using the medium abdominal phantom a significant drop in image quality is observed for three scanners ("Db", "Bb", and "Ba"). For the other scanners, comparable image quality is preserved but again within a large range of $CTDI_{vol}$ values (Fig. 7).

Finally, when using the largest anthropomorphic abdominal phantom, large differences in behaviors were observed (Fig. 8).

Comparison of performance of new and old scanner models from each manufacturer: For all manufacturers but one, major improvements were demonstrated with the newer models. It was possible to reach similar image quality levels at significantly lower $CTDI_{vol}$ levels. For manufacturer A, image quality level was slightly decreased but the patient exposure was reduced by 50%. For manufacturer C, a noticeable improvement in image quality was obtained at less than half the dose from the older model. For manufacturer D, a major improvement of image quality was obtained with a lower dose reduction than manufacturers A and C (30%). The only manufacturer where no major improvement was noted was manufacturer B, where similar image quality level was obtained at a slightly higher $CTDI_{vol}$ value.

Differences between manufacturers (newer CT models): When using the largest size of the phantom (simulating a patient of 100 kg), all newer CT scanners reached a high level of image quality ($AUC_w > 0.850$). Nevertheless, this level of image quality was reached with $CTDI_{vol}$ differences within a range of 300%.

4. DISCUSSION

A full characterization of CT scanner units would require the assessment of a large number of parameters. Among these parameters one could mention: the acquisition time, the standard high- and low-contrast resolutions, the temporal resolution, and the energy resolution when dealing with kV optimization or dual energy imaging.

We chose to use simple task-based image quality assessment methodologies that do not include the whole range of potential performance of the scanners. We assessed the performance regarding image quality of high- and low-contrast structures

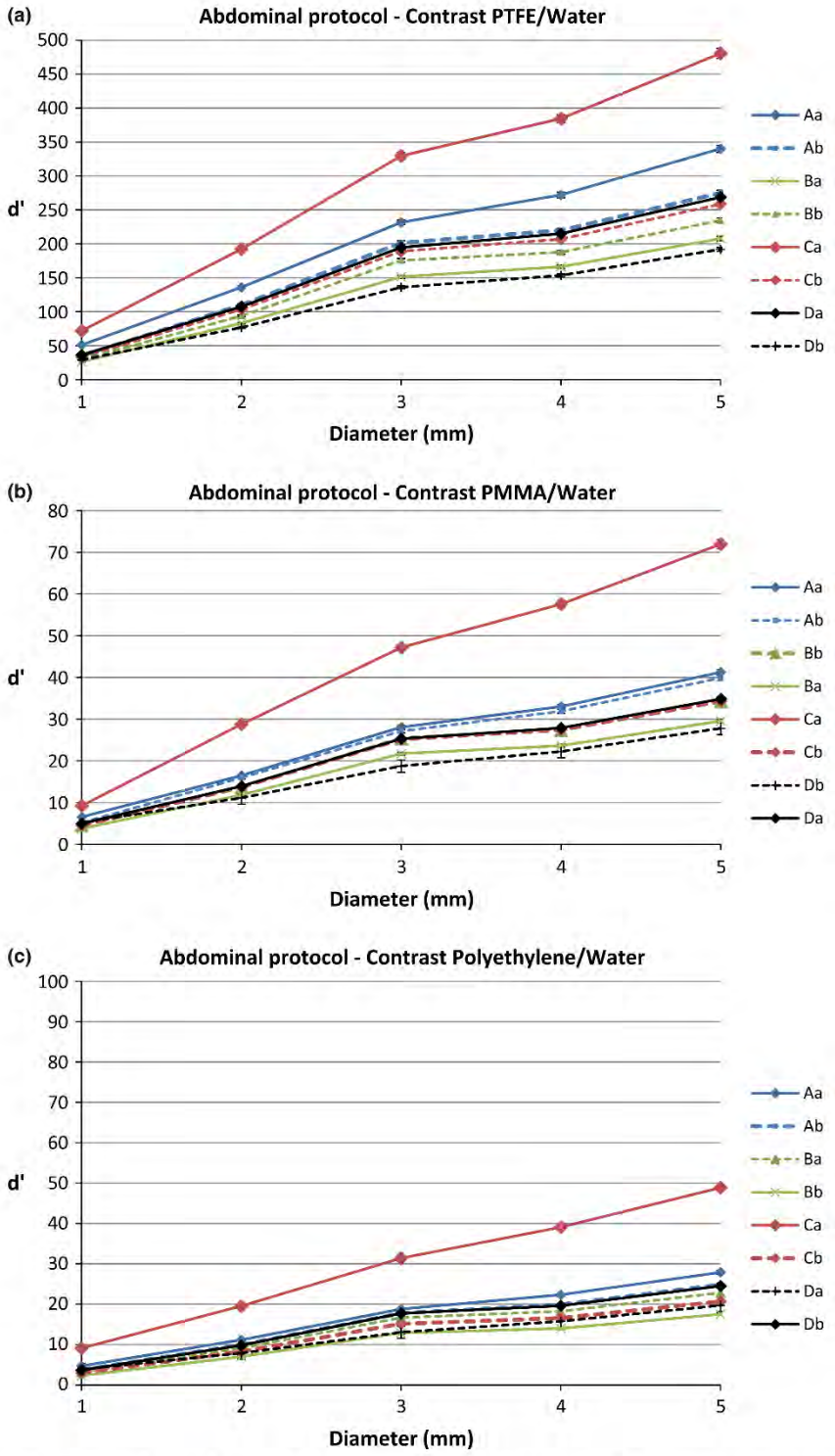


FIG. 4. (a) Detectability index (d′) calculated with the PW model for a nominal contrast of 1080 HU. The horizontal axis represents the diameter of structures. (b) Detectability index (d′) calculated with the PW model for a nominal contrast of 120 HU. The horizontal axis represents the diameter of structures. (c) Detectability index (d′) calculated with the PW model for a nominal contrast of −80 HU. The horizontal axis represents the diameter of structures.

using ideal MO for eight CT systems and three clinical protocols. This benchmark provided a large panel of image quality levels for older (2003–2007) and newer CTs (2011–2014).

The first aim of the study was to investigate if technological improvements over time could be shown using our limited set of image quality criteria. For the high-contrast

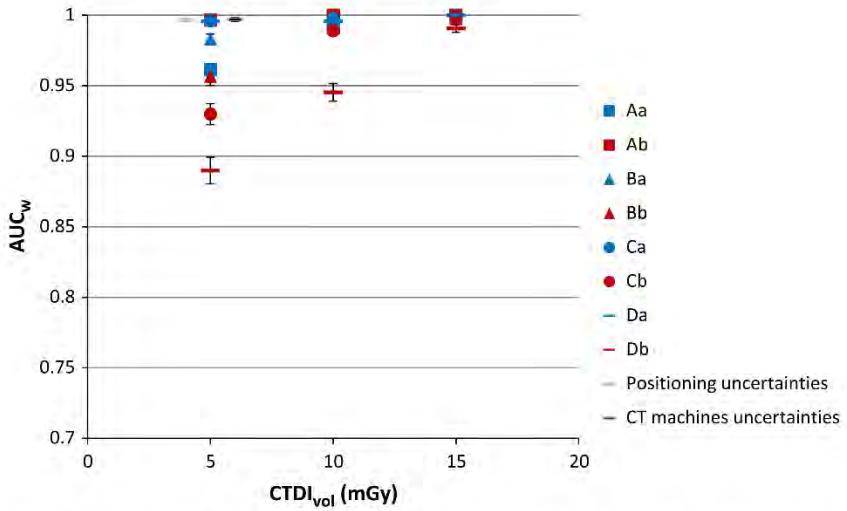


FIG. 5. AUC_w to detect a sphere in soft tissue equivalent at three $CTDI_{vol}$ values using the smallest anthropomorphic abdominal phantom with fixed tube current. The vertical axis is the AUC_w for each of the eight CTs. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, and the blue color represents the newer ones. AUC_w = weighted area under the curve $CTDI_{vol}$ = volume CT dose index.

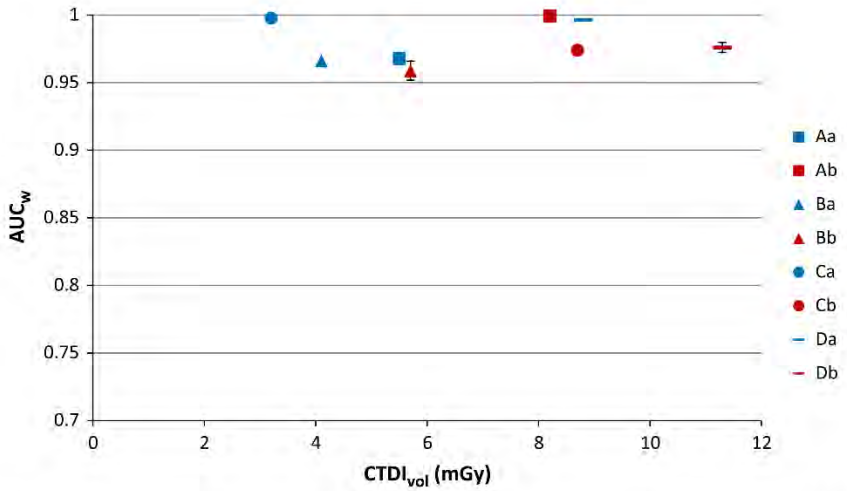


FIG. 6. AUC_w to detect a sphere in soft tissue equivalent using the smallest anthropomorphic abdominal phantom at different $CTDI_{vol}$ levels using automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC_w outcome. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC_w = weighted area under the curve, $CTDI_{vol}$ = volume CT dose index.

detectability, scanners could be discriminated using the PW MO, and performance improvement was noted for manufacturers A, C, and D. The d' values were systematically very high, indicating that the detection of a structure > 2 mm in diameter with such a nominal contrast value was trivial. For better discrimination one could add complexity to the task, for example: the estimation of shape, size, and contrast. Concerning the low-contrast resolution, performance improvements were observed also for three (manufacturers A, C, and D) out of four manufacturers with a drastic dose reduction to reach similar high image quality levels.

The second aim of our study was to investigate if major differences in performance existed between newer CT scanners of various manufacturers. For the limited criteria chosen in this study, manufacturer C reached the highest performance with the chosen reconstruction kernels. However, our measurements have two limitations: the first one deals with the choice of the reconstruction kernel that could not be the same for all manufacturers.[25]

To investigate if some kernels used in this study would give advantage to a particular range of target sizes, we computed the d' values for structures 1–5 mm in diameter. A regular increase in d' with the object diameter was noted for all manufacturers. In addition, the clinical parameters used for the high-contrast resolution would not necessarily represent the maximum theoretical capability of the system. The

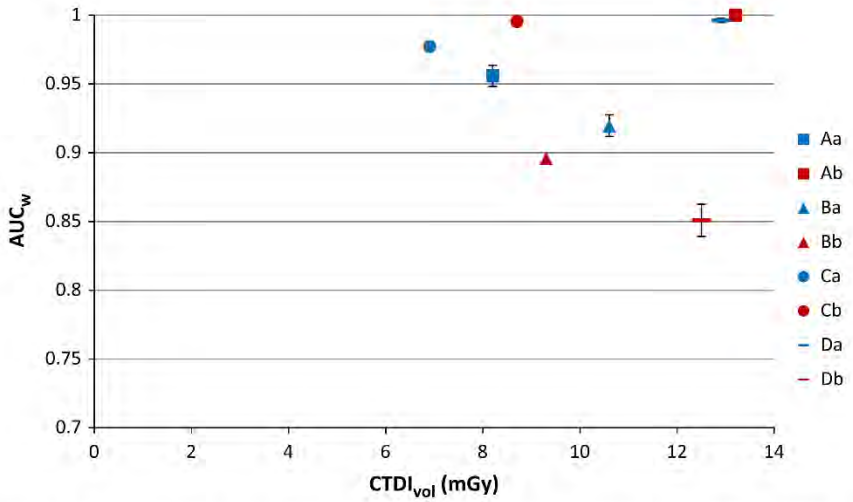


FIG. 7. AUC_w to detect a sphere in soft tissue equivalent using the medium anthropomorphic abdominal phantom at different $CTDI_{vol}$ levels using automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC_w outcome. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC_w = weighted area under the curve, $CTDI_{vol}$ = volume CT dose index.

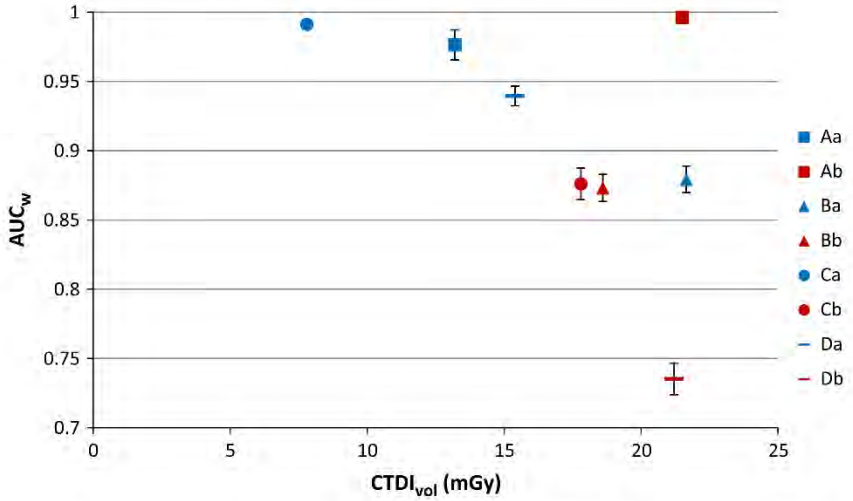


FIG. 8. AUC_w to detect a sphere in liver tissue equivalent using the large phantom at different $CTDI_{vol}$ levels using automatic tube current modulation for the eight different CT scanners. The vertical axis represents the AUC_w outcome. The errors bars represent the 95% confidence intervals. The red color represents the older scanners of each manufacturer, whereas the blue color represents the newer ones. AUC_w = weighted area under the curve, $CTDI_{vol}$ = volume CT dose index.

second limitation deals with the Slice Sensitivity Profile, a parameter that was not considered in our methodology. Nevertheless, comparable reconstruction slice thicknesses (within the range 2.5–3.0 mm) were used, so no major influence of this parameter is expected in our results. Moreover, as the 2D NPS was not isotropic in the phantom, using symmetric channels could impact the performance of the model; but this impact would be minor.

Dealing with the protocol using fixed $CTDI_{vol}$ values, we proposed first to use a range of dose levels. The highest value, 15 mGy, corresponds to our DRL for a "standard" patient of 75 kg. At 15 mGy, no difference between CTs appeared when we used the small abdominal phantom. However, differences were detected at lower dose levels. Thus, such a phantom should be imaged over a lower dose range (e.g., 2–10 mGy), or replaced by a larger version as shown in Figs. 7 and 8 to produce useful results. Finally, with a larger phantom size, the use of a fixed tube current might be a limitation that introduces weaknesses of the image quality which are not relevant for clinical applications where tube current modulation is generally used.

The use of the protocol with tube current modulation on various phantom sizes is certainly more realistic in the framework of patient dose optimization. When comparing the products of one manufacturer, the results were generally straightforward. The major difficulty of such an approach is

to compare various manufacturers who propose different strategies to manage the balance between image quality and patient exposure. In this study we decided to use the local settings. The indication "search for focal liver lesions" requires particularly high image quality. Using the small phantom, a high level of image quality was reached by all CTs using a large range of $CTDI_{vol}$ values. For the larger phantom, a high level of image quality could not be reached by certain units despite the use of a large range of $CTDI_{vol}$ values.

The outcome of this work clearly demonstrates the weaknesses of the DRL concept; indeed, a similar dose level cannot be reached on different scanners without impairing the LCD. To improve this situation, one could associate the DRL to an image quality criterion such as the LCD estimated in a standardized phantom.

5. CONCLUSION

This study shows that MOs can objectively benchmark CT scanners using a task-based image quality method. Such an approach may be useful for quantitatively comparing the clinically relevant image quality among various scanners to aid in the estimation of the potential dose reduction without missing the detection of critical lesions.

ACKNOWLEDGMENTS

This work was supported by the German Radiation Protection Agency (BfS – "UFO-Plan Vorhabens 3613S20007"). The authors thank Prof. Peter Vock and PD Dr Sebastian Schindera, the two other senior radiologists involved in this project for their help when dealing with the choice of the protocols and diagnostic tasks. The authors also thank MSc Pascal Monnin for his advices concerning the estimation of the TTF functions.

CONFLICT OF INTEREST

The authors have no relevant conflicts of interest to disclose.

*D. Racine and A. Viry contributed equally to this work.
[a)]Author to whom correspondence should be addressed. Electronic mail: damien.racine@chuv.ch; Telephone: +41 21 314 92 49.

REFERENCES

1. Brenner DJ, Hall EJ. Computed tomography — An increasing source of radiation exposure. *N Engl J Med*. 2007;357:2277–2284.
2. Schauer DA, Linton OW. National council on radiation protection and measurements report shows substantial medical exposure increase. *Radiology*. 2009;253:293–296.
3. Samara ET, Aroua A, Bochud FO, et al. Exposure of the swiss population by medical x-rays: 2008 review. *Health Phys*. 2012;102:263–270.
4. Coultre RL, Bize J, Champendal M, et al. Exposure of the swiss population by radiodiagnostics: 2013 review. *Radiat Prot Dosimetry*. 2015;169:221–224.
5. Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys Med Biol*. 2015;60:R1.
6. Vennart W. ICRU report 54: medical imaging — The assessment of image quality. *Radiography*. 1996;3:243–244.
7. Racine D, Ott JG, Ba A, Ryckx N, Bochud FO, Verdun FR. Objective task-based assessment of low-contrast detectability in iterative reconstruction. *Radiat Prot Dosimetry*. 2016;169:73–77.
8. Saiprasad G, Filliben J, Peskin A, et al. Evaluation of low-contrast detectability of iterative reconstruction across multiple institutions, CT scanner manufacturers, and radiation exposure levels. *Radiology*. 2015;277:124–133.
9. Vaishnav JY, Jung WC, Popescu LM, Zeng R, Myers KJ. Objective assessment of image quality and dose reduction in CT iterative reconstruction. *Med Phys*. 2014;41:71904.
10. Verdun FR, Racine D, Ott JG, Tapiovaara MJ, et al. Image quality in CT: from physical measurements to model observers. *Phys Medica Eur J Med Phys*. 2015;31:823–843.
11. Schindera ST, Odedra D, Raza SA, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*. 2013;269:511–518.
12. International Electrotechnical Committee. *Medical diagnostic X-ray equipment - Radiation conditions for use in the determination of characteristics*. 1994. Standard IEC # 61267. https://webstore.iec.ch/publication/21937. Published 2015. Accessed September 27, 2016.
13. Office fédéral de la santé publique OFSP. *Notice R-06-06 :Niveaux de référence diagnostiques en tomodensitométrie*. 2010.
14. Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. *Phys Med Biol*. 2014;59:4047–4064.
15. Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with channelized hotelling observer. *Phys Medica Eur J Med Phys*. 2016;32:76–83.
16. Boone JM. Determination of the presampled MTF in computed tomography. *Med Phys*. 2001;28:356–360.
17. Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms. *Med Phys*. 2012;39:4115–4122.
18. International Commission on Radiation Units and Measurements. ICRU report no. 87: radiation dose and image-quality assessment in computed tomography. *J ICRU*. 2012;12:1–149.
19. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci U S A*. 1993;90:9758–9765.
20. Park S, Barrett HH, Clarkson E, Kupinski MA, Myers KJ. Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal. *J Opt Soc Am A Opt Image Sci Vis*. 2007;24:B136–B150.
21. Gallas BD, Barrett HH. Validating the use of channels to estimate the ideal linear observer. *J Opt Soc Am A*. 2003;20:1725–1738.
22. Barrett HH, Myers KJ. *Foundations of Image Science*. Wiley-Interscience; 2004.
23. International Commission on Radiation Units and Measurements. Receiver operating characteristic analysis in medical imaging. In: *ICRU Report n 79*, Vol 79. Bethesda, MD: International Commission on Radiation Units and Measurements; 2008: 79.
24. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: CRC Press; 1994.
25. Solomon JB, Christianson O, Samei E. Quantitative comparison of noise texture across CT scanners from different manufacturers. *Med Phys*. 2012;39:6048–6055.

5.2 Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study.

Published in EJMP (Physica Medica) (2018)[92]

Physica Medica 48 (2018) 111–118

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study

Anais Viry[a,*], Christoph Aberle[b], Damien Racine[a], Jean-François Knebel[c,d], Sebastian T. Schindera[e], Sabine Schmidt[f], Fabio Becce[f,1], Francis R. Verdun[a,1]

[a] Institute of Radiation Physics, Lausanne University Hospital, Rue du Grand-Pré 1, 1007 Lausanne, Switzerland
[b] Department of Radiology, University Hospital Basel, University of Basel, Basel, Switzerland
[c] Laboratory for Investigative Neurophysiology (The LINE), Department of Radiology and Department of Clinical Neurosciences, Lausanne University Hospital, Lausanne, Switzerland
[d] Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM), Lausanne and Geneva, Switzerland
[e] Department of Radiology, Kantonsspital Aarau, Aarau, Switzerland
[f] Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Lausanne, Switzerland

ARTICLE INFO

Keywords:
Computed tomography
Iterative reconstruction algorithms
Low-contrast detectability
Mathematical model observers
Task-based image quality assessment

ABSTRACT

Purpose: To investigate how various generations of iterative reconstruction (IR) algorithms impact low-contrast detectability (LCD) in abdominal computed tomography (CT) for different patient effective diameters, using a quantitative task-based approach.

Methods: Investigations were performed using an anthropomorphic abdominal phantom with two optional additional rings to simulate varying patient effective diameters (25, 30, and 35 cm), and containing multiple spherical targets (5, 6, and 8 mm in diameter) with a 20-HU contrast difference. The phantom was scanned using routine abdominal protocols ($CTDI_{vol}$, 5.9–16 mGy) on four CT systems from two manufacturers. Images were reconstructed using both filtered back-projection (FBP) and various IR algorithms: ASiR 50%, SAFIRE 3 (both statistical IRs), ASiR-V 50%, ADMIRE 3 (both partial model-based IRs), or Veo (full model-based IR). Section thickness/interval was 2/1 mm or 2.5/1.25 mm, except 0.625/0.625 mm for Veo. We assessed LCD using a channelized Hotelling observer with 10 dense differences of Gaussian channels, with the area under the receiver operating characteristic curve (AUC) as a figure of merit.

Results: For the smallest phantom (25-cm diameter) and smallest lesion size (5-mm diameter), AUC for FBP and the various IR algorithms did not significantly differ for any of the tested CT systems. For the largest phantom (35-cm diameter), Veo yielded the highest AUC improvement (8.5%). Statistical and partial model-based IR algorithms did not significantly improve LCD.

Conclusion: In abdominal CT, switching from FBP to IR algorithms offers limited possibilities for achieving significant dose reductions while ensuring a constant objective LCD.

1. Introduction

Over the last twenty years, computed tomography (CT) examinations have been performed with increasing frequency in Western countries, leading to improved patient care. This has, however, substantially increased the impact of CT on the collective effective radiation dose [1].

Automatic tube current modulation (ATCM) has first been proposed to reduce radiation exposure by accounting for the patient's X-ray attenuation. This is one of the most effective tools for optimizing patient exposure, particularly for thoracic and abdominal CT examinations. One of the ATCM strategies, developed by Siemens, is to automatically adapt the tube current level depending on the patient body habitus. Another strategy, used by GE, is to manually set a noise index (NI) level [2,3]. In all cases, radiologists must tolerate higher noise levels in larger patients because of the exponential nature of X-ray attenuation [4,5]. To offer further dose reductions, CT manufacturers have subsequently replaced the traditional filtered back-projection

* Corresponding author.
E-mail address: anais.viry@chuv.ch (A. Viry).
[1] F. Becce and F.R. Verdun contributed equally to this work.

https://doi.org/10.1016/j.ejmp.2018.04.006
Received 28 December 2017; Received in revised form 5 April 2018; Accepted 7 April 2018

(FBP) reconstruction algorithm with iterative reconstruction (IR) algorithms, which can be distinguished into two generations: statistical or hybrid IR, and model-based IR [6]. Model-based IR algorithms include the modeling of physical and optical aspects in the reconstruction process. The modeling complexity impacts the image reconstruction time, such that manufacturers can propose either a fast partial version or a longer full version of their model-based IR algorithm.

When aiming at reducing radiation dose, the optimization process of CT protocols should ensure that the clinical question can still always be answered, while keeping the radiation exposure as low as reasonably achievable [7]. Among all clinical data contained in patient images, one of the most challenging task is the detection of low-contrast lesions, such as in non-contrast CT scans of the abdomen, particularly in large patients. Hence, it appears necessary to assess the low-contrast detectability (LCD) performances of various IR algorithms with the routine clinical CT settings for different patient body habitus before implementing drastic dose reductions with IR [8].

To assess the impact of dose reductions on LCD, one robust and objective approach involves the use of anthropomorphic CT phantoms with simulated low-contrast lesions. Previous studies have assessed the potential improvement of LCD with IR algorithms for one single phantom size, and often applied subjective and/or objective image quality parameters based on the technical efficacy (image noise, contrast-to-noise ratio (CNR), and spatial resolution) [9–11]. However, since it is difficult to correlate these image quality metrics with an accurate assessment of diagnostic performance, it is of paramount importance to use image quality criteria linked to a specific and clinically relevant task, such as LCD, for different phantom sizes [12,13].

In this study, we used a task-based quantitative image quality assessment approach to investigate how the LCD for various patient effective diameters was impacted by the use of various generations of IR algorithms from two CT manufacturers, using routine clinical settings.

2. Materials and methods

2.1. Phantom

A commercially-available tissue equivalent anthropomorphic abdominal phantom (QRM, Moehrendorf, Germany) was used to mimic data obtained from non-contrast CT scans of patients. The phantom contains a 10-cm-diameter central hole into which modules can be inserted to assess image quality. We used a cylindrical module containing low-contrast spheres that mimic hypodense abdominal lesions, and a homogeneous module comprising the same tissue equivalent material as the phantom. The module with hypodense abdominal lesions contained 24 spheres of 8, 6, 5, 4, and 3 mm in diameter, with a 20-HU contrast difference compared to the background material in a single axial section. This contrast difference was obtained using plastic materials of low effective Z numbers and different mass densities. The effective diameter of this phantom is 25 cm, thus simulating the X-ray attenuation of a thin patient weighing ~50 kg (Fig. 1a). To simulate different patient body habitus, additional rings of soft tissue equivalent were added around the core of the phantom. Using 2.5- and 5-cm-thick rings, respectively, we achieved effective diameters of 30 and 35 cm, thus simulating patients weighing ~75 (Fig. 1b) and ~100 kg (Fig. 1c), respectively. The correlation between the phantom effective diameter and patient weight was previously obtained by performing a linear regression between the effective abdominal diameter and patient weight for 500 patients.

2.2. CT protocols

We scanned the phantom both with and without its additional rings using four different CT systems: the SOMATOM Definition Flash and SOMATOM Definition Edge (Siemens Healthineers, Forchheim, Germany), and the Discovery CT750 HD and Revolution (GE Healthcare, Milwaukee, WI, USA). Before data acquisition, we selected on each CT system the routine protocols used for the following clinical task: detection of low-contrast liver lesions. The portovenous acquisition phase was considered. The ATCM settings were those used in routine clinical practice: the quality reference mAs (with the "average" curve strength) were 150 and 130 mAs for the Flash and Edge Siemens systems, respectively; while for both GE systems, NI values were set to 20, 26, and 32 HU, for the small, medium, and large phantoms, respectively. The $CTDI_{vol}$ were retrieved directly from the CTDI dose report. With these settings, the $CTDI_{vol}$ for all CT systems varied between 5.9 and 7.8 mGy for the small phantom, 8.5 and 11.6 mGy for the medium phantom, and 11.1 and 16 mGy for the large phantom. To further simulate clinical practice, the display FOV was adapted for each phantom size, thus yielding different pixel sizes. Data acquisition and image reconstruction parameters are summarized in Table 1:

These settings produced a total of 27 datasets, including three phantom sizes and two image reconstruction algorithms for three of the four tested CT systems, and three phantom sizes and three image reconstruction algorithms for one CT scanner.

To achieve statistical robustness with the utilized image quality metrics, each phantom with the module containing low-contrast spheres was scanned 15 times without moving the phantom between acquisitions. Then, the low-contrast module was replaced with its homogeneous counterpart, and the phantom was scanned three times using the exact same settings, to provide the required number of images.

2.3. Task-based image quality assessment

We used a task-based approach to quantitatively assess the image quality of routine abdominal CT protocols, noting that the detection of low-contrast lesions is the most challenging task for such protocols. LCD was assessed for each image reconstruction algorithm and phantom size for three lesion sizes that were considered as clinically relevant for the investigated protocols (8, 6, and 5 mm in diameter). In clinical practice, smaller focal liver lesions (< 5 mm in diameter) may not be reliably characterized with the routine CT settings due to the poor low-contrast resolution of CT [14].

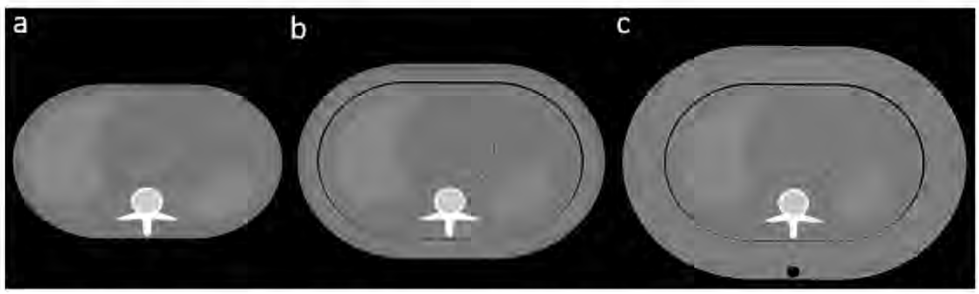


Fig. 1. Axial CT images of the anthropomorphic abdominal phantom simulating patients with weights of 50 (a), 75 (b), and 100 kg (c), respectively.

Table 1
Summary of the utilized CT data acquisition and image reconstruction parameters.

	Siemens Flash	Siemens Edge	GE Discovery CT750 HD	GE Revolution
Data acquisition parameters				
Tube potential (kVp)	120	120	120	120
Gantry revolution time (s)	0.5	0.5	0.6	0.5
Beam collimation (mm)	38.4	38.4	40	40
Pitch	0.6	0.8	0.986	0.975
ATCM settings[1]	Reference mAs = 150 mAs	Reference mAs = 130 mAs	NI = 20/26/32 HU	NI = 20/26/32 HU
Tube charge (mAs)[2]	100/142/215[3]	88/125/182[3]	110/164/225	96/124/160
$CTDI_{vol}$ (mGy)[2]	6.8/9.6/14.6	5.9/8.5/12.2	7.8/11.6/16	6.7/8.6/11.1
Image reconstruction parameters				
Section thickness/interval	2/1 mm	2/1 mm	2.5/1.25 mm 0.625/0.625 mm[4]	2.5/1.25 mm
Pixel size (mm)[2]	0.625/0.723/0.820	0.625/0.723/0.820	0.625/0.723/0.820	0.625/0.723/0.820
Display FOV (mm)[2]	320/370/420	320/370/420	320/370/420	320/370/420
Kernel	B30f (FBP)/I30f (IR)	B30f (FBP)/I30f (IR)	Standard	Standard
Image reconstruction algorithms				
FBP	FBP	FBP	FBP	n.a.[5]
Statistical IR	SAFIRE 3	n.a.	ASiR 50%	n.a.
Partial model-based IR	n.a.	ADMIRE 3	n.a.	ASiR-V 50%
Full model-based IR	n.a.	n.a.	Veo	n.a.

[1] ATCM settings correspond to quality reference mAs for Siemens systems and noise index (NI) values for GE systems. NI values correspond to the small, medium, and large phantom, respectively.
[2] Tube charge, $CTDI_{vol}$, pixel size, and display FOV all varied with the phantom size. Values correspond to the small, medium, and large phantom, respectively.
[3] The mAs values for the two Siemens systems are effective mAs (mAs/pitch).
[4] For the full model-based IR (Veo), 0.625/0.625 mm was the only section thickness/interval reconstruction available at the CT console at the time of the study.
[5] Not available, ASiR-V at a level of 0% was used instead.

From each dataset, because of the different display FOV used depending on the phantom diameter, we extracted square regions of interest (ROIs) containing the low-contrast lesion, which measured 22 × 22, 18 × 18, and 16 × 16 pixels for the small, medium, and large phantoms, respectively. We extracted a total of 60 ROIs per sphere diameter and for the various acquisition/reconstruction conditions. We additionally extracted 300 ROIs of noise only from images of the homogeneous module for each acquisition/reconstruction condition. Due to the non-stationarity of IR algorithms, we selected ROIs of noise only at the same position than signal-containing ROIs.

To provide an objective quantitative metric, we used an anthropomorphic mathematical model observer designed to assess the performance of a classification task (signal present/signal absent). This method is based on a statistical decision theory that aims to mimic the performance of human observers in the detection of low-contrast structures within an image [15–17]. We chose to use the channelized Hotelling observer (CHO) with 10 dense differences of Gaussian channels as a mathematical model observer [18]. The radial frequency of the jth channel is given by .Eq. (1):

$$U_j(\rho) = \exp\left[-\frac{1}{2}\left(\frac{\rho}{W\sigma_j}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{\rho}{\sigma_j}\right)^2\right] \quad (1)$$

In this equation, ρ is the spatial frequency, the multiplicative factor W is the channel bandwidth, and σ_j is the standard deviation of the jth channel with σ_j defined from a starting value σ_0 by $\sigma_j = \sigma_0\alpha^j$. In this study, W = 1.67, σ_0 = 0.005, and α = 1.4 [19].

Each vectorized ROI g, containing either a signal or noise only, was channelized using Eq. (2):

$$\mathbf{v} = \mathbf{U}^T\mathbf{g} \quad (2)$$

where U is a matrix with columns corresponding to channel vectors.

As a human observer, the mathematical observer attributes a grade (also called decision variable λ) to each channelized image. Eq. (3) was used to compute λ:

$$\lambda = \mathbf{w}^T\mathbf{v} \quad (3)$$

And the template w was computed using Eq. (4):

$$\mathbf{w} = \mathbf{K}_\mathbf{v}^{-1}(\langle \mathbf{v_s} \rangle - \langle \mathbf{v_n} \rangle) \quad (4)$$

In this equation, K_v is the covariance matrix of the 300 channelized signal-absent ROIs, $\langle v_s \rangle$ is the mean of the 80 channelized ROIs with signal, and $\langle v_n \rangle$ is the mean of the 300 channelized ROIs with noise only. Using this method, the model observer was trained to calculate the template, and was tested for calculating the decision variables using the same images, which could overestimate the performances (positive bias). The number of signal-absent ROIs and signal-present ROIs were chosen to avoid singularity issues when dealing with the estimation of the covariance matrix inversion and to minimize bias due to the use of a limited number of images [20].

Higher decision variables corresponded to higher confidence of the model observer regarding the presence of a low-contrast lesion in the image. All the decision variables were combined in a histogram, as illustrated as an example in Fig. 2a. Varying a threshold of the decision variable enabled the plotting of a receiver operating characteristic (ROC) curve (Fig. 2b), which was the complete outcome of the model observer. The ROC curve was plotted using 200 thresholds. Next, the image quality was assessed by computing the area under the ROC curve (AUC) using a trapezoidal method.

2.4. Statistical analysis

To each dataset, we applied a bootstrapping technique (random selection with replacement) to estimate the reference range of the results of the model observer [21]. We performed a total of 1000 different bootstrapped samplings using 60 signal images and 300 noise-only images. For each randomization, we calculated the ROC curves and the corresponding AUC values. To assess the statistical differences between the two distributions corresponding to FBP and IR outcomes, we calculated the percentage (PC) of IR outcomes superior to FBP, without making any assumption about the underlying distribution of the data. The improvement of LCD for the IR algorithm compared to FBP on the same CT system was considered statistically significant if PC was > 95% (0.95). This level was set arbitrarily, yet in analogy with the significance levels commonly used for p-values, to be even more specific in terms of discrimination, considering that AUCs > 0.90 are

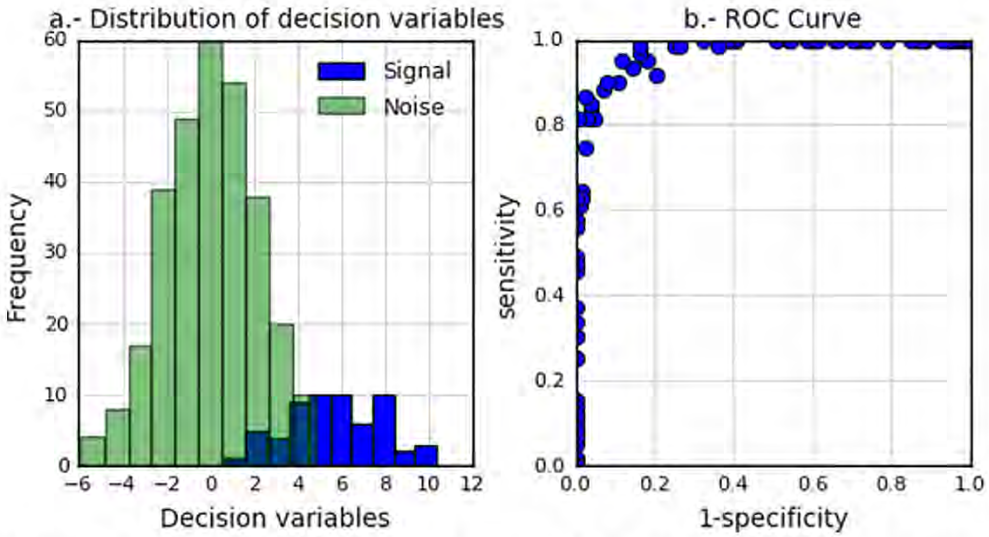


Fig. 2. a) Example of the distribution of decision variables obtained using the CHO model observer for the two categories of images: images with signal (blue bars) and images with noise only (green bars). b) Receiver operating characteristic (ROC) curve showing the outcome of decision variable threshold variation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rated as excellent [22]. In such cases, we calculated the potential mean improvement.

3. Results

3.1. Low-contrast detectability using FBP

Fig. 3 presents the AUC for each sphere size (5, 6, and 8 mm) obtained using the standard FBP algorithm (except ASiR-V 0% for the Revolution system) as a function of the phantom size. The vertical error bars on each data point represent the 95% confidence intervals. As expected based on the NI settings or ATCM behavior, the AUC decreased with increasing phantom size at each sphere diameter. Moreover, this decreasing performance was systematically higher for spheres of smaller diameter, i.e., with increasing task difficulty. For a sphere diameter of 5 mm, AUC decreased by 18%, 20%, 25%, and 19% for the Flash, Edge, Discovery CT750 HD, and Revolution systems, respectively.

3.2. Impact of the various IR algorithms on LCD performances as a function of phantom size

Fig. 4 presents a pairwise comparison of the results of the CHO model observer for detection of the smallest sphere (5 mm in diameter) using the conventional FBP and available IR algorithms for the three phantom sizes. Table 2 summarizes the PC of IR outcomes that were superior to FBP. For the small phantom, the AUC values were nearly 1 for each CT system. In this case, the detection task was very easy, and LCD was not significantly improved with any of the tested IR algorithms relative to FBP. For the medium phantom, no significant LCD improvement was observed with ASiR 50%, SAFIRE 3, ASiR-V 50%, or ADMIRE 3 compared to FBP. However, with the Discovery CT750 HD system, Veo was associated with a significant AUC improvement of 5.8% compared to FBP. Similarly, with the large phantom, only Veo yielded a significant AUC improvement of 8.5% relative to FBP. When quantitatively comparing all CT systems using a standard FBP algorithm and for each phantom size, we found that the AUC values were nearly equal (differences of ± 0.05), indicating comparable objective image quality despite the dose variations.

3.3. Impact of the various IR algorithms on LCD performances as a function of lesion size

Fig. 5 presents the results of the CHO model observer as a function of sphere size with the various IR algorithms. These results were only obtained for the large phantom, as this was the most discriminating case. As expected, the AUC values systematically increased with increasing sphere diameter. Table 3 presents a pairwise comparison of the IR and FBP algorithms for each tested CT system and each sphere diameter. With the smallest sphere diameter (5 mm), LCD was not significantly improved with the tested IR algorithms compared to FBP, except with Veo. With the 6-mm sphere, slight LCD improvements were noted for ADMIRE 3 (2% mean improvement) and ASiR 50% (2.2% mean improvement). With the 8-mm sphere, we found that LCD was slightly improved for SAFIRE 3 (0.9%), ADMIRE 3 (1.1%), and ASiR-V 50% (1.6%). Surprisingly, ASiR 50% showed no benefit with the 8-mm spheres. The greatest benefit was observed with the full model-based IR algorithm Veo, which was available on the Discovery CT750 HD system, with LCD increased by 8.5% with the 5-mm spheres and by 5.2% for the 6-mm spheres, despite the fourfold thinner slices. Notably, the Veo algorithm achieved a higher relative LCD gain with the more difficult tasks (i.e., smaller sphere diameter).

4. Discussion

In our study, we used a task-based quantitative image quality assessment approach to investigate how various generations of IR algorithms impacted LCD with different simulated patient effective abdominal diameters. We found that with each tested CT system, LCD decreased with increasing phantom size. This was expected since, in the dose optimization process proposed by Siemens, a reduction of LCD in larger patients is considered acceptable to avoid high-dose exposure. For GE systems, technical charts have been proposed to adapt NI values to the individual patient body habitus to avoid non-diagnostic image quality for thinner patients and overexposure for larger patients [5]. Moreover, the technical limitations of X-ray tube power could also limit the feasibility of equivalent noise levels for all patient morphotypes. Using these methodologies, it appears that although dose increased when the phantom's diameter increased, the LCD did not remain constant with all phantom sizes. This implies that diagnostic requirements

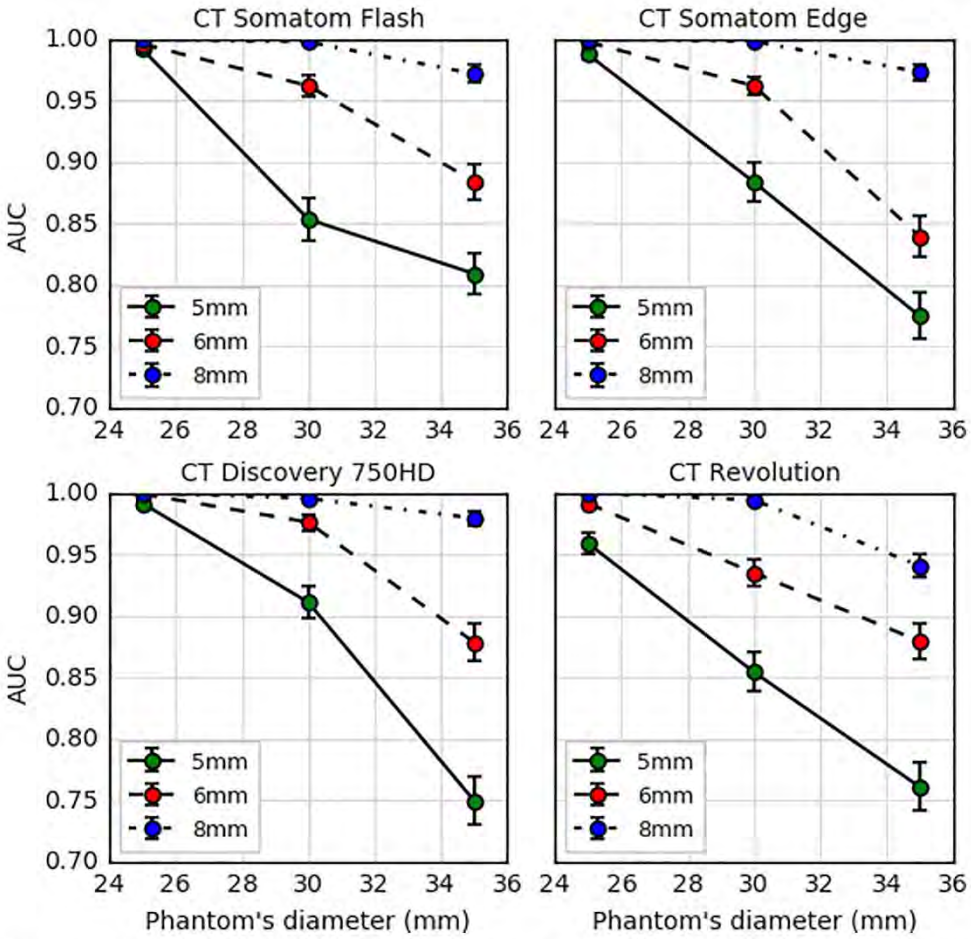


Fig. 3. Low-contrast detectability assessed by means of the area under the receiver operating characteristic curve (AUC) for each computed tomography (CT) system for the three sphere diameters as a function of the phantom diameter for filtered back-projection images.

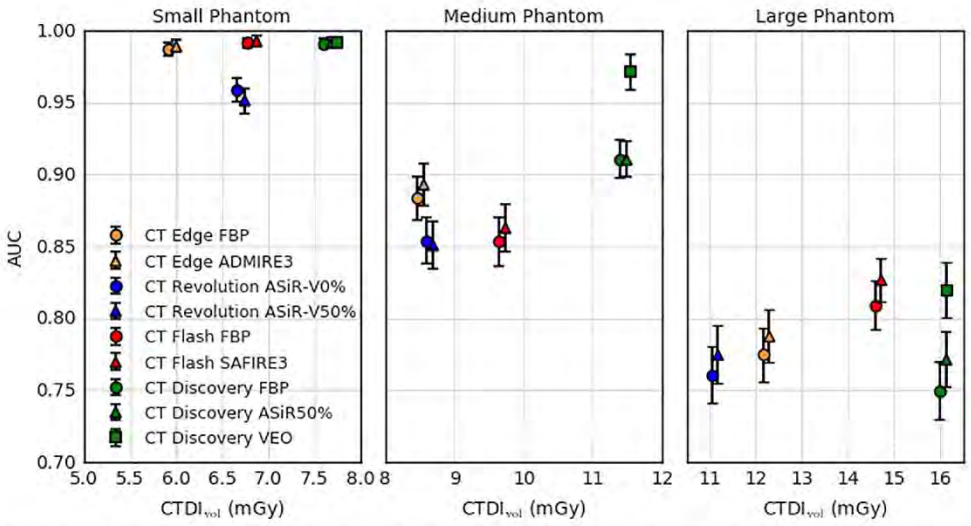


Fig. 4. Impact of the image reconstruction algorithms on low-contrast detectability for 5-mm-diameter spheres and the three phantom sizes. Note that each CT system is reported as a function of the $CTDI_{vol}$ routinely applied in clinical practice.

Table 2
Percentage of iterative reconstruction (IR) outcomes that were superior to filtered back-projection (FBP) for the three phantom sizes for the smallest spheres (5-mm diameter).

Phantom size	Small	Medium	Large
CT Edge (FBP/ADMIRE 3)	76.7%	81.3%	85.0%
CT Flash (FBP/SAFIRE 3)	74.1%	78.4%	93.9%
CT Revolution (ASiR-V 0%/ASiR-V 50%)	9.9%	40.1%	84.6%
CT Discovery CT750 HD (FBP/ASiR 50%)	65.8%	50.1%	94.3%
CT Discovery CT750 HD (FBP/Veo)	64%	**100%***	**100%***

* Statistically significant improvement between FBP and IR algorithm.

Table 3
Percentage of iterative reconstruction (IR) outcomes that were superior to filtered back-projection (FBP) for the three lesion sizes for the large phantom.

Lesion size	5 mm	6 mm	8 mm
CT Edge (FBP/ADMIRE 3)	85.0%	**95.8%***	**99.7%***
CT Flash (FBP/SAFIRE 3)	93.9%	92.5%	**97.2%***
CT Revolution (ASiR-V 0%/ASiR-V 50%)	84.6%	94.3%	**99.4%***
CT Discovery CT750 HD (FBP/ASiR 50%)	94.3%	**99.4%***	85.3%
CT Discovery CT750 HD (FBP/Veo)	**100%***	**100%***	**96.4%***

* Statistically significant improvement between FBP and IR algorithm.

should be defined individually based on both the patient-specific body habitus and clinical task. Thus, an optimization process should account for the patient's body habitus and use an objective image quality metric linked to the clinical task.

Our results further demonstrated that the first and second generations of IR techniques (statistical IR and partial model-based IR, respectively) had no substantial impact on LCD for each phantom size. For the small phantom, the examined task was trivial, and thus no difference was observed between the various image reconstruction algorithms. Notably, the dose levels used enabled a high level of performance (Fig. 4). The use of a lower dose range may have enabled the detection of differences between reconstruction algorithms. For the medium and large phantoms, we still observed no substantial improvement for 5-mm diameter lesions with ASiR 50%, SAFIRE 3, ASiR-V 50%, or ADMIRE 3 compared to FBP. However, the effectiveness of IR algorithms increased with increasing phantom size. Only the full model-based IR Veo, which includes the physical and optical aspects of the CT system in the image reconstruction process, had a significant impact on LCD for the medium and large phantoms. Since Veo further provides a reduced section thickness in comparison with the other algorithms, we secondarily investigated the effect of the reduced section thickness on LCD by averaging thin Veo slices. For the smallest lesion size, no significant impact was noted for each phantom size. As a matter of fact, partial volume effect that may impair LCD should not be significant when dealing with the detection of a 5 mm sphere using a section thickness of 2.5 mm with a reconstruction interval of 1.25 mm. Nevertheless, this reduced section thickness may further improve the longitudinal spatial resolution and the LCD in the other two

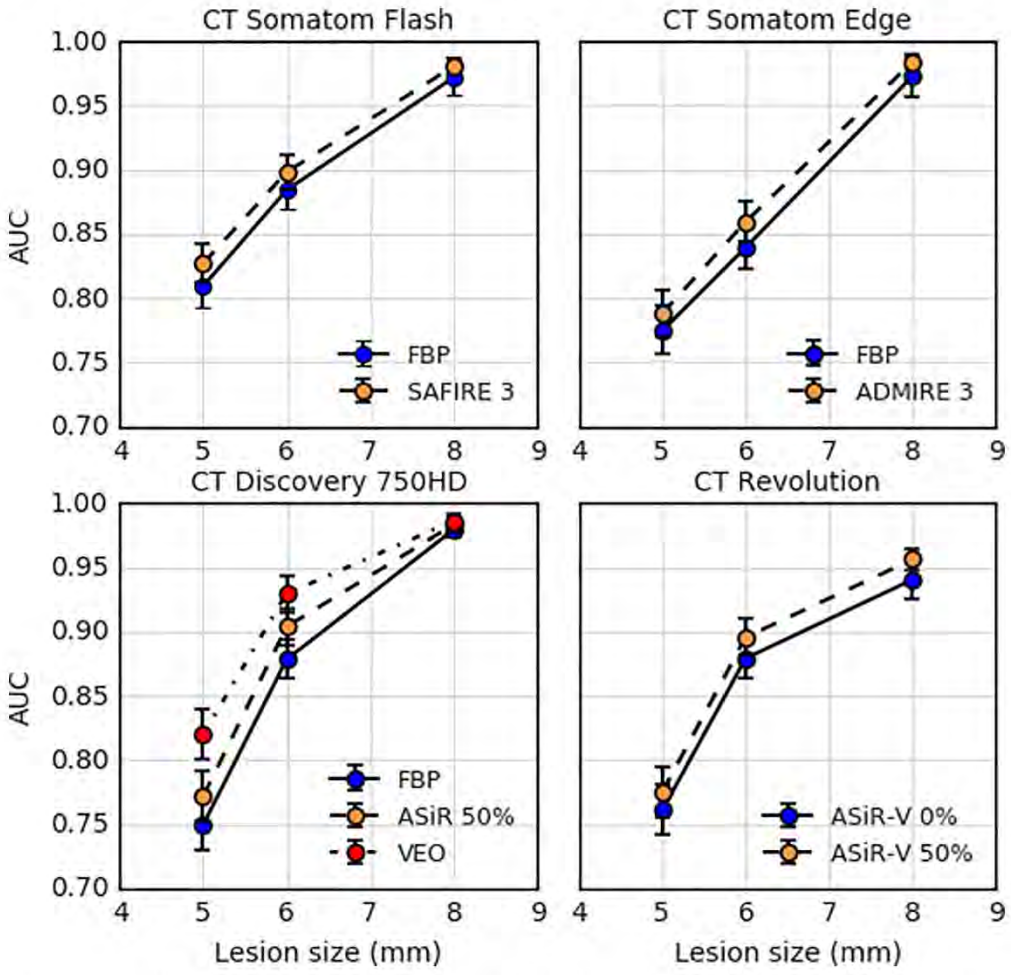


Fig. 5. Low-contrast detectability as a function of lesion size with various image reconstruction algorithms for the large phantom.

reconstruction planes (i.e. coronal and sagittal) when using multiplanar reformations. However, this point was beyond the scope of the present investigation. The main limitation of Veo is that the reconstruction time takes up to 45 min for a single acquisition phase. In summary, our present results clearly demonstrated that, for the selected clinical task, the improvement in LCD by using statistical IRs or partial model-based IRs rather than FBP is very limited. Hence, the potential to achieve substantial dose reduction while maintaining LCD appears to be limited. The use of full model-based IR rather than FBP had greater potential to improve LCD, particularly for large body sizes.

Other investigators have also used CT phantoms to assess the potential improvement of LCD with various IR algorithms. Subjective analyses with human readers (ROC, 4AFC experiment) have indicated that IR techniques (SAFIRE, ASiR, and AIDR) do not statistically improve LCD [9,23–25] or only slightly improve LCD at the same dose level with various IR algorithms [26–28]. These researchers have also applied quantitative image quality parameters (noise and CNR) to validate their use. They report that image noise is significantly reduced with IR algorithms, leading to major CNR improvements. On the other hand, the CNR is not correlated with LCD [29], and should no longer be used as a surrogate of LCD when the noise texture changes, which typically occurs with IR algorithms. This is the case when comparing different CT systems, or when comparing various image reconstruction algorithms on the same CT scanner. Our results confirmed the findings of previous human subjective analyses of statistical IR (ASiR and SAFIRE) [26,27] and partial model-based IR (ASiR-V and ADMIRE) [25,28] using an objective methodology based on a task-specific concept.

Other studies have also applied task-based image quality metrics to LCD assessment using Fourier transform metrics (noise power spectrum and modulation transfer function) to compute a detectability index [30–32] or by directly computing mathematical model observers [16,28,33]. To compute Fourier transform metrics, imaging systems must be linear and shift invariant [34]. However, this is no longer the case with IR algorithms; thus, this method must be applied with caution in such contexts. In our study, we applied a mathematical model observer directly in the image space domain. Our results confirmed the findings of a previous study using Fourier transform metrics for ASiR and Veo [31]. ASiR did not improve LCD as compared to FBP, whereas Veo allowed better improvement compared to both FBP and ASiR.

Our study has several limitations. First, the anthropomorphic abdominal phantom comprised soft-tissue-equivalent rings that do not accurately simulate large patients with fat around their organs in addition to subcutaneous fat. Rather, the large phantom simulates a patient with a higher X-ray attenuation, and we presume that LCD improvement with IR algorithms would be identical or less significant using a phantom with fat-equivalent rings. Second, the technological differences between CT systems and the different ATCM settings, which are based on different principles between the two manufacturers, led to the clinical use of various $CTDI_{vol}$. This was not a problem since the goal of our study was not to benchmark CT systems but rather to investigate the impact of IR over FBP using routine clinical protocols and settings. Despite these differences, we noted that noise levels and LCD of the FBP acquisitions for a common and clinically relevant task were comparable for both manufacturers. Third, we did not investigate the impact of the strength levels of the various IR algorithms on LCD; we only selected one of the most commonly used strength levels proposed by each manufacturer for abdominal CT protocols. Finally, the chosen task (detection of a low-contrast structure at a known position in a homogeneous background) is very simple compared to the clinical reality. However, even in this simple condition, the benefit of IR appears marginal and we should not expect major improvements when dealing with textured backgrounds [35]. To go a step further, we may, in the future, add complexity to the task, and assess the performance of IR algorithms to correctly locate a signal or estimate the size and shape of a lesion [36,37].

5. Conclusions

In conclusion, in routinely used abdominal CT protocols, switching from FBP to IR algorithms does not significantly improve LCD. Dose optimization should be performed by first defining the diagnostic requirement for a relevant clinical task and the patient's specific body habitus, and by subsequently evaluating image quality criteria linked to this task before drastically reducing dose levels.

Source of funding

This work was partly supported by a grant from the Swiss National Science Foundation (SNSF) No. 320030-163344.

Conflict of interest

The authors have no relevant conflict of interest to disclose.

References

[1] Schauer DA, Linton OW. National council on radiation protection and measurements report shows substantial medical exposure increase. Radiology 2009;253:293–6. http://dx.doi.org/10.1148/radiol.2532090494.
[2] Lee CH, Goo JM, Ye HJ, Ye S-J, Park CM, Chun EJ, et al. Radiation dose modulation techniques in the multidetector CT era: from basics to practice. RadioGraphics 2008;28:1451–9. http://dx.doi.org/10.1148/rg.285075075.
[3] Söderberg M, Gunnarsson M. Automatic exposure control in computed tomography – an evaluation of systems from different manufacturers. Acta Radiol 2010;51:625–34. http://dx.doi.org/10.3109/02841851003698206.
[4] McCollough CH. Automatic exposure control in ct: are we done yet? Radiology 2005;237:755–6. http://dx.doi.org/10.1148/radiol.2373051151.
[5] Favazza CP, Yu L, Leng S, Kofler JM, McCollough CH. Automatic exposure control systems designed to maintain constant image noise: effects on computed tomography dose and noise relative to clinically accepted technique charts. J Comput Assist Tomogr 2015;1. http://dx.doi.org/10.1097/RCT.0000000000000221.
[6] Patino M, Fuentes JM, Singh S, Hahn PF, Sahani DV. Iterative Reconstruction Techniques in Abdominopelvic CT: technical Concepts and Clinical Implementation. Am J Roentgenol 2015;205:W19–31. http://dx.doi.org/10.2214/AJR.14.13402.
[7] Trattner S, Pearson GDN, Chin C, Cody DD, Gupta R, Hess CP, et al. Standardization and Optimization of CT Protocols to Achieve Low Dose. J Am Coll Radiol 2014;11:271–8. http://dx.doi.org/10.1016/j.jacr.2013.10.016.
[8] Verdun FR, Racine D, Ott JG, Tapiovaara MJ, Toroi P, Bochud FO, Veldkamp WJH, et al. Image quality in CT: from physical measurements to model observers. Phys Medica Eur J Med Phys 2015;31:823–43. http://dx.doi.org/10.1016/j.ejmp.2015.08.007.
[9] Goenka AH, Herts BR, Dong F, Obuchowski NA, Primak AN, Karim W, et al. Image noise, CNR, and detectability of low-contrast, low-attenuation liver lesions in a phantom: effects of radiation exposure, phantom size, integrated circuit detector, and iterative reconstruction. Radiology 2016;280:475–82.
[10] Schindera ST, Odedra D, Raza SA, Kim TK, Jang H-J, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269:511–8.
[11] Greffier J, Pereira F, Macri F, Beregi J-P, Larbi A. CT dose reduction using Automatic Exposure Control and iterative reconstruction: a chest paediatric phantoms study. Phys Med 2016;32:582–9. http://dx.doi.org/10.1016/j.ejmp.2016.03.007.
[12] Racine D, Viry A, Becce F, Schmidt S, Ba A, Bochud FO, et al. Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners. Med Phys 2017;44:e153–63. http://dx.doi.org/10.1002/mp.12224.
[13] Viry A, Racine D, Ba A, Becce F, Bochud FO, Verdun FR. Characterization of a CT unit for the detection of low contrast structures. In: Kupinski MA, Nishikawa RM, editors. 2017. p. 101361C. https://doi.org/10.1117/12.2250529.
[14] Marrero JA, Rajender Reddy K. ACG clinical guideline: the diagnosis and management of focal liver lesions. Am J Gastroenterol 2014;109:1328–47. http://dx.doi.org/10.1038/ajg.2014.213.
[15] Zhang Y, Leng S, Yu L, Carter RE, McCollough CH. Correlation between human and model observer performance for discrimination task in CT. Phys Med Biol 2014;59:3389–404. http://dx.doi.org/10.1088/0031-9155/59/13/3389.
[16] Ott JG, Ba A, Racine D, Viry A, Bochud FO, Verdun FR. Assessment of low contrast detection in CT using model observers: developing a clinically-relevant tool for characterising adaptive statistical and model-based iterative reconstruction. Z Für Med Phys 2016. http://dx.doi.org/10.1016/j.zemedi.2016.04.002.
[17] Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms. Med Phys 2013;40:041908.
[18] Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci U S A 1993;90:9758–65.
[19] Abbey CK. Human-and model-observer performance in ramp-spectrum noise:

effects of regularization and object variability. JOSA A 2001;18:473–88.
[20] Wunderlich A, Noo F, Gallas BD, Heilbrun ME. Exact confidence intervals for channelized Hotelling observer performance in image quality studies. IEEE Trans Med Imaging 2015;34:453–64. http://dx.doi.org/10.1109/TMI.2014.2360496.
[21] Gagne RM, Gallas BD, Myers KJ. Toward objective and quantitative evaluation of imaging systems using images of phantoms. Med Phys 2006;33:83. http://dx.doi.org/10.1118/1.2140117.
[22] Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. J Pediatr Psychol 2014;39:204–21. http://dx.doi.org/10.1093/jpepsy/jst062.
[23] Miéville FA, Gudinchet F, Brunelle F, Bochud FO, Verdun FR. Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments – A phantom approach. Phys Med 2013;29:99–110. http://dx.doi.org/10.1016/j.ejmp.2011.12.004.
[24] Schindera ST, Odedra D, Raza SA, Kim TK, Jang H-J, Szucs-Farkas Z, et al. Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? Radiology 2013;269:511–8. http://dx.doi.org/10.1148/radiol.13122349.
[25] Euler A, Stieltjes B, Szucs-Farkas Z, Eichenberger R, Reisinger C, Hirschmann A, et al. Impact of model-based iterative reconstruction on low-contrast lesion detection and image quality in abdominal CT: a 12-reader-based comparative phantom study with filtered back projection at different tube voltages. Eur Radiol 2017;27:5252–9. http://dx.doi.org/10.1007/s00330-017-4825-9.
[26] Goenka AH, Herts BR, Obuchowski NA, Primak AN, Dong F, Karim W, et al. Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: an 18-reader study. Radiology 2014;272:154–63. http://dx.doi.org/10.1148/radiol.14131928.
[27] McCollough CH, Yu L, Kofler JM, Leng S, Zhang Y, Li Z, et al. Degradation of CT Low-Contrast Spatial Resolution Due to the Use of Iterative Reconstruction and Reduced Dose Levels. Radiology 2015;276:499–506. http://dx.doi.org/10.1148/radiol.15142047.
[28] Racine D, Ott JG, Ba A, Ryckx N, Bochud FO, Verdun FR. Objective task-based assessment of low-contrast detectability in iterative reconstruction. Radiat Prot Dosimetry 2016;169:73–7. http://dx.doi.org/10.1093/rpd/ncw020.
[29] Urikura A, Hara T, Ichikawa K, Nishimaru E, Hoshino T, Yoshida T, et al. Objective assessment of low-contrast computed tomography images with iterative reconstruction. Phys Med 2016;32:992–8. http://dx.doi.org/10.1016/j.ejmp.2016.07.003.
[30] Richard S, Husarik DB, Yadava G, Murphy SN, Samei E. Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms: towards task-based assessment of CT performance. Med Phys 2012;39:4115–22. http://dx.doi.org/10.1118/1.4725171.
[31] Samei E, Richard S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology: CT task-based performance metrology. Med Phys 2014;42:314–23. http://dx.doi.org/10.1118/1.4903899.
[32] Ott JG, Becce F, Monnin P, Schmidt S, Bochud FO, Verdun FR. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. Phys Med Biol 2014;59:4047–64. http://dx.doi.org/10.1088/0031-9155/59/4/4047.
[33] Racine D, Ba AH, Ott JG, Bochud FO, Verdun FR. Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer. Phys Medica Eur J Med Phys 2016;32:76–83. http://dx.doi.org/10.1016/j.ejmp.2015.09.011.
[34] Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. Phys Med Biol 2015;60:R1. http://dx.doi.org/10.1088/0031-9155/60/2/R1.
[35] Solomon J, Ba A, Bochud F, Samei E. Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms: voxel-based 3D printed textured phantoms. Med Phys 2016;43:6497–506. http://dx.doi.org/10.1118/1.4967478.
[36] Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E. Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 2016;279:185–94. http://dx.doi.org/10.1148/radiol.2015150892.
[37] Takata T, Ichikawa K, Mitsui W, Hayashi H, Minehiro K, Sakuta K, et al. Object shape dependency of in-plane resolution for iterative reconstruction of computed tomography. Phys Med 2017;33:146–51. http://dx.doi.org/10.1016/j.ejmp.2017.01.001.

5.3 In vitro diagnostic performances of a multi-energy spectral photon-counting CT for crystal-related arthropathies

To be submitted in Radiology

TITLE: In vitro diagnostic performances of a multi-energy spectral photon-counting CT for crystal-related arthropathies

ARTICLE TYPE: Technical development

AUTHORS: [a]Anais Viry, [b]Aamir Y. Raja, [c]Tracy E. Kirkbride, [b]Chloe Choi, [d]Lisa K. Stamp, [e]Nicola Dalbeth, [f]Christele Combes, [a]Francis R. Verdun, [b]Nigel G. Anderson, [g]Fabio Becce

AUTHOR AFFILIATIONS -[a]Institute of Radiation Physics, Lausanne Univ. Hospital (Switzerland), [b]Department of Radiology, Univ. of Otago, Christchurch (New Zealand),[c]Ara Institute of Canterbury (New Zealand), [d]Department of Medicine, Univ. of Otago, Christchurch (New Zealand), [e]Department of Medicine, The Univ. of Auckland (New Zealand) ,[f]CIRIMAT, Univ. de Toulouse, CNRS, INPT, UPS, ENSIACET (France), [g]Department of Diagnostic and Interventional, Lausanne Univ. Hospital (Switzerland).

SUMMARY STATEMENT: Multi-energy spectral photon counting CT can accurately differentiate MSU from calcium crystals and moderately differentiate hydroxyapatite from calcium pyrophosphate using a lower energy range 20-30 keV.

IMPLICATION FOR PATIENT CARE: N/A

ABBREVIATIONS LIST:

AUC: Area under ROC curve

BCP: Basic calcium phosphate

CPP: Calcium pyrophosphate

DECT: Dual energy CT

HA: Hydroxyapatite

MSU: Monosodium urate

OCP: Octacalcium phosphate

ROC: Receiver operating characteristic

SPCCT: Spectral photon counting computed tomography

ABSTRACT

Purpose: To determine the in-vitro diagnostic performance of multi-energy spectral photon-counting CT (SPCCT) in crystal-related arthropathies.

Methods: Four crystals (monosodium urate, MSU; calcium pyrophosphate, CPP; calcium hydroxyapatite, HA; octacalcium phosphate, OCP) were synthesized and blended with agar at the following concentrations: 240, 100, 100, and 100mg/mL, respectively. Two crystal suspensions of HA and CPP were prepared each at 50 and 150 mg/mL. Crystal suspensions were scanned on a pre-clinical SPCCT system at 80 kVp using four energy thresholds: 20, 30, 40, and 50 keV. Linear attenuation coefficients of crystals were compared using the receiver operating characteristic (ROC) curve statistics. Area under the ROC curves (AUCs), sensitivities, specificities, positive and negative predictive values (PPV and NPV) were calculated.

Results: For each pairwise crystal comparison, AUCs, sensitivities, specificities, PPV and NPV were all significantly higher in the lower energy range. MSU was very accurately differentiated from CPP (sensitivity, 98.1%; specificity, 99.5%; AUC), OCP (sensitivity, 97.4%; specificity, 99.4%; AUC, 0.98) and HA (sensitivity, 91.1%; specificity, 93.4%; AUC). HA was fairly differentiated from CPP (sensitivity, 70.8%; specificity, 73.6%; AUC, 0.80) and OCP (sensitivity, 68.9%; specificity, 73.1%; AUC, 0.99). CPP failed to be differentiated from OCP. Comparing different concentrations, HA at 150 mg/mL was fairly differentiated from CPP at 150 mg/mL (sensitivity, 67.8%; specificity, 79.6%; AUC, 0.83), and HA at 50 mg/mL was poorly differentiated from CPP at 50 mg/mL (sensitivity, 59.3%; specificity, 66.0%; AUC, 0.68).

Conclusions: Multi-energy SPCCT can differentiate MSU from calcium crystals with excellent diagnostic performance, and HA from CPP and OCP with moderate accuracy. However, distinguishing CPP from OCP is not feasible.

KEYWORDS: Crystal-related arthropathies, spectral photon-counting CT, calcium crystals, monosodium urate, diagnostic performances

INTRODUCTION

Crystal-related arthropathies are characterized by the accumulation and deposition of crystals in or around joints. Gout, characterized by the deposition of monosodium urate (MSU) crystals, is the most common form of crystal arthropathies and the most common inflammatory arthritis (1), with a rising prevalence ranging from 0.1 to 10% (2). Calcium pyrophosphate deposition (CPPD) disease, formerly known as pseudo-gout is the second most common crystal arthropathy and characterized by the accumulation of calcium pyrophosphate (CPP) crystals (3). The clinical picture of acute CPP crystal arthritis is often indistinguishable from acute gout flares. Even though the short-term treatment of these two acute arthritis is similar with non-steroidal anti-inflammatory drugs, the long-term management requires the correct identification of the various crystals involved. Another calcium crystals type involved in arthritis are basic calcium phosphate (BCP) crystals, which include calcium hydroxyapatite (HA) and octacalcium phosphate (OCP). BCP crystals can be involved in various arthritic and periarthritic conditions, such as Milwaukee shoulder syndrome, and may also play a role in the development of osteoarthritis (4).

The gold standard for crystal identification currently remains synovial fluid aspiration followed visualization with polarized light microscopy (5). However, synovial fluid or tophus aspiration may fail in approximately 25% of the cases, is an invasive procedure and may further rarely be complicated by infection. Furthermore, a recent study reported a moderate accuracy of 81% for the diagnosis of gout and only 68% for CPPD, showing that this technique could be suboptimal for the diagnosis of crystal-related arthropathies, especially for CPPD (6). Moreover, the identification of BCP crystals (HA and OCP) with polarized light microscopy is very challenging and non-reliable.

Among imaging techniques, dual-energy CT (DECT) has shown promising results for the identification of MSU crystal deposits in gout (7). However, its performance remains limited in the early stages of the disease (8). Concerning the identification of CPP crystals, a recent ex-vivo study showed a moderate sensitivity (78 %) with a high specificity (94%) (9). However, no imaging technique is capable of discriminating between calcium crystals (CPP, HA and OCP).

Multi-energy spectral photon counting CT (SPCCT) is a novel CT imaging technique capable of discriminating various materials using an energy-resolving photon-counting detector (10). The detector can count the number of photons and discriminate them as a function of their energy. Based on the predominance of photoelectric effect for calcium crystals with low-energy photons, we hypothesized that SPCCT with adapted energy bins could easily discriminate MSU from calcium crystals, but also discriminate between the various calcium crystal types (11,12).

Hence, we aimed to determine the in-vitro diagnostic performance of multi-energy spectral photon-counting CT (SPCCT) in crystal-related arthropathies. Our secondary goal was to study the influence of the sample's concentration on the discrimination performance.

MATERIALS AND METHODS

Experimental design

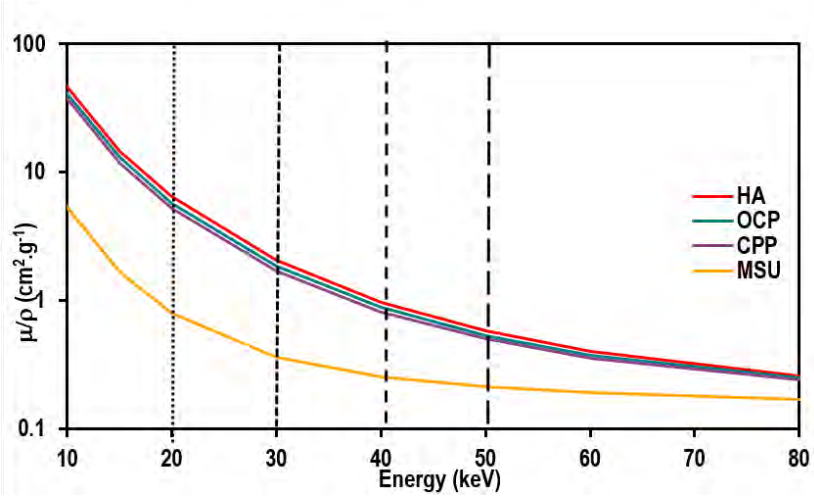
Four suspensions of HA at 100 mg/mL, OCP at 100 mg/mL, CPP at 100 mg/mL, and MSU at 240 mg/mL were prepared by blending synthetic crystal powders with agar. These concentrations were selected based on those encountered on clinical DECT images of patients affected by the various crystal-related arthropathies. A liquid agar solution at 1.5% was prepared by diluting the agar powder in water at 100°C. For each suspension, the exact amount of crystal was poured and blended into Eppendorf vials containing 200 µL of liquid agar. Agar was chosen as the suspension medium because of the quick solidification of agar can ensure that crystal suspensions are evenly dispersed in the volume of the vial. We assumed that the crystal suspension in Agar could be a realistic model of the precipitation of crystals in or around the joint in patients. In addition, the linear attenuation coefficient of agar is close to soft tissue and will thus not interfere with calcium attenuation. To study the influence of the concentration, two supplemental suspensions of HA and CPP, each at 50 mg/mL and 150 mg/mL were also prepared. Supplemental Table 1 reports the effective Z numbers, density, chemical formula and concentration of each crystal suspension in agar. Eppendorf vials containing MSU and the different calcium crystal suspensions were inserted into a customized 25-mm-diameter cylindrical polymethylmethacrylate (PMMA) phantom. Two vials containing only agar were added to serve as a reference.

	Zeff	Density	Chemical Formula	Concentration in agar
Monosodium urate (MSU)	9.5	2.12 g/cm3	C5H3N4O3Na	240 mg/mL
Calcium pyrophosphate (CPP)	14.6	2.56 g/cm3	Ca2P2O7,2H2O	50 mg/mL 100 mg/mL 150 mg/mL
Octacalcium phosphate (OCP)	15.1	2.67 g/cm3	Ca8H2 (PO4)6,5H2O	100 mg/mL
Calcium hydroxyapatite (HA)	15.9	3.16 g/cm3	Ca10(PO4)6(OH)2	50 mg/mL 100 mg/mL 150 mg/mL

Supplemental Table 1. Effective Z numbers, density, chemical formula and concentration in agar of each crystal suspension.

CT parameters and image reconstruction

The phantom was scanned on a pre-clinical multi-energy SPCCT system (MARS Bioimaging Ltd, Christchurch, New Zealand). The scanner is equipped with a polychromatic X-ray tube and a 14x14 mm CZT sensor bump bonded at 110 µm to a Medipix3RX readout chip. The detector chip is an energy-resolving single photon-counting detector that enables spectral x-ray imaging by counting and measuring the energy of individual photons transmitted through an object and assigning them to user-defined energy bins (13). Based on prior experience, tube voltage was set to 80 kVp with a 2 mm additional aluminium filtration (11). Tube current and exposure time were fixed to 27 µA and 220 ms, respectively, to give detector count rates of less than 5 per millisecond and limit the risk of pulse pile-up effect and saturation of the detector. Source-to-object and source-to-image-receptor distances were set to 200 and 250mm, respectively. Four energy thresholds of 20, 30, 40, and 50 keV were used to take advantage of the potential higher discrimination between calcium crystals and MSU for lower energy photons (supplemental Fig. 1). Image acquisition was performed using 720 circular projections over 360°. The slice thickness and field of view were equal to 0.1 mm and 48 x 48mm with 480 pixels, yielding cubic voxels of 0.1 x 0.1 x 0.1 mm^3. Images were reconstructed for each energy range using proprietary MARS Simultaneous Algebraic Reconstruction Technique (Figure 1) (14).



Supplemental Figure 1: Mass attenuation coefficients (µ/ρ, in $cm^2.g^{-1}$) for various crystals involved in crystal-related arthropathies as a function of photon energy (in keV), generated from the National Institute of Standards and Technology database. Vertical dashed lines represent the four user-defined energy thresholds (20, 30, 40 and 50 keV).

Data Analysis

CT images in which pixels are expressed in linear attenuation coefficients were directly analyzed to assess the potential discrimination between the various crystal combinations. The first step was to subtract the mean linear attenuation value

of the two agar-only vials to linear attenuation values of the various crystal suspensions, for each energy bin. Then, for each vial, fifty consecutive circular regions of interest of 8 mm^2 were sampled along the volume, yielding to 50000 pixels data extracted. Histograms of linear attenuation coefficients were then plotted for each energy bin and for each crystal suspension.

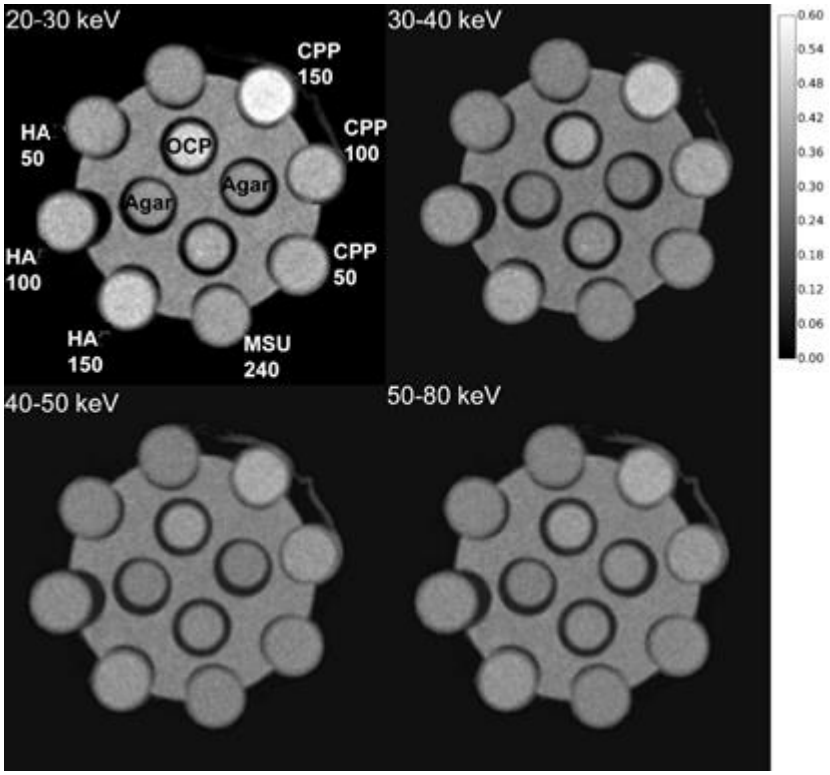


Figure 1: Energy-resolving MARS CT images of the phantom. The quantitative contrast scale corresponds to linear attenuation coefficients. On the top left image, the name and concentration of each crystal suspension were indicated.

Statistical Analysis

Statistical analysis was performed with Python 2.7. To assess the discrimination between pixel values for a pair of samples, a receiver operating characteristic curve (ROC) curve was plotted. The estimated empirical area under the ROC curve (AUC) and 95% confidence interval was calculated using the nonparametric asymptotic method proposed by DeLong (15,16), which made no assumption of data distribution. AUC was used as the figure of merit to assess the differentiation between two crystals considering that the higher the AUC value, the easier would be the discrimination task. To discriminate the AUC values between two energy bins, a z-test taking into account the correlation between the data were used (17). The effect size measure Cohen's d was computed to interpret the statistical significance of the results (18,19). The cut-off that maximized the sum between sensitivity and specificity on the ROC curve was determined in order to calculate best parameter performances: sensitivity and specificity. 95% confidence interval for each parameter was calculated using a bootstrap resampling method with 1000 iterations.

RESULTS

Figure 2 shows the histograms of linear attenuation coefficients for various crystals: MSU at 240 mg/mL, OCP at 100 mg/mL, CPP at 100 mg/mL and HA at 100 mg/mL for each of the four energy bins. Mean linear attenuation values decreased for the higher energy ranges and the overlap of various histograms increased

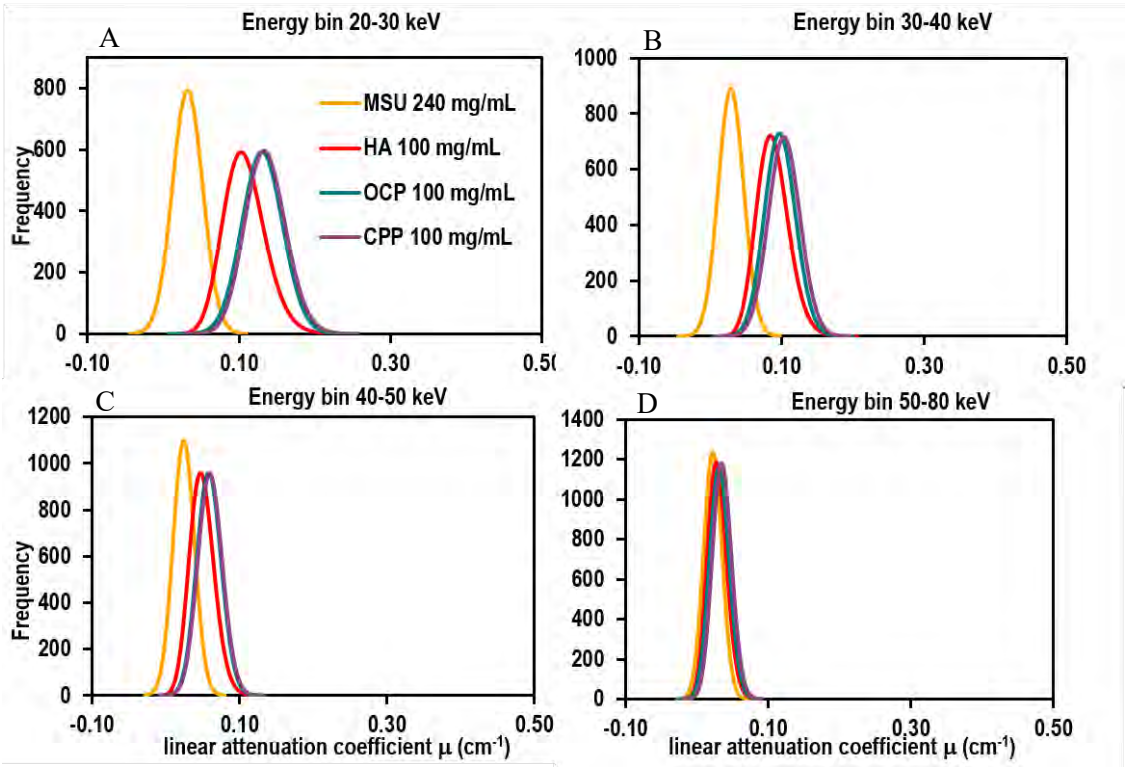


Figure 2: Histograms of linear attenuation coefficients of HA, OCP, CPP and MSU crystals for the four energy bins: (A) 20-30 keV, (B) 30-40 keV, (C) 40-50 keV, (D) 50-80 keV. The curves were fitted using a Gaussian function.

Discrimination between MSU and calcium crystals

In figure 3a, ROC curves show that the discrimination between MSU and the mix of the calcium crystals was easier with decreasing the energy range. In Table 1, AUC and the various ROC parameters values showed that MSU can be accurately differentiated from calcium crystals with an AUC upper to 0.99 for the two first energy bins. For the two higher energy bins, the discrimination is more difficult with an AUC of 0.92.

Comparing MSU and each calcium crystal type for the different energy bins, a similar behavior was observed for the discrimination between (Figure 3b, 3c, 3d). A statistical z test showed that AUC values for the first energy bin was statistically higher than AUC values for the second energy bin ($p<0.0001$) with a z score equal to 71.9, 44 and 34.5 for the combinations MSU vs HA, MSU vs OCP and MSU vs CPP, respectively. The effect size measure, Cohen's d was equal to 0.461, 0.281 and 0.219 for the combinations MSU vs HA, MSU vs OCP and MSU vs CPP, respectively,

indicating that the superiority of the first energy bin to discriminate MSU from calcium crystals was not practically relevant. The best discrimination performances were found between MSU and CPP, and between MSU and OCP yielding an AUC of 0.99. The discrimination between MSU and HA yielded an AUC of 0.98.

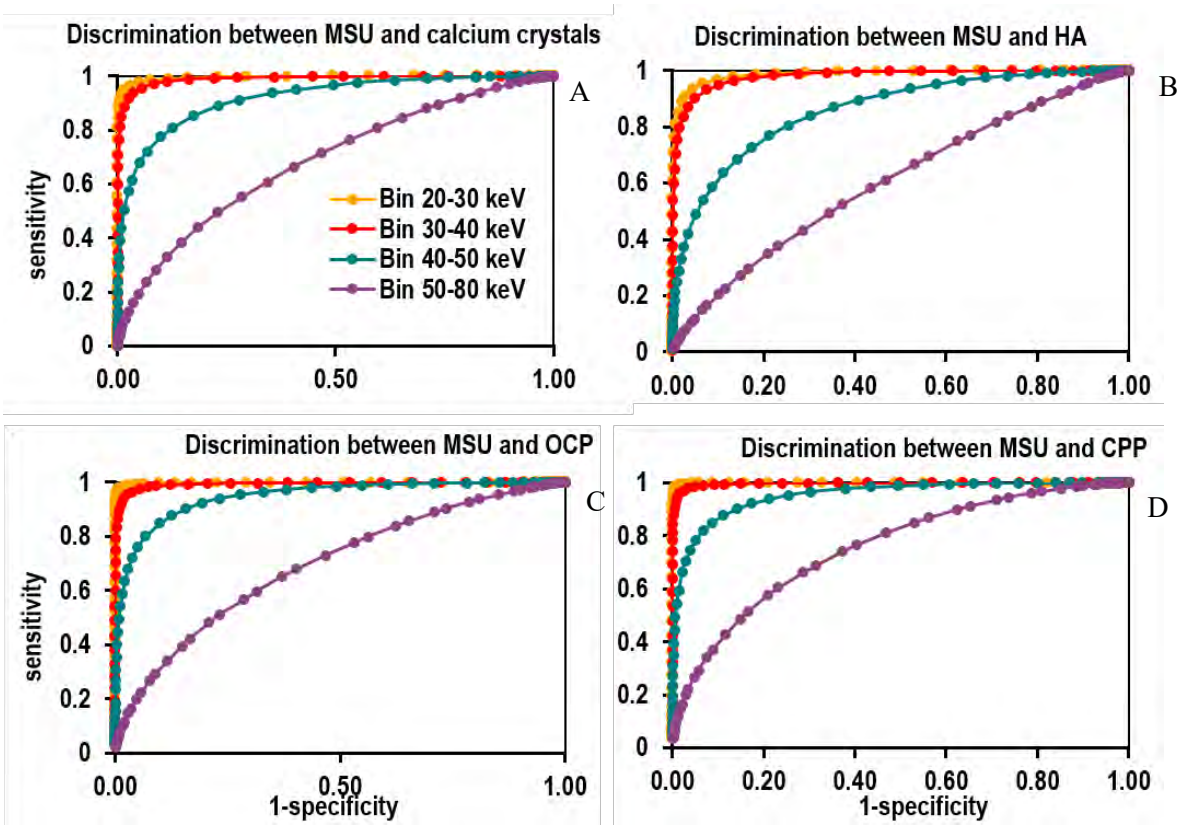


Figure 3: ROC curves for the three combinations between MSU and the various calcium crystals with the four energy bins: (A) MSU and calcium crystals, (B) MSU and HA, (C) MSU and OCP, (D) MSU and CPP.

Discrimination between calcium crystals

In figure 4a, ROC curves show a moderate discrimination between HA and CPP, HA and OCP with an AUC equal to 0.80 and 0.77 respectively. Then, AUC results decreased with increasing the energy range. However, the discrimination between OCP and CPP was poor with an AUC equal to 0.52 for the first energy bin. For this combination, AUC values were not decreasing with the energy range (see Table 1).

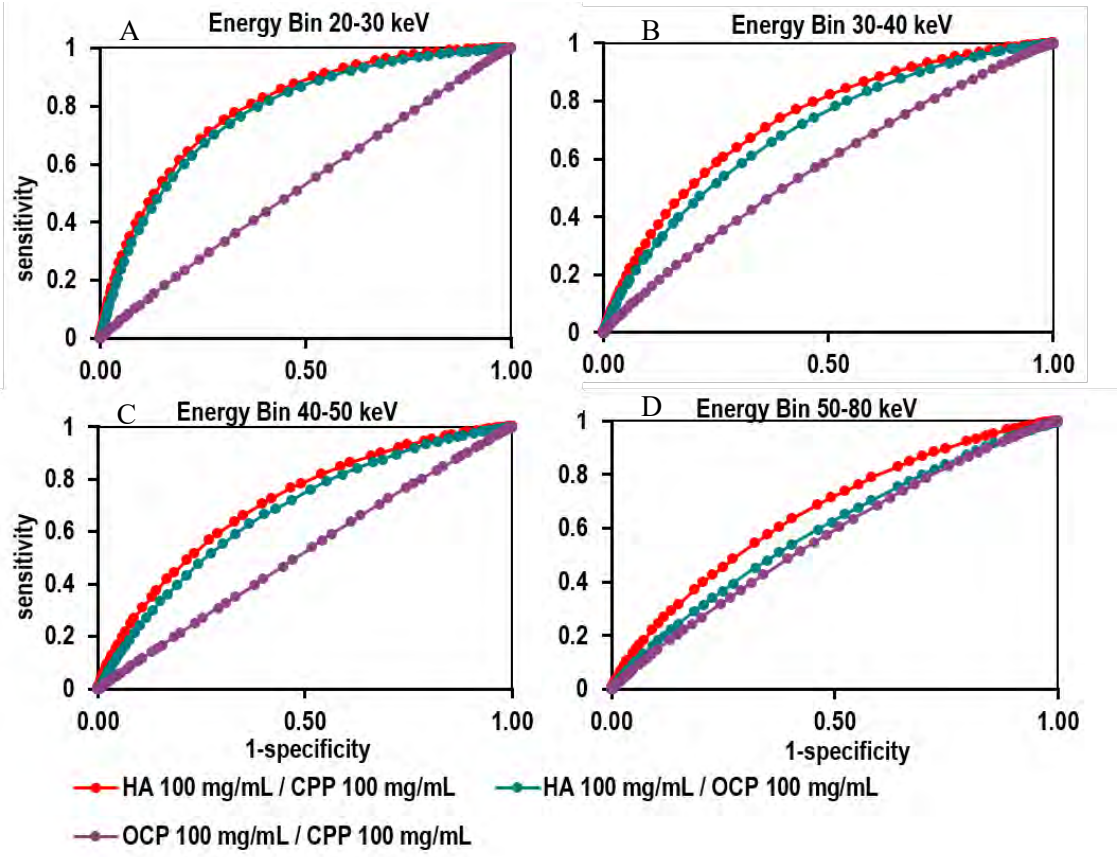


Figure 4: ROC curves for the three combinations between calcium crystals HA and CPP, HA and OCP, OCP and CPP for the four energy bins: (A) 20-30 keV, (B) 30-40 keV, (C) 40-50 keV, (D) 50-80 keV

		AUC	Sensitivity	Specificity
MSU vs Calcium crystals	Bin 20-30 keV	0.995 [0.995-0.996]	0.965 [0.957-0.968]	0.971 [0.968-0.979]
	Bin 30-40 keV	0.99 [0.99-0.991]	0.942 [0.935-0.95]	0.961 [0.953-0.966]
	Bin 40-50 keV	0.921 [0.919-0.923]	0.814 [0.792-0.839]	0.869 [0.842-0.888]
	Bin 50-80 keV	0.686 [0.683-0.69]	0.533 [0.494-0.556]	0.735 [0.712-0.774]
MSU 240 mg/mL vs HA 100 mg/mL	Bin 20-30 keV	0.988 [0.987-0.989]	0.911 [0.9-0.924]	0.969 [0.962-0.975]
	Bin 30-40 keV	0.981 [0.98-0.982]	0.931 [0.92-0.94]	0.926 [0.916-0.936]

	Bin 40-50 keV	0.861 [0.858-0.863]	0.778 [0.746-0.792]	0.778 [0.763-0.808]
	Bin 50-80 keV	0.607 [0.604-0.611]	0.536 [0.491-0.558]	0.615 [0.592-0.661]
MSU 240 mg/mL vs OCP 100 mg/mL	Bin 20-30 keV	0.998 [0.998-0.999]	0.974 [0.969-0.977]	0.994 [0.992-0.995]
	Bin 30-40 keV	0.994 [0.994-0.994]	0.960 [0.954-0.97]	0.968 [0.957-0.973]
	Bin 40-50 keV	0.948 [0.947-0.95]	0.868 [0.848-0.886]	0.884 [0.866-0.904]
	Bin 50-80 keV	0.696 [0.693-0.7]	0.573 [0.51-0.654]	0.709 [0.631-0.769]
MSU 240 mg/mL vs CPP 100 mg/mL	Bin 20-30 keV	0.999 [0.999-0.999]	0.981 [0.975-0.985]	0.995 [0.993-0.996]
	Bin 30-40 keV	0.996 [0.996-0.997]	0.971 [0.967-0.979]	0.975 [0.967-0.978]
	Bin 40-50 keV	0.953 [0.952-0.955]	0.878 [0.862-0.891]	0.886 [0.873-0.902]
	Bin 50-80 keV	0.757 [0.754-0.76]	0.649 [0.63-0.664]	0.727 [0.709-0.747]

Table 1: Mean values for AUC and best parameter performances for combinations between MSU and calcium crystals, MSU and HA, MSU and OCP, MSU and CPP. 95% confidence intervals were detailed in brackets. Concentration of each crystal suspension was indicated in the first column in mg/mL.

Impact of the concentration

Figure 6 shows the histograms of HA and CPP for three concentrations: 50 mg/mL, 100 mg/mL and 150 mg/mL and for the different energy bins. As expected mean linear attenuation value decreased for the lowest concentration and for the higher energy bin. The overlap of various histograms increased for the higher energy ranges. Comparing the pair HA/CPP for identical concentration and for each energy bin, we found that the overlap of histograms increased when decreasing the concentration. For the highest concentration 150 mg/mL, HA can be differentiated from CPP with an AUC of 0.83, yielding a sensitivity of 71% and a specificity of 83%. For the lowest concentration 50 mg/mL, the

discrimination between HA and CPP was more difficult with an AUC of 0.68, yielding a sensitivity of 60% and a specificity of 68% (see supplemental table 2).

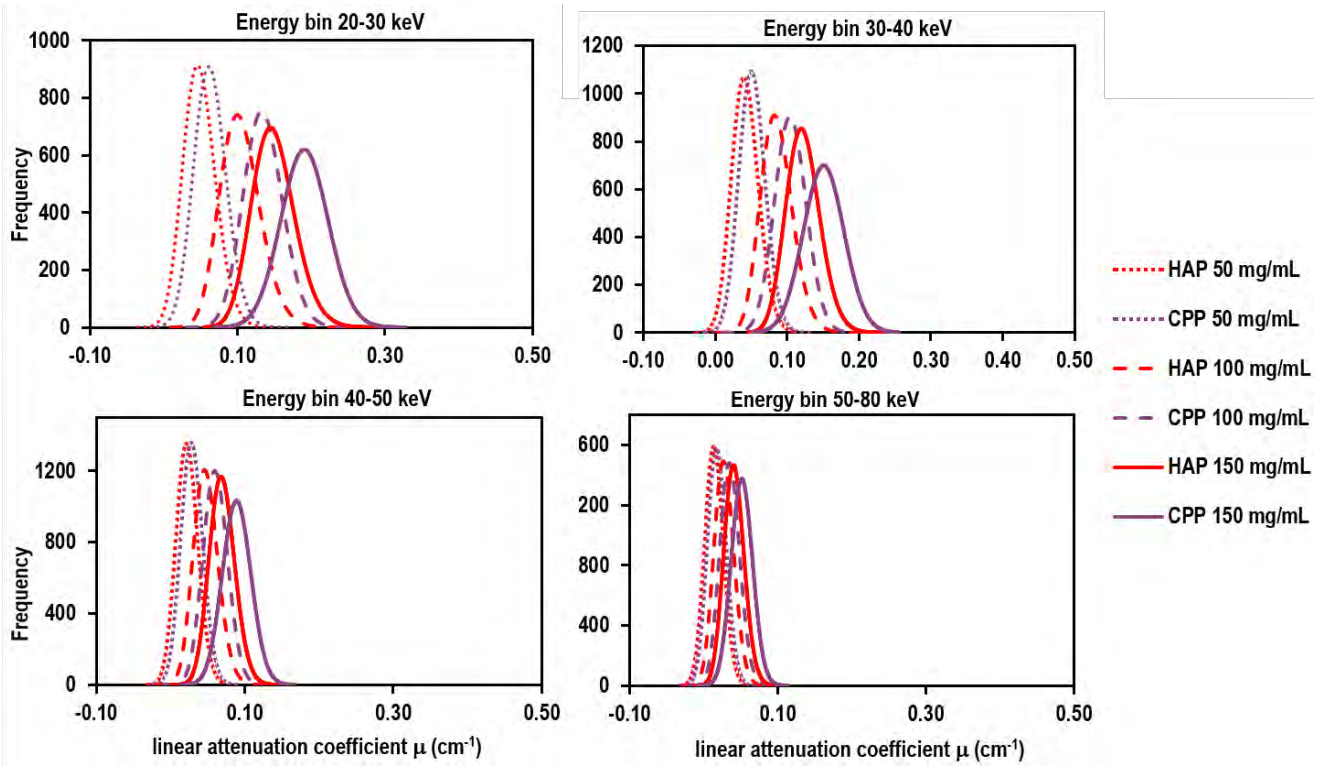


Figure 6: Histograms of linear attenuation coefficients of HA and CPP at 50 mg/mL, 100 mg/mL and 150 mg/mL the four energy bins: (A) 20-30 keV, (B) 30-40 keV, (C) 40-50 keV, (D) 50-80 keV.

	AUC	Sensitivity	Specificity
CPP 150 vs HA 150	0.833 [0.831-0.836]	0.705 [0.678-0.736]	0.826 [0.796-0.846]
CPP 150 vs HA 100	0.974 [0.973-0.975]	0.89 [0.882-0.902]	0.948 [0.939-0.953]
CPP 150 vs HA 50	1 [1-1]	0.989 [0.987-0.992]	0.993 [0.991-0.995]
CPP 100 vs HA 150	0.642 [0.638-0.645]	0.521 [0.462-0.59]	0.675 [0.605-0.73]
CPP 100 vs HA 100	0.795 [0.792-0.797]	0.709 [0.665-0.728]	0.735 [0.718-0.774]
CPP 100 vs HA 50	0.993 [0.993-0.994]	0.956 [0.946-0.963]	0.969 [0.965-0.975]

CPP 50 vs HA 150	0.993 [0.993-0.993]	0.943 [0.933-0.95]	0.97 [0.965-0.976]
CPP 50 vs HA 100	0.885 [0.883-0.887]	0.748 [0.713-0.766]	0.85 [0.836-0.875]
CPP 50 vs HA 50	0.681 [0.677-0.684]	0.593 [0.564-0.659]	0.666 [0.601-0.694]

Supplemental Table 2: Mean values for AUC and best parameter performances for the combinations between CPP and CHA at 50, 100 and 150 mg/mL for the first energy range only (20-30 keV). Concentration of each crystal suspension was indicated in the first column in mg/mL.

DISCUSSION

The in-vitro study shows that the multi-energy SPCCT can accurately discriminate MSU and calcium crystals (HA, CPP, OCP). The results are in line with previous findings using a DECT imaging technique (8). Concerning the discrimination between calcium crystals, only the pair HA and CPP can be moderately discriminate. This result is very promising because so far no other imaging technique have shown similar performances.

Concerning the SPCCT performances, the discrimination was better using the lower energy ranges than the higher energy ranges. The results comes directly from the predominant photoelectric effect for low energy photons. The imaging protocol was currently optimized to avoid the risk of pile-up and saturation of the detector that limit spectral imaging performances. A better tuning of the energy thresholds could achieve a better discrimination between two crystals. The methodology proposed to assess the discrimination between two materials is promising for spectral imaging data. However, it must be applied for each energy bin. A manual selection of the best one to discriminate two crystals is then required. The major limitation of this work dealt with the preparation of crystals suspension samples in agar. Firstly, the samples stability is not guaranteed over a long period. Secondly, even if results showed a good linear relationship between linear attenuation coefficients and concentration of the various materials, the uncertainty on the concentration value is unknown.

The major strength of the work is the in-vitro demonstration of the capability of a new multi-energy SPCCT prototype to discriminate the various crystals involved in crystal-related arthropathies. If the results translate to ex-vivo human samples and latterly to clinical practice, the diagnosis of crystal-related arthropathies could be achieved using a multi-energy SPCCT.

REFERENCES

1. Neogi T, Jansen TLTA, Dalbeth N, et al. 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2015;74(10):1789–1798.
2. Kuo C-F, Grainge MJ, Zhang W, Doherty M. Global epidemiology of gout: prevalence, incidence and risk factors. Nat Rev Rheumatol. 2015;11(11):649–662.
3. Rosenthal AK, Ryan LM. Calcium Pyrophosphate Deposition Disease. Campion EW, editor. N Engl J Med. 2016;374(26):2575–2584.
4. Rosenthal AK. Basic calcium phosphate crystal-associated musculoskeletal syndromes: an update. Curr Opin Rheumatol. 2018;30(2):168–172.
5. Dieppe P, Swan A. Identification of crystals in synovial fluid. Ann Rheum Dis. 1999;58(5):261–263.
6. Berendsen D, Neogi T, Taylor WJ, Dalbeth N, Jansen TL. Crystal identification of synovial fluid aspiration by polarized light microscopy. An online test suggesting that our traditional rheumatologic competence needs renewed attention and training. Clin Rheumatol. 2017;36(3):641–647.
7. Choi HK, Burns LC, Shojania K, et al. Dual energy CT in gout: a prospective validation study. Ann Rheum Dis. 2012;71(9):1466–1471.
8. Bongartz T, Glazebrook KN, Kavros SJ, et al. Dual-energy CT for the diagnosis of gout: an accuracy and diagnostic yield study. Ann Rheum Dis. 2015;74(6):1072–1077.
9. Tanikawa H, Ogawa R, Okuma K, et al. Detection of calcium pyrophosphate dihydrate crystals in knee meniscus by dual-energy computed tomography. J Orthop Surg. 2018;13(1):73.
10. Ronaldson JP, Zainon R, Scott NJA, et al. Toward quantifying the composition of soft tissues by spectral CT with Medipix3: Quantifying the composition of soft tissues. Med Phys. 2012;39(11):6847–6857.
11. Kirkbride TE, Raja AY, Müller K, Bateman CJ, Becce F, Anderson NG. Discrimination Between Calcium Hydroxyapatite and Calcium Oxalate Using Multienergy Spectral Photon-Counting CT. Am J Roentgenol. 2017;209(5):1088–1092.
12. Viry A, Combes C, Verdun FR, et al. Multi-energy spectral photon-counting CT in crystal-related arthropathies: initial experience and diagnostic performance in vitro. In: Chen G-H, Lo JY, Gilat Schmidt T, editors. Med Imaging 2018 Phys Med Imaging. Houston, United States: SPIE; 2018. p. 187https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10573/2293458/Multi-energy-spectral-photon-counting-CT-in-crystal-related-arthropathies/10.1117/12.2293458.full. Accessed November 26, 2018.
13. Ballabriga R, Alozy J, Blaj G, et al. The Medipix3RX: a high resolution, zero dead-time pixel detector readout chip allowing spectroscopic imaging. J Instrum. 2013;8(02):C02016–C02016.
14. Tang ND, de Ruiter N, Mohr JL, Butler APH, Butler PH, Aamir R. Using algebraic reconstruction in computed tomography. Proc 27th Conf Image Vis Comput N Z - IVCNZ 12. Dunedin, New Zealand: ACM Press; 2012. p. 216–221http://dl.acm.org/citation.cfm?doid=2425836.2425881. Accessed December 10, 2018.
15. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–845.
16. Hanley JA, Hajian-Tilaki KO. Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: An update. Acad Radiol. 1997;4(1):49–58.
17. Zhou XH, Gatsonis CA. A SIMPLE METHOD FOR COMPARING CORRELATED ROC CURVES USING INCOMPLETE DATA. Stat Med. 1996;15(15):1687–1693.
18. Lenhard W, Lenhard A. Computation of Effect Sizes. Psychometrica; 2014.http://rgdoi.net/10.13140/RG.2.1.3478.4245. Accessed November 28, 2018.
19. Cohen J. Chi-Square Tests for Goodness of Fit and Contingency Tables. Stat Power Anal Behav Sci. Elsevier; 1977. p. 215–271https://linkinghub.elsevier.com/retrieve/pii/B9780121790608500128. Accessed November 28, 2018.

5.4 **Clinical utility of multi-energy spectral CT in crystal arthritis**

Published in Arthritis and Rheumatology [93]

DR. FABIO BECCE (Orcid ID : 0000-0001-8444-8504)

Article type : Brief Report

Title: Clinical utility of multi-energy spectral photon-counting CT in crystal arthritis

Authors: Lisa K Stamp, FRACP, PhD[1], Nigel G Anderson, RANZCR[2], Fabio Becce, MD[3], Maya Rajeswari[2], Matthew Polson, PhD[4], Olivier Guyen, MD[5], Anais Viry, MSc[6], Chloe Choi[2], Tracy E Kirkbride, PhD[7], and Aamir Y Raja, PhD[2]

Affiliations: [1]Department of Medicine, University of Otago, Christchurch, New Zealand; [2]Department of Radiology, University of Otago, Christchurch, New Zealand; [3]Department of Diagnostic and Interventional Radiology, Lausanne University Hospital, Lausanne, Switzerland; [4]School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand; [5]Department of Orthopaedics and Traumatology, Lausanne University Hospital, Lausanne, Switzerland; [6]Institute of Radiation Physics, Lausanne University Hospital, Lausanne, Switzerland; and [7]Ara Institute of Canterbury, Christchurch, New Zealand

Corresponding author: Professor Lisa K Stamp, Department of Medicine, University of Otago, Christchurch, P.O. Box 4345, Christchurch 8140, New Zealand; ***Phone:*** +64-3-364-0253; ***Fax:*** +64-3-364-0935; ***Email:*** lisa.stamp@cdhb.health.nz

Acknowledgements

We are grateful to MARS research team in Christchurch, New Zealand. The MARS spectral scanner was developed through a project funded by Ministry of Business, Innovation and Employment (MBIE), New Zealand under contract number UOCX0805.

Declaration of interests

Lisa K Stamp reports grants from Health Research Council of New Zealand, during the conduct of the study; grants from Ardea Biosciences and Health Research Council of New Zealand, outside the submitted work; Nigel G Anderson is a shareholder of MARS Bioimaging Ltd. Others have no financial holding in the company, nor any other conflict of interest to disclose.

Abstract

Objective: To determine whether novel multi-energy spectral photon-counting CT (SPCCT) imaging can detect and differentiate between monosodium urate (MSU), calcium pyrophosphate (CPP) and hydroxyapatite (HA) crystal deposits *ex vivo*.

Methods: A finger with a subcutaneous gouty tophus and a calcified knee meniscus excised at the time of surgery were obtained. The finger was imaged using plain X-ray, dual-energy CT (DECT) and multi-energy SPCCT. Plain X-ray and multi-energy SPCCT images of the meniscus were acquired. For validation purposes, samples of the crystals were obtained from the tophus and meniscus, and examined by polarised light microscopy and/or X-ray diffraction. As a further validation, synthetic crystal suspensions of MSU, CPP and HA were scanned using multi-energy SPCCT.

Results: Plain X-ray of the gouty finger revealed bone erosions with overhanging edges. DECT and multi-energy SPCCT both identified MSU crystal deposits; SPCCT was able to detect finer detail. Plain X-ray of the calcified meniscus identified chondrocalcinosis consistent with CPP, while SPCCT identified and differentiated CPP and HA.

Conclusion: Multi-energy SPCCT can not only detect, differentiate and quantify MSU crystal deposits in a gouty finger *ex vivo*, but also specifically detect, identify and quantify CPP within an osteoarthritic meniscus, and distinguish them from HA crystal deposits. There is potential for multi-energy SPCCT to become useful in the diagnosis of crystal arthropathies.

Key words: multi-energy spectral photon-counting CT, monosodium urate, calcium pyrophosphate, calcium hydroxyapatite, crystals, imaging

Introduction

A variety of crystals have been associated with arthritis. Gout, which is due to the presence of monosodium urate (MSU) crystal deposition, is the classical crystal arthritis and one of the most common forms of inflammatory arthritis. However, calcium crystals such as calcium pyrophosphate (CPP) and basic calcium phosphate (BCP) can also produce acute inflammatory arthritis known as "pseudogout" and Milwaukee shoulder syndrome, respectively. The clinical presentation of gout flare and acute CPP crystal arthritis are frequently indistinguishable with a red, hot, swollen and exquisitely painful joint. While the acute management of gout and CPP crystal arthritis are similar, namely non-steroidal anti-inflammatory or corticosteroid therapy, the long-term management of these conditions differs with urate lowering therapy indicated for those with gout but not CPP crystal arthritis. Importantly, the clinical picture of an acute crystal arthritis may be indistinguishable from

septic arthritis, which if not treated appropriately with anti-microbial therapy may result in significant morbidity and mortality. Thus, prompt accurate diagnosis of crystal arthritis is important.

The gold standard for the diagnosis of crystal arthritis requires aspiration of synovial fluid or tophus and direct visualisation of crystals by polarised light microscopy. (1)In the case of gout, MSU crystals appear needle-shaped and negatively birefringent, while CPP crystals are typically monoclinic or triclinic with weak positive or no birefringence. BCP crystals cannot be detected by polarised light microscopy. However, crystal identification is operator dependent and may be suboptimal particularly for crystals other than MSU.(2) For example, in a recent on-line test undertaken by 110 individuals with an interest in crystal arthritis, including rheumatologists, rheumatology trainees and laboratory technicians, MSU crystals were correctly identified by 81%, while CPP crystals by only 68%.(3) Misidentification of crystals may result in inappropriate or lack of appropriate treatments such as urate lowering therapy or antibiotics. Aspiration of an acutely inflamed joint is painful and while rare, joint aspiration may also be associated with complications such as iatrogenic infection. Thus, more reliable and less invasive means of accurately identifying and differentiating between the various types of crystals are of interest.

Other imaging modalities such as ultrasound and dual-energy CT (DECT) are being increasingly used to aid in the diagnosis of gout.(4) A recent study showed that DECT can also detect CPP crystal deposits *ex vivo* with a reported sensitivity and specificity of 78% and 94%, respectively.(5) However, the role of these imaging techniques is less well defined in other crystal deposition diseases and they can not easily distinguish between different calcium crystals. The ability to reliably detect and differentiate MSU, CPP and calcium

hydroxyapatite (HA) crystal deposits using non-invasive imaging would be an advance. Multi-energy spectral photon-counting CT (SPCCT) is a novel imaging technique which uses a standard polychromatic X-ray source and photon-counting detector that records the number and energy of transmitted photons into multiple energy bins,(6) providing a range of energy-dependent Hounsfield units.(7) As X-ray attenuation of each material is energy dependent, multi-energy bin data provides specific identification and quantification of several materials simultaneously.(8, 9) Our previous findings show that multi-energy SPCCT can differentiate MSU, CPP and HA *in vitro*.(10, 11) The aim of this study was to determine whether we could detect and differentiate between MSU, CPP and HA crystal deposits in excised human specimens using multi-energy SPCCT imaging.

Materials and methods

Ethical approval was obtained from the University of Otago Human Health Ethics Committee, The Health and Disability Ethics Committee of New Zealand and from the Institutional Ethics Committee (CER-VD) of Lausanne University Hospital.

A finger with a subcutaneous gouty tophus was obtained from the Christchurch Cancer Tissue Bank (Figure 1A) and imaged using plain X-ray, DECT and a preclinical multi-energy SPCCT scanner (MARS Bioimaging Ltd, Christchurch, New Zealand). For validation purposes, crystals from a core of the tophus were obtained (Figure 1A) and a sample of the tophus material was examined by polarised light microscopy.

A calcified meniscus excised at the time of total knee joint replacement surgery was obtained (Figure 2B) and imaged using plain X-ray and multi-energy SPCCT. For validation purposes, crystals were obtained from scrapings of the surface of the meniscus and examined by

polarised light microscopy and X-ray diffraction (XRD). For XRD, samples were mounted on a nylon loop in a minimum amount of perfluorinated oil at 298 K. Data were collected on an Agilent SuperNova (Dual, copper at zero, Atlas) diffractometer (Agilent Technologies, Yarnton, UK). The images were integrated using CrysAlisPro (1.171.38.43 (Rigaku Oxford Diffraction, 2015)) and the background was manually subtracted using Origin 8.6001. As a further validation step, synthetic crystal suspensions with known concentrations of MSU (240 mg/cm^3), CPP (88 mg/cm^3) and HA (72 mg/cm^3) were prepared as described in greater detail elsewhere, (11) and then scanned using multi-energy SPCCT with the same protocol as described below. The resulting multi-energy SPCCT images (in which pixels are expressed as linear attenuation coefficients, Supplementary Figure 1) were subsequently analyzed according to the method previously described by Viry et al. to determine the diagnostic accuracy for the differentiation of each crystal pair. (11)

Multi-energy images were obtained using a preclinical SPCCT system suitable for small animal models and human specimens of diseased tissue.(12) The scanner incorporated a microfocus poly-energetic X-ray source and a Medipix3RX energy-resolving photon-counting detector within a continuous rotating gantry. We used a 2-mm-thick cadmium zinc telluride sensor (14 × 14 mm^2), bump bonded at 110 μm to a Medipix3RX readout chip. Multi-energy SPCCT imaging of biological specimens and calibration phantoms containing known concentrations of MSU, CPP and HA crystal suspensions was performed using four charge-summing mode energy thresholds of 20, 30, 40 and 50 keV at 80 kVp. (11) Imaging data were reconstructed simultaneously into non-overlapping energy bins across the measured spectrum such as 20-30, 30-40, 40-50 and 50-80 keV, at a spatial resolution of 0.09 × 0.09 × 0.09 mm (Supplementary Figure 1). Images of the finger with subcutaneous gouty tophus and calcified meniscus were calibrated using mass attenuation coefficients estimated

from the calibration phantom data to distinguish MSU, CPP and HA crystal deposits, (11) and furnish material decomposition maps.

Dual-energy images were obtained using a dual-source DECT system (Somatom Flash, Siemens Healthineers, Erlangen, Germany) with two X-ray tubes and two corresponding energy-integrating detectors mounted at 90° to one another and operating at 80 and 140 kVp. Data were acquired and images reconstructed using scanning parameters routinely applied for people with gout, at a spatial resolution of 0.23 × 0.23 × 0.60 mm.(13) The standard proprietary gout setting on a syngo.via workstation (Siemens Healthinners) was used for post-processing.

Results

Plain X-ray of the excised gouty finger revealed bone erosions with overhanging edges and soft tissue changes consistent with tophus material (Figure 1B). DECT (Figure 1C) and multi-energy SPCCT (Figure 1D) both identified MSU crystal deposits; however, SPCCT was able to detect finer detail and higher MSU volume. Polarised light microscopy confirmed the presence of negatively birefringent MSU crystals from the tophus (Figure 1E).

Preoperative plain X-ray of the right knee revealed chondrocalcinosis (Figure 2A). The chondrocalcinosis was confirmed to be within the meniscus on a plain X-ray of the excised lateral meniscus (Figure 2C). Multi-energy SPCCT showed a predominance of CPP crystal deposits in the calcified regions of the excised meniscus (Figures 2D-E). Both CPP (Figure 2E) and HA crystal deposits (Figure 2F) could be identified and differentiated. Multi-energy SPCCT imaging findings of the meniscus were partly validated using polarised light microscopy (Figure 2G) and XRD (Figure 2H) of the crystal scrapings. The diffraction

pattern obtained from the sample was assigned as a mixture of monoclinic and triclinic CPP dihydrate (m-CPPD and t-CPPD).(14) There were no significant peaks to indicate the presence of HA. The diagnostic accuracy measures for differentiating each pair of synthetic crystal suspensions are reported in Table 1 and Supplementary Table 1. Overall, the differentiation of MSU from both calcium crystals was excellent (area under the receiver operating characteristic curve (AUC) > 0.99 for the 20-30 keV energy bin), while the discrimination between CPP and HA was moderate (AUC = 0.78 for the 20-30 keV energy bin).

Discussion

Herein we have shown that multi-energy SPCCT can not only detect and differentiate MSU crystal deposits in an excised gouty finger, but also specifically detect and identify CPP crystal deposits within an excised osteoarthritic knee meniscus, and distinguish them from HA crystal aggregates. This is the first time multi-energy SPCCT has been used to study *ex-vivo* human samples for the presence of pathogenic crystals.

Although DECT can be used to detect and characterise MSU and CPP crystal deposits with a variable diagnostic performance,(15, 16) we have shown that multi-energy SPCCT has the potential to also distinguish other calcium crystal aggregates such as HA.(10, 11) While XRD and polarised light microscopy can be used to validate our findings of MSU and CPP crystals, it is more difficult to validate HA which cannot be identified using these techniques. As an alternate validation method, we have shown that multi-energy SPCCT, especially when using lower energy bins such as 20-30 keV, can differentiate synthetic crystal suspensions with known clinically-relevant concentrations of MSU and CPP with excellent diagnostic accuracy

(Supplementary Table 1), while the accuracy for distinguishing CPP from HA was moderate (Table 1).

Multi-energy SPCCT has potential advantages over DECT in detecting and characterising MSU crystal deposits. The images show a finely detailed punctate pattern of MSU crystal deposition in contrast to the clump-like appearance on DECT. This difference might be due in part to increased sensitivity of SPCCT; however, the finer spatial resolution of a preclinical micro-CT compared to a clinical whole-body CT system may also contribute. Higher spatial resolution and lower partial volume effects certainly also contribute to improved diagnostic accuracy of multi-energy SPCCT (AUC > 0.99 for MSU with 20-30 and 30-40 keV energy bins, Supplementary Table 1) compared to 0.96 AUC for DECT according to a recent meta-analysis. (15) Further work using samples with smaller urate and calcium burden will be required to determine the diagnostic accuracy and lower limit of detection of MSU and calcium crystal deposits using multi-energy SPCCT. With these initial data, we are now in a position to demonstrate that we can replicate the image quality we achieved for *ex-vivo* samples in live subjects with crystal deposition diseases. The first living human images from a healthy volunteer, scanned using a MARS 8-energy SPCCT system, have recently been released. These images have produced much excitement in the short time that the technology has been available. The human MARS scanner is almost ready for clinical trials, while ethics approval is currently being sought. Technical details of human 8-energy SPCCT scanner have just been presented in November 2018.

Author contributions

Lisa K Stamp – study conception and design, data acquisition, analysis and interpretation, literature review, manuscript drafting and revision

Nigel G Anderson – study conception and design, data acquisition, analysis and interpretation, literature review, manuscript drafting and revision

Fabio Becce – study conception and design, data analysis and interpretation, literature review, manuscript revision

Maya Rajeswari – data acquisition, analysis and interpretation, manuscript revision

Matthew Polson – data acquisition, analysis and interpretation, manuscript revision

Olivier Guyen – data acquisition, manuscript revision

Anais Viry – data acquisition and analysis, manuscript revision

Chloe Choi – data acquisition, analysis and interpretation, literature review, manuscript drafting and revision

Tracy E Kirkbride – data acquisition, analysis and interpretation, literature review, manuscript drafting and revision

Aamir Y Raja – study conception and design, data acquisition, analysis and interpretation, literature review, manuscript drafting and revision

References

1. Neogi T, Jansen T, Dalbeth N, Fransen J, Schumacher H, Berendsen D, et al. 2015 Gout Classification Criteria. An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. Arthritis Rheum. 2015;67(10):2557-68.

2. Dieppe P, Swan A. Identification of crystals in synovial fluid. Ann Rheum Dis. 1999;58(5):261-3.

3. Berendsen D, Neogi T, Taylor W, Dalbeth N, Jansen T. Crystal identification of synovial fluid aspiration by polarized light microscopy. An online test suggesting that our traditional rheumatologic competence needs renewed attention and training. Clin Rheumatol. 2017;36:641-47.

4. Neogi T, Jansen T, Dalbeth N, Fransen J, Schumacher H, Berendsen D, et al. 2015 Gout Classification Criteria. An American College of Rheumatology/European League Against Rheumatism Collaborative Initiative. Ann Rheum Dis. 2015;74(10):1789-98.

5. Tanikawa H, Ogawa R, Okuma K, Harato K, Niki Y, Kobayashi S, et al. Detection of calcium pyrophosphate dihydrate crystals in knee meniscus by dual-energy computed tomography. J Ortho Surgery Res. 2018;13:73.

6. Ballabriga R, Alozy J, Blaj G, Campbell M, Fiederle M, Frojdh E, et al. The Medipix3RX: a high resolution, zero dead-time pixel detector readout chip allowing spectroscopic imaging. JINST. 2013;8:C02016.

7. Hurrell M, Butler A, Cook N, Butler P, Ronaldson J, Zainon R. Spectral Hounsfield units: a new radiological concept. Eur Radiol. 2012;22(5):1008-13.

8. Moghiseh M, Panta R, de Ruiter N, Chernoglazov A, Healy J, Butler A, et al. Discrimination of Multiple High-Z Materials by Multi-Energy Spectral CT– A Phantom Study. JSM Biomed Imaging Data. 2016;3(1):1007.

9. Raja A, et al., . . Measuring Identification and Quantification Errors in Spectral CT Material Decomposition. Appl Sci.8(3):467.

10. Kirkbride T, Raja A, Müller K, Bateman C, Becce F, Anderson N. Discrimination Between Calcium Hydroxyapatite and Calcium Oxalate Using Multienergy Spectral Photon-Counting CT. Am J Roentgenol. 2017;209(5):1088-92.

11. Viry A, Raja A, Kirkbride T, Choi V, Stamp L, Dalbeth N, et al. Multi-energy spectral photon-counting CT in crystal-related arthropathies: initial experience and diagnostic performance in vitro. Proc SPIE. 2018;10573:https://doi.org/10.1117/12.2293458.

12. Anderson N, Butler A. Clinical applications of spectral molecular imaging: potential and challenges. Contrast Media Mol Imaging. 2014;9(1):3-12.

13. Dalbeth N, Nicolaou S, Baumgartner S, Hu J, Fung M, Choi H. Presence of monosodium urate crystal deposition by dual-energy CT in patients with gout treated with allopurinol. Ann Rheum Dis. 2018;77(3):364-70.

14. Gras P, Rey C, Marsan O, Sarda S, Combes C. Synthesis and Characterisation of Hydrated Calcium Pyrophosphate Phases of Biological Interest. Eur J Inorg Chem. 2013:5886–95.

15. Lee YH, Song GG. Diagnostic accuracy of dual-energy computed tomography in patients with gout: A meta-analysis. Semin Arthritis Rheum. 2017;47(1):95-101.

16. Mallinson P, Coupal T, McLaughlin P, Nicolaou S, Munk P, Ouellette H. Dual-Energy CT for the Musculoskeletal System. Radiology. 2016;281(3):690-707.

17. Nature. Whales, scales and moons — July's best science images. Nature. 2018;https://www.nature.com/articles/d41586-018-05858-8.

Figure legends

Figure 1. A) An excised gouty finger with a core (arrow) revealing subcutaneous tophus; B) plain X-ray showing bone erosions with overhanging edges and soft tissue changes consistent with tophus; C) DECT and D) multi-energy SPCCT images both depicting MSU crystal deposits, SPCCT detecting finer detail and higher MSU volume; and E) polarised light microscopy of a sample obtained from the core confirming the presence of negatively birefringent MSU crystals.

Figure 2. A) Preoperative plain X-ray of the right knee revealing chondrocalcinosis; B) an excised calcified lateral meniscus from an osteoarthritic knee; C) plain X-ray of the excised meniscus confirming calcifications to be within the meniscus; D) composite colour-coded multi-energy SPCCT images showing a predominance of CPP crystal deposits in the calcified regions of the excised meniscus (blue – water like, red to yellow – CPP like, and white – HA like; quantitative contrast scale on left), with E) CPP and F) HA material decomposition maps; G) polarised light microscopy and H) XRD of the crystal scrappings confirming the presence of weakly positively birefringent CPP crystals with a diffraction pattern assigned as a mixture of monoclinic and triclinic CPP.

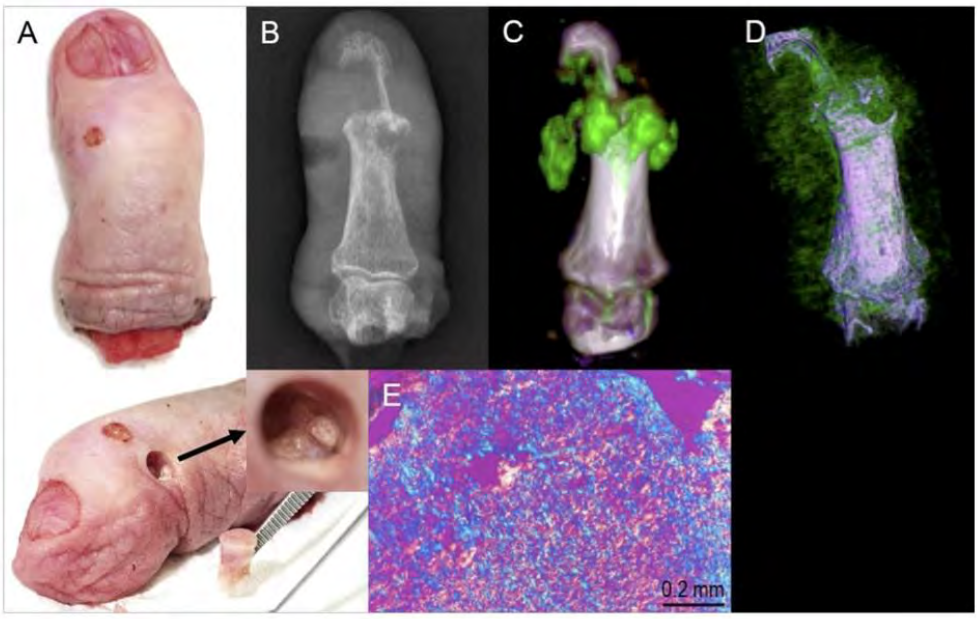
Supplementary Figure 1. Multi-energy SPCCT images of polymethyl methacrylate holder and calibration phantoms showing changes in X-ray attenuation, and thus material differentiation potential, as function of the energy bin (20-30 keV, top left; 30-40 keV, top right; 40-50 keV, bottom left; and 50-80 keV, bottom right) for various materials including synthetic crystal suspensions with known concentrations of MSU (240 mg/cm^3, not shown), CPP (88 mg/cm^3, orange) and HA (72 mg/cm^3, red). Pixels are expressed as linear attenuation coefficients (cm^{-1}) with quantitative contrast scale on right.

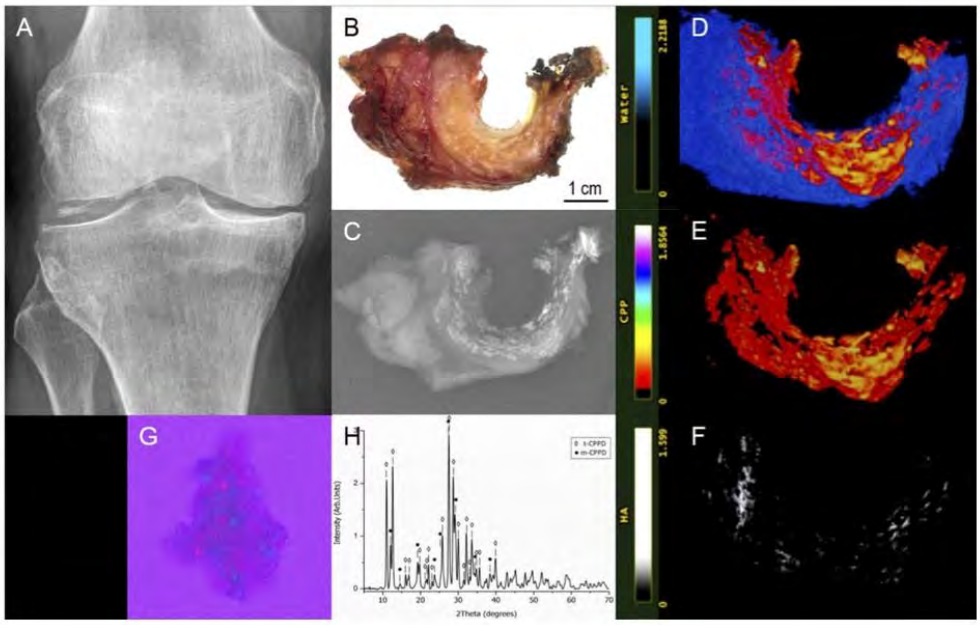
Table 1

Diagnostic accuracy measures for the differentiation of synthetic CPP crystal suspensions from HA

Energy bin (keV)	Sensitivity (%) [95% CI]	Specificity (%) [95% CI]	PPV (%) [95% CI]	NPV (%) [95% CI]	AUC
20-30	67.3 [59.4, 75.3]	78.0 [67.3, 88.8]	75.8 [68.2, 83.3]	70.6 [67.5, 73.6]	0.78 [0.77, 0.79]
30-40	66.5 [57.8, 75.3]	76.1 [64.3, 87.9]	74.0 [66.3, 81.6]	69.6 [66.4, 72.7]	0.76 [0.75, 0.77]
40-50	61.3 [52.2, 70.4]	73.7 [61.3, 86.0]	70.3 [62.7, 77.9]	65.6 [62.8, 68.4]	0.71 [0.69, 0.72]
50-80	61.0 [52.4, 69.5]	70.6 [59.4, 81.9]	67.8 [61.7, 73.9]	64.5 [61.9, 67.0]	0.68 [0.67, 0.69]

AUC = area under the receiver operating characteristic curve, CI = confidence interval, CPP = calcium pyrophosphate, HA = hydroxyapatite, NPV = negative predictive value, PPV = positive predictive value





6. Discussion and perspectives

The thesis deals with the characterisation of image quality in CT to assess the relevant diagnostic information. Task-based image quality methodologies were applied to characterise various clinical tasks for conventional CT (single-energy CT) systems, DECT systems installed in medical imaging departments, and a new prototype of SPCCT under development. The main goal was to characterise diagnostic performances for the various generations of CT.

A full characterisation of diagnostic performances for conventional CT systems would require the assessment of a large number of parameters and clinical tasks for various clinical protocols. Among the major clinical CT indications, lesion detection is the most common task that could be performed. It is also the most critical task when dealing with patient dose reduction. Which is why we chose to investigate a detection of high and low-contrast structures for different clinical protocols.

The detection of high-contrast structures in a uniform background was the task applied to characterise older and newer versions of CT devices using an abdominal protocol for the detection of renal stones. The SNR values, systematically very high, indicated that the detection task was trivial for the various CT devices. Indeed, uniform backgrounds do not take into account the influence of the complexity of the surrounding tissue when detecting lesions. To enhance discrimination between CT, one could assess a detection task in anatomical background or add complexity to the task with a characterisation of lesion shape, size, or density. Indeed, the defined clinical tasks must be adapted to the imaging protocol. For example, a discrimination task with DECT between uric acid and calcium calculi could be more clinically pertinent than the simple detection of calculi for abdominal protocols. Concerning the proposed methodology, the ideal model observer was suitable for the benchmark between different scanners. However, the task-based methodology in the frequency domain applied in this study had some limitations. The PW observer's performance is dependent on the reconstruction kernel. In particular, the observer's performance is always higher with soft kernels than sharp kernels. Hence, this observer should not be used to compare CT devices from different manufacturers because the choice of the reconstruction kernel could not be the same for all manufacturers [94]. An interesting perspective of this study could be the optimisation of the various clinical protocols for each scanner. In this case, it would be preferable to use anthropomorphic observers instead of ideal observers with appropriate anthropomorphic phantoms.

The detection of low-contrast structures was studied in two articles, for the characterisation of older and newer CT versions, and the impact of various IR algorithms compared to FBP, in both cases for an abdominal protocol. The detection task in a uniform background was pertinent to assess abdominal protocols (especially for the detection of focal liver lesions) even if it is a very simple task compared to clinical reality. To go a step further, we could work with anatomical backgrounds. Indeed, assessment of detectability of low contrast lesion in anatomical background should be an essential step for task-based image quality assessment.[95, 96]

The first low-contrast study dealt with a benchmark between different CTs. The ideal model observer was suitable for this purpose. Results showed that no major differences were observed between CTs at a high dose level for an abdominal protocol. However, differences were detected at a lower dose. However, a larger phantom size should be used to achieve a better discrimination at 15mGy because the small abdominal

phantom size cannot mimic the body habitus of a "standard patient" (around 75 Kg) used to establish DRL values. This study showed the limitation of the DRL concept to optimise clinical protocols. Indeed, different generations of CTs cannot achieve a similar dose level without impairing the low-contrast detectability. To overcome this limitation, a new concept could link a DRL value to an "image quality reference level" for different clinical indications. This methodology would ensure that each CT could correctly perform a diagnosis while keeping the dose level as low as achievable with respect to the technology and software solutions implemented in the respective devices.

The second low-contrast study dealt with a comparison between different IR algorithms. As abdominal protocols generally use ATCM to reduce patient exposure, the assessment of low-contrast detectability required the use of different phantom sizes. It appeared that low-contrast detectability decreased when increasing the phantom size, although the dose level also increased. This means that in an optimisation process it would be necessary to study image quality for different body habitus using clinical ATCM settings. Concerning the impact of IR algorithms, the results showed that the different generations of IR algorithms did not significantly improve the low-contrast detectability, except the full IR for the larger phantom size. Hence, in order to ensure a constant low-contrast detectability the possibility of dose reduction remains limited when switching from FBP to IR algorithms in abdominal protocols. An interesting perspective for this study will be to plot an optimisation curve (low-contrast detectability as a function of the dose level) for each CT and phantom size. By setting diagnostic requirements for different patients' body habitus, it would be possible to define ATCM settings for this task and obtain an optimal dose level. However, the phantom also presents some limitations that should be taken into account when dealing with an optimisation process. Firstly, the additional rings are soft-tissue equivalent that could not mimic large patients with fat around their organs. Secondly, the contrast of the various lesions was created using plastic materials of low Z_{eff} number. These materials cannot simulate the contrast of the lesion with an injected protocol. Therefore, this abdominal phantom cannot be used to assess the automatic tube voltage setting proposed by some manufacturers. A selection of tube voltage could change the task-based results obtained in this study with ATCM and IR algorithms [97]. Concerning the task-based methodology, the anthropomorphic model observers in the image domain was suitable for this study and adapted to the abdominal phantom. Moreover, results were successfully compared to another study using model observers in the frequency domain[69].

As shown in the previous studies, model observers are useful for assessing diagnostic accuracy for detection tasks in CT and thus replace the CNR which is limited when pixel size, signal size, or texture noise change. Although model observers show a high performance, their use in a clinical routine can be complex. Indeed, many technical parameters must be chosen which requires a phantom, human resources, and acquisition time. As already mentioned, different model observers (ideal or anthropomorphic) exist and they can be computed in the image or frequency domain. The computation in the image domain enables a direct comparison with human observers but requires a high number of images to train the model and test it (100 acquisitions for the CHO with 10 channels if the phantom contains only one lesion with a given size and contrast). The choice of the number of images for the two steps influences the results and a correct estimation of uncertainties is difficult. In this context, an effort should be made to make uniform the practice of model observers and provide reference results for a given data set of CT images[98].

Even if efforts are currently focused on reducing dose levels for conventional CT with IR techniques, DECT devices are emerging and the number of clinical applications increasing. DECT can offer new diagnostic information to identify, quantify, and discriminate between materials using different post-processing algorithms. As for conventional CT, the assessment of image quality linked to diagnostic performances for DECT systems is essential. The first task is to ensure the accuracy of the quantitative results obtained with various post-processing algorithms. Indeed, it is essential to ensure a good quantification in DECT since quantitative results directly help the diagnosis and could change the management of diseases.

The first quantification study dealt with the accuracy of Z_{eff} map. Results showed a good mean accuracy for the various materials studied: adipose, muscle, water, trabecular bone, and dense bone. The relative error was large for low Zeff materials (adipose, water and muscle) and so some precautions should be taken in order to discriminate materials using their Z_{eff} values. These simple measurements should be performed for annual quality control of DECT systems at different dose levels, for various kernels and reconstruction algorithms in order to ensure the accuracy of the quantitative Z_{eff} measurements for various protocols. To go a step further, specific materials involved in the different pathologies should be used to assess clinical protocols (iron, iodine, calcium crystals, uric acid …). In this case, commercial phantom inserts do not exist, and calibration samples have to be prepared manually with a high accuracy. Collaborating with chemists is therefore often necessary.

The second quantification study dealt with iodine quantification for specific iodine maps. Results showed a relative good concordance between known iodine concentration values and measured concentrations. However, uncertainties on iodine quantification was large, especially for low dose protocols due to the higher noise level. Therefore, radiologists should take into account this point when using this information. A fixed iodine density threshold should not be set to determine whether contrast enhancement is present and characteristic of a malignant lesion. This practice could lead to misdiagnosis situations, especially for low-dose protocols. Instead, two thresholds should be set to define a zone of uncertainty where it is not possible to make a correct diagnosis without other information. The choice of the two thresholds should take into account the uncertainty of the iodine quantification assessed from phantom studies. An interesting perspective of the study could reproduce the experiment with various phantom sizes in order to mimic iodine quantification for different patient sizes. Indeed, faced with increasing X-ray scattering for larger phantom sizes, the accurate quantification of iodine could be more difficult. The optimisation of DECT protocols should thus be based on an appropriate quantification task (for example the iodine quantification for abdominal protocols) and should take various patient morphologies into account.

In parallel to the emerging applications of DECT systems, the development of spectral CT also requires the assessment of image quality to determine new diagnostic performances. For the various clinical applications, pertinent clinical tasks should be defined with radiologists and referring practitioners. Among the new applications under development, the common tasks are the discrimination and the quantification of different materials. Therefore, task-based image quality metrics should be applied for these specific tasks. Methodologies are based on previous works that deal with classification and estimation tasks in CT. The potential of SPCCT devices should be firstly assessed in-vitro, based on phantom measurements. For this specific purpose, it is necessary to develop calibration samples for different material of interests. As SPCCT

are still under development, commercial solutions are not always proposed. In this case, a collaboration with chemists is primordial in order to prepare calibration samples with a high accuracy. The second step is to ex-vivo validate the primary results with human samples and later, in-vivo with human clinical trials.

In the thesis, a discrimination task between different crystals was proposed to assess the SPCCT performances for the diagnosis of crystal-related arthropathies. The study was firstly conducted in vitro, preparing crystal suspension samples, and later ex-vivo, by scanning human specimens. Firstly, in-vitro results showed that MSU can be accurately differentiated from HA and CPP. These results were in line with previous findings of patient studies using a DECT technique [89]. In-vitro results also revealed that CPP could be moderately discriminated from HA with an accuracy equal to 70%. Currently, no existing X-ray imaging technique could reproduce this potential. If this result is clinically validated with patient studies, it could open up the diagnosis and management of different crystal-related arthropathies, not only tophaceous gout.

It is important to note that SPCCT devices are continuously under development: improving the number of detectors, their configuration, the speed of image reconstruction, and the post-processing algorithms. It means that protocols have to be regularly adapted in order to take into account the different technological developments and the various limitations of PCD (pile up effect, charge sharing, K-escape X-ray and Compton scattering). This is a major difference compared to commercial CT devices installed in medical imaging departments, where clinical protocols are well-defined for different indications. With SPCCT devices, the selection of energy thresholds is particularly important because it directly influences the discrimination performances. Contrary to the majority of the current focus of development in SPCCT, in this study the choice of energy threshold is not based on the K-edge because the K-edge of calcium is too low (4 keV). Indeed, the energy selection comes directly from the X-ray attenuation law through various materials, the predominance of photoelectric effect improves the differentiation between the various crystals for low energy photons. This means that the lower the energy range, the better the discrimination. For this study, the choice of the energy thresholds was determined in order to avoid the risk of pulse pile up and saturation of the detector. A better tuning of the number of energy thresholds and the size of energy ranges could thus improve the diagnostic performances. The choice should be focused on low energy ranges while maintaining a sufficient number of photons to avoid too noisy images, which could decrease discrimination performances. Moreover, the filtration should also be optimised to eliminate too low energy photons that will not contribute to the image but only to the patient exposure. Therefore, a radiation exposure study should also be conducted for human clinical trials when optimising clinical protocols. For this purpose, it is essential to find an objective methodology to quantify the discrimination task with a multi-energy SPCCT. The ROC paradigm was proposed to directly assess the discrimination between pixels for different materials. The major limitation of this methodology was that the discrimination performance was assessed for each energy bin. Therefore, a manual selection of the energy bin that maximize the discrimination is required. To overcome this limitation, the LDA methodology was proposed to find the best linear combination between energy bins to maximize the discrimination. Developed in the thesis to discriminate between four materials, the methodology could be easily adapted for discriminating between two materials. An interesting perspective of this study will be to apply the LDA methodology to optimise the clinical protocol, selecting the optimal tube voltage or selecting energy thresholds to maximize

the discrimination between crystals[99]. Based on a previous study, optimising the energy thresholds has a higher impact on the discrimination between calcium crystals than optimising the tube potential [61].

The major limitation of the in-vitro study dealt with the preparation of crystal suspension samples with agar. Firstly, sample stability is not guaranteed over a long period. Secondly, even if results showed a good linear relationship between linear attenuation coefficients and concentration of the various materials, the uncertainty on the concentration value was unknown. In this context, it is necessary to validate the in-vitro diagnostic performances with new calibration samples prepared by a phantom manufacturer that could ensure and furnish the accuracy and the precision of the concentration value.

The ex-vivo study determines whether these promising preliminary results can be translated to excised human specimens. The SPCCT acquisition of the finger with a gouty tophus revealed the presence of MSU deposition and correctly discriminated MSU with HA crystals presents in bone and cartilage. The acquisition of the excised meniscus revealed a majority of CPP crystals in the calcified regions, which was confirmed by polarized light microscopy and X-ray diffraction. To go a step further and confirm the ex-vivo results, other excised samples should be scanned. For example, an acquisition of the tibial plateau could confirm the good discrimination between HA and CPP since it should contain a majority of HA crystals. Following this preliminary study with only one excised human specimen, various calcified medial meniscus excised at the time of joint replacement surgeries will be scanned with an optimised SPCCT protocol. In this way, ex-vivo results of sensitivity and specificity for the discrimination between calcium crystals could be obtained.

These two studies showed that the development of SPCCT devices imposes a multi-disciplinary collaboration between radiologists, referring practitioners, surgeons, physicist, medical physicists, chemists, and phantom manufacturers. Today, even if the introduction of SPCCT devices in the clinical routine is slowed down by technical challenges, these initial in-vitro and ex-vivo studies have shown promising results for the diagnosis of crystal-related arthropathies.

7. Conclusion

The various studies developed in this thesis have shown the importance of characterising image quality to assess diagnostic performances for the various CT generations (conventional CT, DECT, and SPCCT). A first step should involve radiologists and medical physicists to identify pertinent clinical tasks. Then, a task-based image quality assessment should be conducted to fully characterise the task and quantify the performances.

Currently, the optimisation of clinical protocols is still mainly focused on reducing radiation exposure. However, the amount of diagnostic information required should be considered at the first place in the optimisation process instead of just the dose aspects. The remaining challenge is to define the diagnostic requirements for specific tasks and clinical protocols. Moreover, optimising CT protocols also involves optimising many clinical parameters that can change the patient's course in a radiology department (acquisition time, reconstruction time, quantity of contrast media injection, number of phases …). Hence, the process should involve a multi-disciplinary team in the hospital (referring practitioners, radiologists, medical physicists, and radiographers), while maintaining an effective collaboration with CT manufacturers. The different roles and responsibilities should be carefully defined to provide image quality and diagnostic reliability with minimal exposure to patients. The collaboration between each member of the profession is summarized in Figure 37.

This thesis focused on characterising image quality for the different CT systems installed in hospitals or under development. However, it is important to note that equivalent research could be performed for other imaging modalities. This could lead to optimising injected activities in nuclear medicine, patient and personal radiation exposure in interventional radiology or acquisition and reconstruction time in MRI using adapted task-based methodologies to assess the diagnostic performances. The role of the medical physicist thus evolves from checking the conformity of medical units, to being closely involved with the medical team. This ensures that the necessary diagnostic information is contained in the image.

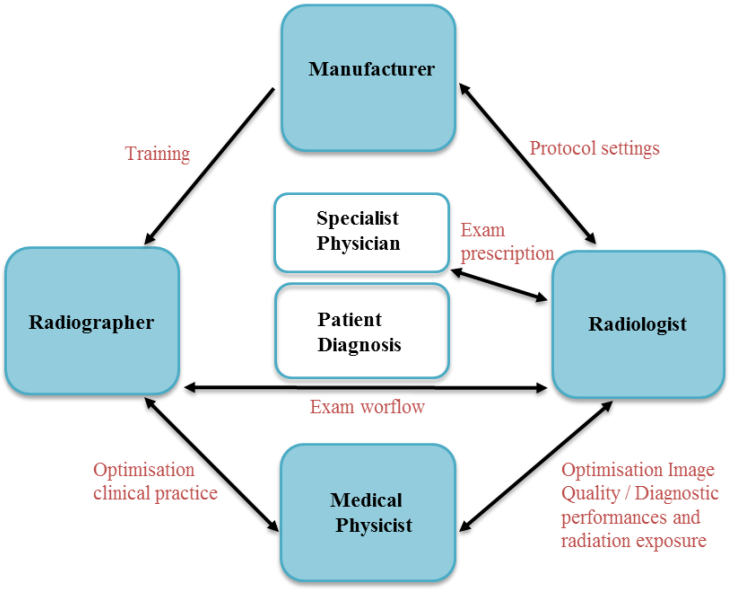


Figure 37: *CT optimisation to improve patient diagnosis, an efficient collaboration between different members.*

8. References

1. Brenner, D. J. & Hall, E. J. Computed Tomography — An Increasing Source of Radiation Exposure. *N. Engl. J. Med.* **357**, 2277–2284 (2007).
2. European Commission. Medical radiation exposure of the European population. (2015).
3. Le Coultre, R. *et al.* EXPOSURE OF THE SWISS POPULATION BY RADIODIAGNOSTICS: 2013 REVIEW. *Radiat. Prot. Dosimetry* **169**, 221–224 (2016).
4. European Council Directive. 2013/59/Euratom on basic safety standards for protection against the dangers arising from exposure to ionising radiation and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom. *OJ of the EU* (2014).
5. European Society of Radiology (ESR). Summary of the European Directive 2013/59/Euratom: essentials for health professionals in radiology. *Insights Imaging* **6**, 411–417 (2015).
6. ICRP. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP Publication 103. *Ann. ICRP 37* (2007).
7. ICRP. Radiological Protection in Medicine. ICRP Publication 105. *Ann. ICRP 37*
8. Sobiecka, A., Bekiesińska-Figatowska, M., Rutkowska, M., Latos, T. & Walecki, J. Clinically Unjustified Diagnostic Imaging – a Worrisome Tendency in Today's Medical Practice. *Pol. J. Radiol.* **81**, 325–330 (2016).
9. Lehnert, B. E. & Bree, R. L. Analysis of Appropriateness of Outpatient CT and MRI Referred From Primary Care Clinics at an Academic Medical Center: How Critical Is the Need for Improved Decision Support? *J. Am. Coll. Radiol.* **7**, 192–197 (2010).
10. Lambert, L. *et al.* Growing number of emergency cranial CTs in patients with head injury not justified by their clinical need. *Wien. Klin. Wochenschr.* **129**, 159–163 (2017).
11. Zargar Balaye Jame, S. *et al.* Indications and Overuse of Computed Tomography in Minor Head Trauma. *Iran. Red Crescent Med. J.* **16**, (2014).
12. Société Française de radiologie & Société Française de Médecine Nucléaire. Guide du Bon Usage des examens d'imagerie médicale. (2013).
13. The Royal College of Radiologists (RCR). IRefer essential radiological investigation guidelines tool.
14. ICRP. Radiological Protection and Safety in Medicine. ICRP Publication 73. *Ann. ICRP 26 (2)* (1996).
15. Office fédéral de la santé publique OFSP. Directive R-06-06 Niveaux de référence diagnostiques en tomodensitométrie. (2018).
16. McCollough, C. H., Bruesewitz, M. R. & Kofler, J. M. CT Dose Reduction and Dose Management Tools: Overview of Available Options. *RadioGraphics* **26**, 503–512 (2006).
17. Kalender, W. A., Wolf, H. & Suess, C. Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements. *Med. Phys.* **26**, 2248–2253 (1999).
18. Rizzo, S. *et al.* Comparison of Angular and Combined Automatic Tube Current Modulation Techniques with Constant Tube Current CT of the Abdomen and Pelvis. *Am. J. Roentgenol.* **186**, 673–679 (2006).
19. Söderberg, M. & Gunnarsson, M. Automatic exposure control in computed tomography – an evaluation of systems from different manufacturers. *Acta Radiol.* **51**, 625–634 (2010).
20. Gilbert, P. Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.* **36**, 105–117 (1972).
21. Thibault, J.-B., Sauer, K. D., Bouman, C. A. & Hsieh, J. A three-dimensional statistical approach to improved image quality for multislice helical CT: Improved statistical reconstruction for multislice helical CT. *Med. Phys.* **34**, 4526–4544 (2007).
22. Beister, M., Kolditz, D. & Kalender, W. A. Iterative reconstruction methods in X-ray CT. *Phys. Med.* **28**, 94–108 (2012).
23. Ehman, E. C. *et al.* Methods for Clinical Evaluation of Noise Reduction Techniques in Abdominopelvic CT. *RadioGraphics* **34**, 849–862 (2014).
24. Patino, M., Fuentes, J. M., Singh, S., Hahn, P. F. & Sahani, D. V. Iterative Reconstruction Techniques in Abdominopelvic CT: Technical Concepts and Clinical Implementation. *Am. J. Roentgenol.* **205**, W19–W31 (2015).
25. Geyer, L. L. *et al.* State of the Art: Iterative CT Reconstruction Techniques. *Radiology* **276**, 339–357 (2015).
26. Ott, J. G. *et al.* Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. *Phys. Med. Biol.* **59**, 4047–4064 (2014).

27. Omotayo, A. & Elbakri, I. Objective performance assessment of five computed tomography iterative reconstruction algorithms. *J. X-Ray Sci. Technol.* **24**, 913–930 (2016).
28. Willemink, M. J. *et al.* Iterative reconstruction techniques for computed tomography Part 1: Technical principles. *Eur. Radiol.* **23**, 1623–1631 (2013).
29. McCollough, C. H., Leng, S., Yu, L. & Fletcher, J. G. Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications. *Radiology* **276**, 637–653 (2015).
30. Hounsfield, G. N. Computerized transverse axial scanning (tomography): Part 1. Description of system. *Br. J. Radiol.* **46**, 1016–1022 (1973).
31. Genant, H. K. & Boyd, D. Quantitative bone mineral analysis using dual energy computed tomography. *Invest. Radiol.* **12**, 545–551 (1977).
32. Omoumi, P. *et al.* Dual-Energy CT: Basic Principles, Technical Approaches, and Applications in Musculoskeletal Imaging (Part 1). *Semin. Musculoskelet. Radiol.* **19**, 431–437 (2015).
33. Goo, H. W. & Goo, J. M. Dual-Energy CT: New Horizon in Medical Imaging. *Korean J. Radiol.* **18**, 555 (2017).
34. Lestra, T. *et al.* Applications of dual energy computed tomography in abdominal imaging. *Diagn. Interv. Imaging* **97**, 593–603 (2016).
35. Lu, G.-M., Wu, S.-Y., Yeh, B. M. & Zhang, L.-J. Dual-energy computed tomography in pulmonary embolism. *Br. J. Radiol.* **83**, 707–718 (2010).
36. Lu, G. M., Zhao, Y., Zhang, L. J. & Schoepf, U. J. Dual-Energy CT of the Lung. *Am. J. Roentgenol.* **199**, S40–S53 (2012).
37. Omoumi, P., Verdun, F., Guggenberger, R., Andreisek, G. & Becce, F. Dual-Energy CT: Basic Principles, Technical Approaches, and Applications in Musculoskeletal Imaging (Part 2). *Semin. Musculoskelet. Radiol.* **19**, 438–445 (2015).
38. Alvarez, R. E. & Macovski, A. Energy-selective reconstructions in X-ray computerized tomography. *Phys. Med. Biol.* **21**, 733–744 (1976).
39. Yveborg, M., Danielsson, M. & Bornefalk, H. Theoretical Comparison of a Dual Energy System and Photon Counting Silicon Detector Used for Material Quantification in Spectral CT. *IEEE Trans. Med. Imaging* **34**, 796–806 (2015).
40. Taguchi, K. & Iwanczyk, J. S. Vision 20/20: Single photon counting x-ray detectors in medical imaging. *Med. Phys.* **40**, 100901 (2013).
41. Ballabriga, R. *et al.* Review of hybrid pixel detector readout ASICs for spectroscopic X-ray imaging. *J. Instrum.* **11**, P01007–P01007 (2016).
42. Koenig, T. *et al.* Imaging properties of small-pixel spectroscopic x-ray detectors based on cadmium telluride sensors. *Phys. Med. Biol.* **57**, 6743–6759 (2012).
43. Hsieh, S. S., Rajbhandary, P. L. & Pelc, N. J. Spectral resolution and high-flux capability tradeoffs in CdTe detectors for clinical CT. *Med. Phys.* **45**, 1433–1443 (2018).
44. Liu, J. & Gao, H. Material Reconstruction for Spectral Computed Tomography with Detector Response Function. *Med. Phys.* **43**, 3834–3834 (2016).
45. Srivastava, S., Cammin, J., Fung, G. S. K., Tsui, B. M. W. & Taguchi, K. Spectral response compensation for photon-counting clinical x-ray CT using sinogram restoration. in (eds. Pelc, N. J., Nishikawa, R. M. & Whiting, B. R.) 831311 (2012). doi:10.1117/12.911394
46. Iwanczyk, J. S. *et al.* Photon Counting Energy Dispersive Detector Arrays for X-ray Imaging. *IEEE Trans. Nucl. Sci.* **56**, 535–542 (2009).
47. Yu, Z. *et al.* Evaluation of conventional imaging performance in a research whole-body CT system with a photon-counting detector array. *Phys. Med. Biol.* **61**, 1572–1595 (2016).
48. Dangelmaier, J. *et al.* Experimental feasibility of spectral photon-counting computed tomography with two contrast agents for the detection of endoleaks following endovascular aortic repair. *Eur. Radiol.* **28**, 3318–3325 (2018).
49. Bennett, J. R. *et al.* Hybrid Spectral Micro-CT: System Design, Implementation, and Preliminary Results. *IEEE Trans. Biomed. Eng.* **61**, 246–253 (2014).
50. Walsh, M. F. *et al.* Spectral CT data acquisition with Medipix3.1. *J. Instrum.* **8**, P10012–P10012 (2013).
51. Walsh, M. F. *et al.* First CT using Medipix3 and the MARS-CT-3 spectral scanner. *J. Instrum.* **6**, C01095–C01095 (2011).
52. Gaind, N. Whales, scales and moons — July's best science images. *Nature* (2018). doi:10.1038/d41586-018-05858-8

53. Schirra, C. O., Brendel, B., Anastasio, M. A. & Roessl, E. Spectral CT: a technology primer for contrast agent development: SPECTRAL CT: A TECHNOLOGY PRIMER FOR CONTRAST AGENT DEVELOPMENT. *Contrast Media Mol. Imaging* **9**, 62–70 (2014).
54. Moghiseh, M. *et al.* Discrimination of multiple high-Z materials by multi-energy spectral CT–A phantom study. *JSM Biomed Imaging Data Pap* **61**, 1007 (2016).
55. Symons, R. *et al.* Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: An *in vivo* study. *Med. Phys.* **44**, 5120–5127 (2017).
56. Anderson, N. G. & Butler, A. P. Clinical applications of spectral molecular imaging: potential and challenges: SPECTRAL MOLECULAR IMAGING - CLINICAL APPLICATIONS. *Contrast Media Mol. Imaging* **9**, 3–12 (2014).
57. Danad, I., Fayad, Z. A., Willemink, M. J. & Min, J. K. New Applications of Cardiac Computed Tomography. *JACC Cardiovasc. Imaging* **8**, 710–723 (2015).
58. Zainon, R. *et al.* Spectral CT of carotid atherosclerotic plaque: comparison with histology. *Eur. Radiol.* **22**, 2581–2588 (2012).
59. Moghiseh, M. *et al.* Spectral Photon-Counting Molecular Imaging for Quantification of Monoclonal Antibody-Conjugated Gold Nanoparticles Targeted to Lymphoma and Breast Cancer: An *In Vitro* Study. *Contrast Media Mol. Imaging* **2018**, 1–9 (2018).
60. Aamir, R. *et al.* MARS spectral molecular imaging of lamb tissue: data collection and image analysis. *J. Instrum.* **9**, P02005–P02005 (2014).
61. Kirkbride, T. E. *et al.* Discrimination Between Calcium Hydroxyapatite and Calcium Oxalate Using Multienergy Spectral Photon-Counting CT. *Am. J. Roentgenol.* **209**, 1088–1092 (2017).
62. Fryback, D. G. & Thornbury, J. R. The Efficacy of Diagnostic Imaging. *Med. Decis. Making* **11**, 88–94 (1991).
63. Nute, J. L., Jacobsen, M. C., Stefan, W., Wei, W. & Cody, D. D. Development of a dual-energy computed tomography quality control program: Characterization of scanner response and definition of relevant parameters for a fast-kVp switching dual-energy computed tomography system. *Med. Phys.* **45**, 1444–1458 (2018).
64. Burgess, A. E. The Rose model, revisited. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **16**, 633–646 (1999).
65. Solomon, J., Marin, D., Roy Choudhury, K., Patel, B. & Samei, E. Effect of Radiation Dose Reduction and Reconstruction Algorithm on Image Noise, Contrast, Resolution, and Detectability of Subtle Hypoattenuating Liver Lesions at Multidetector CT: Filtered Back Projection versus a Commercial Model–based Iterative Reconstruction Algorithm. *Radiology* **284**, 777–787 (2017).
66. Schindera, S. T. *et al.* Iterative reconstruction algorithm for CT: can radiation dose be decreased while low-contrast detectability is preserved? *Radiology* **269**, 511–518 (2013).
67. Barrett, H. H. & Myers, K. J. *Foundations of image science*. **1**, (Wiley-Interscience, 2004).
68. Barrett, H. H., Myers, K. J., Hoeschen, C., Kupinski, M. A. & Little, M. P. Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys. Med. Biol.* **60**, R1 (2015).
69. Samei, E. & Richard, S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology: CT task-based performance metrology. *Med. Phys.* **42**, 314–323 (2014).
70. Racine, D., Ba, A. H., Ott, J. G., Bochud, F. O. & Verdun, F. R. Objective assessment of low contrast detectability in computed tomography with Channelized Hotelling Observer. *Phys. Medica Eur. J. Med. Phys.* **32**, 76–83 (2016).
71. Euler, A. *et al.* Impact of model-based iterative reconstruction on low-contrast lesion detection and image quality in abdominal CT: a 12-reader-based comparative phantom study with filtered back projection at different tube voltages. *Eur. Radiol.* **27**, 5252–5259 (2017).
72. Zhang, Y., Leng, S., Yu, L., Carter, R. E. & McCollough, C. H. Correlation between human and model observer performance for discrimination task in CT. *Phys. Med. Biol.* **59**, 3389–3404 (2014).
73. Hernandez-Giron, I., Calzado, A., Geleijns, J., Joemai, R. M. S. & Veldkamp, W. J. H. Comparison between human and model observer performance in low-contrast detection tasks in CT images: application to images reconstructed with filtered back projection and iterative algorithms. *Br. J. Radiol.* **87**, 20140014 (2014).
74. Vaishnav, J. Y., Jung, W. C., Popescu, L. M., Zeng, R. & Myers, K. J. Objective assessment of image quality and dose reduction in CT iterative reconstruction: Validating dose reduction claims for CT iterative reconstruction. *Med. Phys.* **41**, 071904 (2014).
75. Viry, A. *et al.* Characterization of a CT unit for the detection of low contrast structures. in (eds. Kupinski, M. A. & Nishikawa, R. M.) 101361C (2017). doi:10.1117/12.2250529

76. Barrett, H. H., Yao, J., Rolland, J. P. & Myers, K. J. Model observers for assessment of image quality. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9758–9765 (1993).
77. Abbey, C. K. Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *JOSA A* **18**, 473–488 (2001).
78. Myers, K. J. & Barrett, H. H. Addition of a channel mechanism to the ideal-observer model. *JOSA A* **4**, 2447–2457 (1987).
79. Gallas, B. D. & Barrett, H. H. Validating the use of channels to estimate the ideal linear observer. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1725–1738 (2003).
80. Boll, D. T. *et al.* Renal Stone Assessment with Dual-Energy Multidetector CT and Advanced Postprocessing Techniques: Improved Characterization of Renal Stone Composition—Pilot Study. *Radiology* **250**, 813–820 (2009).
81. Hidas, G. *et al.* Determination of Renal Stone Composition with Dual-Energy CT: In Vivo Analysis and Comparison with X-ray Diffraction. *Radiology* **257**, 394–401 (2010).
82. Shinohara, Y. *et al.* Assessment of carotid plaque composition using fast-kV switching dual-energy CT with gemstone detector: comparison with extracorporeal and virtual histology-intravascular ultrasound. *Neuroradiology* **57**, 889–895 (2015).
83. Manohara, S. R., Hanagodimath, S. M., Thind, K. S. & Gerward, L. On the effective atomic number and electron density: A comprehensive set of formulas for all types of materials and energies above 1keV. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* **266**, 3906–3912 (2008).
84. Murty, R. C. Effective Atomic Numbers of Heterogeneous Materials. *Nature* **207**, 398–399 (1965).
85. Mileto, A. *et al.* Accuracy of Contrast-Enhanced Dual-Energy MDCT for the Assessment of Iodine Uptake in Renal Lesions. *Am. J. Roentgenol.* **202**, W466–W474 (2014).
86. McQueen, F. M., Doyle, A. & Dalbeth, N. Imaging in the Crystal Arthropathies. *Rheum. Dis. Clin. N. Am.* **40**, 231–249 (2014).
87. Neogi, T. *et al.* 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann. Rheum. Dis.* **74**, 1789–1798 (2015).
88. Choi, H. K. *et al.* Dual energy CT in gout: a prospective validation study. *Ann. Rheum. Dis.* **71**, 1466–1471 (2012).
89. Bongartz, T. *et al.* Dual-energy CT for the diagnosis of gout: an accuracy and diagnostic yield study. *Ann. Rheum. Dis.* **74**, 1072–1077 (2015).
90. Berendsen, D., Neogi, T., Taylor, W. J., Dalbeth, N. & Jansen, T. L. Crystal identification of synovial fluid aspiration by polarized light microscopy. An online test suggesting that our traditional rheumatologic competence needs renewed attention and training. *Clin. Rheumatol.* **36**, 641–647 (2017).
91. Racine, D. *et al.* Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners. *Med. Phys.* **44**, e153–e163 (2017).
92. Viry, A. *et al.* Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study. *Phys. Med.* **48**, 111–118 (2018).
93. Stamp, L. K. *et al.* Clinical utility of multi-energy spectral photon-counting CT in crystal arthritis. *Arthritis Rheumatol.* (2019). doi:10.1002/art.40848
94. Solomon, J. B., Christianson, O. & Samei, E. Quantitative comparison of noise texture across CT scanners from different manufacturers. *Med. Phys.* **39**, 6048–6055 (2012).
95. Gong, H. *et al.* Correlation between model observers in uniform background and human observer in patient liver background for a low-contrast detection task in CT. in *Medical Imaging 2018: Image Perception, Observer Performance, and Technology Assessment* (eds. Nishikawa, R. M. & Samuelson, F. W.) 21 (SPIE, 2018). doi:10.1117/12.2294955
96. Ba, A. *et al.* Low contrast detection in abdominal CT: comparing single-slice and multi-slice tasks. in (eds. Kupinski, M. A. & Nishikawa, R. M.) 101360S (2017). doi:10.1117/12.2254237
97. Seyal, A. R. *et al.* CT of the Abdomen with Reduced Tube Voltage in Adults: A Practical Approach. *RadioGraphics* **35**, 1922–1939 (2015).
98. Ba, A. *et al.* Inter-laboratory comparison of channelized hotelling observer computation. *Med. Phys.* **45**, 3019–3030 (2018).
99. Rigie, D. S. & La Rivière, P. J. Optimizing spectral CT parameters for material classification tasks. *Phys. Med. Biol.* **61**, 4599–4622 (2016).

Annex 1: Development of the optimal linear observer for a 2-class problem (Hotelling observer)

A linear observer takes the general form:

$$\lambda_k{}^{(i)} = w^T g_k{}^{(i)}$$

$\lambda_k{}^{(i)}$ is the scalar response of a vector image $g_k{}^{(i)}$ from class k through w (in a 2-class problem, k=0 or 1). We can define $\overline{\lambda_k}$ as the mean scalar responses of vector images from class k and σ_k as the standard deviation of the scalar responses for vector images from class k:

$\overline{\lambda_k} = w^T \overline{g_k}$, with $\overline{g_k}$ the mean vector images for class k.

$\sigma_k{}^2 = \sum_i (w^T g_k{}^{(i)} - w^T \overline{g_k})^2 = \sum_i w^T (g_k{}^{(i)} - \overline{g_k})(g_k{}^{(i)} - \overline{g_k})^T w = w^T \boldsymbol{K_k} w$, with $\boldsymbol{K_k}$ being the covariance matrix of class k, N the number of images in class k.

Then try to find the optimal direction of w to maximize the signal to noise ratio:

$$SNR^2(w) = \frac{2(w^T \overline{g_1} - w^T \overline{g_0})^2}{w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w} = \frac{2w^T \Delta_g \Delta_g{}^T w}{w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w}$$

Maximize SNR(w) by taking the derivative with respect to w and setting it to 0.

$$\frac{dSNR}{dw} = \frac{2w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w * 2\Delta_g \Delta_g{}^T w - 2\, w^T \Delta_g \Delta_g{}^T w * 2(\boldsymbol{K_0} + \boldsymbol{K_1})\mathrm{w}}{\left[w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w\right]^2}$$

$$\frac{dSNR}{dw} = 0 \Rightarrow 2w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w * 2(\mathrm{N}-1)\Delta_g \Delta_g{}^T w - 2\,(\mathrm{N}-1) w^T \Delta_g \Delta_g{}^T w * 2(\boldsymbol{K_0} + \boldsymbol{K_1})\mathrm{w} = 0$$

$$\Rightarrow \frac{w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w \Delta_g \Delta_g{}^T w}{w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w} - \frac{w^T \Delta_g \Delta_g{}^T w}{w^T(\boldsymbol{K_0} + \boldsymbol{K_1})w} * (\boldsymbol{K_0} + \boldsymbol{K_1})\mathrm{w} = 0$$

$$\Rightarrow \Delta_g \Delta_g{}^T w - \mathrm{SNR} * (\boldsymbol{K_0} + \boldsymbol{K_1})\mathrm{w} = 0$$

This leads to solving the generalized eigenvalue problem:

$$\Rightarrow (\boldsymbol{K_0} + \boldsymbol{K_1})^{-1} \Delta_g \Delta_g{}^T w = \mathrm{SNR} * \mathrm{w}$$

By definition $\Delta_g{}^T w$ is a constant λ. Hence, the eigenvalue problem can be written:

$$\Rightarrow (\boldsymbol{K_0} + \boldsymbol{K_1})^{-1} \Delta_g = \frac{\mathrm{SNR}}{\lambda} * \mathrm{w}$$

The optimal direction of w is:

$$\mathrm{w}_{opt} = (\boldsymbol{K_0} + \boldsymbol{K_1})^{-1} \Delta_g$$

Annex 2: Development of the linear discriminant analysis for a 2-class problem

A linear discriminant takes the general form:

$$y_k{}^{(i)} = w^T x_k{}^{(i)}$$

$y_k{}^{(i)}$ is the scalar response of data $x_k{}^{(i)}$ from class k through w (in a 2-class problem, k=0 or 1). We can define $\overline{y_k}$ as the mean scalar responses of data from class k and σ_k as the standard deviation of the scalar responses for vector images from class k:

$\overline{y_k} = w^T \overline{x_k}$, with $\overline{g_k}$ being the mean vector images for class k.

$(\overline{y_1} - \overline{y_0})^2 = [w^T(\overline{x_1} - \overline{x_0})]^2 = w^T(\overline{x_1} - \overline{x_0})^T(\overline{x_1} - \overline{x_0})w = w^T \boldsymbol{S_B} w$ with $\boldsymbol{S_B}$ being the between-class matrix

$\sigma_k{}^2 = \sum_i (w^T x_k{}^{(i)} - w^T \overline{x_k})^2 = \sum_i w^T (x_k{}^{(i)} - \overline{x_k})(x_k{}^{(i)} - \overline{x_k})^T w = w^T \boldsymbol{S_{w,k}} w$ with $\boldsymbol{S_{w,k}}$ being the within class matrix for the class k, N the number of data from class k

Then try to find the optimal direction of w to maximize the signal to noise ratio:

$$J(w) = \frac{w^T \boldsymbol{S_B} w}{w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w}$$

Maximize SNR(w) by taking the derivative with respect to w and setting it to 0.

$$\frac{dJ}{dw} = \frac{w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w * 2\boldsymbol{S_B} w - w^T \boldsymbol{S_B} w * 2(\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})\mathrm{w}}{\left[w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w\right]^2}$$

$$\frac{dSNR}{dw} = 0 \Rightarrow w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w * 2\boldsymbol{S_B} w - w^T \boldsymbol{S_B} w * 2(\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})\mathrm{w} = 0$$

$$\Rightarrow \frac{w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w * \boldsymbol{S_B} w}{w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w} - \frac{w^T \boldsymbol{S_B} w}{w^T (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}}) w} * (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})\mathrm{w} = 0$$

$$\Rightarrow \boldsymbol{S_B} w - \mathrm{J} * (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})\mathrm{w} = 0$$

This leads to solving the generalized eigenvalue problem:

$$\Rightarrow (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})^{-1} \boldsymbol{S_B} \quad w = \mathrm{J} * \mathrm{w}$$

For a two-class problem, $\boldsymbol{S_B} w$ is in the same direction than $\overline{x_1} - \overline{x_0}$, so the eigenvalue problem can be directly solved with:

$$w = (\boldsymbol{S_{w,0}} + \boldsymbol{S_{w,1}})^{-1} (\overline{x_1} - \overline{x_0})$$

Annex 3: Development of the linear discriminant analysis for an L-class problem

Reminder

The data are combined in a matrix X, composed with the data for the different features and the L different classes. The matrix size is NxM, where N is the sum of the number of data in each class and M is the number of features. The LDA will find L-1 projections yi of the data X by means of L-1 projection vectors w, arranged by columns in a projection matrix W.

$$y_i = w^T X \text{ and } Y = W^T X$$

We can define the following matrix from the data X: S_B the between class scatter matrix and S_W the within class scatter matrix.

Development

Now, we need to find W, that maximizes de Fisher criterion J(W). The Fisher criterion represents a measure of the separability between the different classes after projection into the new subspace, taking into account the distance between the mean of each classes (S_B) and the scatter of pixel values for each class (S_W). In a 2-class problem, the Fisher criterion is equivalent to the SNR.

$$J(W) = \frac{|\boldsymbol{W^T S_B W}|}{|\boldsymbol{W^T S_W W}|}$$

After derivation with respect to W (see Annex 1). We can find w (the column vectors of W) solving the following generalized eigenvalue equation:

$$\boldsymbol{S_W}^{-1}\boldsymbol{S_B} w = \lambda w$$

The equation above admits, at most, L-1 solutions ($k \subset [1\,, L-1]$). The solutions are the eigenvectors corresponding to the L-1 non-zero eigenvalues. The size of the resulting matrix Y is N x k. The LDA discriminant can be also useful to reduce the number of features.

List of publications

Peer-reviewed first author

Effects of various generations of iterative CT reconstruction algorithms on low-contrast detectability as a function of the effective abdominal diameter: A quantitative task-based phantom study.

Anais Viry, Christoph Aberle, Damien Racine, Jean-François Knebel, Sebastian T. Schindera, Sabine Schmidt, Fabio Becce &, Francis R. Verdun
Physica Medica, April 2018, doi:10.1016/j.ejmp.2018.04.006

Objective comparison of high-contrast spatial resolution and low-contrast detectability for various clinical protocols on multiple CT scanners

Damien Racine & Anais Viry, Fabio Becce, Sabine Schmidt, Alexandre Ba, François O. Bochud, Sue Edyvean, Alexander Schegerer, Francis R. Verdun
Medical Physics, September 2017, doi:10.1002/mp.12224

Peer reviewed co-author

Assessment of low-contrast detection in CT using model observers: Developing a clinically-relevant tool for characterising adaptive statistical and model-based iterative reconstruction

Julien Ott, Alexandre Ba, Damien Racine, Anais Viry, François O. Bochud, Francis R. Verdun
Zeitschrift für Medizinische Physik, June 2017, doi: 10.1016/j.zemedi.2016.04.002

Task-based quantification of image quality using a model observer in abdominal CT: a multicentre study

Damien Racine, Nick Ryckx, Alexandre Ba, Fabio Becce, Anais Viry, Francis R. Verdun, Sabine Schmidt

European Radiology, June 2018, doi:10.1007/s00330-018-5518-8

Clinical Utility of multi-energy spectral CT in crystal arthritis

Lisa K Stamp, Nigel Anderson, Fabio Becce, Maya Rajeswari, Matthew Polson, Olivier Guyen, Anais Viry, Chloe Choi, Tracy Kirkbride, Aamir Raja

Arthritis Rheumatol. (2019). doi:10.1002/art.40848

Channelized Hotelling observer correlation with human observers for low-contrast detection in liver CT images

Alexandre Ba; Craig K. Abbey; Damien Racine; Anaïs Viry; Francis R. Verdun; Sabine Schmidt; François O. Bochud
J. of Medical Imaging, 6(2), 025501 (2019). doi: 10.1117/1.JMI.6.2.025501

Proceeding conferences first author

Characterisation of a CT unit for the detection of low-contrast structures

Anais Viry, Damien Racine, Alexandre Ba, Fabio Becce, Francis R. Verdun
Proceedings Volume 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment, doi:10.1117/12.2250529

Multi-energy spectral photon-counting CT in crystal-related arthropathies: initial experience and diagnostic performance in vitro

Anais Viry, Aamir Y. Raja, Tracy E. Kirkbride, Chloe Choi, Lisa K. Stamp, Nicola Dalbeth, Christele Combes, Francis R. Verdun, Nigel G. Anderson, Fabio Becce
Proceedings Volume 10573, Medical Imaging 2018: Physics of Medical Imaging; 1057351 (2018) https://doi.org/10.1117/12.2293458

Proceeding conferences co-author

Low contrast detection in abdominal CT: comparing single-slice and multi-slice tasks

Alexandre Ba; Damien Racine; Anaïs Viry; Francis R. Verdun; Sabine Schmidt; François O. Bochud
Proceedings Volume 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment

Characterisation CT unit using a dose efficiency index concept

Damien Racine, Pascal Monin, François O. Bochud, Anais Viry, Alexander Schegerer, Sue Edyvean, Alexander Schegerer, Francis R. Verdun

Proceeding « The 4th International Conference on Image Formation in X-Ray Computed Tomography »

Under Review

In vitro diagnostic performances of a multi-energy spectral photon-counting CT for crystal-related arthropathies

Anais Viry, Aamir Y. Raja, Tracy E. Kirkbride, Chloe Choi, Lisa K. Stamp, Nicola Dalbeth, Christele Combes, Francis R. Verdun, Nigel G. Anderson, Fabio Becce
Submitted in Radiology

List of Attended conferences

Only posters and talks as a first author are indicated

Utilisation d'un modèle d'observateur pour une comparaison de protocoles cliniques en tomodensitométrie.

Anais Viry, Christoph Aberle, Francis R. Verdun
Journées scientifiques de la SFPM 2016, Nancy, 1-3 juin 2016

Characterisation of two CT units for the detection of low-contrast structures (Poster)

Anais Viry, Damien Racine, Alexandre Ba, Fabio Becce, François O Bochud, Francis R Verdun
SPIE Medical imaging, Orlando, 11-16 February 2017

Effect of five iterative reconstruction algorithms on low-contrast detectability in patients with varying abdominal diameters

Anais Viry, Damien Racine, Christoph Aberle, Sabine Schmidt, Sebastian T. Schindera, Francis R. Verdun, Fabio Becce
European Congress of Radiology, Wien, 1-5 March 2017

Impact of various iterative algorithms on low-contrast detectability in patients with varying abdominal diameters: A CT phantom

Anais Viry, Damien Racine, Christoph Aberle, Sabine Schmidt, Sebastian T. Schindera, Francis R. Verdun, Fabio Becce
Swiss Congress of Radiology, Bern, 7-9 June 2017

Multi-energy spectral photon-counting CT in crystal-related arthropathies: initial experience and diagnostic performance in vitro (Poster)

Anais Viry, Aamir Y. Raja, Tracy E. Kirkbride, Chloe Choi, Lisa K. Stamp, Nicola Dalbeth, Christele Combes, Francis R. Verdun, Nigel G. Anderson, Fabio Becce
SPIE Medical Imaging, Houston, 10-14 February 2018

Multi-energy spectral photon-counting CT in crystal-related arthropathies: Initial experience and diagnostic performance in vitro

Anais Viry, Aamir Y. Raja, Tracy E. Kirkbride, Chloe Choi, Lisa K. Stamp, Nicola Dalbeth, Christele Combes, Francis R. Verdun, Nigel G. Anderson, Fabio Becce
Swiss Congress of Radiology, Lausanne 10-12 May 2018

Impact des différentes générations d'algorithmes itératifs sur la détectabilité à bas contraste en tomodensitométrie abdominale

Anaïs Viry, Damien Racine, Fabio Becce, Francis R. Verdun
Journées scientifiques de la SFPM 2018, Toulouse, 13-15 Juin 2018

Première expérience de discrimination in vitro de cristaux impliqués dans des arthropathies microcristallines à l'aide d'un CT spectral multi-énergie

Anaïs Viry, Aamir Y. Raja, Tracy E. Kirkbride, Chloe Choi, Lisa K. Stamp, Nicola Dalbeth, Christele Combes, Francis R. Verdun, Nigel G. Anderson, Fabio Becce
Journées scientifiques de la SFPM 2018, Toulouse, 13-15 Juin 2018

Diagnosis of crystal-related arthropathies with a multi-energy spectral photon counting CT

Anaïs Viry, Aamir Raja, Lisa Stamp, Francis R. Verdun, Fabio Becce
52th SSRMP annual meeting, Lausanne, 22-23 Novembre 2018