



MONASH University

Characterisation of organs in a post-mortem computed tomography database using machine learning

Carlos Andrés Peña Solórzano

Bachelor of Engineering (Electronics)

Master of Engineering (Industrial Automation)

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2019

Department of Medical Imaging and Radiation Sciences
Faculty of Medicine, Nursing and Health Sciences

© The author (2019)

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

ABSTRACT

Medical databases can provide valuable prognostic and diagnostic information for individual patient management and also insights into population level trends. Extraction of the salient information requires sophisticated text and image analysis tools. The vast size of modern databases, in large part due to the incorporation of volumetric imaging data, makes the management and labelling of content an onerous task; this is particularly true in the forensic setting where only a small subsection of the data that is explicitly relevant to the cause of death is labelled by an expert reader. Machine learning (ML) approaches, including both supervised and unsupervised pipelines, offer a potential avenue for alleviating the level of human intervention required in order to label and therefore structure medical databases. The ability of algorithms to sufficiently identify key anatomical and pathological landmarks depends on the condition of the body, the scanner configuration, and the resulting image quality. In this thesis, the effects of a wide range of decedent-presentations on an automated labelling pipeline for analysing post-mortem computed tomography (CT) scans have been investigated. This document is organised into three investigations: development of a ML pipeline applied to forensic data; the development of a tool for generation of synthetic trabecular structures; and, the quantification of the effect of artefacts on image quality caused by dense structures in CT scans. The final chapters present a discussion and conclusion with reference to future work.

The first investigation, presented in this thesis, measured the performance characteristics of a deep learning (DL) pipeline that comprised localisation and labelling steps. Variables that affected the performance of the organ localisation technique, when applied to the femur, were studied and their effect quantified. The automatic labelling of proximal and distal femoral volumes was performed, and the resulting classification was studied using confusion matrices to compare the performance of the proposed approach when additional information about the cases was utilised during training.

In the second investigation, a simple analytical tool for modelling trabecular bone structure was developed. The model provided an accurate way of generating synthetic data for pre-training and transfer learning which ultimately will help identify degraded bone structures more efficiently, e.g. regions affected by osteoporosis.

In the third investigation, the coupling artefacts that arise due to the interplay between the cylindrical symmetry of the CT scanner and the linear organisation of multiple dense objects in the field of view, were investigated. The effect of the artefacts on the performance of the pipeline was quantified.

Finally, the numerical trabecular model and an algorithm designed to generate synthetic coupling artefacts were combined to investigate the limitations of a DL approach with application to denoising of medical images.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 2 original papers published in peer reviewed journals and 2 submitted papers under review. The core theme of the thesis is medical image processing using machine learning techniques. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Medical Imaging and Radiation Sciences under the supervision of Dr. Matthew Dimmock.

(The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.)

In the case of chapters 2 to 5, my contribution to the work involved the following:

Thesis chapter	Publication title	Status	Nature and % of student contribution	Co-author name(s), Nature and % of co-author's contribution	Co-author(s), Monash student Y/N
2	Findings from machine learning in clinical medical imaging applications – lessons for translation to the forensic setting.	Submitted	70%. Review, analyse, and draft the manuscript.	1) D. Albrecht: Proofreading 5%; 2) R. Basset: Proofreading and structure development 10%; 3) M. Burke: Proofreading 5%; 4) M. Dimmock: Concept and input into manuscript 10%	No
3	Semi-supervised labelling of forensic CT data using deep learning.	Submitted	60%. Concept development, data collection and analysis.	1) D. Albrecht: Proofreading 5%, 2) P. Harris: Proofreading 5%, 3) R. Basset: Concept development 10%, 4) J. Gillam: Direction of manuscript 5%, 5) M. Dimmock: Concept and input into manuscript 15%	No
4	Development of a simple numerical model for trabecular bone structures.	Published	55%. Concept development, data collection and analysis.	1) D. Albrecht: Proofreading 5%; 2) D. Paganin: Concept development 10%; 3) P. Harris: Proofreading 10%; 4) C. Hall: Data collection 5%; 5) R. Basset: Proofreading 5%; 6) M. Dimmock: Concept and input into manuscript 10%	No
5	Effect of external fixation rod coupling in computed tomography.	Published	60%. Concept development, data collection and analysis.	1) M. Dimmock: Concept and input into manuscript 5%; 2) D. Albrecht: Proofreading 5%; 3) D. Paganin: Concept and input into manuscript 5%; 4) R. Basset: Proofreading 5%; 5) M. Klein: Veterinary advice 5%; 6) P. Harris: Concept development 15%	No

I have not renumbered sections of submitted or published papers reproduced within this thesis in order to generate a consistent presentation within the thesis.

Student name: Carlos Andrés Peña Solórzano

Student signature:

Date:

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Matthew Dimmock

Main Supervisor signature:

Date:

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisors, Dr. Matthew Dimmock and Dr. David Albrecht for their full support, guidance, encouragement and understanding throughout this endeavour, without which, this candidature would have been unquestionably an overwhelming pursuit.

Secondly, to my publication co-authors, the opportunity to work with an amazing and diverse group of people is an unforgettable experience. It has been an absolute privilege and joy to have been given the opportunity to collaborate. Special acknowledgements go to my non-official but still wonderful associate supervisors, without whom I could not have completed this work: firstly, thanks so much to Richard Bassed for your extensive advice in all matters forensic and all your support in resolving the tedious ethics and data access issues; secondly, a big thank you to Chris Harris, for your wonderful advice in the world of orthopaedic surgery, the long nights at the Australian Synchrotron, and ferrying me backwards and forwards to the Western Hospital with a freezer full of lamb legs in the boot of your car; and finally, to the always wonderful physicist David Paganin for your extensive advice in all things technical and fantastic problem solving abilities.

To the Department of Medical Imaging and Radiation Sciences, the Faculty of Medicine, Nursing and Health Sciences, the Faculty of Information Technology, and the Monash Institute of Graduate Research, thank you for the technical and administrative support given throughout my candidature. In addition, I would like to thank both academic units for the financial support provided as part of the Computational Biomedicine PhD Scholarship, 2016-2019.

Thanks also to the Australian Synchrotron and the Australian Orthopaedic Association for their support in the form of synchrotron beam time on the Imaging and Medical beam-line (IMBL) and grant money from the Victorian Orthopaedic Research Trust (VORT), respectively. In addition, I would like to thank the Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) HPC facility for the continuous technical support.

To my mother, Maria Ludivia Solórzano, and my father, Carlos Alberto Peña, thank you for your undying support, love and patience. This dream would not have been possible without your immense sacrifices. To my sister, Diana Catherine Peña Solórzano, thanks for your never-ending encouragement and guidance.

Finally, I would like to thank Maria Del Rocio Garcia Cruz. Your presence during the development of this project has made life significantly more enjoyable, thank you.

Contents

Abstract	iii
Acknowledgements	viii
1 Introduction	1
1.1 Medical databases	1
1.2 Machine learning and forensic imaging	2
1.3 Thesis Overview	4
2 Findings from machine learning in clinical medical imaging applications – lessons for translation to the forensic setting	8
2.1 Publication	8
3 Femur localisation and structured unsupervised labelling in CT scans using machine learning	26
3.1 Publication	27
4 Numerical model of the trabecular bone structure	43
4.1 Publication	44
4.2 Synthetic trabecular structure applications	56
5 Metal artefacts from the coupling of dense objects in CT	60
5.1 Publication	61
5.2 Deep learning denoising of a synthetically generated leg database	75
6 Discussion	78
7 Conclusion	82
Bibliography	83

1.1 Medical databases

Health records, composed of a patient's medical history, medical encounters, orders/prescriptions, progress notes, and test results (including medical images), are now largely found in electronic form, creating large digital medical databases. The text and image data from an individual record can be linked to yield diagnostic and prognostic indicators, or records can be aggregated to give population level trends [1, 2].

In the last decade, the amount of information that is being generated by health care systems around the world has increased substantially, thanks largely to the acquisition of imaging data [3, 4]. Of particular importance are the medical imaging techniques that generate volumetric representations of the body, e.g. computed tomography (CT) scans and magnetic resonance imaging (MRI). With the continued technical advances achieved in the field of medical imaging, the size of these databases will keep increasing due to improvements in resolution, speed, and reduction in radiation dose [5, 6]. For diagnostic and prognostic applications, the current medical imaging workflows typically involve the radiographer or radiologist labelling small portions of the data, leaving large volumes unannotated [7, 8] and limiting the potential for large scale applications.

Machine learning (ML) approaches offer a potential avenue for alleviating the level of human intervention during labelling, with further applications to automatic diagnosis and auditing images to make the database searchable [9, 10]. ML is widely used for modern image analysis and computer vision applications in areas as diverse as robotics [11, 12], surveillance [13, 14], and quality assurance. ML techniques are utilised to find patterns in data, and/or encode information that can be used for segmentation and classification

tasks. More recently, there has been much progress using ML in the automation of image processing tasks to enhance medical imaging workflows for the key modalities of plain film X-ray [15, 16], CT [17, 18], and MRI [19, 20]. Drawbacks in the use of these techniques include the lack of a gold standard to compare results between different ML approaches, as results giving definitive biomarkers from histo-pathology tests are not always available [21, 22].

The utilisation of ML for improving workflow efficiency requires natural language processing (NLP) to interpret and code both medical and legal documents, and image processing techniques to process images of anatomy, pathology and histology. Image and NLP applications are being widely studied using deep learning techniques (DL) [19, 23–25]. Although the analysis of clinical databases using these techniques is an active field of research, access to the data can be cumbersome due to the long waiting times for the ethics approvals, and the lack of centralised information [26] requiring the ethics applications to be filed in different centres. Furthermore, medical imaging and reporting in the clinical setting is usually anatomically-constrained, limiting its potential application to that of the organ or structure under investigation [27, 28].

In addition to clinical databases, equally important but as yet underutilised resources are those at forensic institutes. Forensic databases provide supporting evidence to the coroner for death investigations [29], and to the police for the crime scene reconstructions [30]. In the forensic setting, whole-body scans are commonly acquired giving significant information about the entire decedent anatomy. The ethics process can also be significantly shorter as the data is held by a single institution and the administration is more straightforward.

1.2 Machine learning and forensic imaging

In forensic investigations, post-mortem CT (PMCT) is the most widely used medical imaging technique for both the initial triage of the body and also for answering more complicated anatomically- and pathologically-motivated questions about cause of death (CoD). The use of clinical image processing approaches for PMCT workflows is non-trivial due to the additional complications relating to decedent decomposition, variability in constitution and positioning of normal anatomical structures resulting from CoD (e.g. incineration), artefacts from foreign bodies, and anatomical positioning due to rigor mortis.

The use of traditional (non-ML, e.g. statistical shape models [31–33]) image processing approaches for anatomical labelling in forensic applications would require a complicated and organ specific decision tree in order to capture the variations within the database. Figure 1.1 shows maximum intensity projections (MIPs) [34] generated from the triage CT scans of five representative cases from the PMCT database of the Victorian Institute of Forensic Medicine (VIFM). For a representative organ such as the femur, Figure 1.1

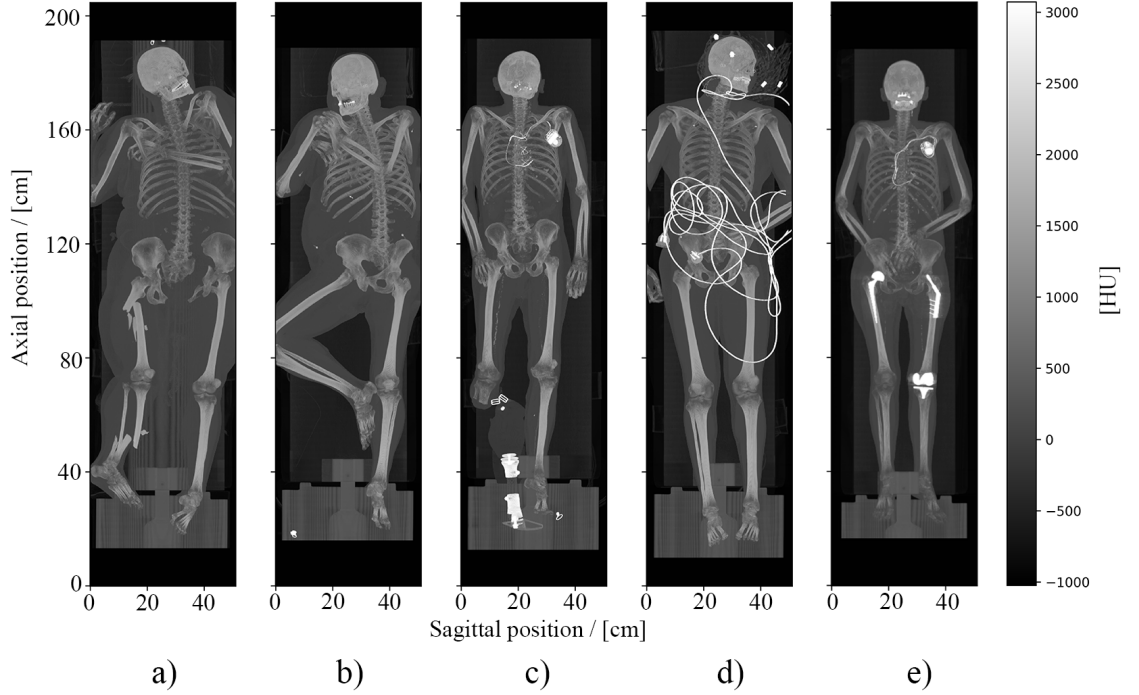


Figure 1.1: Coronal MIPs generated from the triage CT scans of representative cases found in forensic databases. a) High-force fracture. b) Limb outside of FoV. c) Missing limb. d) Foreign objects. e) Metallic implants.

demonstrates the significant variability relating to its identification and classification. Figures 1.1a-e show a high-force fracture, limb outside of field of view (FoV), missing limb, foreign objects, and presence of metallic implants, respectively.

In an attempt to overcome the difficulties in developing traditional algorithmic approaches that could account for the high degree of variability present in the forensic setting, the use of ML techniques has been investigated in this work. As stated previously, the development of ML approaches for analysing clinical imaging data is currently widely reported in the literature. The most common ML approaches used in clinical medical imaging include support vector machines (SVMs) [35], random forests (RFs) [36], and artificial neural networks (ANNs) [37]. ANNs are inspired by the way biological nervous systems process information. ANNs contain a large number of highly interconnected nodes (called neurons) which are separated into layers. These networks can process different pieces of information while taking constraints and relationships into consideration to coordinate internal processing, and to optimise the resulting output [37, 38]. The application of ANNs to medical images has facilitated the automation of needle insertion localisation for lumbar ultrasound images [39], bone age assessment [40], and segmentation of glioma [41] and the human retina [42], to name a few. Depending on network architecture and size, ANNs can have further classifications: DL is the subset of ML that utilises deep neural networks (DNNs) which differ from the more commonplace ANNs by their depth. DNNs have proven to be

immensely successful in problem solving due to their capacity to learn features from large training datasets [43].

There is currently a dearth of literature quantifying the performance of the aforementioned types of ML techniques to forensic medical imaging. The research that constitutes this thesis builds on the existing clinical and forensic literature available [44–46] and quantifies the limitations of the use of a ML pipeline in locating and classifying attributes of the femur. Ultimately, approaches such as that presented in this thesis for automatic labelling of vast medical databases will significantly reduce the burden on pathologists and radiologists in labelling retrospectively collected images. The labelling of retrospective images is of paramount importance for adding structure to existing and future databases in order to improve the access efficiency for answering important anthropological and medico-legal questions.

1.3 Thesis Overview

The first section (Chapter 2) of this thesis reports the findings of an extensive review of the literature. ML techniques were found to have been applied to a large number of tasks that could be used in medical imaging, where the algorithms most widely utilised include RFs, SVMs, and CNNs; CNNs were shown to deliver better performance for applications where different combinations of these approaches were directly contrasted. Techniques that were found to improve the ML performance in medical imaging included data augmentation, improved feature selection and algorithmic combination, e.g. majority voting. Performance was shown to be affected by resizing of the input images and the accuracy of the labels provided with the training data. In addition, the quantitative comparison of results from different groups for solving similar questions in the medical imaging domain was found to be difficult due to the lack of gold-standard labels, as well as the variability in reporting of metrics, and lack of reporting of a diagnostic odds ratio. It was concluded that ML applications that utilised CNNs could be re-purposed to the forensic domain with careful consideration to account for the increased variability and temporal factors, e.g. decomposition, that affect the data used to train the ML techniques. In broad terms, due to the complexity of the autopsy process, key applications of ML to forensic radiology would be to highlight and annotate areas of forensic interest, or to present information to optimally determine the CoD, including differentiation between body cavity fluid accumulations (blood, pus, ascites) and their corresponding volumes, calculation of organ volumes and weights, percentage of coronary artery calcification, identification of subtle fractures especially in critical areas such as the cervical spine, and determination of skeletal completeness and skeletal commingling after mass fatality incidents. These algorithms could also be used to automate decedent identification using dental records or photographs.

In the main study detailed in this thesis (Chapter 3), work pertaining to the localisation and labelling of the femur to provide enhanced structure to a post-mortem CT database was performed. DL algorithms were implemented and tested for localisation and classification tasks when processing PMCT scans. Following the development of the localisation and labelling pipeline, Chapter 4 presents a numerical model for generation of synthetic trabecular bone structures which is intended for applications to transfer learning to optimise the training efficiency of DL approaches. Chapter 5 then presents a study of the effects of metal artefacts from the coupling of dense structures on image quality. The subsequent discussion (Chapter 6) then contextualises the limitations of the DL approaches employed for femur localisation and classification by utilising a DNN trained to reconstruct synthetic trabecular bone structures in the output where the input comprised poor quality images generated using the aforementioned trabecular bone and metal artefact models. A more detailed summary of the findings of each individual section is described below.

1.3.1 Chapter 3: Femur localisation and structured unsupervised labelling in CT scans using machine learning

The first study presented in this thesis documents the implementation and interpretation of the two components of a DL pipeline that was developed to localise and label the femur. The system was trained using 350 full-body CT scans (~ 2000 slices per scan) from the VIFM database with a large variability in terms of body mass index (BMI), time since death, incompleteness of lower limbs, presence of metallic implants, trauma, and the presence of foreign objects. Three separate networks were created and linked for parallel execution in order to determine the presence (or absence) of an organ in a given slice of each of the three orthogonal views (axial, coronal and sagittal) defined by the reconstructed CT scan volume. The predictions of the networks were then subsequently merged in order to generate the 3D bounding box for the organ in the scan from each case. The resulting predicted bounding boxes, with the addition of 100 more cases obtained from a population with a higher probability of presenting metallic components in the femur, were then fed into a DL network. The network was designed for automatic labelling of the organs according to clustering in a latent vector space representation generated from a convolutional autoencoder (CAE). The results were then contextualised with respect to the available forensic information obtained from the associated pathology reports.

For the femur localisation, Dice scores of 0.99, 0.96, and 0.98 and mean absolute errors (MAEs) of 3.2, 7.1, and 4.2 mm were obtained in the axial, coronal, and sagittal views, respectively. Test cases were properly labelled as without-implant (I^-), nail, hip replacement, or knee replacement with an accuracy $>97\%$, where the recall for I^- and hip replacement cases was 1.00, but falling to 0.82 and 0.65 for cases with knee replacements and nails. Factors that were found to be statistically significant in affecting the localisation stage included the presence of metal in the primary femur, and the rotation of the femur

with respect to the movement axis of the scanner bed. The results of the femur localisation and automatic labelling were submitted to the journal *Forensic Science International* in November 2019 and we are still awaiting an outcome from this.

This project was the first to address the utilisation of DL techniques at the VIFM. As such, the work contained in this thesis was the first attempt at accessing large numbers of deidentified CT scans with accompanying decedent’s demographic information, and therefore had to resolve issues of data access. The transfer of retrospective data from the VIFM patient archive and communication system (PACS) was non-trivial, and the exporting of legacy cases was found to interfere with the day-to-day workflow of the forensic pathologists. Therefore, although ideally much larger numbers of cases would have been used, the limited times at which data could be accessed and the unavailability of a batch script to export the data en-mass restricted the workflow.

1.3.2 Chapter 4: Numerical model of the trabecular bone structure

Due to the large dataset requirement of machine learning techniques, and the necessity for customised training data, e.g. osteoporosis and bone fractures, we derived a numerical model for the generation of synthetic bone structures. The model included the generation of projection data for absorption- and propagation-based phase-contrast CT (PB-CT) [47,48] that was then reconstructed. PB-CT is a special case of conventional absorption-based CT and exploits the refraction of X-rays from coherent sources to significantly improve image contrast.

The synthetic structures, when compared with the real trabecular micro-architecture, yielded an average mean thickness error of $2\text{ }\mu\text{m}$, and a mean difference in standard deviation of $33\text{ }\mu\text{m}$ for the humerus, $24\text{ }\mu\text{m}$ for the ulna and radius, and $15\text{ }\mu\text{m}$ for the vertebrae. Simulated absorption- and propagation-based phase contrast CT projection data were generated and reconstructed using the derived mathematical simplifications from the two-material approximation, and the phase-contrast effects were successfully demonstrated. The results of this work were published in *Medical Physics* at the start of 2019.

1.3.3 Chapter 5: Metal artefacts from the coupling of dense objects in CT

Metal artefacts result from beam hardening [49] and photon starvation [50], and significantly affect image quality in both the clinical and forensic settings. The effect of metal artefacts on image quality was investigated and characterised. CT scans of fractured ex-vivo ovine legs with varying degrees of external fixation were acquired using dual-energy CT (DECT) and metal artefact reduction software (MARS) protocols.

Both the use of DECT and MARS were studied and quantified by measuring attributes of the resulting image histograms with and without metal. In the presence of metal structures, the bone distribution at 70 keV presented a change of 80.1% and 30.6% for σ , plus a relative error of 25.34% and 34.91% for μ , in the case of DECT and DECT + MARS, respectively. Higher energies in DECT improved σ and worsened μ of the bone peak, while the error increased for both parameters when both DECT and MARS protocols are used. The results of this work were published in *Strategies in Trauma and Limb Reconstruction* at the end of 2018.

From this research a set of tentative clinical guidelines were established for the positioning of external fixation frames that maintain the patient support but minimise artefacts visible in the volumetric scan data. In the final experimental section of this thesis, the proposed numerical model and the results from the study of metal artefacts from the coupling of dense objects were utilised for the generation of a noisy simulated database. Volumetric data was created to mimic axial slices through a femur, approximated as three concentric elliptical cylinders on a background with appropriate Hounsfield unit (HU) values. The simulated volumes that approximated extremity CT scans were utilised as the output of the network, after Gaussian noise was added [51] and the sizes of the simulated bones were varied to improve the algorithm's capabilities. A simple model was used to include the metal artefacts generated when different numbers of dense objects were added around the simulated leg. A DNN with similar architecture to the CAE, called VNet [52], was utilised to attempt to denoise the retrospective data by removing the artefacts whilst preserving the integrity of the simulated anatomy. V Nets are the most appropriate DL approach to use when reconstructing a volumetric input due to its edge preserving nature.

The results were obtained through comparison of the mean and width parameters of Gaussian curves fitted to the volume's histograms, where the soft tissue was not significantly affected by the addition of the synthetic coupling artefacts. For the bone tissue in the presence of the coupling artefacts, the mean and standard deviation values changed an average of 7% and 51%, respectively, when compared to the artefact-free volumes. The mean and width values of the predicted volumes when compared to the original volumes, differed an average of 3% and 1%, respectively. Although the results have been obtained from a simple model database, the presented artefact (and noise) removal approach utilising DL has potential applications to retrospective denoising of medical databases through direct training or transfer learning [53, 54].

Findings from machine learning in clinical medical imaging applications – lessons for translation to the forensic setting

Findings from machine learning in clinical medical imaging applications – lessons for translation to the forensic setting

C. A. Peña-Solórzano, D. W. Albrecht, R. B. Bassed, M. D. Burke, M. R. Dimmock

Submitted to Forensic Science International, October 15, 2019

This chapter is an exact copy of the paper submitted

A literature review was performed to determine current applications of ML for the analysis of medical images in both the clinical/pre-clinical and forensic settings in order to determine both the breadth and limitations of the identified approaches. It was determined that ML techniques are currently not widely used for either the generation of forensic evidence or the aggregation of data and subsequent analysis to give population level trends. It was also found that ML applications could be transferred from clinical to forensic settings, provided that due consideration was taken to account for the increased variability and temporal factors, e.g. decomposition, that affect the data used to train the algorithms. DNNs were widely reported to outperform other ML techniques in image processing tasks. There was found to be a dearth of quantitative metrics presented for inter-study comparisons. It was also found that a wider use of histopathological gold standard metrics for assessing algorithm performance is required.

2.1 Publication

The following paper has been submitted to Forensic Science International and is currently under review for publication.

Findings from machine learning in clinical medical imaging applications – lessons for translation to the forensic setting

Abstract

Machine learning (ML) techniques are increasingly being used in clinical medical imaging to automate distinct processing tasks. In forensic anatomical imaging, the use of these algorithms presents significant challenges due to variability in organ position, structural changes from decomposition, inconsistent body placement in the scanner, and the presence of foreign bodies. Existing ML approaches in clinical imaging can likely be transferred to the forensic setting with careful consideration to account for the increased variability and temporal factors that affect the data used to train these algorithms. Additional steps are required to deal with these issues, by incorporating the possible variability into the training data through data augmentation, or by using atlases as a pre-processing step to account for death-related factors. A key application of ML would be then to highlight anatomical and gross pathological features of interest, or present information to help optimally determine the cause of death. In this review, we highlight results and limitations of applications in clinical medical imaging that use ML to determine key implications for their application in the forensic setting.

Keywords

Machine learning, clinical medicine, forensic radiology, CT, MRI

Introduction

Forensic radiology is not clinical radiology applied to a deceased person. In the forensic setting, findings that a clinical radiologist may not typically have encountered are commonplace [1], e.g. post-mortem gas formation [2]. Post-mortem computed tomography (PMCT) is widely used in forensic investigations, where acquisition protocols used during clinical CT are not applicable due to rigor mortis and aversion to repositioning the decedent to avoid tampering with evidence. However, CT scans can be acquired with higher doses and there is no patient motion, therefore improving image quality. Additionally, recent developments such as PMCT angiography (PMCTA) with specialized pumps allows the diagnosis of vascular lesions whilst maintaining the integrity of anatomic structures, keeping evidence intact [3, 4].

In order to overcome the limitations of soft tissue contrast and a lack of vascular visualization provided by PMCT [5], postmortem magnetic resonance imaging

(PMMRI) is increasing in impact, albeit in a small way thus far. Whilst PMMRI offers improved soft tissue contrast, for vascular diagnoses it presents similar performance to PMCTA, with higher associated cost. However, applications to cardiac imaging are an exception, due to improved visualization of the coronary arteries and myocardium [5].

Recently, there has been much progress in the automation of image processing tasks to enhance medical imaging workflow [6, 7] for the key modalities of plain film X-ray [8], CT [9, 10], and MRI [11]. In the forensic setting, the completion of these tasks has added complications such as decedent decomposition, trauma, incineration, variability in positioning of normal anatomical structures, and artefacts from foreign bodies. This review synthesizes the information currently available in the literature on the pre-clinical and clinical use of machine learning (ML) applications, and contextualizes the information relative to the forensic setting. The limited literature relating

to the use of ML in forensic anatomical imaging is also addressed.

Whilst the use of MRI is not yet widespread in forensics, its growing popularity and extensive use with ML in clinical imaging yields important conclusions for long-term forensic implementation considerations. In addition, it should be noted that whilst there are extensive applications for the use of ML in both clinical and forensic histopathology, these are not considered.

Machine learning (ML) algorithms

Image processing typically involves segmentation, feature extraction, and classification. Many approaches have been proposed for medical image segmentation, including probabilistic atlases [12, 13], statistical shape models (SSMs) [14, 15], graph-cut (GC) algorithms [16, 17], multi-atlas segmentation (MAS) [18], and recently ML techniques. ML approaches have been proposed for the automation of both image analysis and diagnosis, reducing the burden on radiologists [19, 20].

ML techniques can be categorized as supervised learning, unsupervised learning, and reinforcement learning. In supervised environments, data is composed of input-output patterns, and the task is to find a deterministic function that can predict the output from an observed input. Unsupervised techniques are a type of self-organized learning that extracts structures from the training samples directly, without pre-existing labels [21]. More recently, self-supervised techniques, a type of unsupervised learning where the training data is automatically labelled by exploiting the relations between different input signals, are being studied for better utilizing unlabeled data [22]. Reinforcement learning on the other hand is based in trial-and-error, where the algorithm evaluates a current situation, takes an action, and receives feedback from the environment. This feedback can be positive or negative [23]. The most common ML techniques used in medical applications are summarized below.

Random forest (RF)

RFs operate by creating a multitude of decision trees (Fig. 1) that can be trained for classification and regression tasks, where the output is obtained by majority vote. Extra randomness is achieved by searching for the best feature among a random subset of features while splitting a node, avoiding overfitting and obtaining a better model.

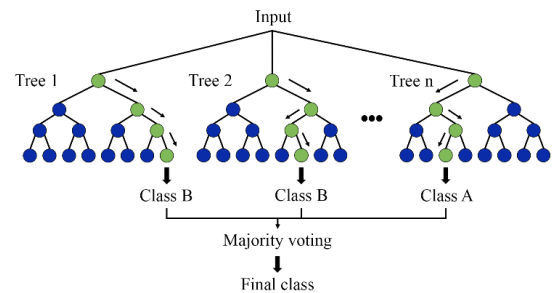


Fig. 1 RF technique showing the decision trees and majority voting for the class prediction of a new sample

k-nearest neighbor (k-NN)

In k -NN, the training samples are divided into classes, and the prediction of a new sample or test point is classified by a majority vote of its neighbors (Fig. 2). The algorithm uses a distance measurement function to search the k closest training samples in the feature space, and assigns the case of the class that is the most common in the subset.

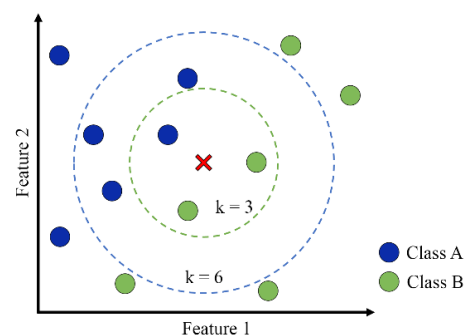


Fig. 2 k -NN technique used to predict the class of a new data point when the number of neighbors (k) for the majority vote changes from 3 to 6

Naive Bayes (NB)

NB is a classifier based on Bayes theorem, which works on conditional probability. Membership probabilities are predicted for each class, indicating the likelihood that a data point belongs to it and assuming feature independence in order to reduce the number of parameters in the model (Fig. 3) [20].

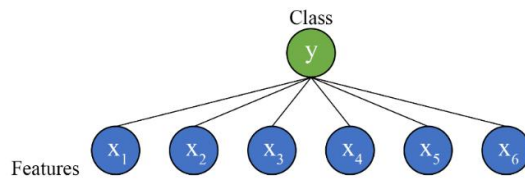


Fig. 3 Naive Bayes model where each node contains a probability table, and the feature values are conditionally independent

Support vector machines (SVMs)

SVMs originated from statistical learning theory [24] and are used for classification as they can model highly non-linear systems. SVMs project the data onto a high-dimensional space and apply a linear classifier on the projected data (Fig. 4) [21, 25].

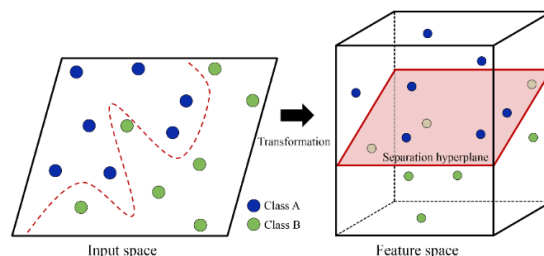


Fig. 4 SVM technique showing data transformation to a high-dimensional space and linear classifier (hyperplane) calculated on the projected data

Artificial neural networks (ANNs)

ANNs are inspired by the biological nervous system. ANNs contain a large number of highly

interconnected nodes (called neurons) separated into layers (Fig. 5), enabling the network to process different pieces of information while considering constraints to coordinate internal processing, and to optimize its final output [21, 26].

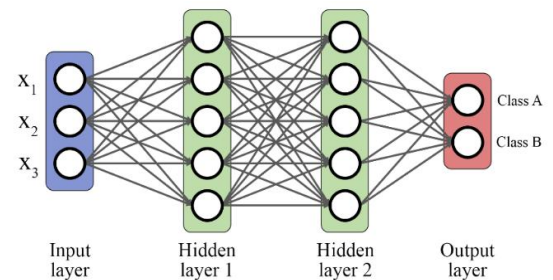


Fig. 5 Artificial neural network with three inputs, two outputs, and two hidden layers of five neurons each

Convolutional Neural Networks (CNN)

CNNs were inspired by the connectivity pattern of the animal visual cortex. Neurons respond to stimuli only in a restricted region (receptive field) of the previous layer, where receptive fields of different neurons partially overlap until they cover the entire visual field (Fig. 6). Unlike other ML techniques, the network learns the filters that are usually “hand crafted”. Also, CNNs exploit the strong spatially local correlation found on images, allowing the features to be detected regardless of their position. In recent years, Deep Neural Networks (DNNs), which differ from ANNs by their depth (the number of neuron layers), have proven to be successful in solving diverse problems, mainly for their capacity to learn features from large datasets [27].

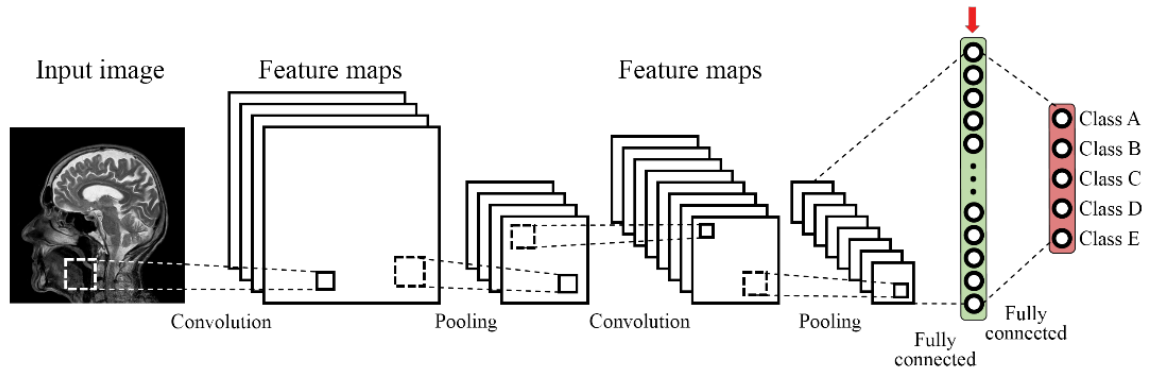


Fig. 6 Convolutional neural network with two fully connected layers for classification into five categories

Current clinical applications

ML techniques have been used for medical applications in the diagnosis and prognosis of diseases as well as for segmentation, classification, and measurement of anatomical structures [20, 28]. In this review, the ML applications have been grouped according to the tissue or organ studied, where brain, lungs, and skeleton were chosen to highlight results and limitations and determine key implications for their application in the forensic setting.

It should be noted that in the following discussion, algorithmic performance is assessed in terms of Dice's coefficient (DC), the modified Hausdorff distance (MHD) and the area under the receiver operating characteristic curve (AUC or AUROC), where possible. The DC quantifies overlap between the segmented image from the technique and a defined ground truth, ranging from zero (no overlap) to unity (identical segmentation). MHD is a measure of similarity between two objects based on their shape attributes. AUROC combines information of the true positive (TP) rate or sensitivity, and false positive (FP) rate or fall-out.

Brain tissue

Traditional atlas-based segmentations require registration to align the atlas images to the unseen image. Whereas, ML approaches can learn the variability between patients, making them especially useful in forensics, where the

image variability is greater than for clinical imaging. ML can also be used in combination with atlas-based approaches or in its own right. As an example of the former, Srhoj-Egekher et al. [29] used atlas-based segmentation for pre-processing T2-weighted MRI neonatal brain images to obtain initial probabilities, subsequently refined using a k -NN approach. Whilst Srhoj-Egekher et al. achieved DCs and MHDs ranging from 77% to 93%, and 0.35 to 2.86 respectively, the approach of assigning a tissue classification to each voxel independently, post atlas registration, meant some voxels were assigned to more than one class, while background voxels were unclassified. Conversely, Zhang et al. [30] opted for purely ML approaches that analyzed image patches for segmentation into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) of infant brains ($n=10$). Four network architectures were tested and, in most cases, the CNN method significantly outperformed SVMs and RFs with overall DC scores and MHDs of 85% and 0.32, respectively. The CNN method also outperformed two other common image segmentation methods, coupled level set (CLS) and majority voting (MV).

Table 1 Summary of selection of papers for non-infant brain tissue segmentation

Authors	Type / No. images	Mean Dice's coefficient (%)
van Opbroek et al. [31]	MRI / 5 training, 12 testing	GM=85, WM=88, CSF=78 (SVM)
Moeskops et al. [32]	MRI / 5 training, 10 testing	GM=91, WM=94, CSF=85 (CNN)
Wachinger et al. [33]	MRI / 20 cases for training, 10 cases for testing (>256 images per case)	All structures=91 (DCNN)

Three further publications were found where the authors segmented similar structures within adult brains. Van Opbroek et al. [31] applied a SVM for pixel-wise classification to registered volumes from a variety of MRI sequences for patients with diabetes and controls. The resulting segmentation of eight different tissue types demonstrated limited success (Table 1). The SVM showed poor performance in low contrast areas, while atlas misregistration caused voxels to be improperly classified. Moeskops et al. [32] used CNNs to process T1-weighted scans to segment the same eight tissue types. With CNNs, the use of different sized patches during training allowed for a smooth segmentation and analysis of local texture. In general, CNNs delivered better segmentation (Table 1), although this was a different patient cohort. A more recent application of 3D DCNNs [33] was used to identify 25 brain structures in T1-weighted MRI scans (n=30). Again, image patches were utilized as input to the network. However, spectral and Cartesian coordinate information relating to the patches was added after the convolutional layers (e.g. see arrow in Fig. 6) in order to introduce spatial information, which substantially increased the segmentation accuracy.

ML can also be used for the assisted diagnosis of neurodegenerative diseases. Salvatore et al. [34] used a combination of PCA with a SVM to classify morphological MRI sequences as patients with Parkinson's disease (n=28), progressive supranuclear palsy (PSP) (n=28), or controls (n=28). The large cohort sizes, inter-class cohort balance, and separation between PSP patients and other parkinsonian

variants were identified as particular strengths, compared to other papers. The performance (accuracy, specificity and sensitivity > 80%) of the model was shown to be limited by the number of principal components (16 to 26) utilized for classification. This dependence is an important consideration when using dimensionality reduction techniques and was also demonstrated for approaches that classified Alzheimer's disease [35].

Finally, ML techniques have also been used to segment and classify brain tumors. Zacharaki et al. [36] used conventional and perfusion MRI from patients with a diagnosis of intra-cranial neoplasm to classify them by type and grade of tumor (n=98). Their approach consisted of region of interest (ROI) definition, feature extraction, feature selection, and classification by SVMs. For comparison, linear discrimination analysis (LDA) and k -NN were also implemented. The mean classification accuracy was 91% for the SVM approach, compared with 81% for LDA and 90% for k -NN. Some of the limitations were related to the lack of features selected that described deformation of healthy structures due to the tumor, and the utilization of ROIs which yielded inter-observer variability.

Once the presence of tumors is verified, one possible subsequent step would be segmentation of the pathology, which is challenging even for experienced neuroradiologists [37]. To address this segmentation problem, a variant of CNNs named U-net is often employed [38]. Beers et al. [39] utilized two 3D U-nets connected sequentially to perform whole tumor,

enhancing tumor, and tumor core segmentation, achieving mean DCs for the test set ($n=95$) of 84%, 70%, and 71%, respectively. When the methodology was implemented on patients from ongoing clinical trials, the mean DCs decreased to 66%, 54%, and 45%, respectively. The lower performances on the clinical trial patients were attributed to scans being post-operative, highlighting the importance of case selection for training.

Studies on brain tissues used mostly MRI data due to the multi-modality information and a good soft-tissue contrast. Whilst the specific pathologies discussed are not all relevant to the forensic setting, the general conclusions deduced from the segmentation and localization of anatomical abnormalities are. Models that utilized dimensionality reduction techniques prior to classification were shown to yield performances dependent on the number of selected components. In addition, the identification of abnormalities in biological tissues required features capable of describing complicated deformations of the healthy structures. In CNNs, the performance of the technique depended significantly on the training set adequately representing expected cases. In general, CNNs outperformed algorithms such as SVMs, RFs, CLSs, and MV in segmentation and classification tasks. Note that some studies used small datasets, which limited statistical power. In addition, as will be demonstrated throughout this review, a combination of the variability in reporting of metrics, the lack of reporting of a diagnostic odds ratio [40], the unavailability of datasets and reference implementations, and the effect of imbalanced data in the classification accuracy, common in medical datasets [41, 42], made it difficult to compare papers quantitatively.

In the topic of brain tissue segmentation, the Multimodal Brain Tumor Segmentation Challenge (BraTS) (<http://braintumorsegmentation.org/>) presents state-of-the-art methods for the segmentation of brain tumors in MRI scans.

Lungs

In machine learning, feature learning refers to the automatic discovery of meaningful representation of features from raw data, in contrast to manual feature engineering, where the features have to be chosen by a domain expert. Feature learning allows for end-to-end learning, where a complex system can be represented by a single model, bypassing the intermediate layers present in traditional workflow designs. Learning a representation of any tissue is a useful process if subsequent classification is required, or if the goal is to find differences between samples in the training data. The representation quality is highly dependent on the learned features.

A restricted Boltzmann machine (RBM) is a generative neural network that can be used to perform automatic feature learning. Li et al. [43] used a Gaussian RBM with a training dataset consisting of different sized patches obtained from high-resolution lung CT images ($n=92$), with the purpose of classifying five tissue types using SVMs. The best accuracy obtained was 84%, with a high rate of FPs caused by the similarity between tissues.

Van Tulder and de Bruijne [44] utilized convolutional RBMs, adding learning objectives that helped the algorithm to extract features for description and training data classification. The training data consisted of CT scans ($n=73$) with five types of tissues classified. Resulting accuracies were $<75\%$, and 85-90% for the classification of lung patches and airway centerlines respectively. The low accuracies were attributed to small training sets and number of extracted filters due to computational restrictions.

Netto et al. [45] utilized examinations ($n=50$) with 198 identified nodules and a SVM to classify the structure as nodule or non-nodule. The resulting accuracy was 91%, with a sensitivity of 86%. The largest errors were reported when the feature was very large or very small, where it could be mistaken for other structures or for being the continuation of one.

Hua et al. [46] used images containing nodules from the Lung Image Database Consortium (LIDC) CT dataset to train both a CNN and a deep belief network (DBN) constructed by stacking RBMs. The performance of the two networks was then compared with two feature-based methods: scale invariant feature transform (SIFT) with local binary patterns (LBPs) and k-NN, and the fractal analysis method together with a SVM (Table 2). The major limitation reported was resizing of the input images, which discarded size cues that were important indicators of malignancy.

Kumar et al. [47] also used the LIDC images (Table 2) to extract features obtained using an autoencoder (AE) and a binary decision tree classifier (BDT). The false positive rate of 39% was attributed to the visual similarity between benign and malignant cases, which can be compared to a 27% rate obtained on The National Lung Screening Trial (NLST) using low-dose CT (LDCT) [48].

A more recent study compared massive-training artificial neural networks (MTANNs) against CNNs [49] using a database of LDCT scans ($n=38$), consisting of 1057 slices. MTANNs are an extension of ANNs, where a large number of overlapping sub-regions are created for each voxel of the original image and used as inputs to the network. The reported AUROC was 0.88 for the MTANN, and 0.78 for the best of the four CNN architectures. The MTANN required a smaller number of training samples than the CNNs for a better classification performance. This was attributed to the hierarchies of the learned features, where the MTANN learned to detect lesions utilizing low-level features, while the CNNs extracted low-, mid- and high-level features, increasing their reliance on irrelevant characteristics.

Studies on lungs generally used CT scans for the segmentation of tissues and tumors, and classification of nodules for early cancer diagnosis. Due to the low contrast between different tissues in the lungs, the approaches reported were reliant on shape, texture, and feature size. The segmentation performance was poor for nodules at the size extremes. Major findings included lower performances due to image resizing, and the importance of reporting false positive rates, which can yield high values in applications that intend to determine nodule malignancy.

Skeleton

Skeletal segmentation usually occurs before measurement and/or diagnosis of bone or articular diseases. Koch et al. [50] segmented MRIs ($n=110$) of the wrist using marginal space learning (MSL) and RFs, where MSL incrementally learned classifiers in marginal spaces of lower dimensions [51]. The segmented images were used to compute the 3D model of every carpal bone, with AUCs of 0.88 for both scan modalities. The approach was an order of magnitude faster than previous work using a semi-automatic method. Similar literature did not report segmentation errors and could not be used for comparison.

Hao et al. [52] used MRI to diagnose intervertebral disk degeneration. Similarity computations, extracted from a training set using texture and shape data, were input to a modelling and learning stage and then to a SVM classifier that took new data from a test set and labelled it as healthy or degenerated. A key outcome of this paper was the introduction of an active learning phase, alleviating human labor by automatically selecting a subset of representative training data for a medical expert. This way, the classification accuracy remained between 90% and 95% while using no more than one third of the labelled data.

Table 2 Summary of selection of papers for lung nodule classification

Authors	Type / dataset	Accuracy (%)	Sensitivity (%)
Hua et al. [46]	CT / 2545 nodules	-	73 (DBN), 73 (CNN), 76 (SIFT)
Kumar et al. [47]	CT / 4323 nodules	75	83 (AE+BDT)

Bone age assessment from plain X-rays is used in pediatrics by comparing the results to chronological age for the evaluation of endocrine and metabolic disorders. A fully automated pipeline was presented by Lee et al. [53] using a pre-trained CNN (transfer learning). Both male and female test X-rays were assigned a bone age within 1 year of the correct value over 90% of the time, and over 98% within 2 years.

X-rays have also been widely used for fracture detection, e.g. of the tibia [54], where texture and shape features were fed into three different ML algorithms: an ANN, k -NN, and SVM, and the outputs fused using a majority vote scheme. The combination of the classifiers using both types of features presented a significant improvement over using just one classifier, or only one feature type. Reported accuracies, precisions, and sensitivities were above 97%. Instead of fusing the results from the classifiers, multi-stage classifiers have also been used. Wels et al. [55] reported a fully automatic system using several RF stages, capable of detecting osteolytic spinal bone lesions from CT volumes, with an average sensitivity of 75%. The performance was affected by differences in contrast and noise characteristics in the data used for training and testing, however, values for accuracy were not presented for further interrogation.

Sharma et al. [56] measured trabecular bone microarchitecture and used the information to discriminate between healthy cases ($n=10$) and patients with Type 1 Gaucher disease ($n=20$). SVMs were used to classify different genotypes of the disease, achieving an average 70% classification accuracy, 74% sensitivity, and 85% precision. The structure of the trabecular bone obtained from MRI have also been used classify

knees with osteoarthritis [57]. The characteristics found to relate to the disease were useful in classifying healthy from affected patients ($n=159$) with an AUC of 0.92, as well as predicting the risk of cartilage loss. In a similar study, the fractal analysis of X-ray images with SVMs enabled the automatic classification of osteoporotic patients ($n=39$) versus controls ($n=38$) with accuracies of up to 95% [58]. Reported limitations from the papers in this section include the small number of cases and the high percentages of patients at early stages of the disease.

Orthopedic ML applications include disease diagnosis, age assessment, and risk prediction e.g. osteoporosis, osteoarthritis. Plain film X-ray and CT were most common; however, MRI studies of joints are being increasingly reported. The performance of ML applications was shown to be affected by the number and selected features, which is significantly influenced by differences in contrast and noise characteristics in the datasets. Comparison or ranking of the results was limited by reported performance metrics and the use of databases that were not representative of the disease stages studied. Other limitations included small patient cohorts and the processing times.

Forensic applications

Three important results for the use of ML in clinically-related applications were found that can also be applied in the forensic setting: firstly, temporal efficiency through the use of transfer learning; secondly, improved accuracy through the combination of ML classifiers using majority voting techniques or multi-stage approaches; and finally, the addition of an active learning phase, where the human labor can be alleviated during labeling.

In terms of the currently reported use of ML in forensic medicine, it is in its infancy. ML has only been trialed in a few specific forensic applications including the automation of bone age assessment [59]; sex determination using bone information [60]; prediction of skull fractures [61]; automatic detection of hemorrhagic pericardial effusion [62]; and automatic forensic dental identification [63]. As far as we are aware, none of these studies has translated into daily forensic practice, despite the potential to streamline case-work in these particular disciplines.

Stern and Urschler [59] utilized MR hand images and random regression forests to fuse growth information from individual bones for automated age estimation. In this case, the MR images were obtained from living volunteers, but the application to the forensic setting was discussed. The resulting mean absolute deviation and standard deviation from the chronological age were 0.82 ± 0.56 years. The results were reported to be similar to clinically established methods.

Arigbabu et al. [60] performed sex estimation utilizing 100 head PMCT scans. They combined and evaluated six local feature representations, two feature learning, and three classification algorithms. The best prediction rate was 86% using mesh local binary pattern (MeshLBP) for features extraction, kernel PCA (KPCA) for dimensionality reduction, and SVM for the classification. The results were within the reported sex prediction range for applications that use cranial features.

Heimer et al. [61] performed skull fracture prediction using 150 head PMCT scans (75 scans for each case: with and without fractures) and deep learning. The skulls were preprocessed through curved maximum intensity projections (CMIPs), where the skull's surface is unfolded into a curved projection that can be observed from a single view. Deep learning was applied and the best-performing selected network yielded an AUC

of 0.965, a sensitivity of 91.4% and a specificity of 87.5%.

Ebert et al. [62] used two separate deep neural networks: the first for classification of images with or without hemopericardium, and the second for the segmentation of the blood content in PMCT. The average DC, recall, and precision for the classification task were 77%, 77%, and 85% respectively. For segmentation, the values obtained were 78%, 78%, and 79%, respectively.

Dental identification traditionally requires a forensic odontologists to find the best match between a post-mortem image and an ante-mortem database, using features such as dental restorations, pathologies, and tooth and bone morphologies. Zhang et al. [63] proposed a new descriptor that encodes the local shape of a person's dental features. They subsequently used a RF classifier to match the features of the unknown person to those in the database ($n=200$). The result yielded 100% accuracy for complete ($n=20$) and incomplete ($n=20$) datasets. Incomplete datasets were derived from cases involving trauma. The method presented was shown to be rotationally and translationally invariant, and was orders of magnitude faster than conventional 2D methods. It is important to note that the database was constructed by using a surface laser scanner on plaster samples in contrast to PMCT scans.

Discussion

Typical goals of ML techniques in medical imaging include the differentiation of healthy from diseased patients or tissues and the localization of pathologies in anatomic structures. Algorithmic performance can be significantly affected when trying to process a new sample that differs significantly from the training dataset. This characteristic is especially important when it comes to applications in forensic medicine, where there is a high variability in the structures and image acquisition protocols, and unclear definition of what normal implies, due to changes occurring because of circumstances of death, tissue

decomposition, trauma, or incineration. However, some applications e.g. organ localization, can be immediately translated to the forensic setting by using the appropriate training data, or by using the clinical medical images for the initial training of CNNs and then fine-tuning using forensic information. This is usually referred to as transfer learning. On the other hand, due to the size and availability of forensic databases, the opposite is also possible, with applications being trained in forensic data and then fine-tuned to the clinical setting.

To improve the capabilities of ML techniques, the training data can be modified, or more informative features can be used as inputs to the algorithms. The selection of features can be optimized using learning objectives [44] or by utilizing an unsupervised technique as a preprocessing step to the classification task [47, 64]. The features selected can also be used to alleviate human labelling, by selecting more representative training data for the medical expert [52, 65]. Another approach to the improvement of ML performance is the combination of several techniques using a majority vote scheme [54], or the use of multi-stage classifiers [39] for segmentation of different spatially related tissues.

A wide range of implemented algorithms were found during the review process, where SVMs outperformed techniques such as LDA and k -NN [36], however the trend in recent works has been the outperformance of CNNs [30, 32]. The main disadvantage of classic ML approaches compared to CNNs is the performance variability due to the quality of the features [34] that must be hand-crafted by an expert according to the goal and dataset. The selected feature pool is commonly processed to lower its dimensionality before training the classifier by using techniques such as PLS or PCA, where PLS has shown to be better for relevant feature extraction [35]. It is important to note that the number of principal components or features selected at the end of this step plays a key role in the classification performance [34].

The performance of the algorithms can also be significantly affected if the labelling process (diagnosis) is prone to error [35]. Furthermore, for medical and forensic applications, the common practice of resizing input images can yield to a loss of information that could be essential for diagnosis purposes [46]. An additional consideration is that some authors use for example a radiologist to classify cases, then benchmark the performance of the algorithm against radiologists. Rajpurkar et al. [66], for instance, presented a CNN that achieved radiologist-level pneumonia detection on a database [67] for which no gold-standard label exist, and listed as limitation the lack of information in the database that affects the radiologists' accuracy. It is also important to note that the lack of reporting of a diagnostic odds ratio [40] and the variability in reporting of metrics makes it difficult to compare papers.

For the task of segmentation, both multi-atlas algorithms and DCNNs with multiple patch sizes showed comparable results [29, 30], demonstrating CNNs were most successful. Patch-based techniques could be a good approach in forensic cases where organs or structures are not localized in the usual anatomic positions [44]. Furthermore, the use of different sized patches in segmentation tasks allows for both a smoother separation and the detailed analysis of local texture [32].

One of the main issues that affects both the clinical and forensic settings is the lack of interpretability of predictions by black-box approaches such as neural networks. This is active area of current research and a current approach to addressing this concern is the use of visual explanations for the class label under consideration, obtained from the convolutional layer feature maps [68, 69], and attention mechanisms [70], able to determine the parts of the input images more relevant for a particular classification. Furthermore, depending on the application, it is not required and could be counter-productive to completely automate a task, for which a human-in-the-loop can be

beneficial by reducing the complexity through human input and assistance [65].

Some applications of ML already found in clinical medicine, that could be repurposed for forensic medicine, include segmentation and classification of organs and structures, including arteries, tiny blood vessels, the liver, spleen, stomach, gallbladder, and pancreas [71, 72]; computation of organ 3D models [50] for virtual autopsies; detection of lesions and calcification on vascular cross-sections [73]; identification of bone and joint atrophies or disorders [52, 56, 57, 58]; fluid volume and composition on body cavities (blood, pus, ascites) [74]; and organ volume estimation, e.g. heart size with respect to body size [75].

Tasks in forensic radiology that to our knowledge have not been tackled using ML include: segmentation and classification of foreign bodies, differentiation between ante-mortem and post-mortem gases, calculation of body mass index, and determination of skeletal completeness after accidents.

For the segmentation and classification of foreign bodies, e.g. bullets, metallic dental fillings, the main challenge becomes finding the object that does not belong inside the body. Furthermore, metallic components can create artefacts such as beam-hardening on CT scans or field distortions in MRI [76], which can also be addressed using deep learning [77].

Differentiation between ante-mortem and post-mortem gases can be difficult using the voxel values of CT scans or MRI, so emphasis should be placed on understanding the expected location and evolution of these gases at different points in time [78]; also, differentiation between acute and remote infarction on the brain, which on a CT scan can be characterized by voxel values and tissue volume changes, can be tackled utilizing existing tissue classification techniques [31, 32, 35], with the addition of new classes to differentiate the types of infarction.

In forensic anthropology, tasks that could be addressed using ML include: determination of skeletal completeness after accidents [79], e.g. plane crashes; 3D reconstruction of incomplete bones, that could be extrapolated from the work by Hermoza and Sipiran [80] on incomplete archaeological objects; and 3D reconstruction of fractured skulls [81, 82], used to infer a cause of death, or to perform facial reconstruction.

Conclusions

ML techniques have been applied to a large number of tasks that could be used in medicine, where the algorithms most widely utilized in applications with medical images include RFs, SVMs, and CNNs. CNNs have shown better performance in the literature.

Techniques to improve the ML performance in radiology include data augmentation, improved feature selection and algorithmic combination, e.g. majority voting. Performance was shown to be affected by resizing of the input images and the accuracy of the labels provided with the training data. In addition, benchmarking was found to be difficult due to the lack of gold-standard labels, as well as the variability in reporting of metrics, and lack of reporting of a diagnostic odds ratio.

ML applications investigated for clinical medicine could be repurposed to the forensic domain with careful consideration to account for the increased variability and temporal factors, e.g. decomposition, that affect the data used to train the ML techniques. Due to the complexity of the autopsy process, a key application of ML to forensic radiology would be to highlight and annotate areas of forensic interest, or to present information to optimally determine the cause of death, including differentiation between body cavity fluid accumulations (blood, pus, ascites) and their corresponding volumes, calculation of organ volumes and weights, percentage of coronary artery calcification, identification of subtle fractures especially in critical areas such as the cervical spine, and determination of skeletal

completeness and skeletal commingling after mass fatality incidents.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Sutherland T, O'Donnell C. The artefacts of death: CT post-mortem findings. *J Med Imag Radiat On.* 2018;62(2):203–210. Available from: <https://doi.org/10.1111/1754-9485.12691>.
2. Bolliger SA, Thali MJ. Imaging and virtual autopsy: looking back and forward. *Philos T Roy Soc B.* 2015;370(1674). Available from: <https://doi.org/10.1098/rstb.2014.0253>.
3. Ross SG, Bolliger SA, Ampanozi G, Oesterhelweg L, Thali MJ, Flach PM. Postmortem CT angiography: capabilities and limitations in traumatic and natural causes of death. *Radiographics.* 2014;34(3):830–846. Available from: <https://doi.org/10.1148/rg.343115169>.
4. Saunders SL, Morgan B, Raj V, Rutty GN. Post-mortem computed tomography angiography: past, present and future. *Forensic Sci Med Pat.* 2011;7(3):271–277. Available from: <https://doi.org/10.1007/s12024-010-9208-3>.
5. Grimm JM, Grabherr S. Future prospects of forensic imaging. In: *Atlas of Postmortem Angiography*. Springer; 2016. p. 583–588. Available from: https://doi.org/10.1007/978-3-319-28537-5_36.
6. Gibson E, Li W, Sudre C, Fidon L, Shakir DI, Wang G, et al. NiftyNet: a deep-learning platform for medical imaging. *Comput Meth Prog Bio.* 2018;158:113–122. Available from: <https://doi.org/10.1016/j.cmpb.2018.01.025>.
7. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digit Med.* 2018;1(1):9.
8. Johnston J, Fauber TL. *Essentials of radiographic physics and imaging*. Elsevier Health Sciences; 2015.
9. Seeram E. *Computed Tomography: Physical Principles, Clinical Applications, and Quality Control*. Elsevier Health Sciences; 2015.
10. Norman N, Dimmock MR, Lee K, Graham J, Bassed R. The applicability of Dual-Energy Computed Tomography (DECT) in forensic odontology—A review. *J Forensic Radiol Imaging.* 2017;10:15–22. Available from: <https://doi.org/10.1016/j.jofri.2017.07.002>.
11. Constantinides C. *Magnetic Resonance Imaging: The Basics*. CRC Press; 2016.
12. Dura E, Domingo J, Ayala G, Marti-Bonmati L, Goceri E. Probabilistic liver atlas construction. *Biomed Eng OnLine.* 2017;16(1):15. Available from: <https://doi.org/10.1186/s12938-016-0305-8>.
13. Park H, Bland PH, Meyer CR. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE T Med Imaging.* 2003 April;22(4):483–492. Available from: <https://doi.org/10.1109/TMI.2003.809139>.
14. Okada T, Shimada R, Hori M, Nakamoto M, Chen YW, Nakamura H, et al. Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model. *Acad Radiol.* 2008;15(11):1390 – 1403. Available from: <https://doi.org/10.1016/j.acra.2008.07.008>.
15. Seim H, Kainmueller D, Heller M, Lamecker H, Zachow S, Hege HC. Automatic

- segmentation of the pelvic bones from CT data based on a statistical shape model. In: *Proceedings of the First Eurographics Conference on Visual Computing for Biomedicine. EG VCBM'08. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2008. p. 93–100.*
16. Boykov YY, Jolly MP. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: *Proceedings Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001. vol. 1. IEEE; 2001. p. 105–112.* Available from: <https://doi.org/10.1109/ICCV.2001.937505>.
 17. Bagci U, Chen X, Udupa JK. Hierarchical scale-based multiobject recognition of 3-D anatomical structures. *IEEE T Med Imaging.* 2012 March;31(3):777–789. Available from: <https://doi.org/10.1109/TMI.2011.2180920>.
 18. Xu Z, Burke RP, Lee CP, Baucom RB, Poulose BK, Abramson RG, et al. Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. *Med Image Anal.* 2015;24(1):18 – 27. Available from: <https://doi.org/10.1016/j.media.2015.05.009>.
 19. Bhargavan M, Kaye AH, Forman HP, Sunshine JH. Workload of radiologists in United States in 2006–2007 and trends since 1991–1992. *Radiology.* 2009;252(2):458–467. Available from: <https://doi.org/10.1148/radiol.2522081895>.
 20. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal.* 2012;16(5):933–951. Available from: <https://doi.org/10.1016/j.media.2012.02.005>.
 21. Camastra F, Vinciarelli A. *Machine learning for audio, image and video analysis.* Springer; 2008.
 22. Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D. Self-supervised learning for medical image analysis using image context restoration. *Med Image Anal.* 2019;58:101539.
 23. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Proc Mag.* 2017;34(6):26–38.
 24. Vapnik VN. *The Nature of Statistical Learning Theory.* Springer-Verlag New York, Inc.; 1995.
 25. Chaplot S, Patnaik L, Jagannathan N. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed Signal Proces.* 2006;1(1):86–92. Available from: <https://doi.org/10.1016/j.bspc.2006.05.002>.
 26. Jiang J, Trundle P, Ren J. Medical image analysis with artificial neural networks. *Comput Med Imag Grap.* 2010;34(8):617–631. Available from: <https://doi.org/10.1016/j.compmedimag.2010.07.003>.
 27. Štern D, Payer C, Lepetit V, Urschler M. Automated age estimation from hand MRI volumes using deep learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2016. p. 194–202. Available from: https://doi.org/10.1007/978-3-319-46723-8_23.
 28. Suzuki K. Pixel-based machine learning in medical imaging. *Int J Biomed Imaging.* 2012;2012:1. Available from: <http://dx.doi.org/10.1155/2012/792079>.
 29. Srhoj-Egekher V, Benders M, Kersbergen KJ, Viergever MA, Isgum I. Automatic segmentation of neonatal brain MRI using atlas based segmentation and machine learning approach. In: *MICCAI Grand Challenge: Neonatal Brain Segmentation; 2012.*
 30. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage.*

- 2015;108:214–224. Available from: <https://doi.org/10.1016/j.neuroimage.2014.12.061>.
31. van Opbroek A, van der Lijn F, de Bruijne M. Automated brain-tissue segmentation by multi-feature SVM classification. In: The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS13); 2013. OA.
 32. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJ, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE T Med Imaging*. 2016;35(5):1252–1261. Available from: <https://doi.org/10.1109/TMI.2016.2548501>.
 33. Wachinger C, Reuter M, Klein T. DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*. 2018;170:434 – 445. Segmenting the Brain. Available from: <https://doi.org/10.1016/j.neuroimage.2017.02.035>.
 34. Salvatore C, Cerasa A, Castiglioni I, Gallivanone F, Augimeri A, Lopez M, et al. Machine learning on brain MRI data for differential diagnosis of Parkinson’s disease and Progressive Supranuclear Palsy. *J Neurosci Meth*. 2014;222:230–237. Available from: <https://doi.org/10.1016/j.jneumeth.2013.11.016>.
 35. Khedher L, Ramrez J, Górriz JM, Brahim A, Segovia F. Early diagnosis of Alzheimer’s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*. 2015;151:139–150. Available from: <https://doi.org/10.1016/j.neucom.2014.09.072>.
 36. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn Reson Med*. 2009;62(6):1609–1618. Available from: <https://doi.org/10.1002/mrm.22147>.
 37. Bonte S, Goethals I, Van Holen R. Machine learning based brain tumour segmentation on limited data using local texture and abnormality. *Comput Biol Med*. 2018;98:39–47. Available from: <https://doi.org/10.1016/j.compbimed.2018.05.005>.
 38. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer; 2015. p. 234–241. Available from: https://doi.org/10.1007/978-3-319-24574-4_28.
 39. Beers A, Chang K, Brown J, Gerstner E, Rosen B, Kalpathy-Cramer J. Sequential neural networks for biologically-informed glioma segmentation. In: *Medical Imaging 2018: Image Processing*. vol. 10574. International Society for Optics and Photonics; 2018. p. 1057433. Available from: <https://doi.org/10.1117/12.2293941>.
 40. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129–1135. Available from: [https://doi.org/10.1016/S0895-4356\(03\)00177-X](https://doi.org/10.1016/S0895-4356(03)00177-X).
 41. Liu T, Fan W, Wu C. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. *Artif Intell Med*. 2019;101:101723.
 42. Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: Experimental evaluation. *Inform Sciences*. 2020;513:429–441.
 43. Li Q, Cai W, Feng DD. Lung image patch classification with automatic feature learning. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE; 2013. p. 6079–6082. Available from: <https://doi.org/10.1109/EMBC.2013.6610939>.
 44. van Tulder G, de Bruijne M. Combining generative and discriminative

- representation learning for lung CT analysis with convolutional restricted boltzmann machines. *IEEE T Med Imaging*. 2016;35(5):1262–1272. Available from: <https://doi.org/10.1109/TMI.2016.2526687>.
45. Netto SMB, Silva AC, Nunes RA, Gattass M. Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Comput Biol Med*. 2012;42(11):1110–1121. Available from: <https://doi.org/10.1016/j.compbiomed.2012.09.003>.
 46. Hua KL, Hsu CH, Hidayati SC, Cheng WH, Chen YJ. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Ther*. 2015;8. Available from: <https://doi.org/10.2147/OTT.S80733>.
 47. Kumar D, Wong A, Clausi DA. Lung nodule classification using deep features in CT images. In: 2015 12th Conference on Computer and Robot Vision (CRV). IEEE; 2015. p. 133–138. Available from: <https://doi.org/10.1109/CRV.2015.25>.
 48. Mehta HJ, Mohammed TL, Jantz MA. The American College of Radiology Lung Imaging Reporting and Data System: potential drawbacks and need for revision. *Chest*. 2017;151(3):539–543. Available from: <https://doi.org/10.1016/j.chest.2016.07.028>.
 49. Tajbakhsh N, Suzuki K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. *Pattern Recogn*. 2017;63:476–486. Available from: <https://doi.org/10.1016/j.patcog.2016.09.029>.
 50. Koch M, Schwing AG, Comaniciu D, Pollefeys M. Fully automatic segmentation of wrist bones for arthritis patients. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2011. p. 636–640. Available from: <https://doi.org/10.1109/ISBI.2011.5872487>.
 51. Zheng Y, Comaniciu D. Marginal Space Learning. In: *Marginal Space Learning for Medical Image Analysis: Efficient Detection and Segmentation of Anatomical Structures*. New York, NY: Springer New York; 2014. p. 25–65. Available from: https://doi.org/10.1007/978-1-4939-0600-0_2.
 52. Hao S, Jiang J, Guo Y, Li H. Active learning based intervertebral disk classification combining shape and texture similarities. *Neurocomputing*. 2013;101:252–257. Available from: <https://doi.org/10.1016/j.neucom.2012.08.008>.
 53. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging*. 2017;30(4):427–441. Available from: <https://doi.org/10.1007/s10278-017-9955-8>.
 54. Umadevi N, Geethalakshmi S. Multiple classification system for fracture detection in human bone X-ray images. In: 2012 Third International Conference on Computing Communication & Networking Technologies (ICCCNT). IEEE; 2012. p. 1–8. Available from: <https://doi.org/10.1109/ICCCNT.2012.6395889>.
 55. Wels M, Kelm BM, Tsybal A, Hammon M, Soza G, Sühling M, et al. Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control. In: *Medical Imaging 2012: Computer-Aided Diagnosis*. vol. 8315. International Society for Optics and Photonics; 2012. p. 1 – 8. Available from: <https://doi.org/10.1117/12.911169>.
 56. Sharma GB, Robertson DD, Laney DA, Gambello MJ, Terk M. Machine learning based analytics of micro-MRI trabecular bone microarchitecture and texture in Type 1 Gaucher disease. *J Biomech*. 2016;49(9):1961–1968. Available from: <https://doi.org/10.1016/j.jbiomech.2016.04.010>.
 57. Marques J, Genant HK, Lillholm M, Dam EB. Diagnosis of osteoarthritis and

- prognosis of tibial cartilage loss by quantification of tibia trabecular bone from MRI. *Magn Reson Med*. 2013;70(2):568–575. Available from: <https://doi.org/10.1002/-mrm.24477>.
58. Taфраouti A, El Hassouni M, Toumi H, Lespessailles E, Jennane R. Osteoporosis diagnosis using fractal analysis and support vector machine. In: 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS). IEEE; 2014. p. 73–77. Available from: <https://doi.org/10.1109/SITIS.2014.49>.
59. Stern D, Urschler M. From individual hand bone age estimates to fully automated age estimation via learning-based information fusion. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE; 2016. p. 150–154. Available from: <https://doi.org/10.1109/ISBI.2016.7493232>.
60. Arigbabu OA, Liao IY, Abdullah N, Mohamad Noor MH. Computer vision methods for cranial sex estimation. *IPSJ Trans Comput Vis Appl*. 2017;9(1):19. Available from: <https://doi.org/10.1186/s41074-017-0031-6>.
61. Heimer J, Thali MJ, Ebert L. Classification based on the presence of skull fractures on curved maximum intensity skull projections by means of deep learning. *J Forensic Radiol Imaging*. 2018;14:16 – 20. Available from: <https://doi.org/10.1016/-j.jofri.2018.08.001>.
62. Ebert LC, Heimer J, Schweitzer W, Sieberth T, Leipner A, Thali M, et al. Automatic detection of hemorrhagic pericardial effusion on PMCT using deep learning-a feasibility study. *Forensic Sci Med Pat*. 2017;13(4):426–431. Available from: <https://doi.org/10.1007/s12024-017-9906-1>.
63. Zhang Z, Ong SH, Zhong X, Foong KW. Efficient 3D dental identification via signed feature histogram and learning keypoint detection. *Pattern Recogn*. 2016;60:189–204. Available from: <https://doi.org/10.1016/-j.patcog.2016.05.007>.
64. Chen M, Shi X, Zhang Y, Wu D, Guizani M. Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE T Big Data*. 2017;(1):1–1. Available from: <https://doi.org/10.1109/-TBDATA.2017.2717439>.
65. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Lect Notes Artif Int*. 2016;3(2):119–131. Available from: <https://doi.org/10.1007/s40708-016-0042-6>.
66. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:171105225*. 2017.
67. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2017. p. 2097–2106.
68. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM: *Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks*. In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE; 2018.
69. shi Zhang Q, chun Zhu S. Visual interpretability for deep learning: a survey. *Front Inform Tech El*. 2018;19(1):27–39.
70. Rebane J, Karlsson I, Papapetrou P. An Investigation of Interpretable Deep Learning for Adverse Drug Event Prediction. In: *IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE; 2019. .
71. Roth HR, Oda H, Hayashi Y, Oda M, Shimizu N, Fujiwara M, et al. Hierarchical 3D fully convolutional networks for multi-organ

- segmentation. arXiv preprint arXiv:170406382. 2017.
72. Wang C, Oda M, Yoshino Y, Yamamoto T, Mori K. Fine segmentation of tiny blood vessel based on fully-connected conditional random field. In: Medical Imaging 2018: Image Processing. vol. 10574. International Society for Optics and Photonics; 2018. p. 105740K. Available from: <https://doi.org/10.1117/12.2293486>.
 73. Zuluaga MA, Magnin IE, Hoyos MH, Leyton EJD, Lozano F, Orkisz M. Automatic detection of abnormal vascular cross-sections based on density level detection and support vector machines. *Int J Comput Ass Rad*. 2011;6(2):163–174. Available from: <https://doi.org/10.1007/s11548-010-0494-8>.
 74. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol*. 2017;52(5):281–287. Available from: <https://doi.org/10.1097/-RLI.0000000000000341>.
 75. Zhen X, Wang Z, Islam A, Bhaduri M, Chan I, Li S. Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. *Med Image Anal*. 2016;30:120–129. Available from: <https://doi.org/10.1016/j.media.2015.07.003>.
 76. Thali MJ, Braun M, Wirth J, Vock P, Dirnhofer R. 3D surface and body documentation in forensic medicine: 3-D/CAD Photogrammetry merged with 3D radiological scanning. *J Forensic Sci*. 2003 November;48(6):1356–1365. Available from: <https://doi.org/10.1520/JFS2003118>.
 77. Gjesteby L, Yang Q, Xi Y, Zhou Y, Zhang J, Wang G. Deep learning methods to guide CT image reconstruction and reduce metal artifacts. In: SPIE Medical Imaging. International Society for Optics and Photonics; 2017. p. 101322W–101322W. Available from: <https://doi.org/10.1117/12.2254091>.
 78. Offiah CE, Dean J. Post-mortem CT and MRI: appropriate post-mortem imaging appearances and changes related to cardiopulmonary resuscitation. *Br J Radiol*. 2015;89(1058):20150851. Available from: <https://doi.org/10.1259/bjr.20150851>.
 79. Iscan MY, Steyn M. The human skeleton in forensic medicine. Charles C Thomas Publisher; 2013.
 80. Hermoza R, Sipiran I. 3D reconstruction of incomplete archaeological objects using a generative adversarial network. In: Proceedings of Computer Graphics International (CGI). CGI 2018. New York, NY, USA: ACM; 2018. p. 5–11. Available from: <https://doi.org/10.1145/3208159.3208173>.
 81. Kettner M, Schmidt P, Potente S, Ramsthaler F, Schrod M. Reverse engineering-rapid prototyping of the skull in forensic trauma analysis. *J Forensic Sci*. 2011;56(4):1015–1017. Available from: <https://doi.org/10.1111/j.1556-4029.2011.01764.x>.
 82. Yu W, Li M, Li X. Fragmented skull modeling using heat kernels. *Graph Models*. 2012;74(4):140–151. Available from: <https://doi.org/10.1016/j.gmod.2012.03.011>.

Femur localisation and structured unsupervised labelling in CT scans using machine learning

Semi-supervised labelling of forensic CT data using deep learning

C. A. Peña-Solórzano, D. W. Albrecht, P. C. Harris, R. B. Bassed, J. Gillam, M. R. Dimmock

Submitted to Forensic and Legal Medicine, November, 2019

This chapter is an exact copy of the paper submitted

This chapter is divided into two subsections. The first subsection describes the development of a tool for localising arbitrary anatomical structures using DL. Localisation is used as a pre-processing step where the individual voxels are not required to be classified, e.g. to determine the presence of the organ, or to reduce the computation required in a subsequent task, e.g. volume calculation or tissues segmentation. The system was trained using 350 full-body CT scans from the VIFM database with a large variability in terms of BMI, time since death, incompleteness of lower limbs, presence of metallic implants, trauma, and the presence of foreign objects. Three separate DNNs, more specifically residual neural networks (ResNets), were created and linked for parallel execution in order to determine the presence (or absence) of the femur in a given slice of each of the three orthogonal views (axial, coronal and sagittal) defined by the reconstructed CT scan volume. The predictions of the networks were then subsequently merged in order to generate the 3D bounding box for the organ in the scan from each case. The femur served as an interesting test case

for the implementation of localisation with a bounding box; the angular variation in positioning of the femur was challenging when compared to other more static organs in the body. However, as the pipeline was intended to be generalisable, it was deemed this was the simplest approach that gave the desired flexibility for the proposed application.

The second subsection details the automated approach to classification of the contents of the femur inside the bounding box as one of four categories: without-implant, nail, hip replacement, or knee replacement. A DNN architecture known as CAE was utilised. The network took the bounding box as input and reconstructed it in the output following both encoding (compression) and decoding (expansion) steps. The output of the encoder and input of the decoder was a compressed vector (latent space) that represented the original input. Additionally, a branch that used the latent space vector for classification was added to leverage additional case information known about the decedent; in this case the age. Both the network with and without the classifier branch were tested, resulting in an improved structure of the latent space and the clustering performance for separating features in the CT volumes when the additional information was utilised.

3.1 Publication

The following paper has been submitted to Forensic and Legal Medicine and is currently under review for publication.

Semi-supervised labelling of forensic CT data using deep learning

Abstract

In order to optimally leverage the information contained in the existing extensive databases of post-mortem computed tomography scans for medico-legal investigations, approaches for automating the labelling of key anatomical and pathological features are required. We have developed a general deep learning pipeline and tested it on classifying metal implants in the femur as a proof-of-principle for labelling content in CT volumes not described in the autopsy reports. A pipeline that used both residual networks and an auto-encoder was developed for automatic organ localization and classification. The pipeline was trained and tested using $n = 350$ full-body PMCT scans. For the localization stage, the ground-truth labels were determined by two medical imaging technologists and validated by two medical specialists. A regression analysis was used to determine significant factors that affected performance. For the classification, an additional $n = 100$ PMCT scans were used to determine the precision and recall rate. For the femur localization task, Dice scores of 0.99, 0.96, and 0.98 and mean absolute errors of 3.2, 7.1, and 4.2 mm were obtained in the axial, coronal, and sagittal views, respectively. Test cases were properly labelled as without-implant (Γ), nail, hip replacement, or knee replacement with an accuracy $>97\%$, where the recall for Γ and hip replacement cases was 1.00, but falling to 0.82 and 0.65 for cases with knee replacements and nails. The semi-automatic technique presented here provides a generalized structure for the efficient labelling of image-based content not reported by the pathologist. The ability to search for and label structures automatically has many applications in accumulating data to answer medico-legal questions with population-level data in forensic radiology.

Keywords: Autoencoder; deep learning; machine learning; organ localization; labelling; CT

Introduction

The ever-increasing number of images and autopsy reports acquired by forensic institutions provides rich opportunities for developing methods that automate the annotation process to unlock their value for answering medical and forensic questions. These databases capture decedents that range in age from 0 to 100+, and incorporate all causes of death, both natural and unnatural – including homicide, and deaths due to accidents and suicide. The utilization of machine learning (ML) to develop new tools for improving workflow efficiency when interrogating the databases requires natural language processing (NLP) to interpret and code medical reports and legal documents, and the use of deep learning (DL) techniques to process images of anatomy, pathology and histology. The training of these NLP/DL algorithms requires detailed and accurate labelling to ensure high sensitivity and specificity and to avoid potential biases. The

automatic labelling of the large numbers of images, which will otherwise take enormous resourcing in terms of person-hours, will significantly reduce the burden on medical experts to provide structure to the databases. The structure is required in order to provide supporting improved evidence to the coroner for death investigations and the police for the crime scene reconstructions [1], as the searchable text only covers a limited number of anatomical features reported in any given pathology report.

Whilst the automation of image processing tasks has been widely reported in the clinical setting for key modalities including plain film X-ray [2, 3], computed tomography (CT) [4-6], and magnetic resonance imaging (MRI) [7-9], there is a dearth of literature reported in the forensic setting [10-15]. The application of medical imaging techniques in forensic investigations raises additional complications to those experienced

clinically, such as decedent decomposition, variability in constitution and positioning of normal anatomical structures resulting from cause of death (e.g. incineration, severe trauma), artefacts from foreign bodies and anatomical positioning due to rigor mortis. In this work we developed a machine-learning pipeline to assist in labelling images in a forensic database containing CT reconstructions, adding increased structure for content-based retrieval. The pipeline utilized both supervised and unsupervised learning, with an initial localization step that defined a bounding box [16-21] around the structure of interest – the femur. After the skeletal element was located, a DL [22] data representation framework was used to identify patterns and relationships associated with metal implants. The results were then contextualized with respect to the available forensic information obtained from the associated pathology reports.

Materials and methods

A total of $n = 350$ whole-body CT scans were obtained retrospectively from the Victorian Institute of Forensic Medicine (VIFM) (ethics approval EC 17/2017). The decedents were selected consecutively from the available 2017 dataset until 50 males from each of seven age groups (20-90 years of age in 10-year increments) were obtained. Cases were included if the cause of death was classified in the autopsy report as accidental, natural or suicide. All decedents could therefore be described as having died due to non-physical trauma.

The post-mortem CT (PMCT) scans were those acquired during the triage process at admission to the VIFM. All PMCT scans were acquired using the same SOMATOM Definition Flash (Siemens Healthcare) scanner configured to operate in dual-source mode at 120 kVp. Each of the reconstructed volumes comprised ~ 1700 contiguous axial slices with 512×512 voxels with 1 mm^3 isotropic resolution. The digital imaging and communications in medicine (DICOM) files were de-identified in accordance with the ethics approval and exported anonymously. The resulting images were then introduced into the

image processing pipeline in order to audit for the presence of different orthopedic implants in the femur as a proof-of-principle validation. The pipeline comprised three steps (see Fig. 1): firstly pre-processing; secondly 3-dimensional (3D) anatomical localization; and finally, semi-automatic classification.

Pre-processing

As all CT reconstructions in the database were acquired with the same scanner, reconstruction technique and parameters, it was determined that the algorithms developed as part of this research project would operate on fixed size volumes in order to simplify the implementation. Each reconstructed volume was padded to $2048 \times 512 \times 512$ (axial \times coronal \times sagittal). The axial slices required for padding (Fig. 2c) were synthetically generated by sampling the first and last slices of each scan (Fig. 2a), generating histograms, fitting the noise profiles (Fig. 2b), and then subsequently randomly resampling according to the aforementioned fit parameters. Commodity graphics processing units (GPUs) were used as a part of the investigation and the limited memory capacity required that the padded volumes were then linearly down-sampled by a factor of two in the axial, coronal, and sagittal directions to $1024 \times 256 \times 256$ voxels (2 mm^3 isotropic resolution). This resampled data were then introduced into the preliminary organ localization algorithm.

Anatomical localization

Two human readers (medical imaging technicians) followed the same multi-step protocol to annotate the data ($n = 350$ CT volumes) and determine the look-up table (LUT) that was subsequently used to train the neural network:

1. display the slices in the axial profile with a window-width (WW) and window-level (WL) of 2000 HU and 0 HU, respectively, to determine the first and last slices that contain femur bone; then

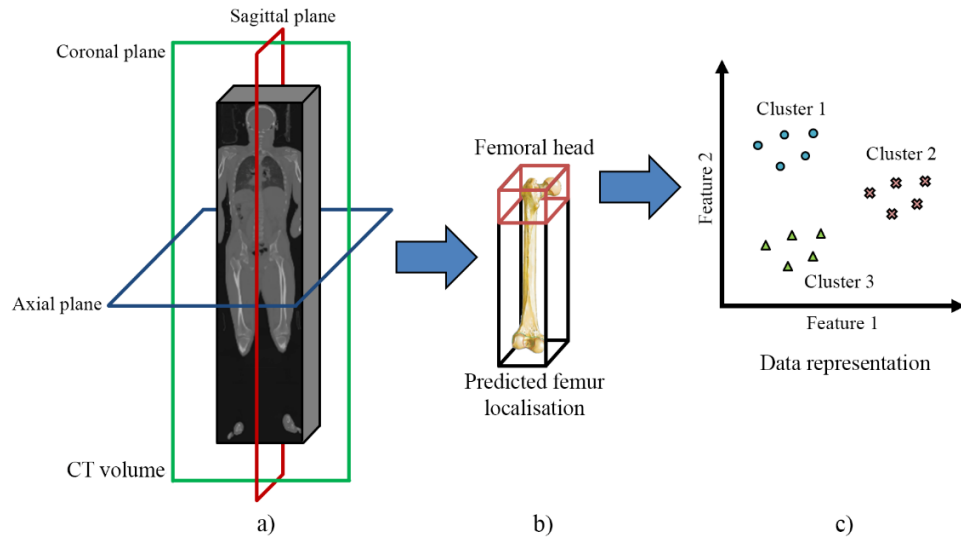


Fig. 1. A schematic view of the image processing pipeline developed for automatic labelling of the femur. a) Each CT scan was processed on each orthogonal view. b) The predictions were brought together to determine the 3D localization of the femur. c) The femoral head was processed through an automatic labelling technique.

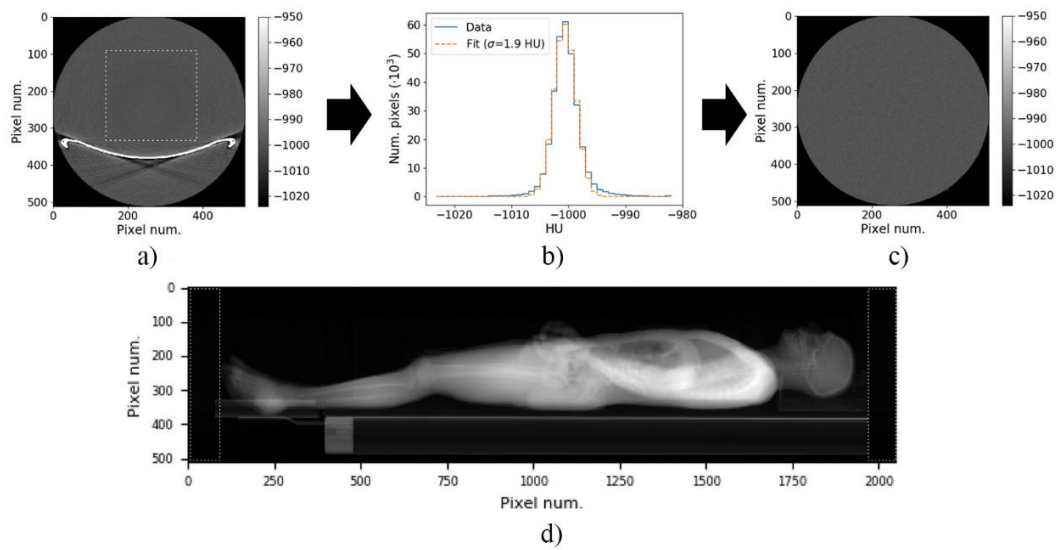


Fig. 2. Padding of the volume on the axial axis. a) Axial slices without the presence of biological tissues were sampled. b) Histogram and fitting obtained for dotted square shown in a). c) Sample slice generated from the obtained distribution and used as padding. d) The generated slices were appended to the axial length of the volume, as indicated by the dotted lines.

2. display the slices in the coronal profile with a WW and WL of 2000 HU and 0 HU, respectively, to determine the first and last slices that display bone; and finally,
3. display the slices in the sagittal profile with a WW and WL of 2000 HU and 0 HU,

respectively, to determine the first and last slices that display bone.

A subset of $n = 20$ CT volumes was then selected for cross-checking by an orthopedic surgeon and odontologist (with 12 and 13 years of experience with CT imaging, respectively). The interrater

reliability was determined through the calculation of an intraclass correlation coefficient (ICC), using a two-way mixed model and absolute agreement [23]. The unlabeled images and LUT were stored on the hard disk for access by the network. All statistical analyses were performed using SPSS version 25.0 (IBM, Armonk, NY, USA).

In this work we used a localization technique that relied on determining the bounding box for the anatomy of interest (femur) in the whole-body CT scan using 3D sub-volumes viewed in the axial, coronal and sagittal planes (see Fig. 3). The implementation required the training of three residual convolutional neural networks (ResNets), one for each orthogonal view. One output class was defined for each network – indicating the presence (1) or absence (0) of the femur in the current slice under consideration in the current direction of visualization. In order to further improve the results, several post-processing steps were implemented [19]. Firstly, the 1D predictions obtained from each orthogonal view were smoothed using a Gaussian filter with a standard deviation (σ) of 10 slices. Secondly, a threshold of 0.5 was defined to convert the filtered output to binary values. Thirdly, connected-component analysis was performed to choose the largest component, dismissing the rest. Finally, the binary outputs of the three networks were joined to obtain the predicted 3D bounding as presented throughout the remainder of this work.

The $n = 350$ CT volumes were split 70:15:15 (as a percentage) for training, validation and testing of performance of the ResNets, respectively. The training set contained 244 scans, while the validation and test sets contained 53 scans each. In order to ensure batch normalization was performed correctly, the Hounsfield unit (HU) values in every CT volume were rescaled by subtracting the mean (μ) and dividing by the σ of the training dataset at initialization.

ResNets are based on residual learning with the introduction of residual blocks [24, 25]. The improved performance of ResNets, when compared to earlier networks, e.g. VGG [26],

arises due to the skip connection that adds the output from the previous layer to the layer ahead. The ResNet stage was defined as two blocks of sequential layers of convolution, batch normalization, activation, convolution, and addition with the skip connection. Fig. 4 shows the architecture utilized for the axial network, consisting of three ResNet stages and 24 filters for the first convolution operations. For the coronal and sagittal networks, one extra ResNet stage and 34 filters were used on the first convolution due to the difference in input sizes. The input to the ResNet for each calculation was the slice of interest in the current view, with the eight surrounding slices – the four immediately preceding and the four immediately following; for example, if slice 27 in the axial direction was under consideration for the presence or absence of a femur, axial slices 23 to 31 inclusive were processed to give volumetric context [21, 27-29]. The slice that was being analyzed was therefore located in the middle of a sub-volume supplied as the input. During the training stage, the LUT was supplied to indicate whether or not the network should detect a femur in the central slice of the current sub-volume.

The ResNets used He-normal and Ridge regression (L_2) as initializer and regularization techniques, respectively, and were trained using random sub-volumes selected from the $n = 350$ CT volumes over five sequential runs in order to maximize statistical strength (similar to a 5-fold cross validation). The batch sizes were 64, 32 and 32 for axial, coronal and sagittal networks, respectively. The difference in batch sizes was due to the differences in shapes of the sub-volumes for the different orthogonal views. During training, data augmentation consisting of random rotations and scaling was applied to reduce overfitting. The possible rotations ranged from -20° to $+20^\circ$ and scaling factors ranged from 0.8 to 1.2. The output layer presented a sigmoid activation, and the loss was calculated using binary cross-entropy. Methods for early stopping and reduction of the learning rate on plateau were used during training to improve the resulting model.

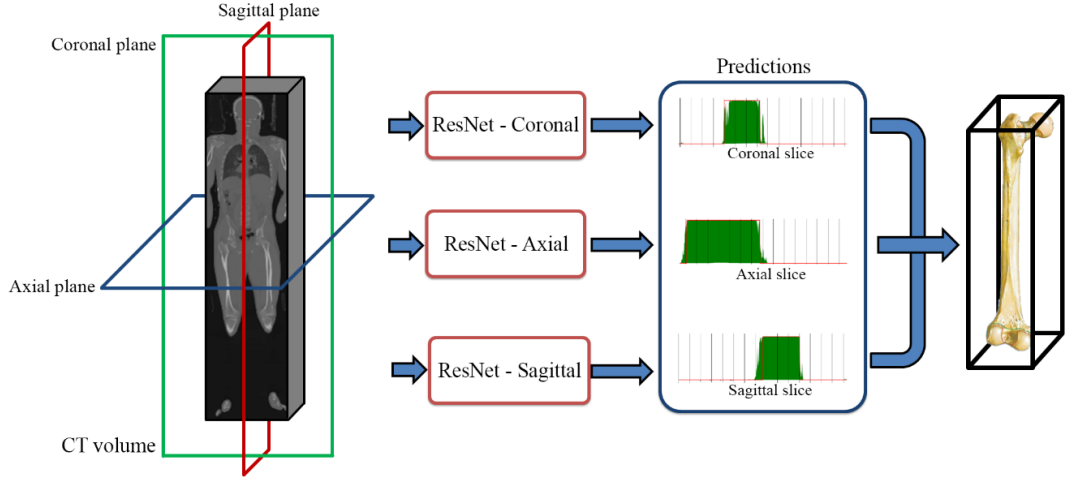


Fig. 3. Diagram of the approach to obtain 3D bounding boxes for femur localization from 3D whole body CT scans.

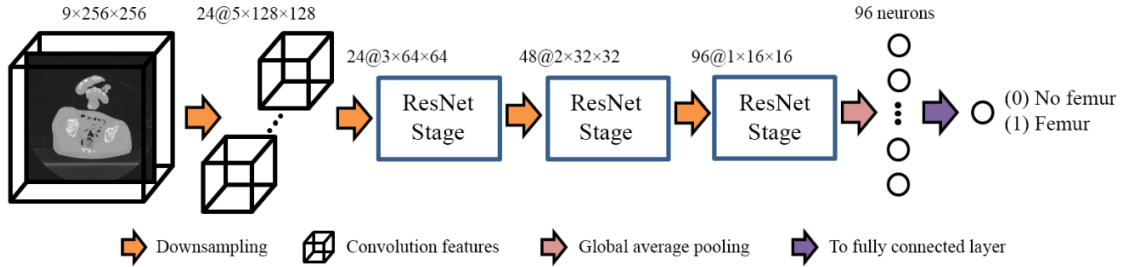


Fig. 4. ResNet architecture utilized for organ localization. For the axial view shown, the sub-volume of the CT scan analyzed for whether its central slice contained a femur was $9 \times 256 \times 256$ voxels. For the coronal and sagittal views, the sub-volumes were $1024 \times 9 \times 256$ and $1024 \times 256 \times 9$, respectively.

In order to evaluate the performance of the organ localization, the predicted 3D bounding boxes were compared to the ground truth (the interrater reliability is presented with the results) using the mean absolute error (MAE), the Dice score (also called the F_1 score) and Jaccard similarity coefficient (also known as intersection over union or IoU). In case of disagreement between raters, the larger box is selected to ensure the complete presence of the femur in the volume. The mean error refers to the difference between the predicted and the real wall slice locations as a physical distance in the CT scan. The Dice score and the Jaccard similarity coefficient are commonly reported in the literature and are defined as [19],

$$DS(A, B) = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN}, \quad (\text{Eq. 1})$$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN}, \quad (\text{Eq. 2})$$

where A and B are the ground truth and predicted volumes, while TP , FP , and FN represent the true positives, false positives, and false negatives, respectively.

A regression analysis was performed to determine which factors were significant in affecting the performance of the localization algorithm. The effect on the MAE (dependent variable) for each of the three views (axial, coronal, and sagittal) and for the combined results for the bounding box was investigated for a total of 11 independent variables:

- age – presented as continuous variable;
- body mass index (BMI) – calculated as a continuous variable;

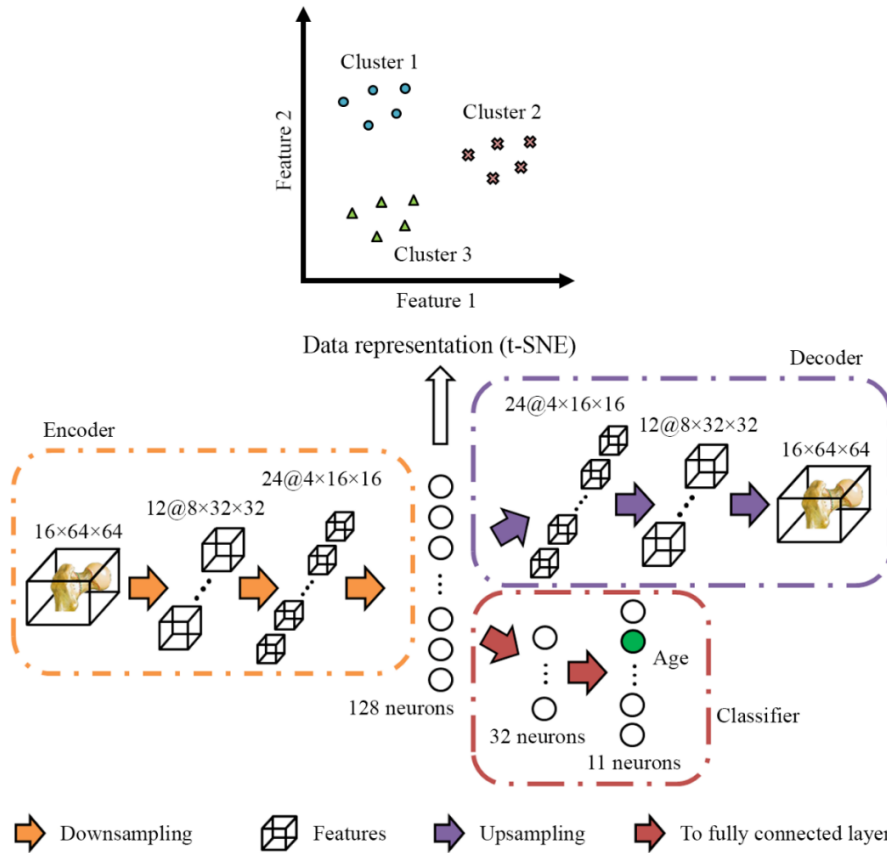


Fig. 5. Diagram of the approach for feature extraction of the femoral epiphyses. The encoder converted the epiphysis to an ND representation that was used for clustering and data analysis. The decoder was an unsupervised stage that ensured the latent space was a good representation of the input. The classifier, branched from the latent space, allowed the system to leverage extra information to improve clustering performance.

- time since death – calculated as a continuous variable showing the number of hours between death and the triage CT scan;
- incomplete femur – binary True in the absence of any section of the femur from the field of view (FoV) of the CT scan;
- incomplete long leg bones – binary True in the absence of any section of lower limb long bone from the CT scan.
- metal on primary femur – binary True in the presence of any metal implant in the primary femur (that being localized);
- metal on secondary femur – binary True in the presence of any metal implant in the secondary femur;
- significant trauma – binary True in the presence of any discontinuity that breaks the femur into more than one section;
- foreign object – binary True in the presence of any non-anatomical object in the body bag;
- rotation – the angle in degrees of the femur from the <001> unit vector that represented the direction of motion of the patient bed; and finally,
- bed – binary True in the presence of the patient bed in the FoV of the CT scan.

Feature classification

Once the femur was localized using the aforementioned ResNet approach, the proximal and distal epiphyses were separated into two $16 \times 64 \times 64$ voxel sub-volumes for feature extraction and classification. The feature extraction was performed by a hybrid convolutional autoencoder (HCAE), where the

clustering of the latent vector space determined by the HCAE was performed using a k -NN analysis.

The traditional implementation of a CAE takes a 2D image as input and reconstructs it in the output following both encoding (compression) and decoding (expansion) steps (see Fig. 5). The output of the encoder and input of the decoder is a compressed vector (latent space) that represents the original input. In this work, the CAE was modified to receive a 3D volume as input. Also, an additional branch that uses the latent space vector for classification was added to leverage additional case information known about the decedent, in this case age was used. This additional step allowed the HCAE to add structure to the latent space which improved clustering. Both the network with and without the classifier branch were tested. Fig. 6 shows three axial slices of a proximal epiphysis determined from the organ localization step being fed into the HCAE. The result of the encoder was the N-dimensional (ND) latent vector space with clusters determined from features detected by the network.

The HCAE was trained using sets with a batch size of 128. Data augmentation analogous to that previously described was also implemented. The activation layers for the decoder and classifier outputs were sigmoid and softmax functions, respectively, while the losses were mean squared error (\mathcal{L}_r) and categorical cross-entropy (\mathcal{L}_c) for the reconstruction and classification. The total loss (\mathcal{L}_t) was a mixture of the two losses with a balancing factor, where $\mathcal{L}_t = \mathcal{L}_r + \lambda \cdot \mathcal{L}_c$, with $\lambda = 0.01$ was chosen empirically after analyzing the magnitude of both loss values.

The ND vectors obtained from the bottleneck were located in the latent space. For data analysis and visualization, t-Distributed Stochastic Neighbour Embedding (t-SNE) was utilized [30]. t-SNE is a widely used nonlinear dimensionality reduction technique for embedding high-dimensional data in a 2D or 3D space. This technique has been shown to recover well-separated clusters [31]. For this work, the output visualization was performed in 2D and the axes will be referred as $z[0]$ and $z[1]$.

The 2D output was post-processed using k -NN [32] in order to partition the observations into k clusters, where the value of k was defined after analyzing the 2D visualization and determining the information that was encoded in the clusters. To present the results more clearly, a confusion matrix was constructed to indicate the performance of the classification model when known features were selected as the true values for the clusters. Using the confusion matrix, several performance metrics were calculated: the accuracy, defined as the ratio of the number of correct classifications to the total number of classifications; the precision, measured as the ratio of the TPs of the class in question to the sum of its TPs and FPs; TP-rate or recall, calculated as the ratio of the TPs of a specific class to the sum of its TPs and FNs; and finally, the F1 score, which conveys the balance between the precision and the recall. At this final stage, the database of $n = 350$ cases was augmented with an additional 100 cases with verified presence of metallic implants.

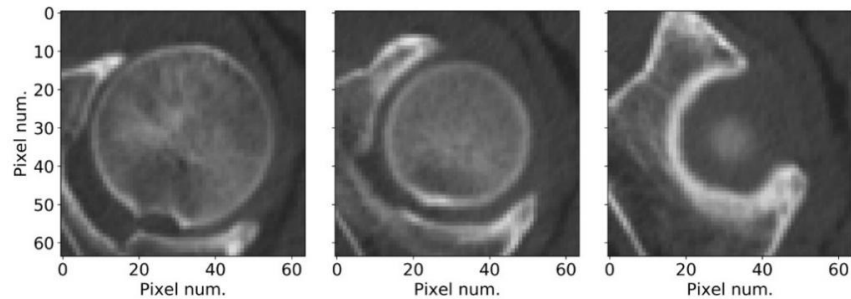


Fig. 6. First (left), middle (center), and last (right) slices of an arbitrarily selected proximal epiphysis sample.

Results

Femur localization

The reproducibility analysis showed an ICC(3,1) >0.99 for the axial, coronal, and sagittal femur localizations, indicating an excellent interrater reliability. The MAEs between mean of the medical imaging technicians and the experts were 1.0 mm, 1.4 mm and 1.5 mm for the axial, coronal and sagittal directions.

Table 1 presents the results from the performance metrics selected for the organ localization task. The results were averaged over the test set of each of the five executions of random processing of subsets with the algorithm. For the Jaccard coefficient and the Dice score, a value of 1.0 indicates a perfect match between the predicted box and the defined ground truth. The MAE measures the wall distance in mm between the two boxes and is 0.0 in cases where the walls are at the same locations. The reader should note that the due to the down-sampling of the data, a prediction error of one slice in any direction gave rise to a MAE of 2 mm.

Table 2 presents the results from the regression analysis demonstrating the significance of factors that affected the calculation of the MAE for each of the three views (axial, coronal, and sagittal) and for the combined results for the bounding box. Because the MAEs between mean of the medical imaging technicians and the experts were less than one slice, cases with a MAE ≥ 2 mm, were

selected for the regression analysis to attempt to explain the remaining variance.

Data representation of femoral epiphyses and knees

Fig. 7 and Fig. 8 present the 2D representation of the latent space for all $n = 450$ cases (the same 350 cases used for femur localization plus 100 cases with high incidence of presence of metallic implants) without and with the use of additional age information, respectively. Each data point was color coded according to age and assigned a “Y” symbol with 90° clockwise rotations to differentiate left proximal femoral epiphysis, left distal femoral epiphysis, right proximal femoral epiphysis, and right distal femoral epiphysis, respectively. Regions that constitute the presence of a hip replacement, nail, or knee replacement are highlighted.

Table 3 and Table 4 present the confusion matrices demonstrating the performance of the classification technique for the CAE and HCAE networks, respectively, where $k=3$. The algorithm was executed 8 times with random subsets to increase statistical power (similar to 8-fold cross validation). The overall accuracy (OA) was 97.7% and 99.1% for the networks without and with the use of age information, respectively. Accuracy, precision, recall, and F1 score are reported for the classes defined: those without implants (I-), those with nails in the femoral head, hip replacements, and knee replacements.

Table 1

Average Jaccard similarity coefficient, Dice score, and mean absolute error for the five executions of femur localization on random test cases.

	Axial ($\mu \pm \sigma$) [avg. length]	Coronal ($\mu \pm \sigma$) [avg. length]	Sagittal ($\mu \pm \sigma$) [avg. length]
Jaccard coefficient	0.99 \pm 0.02	0.93 \pm 0.06	0.96 \pm 0.05
Dice score	1.00 \pm 0.01	0.96 \pm 0.03	0.98 \pm 0.03
MAE [mm]	3.2 \pm 3.9 [459.0]	7.1 \pm 6.0 [95.9]	4.2 \pm 4.2 [117.4]

Table 2

Regression results showing p-values obtained after training the femur localization algorithm and analyzing the MAEs for cases with $MAE \geq 1$ slice (2 mm), $r^2 = 0.26$.

	p-value			
	All	Axial	Coronal	Sagittal
Age	0.64	0.78	0.29	0.41
BMI	0.07	0.09	0.11	0.59
Time since death	0.65	0.94	0.76	0.26
Incomplete primary femur	0.17	0.08	0.45	0.77
Incomplete other long leg bones	0.17	0.81	0.34	0.39
Metal on primary femur	<0.001***	<0.05*	<0.05*	0.31
Metal on secondary femur	0.11	0.78	0.10	0.32
Significant trauma	0.49	0.22	0.39	0.23
Foreign object	0.64	0.96	0.76	0.64
Rotation	<0.01**	0.19	0.60	<0.01**
Bed	0.57	0.57	0.68	0.33

Table 3

The confusion matrix showing the performance of the clustering technique when the age information was not used.

		Predicted label				Acc	Precision	Recall	F1 Score
		I ⁻	Nail	Hip rep.	Knee rep.				
True label	I ⁻	2400	8	0	2	0.98	0.98	1.00	0.99
	Proximal Nail	20	19	7	0	0.99	0.70	0.41	0.52
	Hip rep.	8	0	46	0	0.99	0.87	0.85	0.86
	Knee rep.	14	0	0	3	0.99	0.60	0.18	0.27

Table 4

The confusion matrix showing the performance of the clustering technique when the age information was used.

		Predicted label				Acc	Precision	Recall	F1 Score
		I ⁻	Nail	Hip rep.	Knee rep.				
True label	I ⁻	2400	3	0	0	0.99	0.99	1.00	0.98
	Proximal Nail	15	31	1	1	0.99	0.91	0.65	0.76
	Hip rep.	0	0	61	0	0.97	0.97	1.00	0.98
	Knee rep.	0	0	1	16	1.00	0.94	0.94	0.94

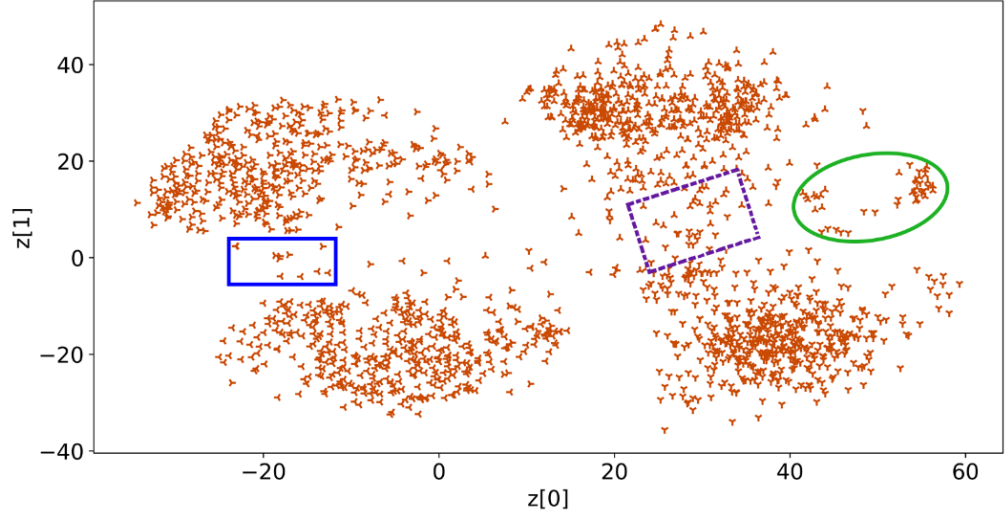


Fig. 7. A 2D data representation of the latent space determined by the CAE after applying t-SNE. The age information was not used. Both training and test cases are plotted. The solid ellipse (green), dashed box (purple), and solid box (blue) represent regions that constitute the presence of a hip replacement, nail, or knee replacement, respectively.

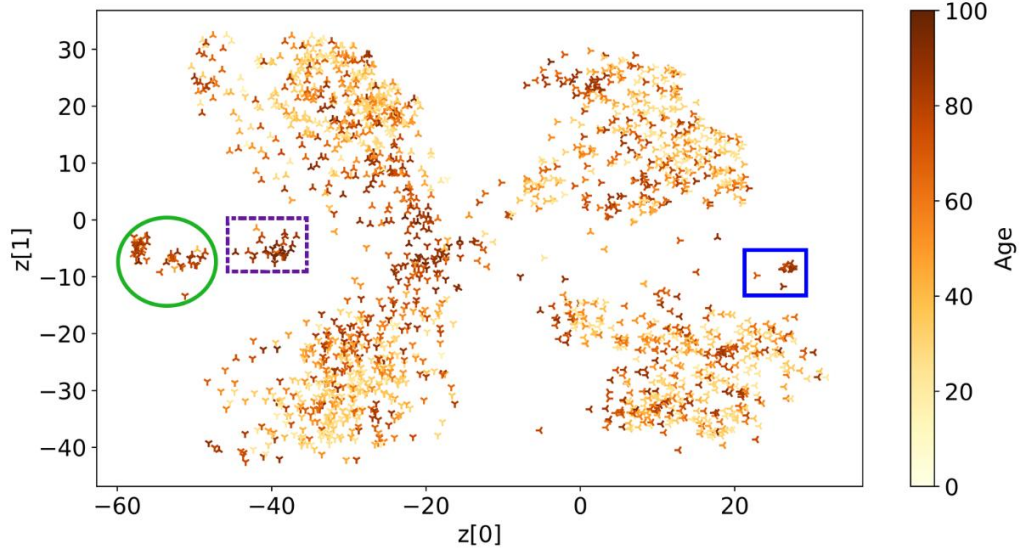


Fig. 8. A 2D data representation of the latent space determined by the HCAE after applying t-SNE. Age information was used to provide structure. Both training and test cases are plotted. The solid ellipse (green), dashed box (purple), and solid box (blue) represent regions that constitute the presence of a hip replacement, nail, or knee replacement, respectively.

Discussion

The femur localization was performed by combining the output of three different networks, where the coronal and sagittal architectures shared the same number of layers. The difference in the size of the networks was related to the size of the input volumes. This size also affected the number of sub-volumes that were obtained from

each case, with the axial network being trained with four times more sub-volumes per case than the sagittal and coronal networks due to the geometry of the reconstructed volumes. In practice, the difference in the number of sub-volumes used for training affected the performance of the localization algorithm, as can be observed in Table 1, where the performance of

the axial network was best for all three defined metrics. In previous works [16, 19, 21, 33], the localization of the femoral head within clinical abdominal CT scans using the bounding box approach has been reported to have errors in the range of 1 to 11 mm. Our results are considered to be consistent with these when allowing for the fact that a patient in a clinical scanner is either manipulated or instructed to lay flat, and the patient bed does not typically include the foreign objects present in forensic scans. Furthermore, the difficulty caused by the variability in the positioning and cause of death is exacerbated by the localization of the whole femur in comparison to other works that only localized the femoral head. Due to this, the MAEs must be analyzed in the context of the mean average size of the box in each direction, especially in the coronal and sagittal views, where MAEs of 7.1 and 4.2 mm are measured with respect to mean lengths of 95.9 and 117.4 mm. This effect can also be observed when considering the MAE between the experts and the medical imaging technicians, with a value almost three times the value reported by Humpire-Mamani et al. [19]. As a visual contrast of the aforementioned differences, Fig. 5 in Humpire-Mamani et al. [19] can be compared to Fig. 9 in this work.

Table 2 demonstrates that the rotation angle of the femur with respect to the scanner bed axis of movement and the presence of metallic implants had statistically significant effects on the performance of the femur localization. The factors referred to in Table 2 were able to explain 26% of the variance for cases with a MAE worse than that obtained by the human readers, with other forms of rotation (pelvis, legs crossing) that were difficult to calculate being suspected of accounting for the remaining observed error. Fig. 9 shows coronal maximum intensity projections (MIPs) that visually demonstrate the key independent variables tested for (other than BMI). The rotation of the femur (Fig. 9a) had a statistically significant effect in the sagittal network, where the view of the angled bone affected the capabilities of the network to

properly classify the slice. The incompleteness of the femur, the presence of significant trauma, the presence of foreign objects, and the presence of the scanner bed in the FoV (see Fig. 9a-d) did not significantly affect the performance of the algorithm in any single orientation and validated the use of the discrete processing of the different views in order to localize the anatomy of interest as a preliminary pipeline step.

After rotation of the femur, the most significant effect in identifying the location of the femur was the presence of metal implants, e.g. hip replacement, surgical nails. This effect can be more clearly observed in the axial images (see Fig. 10a) and is caused by the beam hardening and photon starvation [34]. The most severe form of metal artefacts, that occurs when metal couples with other metal or dense structures such as bone, is of significant interest in forensic CT scans; the requirement for the preservation of evidence or presence of rigor-mortis can mean that repositioning of the body is not feasible. Indeed, this effect can manifest in a range of locations, e.g. the inability to raise the decedent's arms above their head introduces artefacts that distort the HUs of organs in the thorax and fluid in the lungs (see Fig. 10b). Whilst these effects were not explicitly accounted for in this work, DL approaches have been shown to enable reductions in metal streak artefacts [35].

The HCAE component of this work was configured to analyze 3D sub-volumes. Whilst this approach is more computationally expensive than processing 2D images, our aim was to define an automated pipeline that would ultimately be more general than simply auditing the presence of different types of metallic orthopedic implants. In radiology, 3D representations of anatomy are generally preferred due to the enhanced contrast that arises from non-overlapping structures inherent in 2D imaging. From Fig. 9, it can be argued that the use of extra dimensionality is more beneficial as the presence of a wide range of foreign objects makes interpretation of the contents of the body bag more complicated.

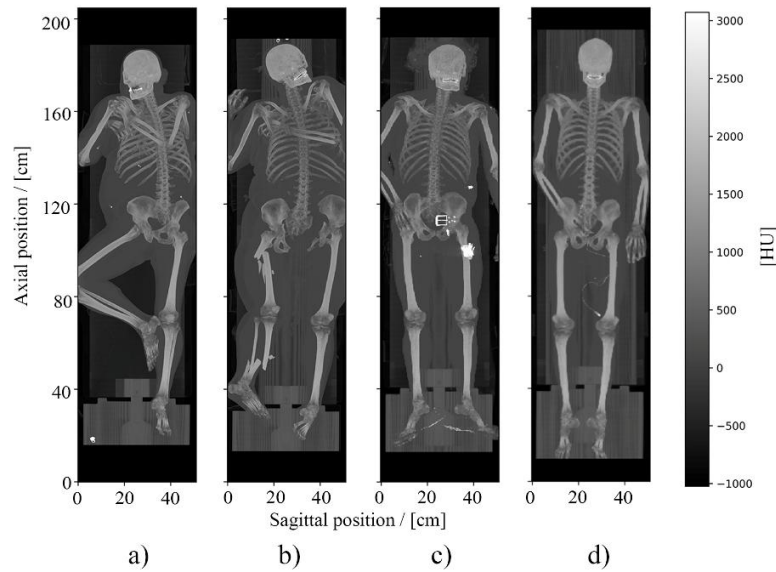


Fig. 9. a-d) Maximum intensity projections (MIPs) generated from the triage CT scans of four representative cases from the PMCT database. a) Limb outside of FoV. b) High-force fracture. c) Presence of metal objects both in front of and behind the body. d) Presence of dense surface of the scanner bed in the FoV.

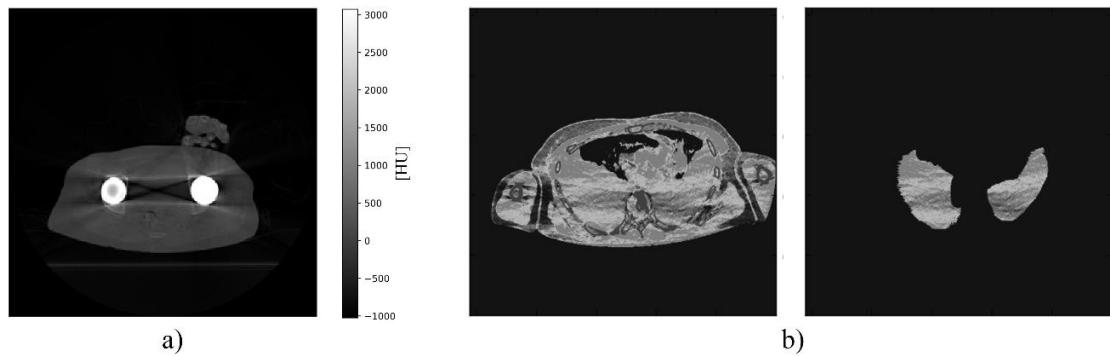


Fig. 10. a) Axial slice for case with bi-lateral total hip replacement. b) Lung HU values being affected by the coupling between the bones in the arms located along the thorax.

In the discipline of Orthopedics, implants in the long leg bones can be classified into three categories: screws or nails, knee replacements, and hip replacements. The presented pipeline was able to cluster cases according to the type of metallic implant without any information relating to these variables during training. Furthermore, the use of age as an output ensured that only the imaging information needs to be used during the testing phase and yielded an increase in the overall accuracy and recall value for the nail and knee replacement classes (Table 3 and Table 4). The clustering algorithm was able to differentiate

the cases with hip replacements, even when the database presented cases with hemiarthroplasty or total hip arthroplasty, which constitute different shapes in the CT scan.

The pipeline presented can be used separately, with the organ localization stage being an important pre-processing step for a number of imaging tasks. It must be noted that the down-sampling of the original volumes affected the performance of the localization due to partial volume effects [36] and loss of resolution. Results could be improved by using the high-resolution

images, having considered the computational costs that this implies. Semi-automatic techniques such as that presented here for the labelling of vast databases could help with finding patterns in the data to improve content-based image retrieval when answering medico-legal questions that arise during forensic investigations.

Conclusion

In this paper we presented a workflow that can be used to automate the labelling of CT images in a forensic database and applied it to the femur. The initial processing step used ResNets to predict the localization of the desired body part, in this case the femur. Dice scores of 1.00, 0.96, and 0.98 and wall distances of 3.2, 7.1, and 4.2 mm were calculated for the localization in the axial, coronal, and sagittal views, respectively.

After localization of the femur, sub-volumes were parameterized using an HCAE for feature extraction. Analysis of the resulting clusters demonstrated the ability to separate left and right I femurs and knees, and the presence of different types of metal implants. Test cases were properly labelled as I, nail, hip replacement, or knee replacement with an accuracy >97%, where the recall for I and hip replacement cases was 1.00, but falling to 0.82 and 0.65 for cases with knee replacements and nails, due to the variability in the number of slices containing metal present in the sub-volume.

Limitations

The use of information different to age during the training of the clustering stage, e.g. height, weight, time since death, could improve the performance of the k -NN technique, but special consideration must be given to the pertinence of the utilized variable and its relationship to the body part being investigated. Only males were included in this work in order to improve statistical power by removing a variable, but the femur localization results were not significantly affected by biological variables. It is expected that the presented technique present favorable results when new variables such as sex or reconstructions

protocols are used. The proposed pipeline should be able to generalize to other clinical/forensic environments, with the use of a normalization technique that keeps the pixel value distribution similar to that of the data used during training.

The performance of the presented pipeline could be improved by utilizing transfer learning to take advantage of networks that have been previously trained in databases with millions of images. The use of this method provides a baseline performance that can be fine-tuned for the specific application, potentially decreasing training time and errors.

References

1. Zhou Y, Zhang J, Huang J, Deng K, Zhang J, Qin Z, et al. Digital whole-slide image analysis for automated diatom test in forensic cases of drowning using a convolutional neural network algorithm. *Forensic Sci Int*. 2019 sep;302:109922.
2. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*. 2017;284(2):574–582.
3. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In: *Proc CVPR IEEE*. IEEE; 2017.
4. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing Machine Learning in Radiology Practice and Research. *Am J Roentgenol*. 2017;208(4):754–760.
5. Singh G, Al'Aref SJ, Assen MV, Kim TS, van Rosendael A, Kolli KK, et al. Machine learning in cardiac CT: Basic concepts and contemporary data. *J Cardiovasc Comput*. 2018;12(3):192–201.
6. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer*. 2018;115:34–41.

7. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pinykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology*. 2018;288(2):318–328.
8. Isn A, Direkoglu C, Sah M. Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods. *Procedia Comput Sci*. 2016;102:317–324.
9. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2017;19(6):1236–1246.
10. Arigbabu OA, Liao IY, Abdullah N, Mohamad Noor MH. Computer vision methods for cranial sex estimation. *IPSJ Trans Comput Vis Appl*. 2017;9(1):19. Available from: <https://doi.org/10.1186/s41074-017-0031-6>.
11. Ebert LC, Heimer J, Schweitzer W, Sieberth T, Leipner A, Thali M, et al. Automatic detection of hemorrhagic pericardial effusion on PMCT using deep learning-a feasibility study. *Forensic Sci Med Pat*. 2017;13(4):426–431. Available from: <https://doi.org/10.1007/s12024-017-9906-1>.
12. Curate F, Umbelino C, Perinha A, Nogueira C, Silva AM, Cunha E. Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers. *J Forensic Leg Med*. 2017;52:75–81.
13. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K. Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study. *J Forensic Leg Med*. 2018;57:41–50.
14. Heimer J, Thali MJ, Ebert L. Classification based on the presence of skull fractures on curved maximum intensity skull projections by means of deep learning. *J Forensic Radiol Imaging*. 2018;14:16 – 20. Available from: <https://doi.org/10.1016/j.jofri.2018.08.001>.
15. Bewes J, Low A, Morphet A, Pate FD, Henneberg M. Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. *J Forensic Leg Med*. 2019;62:40–43.
16. Criminisi A, Robertson D, Konukoglu E, Shotton J, Pathak S, White S, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med Image Anal*. 2013 dec;17(8):1293–1303.
17. Afshari S, BenTaieb A, Hamarneh G. Automatic localization of normal active organs in 3D PET scans. *Comput Med Imag Grap*. 2018;70:111–118.
18. de Vos BD, Wolterink JM, de Jong PA, Leiner T, Viergever MA, Isgum I. ConvNet-Based Localization of Anatomical Structures in 3-D Medical Images. *IEEE T Med Imaging*. 2017;36(7):1470–1481.
19. Humpire-Mamani GE, Setio AAA, van Ginneken B, Jacobs C. Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans. *Phys Med Biol*. 2018;63(8):085003.
20. Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Med Image Anal*. 2018;45:94–107.
21. Xu X, Zhou F, Liu B, Fu D, Bai X. Efficient Multiple Organ Localization in CT Image using 3D Region Proposal Network. *IEEE T Med Imaging*. 2019;p. 1–1.
22. Min E, Guo X, Liu Q, Zhang G, Cui J, Long J. A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture. *IEEE Access*. 2018;6:39501–39514.
23. Kishimoto R, Kikuchi K, Koyama A, Kershaw J, Omatsu T, Tachibana Y, et al. Intra- and inter-operator reproducibility of US point shear-wave elastography in various organs:

- evaluation in phantoms and healthy volunteers. *Eur Radiol.* 2019;29(11):5999–6008.
24. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proc CVPR IEEE*. IEEE; 2016.
25. He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In: *Lect Notes Comput Sc.* Springer International Publishing; 2016. p. 630–645.
26. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*. 2014;p. arXiv:1409.1556.
27. Balagopal A, Kazemifar S, Nguyen D, Lin MH, Hannan R, Owringi A, et al. Fully automated organ segmentation in male pelvic CT images. *Phys Med Biol.* 2018;63(24):245015.
28. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Med Phys.* 2019;46(5):2157–2168.
29. Larsson M, Zhang Y, Kahl F. Robust abdominal organ segmentation using regional convolutional neural networks. *Appl Soft Comput.* 2018;70:465–471.
30. Maaten Lvd, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* 2008;9:2579–2605.
31. Linderman GC, Steinerberger S. Clustering with t-SNE, Provably. *SIAM J Math Data Sci.* 2019;1(2):313–332. Available from: <https://doi.org/10.1137/18M1216134>.
32. Sugimoto K, Kon Y, Lee S, Okada Y. Detection and localization of myocardial infarction based on a convolutional autoencoder. *Knowl-Based Syst.* 2019;178:123–131.
33. Xu X, Zhou F, Liu B, Bai X. Multiple Organ Localization in CT Image Using Triple-Branch Fully Convolutional Networks. *IEEE Access.* 2019;7:98083–98093.
34. Peña-Solórzano CA, Dimmock MR, Albrecht DW, Paganin DM, Bassed RB, Klein M, et al. Effect of external fixation rod coupling in computed tomography. *Strategies Trauma Limb Reconstr.* 2018;13(3):137–149.
35. Gjestebj L, Yang Q, Xi Y, Claus BEH, Jin Y, Man BD, et al. Deep learning methods for CT image-domain metal artifact reduction. In: *Proc SPIE*. SPIE; 2017.
36. Ginat DT, Gupta R. Computed Tomography. In: *Pitfalls in Diagnostic Radiology*. Springer Berlin Heidelberg; 2014. p. 23–38.

Numerical model of the trabecular bone structure

Development of a simple numerical model for trabecular bone structures

C. A. Peña-Solórzano, D. W. Albrecht, D. M. Paganin, P. C. Harris, C. J. Hall, R. B. Bassed, M. R. Dimmock

Published in Medical Physics, Volume 46, Issue 4, Pages 1766–1776, 2019

DOI: 10.1002/mp.13435

This chapter is an exact copy of the journal paper referred above

Chapter 3 demonstrated the capabilities of ML systems for automatically differentiating between the presence or lack of metallic implants and their sub-classification in CT scans. Future applications of the presented pipeline are related to the determination and automatic classification of pathology, including osteoporosis and fractures. As the presence of bone diseases and small fractures can be subtle in CT, the use of models for the training of ML approaches could overcome the difficulties in finding the required data in clinical and forensic databases through the application of transfer learning [55]. ML requires an appropriate dataset to act as a surrogate to initially train a pipeline at identifying the kinds of attributes for which the algorithm will ultimately be used.

This chapter is divided into two parts: firstly, a manuscript that presents a new and simple approach for the generation of synthetic trabecular bone structures that could be customised to the desired application; and secondly, the 3D printing of the generated

models for phantom development. The proposed model only uses seven parameters to capture the pseudo-random structure observed in trabecular bone and can therefore be considered to have more widespread uses than those previously described in the literature [56, 57].

4.1 Publication

The following publication has been published and reprinted from [58]. Copyright 2019, with permission from WILEY.

Development of a simple numerical model for trabecular bone structures

Carlos A. Peña-Solórzano^{a)}

Department of Medical Imaging and Radiation Sciences, Monash University, Melbourne, Vic. 3800, Australia

David W. Albrecht

Faculty of Information Technology, Monash University, Melbourne, Vic. 3800, Australia

David M. Paganin

School of Physics and Astronomy, Monash University, Melbourne, Vic. 3800, Australia

Peter C. Harris

*Department of Orthopaedic Surgery, Western Health, Footscray Hospital, Melbourne, Vic. 3011, Australia
The Royal Children's Hospital Melbourne, Melbourne, Vic. 3052, Australia*

Chris J. Hall

Imaging and Medical Beam Line, ANSTO Australian Synchrotron, Melbourne, Vic. 3168, Australia

Richard B. Bassed

*Victorian Institute of Forensic Medicine, Melbourne, Vic. 3006, Australia
Department of Forensic Medicine, Monash University, Melbourne, Vic. 3800, Australia*

Matthew R. Dimmock

Department of Medical Imaging and Radiation Sciences, Monash University, Melbourne, Vic. 3800, Australia

(Received 15 October 2018; revised 18 January 2019; accepted for publication 1 February 2019; published 5 March 2019)

Purpose: Advances in additive manufacturing processes are enabling the fabrication of surrogate bone structures for applications including use in high-resolution anthropomorphic phantoms. In this research, a simple numerical model is proposed that enables the generation of microarchitecture with similar statistical distribution to trabecular bone.

Methods: A human humerus, radius, ulna, and several vertebrae were scanned on the Imaging and Medical beamline at the Australian Synchrotron and the proposed numerical model was developed through the definition of two complex functions that encode the trabecular thickness and position-dependant spacing to generate volumetric surrogate trabecular structures. The structures reproduced those observed at 19 separate axial locations through the experimental bone volumes. The applicability of the model when incorporating a two-material approximation to absorption- and phase-contrast CT was also investigated through simulation.

Results: The synthetic structures, when compared with the real trabecular microarchitecture, yielded an average mean thickness error of 2 μm , and a mean difference in standard deviation of 33 μm for the humerus, 24 μm for the ulna and radius, and 15 μm for the vertebrae. Simulated absorption- and propagation-based phase contrast CT projection data were generated and reconstructed using the derived mathematical simplifications from the two-material approximation, and the phase-contrast effects were successfully demonstrated.

Conclusions: The presented model reproduced trabecular distributions that could be used to generate phantoms for quality assurance and validation processes. The implication of utilizing a two-material approximation results in simplification of the additive manufacturing process and the generation of synthetic data that could be used for training of machine learning applications. © 2019 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13435]

Key words: numerical model, PB-CT, trabecular

1. INTRODUCTION

Anthropomorphic phantoms are commonly used in radiology for quality assurance (QA)^{1,2} and investigation of the effects of changing experimental settings, for example, mAs and image reconstruction kernel.³ More advanced QA procedures require bone microarchitecture, for example, the validation of high-resolution cone beam computed tomography (CB-CT) images⁴ and the use of high-resolution

peripheral quantitative CT (HR-pQCT).^{5,6} In addition, the planning of experimental design impacts such as object-to-detector propagation distance in phase-contrast synchrotron based studies^{7–9} can be better predicted with realistic phantoms.

An important clinical consideration for high-resolution CT techniques is the requirement for multicenter, temporal consistency checks and validation. In order to ensure such reproducibility, phantoms have been constructed using

cadaveric bone tissue¹⁰ and then mailed between centers; this is a time consuming, inefficient process. With the advent of cost-effective additive manufacturing processes, a solution to overcoming the aforementioned issues could be the use of three-dimensional printed (3DP) trabecular structures,^{11–13} provided that the appropriate bone equivalent material could be identified.^{14,15} Such a process would also require a numerical model from which the appropriate bone pattern could be generated.

Additional benefits of producing phantoms with microarchitecture that strongly represents that observed in medical imaging include the ability to improve clinical radiographic image quality and to generate training sets for machine learning (ML) applications for future radiologic workflow optimization. With respect to clinical radiography, medical imaging technologists (MITs) use trabecular structure in assessing patient positioning and technique factor optimization¹⁶ for plain film x-ray acquisition. For ML applications, the workflows require large numbers of physical or simulated phantoms with feature architecture that can be synthesized in a controlled manner.¹⁷

The trabecular structures in bone can be represented as a quasi-random network comprising interconnected plates and rods.¹⁸ Mathematical models and phantoms of trabecular bone structures have been studied in the field of acoustics in order to determine factors such as ultrasonic attenuation,¹⁹ phase velocity and dispersion,^{20,21} and the experimental observation of ultrasound waves through real trabecular samples vs 3D printed phantoms.²² In these works, authors have used surrogates such as grids of nylon strands.

For x-ray imaging techniques including plain film and high-resolution CT, the cortical structure and trabecular microarchitecture of bone have been characterized through the calculation of cortical thickness,^{23,24} plate-to-rod ratios,²⁵ and more recently tensor-based morphometry²⁶. In these works, the authors have generated simulated microarchitecture using skeletonized continuous spaces comprising unions of sinusoidal, cylindrical, and elliptical surfaces and curves^{18,24}; however, their appearance can be considered somewhat different to that observed in clinical images, for example, see Fig. 3 in Saha *et al.*¹⁸ with respect to Fig. 10 in Liu *et al.*²⁴ In this work we present an alternative model and justify the use of a two-material approximation for ease of fabrication and simplification of simulation for synthetic data generation.

We close this introduction with a brief overview of the organization of this paper. In Section 2, we present a new numerical model for the generation of synthetic trabecular bone structures in two-dimensional (2D) and 3D settings. We also present simplifications to the image reconstruction process that can be achieved assuming that the bone and/or extremity as a whole can be comprised of two materials only. Section 3 presents examples of the trabecular bones generated with our model, results comparing the average trabecular thickness between the synthetic structures and real bone, and the reconstruction of propagation-based phase-contrast CT (PB-CT) images from the generated projection data. Finally,

Sections 4 and 5 present the discussion and conclusions of the paper, respectively.

2. MATERIALS AND METHODS

In this paper, we start by evaluating a clinical CT scan of a currently available anthropomorphic phantom and subsequently develop a numerical model that could produce 3D trabecular structures with similar qualitative appearance and quantitative statistical attributes to those observed experimentally.

2.A. Currently available phantom

We investigated the performance of a current best practice anthropomorphic phantom in order to determine the degree to which bones are accurately simulated. The right leg of the PBU-50 whole body phantom (Kyoto Kagaku Co. Ltd, Kyoto, Japan) was scanned with a Discovery CT750 HD CT scanner (GE Healthcare, Chicago, IL, USA), using 100 kVp energy and with one pixel being equivalent to $0.39 \text{ mm} \times 0.39 \text{ mm}$ of scanned area (Fig. 1). The cortical thickness of the phantom's tibial bone was measured from contours applied to the axial slices for which a gradient function was then used to calculate the normal to the surface for sampling. Figure 1(b) shows the contour and three normal vectors from which sampling occurred. A flat map was then produced to present the calculated cortical bone thickness values.

2.B. Numerical models

Due to the observed deficiency in the microarchitecture in current phantoms, we developed a numerical model that could produce 3D structures that approximate the trabecular structure observed in real bones. The structures were developed with a multi-step process:

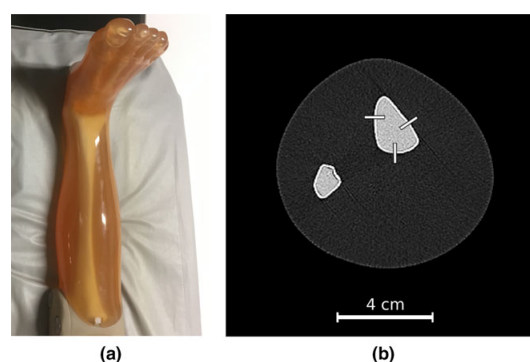


FIG. 1. Right leg of the PBU-50 phantom. (a) Photograph. (b) Axial computed tomography slice showing tibia and fibula bones, and orthogonal vectors for three surface points. No internal trabecular microstructure was observed. [Color figure can be viewed at wileyonlinelibrary.com]

1. implementation of a simple model that reproduces trabecular bone distributions within a single axial slice; then
2. evolution of the single-slice model into a volumetric structure; and
3. evaluation of the efficacy of utilizing a two-material approximation of the bone, and extremity as a whole, specifically with application to PB-CT reconstruction. A two-material approximation could be readily generated using 3DP techniques and is shown to provide both simplified projection data generation in conventional CT and reconstruction in PB-CT.

2.B.1. Trabecular bone model

A series of ex vivo human bones and lamb extremities were imaged with a resolution of 19.8 μm at x-ray energies of 40 and 70 keV on the Imaging and Medical beamline (IMBL) at the Australian Synchrotron (AS). The IMBL provided multiple simultaneous benefits, including: the acquisition of high-resolution projection data; the ability to scan at single energies with effectively parallel rays; and the ability to reconstruct both absorption- and PB-CT images. The aforementioned factors significantly simplified the subsequent analysis and model development, while also providing assessment of the ability to acquire phase-contrast enhancement. The experimentally acquired projection data were processed using X-TRACT²⁷ which facilitated both absorption- and PB-CT reconstruction. The reconstructed images were used to demonstrate how the model could be used with the two-material approximation to generate synthetic data. Figure 2 shows the workflow of the numerical model.

Before proceeding with the mathematical description for our numerical trabecular-bone model, we give a brief qualitative explanation of how our overall theoretical framework works. The overarching desideratum governing this model is that it is simple enough to generate synthetic trabecular structures using a very small number of numerical parameters, yet yield structures that are complex enough to mimic some key features of genuine trabecular bone. In words, the key steps of our numerical model are as follows: (a) Since there is a degree of spatial randomness associated with trabecular bone, our numerical model for generating synthetic trabecular distributions takes a spatially random function as a starting point. For each slice, this input is taken to be white noise, namely a 2D array of pixels, each of which is seeded with an independent random value. (b) Fourier filtration, a form of low-pass filter that suppresses fine detail, is then used to

smooth the white-noise spectrum so as to generate a spatially random structure that has a feature size that can be tuned to match that of a typical trabecular bone structure. This typical feature size is the first numerical parameter used in our method. (c) Since trabecular bone possesses nonrandom as well as random features, for example, the approximate cylindrical symmetry imposed by encasing cortical bone, the previously obtained filtered-noise map needs to be shaped by a deterministic background function, which we incorporate in a simple manner into our model. As we shall see, it is the gradient of this background function that sculpts the trabecular-bone distribution, rather than the background function by itself. (d) Lastly, our simple mathematical model will generate wall thicknesses that are the same on average as the corresponding pore thickness. To introduce a difference between wall and pore thicknesses, a simple power-law transformation is applied to our model; the associated power is a final key parameter that can be tuned.

The first length scale (proportional to the mean trabecular thickness) was encoded by taking the point-wise product of the white noise distribution ($N(j,k)$) with an enveloping function ($M(j,k)$), as,

$$N'(j,k) = M(j,k) \cdot N(j,k), \quad (1)$$

where j and k are pixel coordinates in the y and z transverse directions, respectively. A Gaussian enveloping function ($M(j,k) \rightarrow M_G(j,k)$) centered at the origin of the input white-noise map [Eq. (2)] was utilized for regions where cylindrical statistical symmetry was prevalent; however, since the Fourier transform of a piece of trabecular structure was found to achieve a better fit from a Lorentzian distribution ($M(j,k) \rightarrow M_L(j,k)$), this enveloping function was also investigated [Eq. (3)]. Figure 3(b) shows the enveloping process of the white noise distribution for a Gaussian with $\sigma_{j,k} = 75$ pixels [Figure 3(a)].

Explicitly, the Gaussian and Lorentzian distributions are given by the functions:

$$M_G(j,k) = A \cdot \exp \left[- \left(\frac{j^2}{2\sigma_j^2} + \frac{k^2}{2\sigma_k^2} \right) \right], \quad (2)$$

$$M_L(j,k) = A \left(\frac{\sigma_j}{(j-j_0)^2 + \sigma_j^2} \right) \left(\frac{\sigma_k}{(k-k_0)^2 + \sigma_k^2} \right), \quad (3)$$

where A is an arbitrary scaling constant and (j_0, k_0) is the center of the Lorentzian distribution. In each case, the distributions $M_G(j,k)$ and $M_L(j,k)$ are characterized by σ_j and σ_k (or

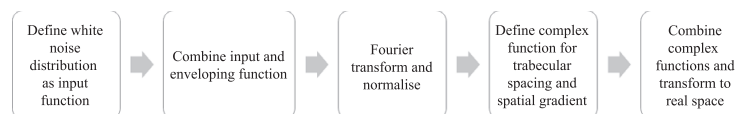


FIG. 2. Workflow of the numerical model that creates synthetic trabecular distributions.

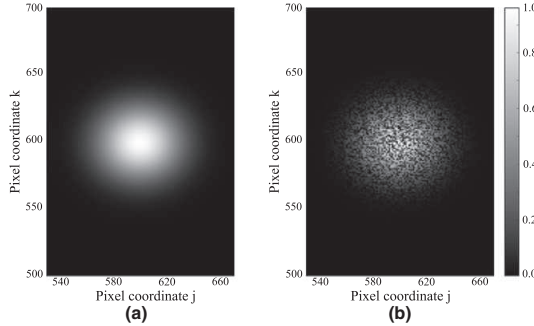


FIG. 3. Combined input and enveloping function for an image of size 1200×1200 pixels (zoom). (a) Gaussian function with $\sigma_{j,k} = 75$ pixels. (b) Combined white noise input and Gaussian enveloping function.

$\sigma_{j,k}$ if it is the same value for both) which controls the standard deviations. Equation (1) was then Fourier transformed over horizontal and transverse pixel coordinates ($F[N'(j, k)]$) to yield a speckle distribution with characteristic spacing $K = 1/\sigma_{j,k}$. The Fourier transform, which will in general be a complex-valued function at each pixel coordinate, was then scaled to have unit modulus, where,

$$N'(q, r) = \frac{F[N'(j, k)]}{|F[N'(j, k)]|}. \quad (4)$$

Here, vertical bars denote the modulus operation. A second complex function $R(q, r) = \exp[-iB(q, r)]$ was introduced to incorporate the trabecular spacing characteristic length scale (L),

$$L = \frac{1}{|\nabla_{\perp} B|}, \quad (5)$$

where ∇_{\perp} denotes the transverse gradient operator. The background function $B(q, r)$ is in general arbitrary, however below it is taken to be an ellipsoid. The function $R(q, r)$ also encodes the spatial gradient, for instance, if an elliptical function is used,

$$R(q, r) = \exp\left(-i \frac{[a(q - q_0)^2 + b(r - r_0)^2]^{0.8}}{L}\right). \quad (6)$$

here, $1/\sqrt{a}$ and $1/\sqrt{b}$ denote the principal axes (i.e., semi-major axis and semi-minor axis) of an ellipse that encodes the spatial gradient. Similarly, q_0 and r_0 denote the offset in the horizontal and vertical directions, respectively, and i is the complex unit $\sqrt{-1}$.

The two complex fields were combined and converted back to real space by raising the modulus to a factor > 2 (i.e., $\eta > 1$) in order to reduce the resulting wall thickness in a tunable manner, hence:

$$T(j, k) = |N'(q, r) + R(q, r)|^{2\eta}. \quad (7)$$

Using a power greater than 2 (i.e., $\eta > 1$) reduces the wall thickness because of the fact that it erodes smaller values and hence narrows the width of a peaked function. This can be seen, for example, by plotting the series of sinusoidal functions $|\sin(x)|^{2\eta}$ vs x , for $\eta = 1, 2, 3$ etc., which demonstrate progressive narrowing as η is increased.

In Eq. (6), if $(q_0, r_0) = (0, 0)$, and $a = b \neq 1$, the separation between structures can be varied asymmetrically [Fig. 4(a)]. If an offset is applied, the center of the distribution generated shifts accordingly. Figure 4(b) shows the variation of the middle location of the synthesized trabecular bone density, when both q_0 and r_0 are negative, while Fig. 4(c) presents the pattern obtained when a and b have different values.

Some regions of bone do not present such strong concentricity as sections of long bones, therefore a region-based approach can be utilized. For these regions, sections of trabecular structure can be generated with a constant spatial gradient, where Eq. (7) can be simplified as,

$$T(j, k) = |N'(q, r) + \exp[-i(cq + dr)]|^{2\eta}, \quad (8)$$

where c and d are constants.

To demonstrate the flexibility of the model, Eq. (7) was used to create concentric distributions with the same mean thickness as axial slices obtained from high-resolution scans of human humerus, radius, and ulna. In addition, synthetic distributions for the vertebrae were generated using Eq. (8) with a Lorentzian input to obtain an idealized model with vertical columns and horizontal struts.²⁸ 19 synthetic trabecular patterns were generated for an image of size 1200×1200 pixels, where the model was constrained according to the resolution achieved through imaging on the IMBL (19.8 μm).

For each of the aforementioned bones, a randomly sampled region was quantitatively assessed in order to determine the slices with the minimum and maximum trabecular thicknesses, and then those slices were reproduced with the model.

2.B.2. Volumetric trabecular structure model

So far, the algorithm presented is able to generate a 2D section of the structure in an axial plane. For a proper analysis of the capabilities of the model, it was extended with the intention of producing a volume similar to what is obtained after scanning a 3D section of bone.

In order to generate a sequence of slices with connected components on the y axis (superiorly or inferiorly along a limb), an iteration loop with the number of desired slices was used, inside which the white noise map was defined as the white noise map from the previous iteration plus a fraction of a new white noise distribution, as,

$$N^{s+1}(j, k) = N^s(j, k) + (d \cdot N(j, k)), \quad (9)$$

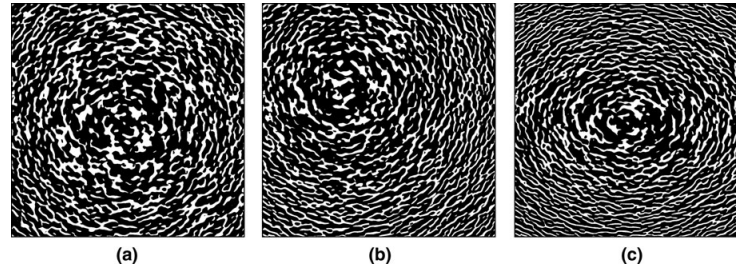


FIG. 4. Representative trabecular distributions obtained using Eq. (7). (a) $a = b \neq 1$ and $q_0, r_0 = 0$. (b) $a, b = 1$ and $q_0, r_0 < 0$. (c) $a \neq b$ and $q_0, r_0 = 0$.

where $d \ll 1$ is an arbitrary positive scaling factor. The scaling factor was used to modify the white noise distribution from the previous iteration, ensuring there are not significant discontinuities in the longitudinal variation of the trabecular structure. Note that the right side of Eq. (9) is always rescaled to have values between 0 and 1, after the addition has been performed. This evolution of the white-noise map gives a slice-by-slice series of inputs into our synthetic trabecular bone model, which then outputs a slice-by-slice series of trabecular bone densities that is subsequently assembled into a 3D density map.

In order to compare the generated trabecular structures against the real tissue, we used the Thickness plug-in from ImageJ,²⁹ which is based in the work from Hildebrand and Rüegsegger.²³ The computation of the local thickness at a given point of the trabecular bone is the diameter of the largest sphere (or circle in the case of 2D images) that includes the point and that can be fitted inside the structure. By fitting maximal spheres to every point in the structure, a mean thickness and the thickness distribution can be calculated. This allowed the model to be compared in a standardized way that can be accessed in an open source platform.

2.B.3. Projection data generation of PB-CT

As stated in Section 1, our numerical model could be applied to both the physical fabrication of anthropomorphic phantoms or incorporation into simulations of experimental CT configurations in order to generate synthetic projection data for the training of ML workflows. In this section, we have incorporated the images obtained with the numerical model into a simplified simulation to demonstrate the usefulness of the assumption of a two-material approximation (Fig. 5). With such an assumption, surrogate bones can be readily produced in additive manufacturing processes and projection data for phase contrast applications can be simulated with significant mathematical simplifications for validating synchrotron configurations.³⁰ Note that due to the ethical difficulties in acquiring cadaveric bone, Fig. 5 shows the histograms and fitted peaks for a femoral axial slice from an ex-vivo ovine sample. The peaks from both 40 and 70 keV were fitted. The distributions demonstrate that a simple two-material approximation is sufficient to characterize an extremity.

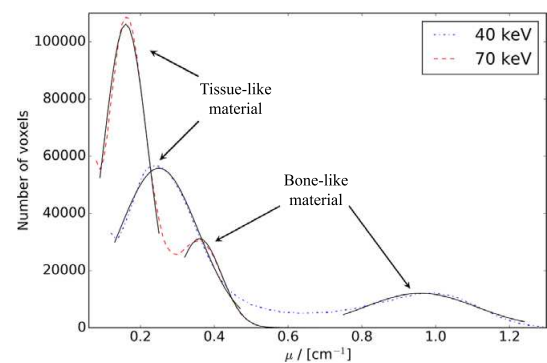


FIG. 5. Histograms and fitted peaks for linear attenuation coefficient of bone and soft tissue for an axial slice from an ex vivo ovine femoral bone imaged at the imaging and medical beamline. The distribution at each energy shows two peaks used as justification for the two-material approximation. [Color figure can be viewed at wileyonlinelibrary.com]

The remainder of this section will show how we simulated the phase-contrast effect that can give rise to edge enhancement in synchrotron imaging [Eqs. (10)–(16)] and simplifications that result from a two-material approximation [Eqs. (17)–(23)].

PB-CT exploits refraction of the x-ray field at tissue interfaces and utilizes the propagation distance from the object to the detector to amplify the interference of the propagating field.^{7,31–33} Figure 6 shows a representation of our simple model, where a flat field of monoenergetic x rays was incident on the trabecular distribution generated using the algorithm outlined in Section 2.B.1. A simple forward model was generated using the ASTRA toolbox^{34,35} to yield absorption-based data that would be obtained at each projection angle for a detector positioned on the plane $z = 0$. The extremity phantom image was generated assuming a two-material approximation. The properties of the two materials are described in Table I. The phantom was defined on a 4096×4096 pixel grid with physical dimension of $16 \text{ mm} \times 16 \text{ mm}$. The projection data sampled on the detection plane was $4096 \times v$, where v was the length expressed as a number of pixels in the longitudinal (y) direction.

In Table I, μ is the linear attenuation coefficient calculated with the energy of the incident photons and the atomic number of the material, $\langle Z/A \rangle$ is the mean ratio of atomic

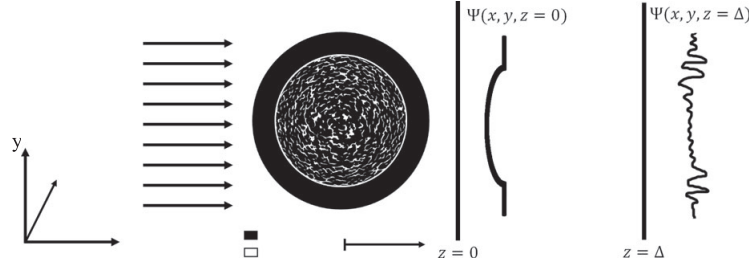


FIG. 6. Simple model of the computed tomography (CT) configuration showing the absorption-based and phase-contrast CT detector locations, $z = 0$ and $z = \Delta$, respectively.

TABLE I. Material properties at 40 keV energy.

	μ (m^{-1})	$\langle Z/A \rangle$ (mole g^{-1})	ρ (g cm^{-3})	δ
Soft tissue	25.0	0.54903	1.127	1.61×10^{-7}
Cortical bone	96.0	0.51478	1.920	2.56×10^{-7}

number Z to atomic weight A for the compound, ρ is the mass density of the material with units of g/cm^3 , and $1-\delta$ is the real part of the complex refractive index, which can be approximated as,³¹

$$\delta(x, y, z) = \frac{C \cdot \rho \cdot \langle Z/A \rangle}{E^2}, \quad (10)$$

where C is a constant of value $4.152 \times 10^{-4} \text{ cm}^3 \text{ keV}^2$, and E is in keV. Throughout, the Cartesian coordinate z is taken to correspond to the optical axis, with (x, y) being coordinates in planes perpendicular to this optical axis.

The complex wave field at the plane $z = 0$, $\Psi(x, y, 0)$ (Fig. 6), can be completely described by the phase $\phi(x, y, 0)$ and amplitude $\alpha(x, y, 0)$ maps at that plane. If the sample is composed of a single material, then the complex wave field at $z = 0$ is determined by the intensity $I(x, y, 0)$ as follows:

$$\begin{aligned} \Psi(x, y, 0) &= \alpha(x, y, 0) \cdot \exp[i\phi(x, y, 0)] \\ &= \sqrt{I(x, y, 0)} \cdot \exp\left(ik \cdot \ln I(x, y, 0) \frac{\delta(x, y)}{\mu(x, y)}\right). \end{aligned} \quad (11)$$

Here, $I(x, y, 0) = |\alpha(x, y, 0)|^2$ is the normalized incident x-ray intensity, k is the wave number, $\delta(x, y)$ is the projection of $\delta(x, y, z)$ along the z direction, and $\mu(x, y)$ is the projection of $\mu(x, y, z)$ along the z direction. Note that an incident z -directed unit-intensity plane wave has been assumed above, together with the paraxial and projection approximations.³⁶ Note also that k refers to the wave number in Eqs. (11)–(14), whereas it refers to the pixel index in the z dimension in all other sections of the paper.

To extend the utility of the model to PB-CT, the Fresnel diffraction approximation³⁶ was applied to calculate the propagated intensity data at the plane $z = \Delta$. The complex wave field distribution at the detector plane ($\Psi(x, y, \Delta)$) can be

obtained from the convolution of $\Psi(x, y, 0)$ with the Fresnel propagator $P(x, y, z)$,

$$\Psi(x, y, \Delta) = \Psi(x, y, 0) * P(x, y, \Delta), \quad (12)$$

where Δ is the distance between the central plane of the sample and the detector plane, and,

$$P(x, y, \Delta) = -\frac{ik \cdot \exp(ik\Delta)}{2\pi\Delta} \exp\left(\frac{ik(x^2 + y^2)}{2\Delta}\right). \quad (13)$$

Practically, this convolution is achieved in Fourier space by application of angular-spectrum diffraction operator,

$$D_\Delta = F^{-1} \exp\left(i\Delta \sqrt{k^2 - k_x^2 - k_y^2}\right) F, \quad (14)$$

giving,

$$\Psi(x, y, \Delta) = D_\Delta \Psi(x, y, 0). \quad (15)$$

The propagated intensity that contains the phase-contrast signal when the detector is positioned at $z = \Delta$ is then given by:

$$I(x, y, \Delta) = |\Psi(x, y, \Delta)|^2. \quad (16)$$

Note that Eq. (14) is in fact more general than the Fresnel propagator described by Eqs. (12) and (13), but for the calculations used in the present paper, the difference is negligible. Equation (15) maps the wave function from a position $z = 0$ to a distance $z = \Delta$, making it possible to simulate the information of the fringes obtained during propagation-based PB-CT (Fig. 6). For the approximation assumed in this work for which the phantom comprises two materials, bone-like and tissue-like (Fig. 5), where the radius (R) of the disk is known (Fig. 6), the thickness of the soft tissue (t_t) can be written as,

$$t_t = 2\sqrt{R^2 - r^2} - t_c, \quad (17)$$

where t_c is the thickness of the cortical bone and $r = \sqrt{x^2 + y^2}$. Assuming monoenergetic x rays incident on the phantom, the Beer–Lambert law can be expressed as,

$$\ln\left(\frac{I_0}{I}\right) = \mu_t t_t + \mu_c t_c \quad (18)$$

where I is the transmitted intensity and I_0 is the incident intensity. Inserting Eq. (17) into (18),

$$\ln\left(\frac{I_0}{I}\right) = \mu_t \left(2\sqrt{R^2 - r^2} - t_c\right) + \mu_c t_c, \quad (19)$$

and solving for t_c then gives,

$$t_c = \frac{\ln\left(\frac{I_0}{I}\right) - 2\mu_t \sqrt{R^2 - r^2}}{\mu_c - \mu_t}. \quad (20)$$

From this, we have now demonstrated that it is possible to generate phase-contrast data using a simple two-material model with arbitrarily complicated shape, where the thickness of the materials can be derived from the intensity distribution on the detector (projection data). To do this, Eq. (20) is substituted into Eq. (21), yielding the following projection-approximation for the phase ϕ of the x rays at the exit surface of the sample,

$$\phi = -k\delta_c t_c - k\delta_t t_t, \quad (21)$$

where $1 - \delta_c$ and $1 - \delta_t$, respectively, denote the real part of the X-ray refractive indices for bone and soft tissue. Combining Eqs. (17), (20), and (21) then gives,

$$\phi = -k\delta_c \left(\frac{\ln\left(\frac{I_0}{I}\right) - 2\mu_t \sqrt{R^2 - r^2}}{\mu_c - \mu_t} \right) - k\delta_t \left(2\sqrt{R^2 - r^2} - \left(\frac{\ln\left(\frac{I_0}{I}\right) - 2\mu_t \sqrt{R^2 - r^2}}{\mu_c - \mu_t} \right) \right), \quad (22)$$

$$\phi = -k \left(\frac{\ln\left(\frac{I_0}{I}\right) - 2\mu_t \sqrt{R^2 - r^2}}{\mu_c - \mu_t} \right) (\delta_c - \delta_t) - 2k\delta_t \sqrt{R^2 - r^2}. \quad (23)$$

Finally, Eq. (23) is substituted into the first equality in Eq. (11) and then propagated using Eqs. (12)–(16). The result is then the intensity distribution at the plane $z = \Delta$. Note that for the detector at $z = 0$ we would emulate conventional CT.

In order to incorporate realistic detector effect, the projection data at $z = \Delta = 6$ m was re-binned to 8 μ m pixels and Poisson noise was added. The same methodology was also applied to the data on the plane $z = 0$ and the two datasets were reconstructed in order to compare the contact and down-stream detector signals. Both datasets were reconstructed using the X-TRACT software.²⁷ A filtered back-projection algorithm was selected for the reconstruction, and phase-retrieval was performed on the down-stream data using single-material transport-of-intensity (TIE) phase retrieval

algorithm developed by Paganin et al.³⁷, and implemented as the “TIE-Hom” option in X-TRACT.²⁷

3. RESULTS

3.A. Currently available phantom

Figure 7 presents the flat map of the PBU-50 phantom with measurements of the cortical thickness and histogram, respectively.

The cortical thickness of the human tibia as measured with current clinical scanners usually ranges between 2 and 6 mm,^{38–40} where the upper values are found on the mid shaft, whereas Fig. 7(b) shows a distribution peak at 0.8 mm. It should be noted that when specialist high-resolution CT scanners are utilized, the cortical thickness on the distal tibia reduces to approximately 0.41 mm for μ CT and 0.7 mm for HR-pQCT.⁴¹ It is therefore evident that the cortical and trabecular structures are insufficient for detailed quality assurance as there is no trabecular structure and the cortical thickness does not present the required range.

3.B. Numerical models

3.B.1. Trabecular bone model

Figures 8 and 9 show visual comparison of real bone vs the synthetic structures generated with our model for the regional [Eq. (8)] and global implementations [Eq. (7)], respectively. The trabecular structures were fitted using ImageJ (see Table II), with brighter areas depicting thicker sections.

For the region-based approach demonstrated in Fig. 8, two synthetic structures [Figs. 8(a) and 8(b)] were generated separately and combined to achieve a model of vertebral trabecular bone comprising columns and struts as motivated by Jensen et al.²⁸ The structures were comparable with differences between mean standard deviations of trabecular thickness of ~ 15 μ m (see Table II).

For the global approach, Fig. 9 shows a section of human humerus and the synthetic trabecular structure cropped to the general shape of the real bone. The structures were comparable with differences between mean standard deviations of trabecular thickness of ~ 33 μ m. The reader should note from Table II that whilst other sections of more concentric bone showed better agreement, this slice was selected for presentation as it demonstrates the difficulty in generating close qualitative matches in regions that are neither highly concentric or grid-like (Fig. 8). However, a more detailed optimization of the parameter space including selection of a more appropriate background function could yield a closer match in these regions.

The measured values for the mean and standard deviation of the trabecular thickness for real trabecular tissue and synthetic structures are presented in Table II. The generated patterns for the humerus, ulna, and radius were created using the

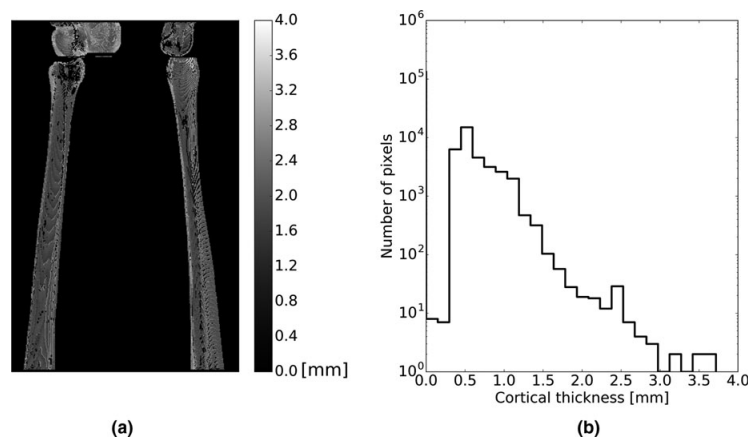


FIG. 7. Right leg of the PBU-50 phantom. (a) Flat map of the cortical thickness derived from the CT scan. (b) Histogram of the flat map with cortical thickness values between 0 and 4 mm.

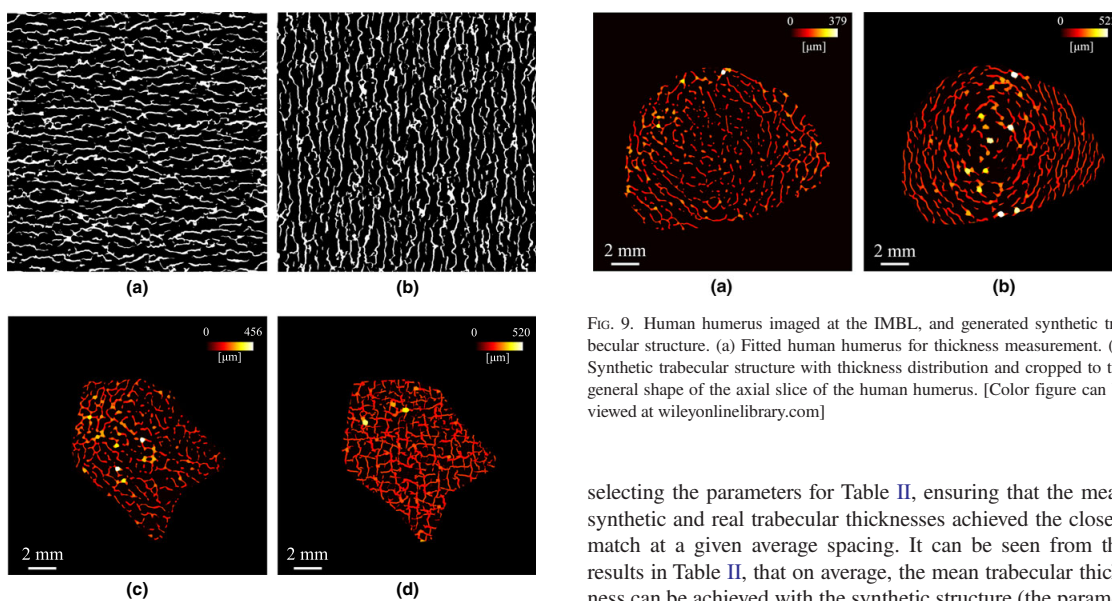


FIG. 8. Visual comparison of synthetic vs human trabecular structure using a region-based approach and implementation of Eq. (8). (a) Synthetic trabecular region with general direction left-right. (b) Synthetic trabecular region with general direction top-bottom. (c) Axial slice of a human vertebra showing thickness distribution. (d) Axial synthetic trabecular structure with thickness distribution after adding the regions (a) and (b), and cropping to the general shape of the vertebra slice. [Color figure can be viewed at wileyonlinelibrary.com]

ellipsoidal function in Eq. (6), and the region-based approach using Eq. (8) was utilized to generate the synthetic vertebral distributions.

For the aforementioned regional and global implementations of the model, an empirical approach was used in

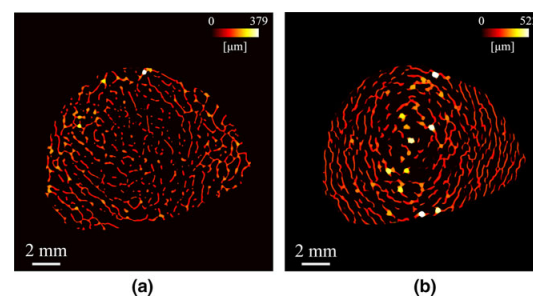


FIG. 9. Human humerus imaged at the IMBL, and generated synthetic trabecular structure. (a) Fitted human humerus for thickness measurement. (b) Synthetic trabecular structure with thickness distribution and cropped to the general shape of the axial slice of the human humerus. [Color figure can be viewed at wileyonlinelibrary.com]

selecting the parameters for Table II, ensuring that the mean synthetic and real trabecular thicknesses achieved the closest match at a given average spacing. It can be seen from the results in Table II, that on average, the mean trabecular thickness can be achieved with the synthetic structure (the parameters can be chosen for an optimal fit by applying small changes to the η value). Also, the patterns generated to mimic the trabecular distribution of the different bones present a mean difference in standard deviation of 33 μm for the humerus, 24 μm for the ulna and radius, and 15 μm for the vertebrae.

3.B.2. Volumetric trabecular structure model

The capability of the numerical model to generate 3D trabecular structures is shown in Fig. 10. Equations (7) and (9) were utilized with $d = 0.1$ and 300 slices. Figure 10(a) shows a render of the trabecular distribution in 3D, while Figs. 10(b) and 10(c) presents axial and sagittal views showing the

TABLE II. Measured trabecular thickness for 19 real and 19 synthetic trabecular distributions. The parameters used for the generation of the synthetic structures are listed. Note that two distributions were used to create the synthetic vertebral images.

Human bone	Model parameters	Trab. thickness (mean \pm SD)		Error (Δstd) (μm)
		Real (μm)	Synthetic (μm)	
Humerus	$\sigma_{jk} = 75$, $\eta = 7.1$	167 \pm 52	166 \pm 83	32
	$L = 1500$, $\eta = 7.5$	170 \pm 53	171 \pm 90	37
	$a = 2.0$, $\eta = 7.0$	174 \pm 63	173 \pm 91	28
	$b = 1.0$, $\eta = 5.0$	205 \pm 69	204 \pm 103	34
	$q_0 = 0.0$, $r_0 = 0.0$			
Ulna and radius	$\sigma_{jk} = 75$, $\eta = 9.0$	146 \pm 51	146 \pm 71	20
	$L = 1500$, $\eta = 8.5$	153 \pm 58	152 \pm 80	22
	$a = 2.0$, $\eta = 6.5$	170 \pm 59	171 \pm 88	29
	$b = 1.0$, $\eta = 6.0$	184 \pm 68	185 \pm 99	31
	$q_0 = 0.0$, $\eta = 2.8$	248 \pm 96	247 \pm 113	17
	$r_0 = 0.0$, $\eta = 2.8$	255 \pm 134	255 \pm 120	14
	$\eta = 2.8$	259 \pm 104	261 \pm 140	35
Vertebrae	$\sigma_{jk} = 30$, $\eta = 10.0$	183 \pm 84	184 \pm 73	11
	$j_0 = [0, -50]$, $\eta = 10.0$	187 \pm 85	187 \pm 75	10
	$k_0 = [50, 0]$, $\eta = 8.0$	202 \pm 107	202 \pm 85	22
	$\eta = 8.0$	205 \pm 93	204 \pm 80	13
	$\eta = 7.0$	214 \pm 100	216 \pm 88	12
	$\eta = 7.0$	220 \pm 117	219 \pm 96	21
	$\eta = 7.0$	222 \pm 100	221 \pm 87	14
	$\eta = 5.7$	238 \pm 111	238 \pm 93	17

variability through the volume whilst maintaining interconnectivity of adjacent slices, which presents similarities to those observed in real bone (Fig. 10(d)).

3.B.3. Projection data generation of PB-CT

In order to demonstrate the applicability of the model and the two-material approximation to the generation of simulated data, the projections were reconstructed for both

absorption- and PB-CT. Figure 11(a) shows the reconstruction from the simulated PB-CT scan. Line profiles for both the absorption- and PB-CT reconstructions for the profile at $y = 0$ are shown in Fig. 11(b). The expected reduction in noise due to the application of phase retrieval was observed. This demonstrates the simplifications derived from Eqs. (17)–(23) due to the two-material approximation can be successfully implemented in generating synthetic data.

4. DISCUSSION

Analysis of the leg of a PBU-50 phantom demonstrated limited reproduction of the anatomical features observed in real bone. Due to the requirement for more representative fabricated and simulated synthetic trabecular structures for applications such as QA for HR-pQCT and μ CT, the training of MLTs, the local analyses of trabecular thickness, and the training of ML algorithms, a numerical model has been proposed.

The numerical model presented can be implemented with a spatial gradient to shape the pattern distribution, for example, concentric patterns using an ellipsoidal function, which more closely resemble the trabecular distribution observed in some sections of the long bones (Fig. 9). Furthermore, bone regions where the trabecular structures are oriented in a general direction can be synthesized and mixed together to obtain known grid-like patterns that more closely resemble the trabecular structure of vertebrae (Fig. 8). This model serves as an alternative to existing computer generated phantoms composed of sinusoidal cross-plates (Fig. 12) where the structure is known,^{18,25} which do not incorporate the pseudo-randomness or concentricity observed.

Four different types of human bone were used during testing: a humerus, radius, ulna, and several vertebrae, yielding an average mean thickness error of 2 μm , and a mean difference in standard deviation of 33 μm for the humerus, 24 μm for the ulna and radius, and 15 μm for the vertebrae. The larger differences observed in standard deviations for the long bones can be primarily attributed to the significant variability in concentricity from slice to slice.

While we have presented the theoretical underpinnings of the model and used some simple parametrization to

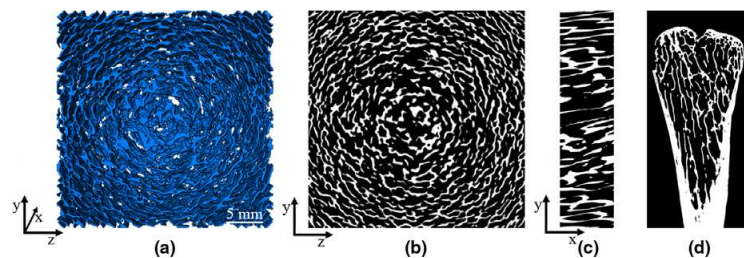


FIG. 10. Generated synthetic three-dimensional (3D) model of the trabecular structure using Eqs. (7) and (9), with $d = 0.1$, $\sigma_{jk} = 75$, $L = 850.0$, $a = 0.5$, $b = 0.5$, $q_0 = 0.0$, $r_0 = 0.0$, and 300 slices. (a) 3D volume render. (b) Axial and (c) sagittal views of the generated 3D structure showing interconnectivity of adjacent slices. (d) Sagittal slice of human radius bone showing distribution similarities to those created by the numerical model. [Color figure can be viewed at wileyonlinelibrary.com]

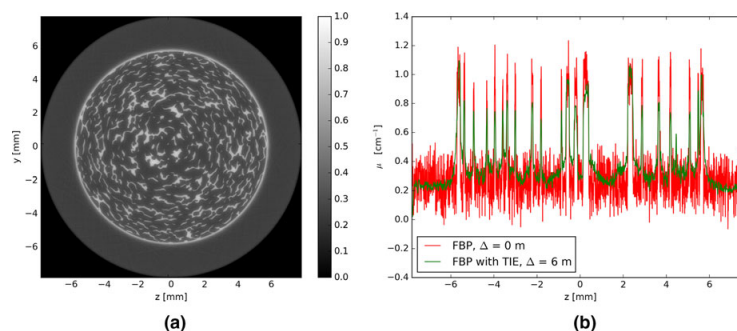


FIG. 11. Reconstructed slice using filtered back projection from projection data generated using the proposed numerical model. (a) Reconstruction using phase-contrast computed tomography (PB-CT) projection data. (b) Line profiles comparing contact ($z = 0$) and PB-CT ($z = \Delta = 6$ m) reconstructions. [Color figure can be viewed at wileyonlinelibrary.com]

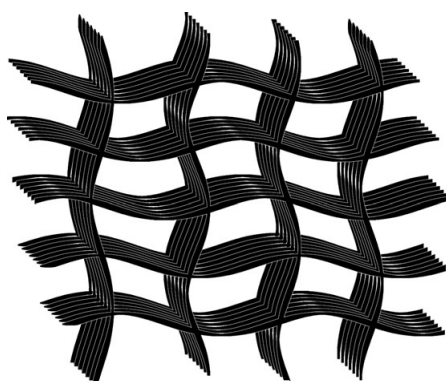


FIG. 12. Computer-generated phantom image of sinusoidal cross-plates.

demonstrate the generalized flexibility, it is left to the reader to perform systematic optimization of the parameters for their specific human or animal imaging task.

The background noise at the heart of our model allows for a more qualitatively natural representation of the trabecular structure, while the capability of seeding the noise generator allows for the reproducibility of the results, which is required for the phantom fabrication to be used in QA and also synthetic simulated projection and reconstruction CT data generation for ML.

When considering the specific problem of multicenter QA, advantages of the proposed approach over the use of phantoms constructed using cadaveric bones include: portability, due to the possibility of creating 3D models that can be printed in the different centers; structural and material stability over time, achieved by proper selection of the printing materials; and realistic and customisable models of human and animal bones, by changing the parameters of the model to control the thickness and spatial distribution.

Note that a real CT scanner gives rise to additional artifacts such as partial volume,^{24,25} which arise when high-density materials, for example, bone, affect the CT values of

surrounding voxels due to a low-resolution detector or a thick acquisition section width. In our simple model of projection data generation of PB-CT we have not accounted for effects such as this.

A further limitation of this study is that the patterns of the synthetic vertebral structures should ideally be influenced by the contour of the cortical bone that encircles the trabecular structure. For our model and its application to the generation of vertebral patterns shown, the trabecular structure is generated and then clipped by the bounding cortical structure. This order of processing generates a somewhat artificial appearance to the intersection of the trabecular and cortical structures.

5. CONCLUSION

In this paper, we presented a numerical method for the generation of high-resolution 3D distributions that reproduce surrogate trabecular structures which are statistically similar to real bone with a mean difference in standard deviation of $33 \mu\text{m}$ for the humerus, $24 \mu\text{m}$ for the ulna and radius, and $15 \mu\text{m}$ for the vertebrae with the parameter values presented here. The developed model could potentially be used for quality assurance in CT systems as recent advances in additive manufacturing will facilitate the printing of 3D trabecular structures. The production of synthetic human and animal bones would also allow the validation of protocols for synchrotron imaging applications to reduce the potential dose to animals and humans. With the advances in machine learning, a model such as the presented may also be used to produce synthetic training data.

ACKNOWLEDGMENTS

This study was supported by the Victorian Orthopaedic Research Trust (VORT) grant (2017 VORT Round One funding application) and the Imaging and Medical beamline (IMBL) at the Australian Synchrotron (reference No. AS172/IM/12180 and AS182/IMBL/13543).

CONFLICTS OF INTEREST

The authors have no relevant conflicts of interest to disclose.

^{a)}Author to whom correspondence should be addressed. Electronic mail: carlos.penasolorzano@monash.edu.

REFERENCES

- Schindera ST, Diedrichsen L, Müller HC, et al. Iterative reconstruction algorithm for abdominal multidetector CT at different tube voltages: assessment of diagnostic accuracy, image quality, and radiation dose in a phantom study. *Radiology*. 2011;260:454–462.
- Schindera ST, Nelson RC, Yoshizumi T, et al. Effect of automatic tube current modulation on radiation dose and image quality for low tube voltage multidetector row CT angiography: phantom study. *Acad Radiol*. 2009;16:997–1002.
- Christianson O, Winslow J, Frush DP, Samei E. Automated technique to measure noise in clinical CT examinations. *Am J Roentgenol*. 2015;205:W93–W99.
- Klintström E, Smedby Ö, Klintström B, Brismar T, Moreno R. Trabecular bone histomorphometric measurements and contrast-to-noise ratio in CBCT. *Dentomaxillofac Rad*. 2014;43:20140196.
- Manske SL, Zhu Y, Sandino C, Boyd SK. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone*. 2015;79:213–221.
- Digby MG, Bishop NJ, Paggiosi MA, Offiah AC. HR-pQCT: a non-invasive ‘biopsy’ to assess bone structure and strength. *Arch Dis Child-hood-E*. 2016;101:268–270.
- Morgan K, Siu KK, Paganin DM. The projection approximation and edge contrast for X-ray propagation-based phase contrast imaging of a cylindrical edge. *Opt Express*. 2010;18:9865–9878.
- Baran P, Pacile S, Nesterets YI, et al. Optimization of propagation-based x-ray phase-contrast tomography for breast cancer imaging. *Phys Med Biol*. 2017;62:2315.
- Yu B, Langer M, Pacureanu A, et al. Assessment of imaging quality in magnified phase CT of human bone tissue at the nanoscale. In: *Proc SPIE*. vol. 10391; 2017. p. 103910L–1–103910L-7.
- Burghardt AJ, Pialat JB, Kazakia GJ, et al. Multicenter precision of cortical and trabecular bone quality measures assessed by high-resolution peripheral quantitative computed tomography. *J Bone Miner Res*. 2013;28:524–536.
- Barak MM, Black MA. A novel use of 3D printing model demonstrates the effects of deteriorated trabecular bone structure on bone stiffness and strength. *J Mech Behav Biomed*. 2018;78:455–464.
- Kuhn V, Ivanovic N, Recheis W. High resolution 3D-printing of trabecular bone based on micro-CT data. *J Orthop Translat*. 2014;4:238.
- Yoon YJ, Moon SK, Hwang J. 3D printing as an efficient way for comparative study of biomimetic structures — trabecular bone and honeycomb. *J Mech Sci Technol*. 2014;28:4635–4640.
- Bibb R, Thompson D, Winder J. Computed tomography characterisation of additive manufacturing materials. *Med Eng Phys*. 2011;33:590–596.
- Hoffmann T, Klink F, Boese A, Fischer K, Beuing O, Rose G. Development of a skull phantom for the assessment of implant X-ray visibility. *Curr Dir Biomed Eng*. 2016;2:351–354.
- Davies AM, Ostensen H, Pettersson H. *The WHO manual of diagnostic imaging: radiographic anatomy and interpretation of the musculoskeletal system*. Geneva, Switzerland: World Health Organization; 2002.
- McCann MT, Jin KH, Unser M. Convolutional neural networks for inverse problems in imaging: a review. *IEEE Signal Proc Mag*. 2017;34:85–95.
- Saha PK, Xu Y, Duan H, Heiner A, Liang G. Volumetric topological analysis: a novel approach for trabecular bone classification on the continuum between plates and rods. *IEEE Trans Med Imaging*. 2010;29:1821–1838.
- Gilbert RP, Guyenne P, Li J. Numerical investigation of ultrasonic attenuation through 2D trabecular bone structures reconstructed from CT scans and random realizations. *Comput Biol Med*. 2014;45:143–156.
- Wear KA. The dependencies of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms. *J Acoust Soc Am*. 2005;118:1186–1192.
- Haia G, Lhemery A, Renaud F, Padilla F, Laugier P, Naili S. Velocity dispersion in trabecular bone: influence of multiple scattering and of absorption. *J Acoust Soc Am*. 2008;124:4047–4058.
- Meziere F, Juskova P, Woittequand J, et al. Experimental observation of ultrasound fast and slow waves through three-dimensional printed trabecular bone phantoms. *J Acoust Soc Am*. 2016;139:EL13–EL18.
- Hildebrand T, Rüegsegger P. A new method for the model-independent assessment of thickness in three-dimensional images. *J Microsc*. 1997;185:67–75.
- Liu Y, Jin D, Li C, et al. A robust algorithm for thickness computation at low resolution and its application to in vivo trabecular bone CT imaging. *IEEE Trans Biomed Eng*. 2014;61:2057–2069.
- Saha PK, Liu Y, Chen C, et al. Characterization of trabecular bone plate-rod microarchitecture using multirow detector CT and the tensor scale: algorithms, validation, and applications to pilot human studies. *Med Phys*. 2015;42:5410–5425.
- Carballido-Gamio J, Bonaretti S, Kazakia GJ, et al. Statistical parametric mapping of HR-pQCT images: a tool for population-based local comparisons of micro-scale bone features. *Ann Biomed Eng*. 2017;45:949–962.
- Gureyev TE, Nesterets Y, Ternovskiy D, et al. Toolbox for advanced X-ray image processing. In: *Proc. SPIE*. vol. 8141; 2011. p. 81410B–1–81410B-14.
- Jensen K, Mosekilde L, Mosekilde L. A model of vertebral trabecular bone architecture and its mechanical properties. *Bone*. 1990;11:417–423.
- Dougherty R, Kunzelmann KH. Computing local thickness of 3D structures with ImageJ. *Microsc Microanal*. 2007;13:1678–1679.
- Hagen CK, Maghsoudlou P, Totonelli G, et al. High contrast microstructural visualization of natural acellular matrices by means of phase-based X-ray tomography. *Sci Rep*. 2015;5:18156.
- Gureyev T, Mayo S, Myers D, et al. Refracting Röntgen's rays: propagation-based X-ray phase contrast for biomedical imaging. *J Appl Phys*. 2009;105:102005.
- Momose A, Takeda T, Itai Y, Yoneyama A, Hirano K. Phase-contrast tomographic imaging using an X-ray interferometer. *J Synchrotron Radiat*. 1998;5:309–314.
- Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Schelokov I. On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev Sci Instrum*. 1995;66:5486–5492.
- Pelt DM, Gürsoy D, Palenstijn WJ, Sijbers J, De Carlo F, Batenburg KJ. Integration of TomoPy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data. *J Synchrotron Radiat*. 2016;23:842–849.
- van Aarle W, Palenstijn WJ, De Beenhouter J, et al. The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography. *Ultramicroscopy*. 2015;157:35–47.
- Paganin DM. *Coherent X-Ray Optics*. Oxford, UK: OUP; 2006.
- Paganin DM, Mayo S, Gureyev TE, Miller PR, Wilkins SW. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J Microsc*. 2002;206:33–40.
- Seebeck J, Goldhahn J, Städele H, Messmer P, Morlock MM, Schneider E. Effect of cortical thickness and cancellous bone density on the holding strength of internal fixator screws. *J Orthop Res*. 2004;22:1237–1242.
- Techawiboonwong A, Song HK, Leonard MB, Wehrli FW. Cortical bone water: in vivo quantification with ultrashort echo-time MR imaging. *Radiology*. 2008;248:824–833.
- Patterson J, Rungprai C, Den Hartog T, et al. Cortical bone thickness of the distal part of the tibia predicts bone mineral density. *J Bone Joint Surg*. 2016;98:751–760.
- Liu XS, Zhang XH, Sekhon KK, et al. High-resolution peripheral quantitative computed tomography can assess microstructural and mechanical properties of human distal tibial bone. *J Bone Miner Res*. 2010;25:746–756.

4.2 Synthetic trabecular structure applications

The developed numerical model was shown to be able to faithfully generate synthetic 3D trabecular structures. The model was subsequently applied to the task of generating a trabecular structure that could be incorporated into a standardised anthropomorphic phantom. The efficacy of 3D printing the generated model was studied in this section, with special consideration to the limitations (resolution) for currently available commercial 3D printing devices. The 3D printing of physical structures was made possible thanks to the successful award of a grant from the Victorian Orthopaedic Research Trust (VORT) in 2017.

The printers utilised were the Form 2, the Object500 Connex3, and the Guider 3D printer. Form 2 (Formlabs) is a stereo-lithography (SLA) 3D printing system, which uses a laser to cure solid parts from a vat of liquid photopolymer resin. The Object500 Connex3 (Stratasys) is a PolyJet printer, which also uses UV light to cure liquid photopolymer, but instead of a vat, a fine print nozzle deposits droplets of the material onto a platform. Finally, the Guider 3D printer (FlashForge) uses a fused filament fabrication (FFF) technology, also called fused deposition modelling (FDM), which lays down layers of material at high temperatures, where each adjacent layer has to first cool down and bond together before the next layer is deposited. The printing materials used in this study were Durable, Flexible, Grey Pro, and Dental SG from Formlabs, VeroClear from Stratasys, Ltd., and ABS and PETG for the Guider 3D printer.

In order to test the resolution limits of the available 3D printers and candidate materials, Derenzo phantom [59] structures were generated (see Figure 4.1a). Figure 4.1b-c show the high- and low-resolution Derenzo phantoms printed using the Dental SG material. The low-resolution phantom had cylinder radii that varied from 500 μm to 1200 μm in 100 μm steps. The high-resolution phantom had cylinder radii that varied from 150 μm to 500 μm in 50 μm steps. The printed structures enabled the study of the feature resolution of the printer and the stiffness of the materials after fabrication.

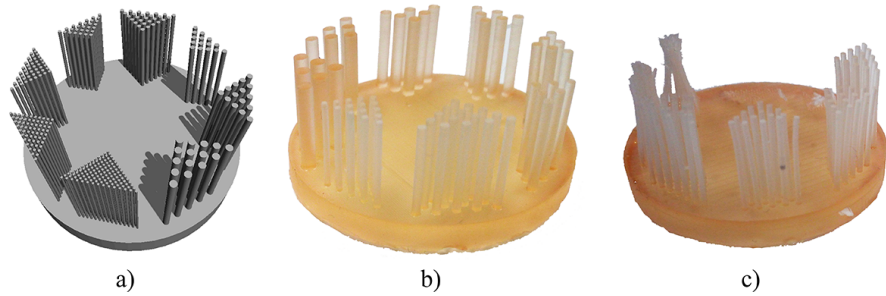


Figure 4.1: Derenzo phantom. a) 3D model. b) Small 3D printed model. c) Large 3D printed model.

Once fabricated, the phantoms were imaged at the IMBL of the Australian Synchrotron. Each object was scanned twice at two different monochromatic energies (40 keV and 70 keV) so that DECT reconstructions could be performed and subsequently transformed into electron density (ρ_e) and effective atomic number (Z_{eff}) bases in order to assess the material characteristics [60]. Whilst DECT imaging can be performed on commercial scanners, the synchrotron offers superior spatial resolution and removes the ambiguity caused by the beam hardening effect [61]. The ρ_e - Z_{eff} transformation can be used as a quantitative approach to characterise materials [62], although high levels of noise are known to be generated in the output image [63].

In order to acquire a test pattern that could be mimicked using the numerical model and then subsequently printed, an appropriate substitute of the human leg, an ovine leg, was scanned at the IMBL, along with solid cylinders of the seven chosen printing materials. Figure 4.2 shows the ρ_e - Z_{eff} transformation of one slice of the scanned lamb legs. The Z_{eff} histogram shows peaks at 9.0 and 13.5 for soft tissue and bone, respectively. The ρ_e histogram presents a peak close to 0.0 for background information, while the peaks for soft tissue and bone are found at 1.0 and 1.35, respectively, as expected from the literature ρ_e .

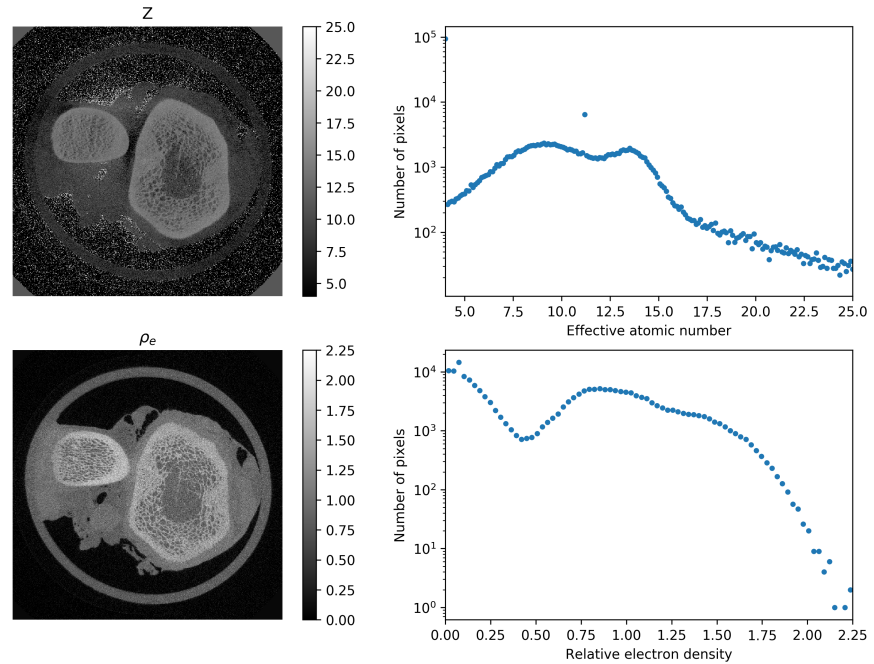


Figure 4.2: ρ_e - Z_{eff} transformation and histogram of lamb leg scan. Left column presents obtained Z_{eff} and ρ_e images, with their respective histograms on the right column.

Figure 4.3 shows the ρ_e - Z_{eff} transformation for the printed cylinders. The Z_{eff} histogram shows two peaks, the first at 6.5 for all the cylinders except the PETG material printed using the Guider 3D, whose peak was found at 9.1. The histogram for the ρ_e image shows three peaks, one close to 0.0 for the background, 1.1 for the regular cylinders, and 1.36 for the PETG cylinder. Although the relative electron density of PETG was closer to

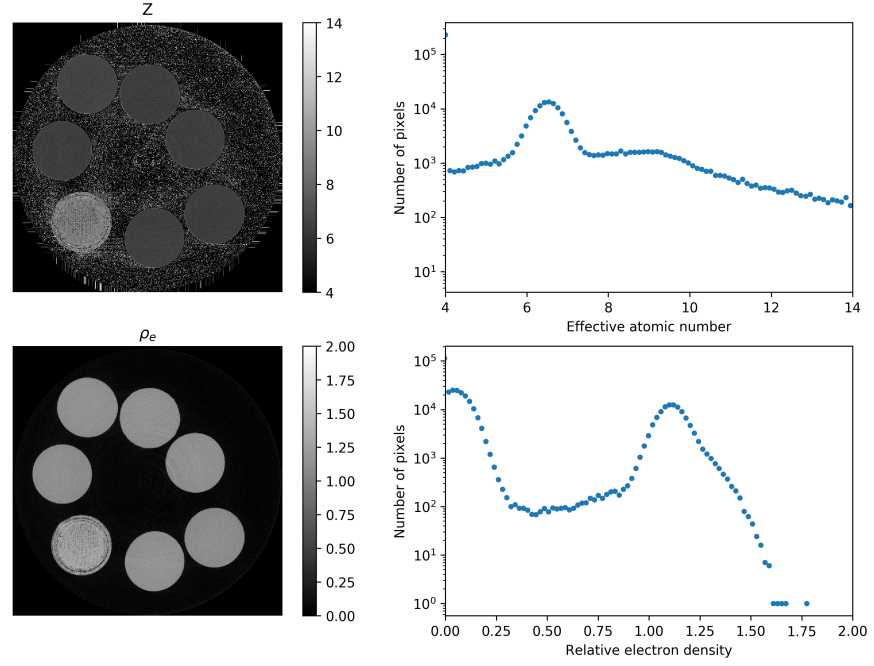


Figure 4.3: ρ_e - Z_{eff} transformation and histogram of material cylinders scan. Left column presents obtained Z_{eff} and ρ_e images, with their respective histograms on the right column.

that of bone, the material could not support structures printed to the resolution of the other systems trialled. The most successful low- and high-resolution Derenzo phantoms were printed using the Dental SG material. This material was selected due to its stiffness and fine details determined through visual inspection of the pillars (Figure 4.4). With a pixel resolution of $19.8 \mu m$ in the reconstructed CT images, the diameters of the smallest properly printed group of cylinders were $475 \mu m$.

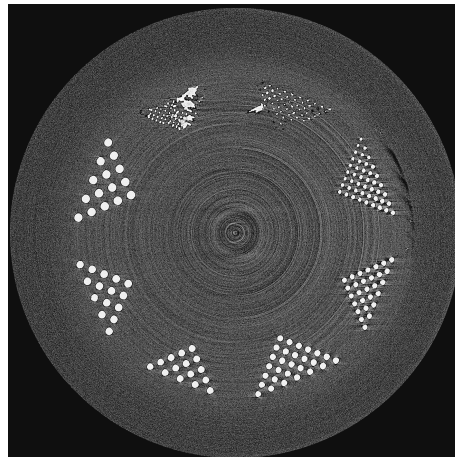


Figure 4.4: Scan of small Derenzo phantom (30 keV) with Dental SG material.

After characterisation of the materials from the synchrotron scans, the model presented in Section 4.1 was used to generate a 3D bone structure that could be printed, as an alternative to current techniques for producing bone phantoms [57, 64, 65]. The parameters

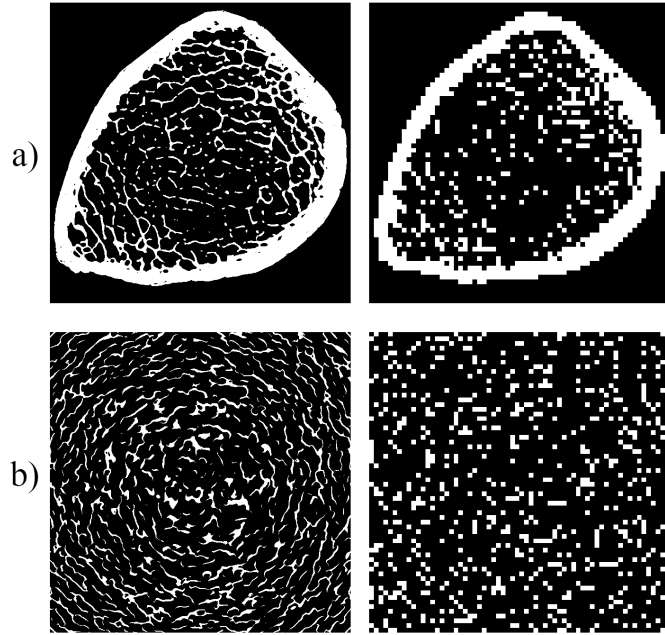


Figure 4.5: synchrotron scan obtained from human bone and synthetic trabecular structure before and after been downsampled from a pixel resolution of $19.8 \mu m$ to the diameter of the smallest properly printed cylinder ($475 \mu m$). a) Human bone. b) Synthetically generated trabecular structure.

of the numerical model were modified to control the thickness and distribution of the trabecular structures. Due to the restrictions of the resolution found with the 3D printers trialled, a limitation commonly found during printing of trabecular bone structures [66–68], a down-sampling technique was utilised to modify both the real and synthetic images of the trabecular bone. Figure 4.5 shows that down-sampling images to the maximum feature resolution of the printer gives rise to discontinuities in the synthetic trabecular structures.

An important limitation of the results obtained in this section relates to the number of materials and printers tested. A comprehensive study of printing systems and materials is necessary to determine the potential resolution that can be achieved with research printing systems, as other authors have shown a vertical, or depth, resolution of $16 \mu m$ [67]. Approaches for micro- and nano-fabrication are being widely investigated [69, 70], and advances in this area are constantly improving the resolution and range of polymers that can be utilised. A further limitation in the work presented in this section arose from insufficient data collection at IMBL. Significant noise can be observed in the $\rho_e Z_{eff}$ transformation images; a higher dwell time at each projection angle when acquiring the synchrotron data was required and a more sophisticated analysis than that presented here requires the data to be collected again at a subsequent beam-time.

Metal artefacts from the coupling of dense objects in CT

Effect of external fixation rod coupling in computed tomography

C. A. Peña-Solórzano, M. R. Dimmock, D. W. Albrecht, D. M. Paganin, R. B. Bassed, M. Klein, P. C. Harris

Published in *Strategies in Trauma and Limb Reconstruction*, Volume 13, Issue 3, Pages 137–149, 2018

DOI: 10.1007/s11751-018-0318-x

This chapter is an exact copy of the journal paper referred above

The physical nature of the acquisition of CT data means that the image quality is influenced by various types of artefacts. The most commonly encountered artefacts in the forensic setting are partial volume [71], beam hardening [49], and photon starvation [50]. Metal artefacts are the extreme combination of beam hardening and photon starvation that result from the most dense type of materials (metal implants, teeth, foreign objects, e.g. bullets) encountered in a CT scanner. Whilst beam hardening and metal artefacts can result from the presence of a single dense object in the scanner, the most significant effect on image quality is observed when two or more dense objects are aligned such that X-rays have to traverse both objects in a single line of sight. Some examples of these types of artefacts can be seen in Figure 5.1.

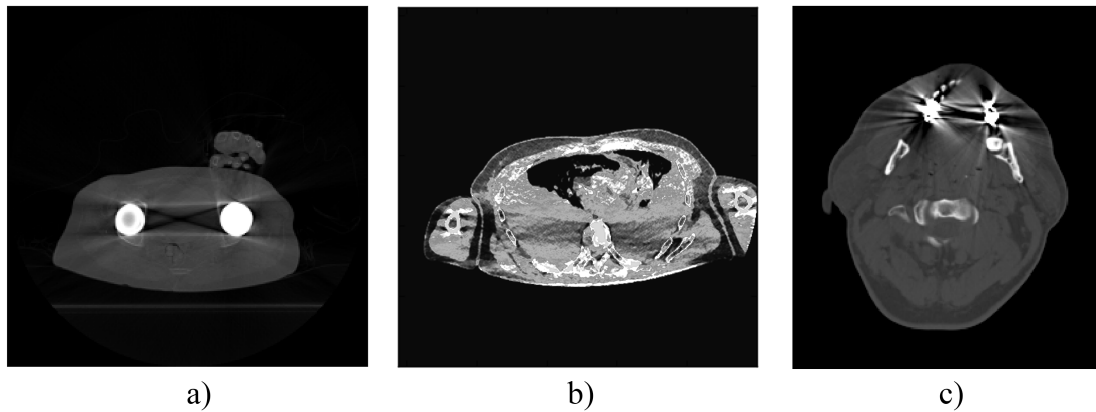


Figure 5.1: Examples of coupling artefacts found in the CT database. a) Between bilateral hip replacement. b) Between the arms, affecting the HU values of the tissues in the thorax. c) Between metal fillings in the teeth.

As can be seen from Figure 5.1, the aforementioned artefacts appear in different anatomical regions. Whilst techniques such as MARS can be applied at the time of acquisition in cases for which it is known that metal is directly affecting the pathological assessment of the CT images, use of these algorithms also affects the visibility of structures in other regions of the body. Therefore, MARS might not be routinely applied to every scan acquired and the retrospective analysis of images already acquired will therefore have these types of artefacts in different regions of the body. In this chapter, the effect of the coupling of different numbers of dense structures on image quality was investigated using external fixation devices (ex-fix) as a surrogate for the types of artefacts shown in Figure 5.1. As the ex-fix uses metal rods positioned at various angular spacings, the artefacts that arise can be referred as those from rod coupling.

This chapter comprises two sections: firstly, a manuscript that quantified the effect of ex-fix and rod coupling on image quality, as assessed by the peak characteristics in the histograms of the images presented; secondly, the denoising of synthetic data generated using a pipeline that combined the models from Chapter 4 and the first section of this chapter.

5.1 Publication

The following paper has been published and reprinted from [51]. Copyright 2018, with permission from Springer (open access).



Effect of external fixation rod coupling in computed tomography

Carlos A. Peña-Solórzano¹ · Matthew R. Dimmock¹ · David W. Albrecht² · David M. Paganin³ · Richard B. Basset^{4,5} · Mitzi Klein⁶ · Peter C. Harris^{7,8}

Received: 19 February 2018 / Accepted: 11 September 2018
 © The Author(s) 2018

Abstract

External fixation is a common tool in the treatment of complex fractures, correction of limb deformity, and salvage arthrodesis. These devices typically incorporate radio-opaque metal rods/struts connected at varying distances and orientations between rings. Whilst the predominant imaging modality is plain film radiology, computed tomography (CT) may be performed in order for the surgeon to make a more confident clinical decision (e.g. timing of frame removal, assessment of degree of arthrodesis). We used a fractured sheep leg to systematically assess CT imaging performance with a Discovery CT750 HD CT scanner (GE Healthcare) to show how rod coupling in both traditional Ilizarov and hexapod frames distorts images. We also investigated the role of dual-energy CT (DECT) and metal artefact reduction software (MARS) on the visualisation of the fractured leg. Whilst mechanical reasons predominantly dictate the rod/strut configurations when building a circular frame, rod coupling in CT can be minimised. Firstly, ideally, all or all but one rod can be removed during imaging resulting in no rod coupling. If this is not possible, strategies for configuring the rods to minimise the effect of the rod coupling on the region of interest are demonstrated, e.g., in the case of a four-rod construct, switching the two anterior rods to a more central single one will achieve this goal without particularly jeopardising mechanical strength for a short period. It is also shown that the addition of DECT and MARS results in a reduction of artefacts, but also affects tissue and bone differentiation.

Keywords Computed tomography · Rod coupling · Dual-energy CT · Metal artefact reduction · Metal artefacts

The work was performed at Western Health, Footscray Hospital, Gordon St, Footscray, Melbourne, VIC 3011, Australia.

✉ Carlos A. Peña-Solórzano
carlos.penasolorzano@monash.edu

Matthew R. Dimmock
matthew.dimmock@monash.edu

David W. Albrecht
david.albrecht@monash.edu

David M. Paganin
david.paganin@monash.edu

Richard B. Basset
richard.basset@vifm.org

Mitzi Klein
Mitzi.Klein@monash.edu

Peter C. Harris
chris.harris@rch.org.au

¹ Department of Medical Imaging and Radiation Sciences, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia

² Clayton School of Information Technology, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia

³ School of Physics and Astronomy, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia

⁴ Victorian Institute of Forensic Medicine, 57-83 Kavanagh St., Southbank, Melbourne, VIC 3006, Australia

⁵ Department of Forensic Medicine, Monash University, Wellington Rd, Clayton, Melbourne, VIC 3800, Australia

⁶ Australian Synchrotron, 800 Blackburn Rd, Clayton, Melbourne, VIC 3168, Australia

⁷ The Royal Children's Hospital Melbourne, 50 Flemington Road, Parkville, Melbourne, VIC 3052, Australia

⁸ Department of Orthopaedic Surgery, Western Health, Footscray Hospital, Gordon St, Footscray, Melbourne, VIC 3011, Australia

Introduction

Circular external fixation is one of several tools available for the orthopaedic surgeon to use in the management of a number of orthopaedic conditions [1], including the treatment of complex fractures, correction of limb deformity, and salvage arthrodesis.

During the post-operative period, imaging is performed to assess, amongst other things, alignment and progress of healing (callus formation, union, etc.). Whilst the predominant imaging modality is plain film radiology, computed tomography (CT) may be performed when plain film radiology does not give sufficient detail of a region of interest (ROI) in order for the surgeon to make a more confident clinical decision (e.g. timing of frame removal, assessment of degree of arthrodesis).

The frame components, being relatively radio-opaque, have the capacity to either obscure the view of the region of interest (predominantly in the case of plain film radiology) or distort the quality of the image to some degree (in the case of CT). For both of these imaging modalities, strategies to minimise this fall under three general categories. Firstly, the surgeon has some degree of choice as to the positioning of certain components (e.g. rings and rods) relative to the region of interest. Secondly, the way in which the patient is positioned during image acquisition can be optimised [2]. Thirdly, some components, such as the connecting rods, can be temporarily removed or repositioned during imaging. With CT there is an additional

strategy; many modern scanners have the facility to reduce the effect of metal artefacts through the use of dual-energy CT (DECT) and/or metal artefact reduction algorithms [3–6]. However, such scanner settings are not specific to the individual patient/frame construct, and so not only do they have varying ability to reduce metallic artefacts, but in addition they may also adversely affect image quality in general.

Since there are situations in which the orthopaedic surgeon may rely heavily on the quality of the CT image, this study focuses on the nature of the adverse effect of metal rods/struts on the quality of the image and on how to optimise the image quality in the presence of metallic components.

Materials and methods

The leg of a sheep that had been euthanised for reasons unrelated to this research was used. A mid-shaft oblique fracture was created by making a single drill hole followed by the use of an osteotome whilst bending the bone. The soft tissues were left in place.

The frame consisted of two 130-mm aluminium rings (Taylor Spatial Frame, Smith & Nephew), spaced 196 mm apart. Two construct types were studied. Construct one represents the traditional Ilizarov frame, where the rings are connected by threaded stainless steel rods (Ilizarov, Smith & Nephew) (see Fig. 1b). Construct two is a hexapod, where

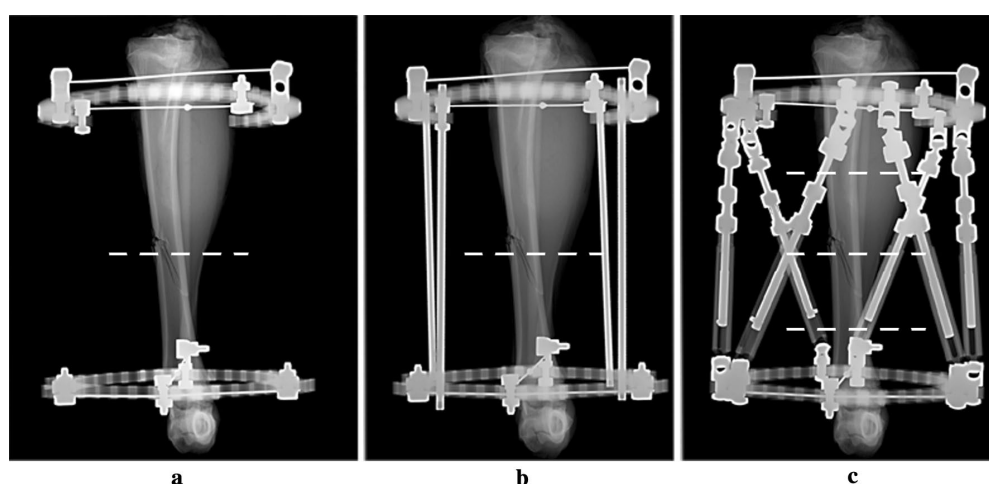


Fig. 1 X-rays (anterior–posterior views) showing various frame constructs. **a** Initial positioning of the rings. **b** Ilizarov frame (4 rods). **c** Taylor Spatial Frame. Dotted lines show the position of the slices used for the study

the rings are connected by six struts attached at the outer mounting holes of the tabs (Taylor Spatial Frame, Smith & Nephew) (see Fig. 1c). For the Ilizarov frame, the rods were varied both in number and in configuration, whilst for the Taylor Spatial Frame (TSF), only one configuration was used, with all 6 struts at the same length. Each ring was attached to the bone using two crossed and tensioned wires. The rings were connected with three plastic threaded rods, which allowed a baseline scan to be performed without any metal between the rings whilst ensuring that the fracture position remained unchanged. The frame was taped to the CT bed in an orthogonal orientation, ensuring that a constant and idealised position was maintained for all study images. To further ensure standardisation of slice acquisition, the proximal ring was used as a reference level for each axial slice (Fig. 2).

The scans were performed with a Discovery CT750 HD CT scanner (GE Healthcare), which has the facility for both DECT and metal artefact reduction software (MARS). The frame/leg construct was scanned initially at 100 kVp, which represents the single-energy setting that would typically be used in the presence of metal. It was then scanned at dual energy (80/140 kVp) with and without MARS. A single axial slice is composed of 512×512 (262,144) pixels, whilst the field of view chosen for this study resulted in one pixel being equivalent to $0.39 \text{ mm} \times 0.39 \text{ mm}$ of scanned area. Each individual pixel is a measure of relative radiodensity and is given a value termed a Hounsfield unit (HU). Water (at standard pressure and temperature) has a HU of zero,



Fig. 2 Photograph of the frame/leg construct in the CT scanner

whilst anything of lesser radiodensity is a negative value and anything of greater radiodensity is a positive value [7].

For the purposes of this study, the pixel values are displayed in two distinct ways: a traditional axial 2D image and a histogram. For the 2D image, the Hounsfield units are displayed as a greyscale, with negative values being increasingly dark and positive values being increasingly white (Fig. 3a, left). We used display settings with a centre of 300 HU and a window width of 2800 HU, which is not a typically used window width, but allows us to better observe the metal artefacts affecting the background of the images. The histogram is a graphical representation of the number of pixels of particular HU that are present in the field of view (Fig. 3a, right). To make the histogram more specific, a ROI around the tibia was defined (Fig. 3b, left); this removes the peak produced by air, making it easier to appreciate any change in the other two peaks (Fig. 3b, right). Whilst the histogram gives no spatial appreciation of individual pixels, it allows a quantitative analysis of the effect of the metal artefact on image quality by using Gaussian fitting of the bone and soft tissue peaks (Fig. 3b; dashed line; note that a Gaussian function appears as an inverted parabola on a logarithmic scale). The parameters of the peak that were analysed for the Gaussian functions were the mean HU value and the width (2.35 times the standard deviation).

Results

Results are displayed as 2D images and ROI histogram, using the scale shown in Fig. 3. The ROI varies according to the slice being analysed, with the shape of an approximately elliptical shape that follows the outer boundary of the bone. Unless otherwise mentioned, the images show scans performed with single-energy CT.

Baseline image

Figure 4a shows the 2D image and ROI histogram of an axial slice through the fracture site in a construct where there is no metal between the rings (although the plastic rods are evident). It represents the best quality image that can be obtained and therefore serves as the baseline for comparison with all subsequent images, where the presence of varying degrees of metal produces some degree of image degradation.

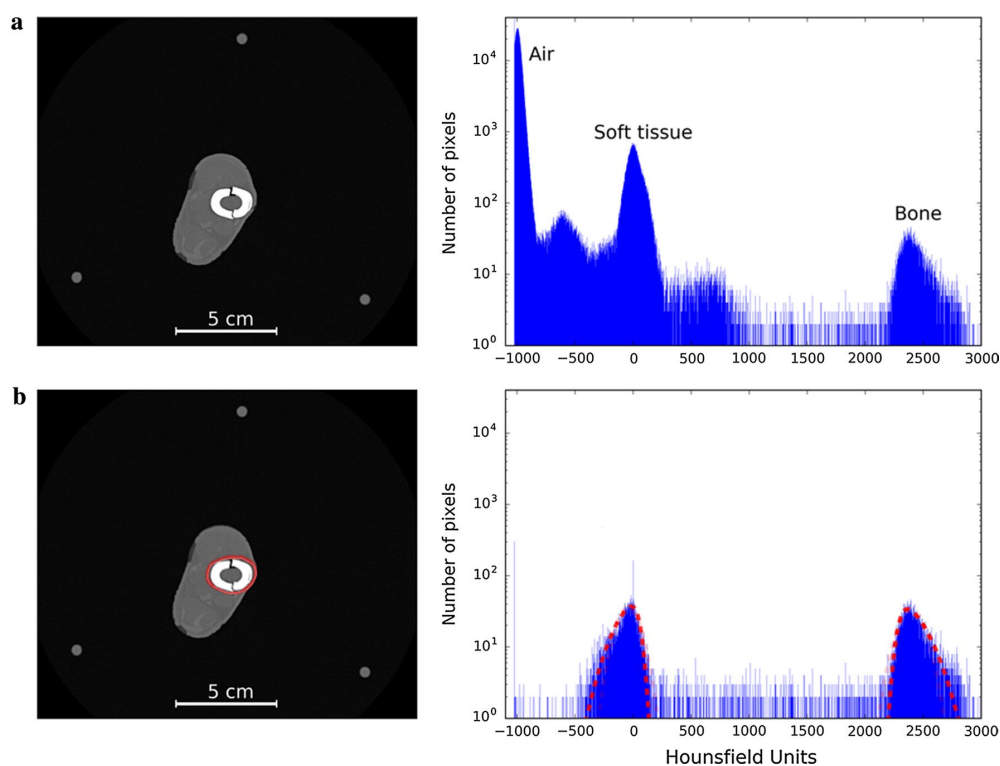


Fig. 3 Example of information used for the analyses. **a** 2D axial slice (left) and histogram of whole image (right), with the three peaks representing air (black), soft tissue (grey), and bone (white). **b** 2D axial slice showing the region of interest circled in red (left) and histogram

of the pixels inside the chosen area (right), where mean HU value and width are determined for bone and soft tissue using Gaussian fitting, shown in dotted lines

Ilizarov frame

The presence of a single metal rod (Fig. 4b) causes fine streaks artefacts to radiate from it, but they have little effect on the quality of the image. On the histogram, the bone and soft tissue peaks are similar to those obtained from the image without rods.

When two rods are used, the effect on the region of interest is highly dependent upon where they are placed (Fig. 5). On the 2D image it can be seen that the rods act as a couple, producing a broader and more noticeable streak that runs between them. This streak has a dark centre and bright edges. Where this streak crosses a part of the image that is

already dark (Fig. 5a), its effect is negligible. However, if it crosses the region of interest, then its effect on detail is more pronounced (Fig. 5b). This can be appreciated on the histogram as a change in the spikes (wider bases and a change in the mean values).

Interestingly, when 3 rods are used, the effect on the region of interest is minimal (Fig. 6a); broad dark streaks with bright edges still occur between rods, but because they do not cross the ROI, they have little adverse effect. For the four-rod construct there are six such streaks, but only two of them cross the region of interest (Fig. 6b). The histogram shows a change in the spikes (wider bases), indicating loss of contrast and therefore image quality.

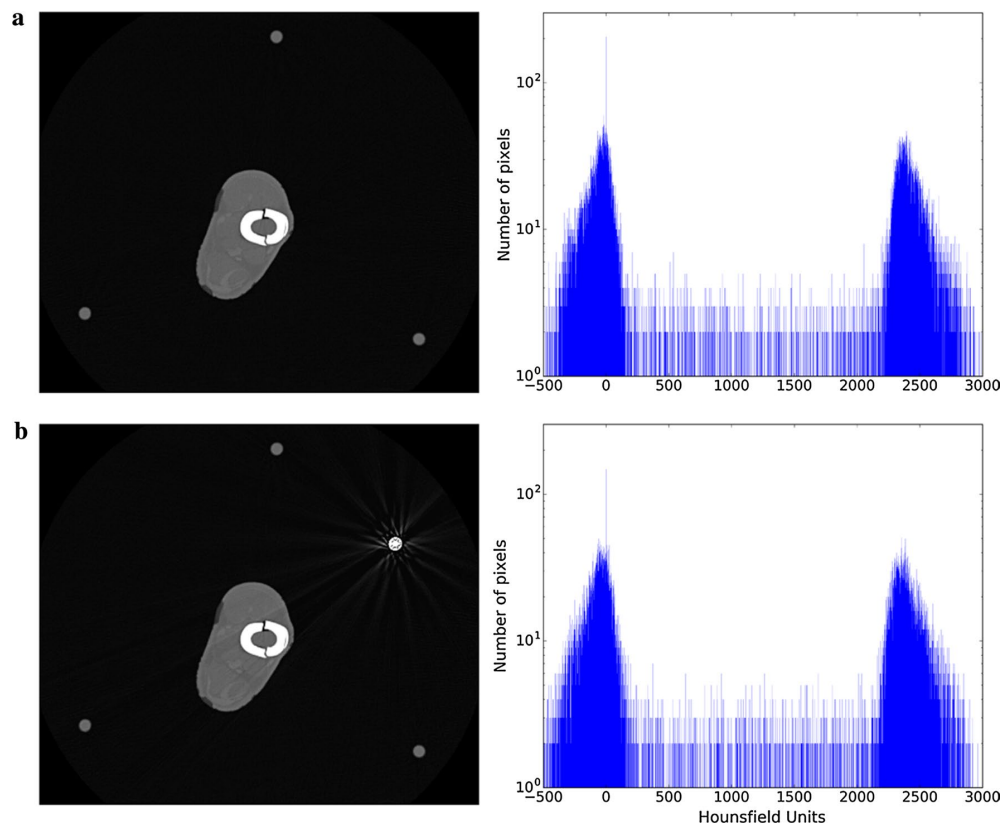


Fig. 4 **a** Axial slice and histogram of the fracture site when there are not metallic components present. **b** Axial slice and histogram of the fracture site when only one rod is present

Taylor Spatial Frame

Geometric shape: The oblique orientation of the struts in the frame results in a distinctive geometric pattern of broad streaks, the nature of which differs according to the level scanned (Fig. 7). There are slices where the ROI is not directly affected by the streak (Fig. 7a), whilst in others, it is crossed several times (Fig. 7b).

Strut material: The makeup of the strut differs along its length, with one end being tubular casing (Fig. 7c) and the

other end being a solid rod (Fig. 7a). In the middle, these two materials overlap to some degree (Fig. 7b). The severity of the broad streaks depends on which part of the strut is present in the particular slice.

Dual-energy scanning/metal artefact reduction software

Figures 8 and 9 show the effect of techniques used to reduce metal artefacts with the four-rod Ilizarov and TSF constructs,

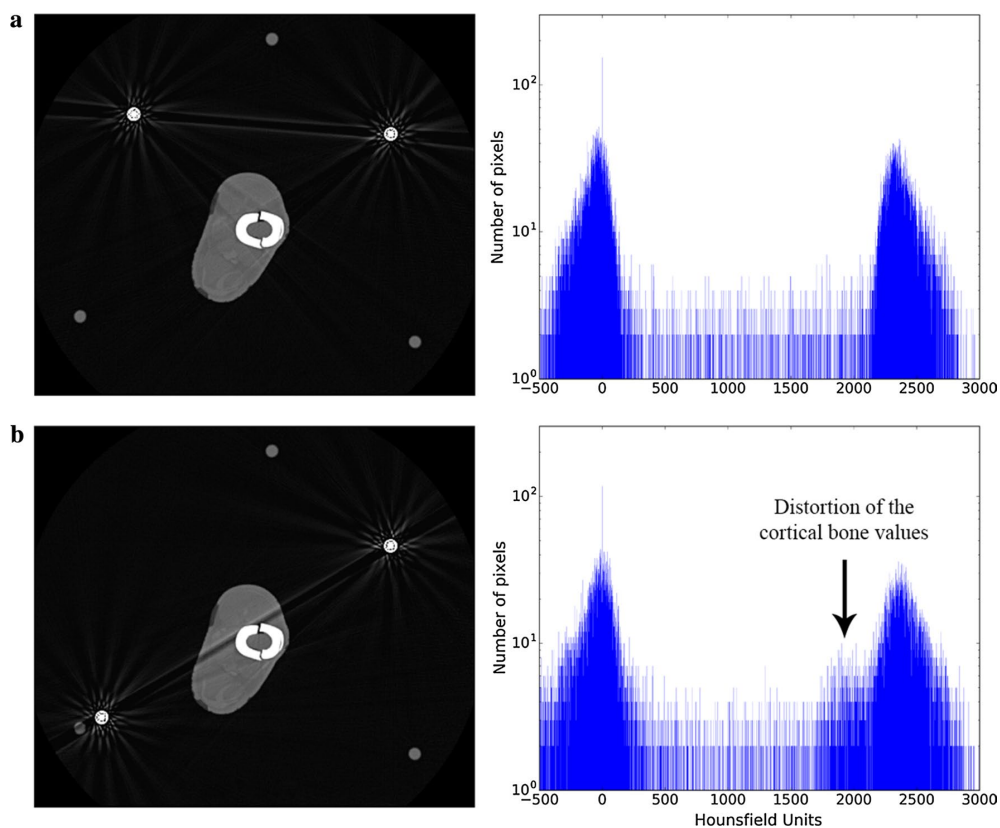


Fig. 5 Axial slice and histogram of the fracture site when there are two metal rods. **a** The rods are located at the same side of the leg. **b** The rods are located at opposite sides of the leg. The rod coupling goes through the ROI

respectively. Figures 8a and 9a present the single-energy scans shown previously, whilst Figs. 8b and 9b are scans performed using DECT, and Figs. 8c and 9c present scans using DECT and MARS. Both of these modalities are associated with a decrease in the overall clarity of the image, particularly so with the MARS. For the histograms, the higher-energy scan has altered the HU value for bone from its mean of 2400 on the single-energy scan to around 1100, whilst the addition of MARS further results in the peaks tending to merge into one.

Discussion

Obtaining good-quality images is a key part of the decision-making process for patients with circular external fixation. Frequently, nearby metal can obscure or have an adverse

effect on details of the images obtained with plain film radiography and CT [3, 8].

The idea for this study came about because it is the practice of the senior author (PCH) to remove as many rods as possible (or exchange struts for rods) when CT imaging. Typically, this is done whilst the patient is lying on the CT bed, so that the temporarily weakened frame is not subjected to undue forces. The rods are then reinserted after the scan has been performed and before the patient gets off the bed. Since the scan itself is relatively quick, the surgeon may end up being present during the whole process of removal, scanning, and reattachment. It requires that the scan is performed at a time when the surgeon is available and is clearly time-consuming, for both the surgeon and the CT department. It therefore raises the question of how much benefit is derived by such a practice; is the improvement in image

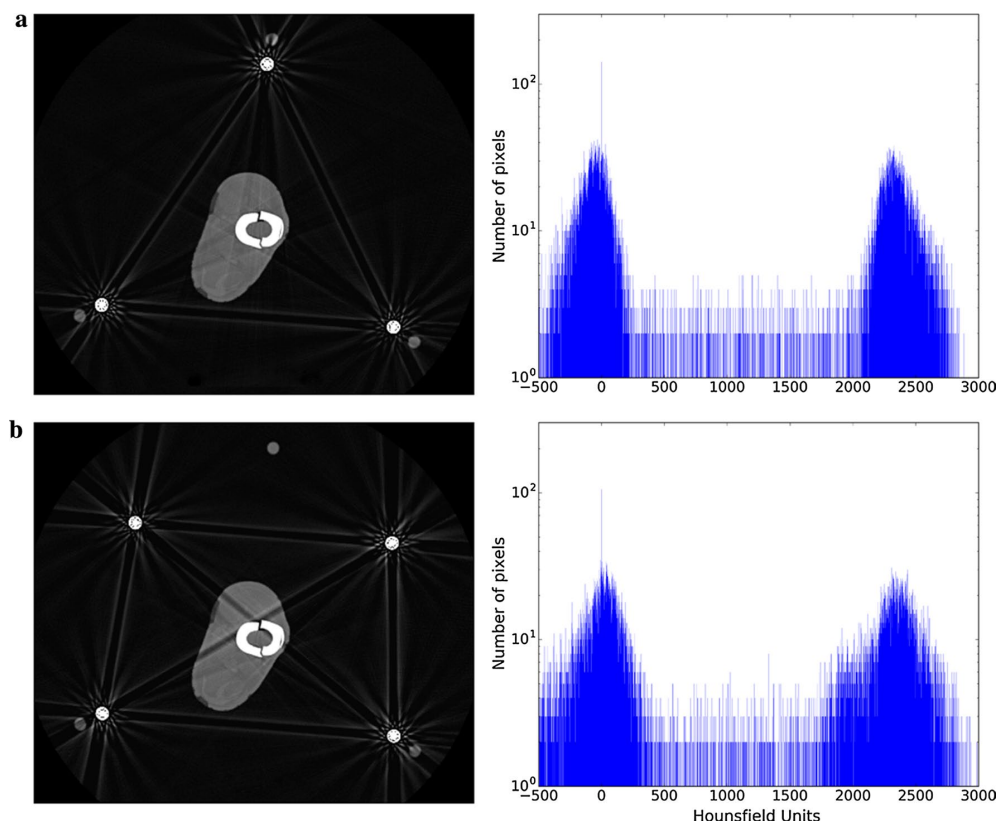


Fig. 6 **a** Axial slice and histogram of the fracture site when there are three metal rods present. **b** Axial slice and histogram of the fracture site in the presence of four metal rods

quality really worth all that effort? The aim of our study was therefore to investigate how the presence of rods/struts affects image quality. We chose a sheep leg over a human subject in order to scan it as many times as needed without concern for the adverse effects that ionising radiation can have on living humans.

The degradation of the CT image arises from the interaction between the poly-energetic X-ray beam and dense structures [4], creating two distinctive effects. The first effect (generalised noise proportional in degree to the overall amount of metal present on the axial slice) is seen as fine dark and bright streaks on the image [9, 10]. In this case, the artefact is distributed fairly proportionately throughout the image. The second effect, which is due to a pairing between rods and struts, causes a more noticeable broad dark streak

with surrounding bright edges, and is a function of the helical manner in which the scan is acquired (Fig. 10). Whilst the general existence of artefacts from geometric considerations has been previously acknowledged [8, 11], our study focuses on its relevance for circular external fixation, for which we have coined the term *rod coupling*. Our study demonstrates that, in the case of circular external fixation, it is rod coupling that is the main factor causing degradation of image quality. The histograms show how changes in the values of bone and soft tissue due to the metal artefacts affect the contrast of the image. The spreading of the peaks in the histograms, which corresponds to bone and soft tissue, in general causes a loss of visibility of the trabecular structure which lays between the peaks, and therefore loses definition.

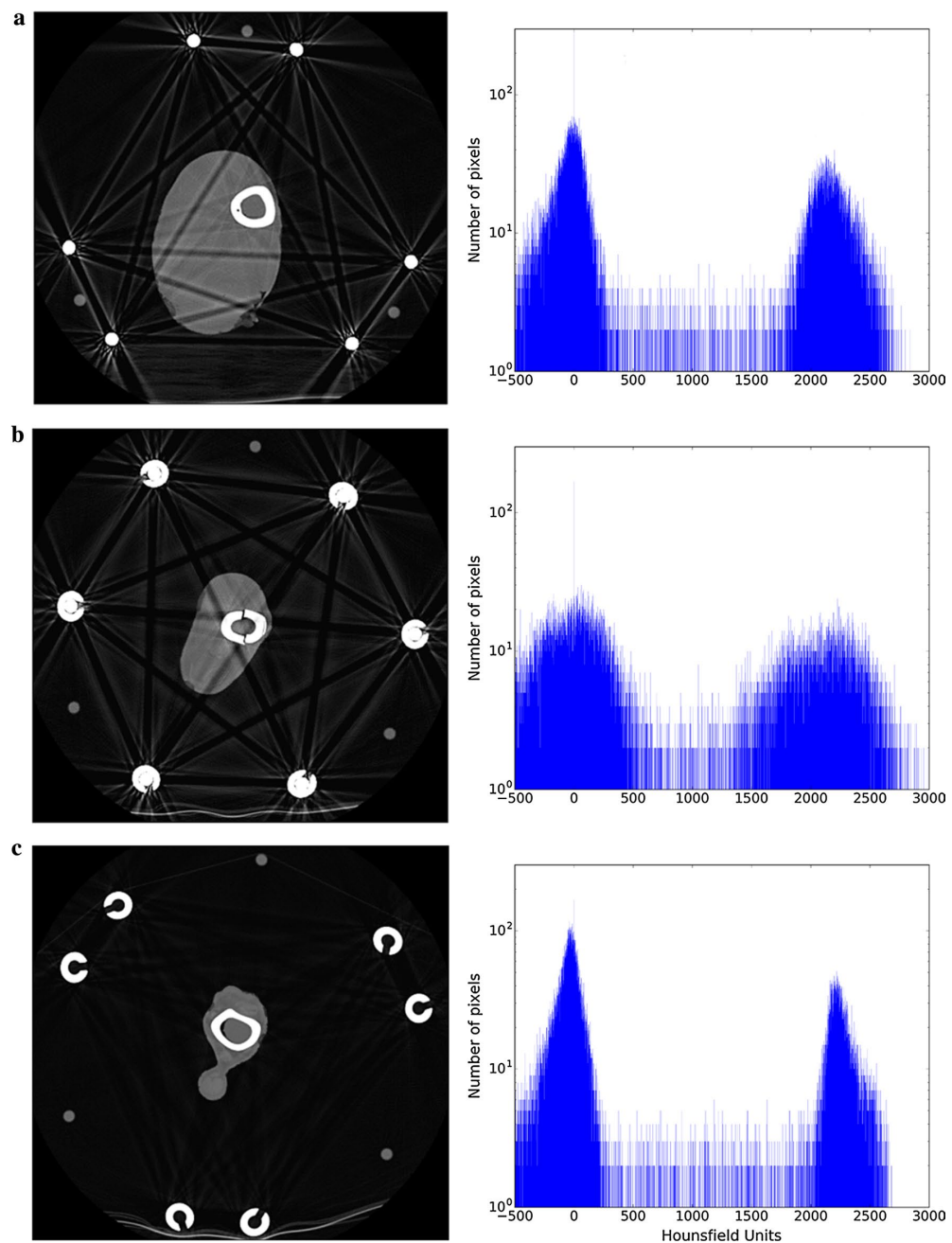


Fig. 7 Axial slices and histograms of the leg when using the TSF. **a** Slice proximal to fracture site, with only the solid rods of the strut, can be observed. **b** Slice of the fracture site, showing a mixture of

the solid rod and the tubular casing. **c** Slice distal to the fracture site, where only the tubular casing of the struts is observed

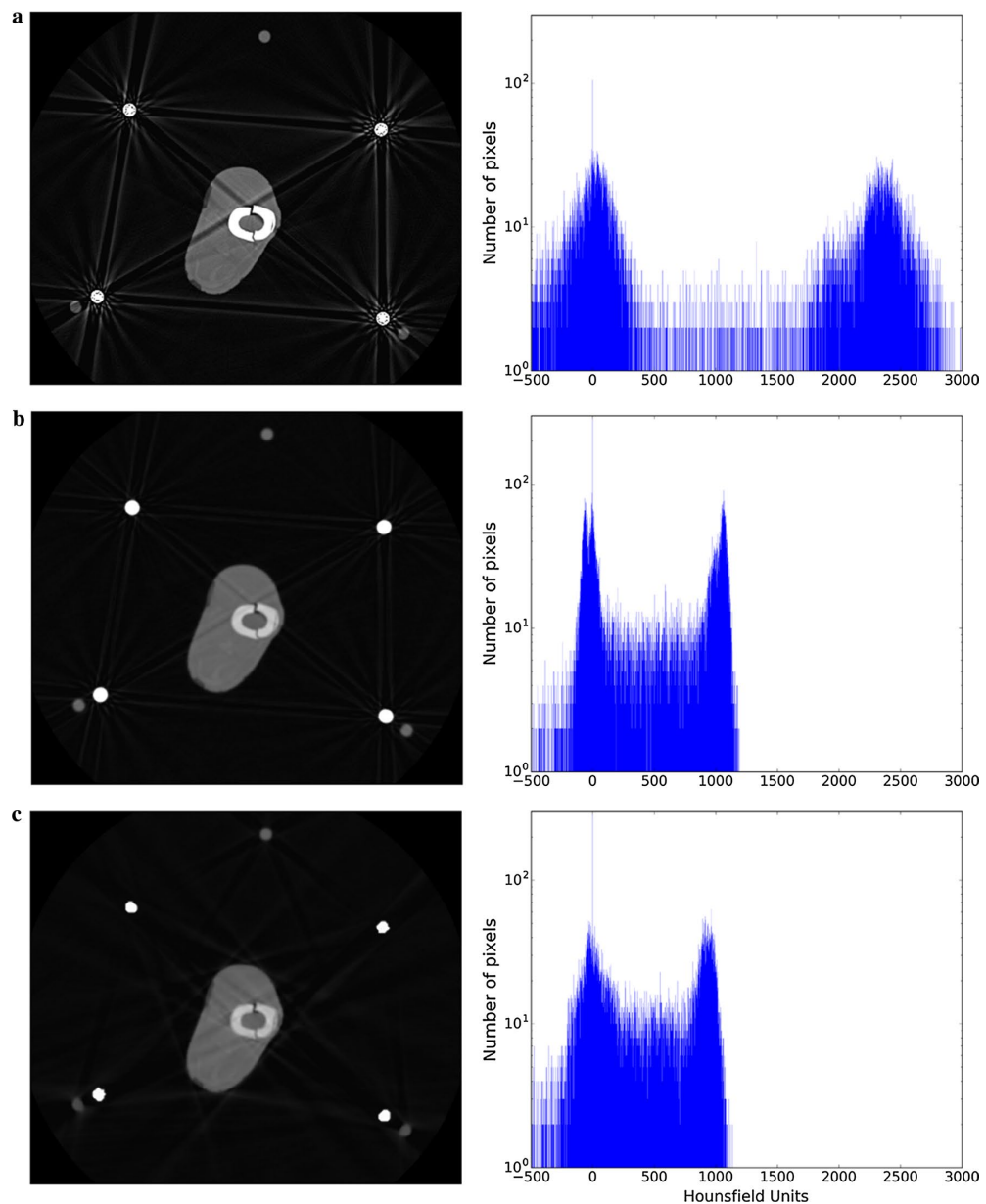


Fig. 8 Effect of metal artefact reduction techniques in image clarity in the presence of 4 metal rods. **a** Single-energy scan. **b** DECT. **c** DECT+MARS. The bone peak moves to the left for DECT and

DECT+MARS, which causes this tissue to become darker in the images when the window is conserved

Whilst we acknowledge that mechanical reasons predominantly dictate the rod/strut configurations when building a circular frame, the effect of rod coupling during CT imaging can be minimised. Firstly, all or all but one rod can be removed during imaging; we see this as the ideal (Fig. 4),

as in these cases there is no rod coupling. If this is not a possibility at that particular time, the second option is to configure the rods in a way that minimises the effect of the rod coupling on the area of interest. It can be appreciated in Fig. 6 that in the case of a four-rod construct, simply

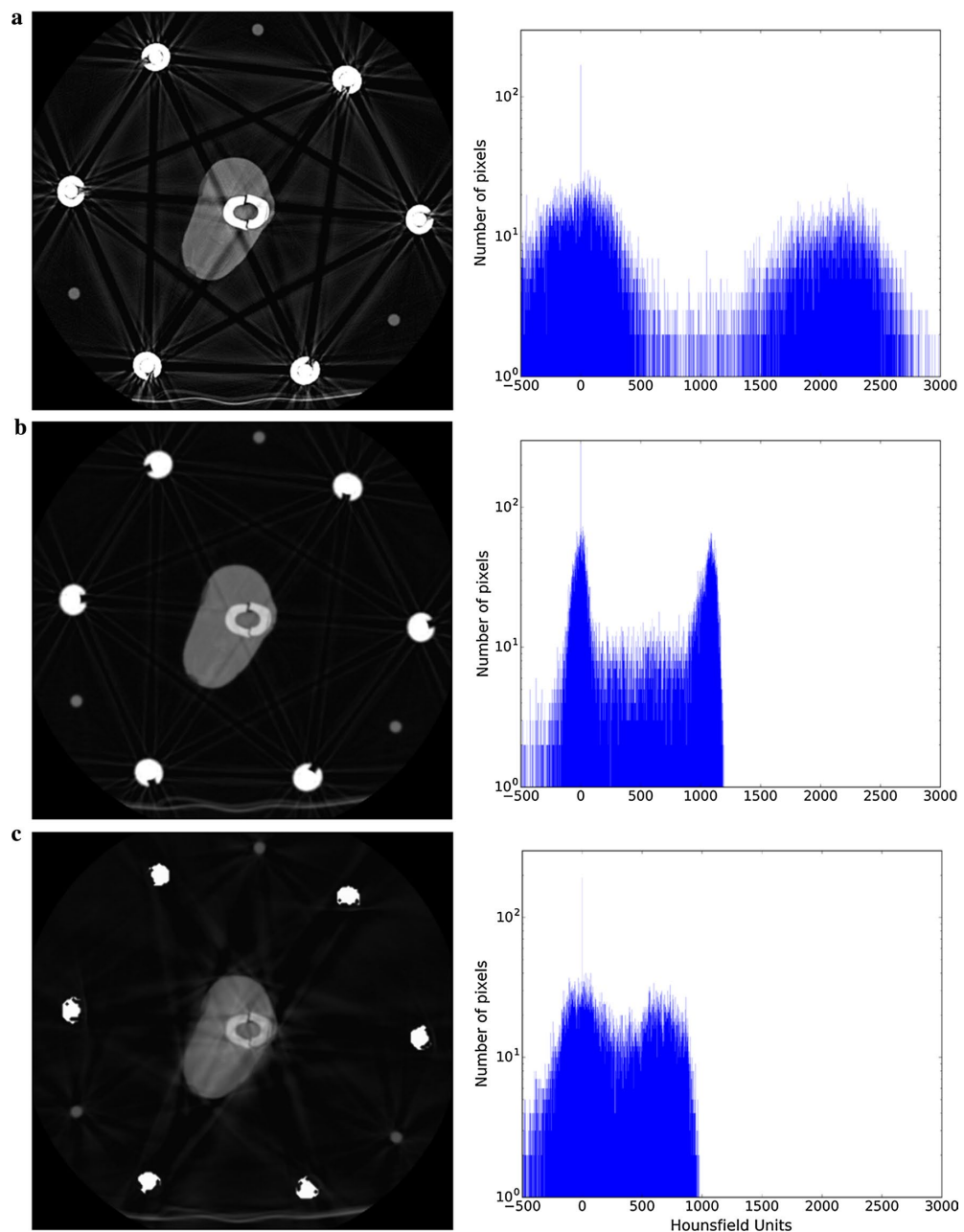


Fig. 9 Effect of metal artefact reduction techniques in image clarity in the presence of the TSF. **a** Single-energy scan. **b** DECT. **c** DECT + MARS

switching the two anterior rods to a more central single one will achieve this goal without particularly jeopardising mechanical strength for a short period. In the case of hexapod frames, whilst temporarily switching struts for rods

would also appear to be a good idea, this can be more time-consuming, as often with hexapods the rings are not parallel, and therefore requiring the use of dished (conical) washers or the construction of hinges.

Fig. 10 Helical nature of the CT scanning process, with X-ray source and detectors shown in three positions A–C. When the source completes a full rotation, a reconstruction algorithm converts the information to a 2D image slice. These slices can be displayed individually or assembled together to generate a 3D volume of the patient

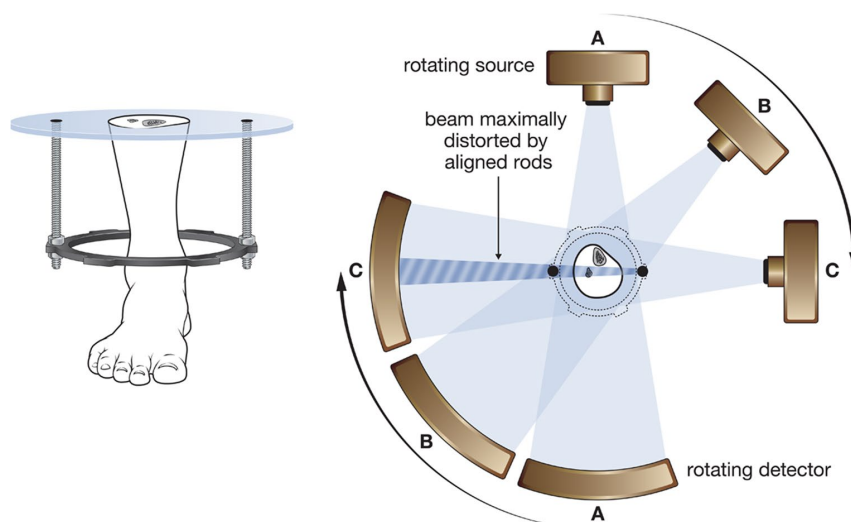


Table 1 Mean and width values for soft tissue and bone peaks as calculated from the histograms when the number of metal rods is changed in single-energy scans

Number of rods/figure number	Soft tissue		Bone	
	Mean (HU)	Width (HU)	Mean (HU)	Width (HU)
0/Figure 4a	–38	207	2396	251
1/Figure 4b	–48	237	2377	280
2 (coupling outside ROI)/Fig. 5a	–44	244	2355	298
2 (coupling crossing ROI)/Fig. 5b	–23	284	2382	388
3/Figure 6a	–50	284	2337	322
4/Figure 6b	16	367	2344	505

Table 2 Mean HU and width values for soft tissue and bone peaks as calculated from the histograms of different slices of the TSF single-energy CT scan

TSF Figure number	Soft tissue		Bone	
	Mean (HU)	Width (HU)	Mean (HU)	Width (HU)
Figure 7a	–18	256	2160	416
Figure 7b	–13	682	2111	846
Figure 7c	–33	190	2231	212

Can the CT technician configure the scan in a way that negates the need to remove or reconfigure rods/struts? In DECT, scans are acquired using two different energy settings and then synthesising pseudo-monochromatic scans at a variety of energies [12, 13]. The advantage of pseudo-monochromatic reconstructions is that the effect of streaking due to beam hardening, typically seen in CT imaging, is greatly reduced. Whilst the higher-energy setting reduces photon starvation and beam hardening, it results in some loss of detail (analogous to an over-penetrated plain X-ray). MARS works by evening out unexpected variation

in pixels, but our study suggests that the image it produces compares poorly with the original; the mean HU values change considerably.

One limitation of our study relates to assessment of the quality of an image. Image quality is determined by, amongst other things, the resolution and the contrast. In this paper, our assessment/comparison of the quality of the 2D images is purely subjective. Whilst efforts are being made to produce an objective scoring system for the quality of digital images in general [14–16], to our knowledge no such score currently exists for radiography. By making our results section predominantly a display of images, we have allowed the reader to draw their own conclusions about the effects of the various constructs and metal artefact reduction modalities on image quality. Whilst Gaussian-fitting analysis of the histograms produces some degree of quantitative data analysis, it is a relatively crude measure of contrast and is not useful in clinical orthopaedic practice. For readers that are familiar with the attributes of Gaussian peak fitting, the quantitative parameters extracted from the peak-fitting procedure are displayed in Tables 1, 2, 3, and 4. The general observation is that as the peaks get

Table 3 Mean HU and width values for soft tissue and bone peaks as calculated from the histograms of four-rod Ilizarov construct scans with single-energy CT, DECT, and DECT+MARS

Ilizarov (4 rods) Figure number/modality	Soft tissue		Bone	
	Mean (HU)	Width (HU)	Mean (HU)	Width (HU)
Figure 8a/single energy	16	367	2344	505
Figure 8b/DECT	− 22	132	1040	141
Figure 8c/DECT+MARS	5	284	922	181

Table 4 Mean HU and width values for soft tissue and bone peaks as calculated from the histograms of TSF construct scans with single-energy CT, DECT, and DECT+MARS

TSF Figure number/modality	Soft tissue		Bone	
	Mean (HU)	Width (HU)	Mean (HU)	Width (HU)
Figure 9a/single energy	− 13	682	2111	846
Figure 9b/DECT	− 4	125	1067	160
Figure 9c/DECT+MARS	13	430	618	486

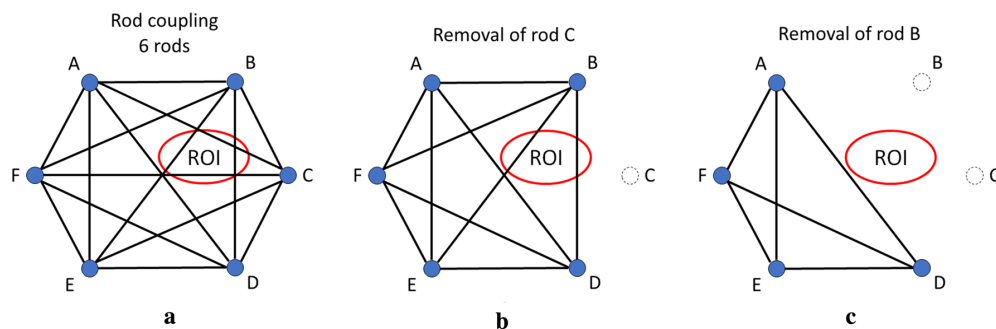


Fig. 11 Schematic representation of rod coupling in the presence of 6 rods and the effect of rod removal on the metal artefacts. **a** Complete graph. **b** Resulting coupling after removal of rod C. **c** Resulting coupling after the removal of rods B and C

broader and closer together, the ability to resolve features relating to those peaks diminishes. It can be seen that for Figs. 8 and 9, the data in Tables 3 and 4 demonstrate that the application of MARS reduces the level of artefacts but also affects the separability between the peaks.

From our perspective, the key point of this study is an appreciation of the phenomenon of rod coupling. Before CT scanning a limb with a circular external fixator, a thoughtful analysis should be made as to the selection of which rods to remove, or if they cannot be removed, how to configure them to avoid the worst effect of the metal artefacts on the region of interest. In our opinion, the best option is to temporarily reduce the number of rods to the bare minimum and/or to avoid rods at opposite sides of the ROI. That way, the surgeon can optimise image quality in the ROI, making it potentially easier to take medical decisions that impact the treatment and/or recovery of the patient.

Removal of rods

A rod-coupling line goes straight from every rod to every other rod (Fig. 11a; case with six rods). When a rod is removed, all lines connected to it are also deleted. One wishes to remove as many rods as are needed, to not have any rod-coupling lines going through the ROI. For example, if the maximum number of rods that can be left remaining is 3, then one could remove B + D + F, so as to have no rod couplings going through the ROI. A systematic process would be as follows:

1. Identify the rod(s) with the greatest number of lines passing through the ROI. In Fig. 11, B and C are the only rods with more than one line passing through the ROI. So, rod C is removed (Fig. 11b).
2. Again, we identify the rod(s) with the greatest number of lines passing through the ROI. Removing rod B will

reduce the number of rod couplings in the ROI by two (Fig. 11c). If more than one rod reduces the number of rod couplings in the ROI by the same amount, a different criterion can be chosen, for example, if a more rigid frame can be obtained by removing one and not the other.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest, and no funds have been received in connection with this study.

Ethical approval This experiment used a sheep leg that had been used in experiments conducted with approval from the Monash University (Australia) Animal Ethics Committees. All experiments were performed in accordance with relevant guidelines and regulations. The sheep was humanely killed in line with approved guidelines and the carcasses scavenged for this experiment.

Informed consent For this type of study (animal research) formal patient consent is not required.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Fragomen AT, Rozbruch SR (2007) The mechanics of external fixation. *HSS J* 3(1):13–29
2. Wright J, Sabah SA, Patel S, Spence G (2017) The silhouette technique: improving post-operative radiographs for planning of correction with a hexapod external fixator. *Strategies Trauma Limb Reconstr* 12(2):127–131
3. Coupal TM, Mallinson PI, McLaughlin P, Nicolaou S, Munk PL, Ouellette H (2014) Peering through the glare: using dual-energy CT to overcome the problem of metal artefacts in bone radiology. *Skelet Radiol* 43(5):567–575
4. Huang JY, Kerns JR, Nute JL, Liu X, Balter PA, Stingo FC, Followill DS, Mirkovic D, Howell RM, Stingo FC, Kry SF (2015) An evaluation of three commercially available metal artifact reduction methods for CT imaging. *Phys Med Biol* 60(3):1047
5. Kotsenas AL, Michalak GJ, DeLone DR, Diehn FE, Grant K, Halaweish AF, Krauss A, Raupach R, Schmidt B, McCollough CH, Fletcher JG (2015) CT metal artifact reduction in the spine: can an iterative reconstruction technique improve visualization? *Am J Neuroradiol* 36(11):2184–2190
6. Morsbach F, Bickelhaupt S, Wanner GA, Krauss A, Schmidt B, Alkadhi H (2013) Reduction of metal artifacts from hip prostheses on CT images of the pelvis: value of iterative reconstructions. *Radiology* 268(1):237–244
7. Hsieh J (2009) *Computed tomography: principles, design, artifacts, and recent advances*. Bellingham, Washington
8. Buckwalter KA, Lin C, Ford JM (2011) Managing postoperative artifacts on computed tomography and magnetic resonance imaging. *Semin Musculoskelet Radiol* 15(4):309–319
9. Gupta A, Subhas N, Primak AN, Nittka M, Liu K (2015) Metal artifact reduction: standard and advanced magnetic resonance and computed tomography techniques. *Radiol Clin* 53(3):531–547
10. Meyer E, Raupach R, Lell M, Schmidt B, Kachelrieß M (2012) Frequency split metal artifact reduction (FSMAR) in computed tomography. *Med Phys* 39(4):1904–1916
11. Berg BV, Malghe J, Maldague B, Lecouvet F (2006) Multi-detector CT imaging in the postoperative orthopedic patient with metal hardware. *Eur J Radiol* 60(3):470–479
12. Lee YH, Park KK, Song HT, Kim S, Suh JS (2012) Metal artifact reduction in gemstone spectral imaging dual-energy CT with and without metal artifact reduction software. *Eur Radiol* 22(6):1331–1340
13. Pessis E, Campagna R, Sverzut JM, Bach F, Rodallec M, Guerini H, Feydy A, Drapé JL (2013) Virtual monochromatic spectral imaging with fast kilovoltage switching: reduction of metal artifacts at CT. *Radiogr* 33(2):573–583
14. Gao F, Tao D, Gao X, Li X (2015) Learning to rank for blind image quality assessment. *IEEE Trans Neural Netw Learn Syst* 26(10):2275–2290
15. Hou W, Gao X, Tao D, Li X (2015) Blind image quality assessment via deep learning. *IEEE Trans Neural Netw Learn Syst* 26(6):1275–1286
16. Xue W, Mou X, Zhang L, Bovik AC, Feng X (2014) Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Trans Image Process* 23(11):4850–4862

5.2 Deep learning denoising of a synthetically generated leg database

The numerical model proposed in Chapter 4 was utilised to generate a synthetic database of lower extremity volumetric data. Each volume approximated the CT reconstruction of a femur and comprised three concentric elliptic cylinders on a background with appropriate Hounsfield unit values; from outermost to innermost ellipses, the values mimicked CT values for soft tissue, bone, and soft tissue, to simulate the bone marrow (Figure 5.2a, left).

The rod coupling observed in Section 5.1, consisting of a pairing between highly-dense objects, causes a noticeable broad dark streak with surrounding bright edges. The dark streaks were reproduced in the synthetic database by incorporating a function in the axial view that generates appropriately weighted dark and light bands between highly-dense cylinders that were placed radially around the leg. The voxels in contact with the generated line were multiplied by a factor smaller than 1, making them darker in the resulting image. For the bright streaks, the width of the lines was increased, and voxels in contact with the expanded section of the line were multiplied by a factor greater than 1. To obtain a pixel distribution similar to the observed in real CT scans of the leg, Gaussian noise was added to the synthetic volumes before the addition of rod coupling (Figure 5.2a, right), and the radii of the different tissues inside the synthetic leg were randomly selected to improve the algorithm’s generalisability.

In order to reduce the noise in medical images, several approaches can be found, including convolutional denoising autoencoders, CNNs with residual learning approaches, and generative adversarial networks (GANs). Convolutional denoising autoencoders [53] are based on CAEs, where the network is trained with the noisy image as input, and the noise-free image as output, intrinsically learning the process to remove the artefacts. CNNs with residual learning approaches [54] learn the noise from the noisy inputs, and the denoised images are obtained by subtracting the learned residual from the noisy image. GANs [72] are used to translate images between different domains, in this case between noisy and clean. A DNN with a similar architecture to the CAE presented in Chapter 3, called VNet [52], was implemented to reduce the metal artefacts of the synthetic volumes. V Nets present edge preserving properties that are not found in regular CAEs. Volumes of $8 \times 512 \times 512$ ($n=360$) were generated and divided into training, validation, and testing with a 70:15:15 ratio. Histograms obtained from the original volumes (before adding the dense objects), for the volumes with metal artefacts, and for the predicted volumes were compared (Table 5.1). The mean values and widths, or 2σ , obtained for soft tissue, were not significantly affected by the addition of the rod coupling. For bone, the average mean value decreased 7% and the width increased 51% when the rod coupling was added. The use of DL for denoising can be seen in Figure 5.2c, where the artefacts are still perceptible, but the Gaussian fit was similar to the histogram of the original volume. The average mean value and

Table 5.1: Mean and width values for Gaussians fitted to the histograms of the synthetic volumes, with and without metal artefacts, and after prediction on the test cases.

	Soft tissue		Bone	
	Mean [HU]	Width [HU]	Mean [HU]	Width [HU]
With artefacts (input)	35	656	2271	1072
Without artefacts (ideal)	38	709	2433	710
De-noised (output)	33	653	2350	704

width of the predicted volumes decreased only 3% and 1%, respectively, when compared to the average mean value and width of the original synthetic volumes.

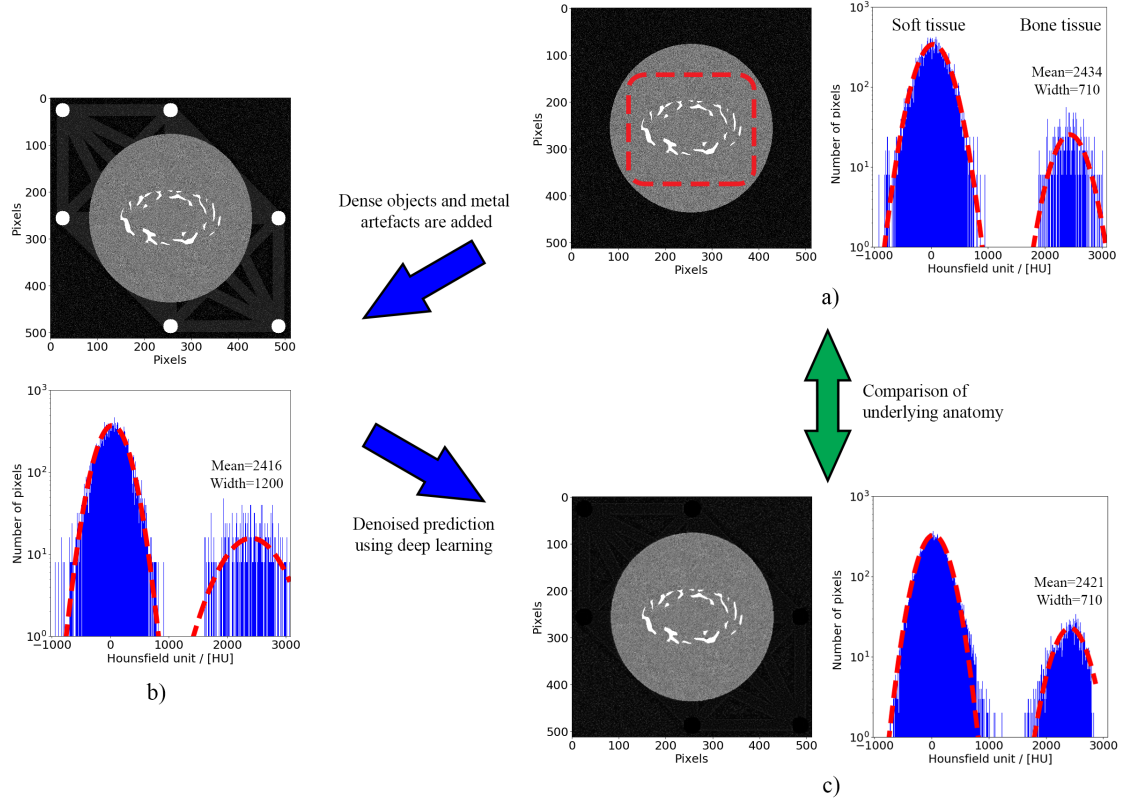


Figure 5.2: Sample of generated synthetic database and prediction using DL. a) Axial slice of synthetic volume next to the histogram and fitted Gaussian curves obtained from dashed patch in the image. The fitting for soft tissue and bone tissue can be observed. b) Axial slice and histogram after the addition of the dense objects and simulated metal artefacts. The standard deviations of the fitted curves increase due to the resulting noise. c) Predicted denoised version of the axial slice shown in b). Although the artefacts are still visible, the standard deviation of the Gaussian fitted to the bone tissue has decreased and is closer to that of the original image shown in a).

The amount of information that is being generated by health care systems around the world, and more specifically imaging data, is continuously increasing the burden on primary healthcare practitioners, e.g. radiologists and pathologists [3, 4]. ML approaches are being applied to automate and guide aspects of diagnostic and prognostic tasks [73–75] to improve workflow efficiency. Interactive or human-in-the-loop approaches [76], where algorithms can optimise their learning by interacting with radiologists or pathologists or even AI augmented radiologists/pathologists [77–81].

In order to achieve an appropriate level of performance, these ML algorithms require large databases of expert-labelled data. The current clinical workflows, typically involving the radiographer or radiologist, only require the labelling of small portions of case data, leaving large volumes unannotated [7, 8]. Since vast databases of images already exist, tools that can be used to standardise metadata formats and workflows, and to label data by means of content-based image retrieval processes (CBMIR) or automatic and semi-automatic classification (labelling) approaches are required. Standardisation, as shown by Gauriau et al. [82] and Willemink et al. [83], has the ability to unlock the potential of large-scale well-structured medical databases, while CBMIR allows for the mining of knowledge from big-data by providing searching capabilities using images as the query [84], and depending on the implementation, can also be used to assign tags to the cases in the database [85]. Both CBMIR and classification approaches have been successfully implemented in clinical databases using ML techniques [10, 86, 87], but the applications are still to be translated to the forensic setting. In contrast to the clinical setting, the use of readily available applications to forensic databases is not only hindered by the different reports accompanying forensic cases (e.g. autopsy reports, interviews, and crime scene related transcripts), but also by the high degree of variability found in forensic images due to positioning and CoD,

as was shown in this thesis.

The DL pipeline presented in this work is ultimately intended to provide a flexible yet robust method for reducing the burden on radiologists and pathologists in retrospectively labelling the extensive archives of CT scans that exist at forensic institutions. Although the localisation errors measured were in the same range as those previously reported in the clinical setting [88–91], the variability in the positioning of the decedent, objective of localising the whole femur (as opposed to just the femoral head) meant that the results of the network should be interpreted as robust. The subsequent automatic labelling of proximal and distal epiphyses according to the presence and type of metallic implant, with the addition of the decedent’s age confirmed the improvement in performance shown in the literature when a small subset of labelled data was added to an unlabelled set [92, 93]. It is important to note that a more exhaustive evaluation of the precision and recall values under different conditions is necessary, due to its potential impact in tagging and searching capabilities of the database.

The two main factors found to influence the localisation and labelling of organs in CT scans were studied: image artefacts, and patient-related variables, e.g. demographic information. Decedent condition can be characterised according to the level of incomplete anatomy in the CT scans and variability presented by the CoD. A review of the existing literature (Chapter 2) demonstrated that typical applications of ML in medical imaging relate to the classification of patients or tissues into healthy or diseased, and the localisation of pathologies in organs or structures observed in volumetric data. The performance of the techniques documented in the literature was found to be poor when tested on cases that were outside the norm (outliers), and the training dataset was not properly selected. In the forensic setting, this characteristic is of special importance due to the high variability in positioning and changes in the structures related to the CoD. The presence of a metallic implants and the rotation of the femur with respect to the motion axis of the scanner bed were found to be a significant factor in the localisation pipeline developed in this work. Demographic information such as age and BMI presented no significance for the presented task, as was also reported by Estrada et al. [94] for the task of adipose tissue segmentation in MRI.

The coupling artefacts commonly found in forensic CT databases [95] affect the visibility of structures throughout the body. The most significant coupling artefacts found in forensic databases result from the prevalence of bilateral knee replacements and hip replacements, whose incidence is expected to grow by 276% and 208%, respectively, by 2030 [96]. Coupling artefacts were also found to affect the thorax due to the presence of the arms besides the body [50], and in the jaw due to metal fillings in teeth [50]. The models developed in this thesis proved valuable in facilitating the quantitative analysis of how the denoising capabilities of a VNet could be used to remove these types of artefacts retrospectively,

improving the anatomical visibility of structures. The results showed that the bone tissue was significantly affected by the artefacts, with the width of the peak-fit increasing 39% when two highly-dense objects were used, compared to the width of the Gaussian when only one dense object was utilised. These types of artefacts have been shown to decrease the contrast between tissues [97], thus potentially affecting diagnostic accuracy and inter-radiologist and -pathologist agreement [98–100].

Coupling artefacts are commonly reduced by the utilisation of MARS and/or virtual mono-energetic (VME) images constructed from the DECT data [101]. In Chapter 5, it was found that these two approaches were not always appropriate due to changes caused in voxel distributions across the image. The results confirmed what was previously documented in the literature, in that the optimal noise reduction approach depended on the desired application, type of metal implant, and anatomical localisation [101, 102]. Furthermore, DECT acquisition must be selected prior to acquisition of the scan, while MARS was most commonly used during reconstruction or before saving the data to the PACS.

The use of ML in medical imaging is proving beneficial in enhancing retrospective data once it has been acquired. The most recent approaches for improving image quality utilising ML have shown promising results [103, 104], however, a major limitation of these techniques is the requirement for vast databases with high variability, to avoid decreased accuracy or non-generalisable models [105]. In order to tackle this limitation, synthetic data generation, data augmentation, and transfer learning (TL) are being investigated. Synthetic data generation is being widely studied using adversarial training with GANs. GANs consist of two ANNs, the generative network that tries to create novel sample candidates (similar to the training database) able to fool the second network, the discriminator, into thinking the sample is not synthetic. With enough training, the generator produces better images, and the discriminator becomes more skilled at flagging synthetic images. GANs can be used e.g. to generate CTs given MRIs [106] or to create synthetic abnormal MRI images with brain tumours [107]. Furthermore, data augmentation can be performed using the previously discussed synthetic data to increase the number of samples during training, or through direct image/voxel manipulation to create variations of a small dataset [108]. TL on the other hand entails the training of the required ML algorithm with a surrogate database from a more general domain, that is subsequently fine-tuned for the specific application [109, 110]. Since a large variability can be difficult to achieve in medical scans due to the number of pathologies, organ sizes, and positioning, the use of synthetic databases is being widely investigated [111–114]. The numerical model presented in this thesis was used to generate synthetic trabecular structures with the definition of only seven parameters, and has potential application for generating customised training data for bone pathologies, e.g. osteoporosis and subtle bone fractures.

In Section 5.2, a synthetic database that mimics CT scans of the human leg was generated. A VNet algorithm for denoising of the synthetic database was implemented. Although the denoised images still displayed subtle residual artefacts, the Gaussian fits to the histograms of the anatomical features in the images were closer to those in the ideal volumes generated prior to the inclusion of artefacts. A similar approach to quantifying the noise reduction capabilities of DL architectures was performed by Yang et al. [115]. In order to compare the results of our denoising approach with those found in the literature, TL would first be required to be implemented in order to fine-tune the network to a medical database.

Three different ML approaches (ResNet, CAE and VNet) and two synthetic models (one for trabecular bone and one for coupling artefacts) were developed and used in order to understand how a pipeline could be constructed to process retrospectively accessed CT volumes from forensic databases.

The performance of a DL pipeline that utilised coupled ResNets and a CAE for the automatic retrospective localisation and labelling of organs in forensic databases was quantified and contextualised with reference to parameters relevant to forensic investigations. The pipeline was applied to femoral bones, with intrinsic features of the localised organ being utilised to automatically classify the forensic cases as healthy, or with the presence of knee replacements, hip replacements, or nails.

The two synthetic models were combined to understand the significant artefacts resulting from the coupling of dense objects in the field of view of the CT scanner and the effect of the use of ML for denoising. Coupling artefacts commonly found in forensic databases can result from metallic implants and positioning of the body on the scanner bed. The VNet approach was found to offer significant improvements in object visibility as quantified by fitting the histograms of the resulting images.

Future work applications include the use of the DL pipeline to the lungs. The lungs were selected for this application due to the added difficulty caused by the low contrast between neighbouring tissues, and the forensic importance due to the potential for automatic labelling of the fluid content that can be related to the CoD. Furthermore, the DL denoising approach will be fine-tuned and tested in real CT scans with the presence of coupling artefacts like those shown in Section 5.1.

Bibliography

- [1] Ouakrim DA, Pizot C, Boniol M, Malvezzi M, Boniol M, Negri E, et al. Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. *BMJ*. 2015;p. h4970.
- [2] Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, et al. Identification of physician-diagnosed Alzheimer’s disease and related dementias in population-based administrative data: a validation study using family physicians’ electronic medical records. *J Alzheimers Dis*. 2016;54(1):337–349.
- [3] Manogaran G, Thota C, Lopez D, Vijayakumar V, Abbas KM, Sundarsekar R. Big Data Knowledge System in Healthcare. In: *Studies in Big Data*. Springer International Publishing; 2017. p. 133–157.
- [4] Wang Y, Kung L, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technol Forecast Soc*. 2018;126:3–13.
- [5] Kim Y, Kim YK, Lee BE, Lee SJ, Ryu YJ, Lee JH, et al. Ultra-Low-Dose CT of the Thorax Using Iterative Reconstruction: Evaluation of Image Quality and Radiation Dose Reduction. *Am J Roentgenol*. 2015;204(6):1197–1202.
- [6] Nakamura S, Kobayashi T, Funatsu A, Okada T, Mauti M, Waizumi Y, et al. Patient radiation dose reduction using an X-ray imaging noise reduction technology for cardiac angiography and intervention. *Heart Vessels*. 2015;31(5):655–663.
- [7] Netto SMB, Silva AC, Nunes RA, Gattass M. Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Comput Biol Med*. 2012;42(11):1110–1121. Available from: <https://doi.org/10.1016/j.compbiomed.2012.09.003>.
- [8] Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. *Med Image Anal*. 2018;45:94–107.

- [9] de Bruijne M. Machine learning approaches in medical image analysis: From detection to diagnosis. *Med Image Anal.* 2016;33:94–97.
- [10] Qayyum A, Anwar SM, Awais M, Majid M. Medical image retrieval using deep convolutional neural network. *Neurocomputing.* 2017;266:8–20.
- [11] Pakhomov D, Premachandran V, Allan M, Azizian M, Navab N. Deep Residual Learning for Instrument Segmentation in Robotic Surgery. In: *Lect Notes Comput Sc.* Springer International Publishing; 2019. p. 566–573.
- [12] Huang H, Lin J, Wu L, Fang B, Wen Z, Sun F. Machine learning-based multi-modal information perception for soft robotic hands. *Tsinghua Sci Technol.* 2020;25(2):255–269.
- [13] Corcoran E, Denman S, Hanger J, Wilson B, Hamilton G. Automated detection of koalas using low-level aerial surveillance and machine learning. *Sci Rep.* 2019;9(1).
- [14] Ramkumar PN, Haeberle HS, Ramanathan D, Cantrell WA, Navarro SM, Mont MA, et al. Remote Patient Monitoring Using Mobile Health for Total Knee Arthroplasty: Validation of a Wearable and Machine Learning–Based Surveillance Platform. *J Arthroplasty.* 2019;34(10):2253–2259.
- [15] Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* 2017;284(2):574–582.
- [16] Wang J, Ding H, Azamian F, Zhou B, Iribarren C, Molloy S, et al. Detecting cardiovascular disease from mammograms with deep learning. *IEEE T Med Imaging.* 2017;36(5):1172–1181.
- [17] Singh G, Al’Aref SJ, Assen MV, Kim TS, van Rosendael A, Kolli KK, et al. Machine learning in cardiac CT: Basic concepts and contemporary data. *J Cardiovasc Comput.* 2018;12(3):192–201.
- [18] Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer.* 2018;115:34–41.
- [19] Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* 2017;19(6):1236–1246.
- [20] Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Panykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology.* 2018;288(2):318–328.
- [21] Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F. Early diagnosis of Alzheimer’s disease based on partial least squares, principal component analysis

- and support vector machine using segmented MRI images. *Neurocomputing*. 2015;151:139–150. Available from: <https://doi.org/10.1016/j.neucom.2014.09.072>.
- [22] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv e-prints*. 2017;p. arXiv:1711.05225.
 - [23] Shin HC, Lu L, Summers RM. Natural Language Processing for Large-Scale Medical Image Analysis Using Deep Learning. In: *Deep Learning for Medical Image Analysis*. Elsevier; 2017. p. 405–421.
 - [24] Savova GK, Danciu I, Alamudun F, Miller T, Lin C, Bitterman DS, et al. Use of Natural Language Processing to Extract Clinical Cancer Phenotypes from Electronic Medical Records. *Cancer Res*. 2019;79(21):5463–5470.
 - [25] Montagnon E, Cerny M, Cadrin-Chênevert A, Hamilton V, Derennes T, Ilinca A, et al. Deep learning workflow in radiology: a primer. *Insights Into Imaging*. 2020 feb;11(1).
 - [26] Skripcak T, Belka C, Bosch W, Brink C, Brunner T, Budach V, et al. Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets. *Radiother Oncol*. 2014;113(3):303–309.
 - [27] Sierink JC, Saltzherr TP, Russchen MJAM, de Castro SMM, Beenen LFM, Schep NWL, et al. Incidental findings on total-body CT scans in trauma patients. *Inj*. 2014;45(5):840–844.
 - [28] Hoang JK, Langer JE, Middleton WD, Wu CC, Hammers LW, Cronan JJ, et al. Managing Incidental Thyroid Nodules Detected on Imaging: White Paper of the ACR Incidental Thyroid Findings Committee. *J Am Coll Radiol*. 2015;12(2):143–150.
 - [29] Ross SG, Bolliger SA, Ampanozi G, Oesterhelweg L, Thali MJ, Flach PM. Post-mortem CT angiography: capabilities and limitations in traumatic and natural causes of death. *Radiographics*. 2014;34(3):830–846. Available from: <https://doi.org/10.1148/rg.343115169>.
 - [30] Colard T, Delannoy Y, Bresson F, Marechal C, Raul JS, Hedouin V. 3D-MSCT imaging of bullet trajectory in 3D crime scene reconstruction: Two case reports. *Leg Med*. 2013;15(6):318–322.
 - [31] Seim H, Kainmueller D, Heller M, Lamecker H, Zachow S, Hege HC. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model.

- In: Proceedings of the First Eurographics Conference on Visual Computing for Biomedicine. EG VCBM'08. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2008. p. 93–100.
- [32] Kainmueller D, Lamecker H, Zachow S, Hege H. An articulated statistical shape model for accurate hip joint segmentation. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2009. p. 6345–6351.
 - [33] Dworzak J, Lamecker H, von Berg J, Klinder T, Lorenz C, Kainmüller D, et al. 3D reconstruction of the human rib cage from 2D projection images using a statistical shape model. *Int J Comput Ass Rad.* 2010;5(2):111–124.
 - [34] Harvey EC, Feng M, Ji X, Zhang R, Li Y, Chen GH, et al. Impacts of photon counting CT to maximum intensity projection (MIP) images of cerebral CT angiography: theoretical and experimental studies. *Phys Med Biol.* 2019;64(18):185015.
 - [35] Chaplot S, Patnaik L, Jagannathan N. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Bio-med Signal Proces.* 2006;1(1):86–92. Available from: <https://doi.org/10.1016/j.bspc.2006.05.002>.
 - [36] Norajitra T, Meinzer HP, Maier-Hein KH. 3D statistical shape models incorporating 3D random forest regression voting for robust CT liver segmentation. In: *Medical Imaging 2015: Computer-Aided Diagnosis.* vol. 9414. International Society for Optics and Photonics; 2015. p. 941406.
 - [37] Jiang J, Trundle P, Ren J. Medical image analysis with artificial neural networks. *Comput Med Imag Grap.* 2010;34(8):617–631. Available from: <https://doi.org/10.1016/j.compmedimag.2010.07.003>.
 - [38] Camastra F, Vinciarelli A. *Machine learning for audio, image and video analysis.* Springer; 2008.
 - [39] Yu S, Tan KK. Classification of lumbar ultrasound images with machine learning. In: *Asia-Pacific Conference on Simulated Evolution and Learning.* Springer; 2014. p. 287–298.
 - [40] Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging.* 2017;30(4):427–441. Available from: <https://doi.org/10.1007/s10278-017-9955-8>.
 - [41] Beers A, Chang K, Brown J, Gerstner E, Rosen B, Kalpathy-Cramer J. Sequential neural networks for biologically-informed glioma segmentation. In: *Medical Imaging 2018: Image Processing.* vol. 10574. International Society for Optics and Photonics; 2018. p. 1057433. Available from: <https://doi.org/10.1117/12.2293941>.

- [42] He Y, Carass A, Jedynak BM, Solomon SD, Saidha S, Calabresi PA, et al. Topology guaranteed segmentation of the human retina from OCT using convolutional neural networks. arXiv e-prints. 2018;p. arXiv:1803.05120.
- [43] Štern D, Payer C, Lepetit V, Urschler M. Automated age estimation from hand MRI volumes using deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016. p. 194–202. Available from: https://doi.org/10.1007/978-3-319-46723-8_23.
- [44] Ebert LC, Heimer J, Schweitzer W, Sieberth T, Leipner A, Thali M, et al. Automatic detection of hemorrhagic pericardial effusion on PMCT using deep learning—a feasibility study. *Forensic Sci Med Pat.* 2017;13(4):426–431. Available from: <https://doi.org/10.1007/s12024-017-9906-1>.
- [45] Arigbabu OA, Liao IY, Abdullah N, Mohamad Noor MH. Computer vision methods for cranial sex estimation. *IPSJ Trans Comput Vis Appl.* 2017;9(1):19. Available from: <https://doi.org/10.1186/s41074-017-0031-6>.
- [46] Heimer J, Thali MJ, Ebert L. Classification based on the presence of skull fractures on curved maximum intensity skull projections by means of deep learning. *J Forensic Radiol Imaging.* 2018;14:16 – 20. Available from: <https://doi.org/10.1016/j.jofri.2018.08.001>.
- [47] Gureyev T, Mayo S, Myers D, Nesterets Y, Paganin D, Pogany A, et al. Refracting Röntgen’s rays: propagation-based X-ray phase contrast for biomedical imaging. *J Appl Phys.* 2009;105(10):102005.
- [48] Morgan K, Siu KK, Paganin D. The projection approximation and edge contrast for X-ray propagation-based phase contrast imaging of a cylindrical edge. *Opt Express.* 2010;18(10):9865–9878.
- [49] Wang G, Snyder DL, O’Sullivan JA, Vannier MW. Iterative deblurring for CT metal artifact reduction. *IEEE T Med Imaging.* 1996;15(5):657–664.
- [50] Barrett JF, Keat N. Artifacts in CT: Recognition and Avoidance. *RadioGraphics.* 2004;24(6):1679–1691.
- [51] Peña-Solórzano CA, Dimmock MR, Albrecht DW, Paganin DM, Bassed RB, Klein M, et al. Effect of external fixation rod coupling in computed tomography. *Strategies Trauma Limb Reconstr.* 2018;13(3):137–149.
- [52] Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: International Conference on 3D Vision 3DV. IEEE; 2016. p. 565–571.
- [53] Gondara L. Medical Image Denoising Using Convolutional Denoising Autoencoders. In: Proc ICDMW. IEEE; 2016. p. 241–246.

- [54] Jifara W, Jiang F, Rho S, Cheng M, Liu S. Medical image denoising using convolutional neural network: a residual learning approach. *J Supercomput.* 2017;75(2):704–718.
- [55] Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. In: *IEEE Int C Int Robots.* IEEE; 2017. p. 23–30.
- [56] Yoon YJ, Moon SK, Hwang J. 3D printing as an efficient way for comparative study of biomimetic structures — trabecular bone and honeycomb. *J Mech Sci Technol.* 2014;28(11):4635–4640.
- [57] Barak MM, Black MA. A novel use of 3D printing model demonstrates the effects of deteriorated trabecular bone structure on bone stiffness and strength. *J Mech Behav Biomed.* 2018;78:455–464.
- [58] Peña-Solórzano CA, Albrecht DW, Paganin DM, Harris PC, Hall CJ, Bassed RB, et al. Development of a simple numerical model for trabecular bone structures. *Med Phys.* 2019;46(4):1766–1776.
- [59] Bunka M, Müller C, Vermeulen C, Haller S, Türlér A, Schibli R, et al. Imaging quality of ^{44}Sc in comparison with five other PET radionuclides using Derenzo phantoms and preclinical PET. *Appl Radiat Isotopes.* 2016;110:129–133.
- [60] Landry G, Reniers B, Granton PV, van Rooijen B, Beaulieu L, Wildberger JE, et al. Extracting atomic numbers and electron densities from a dual source dual energy CT scanner: Experiments and a simulation model. *Radiother Oncol.* 2011;100(3):375–379.
- [61] Torikoshi M, Tsunoo T, Sasaki M, Endo M, Noda Y, Ohno Y, et al. Electron density measurement with dual-energy x-ray CT using synchrotron radiation. *Phys Med Biol.* 2003;48(5):673–685.
- [62] Yu L, Leng S, McCollough CH. Dual-Energy CT-Based Monochromatic Imaging. *Am J Roentgenol.* 2012;199(5_supplement):S9–S15.
- [63] Landry G, Seco J, Gaudreault M, Verhaegen F. Deriving effective atomic numbers from DECT based on a parameterization of the ratio of high and low linear attenuation coefficients. *Phys Med Biol.* 2013;58(19):6851–6866.
- [64] Wood Z, Lynn L, Nguyen JT, Black MA, Patel M, Barak MM. Are we crying Wolff? 3D printed replicas of trabecular bone structure demonstrate higher stiffness and strength during off-axis loading. *Bone.* 2019;127:635–645.
- [65] Amini M, Reisinger A, Pahr DH. Influence of processing parameters on mechanical properties of a 3D-printed trabecular bone microstructure. *J Biomed Mater Res B.* 2019;108B:38–47.

- [66] Farzadi A, Solati-Hashjin M, Asadi-Eydivand M, Osman NAA. Effect of layer thickness and printing orientation on mechanical properties and dimensional accuracy of 3D printed porous samples for bone tissue engineering. *PloS one*. 2014;9(9):e108252.
- [67] Filippou V, Tsoumpas C. Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. *Med Phys*. 2018;45(9):e740–e760.
- [68] Kamona N, Li Q, Berman B, Sahiner B, Petrick N. Reproducibility of CT-based texture feature quantification of simulated and 3D-printed trabecular bone: influence of noise and reconstruction kernel. In: *Medical Imaging 2019: Computer-Aided Diagnosis*. vol. 10950. International Society for Optics and Photonics; 2019. p. 109501S.
- [69] Han D, Lu Z, Chester SA, Lee H. Micro 3D Printing of a Temperature-Responsive Hydrogel Using Projection Micro-Stereolithography. *Sci Rep*. 2018;8(1).
- [70] Zieger MM, Müller P, Blasco E, Petit C, Hahn V, Michalek L, et al. A Subtractive Photoresist Platform for Micro- and Macroscopic 3D Printed Structures. *Adv Funct Mater*. 2018;28(29):1801405.
- [71] Paulis LE, Kroll J, Heijnsens L, Huijnen M, Gerretsen R, Backes WH, et al. Is CT bulletproof? On the use of CT for characterization of bullets in forensic radiology. *Int J Legal Med*. 2019;133(6):1869–1877.
- [72] Guo Y, Wang K, Yang S, Wang Y, Gao P, Xie G, et al. Structure-Aware Noise Reduction Generative Adversarial Network for Optical Coherence Tomography Image. In: *Ophthalmic Medical Image Analysis*. Springer International Publishing; 2019. p. 9–17.
- [73] Kansagra AP, John-Paul JY, Chatterjee AR, Lenchik L, Chow DS, Prater AB, et al. Big data and the future of radiology informatics. *Acad Radiol*. 2016;23(1):30–42.
- [74] Collado-Mesa F, Alvarez E, Arheart K. The role of artificial intelligence in diagnostic radiology: a survey at a single radiology residency training program. *J Am Coll Radiol*. 2018;15(12):1753–1757.
- [75] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30–36.
- [76] Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Lect Notes Artif Int*. 2016;3(2):119–131. Available from: <https://doi.org/10.1007/s40708-016-0042-6>.
- [77] Liew C. The future of radiology augmented with Artificial Intelligence: A strategy for success. *Eur J Radiol*. 2018;102:152–156.

- [78] Serag A, Ion-Margineanu A, Qureshi H, McMillan R, Martin MJS, Diamond J, et al. Translational AI and Deep Learning in Diagnostic Pathology. *Front Med.* 2019;6.
- [79] Do HM, Spear LG, Nikpanah M, Mirmomen SM, Machado LB, Toscano AP, et al. Augmented Radiologist Workflow Improves Report Value and Saves Time: A Potential Model for Implementation of Artificial Intelligence. *Acad Radiol.* 2020;27(1):96–105.
- [80] Tajmir SH, Alkasab TK. Toward Augmented Radiologists. *Acad Radiol.* 2018;25(6):747–750.
- [81] Parwani AV. Next generation diagnostic pathology: use of digital pathology and artificial intelligence tools to augment a pathological diagnosis. *Diagn Pathol.* 2019;14(1).
- [82] Gauriau R, Bridge C, Chen L, Kitamura F, Tenenholtz NA, Kirsch JE, et al. Using DICOM Metadata for Radiological Image Series Categorization: a Feasibility Study on Large Clinical Brain MRI Datasets. *J Digit Imaging.* 2020;.
- [83] Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing Medical Imaging Data for Machine Learning. *Radiology.* 2020;295(1):4–15.
- [84] Zhang S, Metaxas D. Large-Scale medical image analytics: Recent methodologies, applications and Future directions. *Med Image Anal.* 2016;100(33):98–101.
- [85] Liu X, Tizhoosh HR, Kofman J. Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In: *IEEE IJCNN.* IEEE; 2016. p. 2872–2878.
- [86] Gur Y, Moradi M, Bulu H, Guo Y, Compas C, Syeda-Mahmood T. Towards an efficient way of building annotated medical image collections for big data studies. In: *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis.* Springer; 2017. p. 87–95.
- [87] Khatami A, Babaie M, Khosravi A, Tizhoosh HR, Nahavandi S. Parallel deep solutions for image retrieval from imbalanced medical imaging archives. *Appl Soft Comput.* 2018;63:197–205.
- [88] Criminisi A, Robertson D, Konukoglu E, Shotton J, Pathak S, White S, et al. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med Image Anal.* 2013;17(8):1293–1303.
- [89] Humpire-Mamani GE, Setio AAA, van Ginneken B, Jacobs C. Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans. *Phys Med Biol.* 2018;63(8):085003.

- [90] Xu X, Zhou F, Liu B, Fu D, Bai X. Efficient Multiple Organ Localization in CT Image using 3D Region Proposal Network. *IEEE T Med Imaging*. 2019;38(8):1885–1898.
- [91] Xu X, Zhou F, Liu B, Bai X. Multiple Organ Localization in CT Image Using Triple-Branch Fully Convolutional Networks. *IEEE Access*. 2019;7:98083–98093.
- [92] Pu Y, Gan Z, Henao R, Yuan X, Li C, Stevens A, et al. Variational autoencoder for deep learning of images, labels and captions. In: *Adv Neur In*; 2016. p. 2352–2360.
- [93] Mabu S, Kido S, Hirano Y, Kuremoto T. Opacity Labeling of Diffuse Lung Diseases in CT Images Using Unsupervised and Semi-supervised Learning. In: *Deep Learning in Healthcare*. Springer; 2020. p. 165–179.
- [94] Estrada S, Lu R, Conjeti S, Orozco-Ruiz X, Panos-Willuhn J, Breteler MMB, et al. FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. *Magn Reson Med*. 2019;p. 1–13.
- [95] Filograna L, Magarelli N, Leone A, Guggenberger R, Winklhofer S, Thali MJ, et al. Value of monoenergetic dual-energy CT (DECT) for artefact reduction from metallic orthopedic implants in post-mortem studies. *Skeletal Radiol*. 2015;44(9):1287–1294.
- [96] Ackerman IN, Bohensky MA, Zomer E, Tacey M, Gorelik A, Brand CA, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskel Dis*. 2019;20(1).
- [97] Kamath S, Song W, Chvetsov A, Ozawa S, Lu H, Samant S, et al. An image quality comparison study between XVI and OBI CBCT systems. *J Appl Clin Med Phys*. 2011;12(2):376–390.
- [98] Leschied JR, Mazza MB, Davenport MS, Chong ST, Smith EA, Hoff CN, et al. Inter-radiologist agreement for CT scoring of pediatric splenic injuries and effect on an established clinical practice guideline. *Pediatr Radiol*. 2015;46(2):229–236.
- [99] Nishikawa RM, Comstock CE, Linver MN, Newstead GM, Sandhir V, Schmidt RA. Agreement Between Radiologists’ Interpretations of Screening Mammograms. In: *Breast Imaging*. Springer International Publishing; 2016. p. 3–10.
- [100] Szczykutowicz TP, DuPlissis A, Pickhardt PJ. Variation in CT Number and Image Noise Uniformity According to Patient Positioning in MDCT. *Am J Roentgenol*. 2017;208(5):1064–1072.
- [101] Lim P, Barber J, Sykes J. Evaluation of dual energy CT and iterative metal artefact reduction (iMAR) for artefact reduction in radiation therapy. *Australas Phys Eng S*. 2019;42(4):1025–1032.

- [102] Kovacs DG, Rechner LA, Appelt AL, Berthelsen AK, Costa JC, Friberg J, et al. Metal artefact reduction for accurate tumour delineation in radiotherapy. *Radiother Oncol.* 2018;126(3):479–486.
- [103] Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol.* 2018;37(1):73–80.
- [104] Gomi T, Sakai R, Hara H, Watanabe Y, Mizukami S. Development of a denoising convolutional neural network-based algorithm for metal artifact reduction in digital tomosynthesis for arthroplasty: A phantom study. *PloS One.* 2019;14(9):e0222406.
- [105] Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. *J Am Coll Radiol.* 2018;15(3):504–508.
- [106] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. In: *Medical Image Computing and Computer Assisted Intervention - MICCAI*. Springer International Publishing; 2017. p. 417–425.
- [107] Shin HC, Tenenholtz NA, Rogers JK, Schwarz CG, Senjem ML, Gunter JL, et al. Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. In: *Simulation and Synthesis in Medical Imaging*. Springer International Publishing; 2018. p. 1–11.
- [108] Mikolajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: *International Interdisciplinary PhD Workshop (IIPhDW)*. IEEE; 2018. p. 117–122.
- [109] Grassi M, Loewenstein DA, Caldirola D, Schruers K, Duara R, Perna G. A clinically-translatable machine learning algorithm for the prediction of Alzheimer’s disease conversion: further evidence of its accuracy via a transfer learning approach. *Int Psychogeriatr.* 2018;31(07):937–945.
- [110] Hu LS, Yoon H, Eschbacher JM, Baxter LC, Dueck AC, Nespodzany A, et al. Accurate Patient-Specific Machine Learning Models of Glioblastoma Invasion Using Transfer Learning. *Am J Neuroradiol.* 2019;40(3):418–425.
- [111] Mahmood F, Chen R, Sudarsky S, Yu D, Durr NJ. Deep learning with cinematic rendering: fine-tuning deep neural networks using photorealistic medical images. *Phys Med Biol.* 2018;63(18):185012.
- [112] Mahmood F, Chen R, Durr NJ. Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training. *IEEE T Med Imaging.* 2018;37(12):2572–2581.

- [113] Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*. 2018;81(1):116–128.
- [114] Cha KH, Petrick N, Pezeshk A, Graff CG, Sharma D, Badal A, et al. Reducing overfitting of a deep learning breast mass detection algorithm in mammography using synthetic images. In: *Medical Imaging 2019: Computer-Aided Diagnosis*. SPIE; 2019. p. 1095004–1095004–7.
- [115] Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE T Med Imaging*. 2018;37(6):1348–1357.