REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Sceintifique

Université Ferhat Abbas Sétif

Faculté des Sciences

Département d'Informatique

Fusion de données multimodales pour l'apprentissage automatique : application au diagnostic médical

Thèse présentée par :

Wafa Boukellouz

En vue de l'obtention du diplôme de

Doctorat $3^{éme}$ cycle LMD en Informatique

Soutenue le : 19/01/2020, devant le jury composé de :

Melkmi	Kamel Eddine	Prof. Université Mostefa Ben Boulaïd Batna 2	Président
Moussaoui	Abdelouahab	Prof. Université Ferhat Abbas Sétif 1	Directeur
Taleb-Ahmed	Abdelmalik	Prof. Université Polytechnique Hauts de France	Co-Directeur
Saidi	Mohamed	MCA. Université Ferhat Abbas Sétif 1	Examinateur
Meziane	Abdelkrim	Maitre de Recherche -A-. CERIST Alger	Invité

2019-2020

PEOPLE'S DEMOCRATIC REPUBLIC of ALGERIA

Ministry of Higher Education and Scientific Research

Ferhat Abbas University of Sétif

Faculty of Sciences

Department of Computer Science

Multi-modal Data Fusion for Machine Learning : Application to Medical Diagnosis

A dissertation by :

Wafa Boukellouz

Submitted in the fulfillment of the requirement for the degree of

Ph. D in computer science

Board of examiners :

Melkmi	Kamel Eddine	Prof. University of Mostefa Ben Boulaïd Batna 2	President
Moussaoui	Abdelouahab	Prof. University of Ferhat Abbas Sétif 1	Advisor
Taleb-Ahmed	Abdelmalik	Prof. Polytechnical University Hauts of France	Co-Advisor
Saidi	Mohamed	Assoc Prof. University of Ferhat Abbas Sétif 1	Examiner
Meziane	Abdelkrim	Assoc Prof. CERIST Algiers	Invited

2019-2020

Abstract

Nowadays, information about a phenomenon can be acquired using various modalities; and this led to the emergence of the concept of multi-modal data fusion. In fact, one of the application domains that strongly relies on multi-modal data acquisition is radiation therapy (RT), in which computed tomography (CT) imaging and magnetic resonance imaging (MRI) are the main modalities employed. The latter offers superior soft-tissue visualisation, hence it is used for tumour contouring, whereas the former contains the necessary electron density information for radiation dose computation. However, the pipeline of RT treatment planning presents many flaws from which is the excessive radiation exposure due to repetitive CT acquisitions. Recently, an idea to remediate for this main limitation and others has emerged; it consists in synthesising a CT image called pseudo-CT (pCT) from MRI images.

In this dissertation, we aim to meet this objective by analysing existing work and proposing two variants of pCT image estimation methods employing multi-modal data fusion and machine learning. The first approach is designated hybrid; and combines image registration, unsupervised learning, image fusion and a new Hounsfield Unit (HU) values correction technique. This approach aims to predict a pCT image from T2-weighted MR images while attempting to reduce prediction error by using information from MR images. The second approach employs multi-modal shape, texture and spatial feature extraction, reduction and fusion of T1-weighted and T2-weighted MR images. In addition, we designed an ensemble learning model with stacked generalisation that builds a mapping from these features to HU values over two levels of learning. Results showed a significant improvement achieved by the proposed methods.

Keywords: multi-modality; data fusion; machine learning; medical images; computed tomography; magnetic resonance imaging.

Declaration

I, Wafa BOUKELLOUZ, declare that this dissertation, "Multi-modal data fusion for machine learning: application to medical diagnosis", and the work presented in it are my own and has been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work;
- I have acknowledged all main sources of help;
- Where the dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: .....

Date: .....

Acknowledgment

First and foremost, I would like to express my sincerest gratitude to my advisor, Pr. Moussaoui Abdelouahab. Pr. Moussaoui has supported me with his knowledge and guidance while giving me the freedom to work on my own ideas. I also would like to acknowledge my co-advisor Pr. Taleb-Ahmed Abdelmalik from the Polytechnical University Hauts of France, and Dr. Boydev Christine from Institut Curie - Hôpital René Huguenin Saint-Cloud of France for their valuable support and help in the course of preparation of this dissertation.

I wish to extend my gratitude to the members of my dissertation committee who were more than generous with their expertise, guidance and precious time throughout the review of this document.

The greatest debt one owes is to one's parents and family. I am grateful for their patience, love and encouragement. My parents, as always, encouraged me by asking "Is it done yet?". I also must thank my sisters and brother for providing me a suitable environment to pursue my Ph. D.

I would like to thank my dearest friends Kanoune Yahia from the national higher school of computer science (ESI) and Medani Khadidja from the University of Ferhat Abbas Sétif (UFAS) for believing in my abilities and encouraging me to go through hard times, I am grateful for their support, wholehearted patience and encouragement. I extend my sincere thanks and gratitude to my beloved friend Mehibel Nissa from the University of M'hamed Bougara Boumerdès (UMBB) for her limitless generosity, support, kindness and hospitality.

In addition, many thanks go for my colleagues and friends from both Universities of Ferhat Abbas Sétif and M'hamed Bougara Boumerdès, whom I consider myself very fortunate to have.

Finally, I would like to thank all administration members of the faculty of sciences and department of computer science for their kindness, transparency and valuable help regarding administrative tasks.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	two-Dimensional
3D	three-Dimensional
AC	Auto-Context
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANN	Artificial Neural Networks
ARF	Alternative Regression Forest
bTFE	balanced Turbo Field Echo
CIRF	Conditional Inference Random Forest
CT	Computed Tomography
CV	Cross Validation
dB	deciBel
DCAN	Deep Convolutional Auto-encoder Networks
DCNN	Deep Convolutional Neural Networks
D-S	Dempster-Shafer
DSC	Dice Similarity Coefficient
DV	Dose Volume
DVH	Dose Volume Histogram
DWT	Discrete Wavelet Transform
ECG	ElectroCardioGram
ED	Electron Density
EM	Expectation Maximisation
ES	Ensemble learning with Stacked generalisation
FCM	Fuzzy C-Means
FFE	Fast Field Echo
GAN	Generative Adversarial Networks
GD	Gradient Descent
GE	Gradient Echo
GMR	Gaussian Mixture Regression

Gy	Gray
hMRF	hidden Markov Random Field
HOG	Histogram of Oriented Gradients
HU	Hounsfield Unit
JDL	Joint Directors of Laboratories
kNN	k-Nearest Neighbours
LASSO	logistic sparse Least-Absolute-Shrinkage-and-Selection Operator
LBPh	histogram of Local Binary Patterns
LDA	Linear Discriminant Analysis
Linac	Linear accelerator
LNCC	Local Normalized Correlation Coefficient
LOOCV	Leave-One-Out Cross Validation
LR	Linear Regression
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
ME	Mean Error
MI	Mutual Information
ML	Machine Learning
MLR	Multiple Linear Regression
MRI	Magnetic Resonance Imaging
MSD	Mean Squared Difference
MSE	Mean Squared Error
NCC	Normalised Correlation Coefficient
NMI	Normalised Mutual Information
NMR	Nuclear Magnetic Resonance
NSSD	Local Normalized Sum of Square Differences
PCA	Principal Component Analysis
pCT	pseudo-Computed Tomography
PD	Proton Density
PDF	Conditional Probability Function

PET	Positron Emission Tomography
PETRA	Point-wise Encoding Time Reduction with Radial Acquisition
PSNR	Peak Signal-to-Noise Ratio
ReLU	Rectified Linear Unit
RF	Random Forests
RIRE	Retrospective Image Registration and Evaluation project
RFq	Radio Frequency
RMSD	Root Mean Squared Difference
ROI	Region Of Interest
RT	Radiation Therapy
SD	Standard Deviation
SE	Spin Echo
SSIM	Structural SIMilarity
SVM	Support Vector Machines
T	Tesla
T1-w	T1-weighted
T2-w	T2-weighted
tanh	Hyperbolic Tangent
TDM	Tomodensitométrie
TE	Echo Time
TFE	Turbo Field Echo
TOF	Time-Of-Flight
TPS	Treatment Planning System
TR	Repetition Time
TSE	Turbo Spin Echo
UTE	Ultra-short Echo Time
ZTE	Zero-Echo Time

Introduction

The instinctive nature of humans allows them to understand an event using more than one sense at the same time such as sigh, hearing and touch. Humans were able to understand the functioning of this multi-sensing mechanism and attempted to establish the same reasoning for their fabricated environments; this led to the emergence of multi-sensor data generation, where large amounts of different information types can be acquired about the same event using one or several sensors, also called modalities. However, using each modality separately may not help in taking a complete and concise decision. Therefore, a data fusion process needs to take place in order to ensure that each modality provides the necessary information for decision-making. Depending on the application requirements, multi-modal data can be fused at different levels of a system's construction pipeline going from raw-data level to feature level to decision level; each level of fusion deals with different aspects of data and attempts to fulfil a certain objective. One of the main fields that has largely benefited from multi-modal data fusion is machine learning (ML). The latter is a widespread discipline that aims to discover and learn patterns from data. The building block of any ML system is data, from which hidden patterns are to be discovered or on which predictions are to be made. Generally, learning requires increasing amounts of data in order to be able to learn all possible hypotheses and generalise well. Thus, multi-modal data can be of great benefit for building a performant ML model.

One of the major application domains of multi-modal data fusion in ML is medicine. In fact, the increasing development of medical diagnosis and treatment techniques is largely owed to ML. The latter explores multi-modal medical data while attempting to find mappings between data such as drug classification or to discover hidden patterns in data such as medical image segmentation for anomaly detection. More precisely, medical imaging is regarded as an important aspect of medical data representation, because it expresses data in a visual manner such that humans can easily perceive it and interpret it. Among the critical medical applications that have witnessed an increasing employment of multi-modal data integration with ML in the last decade is radiation therapy (RT), it is one of several means for cancer treatment and operates non-invasively by sending high doses of ionising radiations to tumours in order

to destroy them while sparing healthy tissues. In order to establish the RT treatment amount also called radiation dose for a certain patient, an entire treatment planning workflow needs to be followed, starting from multi-modal imaging, tumour definition and contouring, contour propagation, dose calculation and dose delivery. More particularly, in the imaging step, the commonly used modalities are magnetic resonance imaging (MRI) and computed tomography (CT). The former is used for tumour definition and contouring and the latter is essential for dose calculation for the reason of having the necessary information called electron density (ED) reflected by intensity values expressed in Hounsfield Units (HU). In the contour propagation step, defined contours are propagated from MR images to CT images using image registration techniques. At the final step, the contoured CT images are sent to a treatment planning system for dose calculation and delivery. This standard pipeline is repeated every treatment session for dose recalculation and optimisation in order to account for any inter-session bodily or structural changes in tumours, and this is referred to as adaptive RT. This workflow of RT preparation has many pitfalls, from which are the following:

- Multi-modal imaging requires that the patient to be in the exact same position for each modality imaging, which is difficult to guarantee when dealing with body parts presenting variable structures such as the abdomen. In addition, the need for several image acquisitions can be time-consuming and increases workload;
- Registration of multi-modal images is not a simple task to perform because it employs images of different physical proprieties and may introduce registration uncertainties [Ulin et al., 2010];
- Contour propagation has shown to introduce uncertainties reaching 2-5 mm in body regions such as the prostate and rectum [Nyholm et al., 2009, Ulin et al., 2010];
- Due to the adaptive nature of RT, CT scans are acquired prior to each treatment session for treatment optimisation. In this case, the patient is exposed to the risk of excess radiation exposure, thus creating a potential environment for cancer development and causes increase in mortality rates [de Gonzalez and Darby, 2004].

What is more, MRI is gaining a considerable interest in the field of RT thanks to the great advantages offered by this imaging modality depicted by the superior soft tissue visualisation that MRI provides and the non-ionising nature of magnetic resonance, making MRI a great substitute for CT imaging. These facts are encouraging research to switch to an MRI-only radiation treatment planning workflow. Despite the advantageous characteristics of MRI over CT, MRI suffers from the absence of the ED information required for dose calculation. In

fact, due to the different physical proprieties of each imaging modality, MR intensities do not correlate with CT electron densities. As a proof of evidence, [Ito et al., 2017] investigated the feasibility of a direct conversion from magnetic-susceptibility to relative ED. However, the correlation coefficient between aligned magnetic susceptibility maps and ED images was 0.0145 indicating a very weak correlation. Many studies discussed the potential of MRI in replacing CT in RT [Boettger et al., 2008], this can be achieved by estimating a CT image from MRI data, also called pseudo-CT (pCT). In other words, synthesising a pCT image from MRI data can be formulated as an image synthesis problem, which can be solved in many ways.

In this dissertation, the objective is a two-fold; firstly, we perform a detailed review of existing work in the field of pCT image synthesis from MRI data for RT while identifying possible limitations and strengths of each method. In the last few decades, the field of MRI-only RT has known a remarkable advancement in pCT image estimation reflected by the appearance of a wide range of methods. Hence, we aim to analyse these works and potentially classify them in order to generalise the understanding of their mechanism and identify their drawbacks and advantages. Secondly, based on the conclusions drawn from the literature review, we propose two approaches for pCT image synthesis in the context of MRI-only RT. The proposed approaches attempt to find a solution for pCT image synthesis while employing ML and multi-modal data fusion. The proposed approaches are to be compared with some state-of-the-art methods in pCT image estimation from MRI data.

Dissertation outline

The dissertation is organised in two parts, the first part presents basic concepts of machine learning and data fusion, it also prepares the ground for the literature review by presentation of the basics in medical imaging and radiation therapy. Part two presents the main contributions for pCT image estimation from MRI images. Each part contains the following chapters:

Part I: Backgrounds and literature review

Chapter 1: presents the basic concepts of machine learning and describes steps to build any machine learning model starting from data preparation to model training and testing, we also present some of the well-known machine learning algorithms; some of which were used in the implementation of the contributions of this dissertation. The chapter also presents the basics of data fusion;

Chapter 2: gives the necessary background in medical imaging by presenting the elements of medical images; followed with presentation of the two main imaging modalities used in this dissertation, that are MRI and CT. We also present the concept of radiation therapy in brief, discuss the limitations of this latter and present how research is being oriented towards pCT image synthesis from MR images.

Chapter 3: reviews the existing works for pCT image estimation from MRI data. We perform a classification of these approaches into four classes, analyse their limitations and present their advantages. In addition, we discuss the current orientations in this field;

Part II: Proposed approaches for pCT image synthesis from MRI data

Chapter 4: presents two contributions, the first is an evaluation of one parameter of image registration, where we studied the effect of setting different values for this parameter in order to judge on the best settings for image registration. Second, we present the first approach for pCT image synthesis using MR images. The proposed approach is referred to as hybrid, because it groups many approaches and attempts to reduce pCT image prediction error by using image registration, unsupervised learning, image fusion and a new hybrid correction technique that corrects CT HU intensity values using information from MR images;

Chapter 5: presents the second approach for pCT image estimation using multi-modal MR images, feature fusion and ensemble learning. Two image modalities were used in order to provide additional information about tissues; shape, texture and spatial features were extracted in order to summarise all information from the multi-modal MR images. Features were fused and reduced and then were employed to build an ensemble model for pCT image estimation. The proposed ensemble model is built upon two learning levels, the first level trains three heterogeneous learning algorithms and their prediction outputs are used as features for a second level that combines them in a learning fashion. Combining results by learning is called stacked generalisation. The proposed approach was compared to the state-of-the-art ensemble method random forests.

Part I

Backgrounds and literature review

CHAPTER 1

Machine learning and data fusion concepts

1.1 Introduction

THanks to the ongoing technological advances in data acquisition, processing and storage large amounts of data are being generated at a daily basis. However, this data becomes less useful if it is not properly exploited. Indeed, data can be of a great use when it is well explained. That is, underlying knowledge can be indirectly retrieved from data or data can be used for tasks requiring past experience. A strong need for a mechanism that exploits data for knowledge discovery and event forecasting has given birth to machine learning. ML covers techniques that use existing data to build systems that discover new patterns or perform predictions. A particularity of these ML-based systems is that they have the ability to improve over time by what is so-called "learning". In this chapter, ML learning is presented with steps to build a ML model. We focus on a well-known ML technique: ensemble learning, which constitutes the cornerstone for many successful ML models. Finally, we present data fusion with different mathematical frameworks to represent data aspects and apply fusion.

1.2 Machine learning

ML can be defined as the process of following a certain learning approach and using a specific set of data to improve and minimise error for a well-defined task. Tom Mitchell gave a broad definition for ML [Mitchell, 1997]:

> *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."* [Mitchell, 1997]

From this, we can infer three principal components that define ML, namely; task, performance measure and experience. Therefore, defining a ML problem is related to defining these three components. Take as an example:

- Task: temperature prediction;
- Performance measure: how close is the predicted temperature to the real temperature;
- Experience: past temperature data.

One may ask about how learning is performed, the answer is that ML builds upon three main pillars: data processing, theory of statistics and optimisation. Initial data is processed to remove unnecessary elements and transformed into a certain form to be best fitted for learning. Then this processed data is combined with statistics, at this stage learning becomes active. It is performed with a certain objective to attain (minimise error or maximise some gain criterion) under guidance of an optimisation process. The next section describes steps for building an ML model starting from data preparation until output interpretation.

1.3 Building a machine learning model

Machine learning problems can be roughly divided into two types: classification and regression [Alpaydin, 2009]. Classification attempts to identify group memberships for input data. The output is usually categorical and referred to as class or label. A classification problem with two classes is the basic form of classification and designated as binary classification. For instance, a classification problem can be in the form of trying to recognise the facial expression of an image of a human face, in this case classes can be: happy, sad, angry and so on. Regression tries to find a mapping from input data to continuous valued output that can be integers or floating values. As an example, one may wish to predict house prices in the next few years, therefore the input data can be information about houses such as location, surface, number of rooms and the output is the house price. Overall, building an ML model is highly dependent on identifying the problem type. In the next lines, steps to build an ML model are presented and explained.

1.3.1 Data preparation

After definition of the problem to which an ML model must be built, data has to be collected and appropriately prepared. Bellow are the steps to collect and prepare data for learning.

Data collection: The choice of data can strongly affect the performance of the ML model. Thus, there are many types of data; we give a broad categorisation of data types:

- **Categorical data:** data of such type refer to a qualitative representation, it can either be nominal i.e. do not have any defined order such as colours; or ordinal, having a defined order such as week days. Data of this type cannot be correctly interpreted by an ML model unless it is transformed by mapping categorical values to integer values;
- **Numerical data:** represented by numbers with a well-defined order such as age or weight.

Data can come in various forms such as arrays, matrices, images, videos, graphs and so forth [Smola and Vishwanathan, 2008].

Feature extraction: After data collection, comes feature extraction in which most expressive and distinctive information is extracted from data. This information is represented by feature vectors where lines can represent samples and columns represent feature types. The extracted features differ depending on the defined problem and the form of data. We take as an example ML problems that use images such as in medical imaging, computer vision and image compression where texture, shape, edge and spatial features need to be extracted. Features can be extracted at different levels to maximize information gain [Nixon and Aguado, 2012]:

1. **Low level features:** these features are computed at the pixel level such as spatial coordinates or pixel intensity values;
2. **Patch level features:** the image is divided into patches from which features are extracted;
3. **Image features:** features are extracted from the image as a whole.

Feature selection: Feature selection addresses the problem of finding a compact set of features that best describes and represents data. This can be done by reducing the dimensionality of the feature set. A common technique for feature selection is principal component analysis (PCA), it proceeds as follows:

1. **Standardisation:** brings all features to the same scale [0; 1], this is crucial to perform prior to feature selection to guarantee an equal contribution of each feature type. Otherwise, features with higher values would dominate features of lower values. standardisation is performed for each feature value f as follows:

$$f' = \frac{f - \mu_f}{\delta_f}. \tag{1.3.1}$$

Where μ_f and δ_f are mean and standard deviation across all feature values of the same type, repsectively.

2. **Computation of the covariance matrix:** aims to identify correlations between features. Each cell of the matrix denotes the covariance between two features. For instance, if there are n features then the covariance matrix would be of size $n \times n$. The covariance values for each cell can be computed using the following formula for two feature types X and Y:

$$Cov(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y)}{n}. \tag{1.3.2}$$

If the covariance values are positive between two features then this indicates a high correlation, else the variables are not correlated.

3. **Computation of the eigenvectors and eigenvalues:** eigenvectors hold the most of the explained data expressed in high variance, which are regarded as principal components, and eigenvalues give the percentage of variance expressed by each eigenvector. If there are n features then there would be n pairs of eigenvectors and their corresponding eigenvalues. Eigenvalues λ_1, λ_2, ..., λ_n are computed from the covariance matrix by solving the following equation :

$$| \, Cov - \lambda I \, | = 0. \tag{1.3.3}$$

where I is the identity matrix.

Their corresponding eigenvectors e_1, e_2, ..., e_n are computed as follows:

$$(Cov - \lambda_j I)\, e_j = 0. \tag{1.3.4}$$

4. **Selecting principal components:** eigenvectors are ordered in descending order according to their eigenvalues. Later, a choice needs to be made whether to keep all eigenvectors or discard those having less variance. The final feature vector is created by multiplying the original standardised feature set by the selected eigenvectors.

1.3.2 Choice of the learning algorithm

In this step, the learning process takes place. There is a wide range of learning algorithms to choose from that can be used either for classification or regression problems [Mitchell, 1997, Alpaydin, 2009, Smola and Vishwanathan, 2008]. We can find k-nearest neighbours (kNN), k-means, fuzzy c-means (FCM), random forests (RF), artificial neural networks (ANN), decision trees and support vector machines (SVM), to name a few. The choice of the learning algorithm depends on data and the expected performance because each learning algorithm behaves differently with features. Before jumping to training, data must be split into training and validation sets. These sets must not overlap in order to avoid any undesired performance of the ML model. The following steps explain how training is performed using the pre-processed data and the selected learning algorithm.

1.3.3 Training step

The selected ML algorithm attempts to find a mapping between input data (features) and the target by adjusting certain weights. Take as an example the simplest ML algorithm linear regression (LR) formulated as $y{=}ax{+}b$, where y is the target, x is the feature set and a and b are the parameters to be adjusted to fit the data. These parameters are also referred to as weights. The training starts with random weights to perform initial predictions, and then these predictions are compared to the target values using some evaluation measure such as the mean squared error (MSE) for regression. After that, weights are adjusted and used as inputs for the next training iteration. These steps are repeated until model convergence to a satisfactory result or until training is stopped using some stopping criterion such as reaching the limited number of iterations.

1.3.4 Evaluation step

The performance of the ML model is measured using the evaluation set that contains features that the model has not seen in the training step. This allows to judge on the performance of the model on new data to adjust training parameters and eventually improve model accuracy.

1.3.5 Parameters tuning

Also called hyper-parameters tuning. Hyper-parameters can be initial weights values, number of learning iterations, training/evaluation data ratios and the learning algorithm's initial parameters values. Hyper-parameters tuning involves adjusting the training parameters to improve

performance of the model.

1.3.6 Prediction

Also called testing step, a new set of features that hasn't participated in the building of the model is fed into the ML model, predictions are made using the final estimated weights.

Fig. 1.3.1 summarises the process for building an ML model from data collection to prediction.

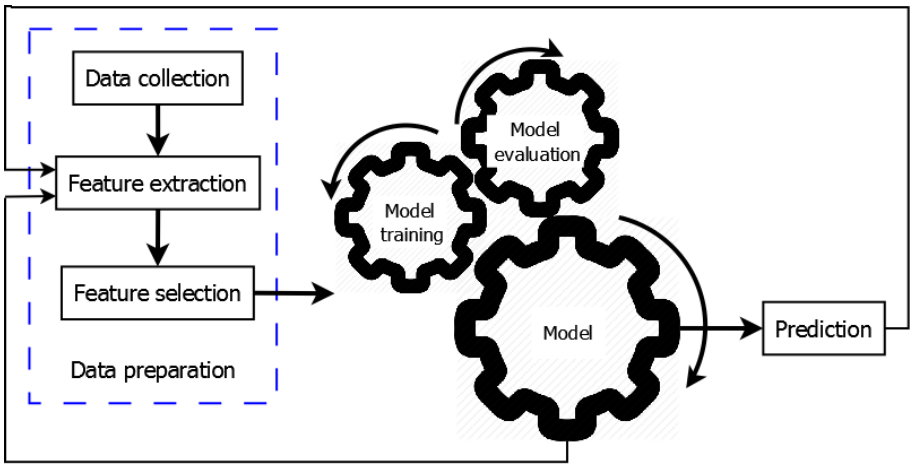

Figure 1.3.1: Pipeline for building an ML model.

1.4 Machine learning algorithms

The process of building an ML model can follow one of two approaches: supervised and unsupervised. In supervised learning, prior knowledge/experience called ground truth is incorporated in the learning process and it refers to what the outputs should be for the input features. The goal is to learn a function that maps input features to the desired output, while minimizing error by finding the best set of learning parameters.

Unlike supervised learning, unsupervised learning does not require any prior knowledge, the goal is to find a specific behaviour in the input data therefore inferring a specific pattern in features. Bellow is a non-exhaustive list of the main supervised and unsupervised ML algorithms that can be used either for classification or regression problems, or for both [Alpaydin, 2009, Smola and Vishwanathan, 2008, Mitchell, 1997].

1.4.1 Multiple linear regression (MLR)

MLR is a simple supervised learning technique that models a data set having two or more independent predictor variables to one dependent response variable. In other words, MLR tries to find the independent variables that explain the dependent variable the most and then assigns them higher coefficients, while the independent variables that explain the dependent variable less are assigned lower coefficients. MLR can be expressed by the following formula:

$$y = a_1x_1 + a_2x_2 + ... + a_nx_n + b. \tag{1.4.1}$$

Where x_1, x_2, $x_3 \ldots$, x_n are predictor variables also called features, a_1, a_2, $a_3 \ldots$, a_n are the regression coefficients to be estimated and b is the slope. y is the target. In the case of having only one predictor variable then the regression problem reduces to a linear regression problem such that:

$$y = ax + b. \tag{1.4.2}$$

1.4.2 K-nearest neighbours (*k*NN)

kNN is an unsupervised learning algorithm. For a new feature vector, the algorithm tries to find its k nearest or most similar feature vectors. Similarity between feature vectors is computed using a distance metric such as the Euclidian distance, correlation, Minkowsky [Batchelor, 2012] and Chi-square [Michalski et al., 1981]. For two feature vectors $f1 = (x_1, x_2, ..., x_n)$ and $f2 = (y_1, y_2, ..., y_n)$ having n features, respective formulas of the afore-mentioned distance metrics are the following:

$$Euclidian\_dist(f1, f2) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}, \tag{1.4.3}$$

$$Minkowski\_dist(f1, f2) = \left(\sum_{i=1}^{n} \mid x_i - y_i \mid^r \right)^{1/r}, r > 0 \tag{1.4.4}$$

$$Correlation\_dist(f1, f2) = \frac{\sum_{i=1}^{n} (x_i - \mu_i)(y_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} (x_i - \mu_i)^2 \sum_{i=1}^{n} (y_i - \mu_i)^2}}, \tag{1.4.5}$$

$$Chi - square\_dist(f1, f2) = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{(x_i + y_i)}. \tag{1.4.6}$$

1.4.3 Fuzzy c-means (FCM)

FCM [Bezdek et al., 1984] is an unsupervised learning technique that allows data to belong to one or more clusters assuming a fuzzy membership to each cluster. It's based on minimizing the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\|x_i - c_j\right\|^2, 1 \leq m < \infty. \tag{1.4.7}$$

Where, m is any real number greater than 1, u_{ij}^m is the fuzzy membership of a data value x_i to the cluster j, c_j is the centre of cluster j, C is the number of clusters and N is the number of data points. The fuzzy membership is updated as follows:

$$u_{ij}^m = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{1-m}}},\ c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}. \tag{1.4.8}$$

The termination criterion is set as $max_{ij}\left\{\left|u_{ij}^{k+1} - u_{ij}^{k}\right|\right\} < \varepsilon$.

1.4.4 K-means

k-means is a type of unsupervised learning, it divides data to a set S of k clusters, each grouping similar data points/features. In other words, k-means finds k cluster centres and assigns each data point to the cluster whose centre is most similar to this data point. The algorithm starts by defining random cluster centres c_j from a set of centres C, then the Euclidian distance is computed between each data point x and centre c_j, and each data point is assigned to a cluster based on:

$$\underset{c_j \in C}{argmin}\, dist(c_j, x)^2, \tag{1.4.9}$$

Then, the cluster centres are updated by computing the mean of data points in each cluster S_j as follows:

$$c_j = \frac{1}{\mid S_j \mid}\sum_{x_i \in S_j} x_i. \tag{1.4.10}$$

This process is repeated until clusters stop changing or a certain number of iterations is reached.

1.4.5 Naive Bayes

It is a probabilistic ML algorithm based on Bayes theorem, which tries to find the probability of a hypothesis given an evidence. In ML, Given X a feature vector denoting the hypothesis and y is the response or the evidence. Bayes theorem takes the following form:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}, \tag{1.4.11}$$

Knowing that X is a feature vector with several features $(x_1, x_2, ..., x_n)$, Bayes theorem can be rewritten as:

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}, \tag{1.4.12}$$

Since all feature values belong to one feature vector (sample), then the denominator does not change for all feature values, hence naïve Bayes can be expressed as:

$$P(y|x_1, ..., x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y). \tag{1.4.13}$$

1.4.6 Linear discriminant analysis (LDA)

LDA is a supervised ML technique that can be used for classification problems. The training set is composed of input features and their corresponding labels. In the training phase, for each feature vector x belonging to a class j, $j=1,..., k$, the mean of the instances belonging to a class j is computed:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i. \tag{1.4.14}$$

where n_j is the number of instances in a class j.

In addition, the variance is computed across all classes:

$$\sigma^2 = \frac{1}{n-k} \sum_{i=1}^{n_j} (x_i - \mu_j)^2. \tag{1.4.15}$$

To predict the label for a new feature vector, naïve Bayes theorem is employed:

$$P_j(X) = \frac{\pi_j f_j(x)}{\sum_{i=1}^{k} \pi_i f_i(x)}. \tag{1.4.16}$$

Where π_j is the base probability of each class observed in the training data (π_j equals the number of instances in a class j divided by the total number of training instances) and

f_j is the estimated probability of x belonging to a class j. The goal is to find the class with the maximum probability to which the feature vector x belongs. When plugging a Gaussian distribution function used for $f_j(x)$ to the above equation, the objective is to maximise:

$$\underbrace{\delta_j(x)}_{discriminant} = x\frac{\mu_j}{\sigma^2} - \frac{\mu_j^2}{2\sigma^2} + log(\pi_j). \tag{1.4.17}$$

$\delta_j(x)$ is called the discriminant function for the class j given an input feature vector x.

1.4.7 Decision trees

A decision tree is a tree-like structure based on the concept of divide and conquer. Decision trees can be used for classification and regression problems. The idea is to divide features based on if-then-else decision rules. That is, the root node is represented by an input feature. As moving down in the tree, each internal node is a test node that is split into two branches leading to other deeper nodes. The decision of an internal test node guides the learning process to take one of the two next nodes. This process of test and split is repeated until a leaf node indicating the final decision is reached. Fig. 1.4.1 illustrates an example of a decision tree for predicting the temperature given whether the season is winter or not.

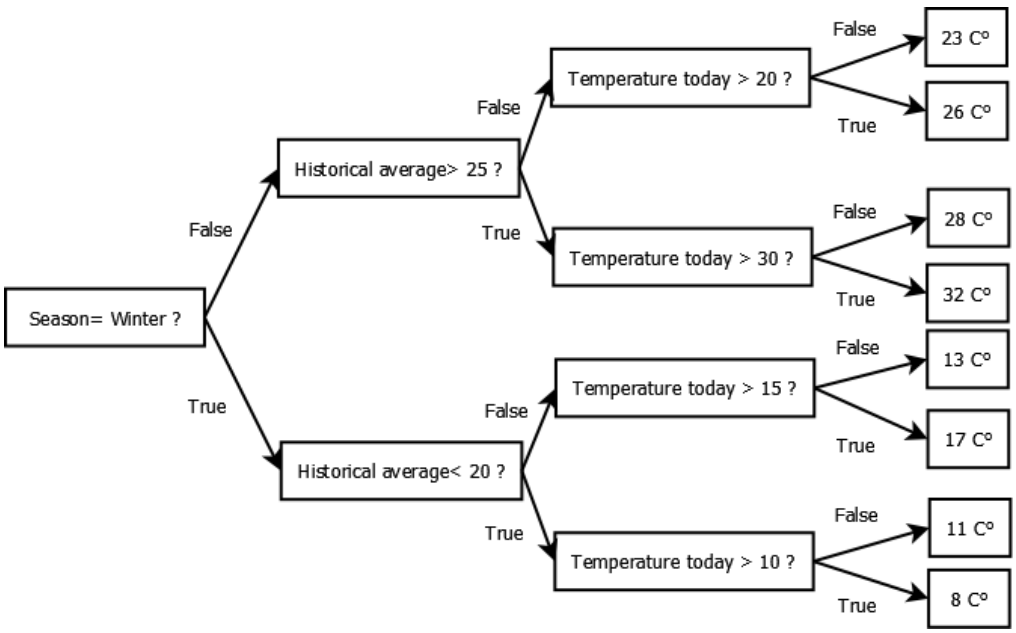

Figure 1.4.1: An example of a decision tree for temperature prediction.

1.4.8 Support vector machines (SVM)

SVM [Vapnik et al., 1996] is a supervised technique that attempts to find a hyper-plane that best separates data while maximising the margin between data and the hyper-plane. SVM are commonly employed for classification problems. However, they can be employed to solve regression problems. A hyper-plane can be formulated as:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n = 0. \tag{1.4.18}$$

For a two-dimensional (2D) space, the hyper-plane takes the following form:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0. \tag{1.4.19}$$

In the training phase, the data is separated according to the label values y, thus, data having the same label fall in one side of the hyper-plane and the other data fall on the other side of the hyper-plane such that:

$$\begin{cases} \beta_0 + \beta_1 x_1 + \beta_2 x_2 > 0 \quad if\, y = 1, \\ \beta_0 + \beta_1 x_1 + \beta_2 x_2 < 0 \quad if\, y = -1. \end{cases} \tag{1.4.20}$$

The hyper-plane margin is computed by finding the smallest perpendicular distance from the data points to the hyper-plane. Fig. 1.4.2 illustrates the main elements of SVM.

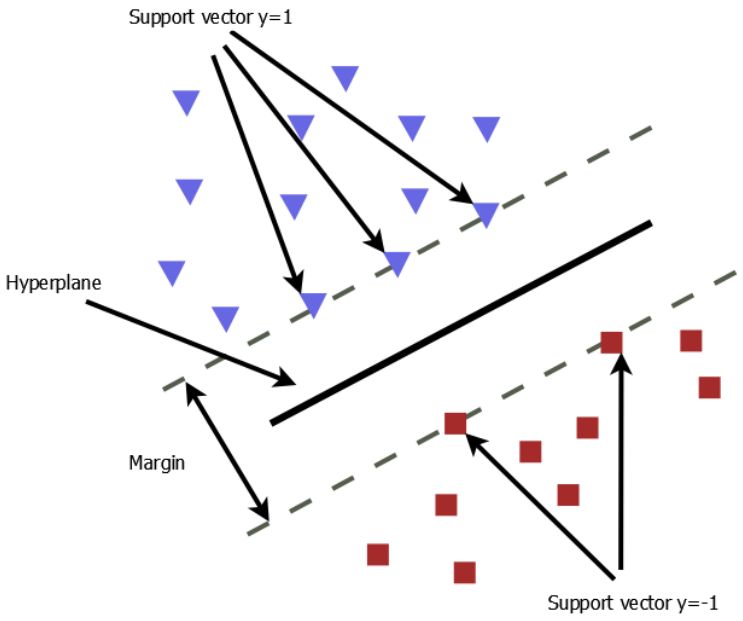

Figure 1.4.2: Main elements of SVM for a binary linear classification problem.Alpaydin [2009]

In few cases, the data is not linearly separable, therefore, SVM introduce kernels to define a non-linear decision boundary.

1.4.9 Artificial neural networks (ANN)

Inspired by the behaviour of human brain neurons, ANN can be used to solve regression and classification problems. Feed forward networks is one common architecture of ANN which can have an input layer, hidden layers and an output layer. In the learning phase, features are passed through the input layer which can have as much neurons as the feature vector size. When moving to the hidden layers, internal node values also called activation values are computed based on values of nodes of the previous layers such that:

For a neurone j in a layer l, its activation value is related to the values of the neurons k in layer $l-1$ as follows:

$$a_j^l = \sigma\left(\sum_k \omega_{jk}^l a_k^{l-1} + b_j^l\right). \tag{1.4.21}$$

Where w_{jk}^l are the weights connecting neurons k in $l-1$ to neurone j in l, a_k^{l-1} is the activation value of neurons k in $l-1$, b_j^l is the bias vector of neurone j at layer l and σ is the activation function. An activation function is used to bind the values of a neuron in order to restrict it to a defined interval depending to the problem type. For a binary classification problem, it would be suitable to use an activation function that transforms neuron values to either 1 or 0 indicating binary class labels. For such a case, the *sigmoid* function can be used:

$$Sigmoid(x) = \frac{1}{1+e^x}. \tag{1.4.22}$$

Other activation functions include The Hyperbolic Tangent *(tanh)* and Rectified linear unit ($ReLU$) functions. For a regression problem, a linear activation function can be employed in order to consider the weighted sum of the neuron. The computed output is then compared to the target values using a loss function. For instance, in regression problems, MSE is computed between the predicted values and the target. The objective is to update weights in a way that the loss is minimised; this is done using the backpropagation algorithm [Rumelhart et al., 1985]. The computed loss is propagated backwards to the previous layers until reaching the input nodes in order to adjust weights. The learning process is repeated until convergence of the model. Fig. 1.4.3 presents the basic components of an ANN.

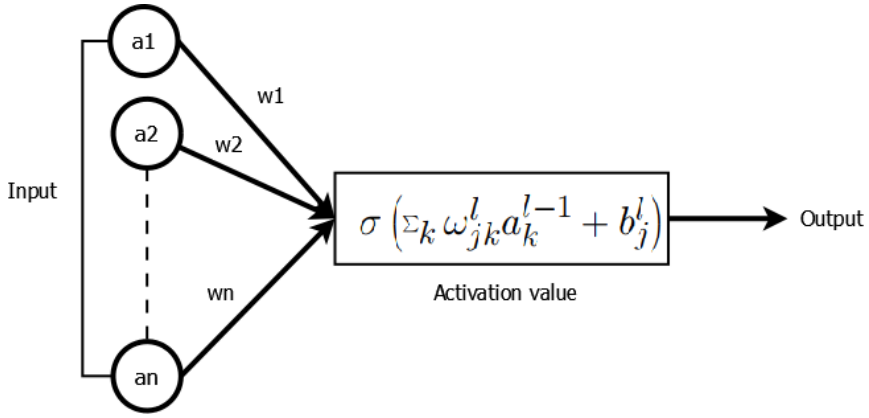

Figure 1.4.3: Basic element of an ANN.

The usefulness of a certain ML algorithm is highly related to the data involved and the application domain. In fact, one ML algorithm can perform very well on a certain data set but performs poorly on a different data set. Therefore, one cannot make prior assumptions on which ML algorithm is the best for a particular problem [Wolpert, 1996]. In addition, the size of the training data set plays an important role in determining the performance of the ML model. From the one hand, a small data set size can result in over-fitting of the model, meaning that the model has fitted the data very well that it memorised all scenarios present in the training set. Thus, the model will perform poorly upon testing on unseen data. On the other hand, a model under-fits if it is unable to sufficiently reduce the prediction error. A learning model is claimed to be good if it generalises well, i.e., have a small prediction error [Zhou, 2012], therefore, a model that over/under-fits data is a model that does not generalise well.

Using several homogeneous or heterogeneous learning algorithms called base learners to build a model is referred to as ensemble learning where prediction results from these learners are combined using different combination methods. Ensemble methods is a widely employed ML scheme that has shown to outperform single learning methods. The next section presents the basics of ensemble learning with a description of the state-of-the-art ensemble method algorithms.

1.5 Ensemble methods

Ensemble methods is a machine learning technique that has been widely employed since their appearance in the 1990s by the works of Hansen and Salamon[Hansen and Salamon, 1990] and Schapire [Schapire, 1990]. It was shown that using several weak learners called base learners, and then combining them results in a strong learner that can reduce error of the weak learners

when considered separately. Fig. 1.5.1 shows the basic ensemble learning architecture [Zhou, 2012].

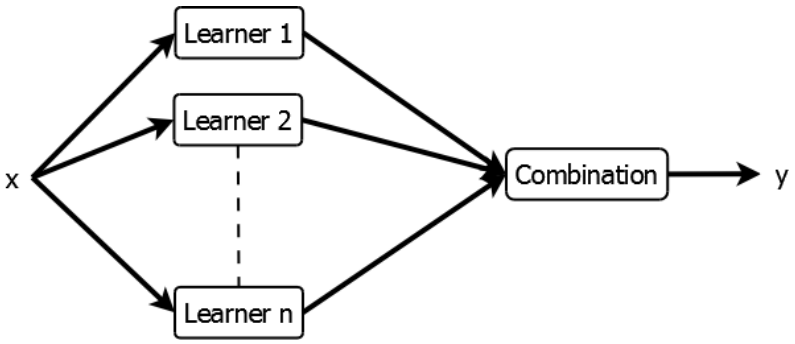

Figure 1.5.1: Basic architecture of an ensemble model [Zhou, 2012].

Ensemble methods were the result of two main motivations: firstly, it was found that combining several base learners can result in a stronger generalisation ability compared to that of the base learners [Hansen and Salamon, 1990]; and secondly, it was theoretically demonstrated that base learners can be boosted to produce a strong learner [Schapire, 1990]. This is advantageous over using one strong learner because the latter is difficult to get due to model complexity and computational cost whereas combining base learners into a strong learner can be less expensive and simpler to get. Ensemble methods have shown to outperform many other learning technique such as in object detection [Paisitkriangkrai et al., 2016], medical diagnosis [Wang et al., 2015] and fraud detection [Zareapoor et al., 2015], among others. An ensemble model is constructed by following two steps: base learners training and combination.

1.5.1 Building an ensemble model

Like any other ML model, ensemble models are built following similar steps. The exception is that ensemble models use several learners and employ a combination scheme to produce the final model. Before building an ensemble model, it is recommended to use base learners that provide a maximum of diversity; this is because nothing would be gained if base learners provide the same output [Chandra and Yao, 2006]. Diversity can be achieved using several approaches, from which:

- Training the base learners on non-overlapping subsets of the training set;
- Training heterogeneous base learners on the whole training set;
- Training several homogeneous base learners, each with different training parameters.

After training of the base learners using the appropriate approach and algorithms, comes the step for combination. Combination methods include averaging, voting or learning. The latter is a mean to combine results from different base learners by learning to generalise from their predictions that are taken as features. Combination of base learners is often less computationally expensive than training the base learners because combination methods are simple and straightforward.

1.5.2 Ensemble-based algorithms

There is a wide range of ensemble-based approaches that use different basic ML algorithms. In this subsection, we present three major categories of ensemble-based algorithms.

1.5.2.1 Bootstrap aggregating

Also known as bagging [Breiman, 2001], it is based on the hypothesis of making the base learners as independent as possible. The main foundations of bagging are bootstrap sampling and aggregating. For a training set of size N and T base learners, bootstrap sampling suggests to employ sampling with replacement, that is, to draw a random subset of size n from the training data set and assign it to the base learner number *i, i=1, 2,. . . , T*. Then, this subset of training data is put back to the training data set and another random subset of the same set is drawn and subsequently assigned to the base learner number *i+1*. This process is repeated until all base learners are assigned subsets of training data. After that, the base learners are trained using the attributed training sets. Upon prediction, a new test data is passed through all trained base learners to make initial predictions and the final prediction result is computed by aggregating the initial predictions by voting for classification or by averaging for regression.

Bagging constitutes the foundation for the state-of-the-art random forests algorithm that trains multiple decision trees on bootstrapped subsets of the training data and then combines their results using aggregation. Random forests can be used for both classification and regression problems. They are implemented as follows:

1. The algorithm selects a subset from the training set by sampling with replacement. Each sample is a feature vector of size l;

2. The tree is grown by performing splits using a subset of features of size $m<l$ that improve training;

3. Upon prediction, a test sample is passed through all trained trees and the final prediction results is computed by aggregating predictions of all trees.

1.5.2.2 Boosting

Boosting was the answer of Schapire in [Schapire, 1990] on a question posed by Kearns and Valiant *"Can a set of weak learners create a single strong learner ?"* [Kearns and Valiant, 1994]. Boosting attempts to boost a weak learner to a strong learner, it works by training base learners in a sequential order and each base learner attempts to correct for mistakes that the previous learners failed to correct. In other words, each base learner focuses more on instances that have not been well-learned by the previous base learners. A boosting ensemble model is constructed as follow:

1. Initially, a base learner is trained on a randomly selected subset of the training data;
2. This base learner is evaluated on the entire training set, thus, some of the instances are not well-predicted. Therefore, more attention should be paid to them in order to improve their prediction;
3. A weighting process is applied to the entire training set by assigning higher weights to those instances that were not well-learned and lower weights to instances accurately predicted;
4. Another base learner is trained on a subset of the training set wherein data is weighted such that the learner attributes more attention to the badly learned instances;
5. Predictions made by both base learners are combined; again, there are some instances that were not well-predicted. However, the instances that were previously assigned high weights are well-predicted;
6. Re-weighting is performed to consider those badly-learned instances and another base learner is trained on a subset of the training data;
7. This process of learning, evaluation, re-weighting and combining is repeated until satisfactory results are obtained.

Boosting groups a wide range of algorithms from which AdaBoost which is very often considered a powerful ensemble algorithm for solving classification problems.

1.5.2.3 Stacked generalisation

Stacked generalisation was introduced by Wolpert in 1992 [Wolpert, 1992]. It is mainly a scheme to minimize the generalization error of one or many base learners. Unlike boosting and bagging, base learners are not combined by voting or averaging, they are rather combined by learning.

The basic idea of stacked generalisation is to train first level base learners on the training data set and then take their predictions and feed them to a second level learner that performs combination and predicts the final results. Stacked generalisation can use both homogeneous or heterogeneous learning algorithms. To avoid over-fitting of the stacked generalisation model, it is suggested to use a subset of data for training the second level learner that has not been employed in the training step of the first level base learners. For instance, one can use cross validation for this purpose.

The following steps describe how to construct an ensemble model with stacked generalisation:

1. The training set is divided into N subsets;
2. Each first level base learner is trained on different *N-1* subsets;
3. Each first level base learner is evaluated on the N^{th} subset that wasn't included in the training phase;
4. Predictions made by these first level learners are used as input features for the second level learner which is trained to perform the combination.

Data subject to learning can be acquired from several modalities/ sensors or can come from the same modality but with different acquisition parameters. However, in order to use this multimodal data for learning, this data needs to be appropriately combined to maximise information gain and guarantee that each modality will contribute to learning. The task of combining multi-modal data is referred to as data fusion. In the next section, data fusion is defined with a description of the main concepts related.

1.6 Data Fusion

With the ever-increasing technological development in hardware and software aspects of data-acquisition systems. It has become feasible to acquire data about a phenomenon using several commensurable or non-commensurable sensors (modalities) such as an electrocardiogram (ECG) and heart ultrasound images. Data fusion is defined as the task of combining data from multiple sources to achieve high levels of system accuracy and robustness and decrease uncertainty. The joint directors of laboratories (JDL) define data fusion as [White, 1991]:

> *"Multilevel, multifaceted process handling the automatic detection, association, correlation, estimation, and combination of data and information from several sources."* [White, 1991]

The use of multi-modal data fusion serves for variety since each modality provides information that is not provided by the other modalities [Lahat et al., 2015]. Mathematically speaking, data fusion can be formulated as [Chang and Bai, 2018]:

$$O = F(X_1, X_2, ..., X_n). \tag{1.6.1}$$

where $(X_1, X_2, ..., X_n)$ denote different data sources, O is the fused output and F is the fusion process.

1.6.1 Data fusion levels

Data fusion can be performed on three major levels of abstraction:

1. Raw-data fusion level: data is provided to the fusion system in its raw form such as the grayscale values of an image. The raw-data fusion can be performed on data coming from commensurable modalities, that is the acquisition systems focus on the same physical proprieties of the phenomenon subject to data acquisition [Hall and Llinas, 1997]. Otherwise, if data is acquired from non-commensurate modalities then the feature level fusion is performed;

2. Feature level fusion: The extracted features from each modality can be fused together into one more discriminative feature vector;

3. Decision level fusion: Combines final decisions from several outputs such as ML models.

In addition, it is possible to incorporate two or more of these fusion levels in the ML model to increase performance and each fusion level would address data aspects that were not addressed by the other levels of fusion. The next subsection describes briefly the different data aspects that should be considered by a fusion method.

1.6.2 Data aspects

Data fusion can be a challenging task to perform because of the various aspects present in data due to imperfection and uncertainty of multimodal acquisition systems. Fig. 1.6.1 presents a non-exhaustive list of data aspects that need to be addressed by a fusion technique [Khaleghi et al., 2013].

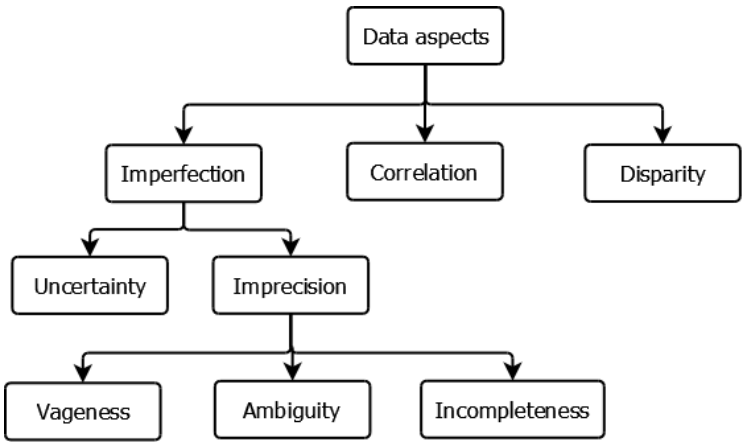

Figure 1.6.1: Different aspects of data [Khaleghi et al., 2013].

1. Imperfection: imperfect data can present at least one of these aspects:

   (a) Uncertainty: data is deemed uncertain when an assigned level of confidence is lower than a pre-defined threshold;

   (b) Imprecision: characterises data that refer to more than one object or phenomenon. Imprecise data can further be ambiguous, vague or incomplete. Data is ambiguous when the attributes of data are well-established and have a level of confidence, yet they are imprecise (for example: an image pixel takes values between 10 and 100, here, the data attribute "pixel" is well-defined, however its value is imprecise). On the other hand, data can have a vague meaning when data attributes are not well-defined and can have various meanings (for example: the moving object is a vehicle, "vehicle" can be variously interpreted). And finally, incomplete data is a data that has missing information.

2. Correlation: data presents dependencies between its attributes, thus some data can be expressed as a function of the other data. This can result in redundancy and dimensionality increase. Therefore correlation should be well-treated prior to fusion;

3. Disparity: data can be acquired from the same or different acquisition systems targeting different physical proprieties of the object under consideration.

Aspects of data can be modelled in various mathematical frameworks. The next subsection describes these frameworks with some fusion operators.

1.6.3 Data fusion techniques

Many fusion techniques exist and can be applied at different levels:

1.6.3.1 Probabilistic fusion

Expresses data uncertainty using probability distribution functions. A common probabilistic fusion approach is the Bayesian fusion that uses Bayes theorem. Let m_1 and m_2 two data instances to be fused. The probability of having a new data estimation m from m_1 and m_2 is:

$$\begin{aligned} P(m|m_1, m_1) &= \frac{P(m,m_1,m_2)}{P(m_1,m_2)} \\ &= \frac{P(m_1|\, m,m_2)P(m,m_2)}{P(m_1,m_2)} \\ &= \frac{P(m_1|\, m,m_2)P(m|\, m_2)P(m_2)}{P(m_1|\, m_2)P(m_2)} \\ &= \frac{P(m_1|\, m,m_2)P(m|\, m_2)}{P(m_1|\, m_2)}. \end{aligned} \tag{1.6.2}$$

1.6.3.2 Evidential belief reasoning

Initiated by Dempster [Dempster, 1968] and later formalised by Shafer [Shafer, 1976], evidential belief reasoning also known as Dempster-Shafer (D-S) theory deals with data uncertainty and imprecision using evidence-based reasoning [Khaleghi et al., 2013]. It is considered as a generalisation to the Bayesian theory. The D-S theory establishes the notion of attributing beliefs and plausibilities to possible measurement hypotheses. In addition, it introduces combination rules for fusion.

Let X represents all possible hypotheses and 2^X the set of all possible subsets of X. D-S theory assigns a belief mass function m to each element E of 2^X in *[0, 1]*. m represents all possible propositions regarding the system state x. proprieties of m are:

$$\begin{cases} m(\phi) = 0 \\ \sum_{E\in 2^X} m(E) = 1 \end{cases} \tag{1.6.3}$$

Hence, *m(E)* is regarded as the proportion of the available evidence that supports the claim that the actual hypothesis belongs to E [Khaleghi et al., 2013]. Thus, a probability for E can be obtained using m by defining belief *bel(E)* and plausibility *pl(E)* bounds such that: $bel(E) \leq P(E) \leq pl(E)$, with $bel(E) = \sum_{B\subseteq E} m(B)$ and $pl(E) = \sum_{B\cap E\neq\phi} m(B)$.

For two different data sources to which are associated two belief mass functions m_1 and m_2, the joint belief mass function m describing the fusion is computed using:

$$\begin{cases} m(E) = (m_1 \oplus m_2)(E) = \frac{1}{1-K} \sum_{B \cap C = E \neq \phi} m_1(B) m_2(C), \\ m(\phi) = 0. \end{cases} \tag{1.6.4}$$

K is the amount of conflict present between the two data sources:

$$K = \sum_{B \cap C = \phi} m_1(B) m_2(C). \tag{1.6.5}$$

1.6.3.3 Fuzzy reasoning

Treats data uncertainty by introducing the notion of partial set membership. Let a fuzzy set $F \subseteq X$ defined by a gradual membership function $\mu_F(x)$ such that:

$$\mu_F(x) \in [0,\, 1] \quad , \forall x \in X \tag{1.6.6}$$

The higher the membership degree, the more x belongs to F. by assigning a membership function to each hypothesis x, these membership functions can be fused using conjunctive combination rules, typically when data sources are considered homogenous and are equally reliable [Khaleghi et al., 2013]. Such combination rules include:

$$\mu_1^{\cap}(x) = min\left[\mu_{F_1}(x),\, \mu_{F_2}(x)\right] \quad , \forall x \in X. \tag{1.6.7}$$

$$\mu_2^{\cap}(x) = \mu_{F_1}(x) \cdot \mu_{F_2}(x) \quad , \forall x \in X. \tag{1.6.8}$$

Representing the intersection and product rules, respectively. On the other hand, disjunctive combination rules can be used when at least a data source is considered reliable but with no provided information about which one is [Khaleghi et al., 2013]. These include union and sum rules given by:

$$\mu_1^{\cap}(x) = max\left[\mu_{F_1}(x),\, \mu_{F_2}(x)\right] \quad , \forall x \in X. \tag{1.6.9}$$

$$\mu_2^{\cap}(x) = \mu_{F_1}(x) + \mu_{F_2}(x) - \mu_{F_1}(x).\mu_{F_2}(x) \quad , \forall x \in X. \tag{1.6.10}$$

respectively.

1.6.3.4 Possibilitic fusion

Founded by Zadeh [Zadeh, 1978], it characterises the uncertain membership of an element x in a known class B with a possibility distribution such that: $\pi_B(x) \in [0, 1]$, $\forall x \in X$. It also imposes a normalisation constraint requiring that at least one value is possible: $\exists\, x^* \in X$, $\pi_B(x^*) = 1$.

For an event U, the possibility measure $T(U)$ defines the extent of plausibility of U:

$$\Pi(U) = max_{x \in U} \{\pi_B(x)\} \quad , \forall U \in X. \tag{1.6.11}$$

and the necessity measure $N(U)$ quantifies the certainty of U:

$$N(U) = min_{x \notin U} \{1 - \pi_B(x)\} \quad , \forall U \in X. \tag{1.6.12}$$

Fusion rules are similar to those employed in fuzzy reasoning considering two possibilistic distributions describing different data sources while keeping in mind the normalisation constraint.

1.6.3.5 Statistical fusion

Data fusion can be regarded as a statistical modelling task. Common statistical fusion approaches include PCA that performs data reduction and data fusion at the same time; it can be better suited for correlated data. In addition, wavelet-based fusion which consists in representing data sources as signals using wavelet functions; and then computes approximation and detail coefficients for each data source that are subsequently fused into one output [Naidu and Raol, 2008]. Average and weighted average are also regarded as simple statistical fusion rules.

1.7 Conclusion

This chapter presented the basic notions of ML with the main algorithms that can be used to build an ML model. In addition, Ensemble learning was presented which is considered an extension of standard ML algorithms and shows to improve an ML model's performance by using the appropriate combination method. Furthermore, data fusion was presented with description of fusion levels and some common fusion techniques. Usually, using data from different sources can help improve prediction accuracy since each data source or modality provides additional or complementary information. Nowadays, machine learning is applied to almost all fields going from computer vision to fraud detection to self-driving cars, to name a few. In particular, medical imaging has an important share of ML applications such as image segmentation and image synthesis. In this dissertation, we are particularly interested in the

application of supervised and unsupervised ML techniques for inter-modality image synthesis in the context of radiation therapy. The next chapter presents the necessary background in radiation therapy and related medical concepts.

CHAPTER 2

Clinical background and medical concepts

2.1 Introduction

The medical field holds an important share of scientific research to which most gratitude is granted for the impressive advances that medicine has achieved. More focus is attributed to medical imaging since the intrinsic understanding of humans mainly relies on sight; therefore acquiring more information about a disease is done throughout acquisition and interpretation of medical images. The more divers these images, the more understanding of a disease is guaranteed. As a result, it has become a clinical practice to use several imaging modalities for disease diagnosis, treatment preparation and monitoring. In this dissertation, a special focus is attributed to the use of multi-modal medical imaging for radiation therapy.

In this chapter, we define medical imaging with focus on MRI and CT imaging. We also present the basic elements of a medical image and explain the process of medical image registration. All these concepts constitute the base for radiation therapy that is further presented in this chapter. In addition, we present limitations of radiation therapy and discuss how research is striving to remove CT imaging for its inherited limitations and replace it with MRI.

2.2 Medical imaging

Standard clinical examination and diagnosis tools are often not enough to make a complete and concise diagnosis for the patient. Therefore, advanced diagnosis tools are provided, among them a vital tool, which is medical imaging, it can be perceived as a mean to non-invasively retrieve information about the internal structures of the body by creating visual representations of anatomical and pathological structures. Medical imaging has undergone many technological advances in the last few decades and this is attributed to the extended understanding of physics

and the ever-increasing computational power on which medical imaging is based. Nowadays, doctors, clinicians and radiologists are provided with cutting-edge imaging technologies to cope with complicated medical cases and improve patient treatment experience. Medical imaging comes with diverse imaging modalities that can be used for different stages of clinical examination, diagnosis, treatment and monitoring. Indeed, these imaging modalities have become a necessity to be included in the clinical workflow. However, one may wonder how these imaging modalities can provide different information about the same body region; in fact, this is owed to the physical proprieties of body tissues and how the imaging modality influences these proprieties to retrieve information and present it in the form of a medical image. Hendee William in his book on medical imaging physics defines medical images as follows:

> *"Medical images are pictures of tissue characteristics that influence the way energy is emitted, transmitted, reflected, and so on, by the human body. These characteristics are related to, but not the same as, the actual structure (anatomy), composition (biology and chemistry), and function (physiology and metabolism) of the body." [Hendee and Ritenour, 2003]*

From this we understand that in order to produce a medical image, a well-defined energy source must be applied on the body region of interest targeting specific physical tissue proprieties which in turn produce valuable information that is processed and presented in the form of an image. Based on this hypothesis, several imaging modalities exist such as CT imaging, MRI, X-ray radiography, ultrasonography, positron emission tomography (PET), mammography and fluoroscopy. Fig. 2.2.1 shows images of the human brain using CT, two variants of MRI and PET demonstrating a remarkable diversity in tissue visualisation that each modality can provide.

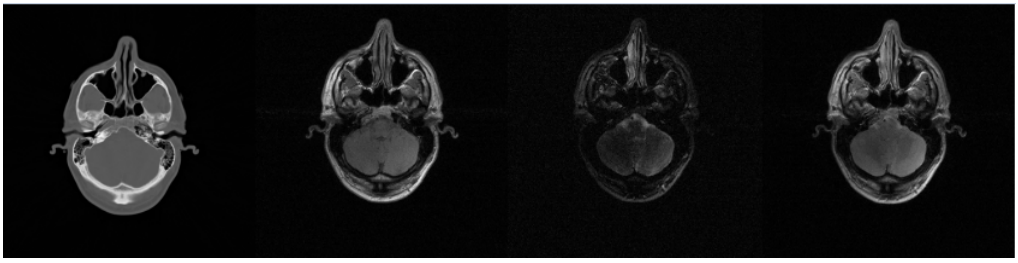

Figure 2.2.1: Images of the human brain using CT, two variants of MRI and PET.

In the context of this dissertation, only MRI and CT imaging modalities are covered in the next subsections, and referred to as tomographic imaging modalities. The essentials of these two imaging modalities are presented and explained to provide the necessary background for the next chapters. The next section presents the fundamental components of a tomographic

medical image. We first define the elements of a medical image in its 2D and three-dimensional (3D) forms.

2.2.1 Medical image components

Unlike general-purpose images, medical images are of a different type and nature since they come from a specific imaging equipment and acquired considering physical proprieties of body tissues. A 2D medical image of size $N \times M$ can be defined in a 2D coordinate space $(\overrightarrow{x}, \overrightarrow{y})$, where each point p is called a pixel and has an intensity taking values in the interval [0; 255] for MR images and [-1000; 2000] for CT images. In addition, each point is characterised by coordinates *(i,j) (i=0,..,N-1 and j=0,. . . , M-1)*. For instance, a pixel *p(23,150)*=200 in an MR image is located in the row number 23 and column 150 and has an intensity value of 200.

A sequence of 2D images placed in a defined order forms a 3D image volume. In addition to the row and column coordinates *(i,* j), a point v in a 3D volume is called voxel and has a slice number k. Thus, each voxel has a row number i, a column number j and a slice number k such that *v (i, j, k)*. Furthermore, each pixel/voxel has a defined size expressed in mm^2/mm^3 which is defined depending on the imaging parameters and scanner settings. Fig. 2.2.2 shows medical image elements of an MR image in 2D and 3D forms.

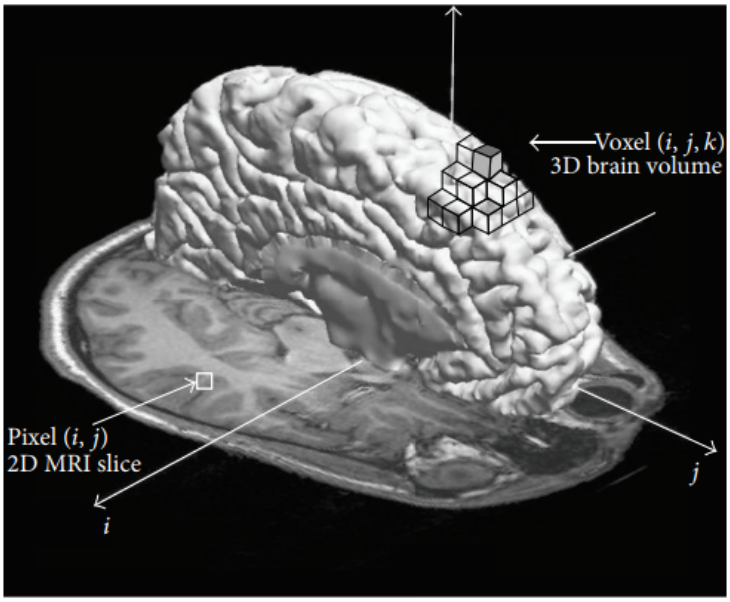

Figure 2.2.2: Medical image elements of a brain MRI [Despotović et al., 2015].

2.2.2 Computed tomography imaging

Also known as X-ray computerised tomography, it combines X-ray radiography with image processing to produce detailed 2D body images called slices, which can be combined into a 3D volume. CT images are acquired using a CT scanner as the one presented in Fig. 2.2.3.

Figure 2.2.3: Siemens SOMATOM Force CT scanner. [1]

To acquire a CT image, the CT scanner is equipped with an X-ray tube that performs series of X-ray electron beam shots toward the body to be scanned. The X-ray tube moves in a circular motion inside the arch shaped opening of the scanner called gantry. To produce one CT slice, a full circular rotation of the X-ray tube is required. When the body under scan receives the X-ray beams, each tissue type absorbs a different amount of X-ray radiation also called radiation dose. The rest of the unabsorbed radiation is collected by detectors placed opposite the X-ray tube, and then sent to a computer for processing and visualisation. Fig. 2.2.4 shows slices of a CT volume of the brain for one subject presented in different view planes (axial, sagittal and coronal).

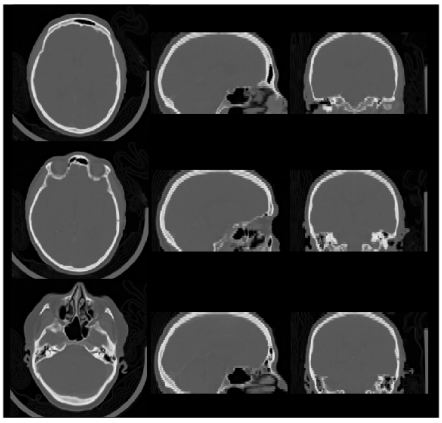

Figure 2.2.4: Axial (left), sagittal (middle) and coronal (right) slices of brain CT images.

[1] `https://www.siemens-healthineers.com/computed-tomography/dual-source-ct/somatom-force`

The amount of the absorbed radiation by a body is referred to as attenuation coefficients (μ) expressed in m^{-1}. In fact, a thick tissue such as bone absorbs more radiation, therefore has a higher attenuation coefficient. In contrast, a thin tissue such as air has a lower attenuation coefficient. In CT imaging, a universal measurement unit is used to characterise CT image intensity values called Hounsfield Unit. In fact, an HU value of a CT image voxel and attenuation coefficients reflect the ED of the specific tissue and both measurment units have a direct relation with ED, therefore when acquiring a CT scan, a linear conversion from ED to HU is performed such that:

$$HU_{Tissue} = 1000 * (ED_{Tissue} - ED_{H_2O})/ED_{H_2O}. \tag{2.2.1}$$

An HU value can vary from -1000 HU to +2000 HU depending on the tissue type. The following table 2.1 shows two possible HU intervals for major tissue types [Onozato et al., 2014, Boydev et al., 2017].

Air	**Soft tissue**		**Bone**	
[-1000; -400] HU	[-399; 600] HU		>601 HU	
Air	**Partial volume**	**Soft tissue**	**Soft bone**	**Bone**
<-870 HU	[-870; -380] HU	[-380; 250] HU	[250; 620] HU	>620 HU

Table 2.1: HU intervals for major tissue types.

2.2.3 Magnetic resonance imaging

MRI is considered as one of the vital imaging modalities employed in clinical practice. It provides indispensable information which other imaging modalities are incapable to capture. The success of MRI is owed to its ability to non-invasively display information about a large scale of tissues. In fact, images produced by MRI exhibit more tissue contrast and reveal bigger body detail compared to other imaging modalities. In contrast to the ionising nature of CT imaging, radiation used in MR imaging is non-ionising thus considered safe for human body and has no side effects.

MRI can be used in a widespread rang of medical applications going from brain function studies, abnormal tissue-growth detection (cancer), Alzheimer monitoring and multiple-tissue visualisation. Unlike CT imaging, MRI uses a powerful magnetic field and radio waves, which make it advantageous over CT imaging since it does not produce any ionising radiation. Fig. 2.2.5 illustrates an example of an MRI scanner.

Figure 2.2.5: Siemens MAGNETOM Sola MRI scanner. [2]

MRI concepts come from quantum physics and originate from studying the nuclear magnetic resonance (NMR) proprieties of hydrogen in tissues [Haacke et al., 1999]. Thus, explaining how MRI works is quite complex. Since the objective of this chapter is to introduce the basic notions for the upcoming chapters, the next lines explain the process of MR image acquisition in a simple manner sparing the reader from complex physics and mathematics.

MRI takes advantage of the presence of the large portion of water in the human body targeting hydrogen nucleus (protons) existing in water molecules. In addition, an interesting propriety of a hydrogen nuclei, which is considered crucial for MR imaging is the "spin" [Moser et al., 2009], this propriety defines an orientation for the hydrogen nuclei. In a normal case, spins of hydrogen nucleus are randomly oriented. Fig. 2.2.6 shows the main components of an MRI scanner.

To acquire an MR image, the MRI scanner performs the following steps:

1. Initially, a strong magnetic field is applied to the patient, the strength of a magnet is measured in Tesla (T);

2. Spins of the hydrogen nucleus within the body line up straight according to the direction of the magnetic field;

3. Another source of energy generated by a radio frequency (RFq) coil is applied to the aligned spins to excite low energy hydrogen nucleus. Many RFq excitations can be released and the time between subsequent RFq excitations is called repetition time (TR);

[2] https://www.siemens-healthineers.com/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner/magnetom-sola

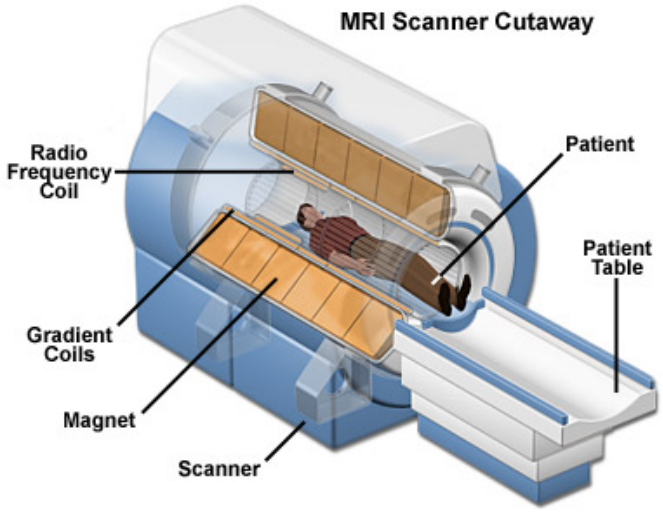

Figure 2.2.6: Main components of an MRI scanner. [3]

4. When the RFq and the magnet are shut down, the excited hydrogen nucleus return to their resting state and release the absorbed energy, this is referred to as relaxation time;

5. Receiver coils intercept the released energy and transmit the signal to computers for processing and visualisation. The time between RFq excitation and signal acquisition is called echo time (TE).

It is the TE and TR parameters that determine the grey level of each tissue type, where the hydrogen nucleus in each tissue type release energy differently and this is what makes MRI feasible. Furthermore, to achieve a slice-by-slice imaging to produce a sequence of MR images, the MRI scanner scans the body section by section. This can be achieved by altering the magnetic field strength using gradient coils along the bodyline going from a stronger field in the head to a weaker field in the feet. An adequate RFq energy to each magnetic field level is applied to target specific hydrogen nucleus in the body. For instance, if one desires to image the head, then a high RFq energy is applied to target the high level of the magnetic field.

When altering acquisition settings of the MRI scanner, different tissues can be targeted, therefore producing several types of MRI sequences. Table. 2.2 presents the main types of MRI sequences employed in clinical practice or under research.

[3] `https://snc2dmri.weebly.com/components--functions.html`

Sequence	Abbreviation	Major tissues visualised
T1-weighted	T1-w	Soft-tissue
T2-weighted	T2-w	Soft-tissue
Proton density	PD	Differentiate between grey matter and white matter
T2-star	T2*	Haemorrhage and calcifications
Fluid attenuated inversion recovery	FLAIR	Fluid suppression with enhancement of grey matter visualisation
Time-of-flight	TOF	Display flow inside vessels
Ultra short echo time	UTE	Bone and bone marrow
Zero echo time	ZTE	Bone
Dixon in phase and out phase	Dixon	Fat and water

Table 2.2: Main MRI sequences with major tissues.

Fig. 2.2.7 shows axial, sagittal and coronal slices of an MRI volume of the brain for one subject presented using three types of MRI sequences.

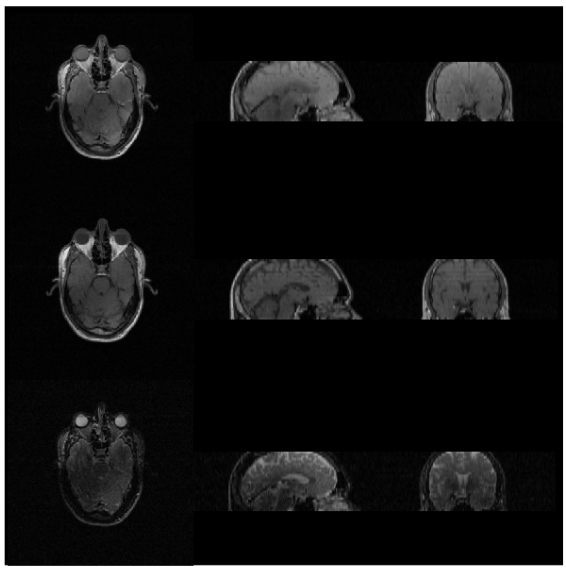

Figure 2.2.7: Axial, sagittal and coronal slices of proton density (PD) (top), T1-w (middle) and T2-w (bottom) MRI sequences.

In some medical cases such as radiation therapy or cancer monitoring, several MR and CT images are used jointly for complementary diagnosis, treatment set-up or monitoring where these images need to be aligned together. However, sometimes the images cover different views of the region of interest or images come from different subjects making the alignment not straightforward and calls for a specialised alignment technique called medical image registration.

The next section describes this technique focusing on types of image registration and steps followed to register medical images together.

2.3 Medical image registration

Notations

I_F	Fixed image
I_M	Moving image
Ω_F	Fixed image spatial domain
Ω_M	Moving image spatial domain
x_i	Voxel number i
(x, y, z)	Spatial coordinates of a voxel
N	Total number of voxels in an image
T	Transform
μ	Registration parameters
$\hat{\mu}$	Registration results
ζ	Cost function
c	Constant
t	Translation
$H(A) = -\int p_A(a) \log p_A(a) da$	Entropy of a random variable
$H(A, B) = \int p_{AB}(a, b) \log p_{AB}(a, b) da\, db$	Joint entropy of two variables

Image registration can be defined as the task of spatially aligning two or more images of the same view (in case of medical image registration: the same anatomical region) at different times and/or acquired by different imaging modalities to obtain a more complete information [Zitova and Flusser, 2003].

Medical image registration can be used for tumour growth monitoring, treatment verification and patient data comparison to other clinical cases. The process of medical image registration is interpreted as applying an appropriate transformation model that transforms a secondary image (moving image) to the coordinate system of a primary image (fixed image) [Devic, 2012], taking into account the degree of alignment between the images, which constitutes the similarity measure that needs to be optimised.

In other words, image registration can be seen as an optimisation task where a similarity measure is minimised throughout several iterations while computing a transform and applying interpolation. It can be formalised as:

$$\hat{\mu} = arg\,\min_{\mu} \zeta\left(\mu;\, I_F,\, I_M\right). \tag{2.3.1}$$

The following diagram on Fig. 2.3.1 presents the main components of an image registration workflow.

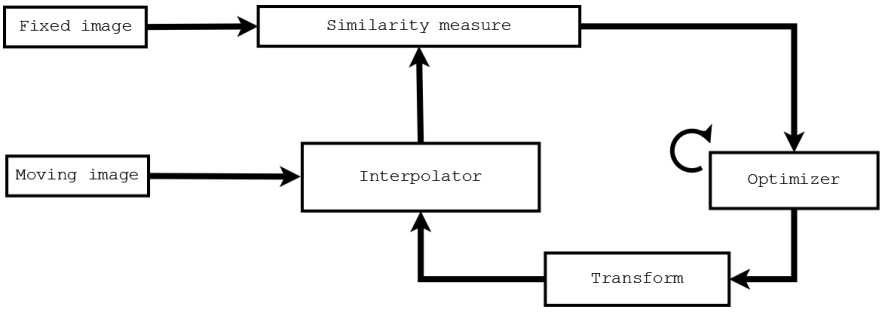

Figure 2.3.1: Main components of an image registration workflow [Johnson et al., 2013].

For the sake of brevity, the main components of image registration are not presented in detail. However, for more details, the reader may refer to the following sources [Brown, 1992, Joseph et al., 2001, Lester and Arridge, 1999, Maintz and Viergever, 1998, Rueckert and Schnabel, 2010].

Moving image: is the image which will be deformed to correspond to the fixed image;

Fixed image: is the image from which deformations will be computed to be applied to the moving image;

Similarity measure: the degree of alignment between images can be measured using several approaches. From which feature-based and voxel-based approaches exist. The former is based on using features such as points, lines, or surfaces and the registration aims to minimize the distance between the corresponding features in both images. However, this technique requires the use of landmarks and application of segmentation techniques. Voxel-based approaches make the use of image intensities from both images and measures the degree of shared information. Among many, the following measures are widely employed [Johnson et al., 2013]:

1. **Mutual information (MI):** considers the probability distributions of image intensities from the fixed and the moving images. Its formula is given by:

$$MI(I_F, I_M) = H(I_F) + H(I_M) - H(I_F, I_M). \tag{2.3.2}$$

MI is more suitable for multi-modal image registration. A variant of MI is normalised MI (NMI), which introduces the normalisation factor $H(I_F,I_M)$.

2. **Mean squared difference (MSD):** incorporates image intensities directly to the computation of the similarity between the fixed and the moving images, thus it's less suitable for multi-modal image registration and best suited to be applied for mono-modal image registration. It is expressed using the following formula:

$$MSD(I_F, I_M) = \frac{1}{N} \sum_{i=1}^{N} \left(I_F(x_i) - I_M(x_i)\right)^2 . \tag{2.3.3}$$

3. **Normalised correlation coefficient (NCC):** assumes a linear relation between image intensities of both fixed and moving images. It can be formulated as follows:

$$NCC(I_F, I_M) = -1 \times \frac{\sum_{i=1}^{N} \left(I_F(x_i) \cdot I_M(x_i)\right)}{\sqrt{\sum_{i=1}^{N} I_F(x_i)^2 \cdot I_M(x_i)^2}}. \tag{2.3.4}$$

Transform: although registration is defined as deforming the moving image to correspond to the fixed image. Mathematically speaking, registration is defined as finding a coordinate mapping from the fixed image domain to the moving image domain [Klein et al., 2010]. This is because transformations are mainly defined from the fixed image, and then applied to the moving image in order to deform it.

A transform tries to map one point from the fixed image to its corresponding location on the moving image. In some cases, the points to interpolate do not lie in the same location; therefore, transformations can be simple or complex depending on the application. The degree of complexity of the transformation is characterized by a number of parameters, which present the degrees of freedom of the transformation [Rueckert and Schnabel, 2010]. Three major types of transformations can be found:

1. **Rigid transform:** used when the geometric relationships between points do not change [Brown, 1992]. An image is considered as a rigid body and manipulated by translations and rotations without any stretching or scaling. It's formulated as:

$$T_\mu(x) = R(x - c) + t + c. \tag{2.3.5}$$

2. **Affine transform:** employed to correct for scaling especially when there is a global gross overall distortion [Devic, 2012]. it is more flexible than the rigid transform wherein images can be translated, rotated, scaled or sheared. Its formula is given by:

$$T_\mu(x) = A(x - c) + t + c. \tag{2.3.6}$$

3. **B-spline transform:** a small number of parameters characterizes the above-mentioned transformation models. Therefore, they are simple to apply and less computationally demanding. However, they are not applicable when important deformations are expected [Rueckert and Schnabel, 2010]. In this case, non-rigid transformation models are applied in which B-spline transform falls. The latter is based on using a set of control points identified in the fixed and moving images and displacements are computed in order to map the locations of control points on the moving image to their corresponding ones on the fixed image. It can be formulised as:

$$T_\mu(x) = x + \sum_{x_k \in N_k} \rho_i \beta^3 \frac{x - x_k}{\sigma}. \tag{2.3.7}$$

x_k are control points, ρ_k are B-Spline control points displacements', σ is B-Spline control point spacing, N_k is the total number of control points and β^3 is is the cubic multidimensional B-Spline polynomial.

Interpolator: After mapping a point from the fixed image to the moving image, image intensity evaluation is required; this is achieved using interpolation. Many interpolation techniques exist from which nearest neighbour interpolation, linear interpolation and B-Spline interpolation.

Optimiser: An optimiser attempts to minimise the similarity measure over several iterations considering transform parameters and following a certain search direction. Gradient descent (GD) is commonly used as an optimiser for image registration tasks.

Both mono-modal and multi-modal images can be considered for image registration. In addition, it is possible to perform image registration over several image resolutions for improved accuracy and performance [Johnson et al., 2013]. It proceeds by going from lower resolutions to register small deformations, and moving on to higher resolutions to register large deformations. However, the set of registration parameters must be carefully chosen since an inappropriate choice of one parameter can lead to undesired results.

Fig. 2.3.2 shows axial slices of two T2-w MR images for two different subjects before and after applying deformable image registration to align the moving image to the fixed image. Registration was performed by using B-spline deformable image registration algorithm with MI

as a similarity measure which was optimised using adaptive stochastic gradient descent over 600 iterations considering four resolution levels.

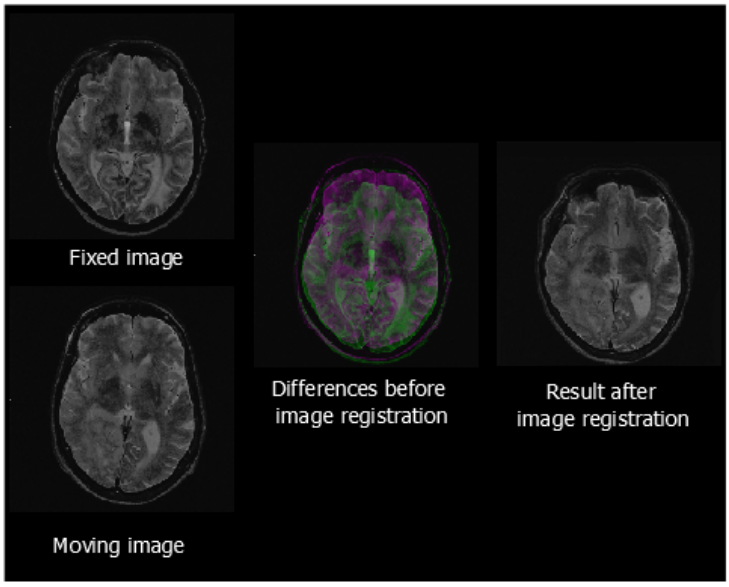

Figure 2.3.2: Axial slices of two T2-w MR images before and after deformable image registration.

One of the major applications of medical image registration is radiation therapy, wherein MR and CT images are aligned together using one or more of the image registration types presented above. Both modalities provide complementary information and each considered crucial for radiation dose calculation and treatment delivery. In the next section, radiation therapy is presented in detail with emphasis on the joint use of MR and CT imaging, its advantages and limitations and first attempts to improve radiation therapy treatment planning workflow.

2.4 Radiation therapy

The treatment of malign or benign tumours can be performed using one or a combination of three treatment options, namely: chemotherapy, surgery and radiation therapy. The latter allows to treat cancer patients by delivering high doses of ionising radiation targeting tumour cells to kill them while sparing healthy tissue. A common type of RT is adaptive external radiation therapy in which the prescribed ionising radiation dose for treatment is fractionated over several treatment sessions and employs a treatment machine called linear accelerator (Linac). This machine operates non-invasively by sending high doses of radiation electron beams at different angles while trying to maximise the delivered dose to the tumours and minimise radiation

to healthy cells. Fig. 2.4.1 shows an example of a linac machine used for adaptive external radiation therapy.

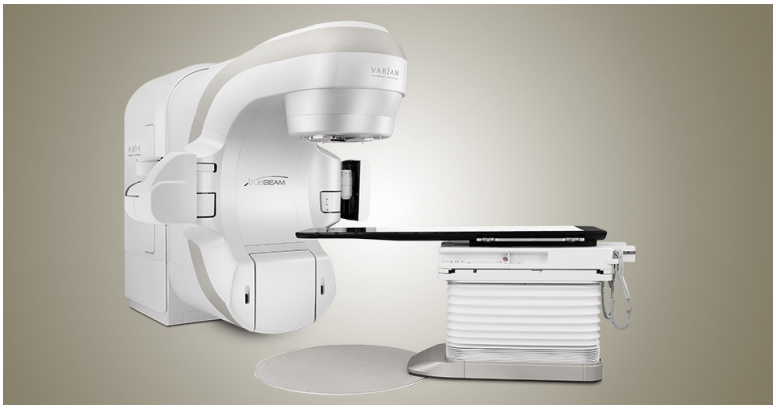

Figure 2.4.1: Varian TRUEBEAM linac machine for external radiation therapy. [4]

To prepare for RT treatment, several steps must be followed:

1. The patient is prescribed with a specific radiation dose expressed in Gray (Gy). It's the distribution and delivery process of this radiation dose that will be planned and computed considering tumour location, size and linac settings;

2. Initially, several 3D imaging modalities are used for tumour localisation and size definition. CT imaging is considered the standard imaging modality for RT since it contains the necessary ED information for treatment calculation. HU values in CT scans are converted to ED values that are necessary for radiation dose calculation [Podgorsak et al., 2005]. However, accurate delineation of the tumour and other organs on CT scans is difficult because of its poor soft tissue contrast. In the case of prostate and rectum cancers, studies have shown that the prostate and rectum volumes are over estimated when contoured on CT scans compared to the delineated volumes on MR images [O'Neil et al., 2009, O'Neill et al., 2014, Rasch et al., 1999]. Therefore, MR images are acquired in parallel for an enhanced tissue visualisation and for tumour contouring;

3. The next step requires that all images to be aligned using either rigid or deformable multi-modal image registration techniques or both. This is done to bring all images to the same coordinate space since images come from different scanners with different physical proprieties;

4. After that, the aligned images are used to perform necessary tumour contouring. In most cases, MR images are used for contouring and the subsequent contours are propagated

[4] `https://www.varian.com/oncology/products/treatment-delivery/truebeam-radiotherapy-system`

to CT scans using image registration [Chao et al., 2008, Thörnqvist et al., 2010, Van Der Put et al., 2009];

5. The contoured CT scans are downloaded to a treatment planning system (TPS) that uses computerised techniques to perform the treatment planning and calculate the dose distribution across the entire volume taking into account the linac parameters settings and the prescribed dose;

6. After treatment planning, all that has left is to deliver the radiation dose as planned to the patient fractionated over several treatment sessions.

2.4.1 Limitations of radiation therapy treatment planning

The presented workflow is valid for conventional external adaptive RT and it is repeated during the course of radiation treatment for dose recalculation and optimisation to account for potential tumour changes in size and location. However, this workflow presents some limitations and pitfalls on different scales:

- Multi-modal imaging requires that the patient to be in the exact same position for each modality imaging. In addition, time between each image acquisition should be as short as possible to avoid any significant structural differences. The need for several image acquisition can be time-consuming and increases workload;

- On the image registration level, alignment of multi-modal images is not as simple as mono-modal image alignment, since the former employs images of different physical proprieties and may introduce registration uncertainties [Ulin et al., 2010];

- Contour propagation has shown to introduce uncertainties reaching 2-5 mm in body regions such as the prostate and rectum [Nyholm et al., 2009, Ulin et al., 2010];

- Due to the adaptive nature of external RT, CT scans are acquired prior to each treatment session for treatment optimisation. In this case, the patient becomes subject to the risk of excessive radiation exposure, thus creating a potential environment for cancer development [de Gonzalez and Darby, 2004].

These facts are pushing research to improve RT workflow. One possible solution is to attempt to remove CT imaging because the main limitations are inherited from CT such as excessive dose exposure and low soft-tissue visualisation and opt for an MRI-only RT workflow. This would completely delete CT imaging from treatment planning and keep MR imaging. MR imaging is characterised by a high soft-tissue visualisation compared to other modalities and there are no

proven side effects for NMR on the human body making MRI a safer and an effective imaging modality to be used solely for RT.

2.4.2 Toward MRI-only RT

Moving toward an MRI-only RT workflow is not straightforward. This is because MR images do not provide the necessary HU information for dose calculation and treatment planning due to the physical proprieties of MR imaging. Physically, hydrogen nucleus (protons) have no direct relation with electron densities making a direct conversion from proton values to electron values infeasible. As a proof of evidence, [Ito et al., 2017] investigated the feasibility of a direct conversion from magnetic-susceptibility to relative ED. However, the correlation between aligned magnetic susceptibility maps and relative ED images was 0.0145 indicating a very weak correlation.

Many studies have discussed the potential of MRI in replacing CT in RT [Beavis et al., 1998, Boettger et al., 2008], this can be achieved by synthesising a CT image from MRI, also called pCT. In other words, specific techniques need to assign ED information to MR images. The first approaches for assigning ED information to MR images consisted of setting the whole body to a uniform ED value usually corresponding to water and assigned a different ED value for bone volume [Beavis et al., 1998, Lee et al., 2003, Jonsson et al., 2010, Lambert et al., 2011]. Despite its simplicity, these approaches cannot generate a reliable pCT image for dose calculation and may lead to erroneous results [Gudur et al., 2014]. Other improved approaches manually segmented the anatomy into different tissue classes and assigned uniform ED values for each tissue class [Lee et al., 2003, Jonsson et al., 2010, Lambert et al., 2011, Eilertsen et al., 2008, Pasquier et al., 2006]. Nevertheless, these approaches suffer from segmentation errors and manual segmentation is time-consuming and tedious.

Since the last decade, more sophisticated approaches have been developed in order to assign the ED information to MRI data. They have proven to synthesise realistic and naturally-looking CT images using MRI data which are promising to improve RT workflow. These approaches are classified and reviewed in detail in the next chapter with discussion of their main advantages and limitations.

2.5 Conclusion

Recently, RT has been attracting researchers and regarded a hot research topic for medical image processing tasks proven by the recent increasing number of published research papers. Indeed, synthesis of a pCT image from MRI data can be formulated as an inter-modality

image synthesis task, which would call for using innovative image processing and other creative techniques to fulfil this task. In this chapter, we presented a walk through clinical notions to set the base for the next chapters. As diving directly into understanding the principal work done would be lacking clearness without having a basic knowledge on the underlying notions and concepts. We also set the ground for the next chapter in which a literature review is presented with emphasis on the main works proposed to synthesise pCT images from MRI data for MRI-only RT with a categorisation of works and a critical analysis.

CHAPTER 3

Literature review on methods for pCT image synthesis from MRI

3.1 Introduction

Thanks to its non-ionising physical proprieties and its excellent soft-tissue visualisation [Moser et al., 2009], MRI is undertaking the lead in the field of RT to replace the baseline CT imaging. This is motivated by the recent introduction of online MRI-guided treatment using MRI linacs. However, implementing an MRI-only RT workflow is challenging because MR intensities have no direct correlation with CT electron densities [Ito et al., 2017] required for dose planning. Alternatively, research is aiming to indirectly assign HU values that characterise CT scans to MRI data. Such an MRI-driven CT image is often referred to as pCT image. In this chapter, we review and classify methods for generating pCT images from MRI data. A classification of these methods with a detailed description of research works involved in each category of methods and some statistical performance results are presented. We further address the advantages and drawbacks of each category and discuss new research orientations to deal with the limitations and difficulties encountered. This chapter is a revised and an extended version of two papers, one presented in [Boukellouz and Moussaoui, 2017a] and the other published in [Boukellouz and Moussaoui, 2018].

First, we present the commonly used performance metrics employed to evaluate the prediction accuracy of the generated pCT images, after that we discuss the availability of MRI-CT data and review in-depth the latest methods for estimating pCT images from MRI data with a comparative analysis and discussion.

3.2 Evaluation measures

To evaluate the performance of a certain method for creating a pCT image from MRI, the generated pCT image is compared to the real acquired CT image using different performance measures. We grouped evaluation measures into statistical and dosimetric measures. The current section presents the common metrics used for evaluation.

3.2.1 Statistical evaluation measures

The following measures can be used.

Mean absolute error (MAE) and mean error (ME) MAE and ME are widely employed in MRI-only RT, they aim to evaluate the voxel-wise differences between the predicted pCT and the real CT images in HU values. Their respective equations are given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|CT(i) - pCT(i)|\,, \tag{3.2.1}$$

and

$$ME = \frac{1}{N}\sum_{i=1}^{N}(CT(i) - pCT(i)). \tag{3.2.2}$$

where N is the total number of voxels in a CT volume.

Mean absolute error histogram ($\mathbf{MAE}_{hist}$) and mean error histogram ($\mathbf{ME}_{hist}$) The MAE_{hist} gives a better understanding regarding the distribution and location of the MAE per tissue type, while the ME_{hist} provides an indication on the presence of a bias towards over-estimation or under-estimation of HU values in each tissue. The HU scale that defines tissue types is divided into consecutive bins of 20 HU; then, MAE and ME are computed inside each bin between the generated pCT image and the real CT scan. Respective formulas of the MAE_{hist} and ME_{hist} are given by:

$$MAE_{hist} = \frac{1}{N}\sum_{i\in Bin}|CT(i) - pCT(i)|\,,\ i = 1...N, \tag{3.2.3}$$

and

$$ME_{hist} = \frac{1}{N}\sum_{i\in Bin}(CT(i) - pCT(i)),\ i = 1...N. \tag{3.2.4}$$

where, N is the number of voxels inside Bin, $pCT(i)$ and $CT(i)$ are the HU values of the voxels number i in both the estimated pCT image and the real CT scan belonging to Bin.

Root mean squared difference (RMSD) Similar to the MAE but is less employed in MRI-only RT, it's given by:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} \left(CT(i) - pCT(i)\right)^2}{N}}. \tag{3.2.5}$$

Where $pCT(i)$ and $CT(i)$ are the HU values of the voxels number i in the estimated pCT image and the real CT scan and N is the total number of voxels in a CT volume.

Dice similarity coefficient (DSC) DSC [Dice, 1945] measures the tissue overlap between the original CT and the pCT volumes. It's commonly applied for bone tissue. It is given by the following formula:

$$DSC = \frac{2 * (V_{CT} \cap V_{pCT})}{V_{CT} + V_{pCT}}. \tag{3.2.6}$$

where V is the binary map for a specific tissue volume in the original CT and pCT images. DSC takes values in the interval [0; 1], a higher DSC indicates a good tissue overlap.

Correlation The correlation between the estimated pCT image and the real CT scan is computed using the following Pearson correlation formula:

$$Correlation = \frac{\sum_{i=1}^{N} \left(CT(i) - \overline{CT}\right)\left(pCT(i) - \overline{pCT}\right)}{\sqrt{\left(\sum_{i=1}^{N} \left(CT(i) - \overline{CT}\right)^2\right)\left(\sum_{i=1}^{N} \left(pCT(i) - \overline{pCT}\right)^2\right)}}. \tag{3.2.7}$$

Where N is the total number of voxels in an image and $\overline{CT}$ and $\overline{pCT}$ denote the mean HU value of the real CT image and the estimated pCT image, respectively. Correlation takes values in the interval [0; 1], a higher value indicates a strong correlation.

Structural similarity (SSIM) SSIM measures the similarity between two images and evaluates the quality of the predicted image. It's given by:

$$SSIM = \frac{\left(2\overline{CT}\,\overline{pCT} + c_1\right)\left(2\delta_{CT_{pCT}} + c_2\right)}{\left(\overline{CT}^2 + \overline{pCT}^2 + c_1\right)\left(\delta_{CT}^2 + \delta_{pCT}^2 + c_2\right)}. \tag{3.2.8}$$

Where $\overline{CT}$, $\overline{pCT}$, δ_{CT}^2, δ_{pCT}^2 and $\delta_{CT_{pCT}}$ are the means, standard deviations, and cross-covariance for images CT and pCT, respectively. c_1 and c_2 are parameters used in stabilizing

the equation.

Peak Signal-to-Noise Ratio (PSNR) Similar to SSIM, PSNR is a measure to compare the quality of the estimated pCT image to its corresponding CT image. it's given by the following formula:

$$PSNR = 10log_{10}\left(Q^2 \diagup \left(\frac{1}{N}||CT - pCT||_2^2\right)\right). \tag{3.2.9}$$

where Q is the highest intensity value between the CT and pCT images, N is the total number of voxels in an image and $||.||_2^2$ is the mean squared error. A higher PSNR indicates a good image quality. The result is expressed in decibel (dB). SSIM and PSNR measures are less commonly employed in MRI-only RT.

Standard Deviation (SD) Measures the depression or variation of a set of data points relative to its mean, a lower SD indicates that data points are close to the mean and a higher SD indicates that data points are far from the mean. It's given by the following formula:

$$SD = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}{N-1}}. \tag{3.2.10}$$

where $\overline{x}$ is the mean of the data points, x_i is the data point number i and N is the total number of data points. SD is expressed in the same unit as the data, for example, if we calculate the SD of the MAE values then the SD is expressed in HU.

Box and whiskers plots for the MAE and ME distributions Box and whiskers plots aid at better understanding the distributional characteristics of MAE and ME values across the error range. A box plot is constructed by ordering data values following an increasing order, then the median value is determined and four quartiles are defined, each grouping 25% of the data to plot. First and third quartiles define the upper and lower bounds of data and they are represented by whiskers, while the inter quartile range is represented by the middle 50% of data and illustrated by a box.

Cross-validation (CV) Cross validation is a strategy used to evaluate and validate the performance of a predictive model. The available data set is partitioned into a fixed number k of groups or folds, then the predictive model is constructed using *k-1* folds and evaluated on the remaining fold. This process is repeated for each fold of the data set. A special case when the fold contains only one data item is called leave-one-out cross validation (LOOCV).

3.2.2 Dosimetric evaluation measures

Metrics in this category evaluate the dose differences of the generated treatment plans between the original CT and the pCT images. One common metric is the gamma index γ [Low et al., 1998] that evaluates the dose distribution differences between each voxel in the original CT and the estimated pCT images. Another measure is the dose volume histogram (DVH), which compares dose distributions by expressing the minimal dose (DV) that a volume V (expressed in percent) receives. For example, D_{98} is the minimal dose delivered to 98% of the volume of interest.

Note: the underlying work in this dissertation does not incorporate dosimetric evaluation because it's of less interest in the field of computer science to which this dissertation work belongs, and it would be of more interest to be considered in a medical-physics targeted dissertation work. Yet, in clinical practice a dosimetric evaluation is considered crucial to be performed in order to assess the effectiveness of a certain pCT estimation method.

3.3 MRI-CT image data availability

As all multi-modal approaches, MRI-only workflows require multi-source imaging data, at least one MRI modality for building the framework and a mandatory CT modality used for evaluation and possibly for implementation. In addition, it is required that each subject has at least one MR image and its corresponding CT scan and time between acquisition of MR and CT images for each patient should be as short as possible in order to avoid any structural or anatomical changes. The majority of works in MRI-only RT use clinical imaging data from hospitals acquired with clinical review board approval. Generally, these images are not made publicly available.

To the best of our knowledge, there are three publicly available MRI-CT data sets that can be used for MRI-only RT: Multi-modality vertebra data sets [1]: the data set contains 20 MRI-CT 3D spine images dedicated for spine recognition. So far, no published work has implemented an MRI-only RT workflow based on these data sets; Alzheimer's Disease Neuroimaging Initiative (ADNI) database[2] which contains 16 CT and MRI brain images and The Retrospective Image Registration Evaluation (RIRE) Project[3] which offers data sets known as the Vanderbilt database, it contains 17 image volumes for four modalities: T1-w and T2-w MRI, CT and PET. The data sets are mainly provided to compare CT-MRI and PET-MRI registration techniques.

[1] `http://spineweb.digitalimaginggroup.ca/spineweb/index.php`
[2] `www.adni-info.org`
[3] `https://www.insight-journal.org/rire/`

However, they can be used for MRI-only RT.

For the sake of this dissertation, a subgroup of images from the Vanderbilt database was used, because it provides multi-modal MR images which are essential for this work.

3.4 Classification of pCT image generation methods

The existing work in deriving a pCT image from MR images can be broadly classified into four categories: segmentation-based, learning-based, atlas-based and hybrid methods. We established the choice of this classification scheme by considering the general technique applied in the workflow of generating pCT images. Other classifications exist, where three categories were established, each grouping segmentation-based, atlas-based and hybrid approaches [Edmund and Nyholm, 2017]. Others [Johnstone et al., 2018] classified MRI-only methods into segmentation, atlas and voxel-based methods. We preferred to make a broad categorization in order to give more understanding about the methods. This section provides a detailed overview of each category with related works. A comparative analysis is further presented and insights for possible improvements are discussed.

3.4.1 Segmentation-based methods

These approaches rely on segmenting MR images into several tissue classes: usually three, 4 or 5 classes; for example, bone, air, soft tissue and fat; and may employ specialised MRI sequences conceived to enhance tissue visualisation, or using fuzzy logic algorithms such as FCM. Then, each tissue class is assigned a specific HU value representing the targeted tissue type in order to obtain the final estimated pCT image.

[Zaidi et al., 2003] presented an MRI segmentation approach by generating a patient-specific attenuation map whose attenuation values are linearly correlated to HU values. After extraction of the skull and scalp from T1-w MR images, these MR images were segmented using FCM to produce four tissue classes. Each class was assigned an attenuation coefficient to obtain the final MR segmented attenuation map. Overall, This approach follows many processing steps, which may introduce a long processing time. In addition, the intensity inhomogeneity artifacts present in the MR images introduced a shading effect after the segmentation process. This latter issue needed manual intervention from the operator.

Boettger et al. [2008]identified three tissue classes by segmenting sets of two MRI sequences, namely: UTE sequence to enhance visualisation of bony structures and an enhanced contrast sequence for soft tissue visualisation. A pCT image was generated by assigning HU values to each tissue class (air, bone and soft tissue). The limitation of this approach is that the bone

segmentation technique can be used only on UTE sequences since bone visualisation is favoured by these sequences. Moreover, this technique showed undesired results of over-segmenting and under-segmenting partial regions of the brain.

[Berker et al., 2012] presented a four-class tissue segmentation approach, in which they employed an UTE triple echo sequence for an enhanced cortical bone segmentation. Additionaly, soft tissue and adipose tissues were segmented using a Dixon-based decomposition technique [Dixon, 1984, Glover and Schneider, 1991]. The approach presented a misclassification between bone and soft tissue voxels.

[Hsu et al., 2013] presented a classification approach to generate a pCT image from several MR images such as T1-w, T2-w, and an Ultra-short Echo Time (UTE) sequence. Fat and water images were calculated using a Dixon-based method. These images were used to distinguish the major tissue types of bone, fat, solid tissue, fluid and air by applying FCM classification. Each tissue class was assigned a fuzzy membership probability and appropriate HU values. The final pCT image was generated by summing HU values of each voxel. Results showed that the FCM classifier tended to misclassify air as bone and the use of UTE sequences does not completely separate bone from air.

[Rank et al., 2013] employed discriminant analysis to derive a pCT image from eight different MRI contrasts. The MR images were resampled and aligned to their corresponding CT images. Afterward, a threshold mask was used to limit the area of interest and cross-validation was performed for parameter optimisation. Optimisation selected a Turbo Spin Echo (TSE) sequence with an UTE sequence as the best MRI contrast combination with two spatial features. Discriminant analysis was implemented to assign MRI feature vectors of a given voxel to predefined classes using decision rules. Each CT class is a 35 HU rang in the HU scale. Their proposed approach gave an MAE of 81 HU, 95.2 HU and 90.1 HU for each phantom used, respectively. SD of the MAE was in the range of 130 to 152 HU, which is significantly large.

[Su et al., 2015] proposed a method for generation of pCT images based on a single acquisition under-sampled UTE-m Dixon pulse sequence. This later was used to extract different contrast information which was provided as inputs for an unsupervised clustering algorithm to estimate five tissue classes namely: air, brain, fat, fluid and bone. Subsequently, each class was assigned a specific HU value producing the final pCT image. MAE$\pm$SD and ME$\pm$SD were 130$\pm$16 HU -22$\pm$29 HU, respectively.

[Khateri et al., 2014] used a short-TE sequence for bone identification and a Dixon MRI sequence with FCM segmentation. Fat, water and soft-tissue masks were derived from Dixon sequences while bone was segmented from the short-TE sequence using FCM. Each of these tissue classes was assigned a particular HU value.

[Liu et al., 2017b] employed T1-Dixon MRI sequences in a classification scheme combined with shape analysis to create a pCT image. First, an initial mask was generated by combining T1, water and fat images to detect bone and air voxels. Then, a bone shape model [Liu et al., 2017b] was applied to this mask in order to identify the pelvic bone. Femur bone was detected using 3D connected component analysis. To segment the rest of the body, a modified version of FCM algorithm was applied and five tissue classes were identified including compact bone, fat, muscle, and the combination of fat interfaces with bone marrow. The pCT image was generated by assigning a specific HU value for each tissue class. MAE±SD was 274.9±26.9 HU for solid bone averaged over nine patients while muscle had an MAE of 13.7±1.8 HU.

[Bredfeldt et al., 2017] was the only one to work on the liver by employing T1- Dixon MRI sequences with FCM segmentation. Intensity thresholding was applied on T1, water and fat images resulting in an air mask. In addition, the contours of the vertebral bone were segmented from fat images using a learning approach [Peng et al., 2006, Huang et al., 2009]. T1, water, and segmented fat images were used as inputs for a modified FCM algorithm [Hsu et al., 2013], which resulted in different tissue classes depending on their intensity (mid-intensity, lower intensity, bone or marrow in the abdomen). MAE ranged from 0 to 160 HU. The bone outside the segmented anterior vertebra bodies was misclassified as mid/lower intensity tissue. Nevertheless, this misclassification did not seem to affect the dose calculation very much as was reported by the authors.

A recent work by [Su et al., 2019], wherein authors attempted to estimate a pCT image for a body region that is rarely considered by MRI-only works which is the thorax. Authors reconstructed seven MR contrast image from an UTE-mDixon sequence each image favouring a specific tissue visualisation covering air, lung, fat, soft tissue, low density bone and dense bone. Air inside and outside the body was determined by intensity-based thresholding and the rest of the other body tissues were determined using FCM segmentation of the previously reconstructed seven MR contrasts. Finally, each segment was assigned a corresponding HU value describing the appropriate tissue such that: air= -1000 HU, body air=-850 HU, fat=-98 HU, water=0 HU, soft tissue=40 HU, low density bone=385 HU, dense bone= 657 HU.

3.4.2 Atlas-based methods

Approaches in this class consist of using deformable image registration algorithms and a database containing one or several aligned CT-MR images called atlas set. Deformable image registration algorithms are used to register/deform the atlas MR image(s) to the target MR image in order to capture the target's anatomy. Then the resulting registration transformations are applied to the CT atlas to generate the final pCT image.

The atlas-based approaches aim for a mono-modal deformable image registration and try to avoid a multi-modal one because the former considers images having the same physical proprieties (MRI-MRI), therefore, optimising a similarity measure based on the same intensity scale is less error-prone than the case of registering images of different physical proprieties (CT-MRI). However, this was not the case for [Schreibmann et al., 2010] who followed a slightly modified atlas-based approach where an atlas CT image was directly registered to the patient's MR image combining parametric and non-parametric registration algorithms namely B-Spline and Hermosillo, respectively. B-Spline was used to account for large deformations while the Hermossilo algorithm was employed to refine results from the B-Spline registration. To evaluate their results, authors used a 3D surface comparison tool to compare the differences between the pCT and the real CT images. Results showed that 20% of surfaces had an error larger than 2 mm and less than 1% of the surfaces presented errors larger than 1 cm for the bone.

[Greer et al., 2011] presented a complete MRI-only RT workflow, the subject's MR image was automatically segmented by registering it to an averaged set of manually segmented MR images used as an atlas. A matching CT atlas set was used for an automatic estimation of ED values for each segment, therefore, mapping HU values to MR images.

[Dowling et al., 2012] presented an atlas-based ED mapping method for auto-segmentation of MR images and pCT image generation. A composite MRI-CT atlas pair was constructed; that is, a set of training MRI-CT pairs was co-registered using rigid and affine registration. The atlas MR image is generated by iteratively registering training contoured MR images and averaging the results to obtain the atlas. The deformations from the previous registrations were applied to the CT atlas set to generate the CT atlas. In order to estimate a pCT image, the new patient's MR image is registered to the atlas MR image for auto-segmentation, then the same deformations are applied to the CT atlas.

[Gudur et al., 2014] used atlas registration between T1-w MR images and an atlas patient composed of matched pairs of CT-MR images to determine the geometry of each voxel (position) given its intensity in the T1-w MR image. This information is represented by a conditional probability function (PDF) and combined into a unifying posterior PDF. This PDF is used to predict the unknown ED values for the corresponding MR image.

Kraus et al. [2017] used landmark image registration to predict a pCT image based on a rigidly registered pair of MR and CT images. The atlas MR image was registered to the target MRI using a marker-based rigid and deformable image registration. This transformation was applied to the atlas CT producing the pCT image. MAE$\pm$SD ranged from 29.9$\pm$53.8 HU to 37.6$\pm$82.6 HU for the body and from 31.3$\pm$27.3 HU to 37.3$\pm$35.8 HU for the prostate.

Other sophisticated atlas-based approaches use several atlas pairs instead of one pair of MR-CT images and introduce an image fusion scheme in order to improve pCT image estimation

[Gudur et al., 2014, Dowling et al., 2012, Burgos et al., 2014, Uh et al., 2014, Sjölund et al., 2015, Arabi et al., 2016, Mehranian et al., 2016, Ren et al., 2017]. In this context, [Burgos et al., 2014] proposed a multi-atlas method to synthesize a pCT image. They used a set of atlas images composed of T1-w MRI and CT scans. The first step aims at registering atlas images to the target MRI using symmetric global registration and B-spline non-rigid registration. Two similarity measures were used: the local normalized correlation coefficient (LNCC) and the local normalized sum of square differences (NSSD). A ranking scheme was used to assign weights to the registered images, i.e., the better the registration, the higher the weight. This step was followed by application of the deformations to the atlas CT scans and the final pCT image was created by doing a weighted fusion of these deformed CT atlases.

[Uh et al., 2014] used atlas registration to create a pCT image using T2-w TSE MRI sequences based on multiple atlas images. The atlas set was composed of six pairs of aligned MRI-CT volumes. Atlas MRI volumes were registered to the patient's MRI volume using non-linear spatial registration. Then, the same deformations were applied to the atlas CT volumes. To generate the pCT image, authors used three schemes to combine the deformed CT images, namely: arithmetic mean of each voxel in each of the deformed CT images (Mean6), pattern recognition with Gaussian process on 6 atlases, and on 12 atlases (PRGP6 and PRGP12). pattern recognition with Gaussian process computes the intensity value of each voxel as a weighted average of the corresponding voxels of the deformed CT images. Results showed that the generated pCT images through PRGP12 performed better in terms of RMSD compared to the real CT images. However, all three fusion schemes showed a smoothing effect where the high intensities in bone smear into the surrounding tissue.

[Sjölund et al., 2015] proposed an atlas-based regression technique. T1-w MRI and CT atlases were aligned with rigid registration. Then, the MRI atlas volumes were registered to the target MR image using deformable image registration algorithm Morphon [Knutsson and Andersson, 2005] and a binary mask to limit the registration on the brain. Morphon algorithm uses local phase differences between signals of similar local frequency to estimate the spatial shift. The registration was performed on different scales to capture large global displacements and small local deformations. The resulting deformations were then applied to the atlas CT scans. The pCT image was created by fusing the deformed atlases by iteratively registering the set to its joint mean.

[Mérida et al., 2015] presented a maximum probability approach. Authors initially performed a standard multi-atlas registration to produce deformed atlas CT images. After that, for each voxel in the deformed CT volumes, a maximum probability class label was calculated considering three tissue classes (soft tissue, air, and bone) and the final pCT voxel value was calculated by averaging HU values belonging to the maximum probability class of the corres-

ponding voxels in all the deformed CT images.

[Ren et al., 2017] proposed a unifying probabilistic multi-atlas scheme where PDFs were used to express the unknown ED value of a certain voxel given its corresponding MR intensity and its location on a reference MRI.

3.4.3 Learning-based methods

Learning-based approaches assume a one-to-one correspondence between MRI and CT voxels which can be modelled by a mapping from MR intensity values to CT HU values. For this, machine learning algorithms are used to build regression/prediction models for MRI to CT mapping.

Johansson et al. [Johansson et al., 2011, 2012] used three MRI sequences, namely: T2-w 3D Spin Echo (SE) sequence and two UTE sequences. Their approach used a Gaussian mixture regression (GMR) model to link MR and CT intensities. The pCT image voxel values were calculated based on the estimated regression model. Streak artefacts in MR images caused the major differences between the original CT and the pCT images. Furthermore, despite the use of UTE sequences to differentiate air and bone, their suggested approach had large deviations between the real CT and the pCT images located at the air-soft tissue and bone-soft tissue interfaces. To address this problem, authors extended their work in [Johansson et al., 2013] by incorporation of spatial information, namely: x, y and z coordinates of each voxel and the shortest distance from each voxel to the external contour of the anatomy. Results showed an improvement in complicated small structures, but no improvement was observed in larger anatomical volumes.

[Navalpakkam et al., 2013] performed a voxel-wise addition of UTE and Dixon water and fat images. The resulting image was used to identify air regions by applying k-means clustering algorithm. To learn the relations between these processed MR images and their corresponding CT images. Five patients were integrated into the learning phase and SVM regression was used. The generated regression model was applied to five new MRI patients to predict their pCT images.

[Kapanen and Tenhunen, 2013] used a T1/T2*-w 3D Gradient Echo (GE) MRI sequences focusing on pelvic bone estimation. The relations between MR intensities and HU values were expressed using a polynomial model that is based on MR signal intensity and fitting parameters. The voxels of the segmented bone structures were grouped into sixteen subgroups based on their MR intensity values. The polynomial model was then applied to each of these subgroups to convert the mean MR intensities to HU values.

[Kim et al., 2015] presented a voxel-based weighted summation approach for generating a

pCT image from four MRI sequences (3D T1-w Fast Field Echo (FFE) sequence, 3D T2-w TSE sequence, 3D balanced Turbo Field Echo sequence (bTFE) and an inverse intensity volumetric image) for prostate cancer patients. Authors manually contoured bone on T2-w sequences. To determine other tissues, each HU value was calculated using a weighted summation of the intensities of the corresponding voxels of the four MRI sequences. Weight optimization was performed to reduce errors starting with random weights to generate an initial pCT image and optimizing by reducing the Euclidian distance of voxel value differences between the pCT and real CT images over 1000 iterations. Afterward, the resulting optimized weights were used for the next pCT image generation. Average MAE was 74.3 $\pm$ 10.9 HU for nine subjects; errors were located around bone contour borders. However, manual bone contouring may introduce uncertainty and it is time-consuming.

[Korhonen et al., 2014] presented a dual HU conversion model for deriving a pCT image from MRI sequences based on two conversion models. The first model was used to convert soft tissue MR signals to HU values: the idea is to mark regions of interest (ROIs) on CT scans (100 ROIs for each patient) covering muscle, urine, fat, prostate and rectal wall. These ROIs were then transferred to MR images via image registration. The MR intensity scale was divided into ranges, each describing signals from different tissues (muscle, urine, fat, prostate and rectal wall) and each MR intensity value belonging to a specific signal rang is converted to the corresponding HU value. The second conversion model used the approach of [Kapanen and Tenhunen, 2013] for converting signals from the bone tissue. Average ME between the pCT and real CT images ranged from -2 to 5 HU for soft tissue, and from 22 to 78 HU for bone. However, the conversion model overestimated HU values for low ED structures, and the manual bone segmentation introduces a significant time. Moreover, the approach required manual adjustments of the conversion model's parameters for each patient, which is not practical in clinical routine.

[Roy et al., 2014] followed a Bayesian scheme to generate a pCT image from two UTE MRI sequences, each voxel of the reference images (two UTE sequences and one CT scan) and the subject data were represented by a feature vector called a patch. The subject and reference patches represent a local pattern of intensities that have been scaled to a similar intensity rang. Once the pairs of CT and MRI for both subject and reference are linked using a Gaussian distribution, the unknown pCT patches for the subject are predicted and combined through the use of Bayesian networks.

[Zhong et al., 2016] used kNN regression with learned local descriptors to predict a pCT image from T1-w and T2-w MRI sequences. The approach consisted in extracting local descriptors or features for each region in the MRI sequences; transforming these local descriptors to a compact form using a supervised descriptor technique [Zhen et al., 2015] and eventually predicting

a pCT image using kNN regression; where, for each point x in a subject's MR image, a local search window centre is defined in the same location in the training CT-MRI pairs. The prediction of the pCT image was done by searching the k-nearest neighbours of each point descriptor in the test MR image, the result is k MRI descriptors, by considering the alignment propriety of the MR-CT pairs, the k samples of the corresponding CT patches are obtained. The final pCT image is generated by performing a weighted average on the overlap CT patches.

[Ghose et al., 2017a] used pairs of registered point-wise encoding time reduction with radial acquisition (PETRA) MR and CT images to build a regression model, where a tissue-specific RF was trained using probability maps obtained as a result of applying expectation maximisation (EM) clustering on MR images. Following the same concept, authors proposed in another work [Ghose et al., 2017b], a regression and statistical shape model where muscle and fat were separated using EM clustering while the bladder and bone were identified based on a statistical shape and appearance model.

Even though ensemble methods are widely employed in computer vision and other medical fields, this ML discipline is relatively new to MRI-only RT. Indeed, it wasn't until 2016 that the first ensemble method was proposed by [Huynh et al., 2016] who used structured RF and auto-context (AC) model to estimate pCT images. Their method consisted in partitioning each MR image into sets of patches where each patch is characterized by four features extracted at three levels (voxel level, sub-region level and whole patch level) including spatial information. To learn the relations between MRI and CT patches, structured RF were employed to train the MRI features. The resulting forests were used to predict initial pCT patches. An AC model was employed to enhance pCT image prediction. The previously estimated pCT patches were used jointly with MR features to be trained for new sets of RF to yield the final pCT image.

[Andreasen et al., 2016] presented a similar work where spatial, texture and edge features were extracted from ten pelvis T1-w MR images. Initially, classification RF were trained to classify tissues into air, fat, soft-tissue and bone. Then, AC features were extracted from the probability maps, and were subsequently combined with the initial MRI features to train another RF. This process was repeated three times. Finally, a regression RF for pCT image estimation was trained on the initial MRI features combined with the AC features extracted at the final iteration.

[Largent et al., 2017] extracted three types of image features considering spatial, texture and edge information from ten pelvis T2-w MR images. The feature maps were then decomposed into patches, each of size 20 x 20 x 20 to constitute inputs for a conditional inference RF (CIRF). pCT image patch values were estimated by computing the corresponding HU values of the averaged weights from 30 conditional inference trees. Authors stated that bone was not well estimated and suggested incorporating an UTE sequence for bone visualisation rendering.

[Hu and Zhang, 2018] proposed a 3D group feature extraction method with alternative regression forests (ARF). Initially, MR images were divided into 3D patches from which four features were extracted. An ARF was trained to map 3D MRI features to 3D CT patches to predict a pCT image.

[Yang et al., 2017] extracted two sets of features from T1-w MRI patches and reduced the feature set using patch preselection based on logistic sparse least-absolute-shrinkage-and-selection operator (LASSO) [Aseervatham et al., 2011]. A regression RF was trained on the preselected features to eventually predict a pCT image.

[Lei et al., 2019b] got inspired by the works of [Huynh et al., 2016] and [Yang et al., 2017] to estimate a pCT image for brain MRI-only RT. Initially, extracted patch MR features were employed into a classification step using classification RF to identify three tissue classes, namely: air, soft tissue and bone. The output of this segmentation are probability maps that were used to extract semantic information characterised by the surrounding probabilities at 27 different locations for each voxel. Together with the initial MR features these semantic features were trained for another classification RF and iteratively repeated this process until convergence. Probability maps obtained at the final step are used jointly used with the initial MR features to train a regression RF. Identically to the training of the classification RF. A first prediction of pCT image was used to extract AC features that were subsequently combined with the initial MR features and then iteratively trained regression RF until convergence of the model. Despite the good results achieved (MAE=57.45±8.45 HU), authors reported that the model required 30 mn to predict a pCT volume which is significantly large compared to other ML models that require few seconds.

Deep learning [Goodfellow et al., 2016] is an ML discipline that has emerged from the advances in ANN and has recently paved its way to the field of MRI-only RT providing novel methods for pCT image estimation. The first work to examine the potential of deep learning in MRI-only RT was of [Han, 2017] who proposed a deep convolutional neural network (DCNN) model that takes a 2D MRI slice and produces its corresponding 2D CT slice. The proposed DCNN is an adaptation of an existing U-Net architecture from the work of [Ronneberger et al., 2015]. This model was trained on 18 3D T1-w MR and CT images each with 160 slices in a six-fold CV strategy. Results showed that the DCNN model accurately predicted pCT images of a good quality. However, these images presented large differences in interfaces between tissue types compared to the real CT scans.

Inspired by the work of [Han, 2017], [Fu et al., 2018] modified the SegNet CNN architecture [Badrinarayanan et al., 2017] in order to create two deep CNN models for pCT image estimation. The first is a 2D model that takes a 2D T1-w MR slice and outputs its corresponding CT slice, and the second is a 3D model that takes an entire pelvis MR image volume and estimates its

corresponding 3D CT volume. Both models were trained on 20 pelvis subjects, each having 30 MRI slices and validated using a five-fold CV strategy. Results showed that the 3D model performed better than the 2D version giving MAE values of 37 HU and 40 HU, respectively. [Liu et al., 2017a] employed a deep convolutional auto-encoder network (DCAN) to convert MRI sequences to CT scans through tissue segmentation, where 30 3D T1-w MR images where trained to separate tissues into three classes : air, soft-tissue and bone. The target training tissue classes (masks) were created through intensity-based thresholding of the real CT scans. Upon training, the resulting tissue maps were used to predict the final pCT scan by assigning to each tissue class a specific HU value (air=-1000 HU, soft-tissue=42 HU and bone=939 HU). DSC was calculated between the segmented images and the ground truth CT masks. Results showed that a good overlap was observed between different tissues. However, authors did not include any statistical or dosimetric assessment of HU prediction accuracy since HU value correctness is a key factor for accurate pCT image estimation.

Using specialised MR imaging, [Leynes et al., 2017b] incorporated a U-Net architecture with 13 layers in their approach. ZTE and Dixon water and fat MR images were decomposed into patches each of size 32 pixels x 32 pixels x 16 pixels x 3 channels (each channel refers to ZTE, fat and water image intensities, respectively). Similarly, CT images were decomposed into patches of the same size with one channel referring to the HU value. The model was trained on 10 3D pelvis images, each image comprised of 89 slices and was tested on other 16 3D MR images. The approach seemed to produce natural looking pCT images with an ME=-12$\pm$78 HU. Yet, the model had difficulties in predicting finer bone structures in the spine.

[Xiang et al., 2018] used a deep embedding CNN for pCT image estimation. The model was fed with MRI patches of size 64 x 64 x 3 to generate feature maps that were subsequently used to compute a tentative pCT image in the midway of the network architecture. This tentative pCT image was embedded back to the feature maps in order to improve prediction. Training of the network was performed on 16 brain subjects and 22 prostate subjects and required 2-3 days due to the extremely deep nature of the network (11 layers for brain data and 17 layers for prostate data).

[Dinkla et al., 2018] used dilated CNN by adding a dilated parameter to convolutional layers in the aim of increasing the amount of contextual information in the training step. Differently from other deep learning approaches presented in this review, authors trained the model with 2D axial, sagittal and coronal slices of 26 brain subjects rather than training on axial slices alone. In addition, the model was tested on other 26 subjects in a two-fold CV method. This approach showed to predict accurate pCT scans presenting low MAE=67$\pm$11 HU. A recent work carried out by [Spadea et al., 2019] who trained three U-net architectures separately, each on axial, sagittal and coronal head MR slices. To reconstruct the final pCT image: for each

voxel of interest, its HU values is computed by taking the median of the corresponding voxel values in the predicted multi-plane slices. Evaluation showed that the proposed DCNN model accurately predicted pCT images of a good quality giving MAE a value of 54±7 HU.

Other works used more complicated deep network architectures such as the generative adversarial networks (GAN) that train competing networks for an improved prediction accuracy [Emami et al., 2018, Maspero et al., 2018, Lei et al., 2019a]. More precisely, GAN train a generator and a discriminator. The generator is trained to generate a pCT image and the discriminator is trained to identify whether the estimated image is real or fake. For instance, in [Kazemifar et al., 2019], authors proposed a GAN model in which the generator is characterised by a ten-layered U-net taking as input 2D MR slices and producing their corresponding 2D pCT slices. On the other hand, the discriminator consisted of six convolutional layers. The GAN model was trained on 77 subjects in a five-fold CV with unaligned MR-CT images taking MI as the loss function instead of the MAE. [Kaiser and Albarqouni, 2019] worked on the RIRE dataset (the same we used in the implementation of the proposed approaches presented in this dissertation) and proposed a GAN model that inputs 3D MR patches of size 32×32×32 and outputs corresponding CT patches of size 16×16×16.

3.4.4 Hybrid methods

Works in this category combine methods of the above-described approaches to combine their strengths, reduce their weaknesses and palliate for limitations, approaches in this class are termed hybrid.

The earliest known hybrid approach was proposed by [Hofmann et al., 2008] who combined atlas registration with pattern recognition to predict a pCT image. The approach consists of learning a model on a data set of registered MRI-CT atlases to define a mapping from MR intensities to CT intensities. The first step consists of using B-spline deformable image registration with MI as a similarity measure to construct the MRI-CT atlas database. Then the MRI atlas images were registered to the test MRI volume using spatial normalization and the same transformations were applied to the CT atlases. These transformed atlases were subsequently used to extract pairs of MRI patches and their corresponding CT values at each voxel location. A regression process that incorporates prior knowledge from the transformed images and uses a kernel function to measure the degree of similarity between MRI patches was used to predict HU values for a new MRI patch and generate the final pCT image. The approach overestimated the bone in case of the presence of a proximate lesion.

[Chen et al., 2014] estimated a pCT image from standard T1-w MRI sequences of the head taking into account distinction between bone and air. To do this, authors generated air masks

from CT scans and rigidly registered them to their corresponding MR images. This was done for all the atlas images of the training set, and then all air masks were registered to the patient MR image to locate air. Once this was done. For each air voxel, an average HU value was assigned from atlas voxels of the same location. At tissue location, hidden Markov random field (hMRF) and sparse regression were used to predict HU values. The idea is to measure the resemblance between the MRI atlas patches and the patient MRI patch to assign a high weight if the resemblance is strong, assuming that similar MRI patches correspond to similar CT patches. The final value of the pCT patch is calculated using the weighted average of the overlap CT atlas patches. The approach presented some misclassification where some bone was classified as air.

[Andreasen et al., 2015] presented a patch-based approach for predicting a pCT image from conventional MRI sequences. Five aligned brain MR and CT images constituted a patch database. Each patch *P(x)* in an x MRI location corresponds to a *T(x)* value defining an HU value in the corresponding aligned CT scan. The pCT image for the test patient is predicted using an intensity-based nearest neighbour search in the patch database. For each test MRI patch, the goal is to find the minimized squared L2-norm between the test patch and the database of patches. k most similar CT patches are found and the final pCT patch value is obtained by computing a similarity-weighted average. The structural similarity measure was used to discard the high dissimilar patches and reduce the search space. This approach depends on anatomical similarities and the assignment of the average of the surrounding patches can cause problems when the surrounding tissues are different from the one being calculated.

[Siversson et al., 2015] used an automatic tissue classification approach using statistical decomposition algorithm with image registration to predict pCT images for the pelvic region. A database of contoured MRI-CT pairs were used to register the atlas MR images to the patient's MR image. Each pair registration yielded a candidate organ segmentation and the final MRI segmentation was calculated using a weighted voting method. Next, deformable image registration was applied to each of the atlas MR images in order to align the segmented structures to the newly segmented MR image. The resulting deformations were applied to their corresponding CT images and the final pCT image was generated by doing a weighted fusion all the deformed CT images. MAE was 36$\pm$4.1 HU within the body contour.

[Wu et al., 2015] used local sparse correspondence combination. They assumed that MRI and CT patches (each patch is a feature vector) are located on two non-linear manifolds. The approach consists of constructing initial MRI and CT patch dictionaries by collecting patches within a local search window. Next, kNN algorithm was applied to restrict the number of patches in the dictionaries. The final pCT patches were predicted by doing a weighted average of the corresponding CT patches.

[Demol et al., 2016] introduced an approach that uses deformable image registration combined with MR intensity information to generate a pCT image from T1-w MR images of the brain. The idea behind using MR intensity information is to incorporate MR intensity values into the generation of the pCT image. Firstly, an atlas pair was constructed by registering an atlas MR image to its corresponding atlas CT image via rigid registration with MI as a similarity measure. Then, the atlas MR image was non-rigidly registered to the MR image of the patient using deformable image registration algorithm ANACONDA. The same deformations were applied to the CT atlas. Later, voxel intensity computations were performed on the deformed CT image. i.e., in the deformed MR image, a certain group of similar voxels was selected by doing a local search inside a 9×9×9 box centred at each voxel with the selection threshold set to 10%. The final HU value for the pCT image is calculated by averaging the CT values at the corresponding voxel location of the deformed CT. Comparison with the real CT resulted in an MAE of 150 HU. The main limitation of the approach is that the MR intensity search introduced incoherent CT numbers due to increasing the search area when no corresponding voxels are found inside the search box. Furthermore, noise near bone and air cavities was detected. An extension of this work was carried out by [Boydev et al., 2017], where authors investigated whether the implementation of the hybrid approach using ZTE MRI sequences would improve pCT image correctness. Results showed that the use of ZTE imaging reduced the MAE with about 110 HU in bone regions compared to T1-w MRI-based propagation and correction. However, using a threshold may cause selection of voxels belonging to tissues different of that of the voxel to be corrected.

[Burgos et al., 2017] used a database containing delineated T1-w, T2-w and CT images with a manually segmented image to create a pCT image by combining atlas registration with segmentation in a single iterative framework. First, the atlas T1-w and T2-w images were registered to the target MR image to obtain an initial segmentation, and the transformations were applied to the atlas CT images to obtain an initial pCT image. The initial segmentation and pCT image were used alongside with the atlas database as inputs for the next iteration, and the process was repeated until satisfactory results were obtained. DSC was calculated to evaluate tissue overlap, and values ranged from 0.73 to 0.9 for prostate, bladder, rectum and femur heads. MAE±SD was 45.7±4.6 HU.

[Leynes et al., 2017a] focused on bone identification by employing ZTE imaging into a hybrid process of pCT image estimation. A bone mask was derived from ZTE sequences following several thresholding and processing steps including manual correction. After that, a linear conversion model was built taking ZTE bone intensity values as inputs and outputting their corresponding HU values. HU values of tissues other than bone were estimated using an existing Dixon-based approach [Wollenweber et al., 2013]. The presented approach required

manual correction of bone segmentation, which is time-consuming and tedious.

[Speier et al., 2018] proposed two multi-modal pCT image estimation methods. The first one performed a voxel-wise conversion by initially creating look-up tables. These look-up tables were created by segmenting T1-w and T2-w brain MR images into six tissue classes. Furthermore, the brain was divided into eleven regions in a so-called regionalisation step. Therefore, each voxel would have a tissue class and a region number in the look-up table. For a test MR image, similar voxels to the voxel of interest were retrieved using the look-up table, and the corresponding HU value was taken to form the pCT image. The second method is a recursive patch search, a database of MRI-CT patches was constructed by extracting T1-w and T2-w MRI patches with their corresponding CT patches. For a test MRI image, each MRI patch pair was compared to the patches in the database. CT patches corresponding to the MRI patches similar to the test patches were averaged to produce an initial pCT patch. The patch search was recursively performed taking the initial pCT image prediction with the MRI patches to refine results.

[Aouadi et al., 2017] presented a multi-scale and dual contrast patch-based non-local mean approach using T1-w and T2-w MRI sequences. For each training pair, most similar patches were selected in a sub-region considering two resolution levels. These patches together constituted multi-scale and dual contrast patches. To predict a pCT patch, given its MRI patches, corresponding CT patches from the training set were selected and the final pCT patch value was calculated using the non-local mean fusion as a weighted average.

[Wiesinger et al., 2016] proposed a hybrid method combining tissue classification and linear regression. Soft-tissue and air were determined using intensity thresholding of ZTE MR images and were assigned discreet HU values of 42 HU and -1000 HU, respectively. Whereas bone HU values were determined using a linear regression model that maps a ZTE intensity value to its corresponding HU value.

3.5 Comparative Analysis

Tables 3.1, 3.2, 3.3, and 3.4 present some of the statistical results for a group of methods for pCT estimation approaches as was reported by the authors of the papers. We further included in the tables information about data set size, body region and MRI sequences that were used in the implementation of each method. In each table, works are grouped following the classification presented in this review.

Work	Region	Subjects	MRI sequences	Reported results
[Hsu et al., 2013]	Brain	10	3D T1-w and 3D T2-w; In-phase and out-phase water and fat images; Ultra short echo times UTE sequence; 2D multi-plane time-of-flight (TOF) angiography.	Correlation=0.99
[Su et al., 2015]	Brain	9	Undersampled single acquisition UTE-m Dixon sequence.	MAE±SD=130±16 HU ME±SD=−22±29 HU
[Khateri et al., 2014]	Brain	2	ShTE Dixon MRI sequence.	Bone segmentation (mean±SD) Accuracy=0.977±0.01 Sensitivity=0.947±0.02 Specificity=0.987±0.01
[Liu et al., 2017b]	Pelvis	30	3D gradient echo sequence VIBE Dixon (T1-w, fat and water images).	$MAE_{Solidbone}$±SD=274.9±26.9 HU MAE_{Muscle}±SD=13.7±1.8 HU
[Bredfeldt et al., 2017]	Liver	16	3D gradient dual echo sequence.	MAE∈[0; 160] HU
[Berker et al., 2012]	Brain	7	UTE Triple Echo sequence; 2D mDixon reconstructed water and fat images	Correlation=0.9958 Classification accuracy=81.1%
[Stanescu et al., 2008]	Brain	4	3D T1 Turbo Field Echo (TFE) sequence; 3D T2 TSE sequence.	-
[Zaidi et al., 2003]	Brain	10	3D T1-w MRI sequence.	Correlation=0.91
[Rank et al., 2013]	Phantom	3	2D TSE sequence; 3D UTE sequence.	MAE∈[81; 95] HU
[Khalifé et al., 2017]	Brain	16	ZTE MRI sequence.	Correlation=0.92
Su et al. [2019]	Thorax	25	UTE-mDixon MRI sequence.	MAE_{Body}<50 HU MAE_{Bone}~100 HU ME_{Body}~−10 HU ME_{Bone}~120 HU $Correlation_{Body}$~0.95 $Correlation_{Bone}$~0.8

Table 3.1: Segmentation-based approaches for generating pCT images from MRI data.

Work	Region	Subjects	MRI sequences	Reported results
[Arabi et al., 2016]	Pelvis	12	Dixon volumetric interpolated T1-w MRI sequence.	ME_{Fat}± SD=2.2 ± 5.5 HU $ME_{Soft tissue}$± SD=−2.0 ± 4.1 HU ME_{Bone}± SD= −1.5±5.0 HU MAE_{Fat}± SD=5.6 ± 4.5 HU $MAE_{Soft tissue}$± SD=5.0 ± 3.9 HU MAE_{Bone}± SD= 4.1±2.2 HU
[Burgos et al., 2014]	Brain	53	T1-w MRI sequence.	MAE±SD=121±17 HU ME±SD=-7.3±23 HU
[Dowling et al., 2012]	Pelvis	37	T2 fast spin echo FSE sequence; T2* gradient echo sequence; Whole pelvic fast spin echo FSE sequence.	MAE_{Rectum}±SD= 54 ±143 HU $MAE_{Bladder}$±SD=9 ±6 HU MAE_{Bone}±SD=340 ±85 HU $MAE_{Prostate}$±SD=42 ±25 HU
[Gudur et al., 2014]	Brain	9	T1-w rapid gradient echo sequence.	MAE± SD=126±25 HU
[Uh et al., 2014]	Brain	26	T2-w MRI sequence prior to contrast agent injection.	$RMSD_{MEAN6}$±SD=224±36 HU $RMSD_{MEAN12}$±SD=219±35 HU $RMSD_{PRGP12}$±SD= 207±33 HU $Correlation_{MEAN6}$±SD=0.787±0.06 $Correlation_{MEAN12}$±SD=0.798±0.059 $Correlation_{PRGP12}$±SD=0.819±0.058
[Kops and Herzog, 2007]	Brain	4	High-resolution 3D T1 w volumetric MRI sequence.	-
[Schreibmann et al., 2010]	Brain	17	-	Mean distance $differences_{External\ contour}$=1,26 mm Mean distance $differences_{Bone}$=2.15 mm MAE< 2 HU
[Greer et al., 2011]	Pelvis	39	T2-w MRI sequence	-
[Mérida et al., 2015]	Brain	27	T1-w MRI sequence.	Classification error±SD =7.81% ± 1.00%
[Kraus et al., 2017]	Phantom	3	T2-w MRI sequence.	MAE±SD=31.3±27.3 HU to 37.3±35.8 HU
[Sjölund et al., 2015]	Brain	10	T1-w 3D spoiled gradient recalled echo sequence.	MAE ± SD=113.4±17.6 HU
[Ren et al., 2017]	Brain	10	T1-w and T2-w MRI sequences.	MAE=119 HU

Table 3.2: Atlas-based approaches for generating pCT images from MRI data.

Work	Region	Subjects	MRI sequences	Reported results
[Zhong et al., 2016]	Brain	13	T1-w and T2-w MRI sequences.	MAE±SD=80.80±22.33 HU PSNR±SD=30.30±1.91 dB
[Huynh et al., 2016]	Brain	16	/	MAE±SD=99.9±14.2 HU PSNR±SD=26.3±1.4 dB Classification accuracy= 91%
	Prostate	22	-	MAE±SD=48.1±4.6 HU PSNR±SD=32.1±0.9 dB Classification accuracy=79%
[Roy et al., 2014]	Brain	5	UTE sequence.	Correlation±SD=0.79±0.13 PSNR±SD=21.92±1.07 dB
[Johansson et al., 2013]	Brain	5	UTE sequence T2 w 3D SPACE sequence.	MAE=137 HU
[Navalpakkam et al., 2013]	Brain	10	Anatomical T1-w TSE sequence; T2-TSE sequences; Echo-planar-diffusion weighted sequence. 3D Dixon-VIBE sequence;	MAE±SD in PET quantification 2.4%±3.86%
[Kapanen and Tenhunen, 2013]	Pelvic bone	20	T1/T2*-w (spoiled steady state) 3D GE sequence LAVA Flex.	Correlation=0.913 MAE±SD=135±121 HU Max differences=578 HU
[Kim et al., 2015]	Pelvis	9	3D T1- w FFE; 3D T2-w TSE; 3D bTFE.	MAE±SD=74±10.9 HU
[Korhonen et al., 2014]	Pelvis	20	T1/T2*-w in-phase MRI obtained with a 3D fast RF-spoiled dual gradient echo sequence (LAVA FlexR).	Average soft tissue ME range=-2 to 5 HU Average bone ME range=22 to 87 HU
[Hu and Zhang, 2018]	Brain	-	T1-w MRI sequences	PSNR∈[42.06;43.84] dB SSIM∈[0.91;0.94]
[Yang et al., 2017]	Brain	5	T1-w MRI sequences	PSNR±SD=27.27±1 dB
[Ghose et al., 2017b]	Pelvis	39	T1-w, T2-w and T2* MRI sequences	MAE±SD=108±23 HU DSC±SD=0.92±0.04
[Han, 2017]	Brain	18	T1-w 3D spoiled gradient recalled echo	MAE±SD=84.4±17.3 HU Correlation=0.906±0.03
[Liu et al., 2017a]	Brain	40	T1-w MRI sequences	DSC_{Air}±SD=0.971±0.005 $DSC_{Soft-tissue}$±SD=0.936±0.0111 DSC_{Bone}±SD=0.803±0.021
[Leynes et al., 2017b]	Pelvis	26	ZTE Dixon fate and water images	ME±SD=-12±78 HU
[Largent et al., 2017]	Pelvis	10	T2-w isotropic SPACE MRI sequences	MAE=45.79 HU ME=9.11 HU
[Fu et al., 2018]	Pelvis	20	T1-w MRI sequences	MAE±SD=40.5±5.4 DSC±SD=0.81±0.04
				MAE±SD=37.6±5.1 DSC±SD=0.82±0.04
[Xiang et al., 2018]	Pelvis	22	T1-w MRI sequences	MAE±SD=42.5±3.1 HU PSNR±SD=33.5±0.8 dB
	Brain	16		MAE±SD=85.4±9.24 PSNR±SD=27.3±1.1
[Dinkla et al., 2018]	Brain	52	T1-w gradient echo MRI sequences	MAE±SD=67±11 HU ME±SD=13±9 HU
[Emami et al., 2018]	Brain	15	T1-w MRI sequences	MAE±SD=89.3±10.3 HU PSNR±SD=26.6±1.2 dB SSIM±SD=0.83±0.03
Lei et al. [2019b]	Brain	14	T1-w MRI sequences	MAE±SD=57.45±8.45 HU PSNR±SD=28.33±1.68 dB Correlation±SD=0.97±0.01
Spadea et al. [2019]	Head	15	T1-w MRI sequences	MAE±SD=54±7 HU ME±SD=-4±17 HU DSC_{Air}±SD=0.92±0.03 DSC_{Bone}±SD=0.93±0.02
Kazemifar et al. [2019]	Brain	77	T1-w MRI sequences	MAE±SD=47.2±11 HU DSC_{Bone}±SD=0.8±0.06

Table 3.3: Learning-based approaches for generating pCT images from MRI data.

Work	Region	Subjects	MRI sequences	Reported results
[Chen et al., 2014]	Brain	10	T1-w MRI sequence.	-
[Siversson et al., 2015]	Pelvis	10	T2-w MRI sequence;	MAE±SD=36±4.1 HU
[Wu et al., 2015]	Brain	13	T1-w and T2-w MRI sequences.	MAE_{T1}±SD= 124.6 ±14.2 HU MAE_{T2}±SD=123.9 ± 17.1 HU MAE_{T1_T2}±SD=113.8 ± 16.8 HU
[Demol et al., 2016]	Brain	22	T1-w MRI sequence.	MAE=150 HU
[Burgos et al., 2017]	Pelvis	15	T1-and T2-w MRI sequences.	MAE±SD=45.7±4.6 HU
[Andreasen et al., 2015]	Brain	5	T1-w FFE MRI sequence; Two dUTE MRI sequences.	MAE±SD=85±14 HU ME±SD=1 ±14 HU
[Hofmann et al., 2008]	Brain	17	T1-w SE MRI sequence.	Correlation±SD=0.968±0.011
[Boydev et al., 2017]	Brain	3	ZTE MRI sequence.	$MAE_{Bone} \in$[275;800] HU $MAE_{Soft\,tissue} \in$[180;380] HU $MAE_{Air} \in$[300;400] HU
[Aouadi et al., 2017]	Brain	13	T1-w T2-w MRI sequences.	MAE±SD=99.69±11.07 HU DSC±SD=0.83±0.03
[Leynes et al., 2017a]	Pelvis	6	Dixon ZTE MRI sequences.	-
[Speier et al., 2018]	Brain	15	T1-w T2-w MRI sequences	MAE_{LUT}±SD=118.7±10.4 HU ME_{LUT}±SD=-3.3±23.3 HU DSC_{LUT}±SD=0.83±0.02 MAE_{Patch}±SD=73±6.4 HU ME_{Patch}±SD=-2.4±17.6 HU DSC_{Patch}±SD=0.89±0.02
[Wiesinger et al., 2016]	Brain	5	ZTE MRI sequences	MAE±SD=123±25 HU DSC±SD=0.73±0.08

Table 3.4: Hybrid approaches for generating pCT images from MRI data.

From the tables, it is clear that most of the research focused primarily on the brain and secondly on the pelvis whereas poor work concentrated on the liver or other body parts. We notice a recent increasing interest in the Dixon MRI sequence since this particular type of sequences is advantageous by producing multiple contrast images within a single acquisition, and this can be of a great interest to be employed in a segmentation-based workflow.

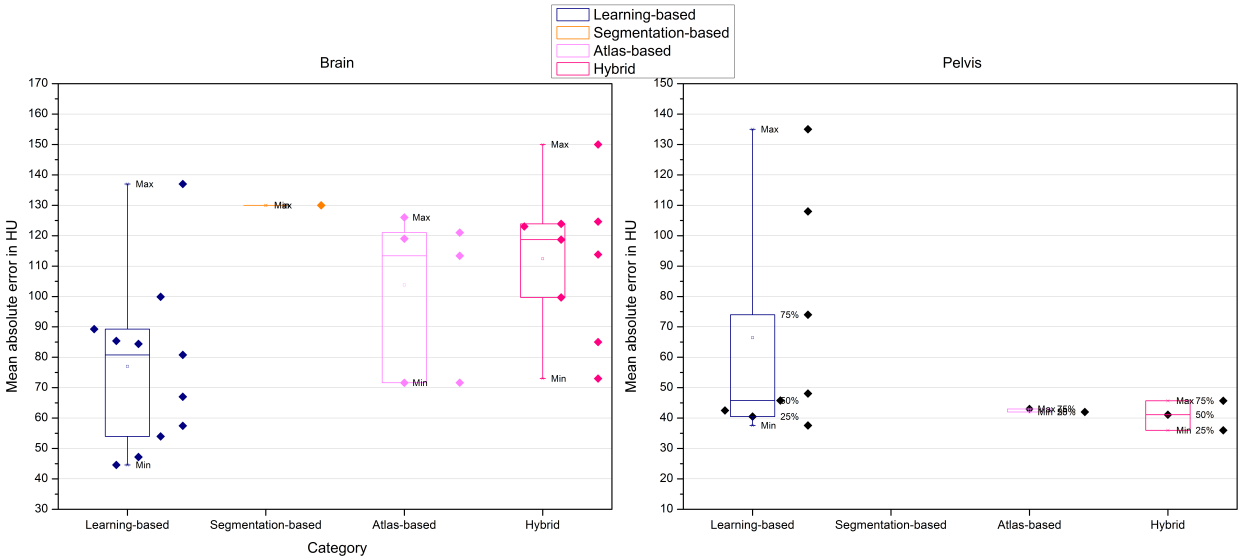

Figure 3.5.1: MAE distribution for each category of approaches considering the brain and the pelvis.

Fig. 3.5.1 shows box and whiskers plots for the MAE distribution for each category of approaches considering the brain and pelvis regions. For the brain, MAE ranged from 44.58 HU to 137 HU, from 71.66 HU to 126 HU and from 73 HU to 150 HU for the learning-based, atlas-based and hybrid approaches, respectively showing that methods relying on machine learning and the hybrid ones produced more accurate pCT images for the brain. For the pelvis, the lowest MAE was scored by a hybrid approach with a value of 36 HU. Learning-based methods produced variable MAE results and this is due to the complex and the uncertain nature of the pelvis from one patient to another making the construction of an accurate pCT image estimation model challenging.

Despite the clinically acceptable performance results of these approaches, each class of methods presents some drawbacks and limitations that should be taken into consideration and could extend the field of research in MRI-only RT. Approaches that are based on using segmentation algorithms suffer from segmentation errors, which can result in a misclassification of voxels. Moreover, bone segmentation is a serious issue because bone and air reflect no signal from standard MRI sequences while they have different HU values on CT images (<-1000 HU for

air and >600 HU for bone). Some research works have investigated the use of other specialized MRI sequences that are capable of detecting signals from the bone using two times of echo after the RF excitation. This type of sequences is the UTE sequence. Another specialized MRI sequence for detecting the bone signal is ZTE [Delso et al., 2015, Wiesinger et al., 2016]. [Delso et al., 2015] showed that the bone segmentation using ZTE sequences performed better than the UTE-based segmentation. Likewise, [Khalifé et al., 2017] attempted to establish a linear relationship between histogram-normalized ZTE intensity for bone and bone normalized CT density. The coefficient of determination was 0.92 indicating a strong correlation. However, ZTE and UTE sequences are still in development/research stages and they are less employed in clinical practice due to their expensive implementation requirements. Furthermore, whole body UTE imaging results in increased noise [Kim et al., 2012].

Other attempts to detect the bone combined short-TE sequences with Dixon pulse sequences in a fuzzy framework to determine tissue classes including cortical bone . Techniques incorporating machine learning schemes to segment the bone seem very promising. For instance, [Bredfeldt et al., 2017] segmented the contours of the vertebra bone from fat images using a learning approach [Peng et al., 2006, Huang et al., 2009] and [Liu et al., 2017b] applied a bone shape model to detect the pelvic bone; in addition, femur bone was segmented using 3D connected component analysis.

Learning-based approaches attempt to establish a link between MRI voxel values and HU values. However, characterising such a link is not straightforward [Ito et al., 2017]. This is because tissues like air and bone have different CT values whereas they present no signal on MR images [Ito et al., 2017]. Nonetheless, with the increasing development of machine learning algorithms and their success proven in various application fields, machine learning models have demonstrated to effectively define complex mappings between multi-source data as in the case of MRI-only RT. Indeed, as shown in Fig. 3.5.1, the lowest MAE values were scored by approaches belonging to the learning-based category compared to MAE values scored by other approaches of the remaining categories, proving that machine learning algorithms can produce flexible prediction models regardless of data complexity. In addition, good results were observed by ensemble methods and deep learning methods showing that using advanced learning models can have a great impact in improving prediction quality. Yet, machine learning algorithms are prone to over-fitting and bias, therefore they can result in an undesirable behaviour of the prediction model. In addition, with the increasing complexity of these models, more training data is required, thus training time can increase significantly. For instance, the proposed CNN model by [Xiang et al., 2018] required 2-3 days for training. [Liu et al., 2017a] and [Dinkla et al., 2018] reported a training time of 34 hours and 30 hours, respectively, whereas other models required less with a training time ranging from 4 to 11 hours [Fu et al., 2018, Leynes

et al., 2017b, Emami et al., 2018, Maspero et al., 2018].

Approaches relying on deformable image registration algorithms produce naturally looking pCT images and make the success of pCT image generation related to the accuracy of image registration. In fact, inaccurate image registration introduces geometric uncertainties due to inter-patient variations and abnormalities. To assess the quality of deformable image registration, [Schreibmann et al., 2010] developed a 3D surface comparison tool that extracts a specific surface from the pCT and the real CT images to compare the differences. [Demol et al., 2016] studied a specific case in detail with the aim of testing the performance of their proposed hybrid method with atypical body structures. The studied case had a part of the skull bone removed surgically. The generated pCT image presented additional bone in the location where the patient's MR image did not have any. Furthermore, the number of atlas images used to generate the pCT image should be carefully selected [Aljabar et al., 2009] and the acquired atlas data set should be representative of standard anatomy. The use of multi-atlas fusion techniques appears to give improved results compared to the single atlas technique since the pCT image values will be predicted from several CT atlas images. [Mérida et al., 2015] evaluated four multi-atlas methods for pCT image generation including a proposed maximum probability approach and three multi-Atlas approaches. Comparison results showed that using an atlas set composed of ten MR-CT pairs outperforms the other methods producing the lowest voxel classification error 7.57 ± 1.05 % compared to 7.95 ± 1.00 %, 7.81 ± 1.00 % and 7.69 ± 1.04 %.

One concern that should be considered is geometric distortions related to non-linearity in magnetic field. The quality of the generated pCT image could be negatively affected if the used MRI sequences contain distortions. Many research papers presented techniques to correct for geometric distortion [Korsholm et al., 2014, Walker et al., 2014]. Overall, The reported results for most of the above-mentioned research works are in acceptable ranges and confirm the success of pCT image in replacing the original CT in RT. therefore, this field of research is emerging toward the use of MRI as the only modality in RT.

To sum up, we plotted the number of published papers each year for each category of methods for generating pCT images from MRI data and results are illustrated in Fig. 3.5.2.

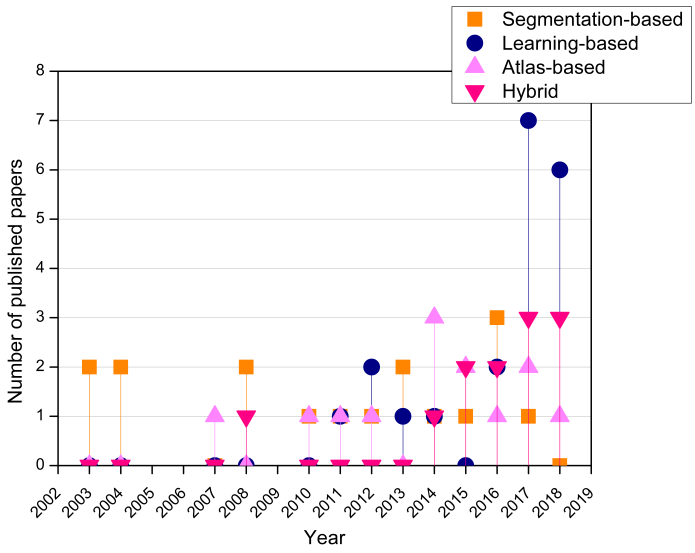

Figure 3.5.2: Number of published papers each year for each category of methods.

It can be seen that segmentation-based methods were intuitively the first ones to be implemented following the manual segmentation methods that were considered the first attempts for establishing an MRI-only workflow. However, their use decreased with the appearance of atlas methods. In addition, it has become more suitable to include image segmentation as a step in pCT image estimation workflow rather than using it for the entire process [Leynes et al., 2017a]. It can be seen that machine learning and hybrid methods have not been well established in the field of MRI-only RT until the beginning of 2016 depicted by an increasing number of published papers. This is owed to the success of these approaches in reducing prediction error and estimating reliable pCT images.

3.6 Conclusion

This chapter aimed to review approaches for pCT image generation from MR images for an MRI-only RT workflow. We presented a classification of these approaches, discussed their strengths and weaknesses. Lately, research is being oriented oriented toward including specialised MRI sequences such as UTE and ZTE despite their complex implementation. Furthermore, new approaches and techniques are being integrated to deal with the bone visualisation issues. We also notice that learning-based and hybrid approaches are able to produce accurate and reliable pCT images compared to the other methods. We conclude that the quality of the generated pCT image depends strongly on the applied approach and the incorporated MRI

sequences. Although many approaches seem promising, one cannot choose the best method compared to the others; this is because different MRI sequences, parameters, data sets and test metrics are used to generate the pCT image and assess its quality. Hence, a benchmark is strongly needed to set in-common performance metrics.

According to the comparative analysis performed in this literature review, one of the main conclusions that one may draw is the remarkable success of the learning-based and hybrid methods for pCT image synthesis from MRI data, which was confirmed by the recent increase in the number of published papers focussing on these two types of methods. In addition, these methods have shown a significant decrease in the prediction error whether for the brain or the pelvis regions. In the light of these conclusions, we propose two approaches to synthesise a pCT image from MRI images for brain cancer subjects. The first contribution is a hybrid method that combines multi-metric and multi-resolution deformable images registration, unsupervised learning for tissue classification, image fusion and a new hybrid correction technique and the second contribution is based entirely on machine learning and inspired from the recent works in ensemble methods.

Part II

Proposed approaches for pCT image synthesis from MRI data

CHAPTER 4

A hybrid method for pCT image estimation

4.1 Introduction

Recently, hybrid methods for pCT image estimation have shown to take a lead in producing reliable pCT images from several types of MRI sequences by combining methods and techniques belonging to the previously presented categories of approaches. Hybrid approaches aim to palliate for the limitations of those approaches when used alone, this is done in a way such that one component of a certain hybrid method corrects further HU values. This can result in a more consistent pCT image estimation method. Inspired by the recent hybrid works in the context of MRI-only RT, we present in this chapter a hybrid method that combines atlas registration, unsupervised learning, a hybrid HU number correction method and a multi-atlas image fusion technique. The presented hybrid method for pCT image estimation in this chapter is an extended version of a paper that we published in [Boukellouz et al., 2019].

One of the main components of the proposed hybrid method is deformable image registration; it plays a key-role in guiding the atlas-based approaches for pCT image estimation [Arabi et al., 2016, Demol et al., 2016, Ren et al., 2017, Sjölund et al., 2015]. Thus setting the right parameters for deformable image registration can highly improve prediction accuracy. One of the relevant parameters of image registration is the similarity measure, which is optimised during registration to achieve a maximised resemblance between the fixed image and the moving image. In order to choose the best parameters for deformable image registration intended for the proposed pCT image estimation hybrid approach in this chapter, we first evaluate the effect of the choice of similarity measure on mono-modal deformable image registration by examination of three common similarity measures and their combinations. After that and based on results of this similarity measure evaluation, we present the proposed contribution for hybrid pCT image estimation.

4.2 Assessing the impact of similarity measure choice on deformable image registration

Our concern is to evaluate the effect of the choice of similarity measures on the accuracy of deformable image registration. The evaluation and validation of deformable image registration is an ill-posed task since no standard validation methodology exists. We evaluated the effect of choice of the similarity measure on the performance of B-spline mono-modal deformable image registration. More precisely, we investigated on three well-known similarity measures that were previously presented in chapter two of this dissertation, namely: MI, MSD, and NCC. Furthermore, to investigate the performance of registration when using more than one similarity measure, we used three combinations of the aforementioned measures: MI-MSD, MI-NCC, and NCC-MSD by considering equal weights for each measure. The presented work in this section is an extension to a paper that we presented in [Boukellouz and Moussaoui, 2017b].

4.2.1 Data and pre-processing

Brain images from the Vanderbilt repository for six subjects were used in this work. 3D T1-w MR images were acquired with prior injection of contrast agent Gadolinium with an SE sequence with TE/TR = 15/800 ms. Voxel size ranged from $0.82\times0.82\times3$ mm^3 to $1.25\times1.25\times4$ mm^3 and number of slices varied from 26 to 52.

All MR images were resampled to the same resolution of $512\times512\times40$. T1-w MR images were bias corrected using the N4 bias-field correction method [Tustison and Gee, 2009] (Bspline order=3, number of iterations at each level=100, convergence threshold=0.001, shrink factor=3) followed by histogram matching to a random T1-w MRI in order to unify MR intensities across all subjects (levels= 1024, match points=7, threshold at mean intensity). To bring all images to the same coordinate system, T1-w MR images were aligned to one randomly chosen T1-w MR image using mutual information based multi-resolution affine registration [Klein et al., 2010].

4.2.2 Similarity measure based mono-modal deformable image registration

The six image volumes consisted of 240 2D slices that were used in a LOOCV strategy, meaning that one subject (40 slices) is considered as the fixed image and the rest of the five image volumes (200 slices) are considered as the moving images, which will be deformed in order to fit the fixed image. This process was repeated for each subject using all six similarity measures, resulting in 30 deformable registration tasks for each subject and 180 registrations for the entire data set.

For deformable image registration, we employed B-spline algorithm, which is based on using a grid of control points and the deformations are computed using B-Spline interpolation from the deformation values of points located in the grid. For our evaluation, the grid spacing is halved after every resolution, such that the final grid spacing is obtained in the last resolution level. We fixed the final grid spacing to 16 voxels. The grid spacing helps to constrain or loosen the registration, that is, a small grid spacing allows the registration to be loose and targets small complex deformations and a large grid spacing constrains the registration and helps to consider large deformations. Furthermore, we performed registration on four resolution levels, starting on a coarse level to register large global deformations and moving on to finer levels to register smaller local deformations. Optimization followed an adaptive stochastic gradient descent approach over 600 iterations. Registration parameters in this work other than the similarity measure were set based on a trial/error strategy for best parameters tuning.

4.2.3 Results

To evaluate registration results, we used three techniques; the most obvious one is visual inspection. Where anatomical differences between the moving image and the fixed image are visualized. For quantitative evaluation, we calculated the RMSD and correlation between the fixed and the moving images after registration to measure the resemblance between images. Note that all MR intensity values were normalised to the interval [0; 1] to account for any intensity differences upon evaluation.

Fig. 4.2.1 shows axial slices for subject #5 considered as the fixed image, other subjects considered as the moving image shown before applying image registration and after applying registration for each similarity measure employed. The rectangles placed over the images show the major observed differences between the fixed image and the registered moving images. The moving images were well-deformed to fit the fixed image which is the case when applying either of the six similarity measures. Subject #4 shows the largest differences typically in the lower right regions of the registered images exhibited by intensity sheering and slight tissue deformations.

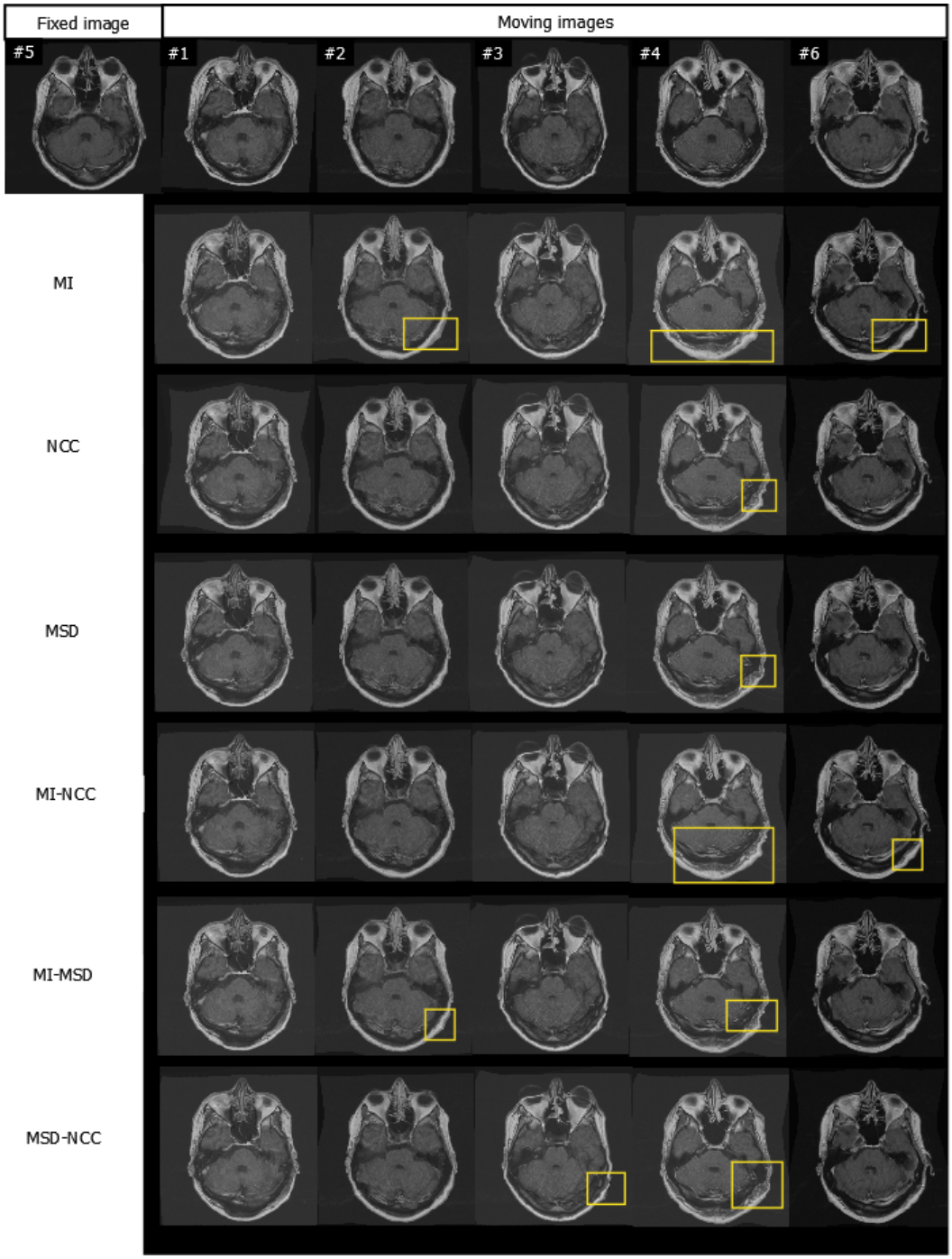

Figure 4.2.1: Axial slices of the target image for subject #5 and fixed images for the other subjects before (top row) and after deformable image registration considering all six similarity measures.

For quantitative evaluation, table 4.1 shows the RMSD results for each subject considering all registration tasks, in addition average RMSD±SD were computed considering all registration tasks for each similarity measure. RMSD ranged from 0.08 to 0.13; from 0.081 to 0.132; from 0.080 to 0.127; from 0.080 to 0.126; from 0.0824 to 0.114 and from 0.0807 to 0.125 for registration tasks based on MI, NCC, MSD, MI-MSD, MI-NCC and NCC-MSD, respectively; showing relatively close values. Similarly, average values are close and deformable registration based on the combined similarity measure MSD-NCC scored the lowest RMSD value with an RMSD±SD=0.103±0.012.

On the other hand, table 4.2 shows correlation results for all registration tasks with average± SD. Correlation ranged from 0.674 to 0.86; from 0.669 to 0.827 and from 0.669 to 0.828 for registration tasks based on MI, NCC and MSD, respectively . For combined similarity measure deformable registration, correlation ranged from 0.666 to 0.826; from 0.667 to 0.841 and from 0.669 to 0.825 when using MI-MSD, MI-NCC and NCC-MSD, respectively. Average values are close and deformable registration based on MSD similarity measure scored the lowest correlation value with an average correlation±SD=0.74±0.04.

Fixed image	Moving image						Average±SD
	#1	#2	#3	#4	#5	#6	
MI							
#1	-	0.093	0.125	0.134	0.115	0.111	0.107±0.015
#2	0.085	-	0.127	0.134	0.110	0.100	
#3	0.114	0.095	-	0.103	0.084	0.098	
#4	0.123	0.106	0.113	-	0.100	0.097	
#5	0.121	0.109	0.105	0.123	-	0.086	
#6	0.128	0.113	0.114	0.123	0.089	-	
NCC							
#1	-	0.096	0.120	0.123	0.103	0.102	0.105±0.013
#2	0.081	-	0.117	0.119	0.096	0.091	
#3	0.114	0.096	-	0.102	0.082	0.094	
#4	0.115	0.111	0.111	-	0.096	0.094	
#5	0.112	0.097	0.105	0.107	-	0.082	
#6	0.123	0.113	0.120	0.133	0.090	-	
MSD							
#1	-	0.096	0.120	0.127	0.105	0.101	0.104±0.013
#2	0.081	-	0.111	0.123	0.094	0.091	
#3	0.109	0.090	-	0.114	0.081	0.094	
#4	0.122	0.110	0.112	-	0.096	0.093	
#5	0.116	0.097	0.101	0.103	-	0.084	
#6	0.125	0.109	0.117	0.116	0.089	-	
MI-NCC							
#1	-	0.088	0.122	0.122	0.110	0.106	0.107±0.014
#2	0.082	-	0.111	0.133	0.100	0.095	
#3	0.114	0.097	-	0.108	0.084	0.095	
#4	0.121	0.113	0.113	-	0.099	0.095	
#5	0.116	0.098	0.106	0.115	-	0.084	
#6	0.127	0.111	0.121	0.125	0.093	-	
MI-MSD							
#1	-	0.090	0.121	0.125	0.110	0.102	0.104±0.013
#2	0.081	-	0.118	0.118	0.096	0.091	
#3	0.110	0.099	-	0.095	0.081	0.095	
#4	0.118	0.108	0.107	-	0.096	0.094	
#5	0.119	0.104	0.096	0.103	-	0.082	
#6	0.127	0.108	0.116	0.126	0.088	-	
MSD-NCC							
#1	-	0.090	0.112	0.116	0.103	0.101	0.103±0.012
#2	0.081	-	0.108	0.120	0.093	0.091	
#3	0.110	0.095	-	0.107	0.083	0.095	
#4	0.122	0.103	0.111	-	0.096	0.093	
#5	0.117	0.097	0.104	0.107	-	0.082	
#6	0.126	0.109	0.115	0.119	0.088	-	

Table 4.1: RMSD values for all deformable image registration tasks with the average and SD.

Fixed image	Moving image						Average±SD
	#1	#2	#3	#4	#5	#6	
MI							
#1	-	0.814	0.684	0.781	0.760	0.819	0.75±0.046
#2	0.810	-	0.747	0.785	0.792	0.860	
#3	0.709	0.753	-	0.709	0.795	0.715	
#4	0.701	0.726	0.674	-	0.717	0.741	
#5	0.683	0.732	0.792	0.740	-	0.776	
#6	0.676	0.734	0.726	0.765	0.781	-	
NCC							
#1	-	0.827	0.680	0.764	0.734	0.779	0.743±0.04
#2	0.823	-	0.739	0.743	0.764	0.818	
#3	0.728	0.769	-	0.697	0.784	0.690	
#4	0.734	0.699	0.669	-	0.742	0.721	
#5	0.712	0.751	0.781	0.730	-	0.759	
#6	0.702	0.748	0.704	0.746	0.761	-	
MSD							
#1	-	0.828	0.682	0.763	0.725	0.773	0.74±0.04
#2	0.823	-	0.737	0.740	0.756	0.819	
#3	0.725	0.767	-	0.698	0.783	0.689	
#4	0.717	0.739	0.669	-	0.697	0.728	
#5	0.701	0.750	0.780	0.715	-	0.759	
#6	0.691	0.742	0.706	0.738	0.759	-	
MI-NCC							
#1	-	0.824	0.684	0.771	0.754	0.802	0.747± 0.044
#2	0.820	-	0.742	0.760	0.789	0.842	
#3	0.725	0.765	-	0.698	0.789	0.694	
#4	0.714	0.739	0.668	-	0.706	0.741	
#5	0.694	0.743	0.789	0.738	-	0.764	
#6	0.689	0.740	0.708	0.757	0.770	-	
MI-MSD							
#1	-	0.825	0.680	0.763	0.735	0.780	0.743±0.042
#2	0.825	-	0.738	0.747	0.759	0.826	
#3	0.728	0.771	-	0.699	0.786	0.688	
#4	0.722	0.741	0.667	-	0.698	0.733	
#5	0.70	0.753	0.785	0.719	-	0.763	
#6	0.692	0.745	0.707	0.747	0.765	-	
MSD-NCC							
#1	-	0.826	0.683	0.762	0.731	0.774	0.741±0.040
#2	0.823	-	0.738	0.742	0.756	0.819	
#3	0.727	0.768	-	0.698	0.783	0.689	
#4	0.719	0.739	0.670	-	0.696	0.730	
#5	0.702	0.751	0.780	0.714	-	0.759	
#6	0.693	0.743	0.705	0.741	0.761	-	

Table 4.2: Correlation values for all deformable image registration tasks with the average and SD.

In addition, we plotted box and whiskers plots to visualise the RMSD and correlation distributions considering all deformable registration tasks, results are shown in Fig. 4.2.2.

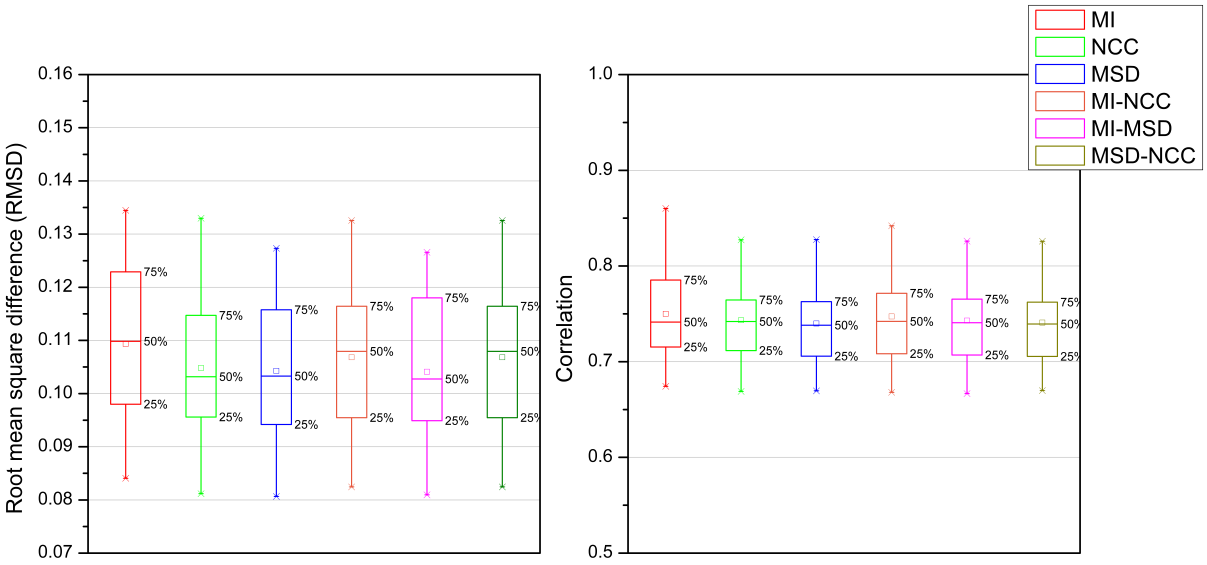

Figure 4.2.2: Box and whiskers plots for the distribution of RMSD and correlation for all the similarity measure based deformable image registration tasks.

The RMSD box and whiskers plot shows that MI registration scored a high percentage of errors over 0.11 compared to other similarity measure based registrations. NCC-based deformable registration shows to score more than 60% of the RMSD lower than 0.11. For correlation evaluation, MI seems to score the best results. All three combined similarity measure registrations show to have an equally similar distribution of the RMSD and correlation..

4.2.4 Discussion

This work presents an evaluation of B-spline deformable image registration based on different similarity measures for T1-w MR images of the brain. We conclude that there is no significant difference between all six measures. However, MI similarity measure showed better performance regarding the correlation between the registered and the fixed images. Furthermore, the MSD-NCC measure showed the minimal RMSD value. For multi-metric-based deformable registration, when setting MI and NCC as similarity measures for registering MR images, best results are obtained compared to combining other measures.

Due to the importance of image registration in the medical field, many works investigated the performance of registration when choosing different similarity metrics. One recent work of [García et al., 2017] that evaluated three similarity measures, namely: MI, sum of squared

distances and NCC for affine and B-spline registrations applied to breast density maps. Authors based their evaluation on the protocol of [Škerl et al., 2008]. Their results show that the NCC metric is a robust and stable metric to lead the affine registration while the sum of squared distances performed better for B-spline registration. For geometrical evaluation of image registration. [Schreibmann et al., 2010] presented a 3D comparison tool that extracts a specific surface from one CT image and another CT image to compare geometrical differences.

Overall, the choice of the similarity measure does not seem to affect strongly B-spline deformable image registration with the current registration parameters, thus other parameters should be investigated such as the transform, which is considered crucial for constraining the deformations applied to the moving images.

4.3 Proposed hybrid approach for pCT image estimation, correction and fusion

In this section, we estimate a pCT image from T2-w brain MR images combining image registration, unsupervised image segmentation, a 3D hybrid HU values correction technique that uses information from MR images and multi-atlas fusion. Firstly, multi-resolution and multi-metric deformable registration algorithm B-spline is employed to deform the atlas MR images onto the target MRI. The deformations resulting from this step are applied to the atlas CT images. Secondly, the target MR and the deformed atlas MR images are used as inputs for the unsupervised clustering FCM algorithm; the segmentation was used to support the 3D hybrid HU values correction mechanism. This mechanism uses information from the target MRI, deformed atlas MRI and tissue maps from the segmented MRI images in a 3D restricted search space considering grey-level information, spatial information and tissue class of the voxel to be corrected in order to select the most convenient and contributing MRI voxels with their corresponding CT voxels on the uncorrected pCT image. The selected CT voxels are combined using weighted average yielding the corrected pCT image. These corrected pCT images are further fused into a final pCT image either by taking their mean or median. Fig.4.3.1 presents a flowchart for the proposed approach.

Each element of the proposed approach was chosen for the following purposes:

1. Deformable image registration is widely used by the atlas-based methods and has shown to produce realistic pCT images with continues valued HU;

2. Unsupervised learning serves to provide a support for the proposed hybrid approach to produce accurate HU values and we characterise it by FCM clustering;

3. The proposed hybrid correction approach aims to correct for HU values using as a support the results of the unsupervised learning algorithm that we employ and MR intensity information from the subject of interest; this is because an additional information about tissue classes can be of a great importance in improving the final results. In addition, we integrate in the hybrid correction method a specific weighted average formula that combines spatial and intensity information to better emphasise on the contribution of each voxel in the correction of HU values;

4. Image fusion is employed as a decision fusion level since multiple decisions contribute to a better output rather than taking a single decision, we adopted two fusion techniques applied at the final stage of pCT image estimation namely: fusion by taking the mean and fusion by taking the median.

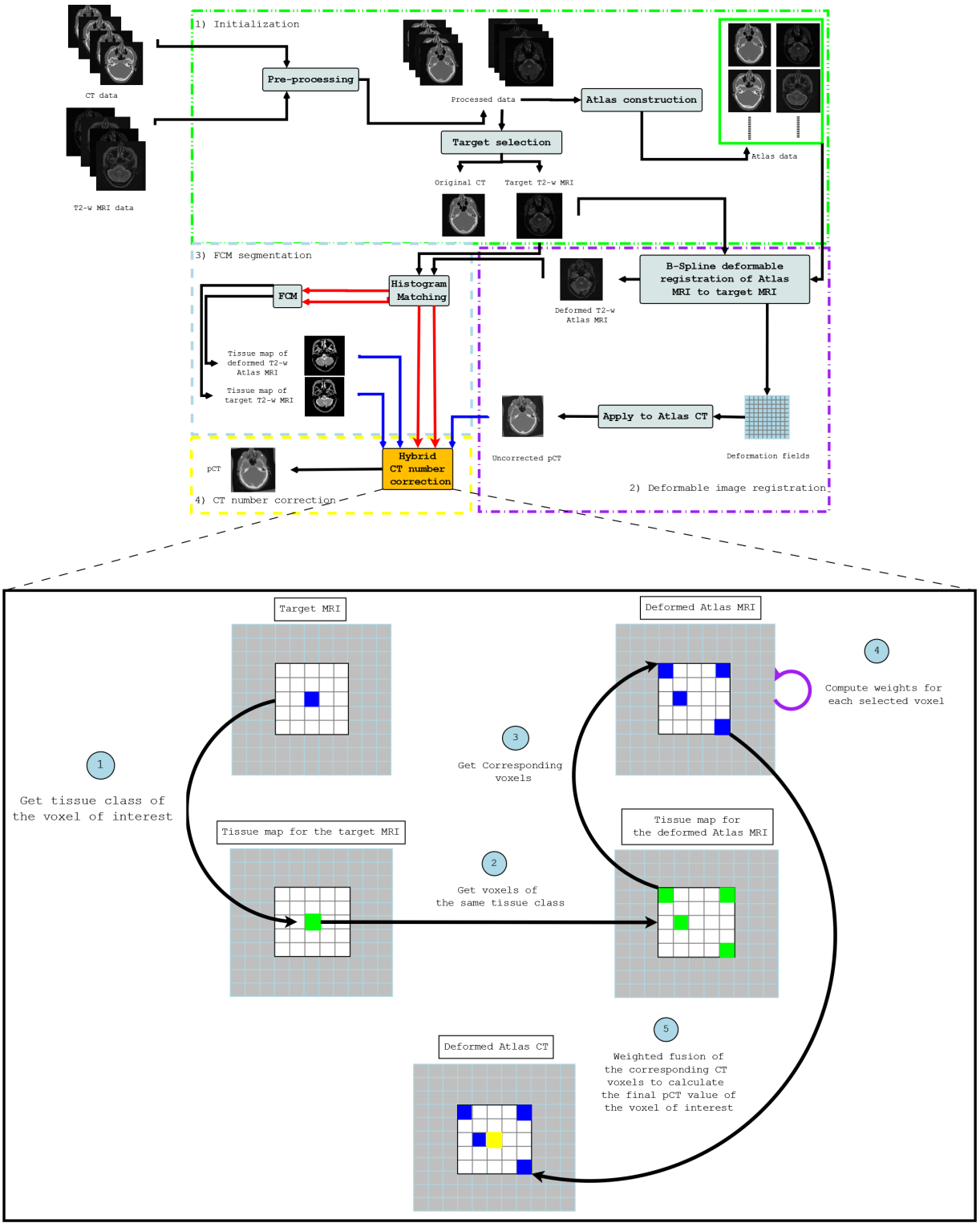

Figure 4.3.1: Steps for pCT image estimation: 1.Data initialisation; 2. Deformable image registration; 3. Unsupervised FCM clustering; 4. Hybrid CT number correction technique.

4.3.1 MRI and CT data

In this work, we used a set composed of eleven pairs of brain T2-w MRI sequences and their corresponding CT scans; acquired from the publicly available Vanderbilt database. 3D SE T2-w MRI sequences were acquired with prior injection of contrast agent Gadolinium; with a voxel size ranging from 0.82× 0.82× 4 mm^3 to 1.25× 1.25× 3 mm^3 , a matrix size of 256 × 256 × 26 or 256×256×52.

CT scans were acquired using a tube voltage of 120 kVp, exposure time of 2 ms, an exposure rate of 250, a voxel size ranging from 0.40× 0.40 × 3 mm^3 to 0.65× 0.65 × 4 mm^3 and a matrix size of 512 × 512 with number of slices varying from 28 to 49.

4.3.2 Data pre-processing and atlas database construction

At first, the imaging bed was removed from CT scans. MR and CT images were all resampled to the size of 512×512×40 and aligned together using MI-based affine registration. Later, MR images were pre-processed using the N4 bias-field correction method [Tustison and Gee, 2009] (Bspline order=3, number of iterations at each level=100, convergence threshold=0.001, shrink factor=3). This was followed by histogram matching to a randomly selected MR image (levels= 1024, match points=7, threshold at mean intensity).

MRI and CT data were divided into two sets; the first set contained one subject (40 slices) for which pCT images will be estimated and the second data set considered as the atlas set (400 slices) consisting of the remaining ten images was used to construct the pCT images in a LOOCV framework. Meaning that each subject will have ten uncorrected pCT images generated, these images will undergo the hybrid HU values correction step and will be fused into one pCT image using two fusion techniques. The result is 100 uncorrected pCT images and 20 corrected fused pCT images.

4.3.3 pCT image generation process

After pre-processing and atlas construction, the atlas and the target images underwent several steps in order to initialize for the hybrid HU values correction and estimate the final pCT images. For the rest of this chapter, the following notations are used:

$\mathbf{T}_{MRI}$	Target MRI
$\mathbf{A}_{def\_MRI}$	Deformed atlas MRI
$\mathbf{M}_{T}$	Tissue maps for the target MRI
$\mathbf{M}_{def\_A}$	Tissue maps for the deformed MRI
pCT_u	Uncorrected pCT image
pCT_h	Corrected pCT image
pCT_h_mn	Corrected fused pCT image using the mean
pCT_h_md	Corrected fused pCT image using the median

4.3.3.1 Deformable image registration

Firstly, the atlas MR images were registered to the target MRI with MI-based affine registration in order to achieve an initial alignment. Secondly, due to the large deformations present between subjects, a non-rigid registration was necessary. To do so, we opted for B-spline deformable registration [Rueckert et al., 1999] considering four resolution levels going from a grainy level to register large deformations and moving on to finer levels to register small deformations. Adaptive stochastic gradient descent was used for optimisation over 600 iterations. Furthermore, the deformable image registration was performed based on two similarity measures namely: MI and NCC. The two metrics were combined by multiplying each by a metric weight. The metric weight value was set to 0.5 for each metric as was presented in the privious section.

After deformable image registration, the subsequent deformations were applied to the atlas CT images in order to generate the uncorrected pCT_u images that will undergo hybrid HU values correction and fusion.

4.3.3.2 Unsupervised Fuzzy C-means clustering

Proposed methods to estimate a pCT image from MRI data belonging to the segmentation-based category [Khateri et al., 2015, 2014, Su et al., 2015] used FCM clustering mainly to segment MR images into several tissue classes and assigned each of them a discreet HU value. Inspired by these works, we introduce FCM clustering to the proposed hybrid version of pCT image estimation to support the process of HU values correction since an additional information about the tissue type can be of a great help to guide the pCT image correction process. For our implementation we set m=2, ε=10^{-5}, and C=3 considering lower intensity, mid intensity and higher intensity tissues. Before segmentation, deformed atlas MR images were histogram matched to the target MRI in order to correct for any intensity changes that could have occurred during the deformable image registration. The result of the FCM clustering is a tissue map that will be used to correct pCT image HU values.

4.3.3.3 Hybrid HU values correction

This step aims to correct HU values of the initially generated uncorrected pCT_u images. The correction step aims to incorporate MRI image information in the process of correcting HU values where MR intensity, tissue label and spatial information from the target and the atlas images are combined in a weighted average process within a pre-defined 3D search space. Fig. 4.3.1 (bottom) illustrates a simplified representation for the process of hybrid HU number correction in 2D mode.

To begin with, a 3D search box of size 5×5×5 (the size of the box was selected based on varying values from 2×2×2 to 10×10×10 and selecting the size that best improves HU values correctness) is centred over the voxel of interest in the target MRI, and the tissue class of the voxel under consideration is retrieved from its tissue map. All k voxels having the same tissue class as the voxel under consideration are selected within the 3D search box in the deformed atlas MR image, then a combined weight is calculated for the contribution of each voxel of the deformed atlas MR image. This combined weight considers the intensity and spatial information from the deformed atlas MR image and the target MRI. It is given by the following formulas:

$$W(j) = W^I(j) * W^L(j), \quad j = 1, ...k, \tag{4.3.1}$$

Where,

$$W^I(j) = \frac{1}{|T_{MRI}(i) - A_{def\_MRI}(j)|}, \quad j = 1, ...k, \tag{4.3.2}$$

and

$$W^L(j) = \frac{1}{\sqrt[2]{(x_{T_{MRI}} - x_{A_{def\_MRI}})^2 + (y_{T_{MRI}} - y_{A_{def\_MRI}})^2 + (z_{T_{MRI}} - z_{A_{def\_MRI}})^2}}, \; j = 1, ...k. \tag{4.3.3}$$

Where, i is the voxel under consideration in the target MR image T_{MRI}, $A_{def\_MRI}(j)$ is the intensity value at voxel number j from the selected k voxels from the deformed atlas MRI satisfying the tissue class condition, $W^I(j)$ is the inverse of the absolute intensity difference between the target voxel i and atlas voxel j, and $W^L(j)$ is the inverse of the distance between the selected voxels j and i.

Formula (4.3.1) is an adaptation of the formula presented in [Boydev et al., 2017, Demol et al., 2016](see chapter three for details) to calculate the weight for each voxel under consideration. Here, we avoid using a threshold in the search process since selecting voxels of the same tissue class allows to limit the correction to only voxels within the same intensity range.

The final HU value of the voxel under consideration is computed by selecting the corresponding CT voxels from the uncorrected pCT_u image that are combined following a weighted average technique:

$$pCT\_h(i) = \sum_{j=1}^{k} W(j) * pCT\_u(j), \quad j = 1, ...k. \tag{4.3.4}$$

The following algorithm summarises the process of the hybrid CT number correction technique.

Algorithme 4.1 The hybrid HU number correction technique.

Inputs: T_{MRI}, A_{def_MRI}, M_T, M_{def_A}, pCT_u;
Output: pCT_h;
Start
  Foreach Voxel i in T_{MRI} , do
    Insert a box of size 5×5×5 over i;
    c← $M_T(i)$;
    Foreach Voxel a in A_{def_MRI}, do
      If ($M_{def\_A}(a)$=c) then
        Calculate W(a);
        $Weight\_set = Weight\_set \bigcup (W(a))$;
        $Voxel\_set = Voxel\_set \bigcup pCT\_u((a))$;
        $k = k + 1$;
      endif;
    Endforeach;
    For j=1 to k, do
      $S = S + (pCT\_u(j) * Weight\_set(j))$;
    endfor;
    $pCT\_h(a) = S/Sum(Weight\_set)$;
    k=0;
    $S = 0$;
  Endforeach;
End;

4.3.3.4 pCT images generated through atlas fusion

After correction of the pCT images for each subject and in order to generate the final pCT images, each set of ten corrected pCT_h for each subject were fused into one pCT image either by taking their median forming pCT_h_md or their mean forming pCT_h_mn.

4.3.4 Results

In the current subsection, we compare results for the generated pCT images using the proposed approach to the uncorrected pCT images. Fig. 4.3.2 shows an axial slice of the target MRI for subject #10 and its tissue map. The figure also shows axial slices of the atlas MR images with their deformed version after registration to the target, as well as their corresponding tissue maps. Visually, the atlas T2-w MR images were well-deformed to correspond to the target and they look very similar to it in terms of anatomical correspondence. Segmentation of the MR images based on FCM algorithm resulted in three tissue classes: lower intensity, mid intensity and higher intensity tissues.

Fig. 4.3.3 shows axial, sagittal and coronal slices of the original CT for subject #10 and the estimated uncorrected pCT_u images for the same subject, and Fig. 4.3.4 shows axial, sagittal and coronal slices of the fused corrected pCT images for the eleven subjects considering the two fusion techniques compared to the real CT scan. The hybrid HU values correction approach with fusion produced images visually similar to the original CT scans where sinus regions were well reproduced in the fused pCT images better than their uncorrected unfused counterparts. However, the fused pCT images using the mean presented a visible blurring in all regions, which was not the case for the fused images using the median.

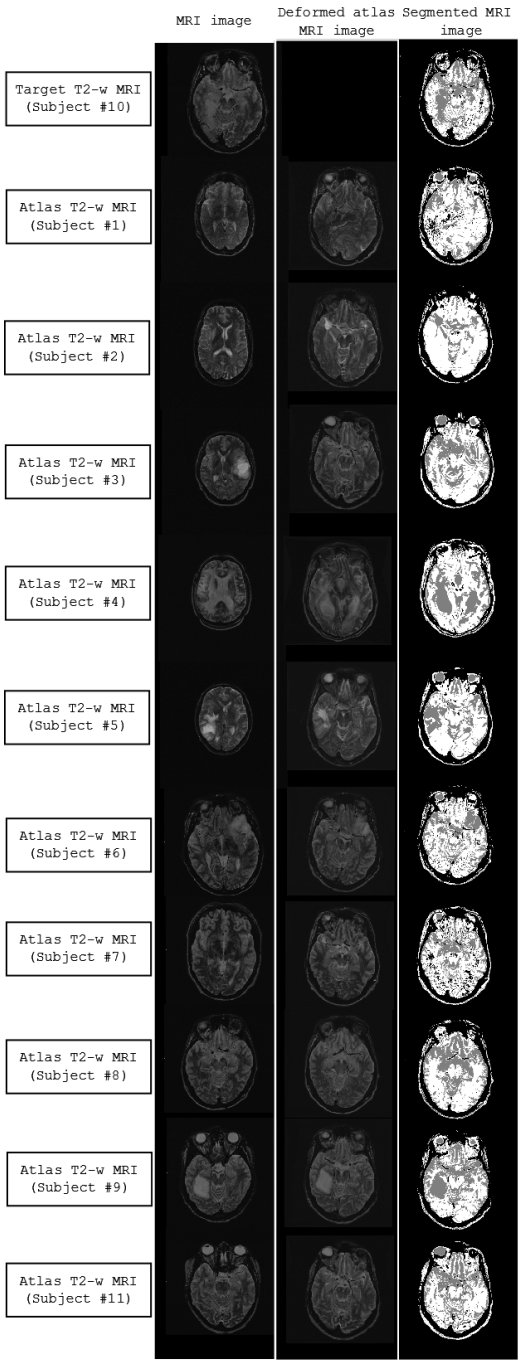

Figure 4.3.2 : Axial, sagittal and coronal slices of the target T2-w MR image for subject number 10, its tissue map, atlas MR images before and after registration with their corresponding tissue maps.

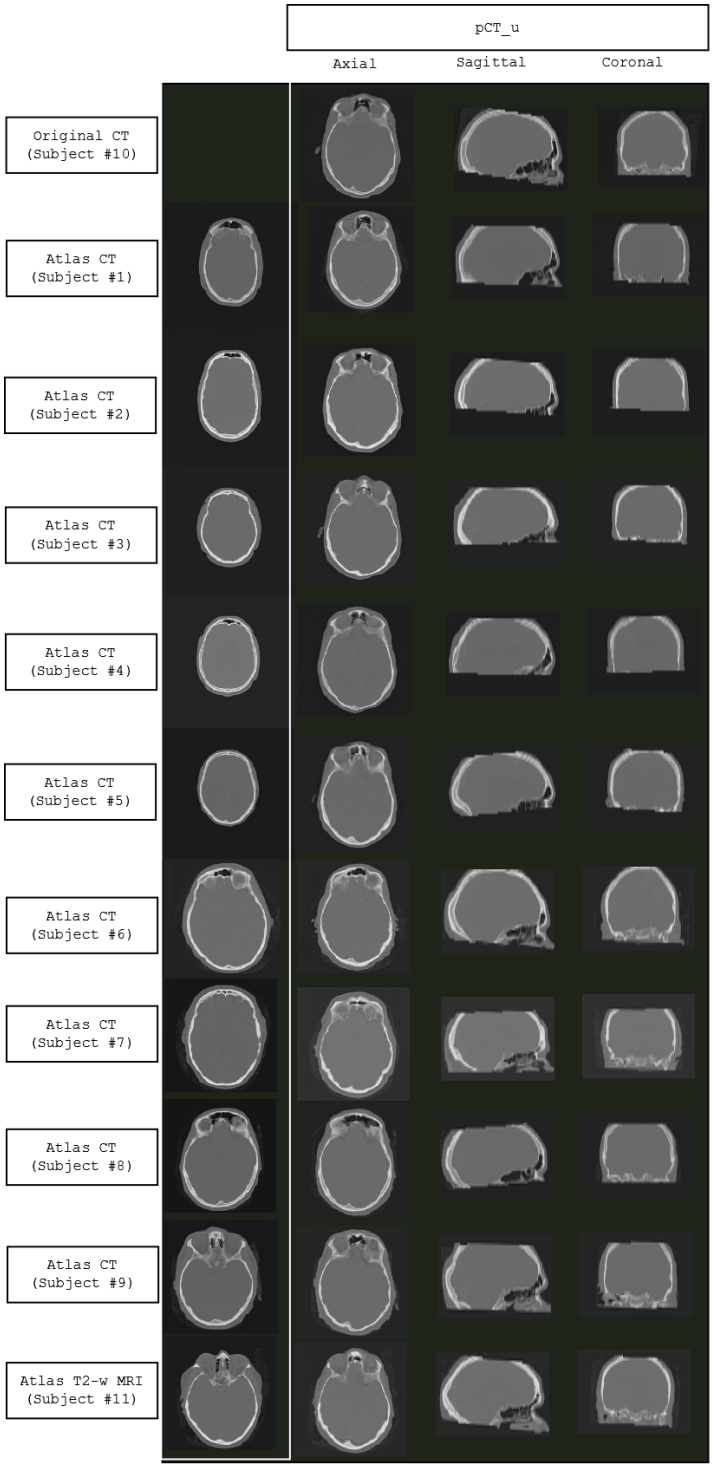

Figure 4.3.3: Axial, sagittal and coronal slices of the estimated uncorrected pCT_u images for subject #10 compared to its real CT scan.

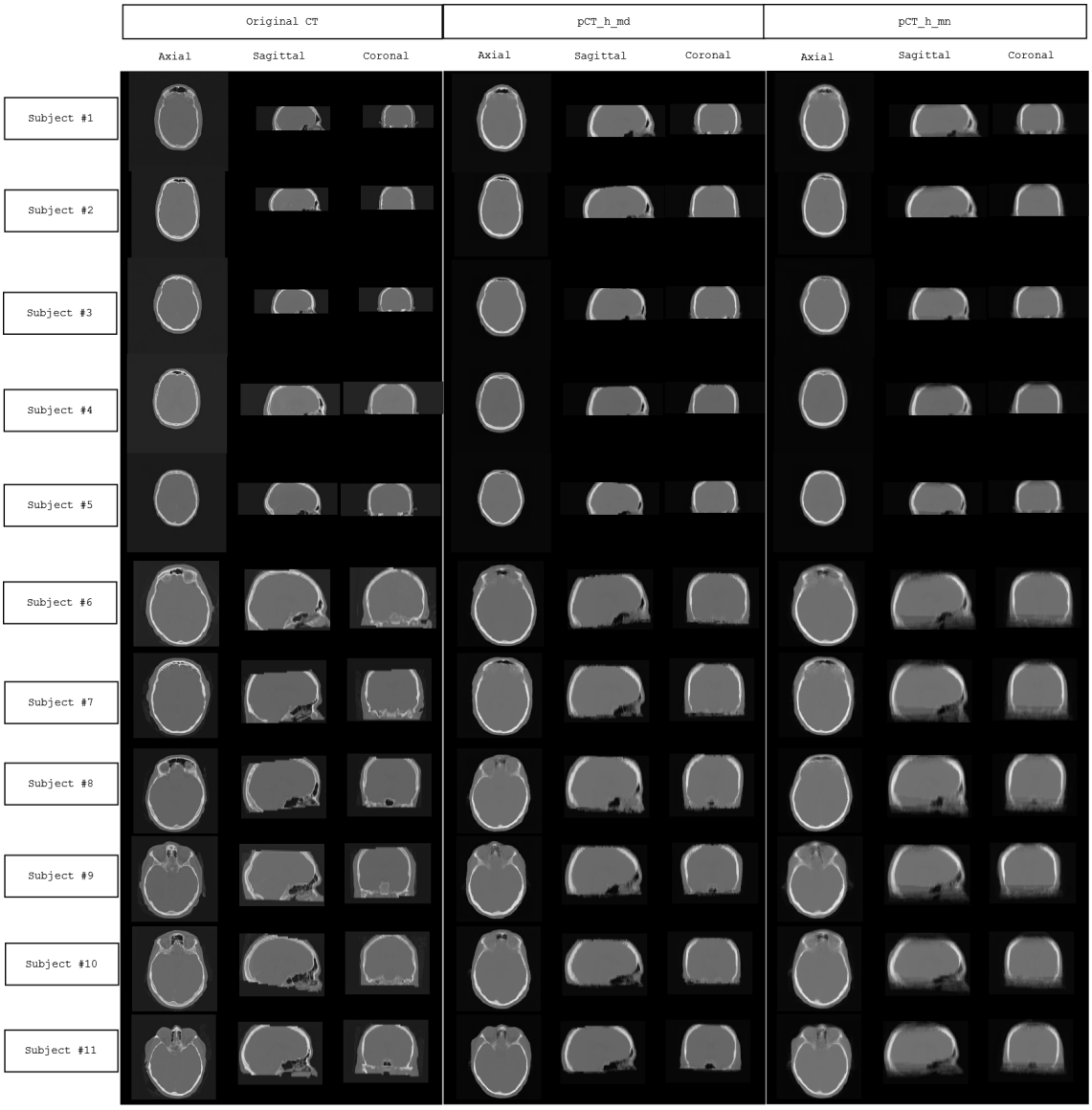

Figure 4.3.4: Axial, sagittal and coronal slices of the real CT scan and the estimated pCT images using the fusion of the corrected pCT images.

Fig. 4.3.5 and Fig. 4.3.6 show axial, sagittal and coronal slices for images that represent the differences between the original CT image and the generated pCT images for subject #10 and all subjects, respectively. Large differences are presented in orange colour. Most differences are located at air-bone interfaces in all pCT images. However, they are more pronounced in the uncorrected unfused pCT images and less present in the pCT images generated through the hybrid correction and fusion techniques. pCT images fused by taking their median seem to present less differences compared to those generated by taking their mean.

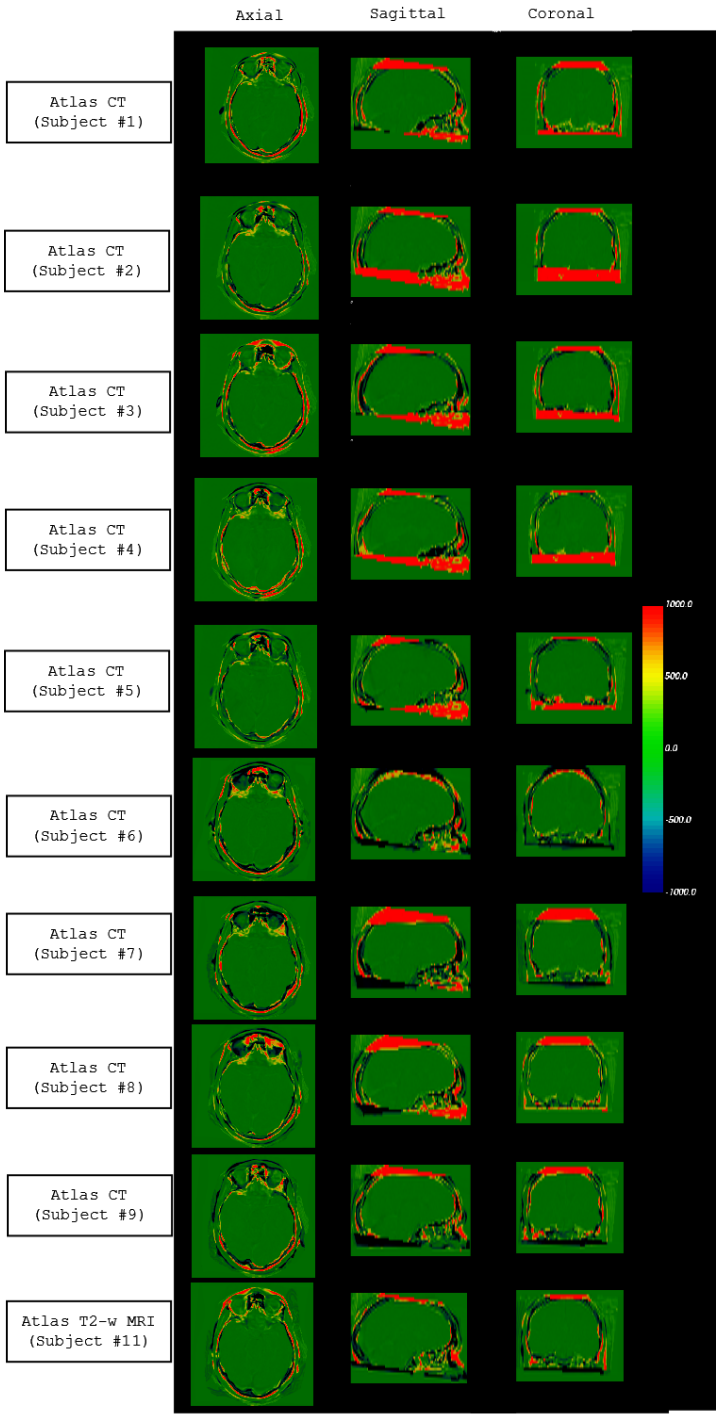

Figure 4.3.5: Axial, sagittal and coronal slices of the difference images for the uncorrected pCT images for subject #10.

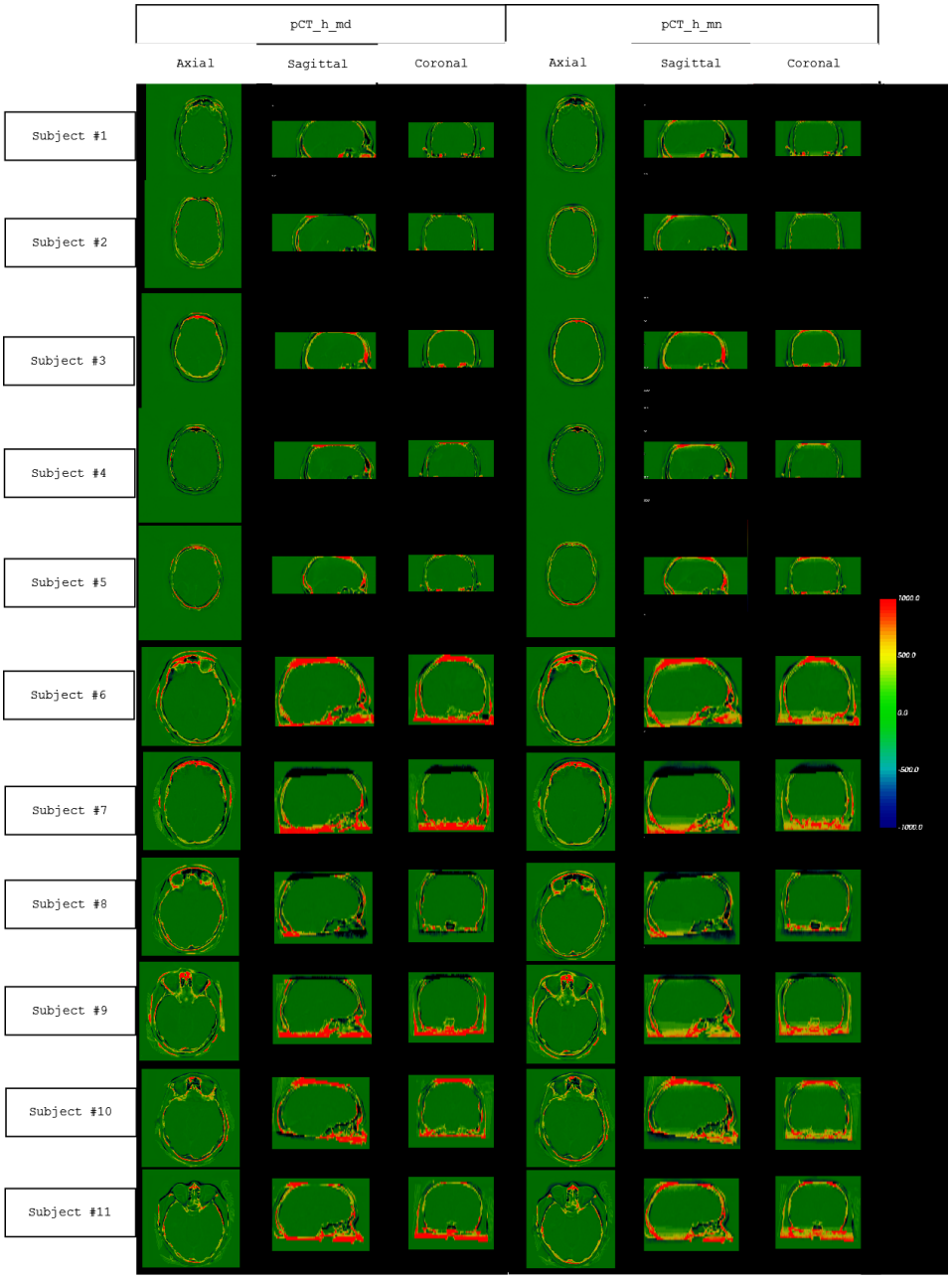

Figure 4.3.6: Axial, sagittal and coronal slices of the difference images for the fused corrected pCT images.

The edges of the estimated pCT images by all approaches presented large differences. In fact, these difference where due to imaging of the head, resulting in some images covering more of the head compared to others that covered less. Therefore, deformable image registration of one image covering less of the head to an image covering more of the head resulted in notable differences in the edges. This issue can be resolved by the use of images covering the same range of the head.

Fig. 4.3.8 shows MAE histograms for the uncorrected pCT_u images and the corrected fused pCT images (pCT_h_mn and pCT_h_md). Each MAE histogram compares differences averaged over the ten atlas images for each subject for the uncorrected pCT_u images against one subject representing corrected fused pCT images. In addition, Fig. 4.3.7 on top left shows the MAE histogram averaged over all generated pCT scans for all subjects considering the uncorrected unfused pCT_u images, corrected fused pCT_h_md images with the median and corrected fused pCT_h_mn images with the mean.

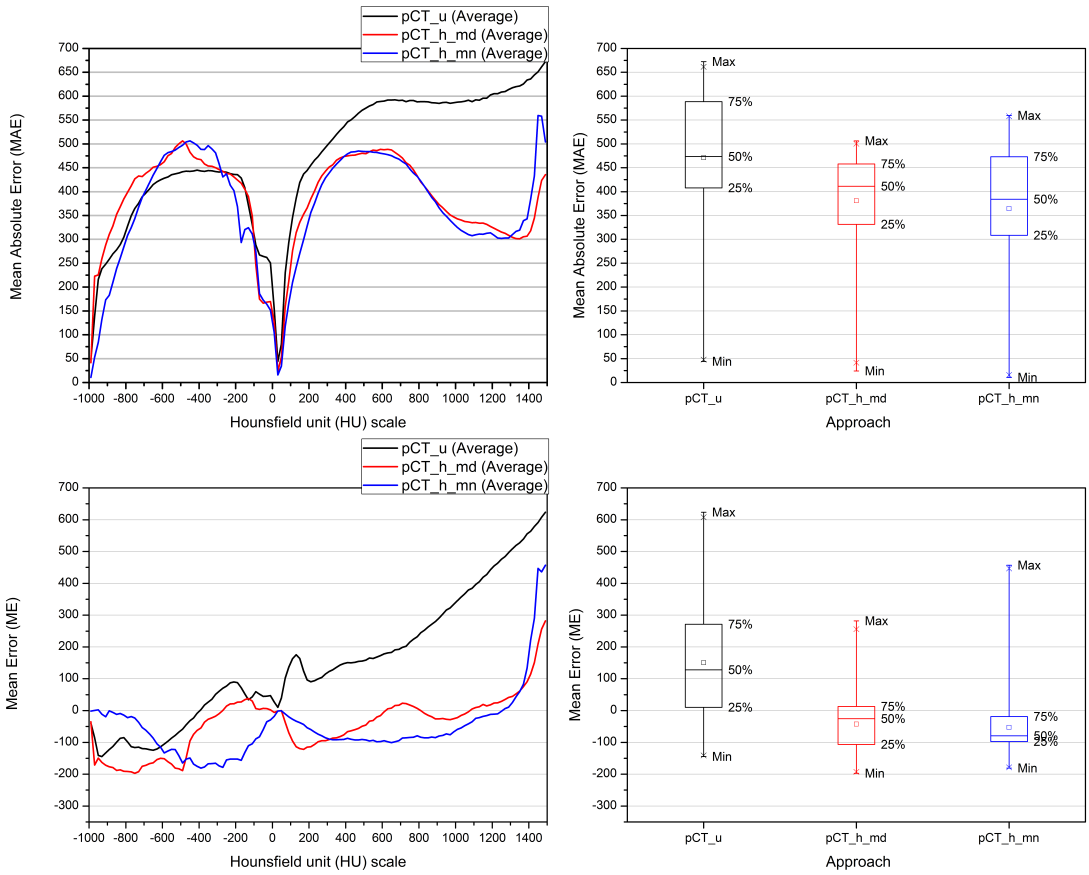

Figure 4.3.7: MAE (top left) and ME (bottom left) histograms all over tissue classes (HU scale) averaged over all subjects with box and whiskers plots for the MAE distribution (top right) and ME distribution (bottom right). The median is represented by a horizontal line inside the box , the maximum and minimum are located at the extremities of the whiskers and the mean is shown by □ symbol.

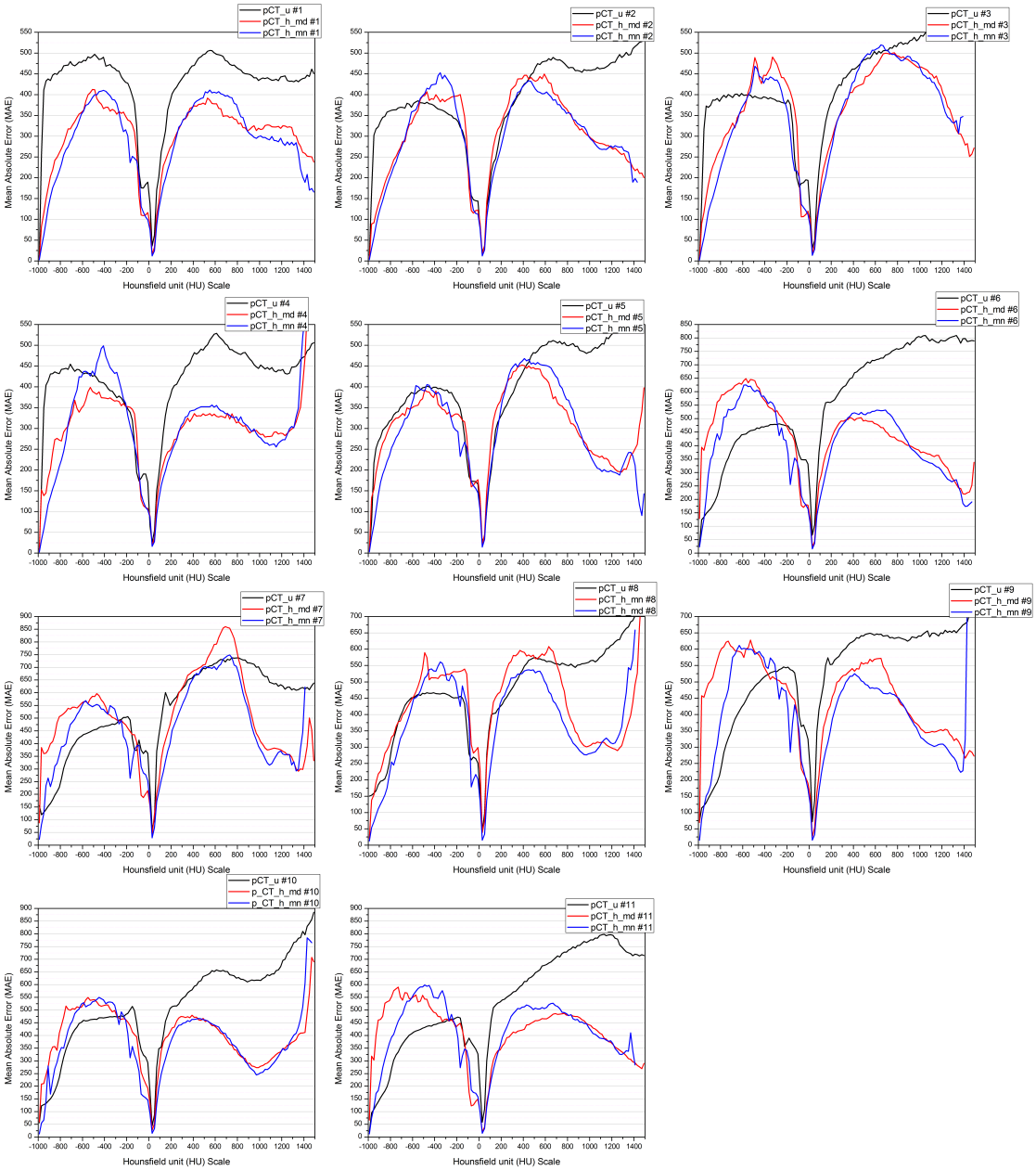

FIGURE 4.3.8 : MAE histogram all over tissue classes (HU scale) for each subject.

Results for the MAE are in similar ranges as reported in [Andreasen et al., 2015, Aouadi et al., 2017, Boydev et al., 2017, Demol et al., 2016]. The uncorrected pCT_u images and the fused corrected pCT images presented close errors reaching 500 HU in the interval of [-870; -380] HU which refers to the region of partial volume. However, MAE was higher than 650 HU for bone (bone>620 HU) in the uncorrected pCT images, while it varied between 300 HU and

450 HU for the proposed correction and fusion approach.

In addition, the top right of Fig. 4.3.7 shows box and whiskers plot for MAE distribution. From the one hand, the lower whisker shows that the minimum MAE is close to 25 HU, ~0 HU and ~50 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. On the other hand, the upper whisker shows that the maximum MAE is in the ranges of 500 HU, 550 HU and 650 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. The first quartile shows that 25% of MAE values are below 325 HU, 300 HU and 400 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. The third quartile describing 75% of MAE having values slightly above 450 HU, 425 HU and close to 600 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. The inter-quartile range represented by the box describes the 50% of MAE values falling between the first and third quartiles. 50% of MAE values fall inside the range of [325; 450] HU, [300; 475] HU and [400; 590] HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively..

By looking at the distribution of MAE values for the three approaches, we conclude that the proposed fusion and correction approach *pCT_h_md* outperformed the other two approaches where most of MAE values were low. Furthermore, discreet values for the MAE computed over the entire data set considering the whole volume were 102.21 HU, 84.70 HU and 78.13 HU for the uncorrected *pCT_u*, corrected fused *pCT_mn* and corrected fused *pCT_md* showing that the proposed correction and fusion approach using the median outperforms the other approaches.

Fig. 4.3.9 shows ME histograms for the uncorrected *pCT_u* images and the corrected fused pCT images (*pCT_h_mn* and *pCT_h_md*). Each ME histogram compares differences averaged over the ten atlas images for each subject for the uncorrected *pCT_u* images against one subject representing corrected fused pCT images. Fig. 4.3.7 on the bottom left shows the ME histogram averaged over all generated pCT scans for all subjects considering the uncorrected *pCT_u* images, fused *pCT_h_md* images with median and fused *pCT_h_mn* images with mean.

Partial volume was over-estimated with 150 HU in the uncorrected *pCT_u* images. However, this was corrected in the fused corrected pCT images but soft tissue HU values were over-estimated by values reaching 150 HU. Bone was significantly under-estimated in the uncorrected *pCT_u* images with values of ME ranging from 100 HU to 600 HU. This issue was tackled when images where corrected and fused by taking their median showing values of ME ranging from -20 HU to 250 HU. In addition, fusing the corrected pCT images by taking their mean seems to over-estimate bone presenting ME values reaching -100 HU.

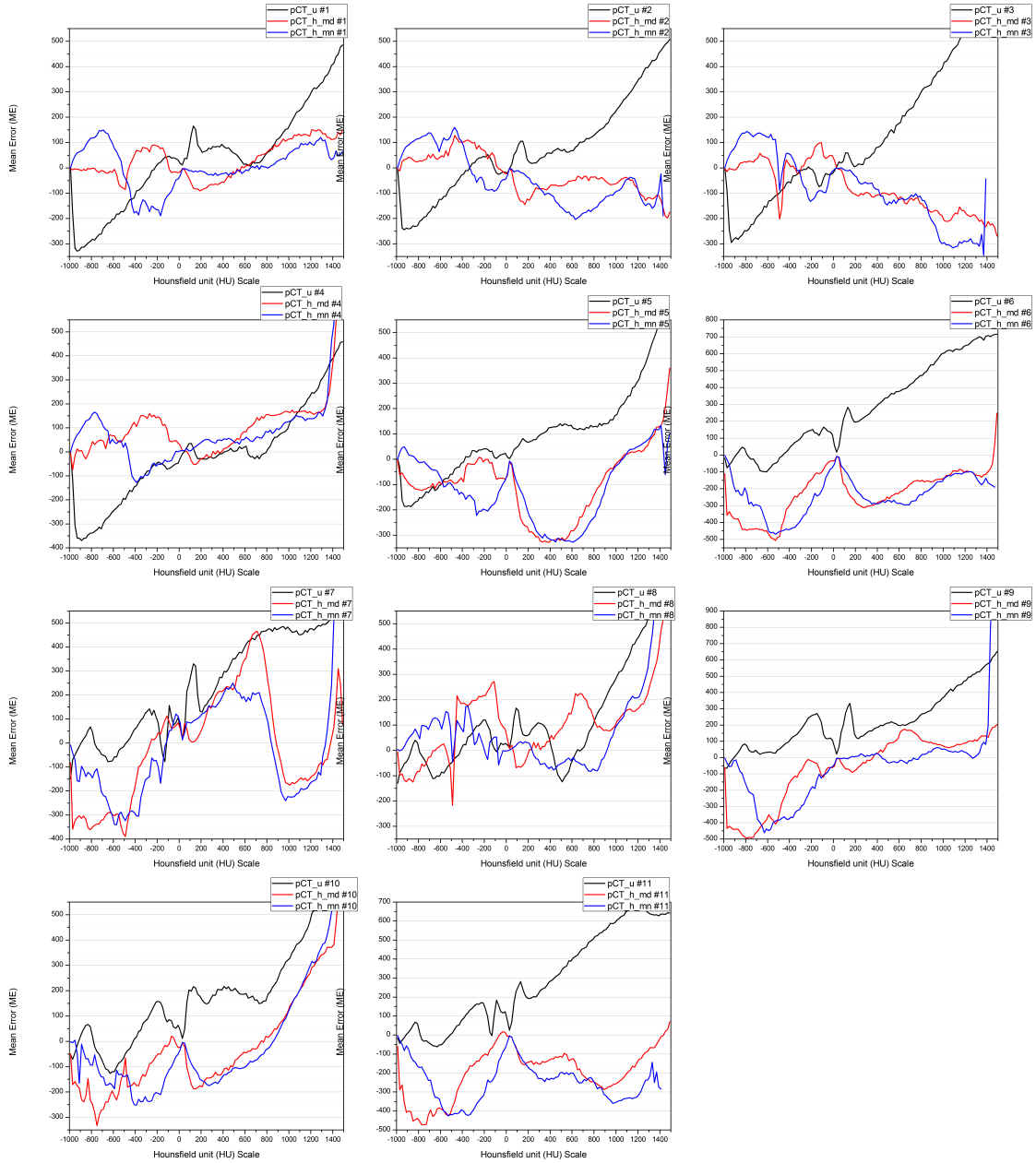

FIGURE 4.3.9 : ME histogram all over tissue classes (HU scale) for each subject.

The bottom right of Fig. 4.3.7 shows box and whiskers plot for ME distribution. The lower whisker shows that the minimum ME is close to -200 HU, ~180 HU and ~150 HU for pCT_h_md, pCT_h_mn and pCT_u, respectively. In addition, the upper whisker shows that the maximum ME is in the ranges of 300 HU, 450 HU and 620 HU for pCT_h_md, pCT_h_mn and pCT_u, respectively. The first quartile shows that 25% of the ME values are

below -100 HU, -100 HU and 0 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. The third quartile describing 75% of ME groups values in the ranges of 0 HU, -25 HU and close to 275 HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. The third quartile can be interpreted that 75% of the ME was caused by over-estimation of HU values by the proposed approach with its two versions and under-estimation of HU values by the single atlas-based estimation with no HU values correction. The inter-quartile range shows that 50% of the ME values fall inside the range of [-100; 0] HU, [-100; -95] HU and [0; 275] HU for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively.

Table 4.3 and Fig. 4.3.10 show the calculated correlation between the estimated pCT images and the real CT images for all subjects. A better correlation is achieved with the fused corrected pCT images with values going from 0.82 to 0.96 compared to the uncorrected *pCT_u* images giving average values ranging from 0.77 to 0.93 for each subject.

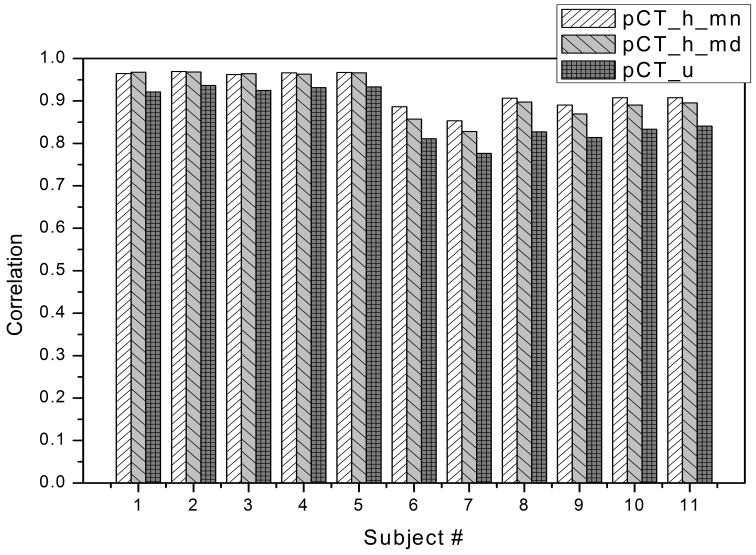

Figure 4.3.10: Calculated correlation between the estimated pCT images and the real CT images for each subject.

Furthermore, averaged correlation±SD over all subjects with standard deviation was 0.91±0.04, 0.92±0.03 and 0.86±0.06 for the corrected fused for *pCT_h_md*, *pCT_h_mn* and *pCT_u*, respectively. Overall, the average correlation shows that the corrected *pCT_h_mn* images correlated better with the original CT images than the corrected *pCT_h_md* and uncorrected *pCT_u*.

Atlas #	Target #										
	1	2	3	4	5	6	7	8	9	10	11
pCT_u											
1	-	0.89	0.92	0.86	0.92	0.93	0.89	0.91	0.94	0.94	0.94
2	0.95	-	0.93	0.94	0.94	0.94	0.89	0.92	0.93	0.93	0.94
3	0.94	0.87	-	0.88	0.94	0.94	0.89	0.91	0.92	0.94	0.97
4	0.93	0.94	0.93	-	0.94	0.93	0.87	0.92	0.91	0.94	0.95
5	0.81	0.72	0.79	0.72	-	0.82	0.81	0.82	0.83	0.89	0.87
6	0.81	0.72	0.79	0.72	0.82	-	0.81	0.82	0.83	0.89	0.87
7	0.78	0.68	0.75	0.67	0.75	0.80	-	0.84	0.84	0.78	0.83
8	0.85	0.75	0.81	0.80	0.84	0.83	0.84	-	0.84	0.84	0.83
9	0.85	0.72	0.77	0.71	0.79	0.86	0.85	0.87	-	0.83	0.86
10	0.85	0.78	0.83	0.83	0.85	0.86	0.79	0.83	0.80	-	0.88
11	0.84	0.77	0.84	0.79	0.83	0.88	0.83	0.83	0.86	0.88	-
pCT_h_mn											
-	0.96	0.96	0.96	0.96	0.96	0.88	0.85	0.90	0.89	0.90	0.90
pCT_h_md											
-	0.96	0.96	0.96	0.96	0.96	0.85	0.82	0.89	0.86	0.89	0.89

Table 4.3: Calculated correlation between the estimated pCT images and the real CT scans for all subjects.

An important evaluation is to measure the overlap of bone (DSC_{Bone}) between the estimated pCT images and the real CT scans. Table 4.4 and Fig. 4.3.11 show results for the DSC_{Bone} for all pCT images using the three approaches. Bone was defined by taking only voxels whose values are over 620 HU.

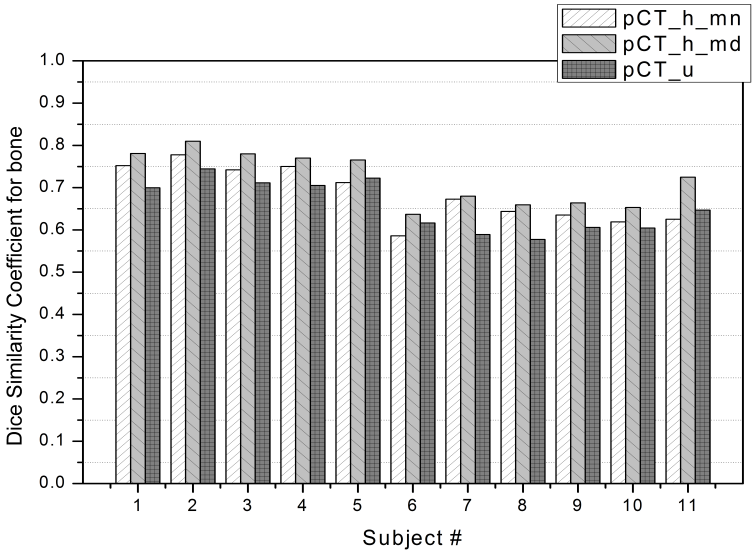

Figure 4.3.11: Calculated DSC_{Bone} between the estimated pCT images and the real CT images for each subject.

Atlas #	Target #										
	1	2	3	4	5	6	7	8	9	10	11
pCT_u											
1	-	0.71	0.73	0.67	0.69	0.69	0.65	0.64	0.71	0.70	0.76
2	0.73	-	0.74	0.73	0.75	0.76	0.68	0.74	0.74	0.74	0.79
3	0.73	0.55	-	0.67	0.74	0.74	0.68	0.68	0.70	0.75	0.82
4	0.69	0.73	0.69	-	0.72	0.70	0.64	0.69	0.67	0.71	0.75
5	0.68	0.74	0.70	0.68	-	0.76	0.67	0.73	0.70	0.74	0.77
6	0.59	0.59	0.62	0.56	0.68	-	0.59	0.59	0.57	0.65	0.67
7	0.58	0.55	0.60	0.56	0.60	0.58	-	0.61	0.54	0.58	0.64
8	0.56	0.57	0.56	0.46	0.64	0.59	0.60	-	0.54	0.61	0.59
9	0.65	0.59	0.62	0.57	0.63	0.61	0.58	0.58	-	0.55	0.64
10	0.57	0.60	0.63	0.63	0.67	0.56	0.57	0.59	0.49	-	0.68
11	0.62	0.64	0.71	0.64	0.68	0.65	0.62	0.58	0.59	0.68	-
pCT_h_mn											
-	0.75	0.77	0.74	0.75	0.71	0.58	0.67	0.64	0.63	0.61	0.62
pCT_h_md											
-	0.78	0.80	0.78	0.77	0.76	0.63	0.67	0.65	0.66	0.65	0.72

Table 4.4: Calculated DSC_{Bone} between the estimated pCT images and the real CT scans for all subjects.

Results for the DSC_{Bone} were significantly low for the uncorrected *pCT_u* images. The

proposed correction and fusion approach improved bone estimation which was already shown by the ME histogram and confirmed by the DSC_{Bone} results, giving values ranging from 0.58 to 0.77 for the corrected fused pCT_h_mn, from 0.78 to 0.80 for the corrected fused pCT_h_md giving the best results in comparison to values ranging from 0.57 to 0.74 for the uncorrected pCT_u. In the same way, we calculated the average DSC for bone, soft-tissue and adipose with standard deviation considering the whole set of subjects for the three approaches. Soft-tissue and adipose were obtained by taking only voxels within the ranges of [0; 600] HU and [-400; 0] HU[Onozato et al., 2014], respectively. Average DSC_{Bone} was 0.72±0.06, 0.68±0.06 and 0.65±0.07 for the corrected fused pCT_h_md, corrected fused pCT_h_mn and the uncorrected pCT_u, respectively indicating that a better bone overlap is achieved with the proposed approach when fusing the corrected pCT images by taking the median. In addition, The average $DSC_{Soft-tissue}$ was 0.86±0.03, 0.81±0.04 and 0.63±0.09 for corrected fused pCT_h_md, corrected fused pCT_h_mn and the uncorrected pCT_u, respectively. Again, a better overlap of soft-tissue is achieved with the proposed correction and fusion approach. Finally, The average $DSC_{Adipose}$ was 0.43±0.08, 0.34±0.08 and 0.1±0.14 for corrected fused pCT_h_md, corrected fused pCT_h_mn and the uncorrected pCT_u, respectively. Results for the $DSC_{Adipose}$ are significantly low and this can be explained by the thin structure of adipose which constitutes a small percentage of the brain and located mostly around the head, therefore obtaining a good overlap of adipose can be challenging.

To assess the effectiveness of the proposed hybrid correction technique, we implemented a simple multi-atlas fusion approach that performs the fusion of the uncorrected pCT_u images into one single pCT image, we presented this approach in [Boukellouz et al., 2017]. A target T1-w MR image was deformedly registered to a set of atlas MR images using multi-resolution and multi-metric B-spline registration. The subsequent deformations were applied to their corresponding CT atlas images in order to produce initial pCT images. The latter images were eventually fused into a single uncorrected pCT image using simple averaging by taking their mean. MAE was 283.39 HU compared to 78.13 HU and 84.7 HU for the corrected fused pCT_md and pCT_mn, respectively. Additionally, correlation was 0.81 for the fused uncorrected pCT compared to 0.91 and 0.86 for the corrected fused pCT_md and pCT_mn, respectively. This indicates that the proposed hybrid correction approach has effectively improved HU values correctness.

4.3.5 Discussion

The proposed hybrid approach combines deformable image registration, image segmentation, a hybrid HU values correction technique and multi-atlas fusion. Deformable image registration was performed on different resolution levels in order to cover a large scale of inter-patient variations. Furthermore, FCM algorithm was used to support the hybrid HU values correction process. This correction technique helped to improve pCT image prediction quality for each voxel of interest by inclusion of additional information about the surrounding voxels in a constrained neighbourhood, but the hybrid HU values correction process is still guided by the voxel of interest. Furthermore, the integration of tissue label information to compute a combined weight is useful to include only voxels of the same tissue class and eliminate voxels that might be close to the voxel under consideration but of a different tissue class. Moreover, the fusion of the generated pCT images into a composite pCT image by taking their median rather than the mean showed to significantly improve HU values correctness compared to the uncorrected pCT images where bone estimation was enhanced and this was shown by the increased values of DSC_{Bone} (average DSC_{Bone} was 0.72 for pCT_h_md compared to 0.68 and 0.65 for pCT_h_mn and pCT_u, respectively).

Our proposition can compete with recent hybrid approaches that focused on the head region, such as the work of [Andreasen et al., 2015], [Demol et al., 2016], [Boydev et al., 2017] and [Aouadi et al., 2017], to name a few. [Demol et al., 2016] proposed a single-atlas propagation scheme with a correction technique that uses MR intensity values. Most similar MRI voxels were selected in a limited neighbourhood based on a threshold set to 10%. Their corresponding CT values were averaged to yield the final pCT image, MAE was 150 HU. However, using a threshold may cause selection of voxels belonging to tissues different of that of the voxel to be corrected. This issue was tackled in the present work by inclusion of tissue class information where only voxels of the same tissue label were selected and incorporated in the correction process of the HU value of the voxel under consideration. Yet, there are some limitations our proposed contribution that should be considered:

- The approach presented low overlap between bone in pCT images and bone in the original CT (best DSC_{Bone} was 0.72) and this is largely owed to low T2 relaxation time of T2-w MR imaging which results in an unability to capture bone signal. As a consequence, lack of precise bone information in MR images resulted in inaccurate segmentation of bone, therefore bone HU values correction was affected by a misclassification of bone as soft-tissue or air in FCM segmentation giving low values of DSC_{Bone} and high rates of MAE as presented in Table 4.3 and Fig. 4.3.7;

- Deformable image registration is regarded a computationally expensive task typically when implemented over several image resolutions. For instance, deformable image registration of one atlas MR image to the target MR image in this work required approximatively 25 mn, which is significantly large when dealing with a multi-atlas approach as the registration time will multiply proportionally with the increased number of atlas images. Generally, deformable image registration for clinical applications requires the integration of a graphical processing unit to accelerate the process and reduce computation time;

In a future work, it would be more relevant to compare the proposed correction and fusion approach with the recent hybrid approaches using similar MRI sequences and same data set size. In addition, evaluation of image segmentation results in this work using a ground truth segmentation is important to be considered because segmentation errors can affect negatively the accuracy of the proposed hybrid HU values correction technique.

4.4 Conclusion

We proposed in this chapter, a multi-atlas hybrid approach to estimate a pCT image from T2-w brain MRI in the context of MRI-only RT. We used multi-metric, multi-resolution deformable image registration with a 3D hybrid HU values correction technique supported by FCM segmentation. This correction technique incorporates information about MR intensity, spatial location and tissue label from both the atlas and the target MRI in a neighbourhood of a predefined size. Furthermore, the corrected pCT images were fused either by taking their mean or their median. Comparison of the resulting fused corrected pCT images to the uncorrected unfused pCT images showed satisfactory results with a significant improvement.

The next chapter presents the second contribution for pCT image synthesis from MRI data. It is a learning-based approach that differs from the first contribution in many aspects as it is less computationally expensive and builds a data-independent prediction model.

CHAPTER 5

Ensemble learning with stacked generalisation for pCT image estimation

5.1 Introduction

Ensemble methods have been widely employed since their appearance in the 1990s by the works of [Hansen and Salamon, 1990] and [Schapire, 1990]. However, it was not until late 2016 that the first ensemble method for pCT image estimation from MRI was proposed. Since then, a number of ensemble-based approaches using random forests has appeared, these methods have shown to generate reliable pCT images. Inspired by these works, we present in this chapter, an ensemble learning approach with stacked generalisation (ES) to estimate a pCT image from multi-modal MR images using patch-based feature extraction, fusion and reduction. Unlike the aforementioned ensemble methods, here we employ heterogeneous learners to build the first level ensemble model. The current chapter is an extension to a paper that we published in [Boukellouz and Moussaoui, 2019]. To the best of our knowledge, this work is considered the first attempt to implement stacking in the context of MRI-only RT. Fig. 5.1.1 shows a flowchart of the proposed workflow.

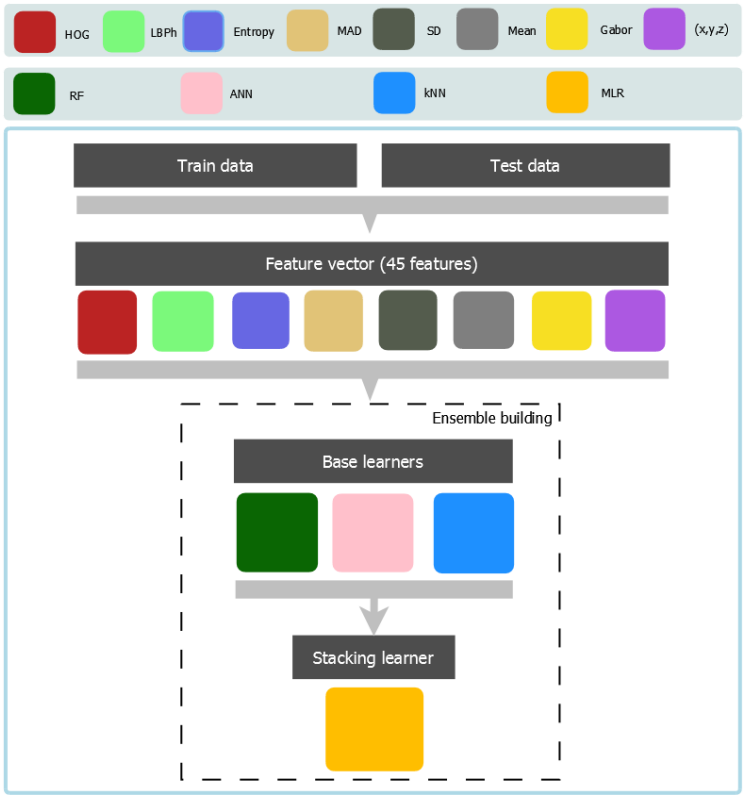

Figure 5.1.1: System overview.

Elements of the proposed learning method were chosen for the following reasons:

1. The use of multi-modal MRI sequences has shown in many works to provide valuable information about different tissues that one imaging sequence may not be able to provide [Speier et al., 2018];

2. Feature extraction focused on texture, shape and spatial characteristics to cover a large amount of information embedded in the original MR images efficiently;

3. Feature fusion and reduction were performed for dimensionality reduction and to fuse correlated features to choose only informative and descriptive features, reduce computation time and combine features;

4. The first level consisted of three heterogeneous base learners that have shown to previously provide good results in pCT image estimation [Zhong et al., 2016, Hu and Zhang, 2018, Yang et al., 2017]. In addition, as was presented in chapter one, heterogeneity of the base learners can results in diversity of predictions [Chandra and Yao, 2006, Liu and Yao, 1999];

5. Stacking was employed instead of other ensemble methods, mainly because this ensemble method has not been investigated before in the field of MRI-only RT, thus it would be of a great interest to investigate the performance of stacking in pCT image estimation. Furthermore, Most of the ensemble learning methods attempt to learn complex mappings between MR and CT values and combine their predictions using either averaging or weighted averaging. Different from these works, in stacking, the second level learner is generally a simple learning algorithm that performs the combination in less time and more efficiently than using averaging or weighted averaging technique. Moreover, combination by learning takes into consideration the underlying data proprieties; therefore, produces a stronger learning model.

5.2 Data description and pre-processing

Brain images from ten subjects were used in this work. 3D T1-w and T2-w MR images were acquired with prior injection of contrast agent Gadolinium with an SE sequence (TE/TR = 15/800 ms and TE/TR = 90/3000 ms for T1-w and T2-w, respectively). Voxel size ranged from $0.82\times0.82\times3$ mm^3 to $1.25\times1.25\times4$ mm^3 and number of slices varied from 26 to 52.

3D CT images were acquired using a tube voltage of 120 kVp, exposure time of 2 ms, and exposure rate of 250. Voxel size ranged from $0.40\times0.40\times3$ mm^3 to $0.65\times0.65\times4$ mm^3 with a number of slices varying from 28 to 49.

All MR and CT images were resampled to the same resolution of $512\times512\times40$. T1-w and T2-w MR images were bias corrected using the N4 bias-field correction method [Tustison and Gee, 2009] (Bspline order=3, number of iterations at each level=100, convergence threshold=0.001, shrink factor=3) followed by histogram matching to unify MR intensities across all subjects (levels= 1024, match points=7, threshold at mean intensity). T2-w MR images were aligned to one randomly chosen T2-w MR image using MI-based multi-resolution affine registration [Klein et al., 2010]. The subsequent deformations were applied to the T1-w and CT images.

To validate the performance of the proposed learning model, two-fold cross validation was performed in the train and test phases, meaning that five subjects (200 image slices) were used for training and the rest of the other five subjects were used for testing and vice versa. The two-fold cross validation was chosen due to hardware and memory limitations. Otherwise, a five/ten-fold cross validation would have been more suitable.

5.3 Patch-based multi-modal feature extraction

In order to cover a large amount of information and represent the original MR images efficiently, shape features were chosen to represent the subject-specific shape because the brain shape varies from one subject to another, thus, histogram of oriented gradients (HOG) [Dalal and Triggs, 2005] features were suitable for this type of image characteristics. Texture features allow to represent information about tissues in MR images which are crucial to define tissue types in CT images. For this purpose, histogram of local binary patterns (LBPh) [Ojala et al., 2002], Gabor [Gabor, 1946] and discrete wavelet transform (DWT) [Mallat, 1989] features were chosen. Finally, spatial features were employed to consider the distribution of tissues. Different tissues can be present at different locations. At first, a non-overlapping sliding window of size 8×8 was placed over each T1-w and T2-w MR image defining a patch from which the following features were extracted:

- **LBPh**: it is a texture descriptor largely used in facial expression recognition and human detection. It is mainly based on labelling the pixels of a $d \times d$ neighbourhood (8×8 in our case) by thresholding each pixel with the centre value and considering the results as a binary number. The result is often expressed in a decimal value. We considered a radius of eight pixels to compute the LBP features. For each pixel at a location *(x,y)*, its LBP was computed as follows:

$$LBP(x,y) = \sum_{i=0}^{P} s\left(g_i - g_c\right) 2^P. \tag{5.3.1}$$

  Where g_c and g_i are grayscale values of the centre pixel and surrounding pixels, respectively. P is the number of pixels in the neighbourhood and $s\left(g_i - g_c\right)$ is 1 if $g_i - g_c \geq 0$ and 0 otherwise. Next, the LBPh was computed by counting the number of occurrences of each LBP feature in 26 bins each of size 20 (0~255) resulting in a feature array of size 1×13;

- **DWT:** it is a signal representation method mainly used in signal processing and data compression originally proposed by [Mallat, 1989]. We used a wavelet decomposition of two levels considering determination coefficients at three orientations (horizontal, vertical and diagonal) and approximation coefficients. A DWT can be described as follows:

$$\psi_{j,k}(t) = 1^{j/2}\psi(2^j t - k). \tag{5.3.2}$$

  j and k are scale and translation parameters. Using the resulting coefficients from the

DWT, we calculated the following features: entropy, mean absolute deviation (MAD), SD and mean. The DWT feature vector is of size 1×32;

- **HOG:** A patch is divide into overlapping blocks of size 2×2 and each block is divided into small cells of size 4×4. Horizontal and vertical gradients are calculated for each pixel within the cell as follows:

$$G_x(y,x) = g(y,x+1) - g(y,x-1);\ G_y(y,x) = g(y+1,x) - g(y-1,x). \tag{5.3.3}$$

  $g(y,x)$ is the grayscale value at locations x and y. $G_x(y,x)$ and $G_y(y,x)$ are the horizontal and vertical gradients, respectively.After that, the histograms were determined according to nine orientation bins in 0°-180°. The HOG feature vector is of size 1×36;

- **Gabor filters:** A specific number of filters with certain orientations is applied to the pixels of a patch in order to extract edges and shapes. A Gabor filter can be modelled as the product of a Gaussian kernel times a sinusoid function as follows :

$$g(t) = ke^{j\theta}\omega(at)s(t). \tag{5.3.4}$$

  We applied four Gabor filters with four orientations resulting in a feature vector of size 1×16;

- Spatial coordinates: (x,y,z) coordinates of the centre of the patch were considered.

In total, we extracted a feature array of size 1638400 × 100 from each modality. The two T1-w and T2-w feature vectors were concatenated together resulting in a feature vector of size 1638400 × 197. The dimensionality of this feature vector is high and can result in what so called "the curse of dimensionality" where a large set of features can result in a degradation of the performance of the trained model. Thus, feature reduction is necessary, for this reason we opted for PCA. In fact, the role of PCA is two-fold: first, PCA can help in reducing the feature space by selecting the most contributing features as was explained in chapter one. Therefore, reducing complexity and computational burden, and second, it serves as a statistical feature fusion level tool, where features from both modalities are combined into one feature vector. PCA was applied to the multi-modal feature vector taking 45 principal components. On the other hand, The target set was built by taking HU values of the real CT scans so that each multimodal feature vector corresponds to one HU value. Fig. 5.3.1 shows a simplified illustration for the feature extraction, fusion and reduction workflow.

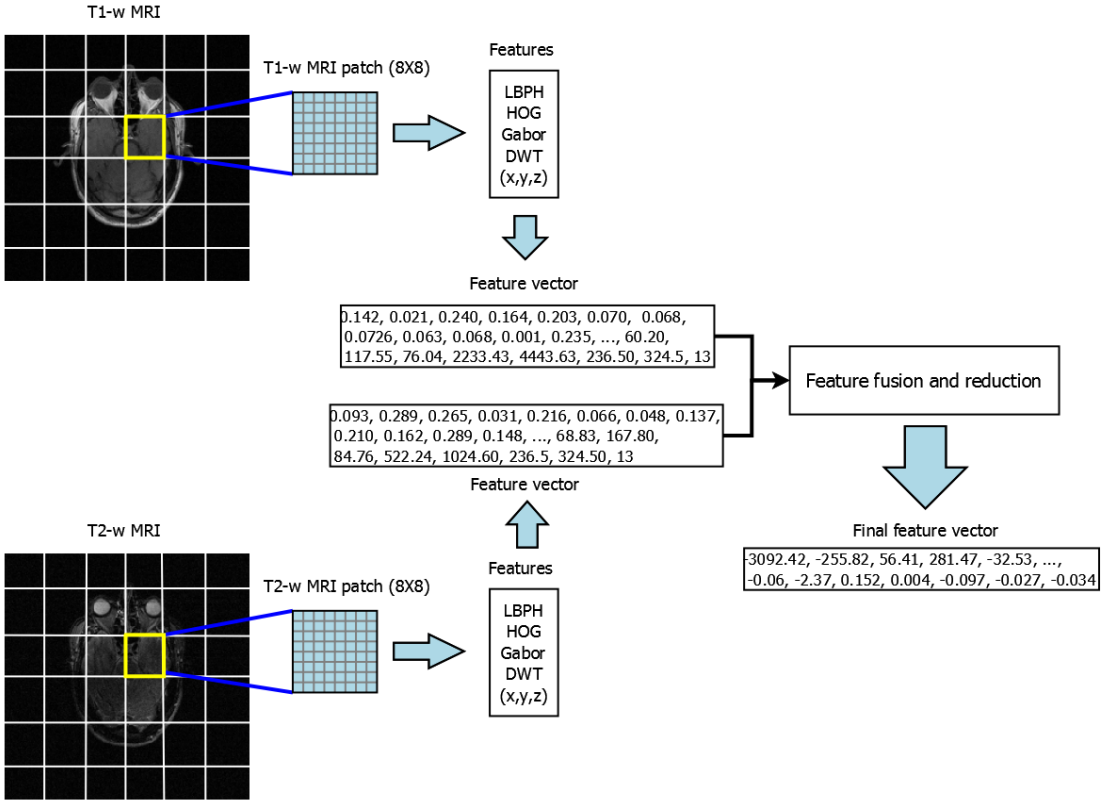

Figure 5.3.1: Simplified workflow for feature extraction, fusion and reduction from T1-w and T2-w MR images.

5.4 Building the ensemble model

Ensemble learning trains several heterogeneous or homogeneous base learners and prediction results from these learners are combined using either averaging, voting or learning. The latter combines results by learning to generalise from the initial predictions. We distinguish stacking[Wolpert, 1992] and infinite ensemble [Lin and Li, 2008]. Stacking introduces a second level learner that takes the output predictions of the first level base learners as input features and learns a mapping between these features and the target values. Whereas, the concept of infinite ensembles is based on learning the combination weights for all possible hypotheses using SVM, which is more suitable for non-linear learning problems. Since the output of the base learners and the target HU values are of the same nature (HU values) then our problem reduces to a linear mapping problem which we solve by introducing a stacking level represented by MLR.

.

5.4.1 Base learners training

To construct the first level of the proposed ensemble model, we employed three common ML algorithms, namely: ANN, RF and kNN that were used as base learners to take advantage of their diversity. It is believed that diversity of the base learners is a key contributor for building a performant ensemble model [Zhou, 2012]. For each base learner, we used the following parameters based on a trial/error strategy:

1. A fitting ANN with an input layer having 15 nodes, four internal layers having nodes of sizes 10, 10, 10, 10, respectively and an output layer composed of one node which reffers to the target HU value. Scaled conjugate gradient backpropagation [Werbos, 1974] was used as the training function, mean squared error (MSE) as the performance measure and tan-sigmoid as the transfer function for the internal layers and a linear transfer function for the output layer.

2. A regression RF having 30 decision trees, each with a minimum leaf size of 20 and 40 predictors per leaf. For a test feature vector x, its value $\hat{y}$ is predicted using the average predictions of all trees such that:

$$\hat{y} = \frac{1}{B}\sum_{i=1}^{B} RF_i\left(x\right), \tag{5.4.1}$$

   Where B is the total number of decision trees (B=30) and RF_i is the trained tree number i.

3. Five nearest neighbours were selected for each feature vector using the Euclidian distance as the similarity measure.

5.4.2 Stacked generalisation

We used MLR as the second level stacking learner. MLR models a data set having two or more independent predictors to one dependent response. When adjusting formula (1.4.1) representing the general form of an MLR model to our case of study, we obtain:

$$Target = a_1x_1 + a_2x_2 + a_3x_3 + b. \tag{5.4.2}$$

Where x_1, x_2 and x_3 are the prediction results from ANN, RF and kNN, respectively. The response variable is represented by the HU values of the original CT scans.

Before feeding the initial predictions to the MLR model, a collinearity analysis is necessary. Collinearity occurs when at least two predictors are highly correlated one to the other, which

can result in inconsistent regression parameters estimation [Liu and Yao, 1999, Chandra and Yao, 2006]. Table 5.1 shows collinearity results.

	x_1	x_2	x_3	Target
x_1	1	-	-	0.91
x_2	0.92	1	-	0.87
x_3	0.79	0.81	1	0.74

Table 5.1: Correlation matrix.

We observe a high correlation between predictors x_1 and x_2 that were generated by ANN an RF predictions, respectively. One way to overcome this issue is to remove one of these two predictors. Hence, we chose to omit x_2, because it showed a lower correlation with the target compared to that of x_1. Similar to the base learners training, the MLR model was trained using a two-fold cross validation approach.

5.4.3 pCT images generated using random forests

For comparison purposes, we generated pCT images with a random forest composed of 30 trees, a minimum leaf size of 20 and 40 predictors per leaf. RF method was chosen for comparison because it is a widely known ensemble method and has proven to be successful in many fields as well as in the field of MRI-only RT as we presented in literature review of this dissertation, therefore, it is considered a state-of-the-art method in ensemble learning. The implemented RF method is close to the one proposed by [Yang et al., 2017] to a certain extent in the exception of the type of the selected features and the use of a three-dimensional patch for feature extraction. In addition, we present some statistical evaluation results when considering prediction made by the base learners alone without stacking in the aim of emphasising the impact of this latter on improving prediction quality.

5.5 Results

Notations

pCT_es	pCT image predicted by the proposed ensemble stacking method
pCT_rf	pCT image predicted by the RF method
pCT_b	Initial predictions made by the base learners

After training the MLR model, the estimated coefficient values and the model equation are as follow:

$$Target = 0.93x_1 + 0.05x_3 + 0. \tag{5.5.1}$$

We notice that ANN predictors contributed largely to the estimation, while kNN predictors' contribution is considerably small.

Fig. 5.5.1 shows axial, sagittal and coronal slices of the original CT scans, predicted pCT_rf using the RF method and pCT_es using the proposed ES method considering all subjects.

It can be seen that the estimated pCT images using both approaches preserved the shape thanks to the shape features employed in the learning process. In addition, air cavities were well-reproduced by the proposed ES approach compared to the results presented by RF. Furthermore, RF method generated noise around the brain characterised by a miss-prediction of bone voxels instead of air. The noise is more pronounced in subjects #3, #5 and #10.

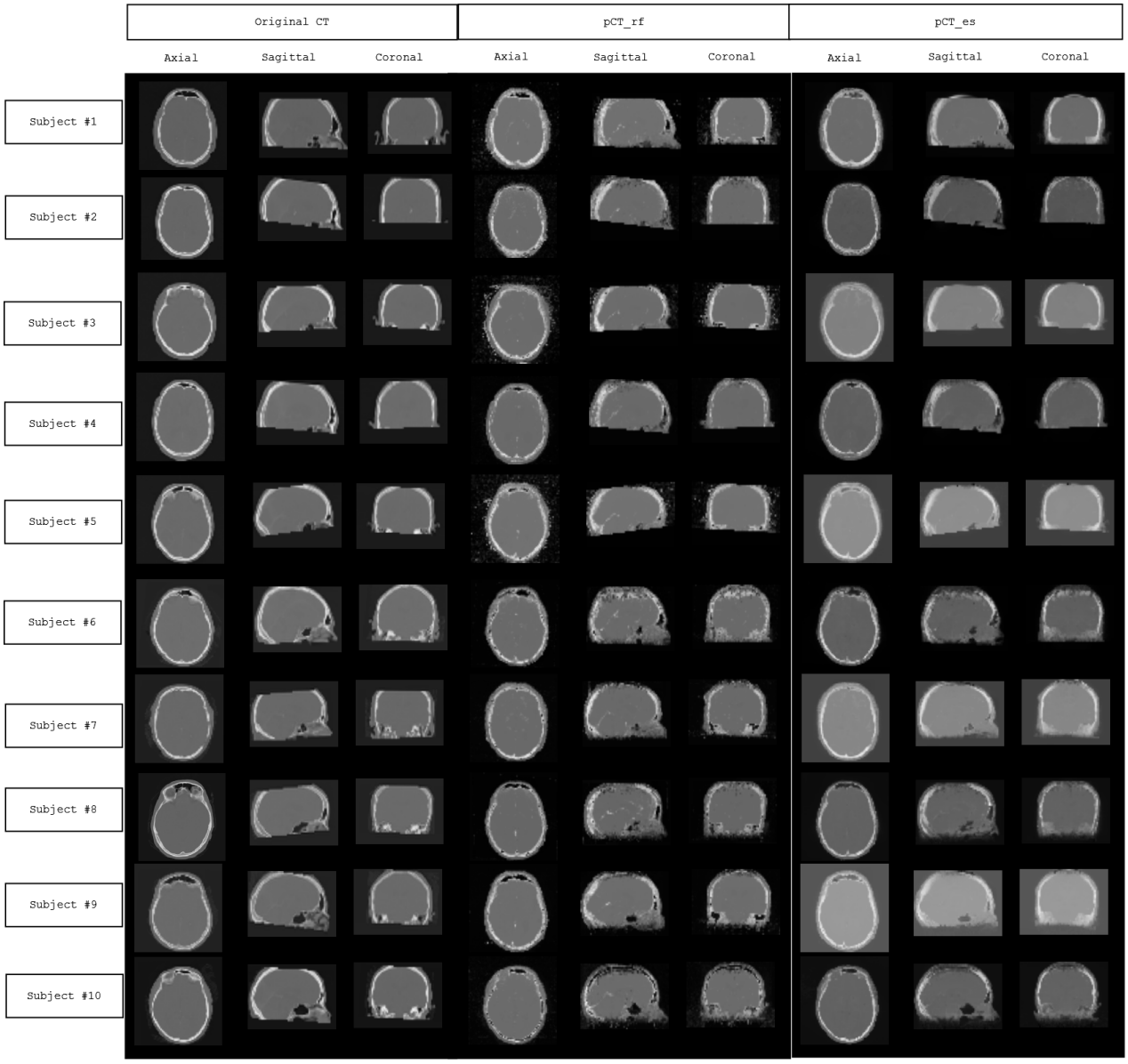

Figure 5.5.1: Axial, sagittal and coronal slices of the original CT images (left) with pCT images estimated using random forests (middle) and the estimated pCT images using the proposed approach (right).

Fig. 5.5.2 shows axial, sagittal and coronal slices of the difference images between the original CT scans and the predicted *pCT_rf* using the RF method; and between the original CT scans and the predicted *pCT_es* using the proposed ES method considering all subjects.

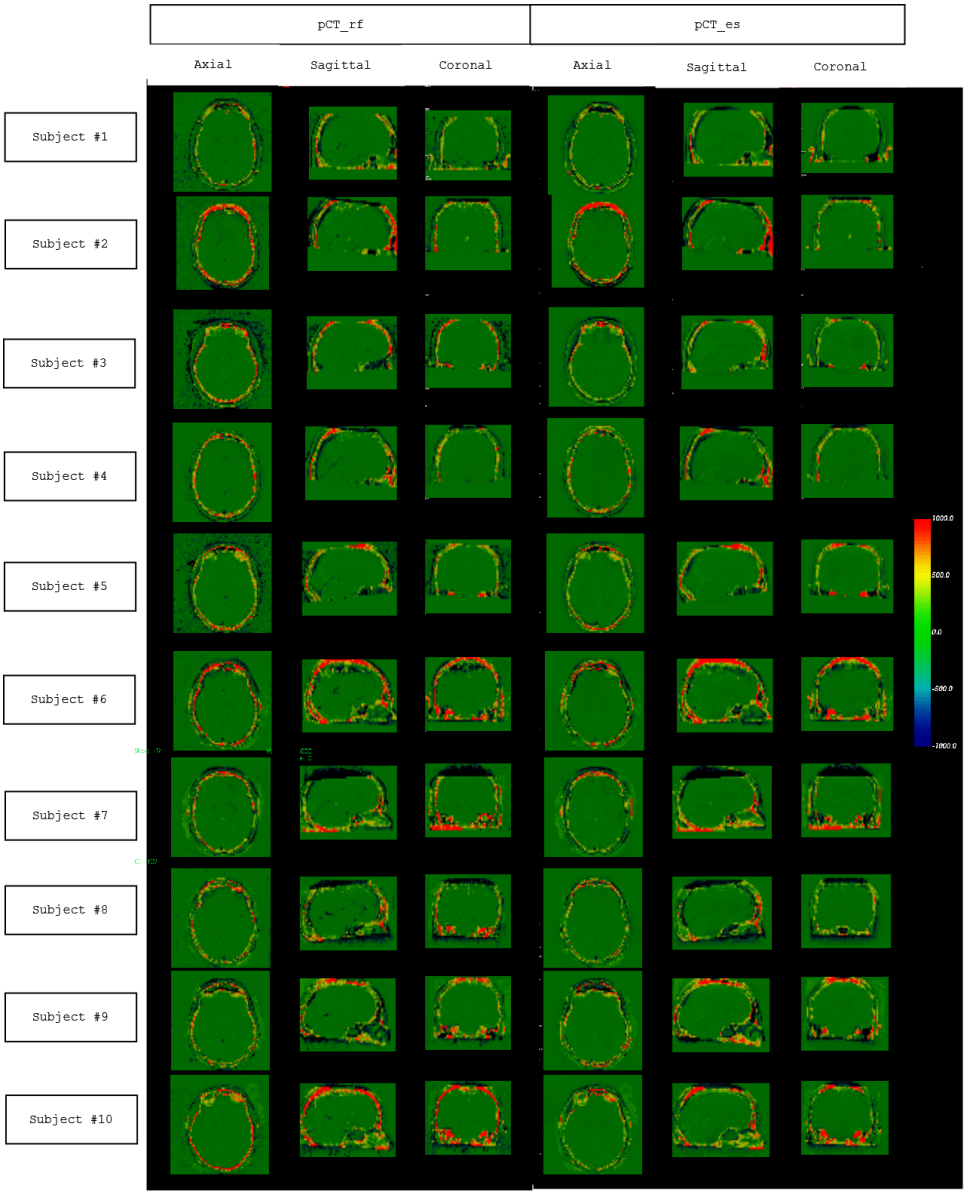

Figure 5.5.2: Axial, sagittal and coronal slices of the difference images between the proposed approach and the original CT scans (right) and btween pCT images estimated using randomm forests and the original CT scans (left).

Difference images show that both approaches present noticeable differences in bone-soft tissue interfaces and bone-air interfaces depicted by the orange colour. However, differences are larger in pCT images generated by RF.

Fig. 5.5.3 shows MAE_{hist} (top) with box and whiskers plot (bottom) for the MAE distribution considering values averaged over all subjects. In addition, individual MAE_{hist} plots for each subject are presented in Fig. 5.5.4.

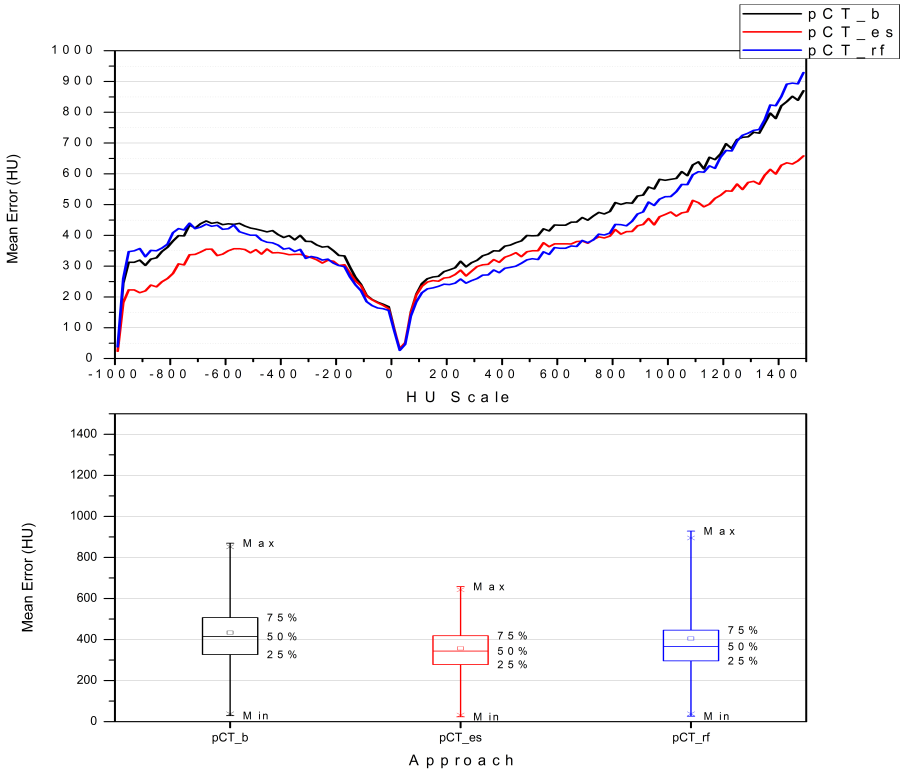

Figure 5.5.3: MAE histogram (top) averaged over all subjects for with box and whiskers plot MAE error distributuion (bottom) for the ES, base learners estimation and RF methods.

MAE_{hist} on Fig. 5.5.3 (top) and Fig. 5.5.4 show that the base learners had MAE values in soft-tissue [-399; 600] HU ranging from 0 to 450 HU. The introduction of a stacking level shows to reduce MAE in soft-tissue with more than 100 HU giving MAE values between 0 and 350 HU. MAE presented by the RF approach are relatively close to those presented by the proposed ES approach. For bone (>601 HU), base learners prediction presented MAE ranging from 425 to 875 HU. However, this error was significantly reduced with the MLR model where bone estimation was improved by reduction in MAE with about 225 HU giving MAE values in bone varying from 350 to 650 HU. The RF method presented larger MAE in bone compared to the proposed ES approach with differences in MAE reaching 300 HU. The bottom plot of Fig. 5.5.3 shows that the first quartile of the MAE distribution (25%) has MAE values going from 0 to 250 HU for all three approaches. However, the third quartile represents 75% of MAE values reach values of 500 HU. It can be seen that the proposed approach helped to significantly reduce 25% of MAE values to less than 400 HU.

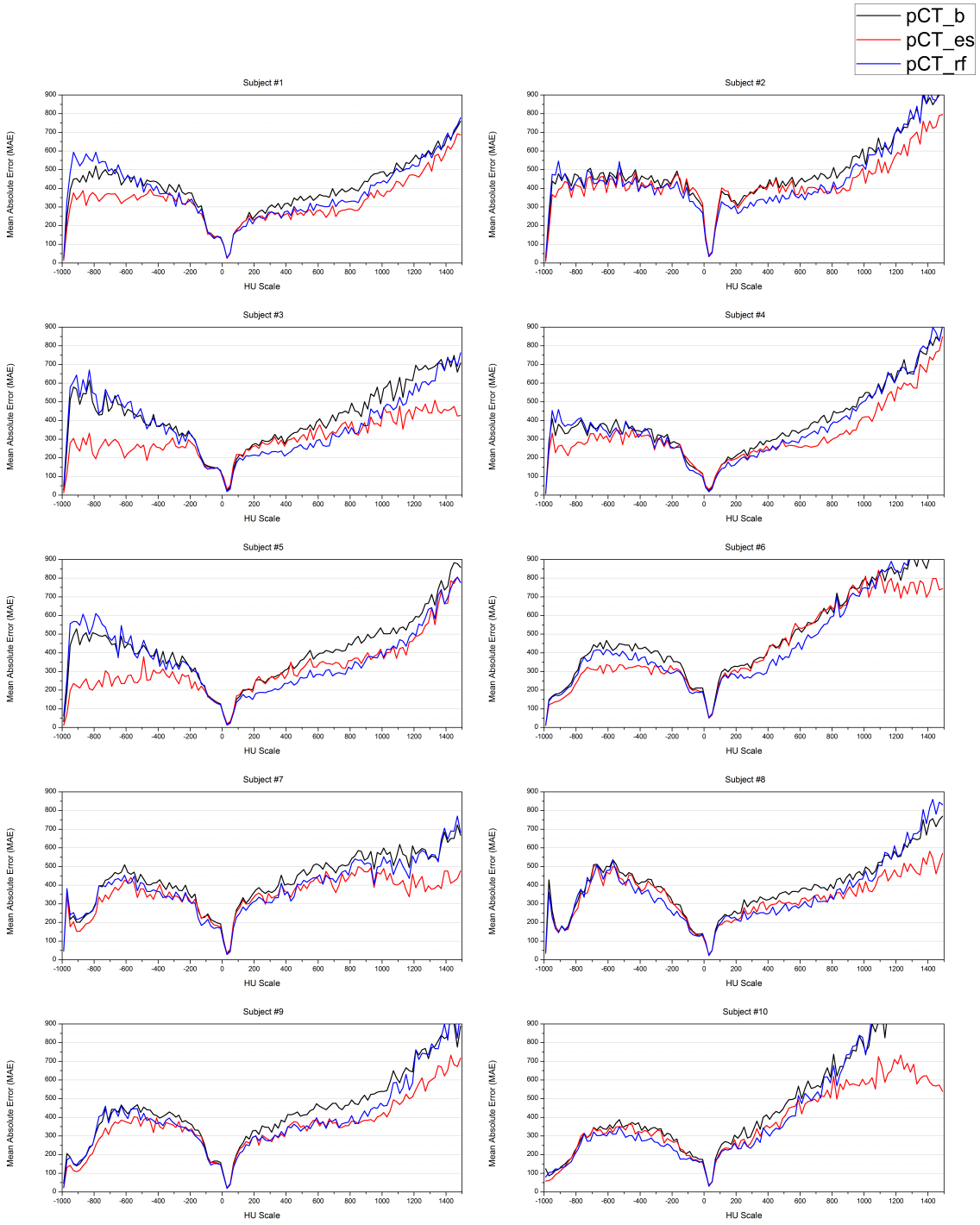

Figure 5.5.4: MAE histogram all over tissue classes (HU scale) for each subject.

Fig. 5.5.6 shows ME_{hist} (top) with box and whiskers plot (bottom) for the ME distribution considering values averaged over all subjects. In addition, individual ME_{hist} plots for each subject are presented in Fig. 5.5.5.

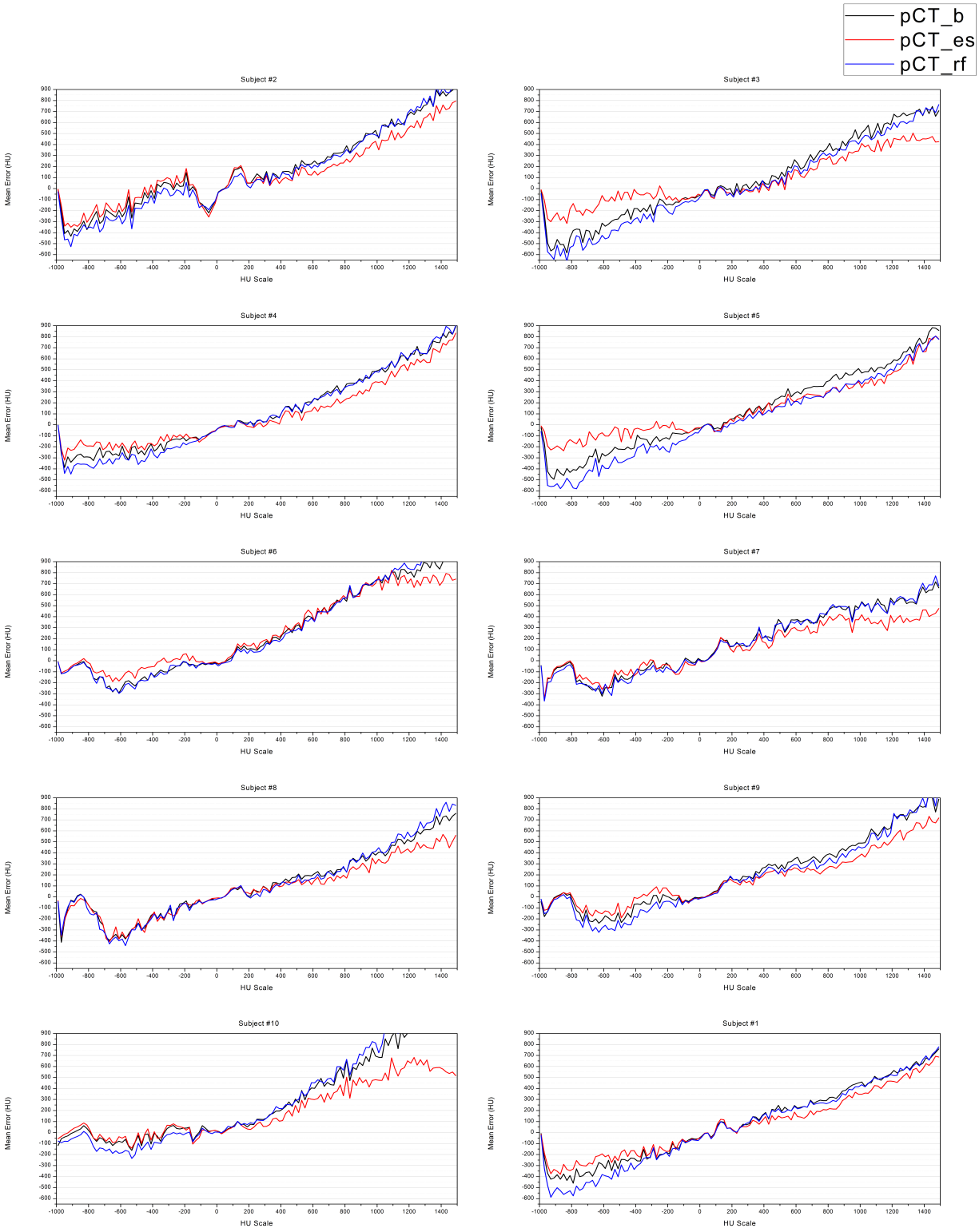

FIGURE 5.5.5 : ME histogram all over tissue classes (HU scale) for each subject.

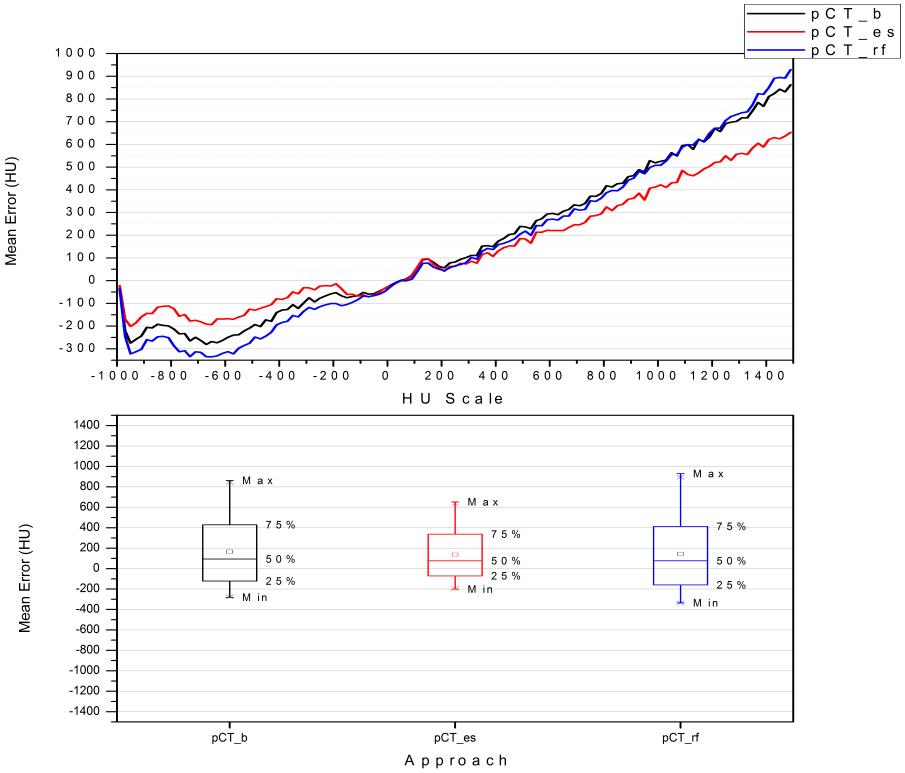

Figure 5.5.6: ME histogram (top) averaged over all subjects for with box and whiskers plot for ME error distributuion (bottom) for the ES, base learners estimation and RF methods.

ME_{hist} presented in Fig. 5.5.6 (top) and Fig. 5.5.5 show that soft-tissue was over-estimated in the interval [-399; 0] HU and under-estimated in the interval [0; 600] HU by all approaches. However, the proposed ES approach presented the lowest ME values. Bone was under-estimated by the base learners and RF with a ME reaching 950 HU. This under-estimation was circumvented by the proposed approach reducing the under-estimation to less than 650 HU. The bottom plot in Fig. 5.5.6 shows that the inter-quartile range for proposed approach is more condensed than the other two approaches meaning that most ME values are located in a small interval [-100; 300] HU compared to the base learners predictions [-120; 400] HU and RF [-200; 400] HU. In summary, all three approaches under-estimated bone HU values. However, the proposed ES approach succeeded to improve this under-estimation.

Table 5.2 shows discreet MAE and ME values for each subject with the average±SD of the entire set. Results show that the proposed ES approach outperformed the RF method. In addition, passing from the base learners to the stacking model seems to improve prediction quality remarkably. For instance, subject #5 presented a value of MAE of 152.30 HU upon first prediction. This was reduced when passing to the MLR model giving an MAE value of 100.86

HU. The average MAE±SD was 87.56±19.7 HU for the proposed ES approach compared to 106.65±18.4 HU for RF method.

Subject #	*pCT_b*		*pCT_rf*		*pCT_es*	
	MAE	ME	MAE	ME	MAE	ME
1	103.73	1.89	108.54	-14.33	90.43	1.81
2	68.474	0.46	66.73	-4.72	60	-5.75
3	122.19	24.73	118.96	21.57	111.44	25.48
4	112.26	-34.98	103.84	-29.83	99.92	-34.31
5	152.30	-0.04	140.18	17.70	100.86	4.11
6	79.19	-1.65	87.44	-18.58	72.39	-8.23
7	93.48	-16.48	109.75	-43.79	62.88	-8.01
8	96.99	-7.944	115.73	-47.65	64.99	8.99
9	121.87	-17.12	112.23	-23.78	115.09	-28.65
10	118.30	-2.74	103.07	-3.52	97.97	-2.30
Average±SD	106.65±18.4	-14.69±22.02	106.88±33.2	-5.38±20.77	87.60±19.7	-4.68±16.43

Table 5.2: MAE and ME values expressed in HU for all subjects with average and SD considering all three approaches.

Pearson Correlation values were 0.92, 0.87 and 0.89 for the proposed ES approach, base learners and RF method, respectively. This demonstrates that pCT images estimated using the proposed approach correlate the best with the real CT scans.

In order to assess bone estimation, DSC_{Bone} was computed between the bone in the estimated pCT images using the two approaches and bone in the original CT scans. Bone was obtained from CT images by taking only voxels whose values are above 620 HU which is the bone interval. Results are presented in table 5.3.

Subject # / Approach	1	2	3	4	5	6	7	8	9	10	Average±SD
pCT_rf	0.61	0.53	0.67	0.51	0.65	0.39	0.65	0.61	0.58	0.35	0.55±0.10
pCT_es	0.66	0.62	0.76	0.62	0.68	0.53	0.70	0.68	0.64	0.67	0.66±0.06

Table 5.3: Calculated DSC_{Bone}values for bone overlap between the estimated pCT images using RF method and the proposed ES approach for all subjects with the average and SD.

From the above table, it is clear that the RF method has a very weak bone overlap between the estimated pCT images and the original CT images concluding that the RF method had difficulties in predicting bone; the DSC_{Bone} for the RF method ranged from 0.35 to 0.67 with an average DSC_{Bone} of 0.55±0.10 . On the other hand, the proposed ES approach also had difficulties in predicting bone depicted by low DSC_{Bone} values ranging from 0.53 to 0.76 and an average DSC_{Bone} of 0.66±0.06 . However, this inability to correctly predict bone values is

not as severe as the case of RF method. This can be largely owed to the nature of T1-w and T2-w MR images since these two imaging modalities do not favour bone visualisation, which was the same case for the hybrid method presented in chapter four. Lacking bone information on MR images makes the feature extraction step unable to extract the descriptive features for bone, thus, bone tissue will be less represented by these features.

5.6 Discussion

In this chapter, we presented an ensemble learning approach with stacking generalisation and multi-modal MRI feature extraction to estimate a pCT image in the context of MRI-only RT. Our model can use data coming from different MRI scanners and can cope with differences in intensities thanks to the histogram-matching step present in the pipeline of the approach, which aims to bring all MR intensities to one scale. The proposed approach uses features from T1-w and T2-w MR images since these two modalities provide different information about tissues in the brain due to their different relaxation parameters. Therefore, more descriptive features are extracted. PCA was employed in order to reduce computational burden, reduce memory consumption and fuse the features. A two level ensemble learning approach was followed in this work by initially feeding the extracted multi-modal features to three base learners (ANN, RF and kNN) for learning the first level prediction models. The latter models were used to perform an initial prediction, which represented the input features for the second level learner that is in this work MLR.

Statistical comparison results showed that our method can effectively predict a reliable pCT image while outperforming RF giving an average MAE of 87.56±19.7 HU and ME of -4.68±16.43 HU compared to an MAE of 106.65±18.4 HU and ME of -14.69±22.02 HU for RF prediction. Furthermore, the proposed ES approach can compare well with other recent contributions for MRI-only RT in the head region. For instance, [Huynh et al., 2016] achieved an MAE of 99.9 HU using SRF when averaged over 16 brain subjects. Multi-atlas approaches of [Burgos et al., 2014] and [Sjölund et al., 2015] achieved MAE values of 102 HU and 113.4 HU, respectively. In addition, [Su et al., 2015] proposed a clustering-based approach, which gave an MAE of 130 HU. [Hofmann et al., 2008] combined pattern recognition with image registration in a hybrid workflow and obtained an MAE of 100.7 HU. [Johansson et al., 2013] obtained an MAE of 130 HU when training a Gaussian mixture regression model on MRI data, and [Andreasen et al., 2015] obtained an MAE of 85 HU using a patch-based approach. Furthermore, the works of Han [2017], [Emami et al., 2018] and [Dinkla et al., 2018] achieved with their deep learning models MAE values of 84.8 HU, 89.3 HU and 67 HU, respectively. However, these results cannot be directly compared because of the differences in type and size

of image data used. It is worth to note that the two major limitations of this approach are:

- The use of the non-overlapping sliding window for feature extraction resulted in low resolution pCT images where details were not well-preserved. However, this was only used due to hardware and memory limitations. An alternative is to use an overlapping sliding window in order to capture more details and preserve context;
- In addition, bone was under-estimated leading to high values of MAE as was shown in Fig. 5.5.3 and table 5.3; one attempt to overcome this issue is to introduce a bone segmentation step prior to ensemble model building or to use specialised MRI sequences that enhance bone visualisation (ZTE or UTE sequences).

It would be more relevant to use only supervised learning for pCT image estimation tasks, this is because unsupervised learning algorithms are generally lazy learners that require the training data to be permanently present and need more time for prediction. In this work, kNN estimation for one pCT volume took approximatively 8 minutes while prediction using the other two supervised base learners was performed in few seconds. A fast pCT image estimation would be of more efficiency to be integrated for online MRI-guided treatment using MRI-linac. Our work suggests that using more diverse and uncorrelated learners can yield a better stacking model therefore improving pCT image prediction accuracy.

5.7 Conclusion

In this chapter, we presented a simple yet effective approach to estimate a pCT image from MRI data in the context of MRI-only RT. Evaluation results showed that the approach successfully predicted pCT images similar to the original CT scans. However, bone estimation is challenging and we strongly advice the introduction of a specialised MR imaging sequence in order to enhance bone visualisation or to perform an initial bone segmentation step, therefore bone features would be more discriminative. The proposed contribution in this chapter presents a valuable add to ensemble methods for pCT image synthesis from MRI data and constitutes a starting point to investigate stacking methods in the field of MRI-only RT . Comparison results in this work show that in some cases stacking can be better suited for a particular problem than bagging. We also confirm that the diversity of the base learners plays a key-role in training the second level learner, where a correlation between predictors can cause the deletion of one predictor in order to preserve model consistency.

Conclusion and perspectives

Machine learning techniques have shown an enormous success in solving many medical-related problems. This is depicted by the increasing number of published works over the last decades focusing on solving divers medical image segmentation, image synthesis and tumour classification tasks, to name a few. In fact, machine learning has found its path for success in medical imaging thanks to the increasing amount of imaging data generated on a daily basis. This diversity in medical data has led to multi-modal data generation and incorporation in clinical practice. More precisely, the ongoing development of radiation therapy is striving to eliminate many of its limitations such as the excess radiation exposure due to the repeated CT imaging for treatment optimisation and the inherent errors related to the standard workflow for radiation treatment planning. Therefore, research is being directed toward using MRI as the only modality for radiation therapy for its numerous advantages by formulating this orientation as an image synthesis problem. To solve such a problem, many works have been proposed using different methods and techniques.

In this dissertation, we performed a review of these works and classified them into four classes in order to analyse their strengths and limitations to eventually propose two main approaches for pCT image synthesis using MRI data. The proposed approaches are the fruit of a deep analysis of existing work where we attempted to overcome some of the limitations of the existing work, such that: the first proposed approach for pCT image synthesis belongs to the hybrid methods class and combines several methods that have shown a considerable success when used alone, we included: deformable image registration, unsupervised FCM clustering and image fusion. In addition, we integrated a new hybrid correction technique that incorporates MRI information (intensity, location and tissue class) to correct HU values, thus improve pCT image prediction correctness. Comparison to a base-line atlas propagation method showed that each element of the proposed hybrid method contributed to reduction of error and improving prediction accuracy. The second proposed method for pCT image synthesis belongs to the learning-based approaches and it is considered a first attempt to implement stacked generalisation with ensemble learning for pCT image estimation. This proposed approach was inspired by the ever-increasing number of works employing the state-of-the-art ensemble method random

forests and comparing our proposition to a similar random forest method confirms that stacking can compare well with bootstrap aggregating.

Despite the success of methods for pCT image synthesis from MRI data, there are still many challenges that constitute a hot research topic in the field of MRI-only RT. From the one hand, the absence of a common benchmark to assess and compare methods for pCT image estimation is regarded as a serious issue. Indeed, a comparison benchmark is crucial to decide on the effectiveness of any method. What is more, most of the data used in the field of MRI-only RT is acquired from hospitals and it is not publically available. Therefore, providing an open test benchmark can be of a great use for the validation of any pCT estimation approach. On the other hand, bone estimation is considered very challenging especially when conventional MRI sequences are employed since the latter do not capture any signal from bone. For instance, the proposed approaches in this dissertation failed to accurately estimate bone due to the nature of T1-w and T2-w sequences. Furthermore, tissue interfaces such as bone-air or soft tissue-bone interfaces are difficult to be precisely estimated because of the complex heterogeneity in tissue distributions present at these locations. The same issue is valid for air cavities present at the nasal areas.

In a future work, these points are to be considered:

- Investigate the effectiveness of the hybrid approach on specialised MRI sequences conceived for enhancing bone visualisation ;
- Re-implement the ensemble stacking method on powerful hardware in order to consider overlapping window for feature extraction in the aim of preserving image details and boost the performance;
- Consider testing other heterogeneous base learners on a larger image set to introduce more variety in results and therefore the stacking model would take advantages of each base learner's contribution;
- Study the benefit of MRI multi-modality for pCT image estimation using different combinations of MRI sequences.

In recent few years, the fields of computer vision and image processing have been flooded with a tremendous amount of research papers focusing on solving many image-related problems using cutting-edge novel deep learning methods. Indeed, this particular machine learning discipline has opened many research directions more particularly in the field of image synthesis. Many deep network architectures have been proposed which have shown to synthesise reliable and accurate images. More particularly, generative adversarial networks is a novel network architecture that has proven its success in image synthesis depicted by the increasing number of

developed applications going from "image generation from scene graphs"[Johnson et al., 2018] to "text to image synthesis"[Zhang et al., 2018] to "natural image synthesis"[Brock et al., 2018] to "multi-contrast MRI image synthesis"Dar et al. [2019], among others. Based on this, it would be relevant to investigate the performance of GAN in the field of MRI-only RT. We envision implementing a GAN model to synthesis a pCT image from multi-modal MRI data considering different image datasets and body regions. Due to the lack of large sets of MRI-CT images, an alternative is to opt for image augmentation prior to training the GAN model to ensure robustness and efficiency.

Contributions

1. The first contribution is a study of the existing works in the field of MRI-only RT with a classification of methods into four categories. The contribution includes a comparative analysis in order to retrieve advantages and drawbacks of each class of methods;

2. The second contribution is an evaluation of one important parameter of deformable image registration which is the similarity measure. This contribution assesses the effect of choice of a certain similarity measure on the course of B-spline mono-modal deformable image registration;

3. The third contribution is a hybrid approach to estimate realistic patient-specific pCT image from conventional T1-w MR images. The contribution helped to significantly reduce prediction error by incorporation of a new correction technique that uses information from MR images;

4. The fourth contribution aims to predict a pCT image from multi-modal MRI data by using handcrafted features with feature level fusion. The contribution is considered a first work that investigates the usefulness of stacked generalisation in MRI-only RT.

List of included publications

Journal papers

1. **Wafa Boukellouz** and Abdelouahab Moussaoui. *Magnetic resonance-driven pseudo CT image using patch-based multi-modal feature extraction and ensemble learning with stacked generalisation*. Journal of King Saud University Computer and Information Science (JKSUCIS). Elsevier. 2019. https://doi.org/10.1016/j.jksuci.2019.06.002

2. **Wafa Boukellouz**, Abdelouahab Moussaoui, Abdelmalik Taleb-Ahmed, and Christine Boydev. *Multiatlas fusion with a hybrid CT number correction technique for subjectspecific pseudo-CT estimation in the context of MRI-only radiation therapy*. Journal of Medical Imaging and Radiation Sciences (JMIRS). Elsevier, 2019. https://doi.org/10.1016/j.jmir.2019.03.184

3. **Wafa Boukellouz** and Abdelouahab Moussaoui. *A review on methods to estimate a CT from MRI data in the context of MRI-alone RT*. Medical Technologies Journal, 2(1) :150–178, 2018. https://doi.org/10.26415/2572-004X-vol2iss1p150-178

Conference proceedings

1. **Wafa Boukellouz**, Abdelouahab Moussaoui, and Abdelmalik Taleb-Ahmed. *Pseudo-CT generated through multi-metric image registration and atlas fusion (application to T1-weighted brain MRI)*. In 2017 European Conference on Electrical Engineering and Computer Science (EECS), pages 230–235. IEEE, 2017. 10.1109/EECS.2017.50

2. **Wafa Boukellouz** and Abdelouahab Moussaoui. *Evaluation of several similarity measures for deformable image registration using T1-weighted MR images of the brain*. In 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B), 2017 , pages 1–5. IEEE, 2017. 10.1109/ICEE-B.2017.8192049

3. **Wafa Boukellouz** and Abdelouahab Moussaoui. *Classification of methods for generating pseudo-CT from MRI images for MRI-alone RT*. Medical Technologies Journal, 1(3):54–54, 2017. https://doi.org/10.26415/2572-004X-vol1iss3p54-54

Bibliography

Aljabar, Paul, Heckemann, Rolf A, Hammers, Alexander, Hajnal, Joseph V, and Rueckert, Daniel. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2009.

Andreasen, Daniel, Van Leemput, Koen, Hansen, Rasmus H, Andersen, Jon AL, and Edmund, Jens M. Patch-based generation of a pseudo ct from conventional mri sequences for mri-only radiotherapy of the brain. *Medical physics*, 42(4):1596–1605, 2015.

Andreasen, Daniel, Edmund, Jens M, Zografos, Vasileios, Menze, Bjoern H, and Van Leemput, Koen. Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features. In *Medical Imaging 2016: Image Processing*, volume 9784, page 978417. International Society for Optics and Photonics, 2016.

Aouadi, Souha, Vasic, Ana, Paloor, Satheesh, Torfeh, Tarraf, McGarry, Maeve, Petric, Primoz, Riyas, Mohamed, Hammoud, Rabih, and Al-Hammadi, Noora. Generation of synthetic ct using multi-scale and dual-contrast patches for brain mri-only external beam radiotherapy. *Physica Medica: European Journal of Medical Physics*, 42:174–184, 2017.

Arabi, Hossein, Koutsouvelis, Nikolaos, Rouzaud, Michel, Miralbell, Raymond, and Zaidi, Habib. Atlas-guided generation of pseudo-ct images for mri-only and hybrid pet–mri-guided radiotherapy treatment planning. *Physics in Medicine and Biology*, 61(17):6531, 2016.

Aseervatham, Sujeevan, Antoniadis, Anestis, Gaussier, Éric, Burlet, Michel, and Denneulin, Yves. A sparse version of the ridge logistic regression for large-scale text categorization. *Pattern Recognition Letters*, 32(2):101–106, 2011.

Badrinarayanan, Vijay, Kendall, Alex, and Cipolla, Roberto. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

Batchelor, Bruce G. *Pattern recognition: ideas in practice*. Springer Science & Business Media, 2012.

Beavis, AW, Gibbs, P, Dealey, RA, and Whitton, VJ. Radiotherapy treatment planning of brain tumours using mri alone. *The British journal of radiology*, 71(845):544–548, 1998.

Berker, Yannick, Franke, Jochen, Salomon, André, Palmowski, Moritz, Donker, Henk CW, Temur, Yavuz, Mottaghy, Felix M, Kuhl, Christiane, Izquierdo-Garcia, David, Fayad, Zahi A, et al. Mri-based attenuation correction for hybrid pet/mri systems: a 4-class tissue segmentation technique

using a combined ultrashort-echo-time/dixon mri sequence. *Journal of nuclear medicine*, 53(5): 796–804, 2012.

Bezdek, James C, Ehrlich, Robert, and Full, William. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

Boettger, Thomas, Nyholm, Tufve, Karlsson, Magnus, Nunna, Chandrasekhar, and Celi, Juan Carlos. Radiation therapy planning and simulation with magnetic resonance images. In *Medical Imaging*, pages 69181C–69181C. International Society for Optics and Photonics, 2008.

Boukellouz, Wafa and Moussaoui, Abdelouahab. Classification of methods for generating pseudo-ct from mri images for mri-alone rt. *Medical Technologies Journal*, 1(3):54–54, 2017a.

Boukellouz, Wafa and Moussaoui, Abdelouahab. Evaluation of several similarity measures for deformable image registration using t1-weighted mr images of the brain. In *In 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B) 2017*, pages 1–5. IEEE, 2017b.

Boukellouz, Wafa and Moussaoui, Abdelouahab. A review on methods to estimate a ct from mri data in the context of mri-alone rt. *Medical Technologies Journal*, 2(1):150–178, 2018.

Boukellouz, Wafa and Moussaoui, Abdelouahab. Magnetic resonance-driven pseudo ct image using patch-based multi-modal feature extraction and ensemble learning with stacked generalisation. *Journal of King Saud University - Computer and Information Sciences*, 2019. ISSN 1319-1578. doi: https://doi.org/10.1016/j.jksuci.2019.06.002.

Boukellouz, Wafa, Moussaoui, Abdelouahab, and Taleb-Ahmed, Abdelmalik. Pseudo-ct generated through multi-metric image registration and atlas fusion (application to t1-weighted brain mri). In *2017 European Conference on Electrical Engineering and Computer Science (EECS)*, pages 230–235. IEEE, 2017.

Boukellouz, Wafa, Moussaoui, Abdelouahab, Taleb-Ahmed, Abdelmalik, and Boydev, Christine. Multiatlas fusion with a hybrid ct number correction technique for subject-specific pseudo-ct estimation in the context of mri-only radiation therapy. *Journal of Medical Imaging and Radiation Sciences*, 2019.

Boydev, C, Demol, B, Pasquier, D, Saint-Jalmes, H, Delpon, G, and Reynaert, N. Zero echo time mri-only treatment planning for radiation therapy of brain tumors after resection. *Physica Medica: European Journal of Medical Physics*, 42:332–338, 2017.

Bredfeldt, Jeremy S, Liu, Lianli, Feng, Mary, Cao, Yue, and Balter, James M. Synthetic ct for mri-based liver stereotactic body radiotherapy treatment planning. *Physics in Medicine and Biology*, 62(8):2922, 2017.

Breiman, Leo. Random forests. *Machine learning*, 45(1):5–32, 2001.

Brock, Andrew, Donahue, Jeff, and Simonyan, Karen. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

Brown, Lisa Gottesfeld. A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376, 1992.

Burgos, Ninon, Cardoso, M Jorge, Thielemans, Kris, Modat, Marc, Pedemonte, Stefano, Dickson, John, Barnes, Anna, Ahmed, Rebekah, Mahoney, Colin J, Schott, Jonathan M, et al. Attenuation correction synthesis for hybrid pet-mr scanners: application to brain studies. *IEEE transactions on medical imaging*, 33(12):2332–2341, 2014.

Burgos, Ninon, Guerreiro, Filipa, McClelland, Jamie, Presles, Benoît, Modat, Marc, Nill, Simeon, Dearnaley, David, Oelfke, Uwe, Knopf, Antje-Christin, Ourselin, Sébastien, et al. Iterative framework for the joint segmentation and ct synthesis of mr images: application to mri-only radiotherapy treatment planning. *Physics in Medicine and Biology*, 62(11):4237, 2017.

Chandra, Arjun and Yao, Xin. Evolving hybrid ensembles of learning machines for better generalisation. *Neurocomputing*, 69(7-9):686–700, 2006.

Chang, Ni-Bin and Bai, Kaixu. *Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing*. CRC Press, 2018.

Chao, Ming, Xie, Yaoqin, and Xing, Lei. Auto-propagation of contours for adaptive prostate radiation therapy. *Physics in medicine and biology*, 53(17):4533, 2008.

Chen, Yasheng, Juttukonda, Meher, Lee, Yue Z, Su, Yi, Espinoza, Felipe, Lin, Weili, Shen, Dinggang, Lulash, David, and An, Hongyu. Mri based attenuation correction for pet/mri via mrf segmentation and sparse regression estimated ct. In *IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014*, pages 1364–1367. IEEE, 2014.

Dalal, Navneet and Triggs, Bill. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

Dar, Salman UH, Yurt, Mahmut, Karacan, Levent, Erdem, Aykut, Erdem, Erkut, and Çukur, Tolga. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 2019.

de Gonzalez, Amy Berrington and Darby, Sarah. Risk of cancer from diagnostic x-rays: estimates for the uk and 14 other countries. *The lancet*, 363(9406):345–351, 2004.

Delso, Gaspar, Wiesinger, Florian, Sacolick, Laura I, Kaushik, Sandeep S, Shanbhag, Dattesh D, Hüllner, Martin, and Veit-Haibach, Patrick. Clinical evaluation of zero-echo-time mr imaging for the segmentation of the skull. *Journal of Nuclear Medicine*, 56(3):417–422, 2015.

Demol, Benjamin, Boydev, Christine, Korhonen, Juha, and Reynaert, Nick. Dosimetric characterization of mri-only treatment planning for brain tumors in atlas-based pseudo-ct images generated from standard t1-weighted mr images. *Medical Physics*, 43(12):6557–6568, 2016.

Dempster, Arthur P. A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):205–232, 1968.

Despotović, Ivana, Goossens, Bart, and Philips, Wilfried. Mri segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, 2015, 2015.

Devic, Slobodan. Mri simulation for radiotherapy treatment planning. *Medical physics*, 39(11):6701–6711, 2012.

Dice, Lee R. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

Dinkla, Anna M, Wolterink, Jelmer M, Maspero, Matteo, Savenije, Mark HF, Verhoeff, Joost JC, Seravalli, Enrica, Išgum, Ivana, Seevinck, Peter R, and van den Berg, Cornelis AT. Mr-only brain radiotherapy: Dosimetric evaluation of synthetic cts generated by a dilated convolutional neural network. *International Journal of Radiation Oncology* Biology* Physics*, 2018.

Dixon, W Thomas. Simple proton spectroscopic imaging. *Radiology*, 153(1):189–194, 1984.

Dowling, Jason A, Lambert, Jonathan, Parker, Joel, Salvado, Olivier, Fripp, Jurgen, Capp, Anne, Wratten, Chris, Denham, James W, and Greer, Peter B. An atlas-based electron density mapping method for magnetic resonance imaging (mri)-alone treatment planning and adaptive mri-based prostate radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 83(1): e5–e11, 2012.

Edmund, Jens M and Nyholm, Tufve. A review of substitute ct generation for mri-only radiation therapy. *Radiation Oncology*, 12(1):28, 2017.

Eilertsen, Karsten, Nilsen Tor Arne Vestad, Line, Geier, Oliver, and Skretting, Arne. A simulation of mri based dose calculations on the basis of radiotherapy planning ct images. *Acta Oncologica*, 47 (7):1294–1302, 2008.

Emami, Hajar, Dong, Ming, Nejad-Davarani, Siamak P, and Glide-Hurst, Carri. Generating synthetic ct s from magnetic resonance images using generative adversarial networks. *Medical physics*, 2018.

Fu, Jie, Yang, Yingli, Singhrao, Kamal, Ruan, Dan, Low, Daniel A, and Lewis, John H. Male pelvic synthetic ct generation from t1-weighted mri using 2d and 3d convolutional neural networks. *arXiv preprint arXiv:1803.00131*, 2018.

Gabor, Dennis. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

García, Eloy, Oliver, Arnau, Diez, Yago, Diaz, Oliver, Lladó, Xavier, Martí, Robert, and Martí, Joan. Similarity metrics for intensity-based registration using breast density maps. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 217–225. Springer, 2017.

Ghose, Soumya, Dowling, Jason A, Rai, Robba, and Liney, Gary P. Substitute ct generation from a single ultra short time echo mri sequence: preliminary study. *Physics in Medicine and Biology*, 62 (8):2950, 2017a.

Ghose, Soumya, Greer, Peter B, Sun, Jidi, Pichler, Peter, Rivest-Henault, David, Mitra, Jhimli, Richardson, Haylea, Wratten, Chris, Martin, Jarad, Arm, Jameen, et al. Regression and statistical shape model based substitute ct generation for mri alone external beam radiation therapy from standard clinical mri sequences. *Physics in Medicine & Biology*, 62(22):8566, 2017b.

Glover, GH and Schneider, E. Three-point dixon technique for true water/fat decomposition with b0 inhomogeneity correction. *Magnetic resonance in medicine*, 18(2):371–383, 1991.

Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. *Deep learning*. MIT press, 2016.

Greer, Peter B, Dowling, Jason A, Lambert, Jonathon A, Fripp, Jurgen, Parker, Joel, Denham, James W, Wratten, Chris, Capp, Anne, and Salvado, Olivier. A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer. *Medical Journal of Australia*, 194(4): S24, 2011.

Gudur, Madhu Sudhan Reddy, Hara, Wendy, Le, Quynh-Thu, Wang, Lei, Xing, Lei, and Li, Ruijiang. A unifying probabilistic bayesian approach to derive electron density from mri for radiation therapy treatment planning. *Physics in medicine and biology*, 59(21):6595, 2014.

Haacke, E Mark, Brown, Robert W, Thompson, Michael R, Venkatesan, Ramesh, et al. *Magnetic resonance imaging: physical principles and sequence design*, volume 82. Wiley-Liss New York:, 1999.

Hall, David L and Llinas, James. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23, 1997.

Han, Xiao. Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical physics*, 44(4):1408–1419, 2017.

Hansen, Lars Kai and Salamon, Peter. Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):993–1001, 1990.

Hendee, William R and Ritenour, E Russell. *Medical imaging physics*. John Wiley & Sons, 2003.

Hofmann, Matthias, Steinke, Florian, Scheel, Verena, Charpiat, Guillaume, Farquhar, Jason, Aschoff, Philip, Brady, Michael, Schölkopf, Bernhard, and Pichler, Bernd J. Mri-based attenuation correction for pet/mri: a novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine*, 49(11):1875–1883, 2008.

Hsu, Shu-Hui, Cao, Yue, Huang, Ke, Feng, Mary, and Balter, James M. Investigation of a method for generating synthetic ct models from mri scans of the head and neck for radiation therapy. *Physics in medicine and biology*, 58(23):8419, 2013.

Hu, Yongsheng and Zhang, Liyi. Pseudo ct generation based on 3d group feature extraction and alternative regression forest for mri-only radiotherapy. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(06):1855009, 2018.

Huang, Szu-Hao, Chu, Yi-Hong, Lai, Shang-Hong, and Novak, Carol L. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE transactions on medical imaging*, 28(10):1595–1605, 2009.

Huynh, Tri, Gao, Yaozong, Kang, Jiayin, Wang, Li, Zhang, Pei, Lian, Jun, and Shen, Dinggang. Estimating ct image from mri data using structured random forest and auto-context model. *IEEE transactions on medical imaging*, 35(1):174–183, 2016.

Ito, Kengo, Kadoya, Noriyuki, Nakajima, Yujiro, Saito, Masahide, Sato, Kiyokazu, Nagasaka, Tatsuo, Yamanaka, Kazuomi, Dobashi, Suguru, Takeda, Ken, Matsushita, Haruo, et al. Feasibility of a direct-conversion method from magnetic susceptibility to relative electron density for radiation therapy treatment planning. *Oncology*, 6:252–265, 2017.

Johansson, Adam, Karlsson, Mikael, and Nyholm, Tufve. Ct substitute derived from mri sequences with ultrashort echo time. *Medical physics*, 38(5):2708–2714, 2011.

Johansson, Adam, Karlsson, Mikael, Yu, Jun, Asklund, Thomas, and Nyholm, Tufve. Voxel-wise uncertainty in ct substitute derived from mri. *Medical physics*, 39(6):3283–3290, 2012.

Johansson, Adam, Garpebring, Anders, Karlsson, Mikael, Asklund, Thomas, and Nyholm, Tufve. Improved quality of computed tomography substitute derived from magnetic resonance (mr) data by incorporation of spatial information–potential application for mr-only radiotherapy and attenuation correction in positron emission tomography. *Acta Oncologica*, 52(7):1369–1373, 2013.

Johnson, Hans J, McCormick, Matt, and Ibanez, Luis. The itk software guide third edition updated for itk version 4.5. *Insight Software Consortium, Tech. rep*, 2013.

Johnson, Justin, Gupta, Agrim, and Fei-Fei, Li. Image generation from scene graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1219–1228, 2018.

Johnstone, Emily, Wyatt, Jonathan J, Henry, Ann M, Short, Susan C, Sebag-Montefiore, David, Murray, Louise, Kelly, Charles G, McCallum, Hazel M, and Speight, Richard. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *International Journal of Radiation Oncology * Biology * Physics*, 100(1): 199–217, 2018.

Jonsson, Joakim H, Karlsson, Magnus G, Karlsson, Mikael, and Nyholm, Tufve. Treatment planning using mri data: an analysis of the dose calculation accuracy for different treatment regions. *Radiation Oncology*, 5(1):1, 2010.

Joseph, V Hanjal, Hajnal Derek, LG, and Hill David, J Hawkes. Medical image registration. *Biomedical Engineering Series. CRC Press LLC, Boca Raton*, 2001.

Kaiser, Bodo and Albarqouni, Shadi. MRI to CT translation with gans. *CoRR*, abs/1901.05259, 2019.

Kapanen, Mika and Tenhunen, Mikko. T1/t2*-weighted mri provides clinically relevant pseudo-ct density data for the pelvic bones in mri-only based radiotherapy treatment planning. *Acta Oncologica*, 52(3):612–618, 2013.

Kazemifar, Samaneh, McGuire, Sarah, Timmerman, Robert, Wardak, Zabi, Nguyen, Dan, Park, Yang, Jiang, Steve, and Owrangi, Amir. Mri-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic ct images generated using a deep learning approach. *Radiotherapy and Oncology*, 136: 56–63, 2019.

Kearns, Michael and Valiant, Leslie. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, 41(1):67–95, 1994.

Khaleghi, Bahador, Khamis, Alaa, Karray, Fakhreddine O, and Razavi, Saiedeh N. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1):28–44, 2013.

Khalifé, Maya, Fernandez, Brice, Jaubert, Olivier, Soussan, Michael, Brulon, Vincent, Buvat, Irène, and Comtat, Claude. Subject-specific bone attenuation correction for brain pet/mr: can zte-mri substitute ct scan accurately? *Physics in Medicine & Biology*, 62(19):7814, 2017.

Khateri, Parisa, Rad, Hamidreza Saligheh, Jafari, Amir Homayoun, and Ay, Mohammad Reza. A novel segmentation approach for implementation of mrac in head pet/mri employing short-te mri and 2-point dixon method in a fuzzy c-means framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 734: 171–174, 2014.

Khateri, Parisa, Rad, Hamidreza Saligheh, Jafari, Amir Homayoun, Kazerooni, Anahita Fathi, Akbarzadeh, Afshin, Moghadam, Mohsen Shojae, Aryan, Arvin, Ghafarian, Pardis, and Ay, Mohammad Reza. Generation of a four-class attenuation map for mri-based attenuation correction of pet data in the head area using a novel combination of ste/dixon-mri and fcm clustering. *Molecular Imaging and Biology*, 17(6):884–892, 2015.

Kim, Joong Hyun, Lee, Jae Sung, Song, In-Chan, and Lee, Dong Soo. Comparison of segmentation-based attenuation correction methods for pet/mri: evaluation of bone and liver standardized uptake value with oncologic pet/ct data. *Journal of Nuclear Medicine*, 53(12):1878–1882, 2012.

Kim, Joshua, Glide-Hurst, Carri, Doemer, Anthony, Wen, Ning, Movsas, Benjamin, and Chetty, Indrin J. Implementation of a novel algorithm for generating synthetic ct images from magnetic resonance imaging data sets for prostate cancer radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 91(1):39–47, 2015.

Klein, Stefan, Staring, Marius, Murphy, Keelin, Viergever, Max A, and Pluim, Josien PW. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29 (1):196–205, 2010.

Knutsson, Hans and Andersson, Mats. Morphons: Segmentation using elastic canvas and paint on priors. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–1226. IEEE, 2005.

Kops, Elena Rota and Herzog, Hans. Alternative methods for attenuation correction for pet images in mr-pet scanners. In *2007 IEEE Nuclear Science Symposium Conference Record*, volume 6, pages 4327–4330. IEEE, 2007.

Korhonen, Juha, Kapanen, Mika, Keyriläinen, Jani, Seppälä, Tiina, and Tenhunen, Mikko. A dual model hu conversion from mri intensity values within and outside of bone segment for mri-based radiotherapy treatment planning of prostate cancer. *Medical physics*, 41(1):011704, 2014.

Korsholm, Marie E, Waring, Line W, and Edmund, Jens M. A criterion for the reliable use of mri-only radiotherapy. *Radiation Oncology*, 9(1):16, 2014.

Kraus, Kim Melanie, Jäkel, Oliver, Niebuhr, Nina I, and Pfaffenberger, Asja. Generation of synthetic ct data using patient specific daily mr image data and image registration. *Physics in medicine and biology*, 62(4):1358, 2017.

Lahat, Dana, Adali, Tülay, and Jutten, Christian. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.

Lambert, Jonathan, Greer, Peter B, Menk, Fred, Patterson, Jackie, Parker, Joel, Dahl, Kara, Gupta, Sanjiv, Capp, Anne, Wratten, Chris, Tang, Colin, et al. Mri-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of mri-based dose planning. *Radiotherapy and Oncology*, 98(3):330–334, 2011.

Largent, Axel, Nunes, Jean-Claude, Saint-Jalmes, Hervé, Simon, Antoine, Perichon, Nicolas, Barateau, Anais, Hervé, Chloé, Lafond, Caroline, Greer, Peter B, Dowling, Jason A, et al. Pseudo-ct generation by conditional inference random forest for mri-based radiotherapy treatment planning. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 46–50. IEEE, 2017.

Lee, Young K, Bollet, Marc, Charles-Edwards, Geoffrey, Flower, Maggie A, Leach, Martin O, McNair, Helen, Moore, Elizabeth, Rowbottom, Carl, and Webb, Steve. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiotherapy and oncology*, 66(2):203–216, 2003.

Lei, Yang, Harms, Joseph, Wang, Tonghe, Liu, Yingzi, Shu, Hui-Kuo, Jani, Ashesh B., Curran, Walter J., Mao, Hui, Liu, Tian, and Yang, Xiaofeng. Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. *Medical Physics*, 2019a.

Lei, Yang, Harms, Joseph, Wang, Tonghe, Tian, Sibo, Zhou, Jun, Shu, Hui-Kuo, Zhong, Jim, Mao, Hui, Curran, Walter J, Liu, Tian, et al. Mri-based synthetic ct generation using semantic random forest with iterative refinement. *Physics in Medicine & Biology*, 64(8):085001, 2019b.

Lester, Hava and Arridge, Simon R. A survey of hierarchical non-linear medical image registration. *Pattern recognition*, 32(1):129–149, 1999.

Leynes, Andrew P, Yang, Jaewon, Shanbhag, Dattesh D, Kaushik, Sandeep S, Seo, Youngho, Hope, Thomas A, Wiesinger, Florian, and Larson, Peder EZ. Hybrid zte/dixon mr-based attenuation correction for quantitative uptake estimation of pelvic lesions in pet/mri. *Medical physics*, 44(3): 902–913, 2017a.

Leynes, Andrew P, Yang, Jaewon, Wiesinger, Florian, Kaushik, Sandeep S, Shanbhag, Dattesh D, Seo, Youngho, Hope, Thomas A, and Larson, Peder EZ. Direct pseudoct generation for pelvis pet/mri attenuation correction using deep convolutional neural networks with multi-parametric mri: zero echo-time and dixon deep pseudoct (zedd-ct). *Journal of Nuclear Medicine*, pages jnumed–117, 2017b.

Lin, Hsuan-Tien and Li, Ling. Support vector machinery for infinite ensemble learning. *Journal of Machine Learning Research*, 9(Feb):285–312, 2008.

Liu, Fang, Jang, Hyungseok, Kijowski, Richard, Bradshaw, Tyler, and McMillan, Alan B. Deep learning mr imaging–based attenuation correction for pet/mr imaging. *Radiology*, 286(2):676–684, 2017a.

Liu, Lianli, Jolly, Shruti, Cao, Yue, Vineberg, Karen, Fessler, Jeffrey A, and Balter, James M. Female pelvic synthetic ct generation based on joint intensity and shape analysis. *Physics in Medicine and Biology*, 62(8):2935, 2017b.

Liu, Yong and Yao, Xin. Ensemble learning via negative correlation. *Neural networks*, 12(10):1399–1404, 1999.

Low, Daniel A, Harms, William B, Mutic, Sasa, and Purdy, James A. A technique for the quantitative evaluation of dose distributions. *Medical physics*, 25(5):656–661, 1998.

Maintz, JB Antoine and Viergever, Max A. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.

Mallat, Stephane G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

Maspero, Matteo, Savenije, Mark HF, Dinkla, Anna M, Seevinck, Peter R, Intven, Martijn PW, Jurgenliemk-Schulz, Ina M, Kerkmeijer, Linda GW, and van den Berg, Cornelis AT. Dose evaluation of fast synthetic-ct generation using a generative adversarial network for general pelvis mr-only radiotherapy. *Physics in Medicine & Biology*, 63(18):185001, 2018.

Mehranian, Abolfazl, Arabi, Hossein, and Zaidi, Habib. Quantitative analysis of mri-guided attenuation correction techniques in time-of-flight brain pet/mri. *Neuroimage*, 130:123–133, 2016.

Mérida, Inés, Costes, Nicolas, Heckemann, Rolf A, Drzezga, Alexander, Förster, Stefan, and Hammers, Alexander. Evaluation of several multi-atlas methods for pseudo-ct generation in brain mri-pet attenuation correction. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015*, pages 1431–1434. IEEE, 2015.

Michalski, Ryszard S, Stepp, Robert E, and Diday, Edwin. A recent advance in data analysis: Clustering objects into classes characterized by conjunctive concepts. In *Progress in pattern recognition*, pages 33–56. Elsevier, 1981.

Mitchell, Tom M. *Machine learning*. McGraw Hill series in computer science. McGraw-Hill, 1997. ISBN 978-0-07-042807-2.

Moser, Ewald, Stadlbauer, Andreas, Windischberger, Christian, Quick, Harald H, and Ladd, Mark E. Magnetic resonance imaging methodology. *European journal of nuclear medicine and molecular imaging*, 36(1):30–41, 2009.

Naidu, VPS and Raol, Jitendra R. Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal*, 58(3):338–352, 2008.

Navalpakkam, Bharath K, Braun, Harald, Kuwert, Torsten, and Quick, Harald H. Magnetic resonance–based attenuation correction for pet/mr hybrid imaging using continuous valued attenuation maps. *Investigative radiology*, 48(5):323–332, 2013.

Nixon, Mark and Aguado, Alberto S. *Feature extraction and image processing for computer vision*. Academic Press, 2012.

Nyholm, Tufve, Nyberg, Morgan, Karlsson, Magnus G, and Karlsson, Mikael. Systematisation of spatial uncertainties for comparison between a mr and a ct-based radiotherapy workflow for prostate treatments. *Radiation Oncology*, 4(1):1, 2009.

Ojala, Timo, Pietikainen, Matti, and Maenpaa, Topi. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

O'Neil, BDP, Salerno, G, Thomas, K, Tait, DM, and Brown, G. Mr vs ct imaging: low rectal cancer tumour delineation for three-dimensional conformal radiotherapy. *British journal of radiology*, 82 (978):509–513, 2009.

O'Neill, BDP, Salerno, G, Thomas, K, Tait, DM, and Brown, G. Mr vs ct imaging: low rectal cancer tumour delineation for three-dimensional conformal radiotherapy. *The British journal of radiology*, 2014.

Onozato, Yusuke, Kadoya, Noriyuki, Fujita, Yukio, Arai, Kazuhiro, Dobashi, Suguru, Takeda, Ken, Kishi, Kazuma, Umezawa, Rei, Matsushita, Haruo, and Jingu, Keiichi. Evaluation of on-board kv cone beam computed tomography–based dose calculation with deformable image registration using hounsfield unit modifications. *International Journal of Radiation Oncology* Biology* Physics*, 89 (2):416–423, 2014.

Paisitkriangkrai, Sakrapee, Shen, Chunhua, and van den Hengel, Anton. Pedestrian detection with spatially pooled features and structured ensemble learning. *IEEE transactions on pattern analysis and machine intelligence*, 38(6):1243–1257, 2016.

Pasquier, D, Betrouni, N, Vermandel, M, Lacornerie, T, Lartigau, E, and Rousseau, J. Mri alone simulation for conformal radiation therapy of prostate cancer: technical aspects. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 160–163. IEEE, 2006.

Peng, Zhigang, Zhong, Jia, Wee, William, and Lee, Jing-huei. Automated vertebra detection and segmentation from the whole spine mr images. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 2527–2530. IEEE, 2006.

Podgorsak, Ervin B et al. Radiation oncology physics. *Vienna: International Atomic Energy Agency*, pages 123–271, 2005.

Rank, Christopher M, Tremmel, Christoph, Hünemohr, Nora, Nagel, Armin M, Jäkel, Oliver, and Greilich, Steffen. Mri-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. *Radiation Oncology*, 8(1):1, 2013.

Rasch, Coen, Barillot, Isabelle, Remeijer, Peter, Touw, Adriaan, van Herk, Marcel, and Lebesque, Joos V. Definition of the prostate in ct and mri: a multi-observer study. *International Journal of Radiation Oncology* Biology* Physics*, 43(1):57–66, 1999.

Ren, Shangjie, Hara, Wendy, Wang, Lei, Buyyounouski, Mark K, Le, Quynh-Thu, Xing, Lei, and Li, Ruijiang. Robust estimation of electron density from anatomic magnetic resonance imaging of the brain using a unifying multi-atlas approach. *International Journal of Radiation Oncology* Biology* Physics*, 97(4):849–857, 2017.

Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Roy, Snehashis, Wang, Wen-Tung, Carass, Aaron, Prince, Jerry L, Butman, John A, and Pham, Dzung L. Pet attenuation correction using synthetic ct from ultrashort echo-time mr imaging. *Journal of Nuclear Medicine*, 55(12):2071–2077, 2014.

Rueckert, Daniel and Schnabel, Julia A. Medical image registration. In *Biomedical Image Processing*, pages 131–154. Springer, 2010.

Rueckert, Daniel, Sonoda, Luke I, Hayes, Carmel, Hill, Derek LG, Leach, Martin O, and Hawkes, David J. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

Rumelhart, David E, Hinton, Geoffrey E, and Williams, Ronald J. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Schapire, Robert E. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.

Schreibmann, Eduard, Nye, Jonathon A, Schuster, David M, Martin, Diego R, Votaw, John, and Fox, Tim. Mr-based attenuation correction for hybrid pet-mr brain imaging systems using deformable image registration. *Medical physics*, 37(5):2101–2109, 2010.

Shafer, Glenn. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

Siversson, Carl, Nordström, Fredrik, Nilsson, Terese, Nyholm, Tufve, Jonsson, Joakim, Gunnlaugsson, Adalsteinn, and Olsson, Lars E. Technical note: Mri only prostate radiotherapy planning using the statistical decomposition algorithm. *Medical physics*, 42(10):6090–6097, 2015.

Sjölund, Jens, Forsberg, Daniel, Andersson, Mats, and Knutsson, Hans. Generating patient specific pseudo-ct of the head from mr using atlas-based regression. *Physics in medicine and biology*, 60(2): 825, 2015.

Škerl, Darko, Likar, Boštjan, and Pernuš, Franjo. A protocol for evaluation of similarity measures for non-rigid registration. *Medical image analysis*, 12(1):42–54, 2008.

Smola, Alex and Vishwanathan, SVN. Introduction to machine learning. *Cambridge University, UK*, 32:34, 2008.

Song, Shuang, Zheng, Yuanjie, and He, Yunlong. A review of methods for bias correction in medical images. *Biomedical Engineering Review*, 1(1), 2017.

Spadea, Maria Francesca, Pileggi, Giampaolo, Zaffino, Paolo, Salome, Patrick, Catana, Ciprian, Izquierdo-Garcia, David, Amato, Francesco, and Seco, Joao. Deep convolution neural network (dcnn) multi-plane approach to synthetic ct generation from mr images-application in brain proton therapy. *International Journal of Radiation Oncology* Biology* Physics*, 2019. doi: doi:10.1016/j.ijrobp.2019.06.2535.

Speier, Christoph, Pileggi, Giampaolo, Izquierdo-Garcia, David, Catana, Ciprian, Sharp, Gregory C, Spadea, Maria Francesca, Bert, Christoph, and Seco, Joao. Advanced multimodal methods for cranial pseudo-ct generation validated by imrt and vmat radiation therapy plans. *International Journal of Radiation Oncology* Biology* Physics*, 102(4):792–800, 2018.

Stanescu, T, Jans, HS, Pervez, N, Stavrev, P, and Fallone, BG. A study on the magnetic resonance imaging (mri)-based radiation treatment planning of intracranial lesions. *Physics in medicine and biology*, 53(13):3579, 2008.

Su, Kuan-Hao, Hu, Lingzhi, Stehning, Christian, Helle, Michael, Qian, Pengjiang, Thompson, Cheryl L, Pereira, Gisele C, Jordan, David W, Herrmann, Karin A, Traughber, Melanie, et al. Generation of brain pseudo-cts using an undersampled, single-acquisition ute-mdixon pulse sequence and unsupervised clustering. *Medical physics*, 42(8):4974–4986, 2015.

Su, Kuan-Hao, Friel, Harry T, Kuo, Jung-Wen, Al Helo, Rose, Baydoun, Atallah, Stehning, Christian, Crisan, Adina N, Traughber, Melanie S, Devaraj, Ajit, Jordan, David W, et al. Ute-mdixon-based thorax synthetic ct generation. *Medical physics*, 2019.

Thörnqvist, Sara, Petersen, Jørgen BB, Høyer, Morten, Bentzen, Lise N, and Muren, Ludvig Paul. Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration. *Acta Oncologica*, 49(7):1023–1032, 2010.

Tustison, N and Gee, J. N4itk: Nick's n3 itk implementation for mri bias field correction. *Insight Journal*, 9, 2009.

Uh, Jinsoo, Merchant, Thomas E, Li, Yimei, Li, Xingyu, and Hua, Chiaho. Mri-based treatment planning with pseudo ct generated through atlas registration. *Medical physics*, 41(5):051711, 2014.

Ulin, Kenneth, Urie, Marcia M, and Cherlow, Joel M. Results of a multi-institutional benchmark test for cranial ct/mr image registration. *International Journal of Radiation Oncology* Biology* Physics*, 77(5):1584–1589, 2010.

Van Der Put, RW, Kerkhof, EM, Raaymakers, BW, Jürgenliemk-Schulz, IM, and Lagendijk, JJW. Contour propagation in mri-guided radiotherapy treatment of cervical cancer: the accuracy of rigid, non-rigid and semi-automatic registrations. *Physics in medicine and biology*, 54(23):7135, 2009.

Vapnik, Vladimir, Golowich, Steven E., and Smola, Alexander J. Support vector method for function approximation, regression estimation and signal processing. In *NIPS*, pages 281–287. MIT Press, 1996.

Walker, Amy, Liney, Gary, Metcalfe, Peter, and Holloway, Lois. Mri distortion: considerations for mri based radiotherapy treatment planning. *Australasian Physical & Engineering Sciences in Medicine*, 37:103–113, 2014.

Wang, Shuangling, Yin, Yilong, Cao, Guibao, Wei, Benzheng, Zheng, Yuanjie, and Yang, Gongping. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, 149:708–717, 2015.

Werbos, PJ. Beyond regression: New tools for prediction and analysis in the behavioral sciences. ph. d. thesis, harvard university, cambridge, ma, 1974. 1974.

White, Franklin E. Data fusion lexicon. Technical report, Joint Directors of Labs Washington DC, 1991.

Wiesinger, Florian, Sacolick, Laura I, Menini, Anne, Kaushik, Sandeep S, Ahn, Sangtae, Veit-Haibach, Patrick, Delso, Gaspar, and Shanbhag, Dattesh D. Zero te mr bone imaging in the head. *Magnetic resonance in medicine*, 75(1):107–114, 2016.

Wollenweber, SD, Ambwani, S, Lonn, AHR, Shanbhag, DD, Thiruvenkadam, S, Kaushik, S, Mullick, R, Qian, H, Delso, G, and Wiesinger, F. Comparison of 4-class and continuous fat/water methods for whole-body, mr-based pet attenuation correction. *IEEE Transactions on Nuclear Science*, 60 (5):3391–3398, 2013.

Wolpert, David H. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.

Wolpert, David H. The lack of a priori distinctions between learning algorithms. *Neural computation*, 8(7):1341–1390, 1996.

Wu, Yao, Yang, Wei, Lu, Lijun, Lu, Zhentai, Zhong, Liming, Yang, Ru, Huang, Meiyan, Feng, Yanqiu, Chen, Wufan, and Feng, Qianjin. Prediction of ct substitutes from mr images based on local sparse correspondence combination. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 93–100. Springer, 2015.

Xiang, Lei, Wang, Qian, Nie, Dong, Zhang, Lichi, Jin, Xiyao, Qiao, Yu, and Shen, Dinggang. Deep embedding convolutional neural network for synthesizing ct image from t1-weighted mr image. *Medical image analysis*, 47:31–44, 2018.

Yang, Xiaofeng, Lei, Yang, Shu, Hui-Kuo, Rossi, Peter, Mao, Hui, Shim, Hyunsuk, Curran, Walter J, and Liu, Tian. Pseudo ct estimation from mri using patch-based random forest. In *Medical Imaging 2017: Image Processing*, volume 10133, page 101332Q. International Society for Optics and Photonics, 2017.

Zadeh, Lotfi Asker. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1(1):3–28, 1978.

Zaidi, Habib, Montandon, Marie-Louise, and Slosman, Daniel O. Magnetic resonance imaging-guided attenuation and scatter corrections in three-dimensional brain positron emission tomography. *Medical physics*, 30(5):937–948, 2003.

Zareapoor, Masoumeh, Shamsolmoali, Pourya, et al. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia computer science*, 48(2015):679–685, 2015.

Zhang, Zizhao, Xie, Yuanpu, and Yang, Lin. Photographic text-to-image synthesis with a hierarchically-nested adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6199–6208, 2018.

Zhen, Xiantong, Wang, Zhijie, Yu, Mengyang, and Li, Shuo. Supervised descriptor learning for multi-output regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1211–1218, 2015.

Zhong, Liming, Lin, Liyan, Lu, Zhentai, Wu, Yao, Lu, Zixiao, Huang, Meiyan, Yang, Wei, and Feng, Qianjing. Predict ct image from mri data using knn-regression with learned local descriptors. In *IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016*, pages 743–746. IEEE, 2016.

Zhou, Zhi-Hua. *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2012.

Zitova, Barbara and Flusser, Jan. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

Appendices

Appendix A. Image pre-processing and software

Image pre-processing

Image pre-processing is a critical step that needs to be performed before proceeding to building any system that relies on the use of images. Image pre-processing uses specific algorithms to enhance an image characteristics and supress any undesired distortions or noise. In MR imaging, noise or distortions can be caused by magnet field inhomogeneity or patient-related movements. Therefore, a correction is necessary, among the image pre-processing techniques that were used to correct for MR images employed in this work are the following :

Bias Correction

A low frequency intensity non-uniformity can be present in an image in the form of a bias and characterised by a non-uniformity in image intensities. There are several methods for bias correction such phantom-based calibration, filtering based, segmentation based method and histogram based methods Song et al. [2017]. In this work, we employed the N4 bias correction method of Tustison and Gee [2009] belonging to the latter category. It seeks to correct for the bias by using a non-parametric non-uniform normalisation of the intensity distribution and employs B-spline for generalisation. This is performed in an iterative fashion and uses results of the previous iterations for correction.

Histogram matching

Histogram matching was applied to MR image in order to bring them to the same intensity range. The basic idea behind histogram matching is to normalise the grayscale values of a source image relying on the grayscale values of a reference image. A histogram is a distribution of grayscale values frequencies in the range [0;255] of an image.

For an image I and a grayscale value G_1, its probability is given by:

$$P_{G_1}(G_1) = \frac{n_i}{n}$$

where, n_i is the frequency of a grayscale value G_{1i} and n is the total number of pixels in an image. The probability of the target grayscale value based on the reference image R is given by: $P_{G_2}(G_2)$. To transform $P_{G_1}(G_1)$ to $P_{G_2}(G_2)$, the cumulative distribution function for each grayscale value is defined as:

$$F1(G_{1k}) = \sum_{j=0}^{k} P_{G_1}(G_{1j}), \qquad k = 0, 1, ..., L.$$

$$F2(G_{2k}) = \sum_{j=0}^{k} P_{G_2}(G_{2j}), \qquad k = 0, 1, ..., L.$$

where, L is the total number of gray levels. The objective is to find a certain mapping function M such that $M(G_1) = G_2$.

Software

Implementation of the presented approaches in part II was performed on a dual core CPU 2.4 GHz with 8 GB of RAM. The following software was used :

1. Fiji imageJ[1] : was used to remove the imaging bed from CT scans ;
2. 3D Slicer[2] : was used for image pre-processing including histogram matching and bias correction ;
3. Elastix[3] : was employed for all the types of image registration tasks included in this dissertation ;
4. Matlab[4] : was used for implementing the two proposed approaches for pCT image synthesis including the hybrid correction technique, image fusion, machine learning algorithm's traning and stacking ;
5. VViewer[5] : was used to visualise MR and CT images and create the difference images.

[1] `https://fiji.sc/`
[2] `https://www.slicer.org`
[3] `http://elastix.isi.uu.nl/`
[4] `https://www.mathworks.com`
[5] `https://www.creatis.insa-lyon.fr/rio/vv`

Abstract
Nowadays, information about a phenomenon can be acquired using various modalities; and this led to the emergence of the concept of multi-modal data fusion. In fact, one of the application domains that strongly relies on multi-modal data acquisition is radiation therapy (RT), in which computed tomography (CT) imaging and magnetic resonance imaging (MRI) are the main modalities employed. The latter offers superior soft-tissue visualisation, hence it is used for tumour contouring, whereas the former contains the necessary electron density information for radiation dose computation. However, the pipeline of RT treatment planning presents many flaws from which is the excessive radiation exposure due to repetitive CT acquisitions. Recently, an idea to remediate for this main limitation and others has emerged; it consists in synthesising a CT image called pseudo-CT (pCT) from MRI images. In this dissertation, we aim to meet this objective by analysing existing work and proposing two variants of pCT image estimation methods employing multi-modal data fusion and machine learning. The first approach is designated hybrid; and combines image registration, unsupervised learning, image fusion and a new Hounsfield Unit (HU) values correction technique. This approach aims to predict a pCT image from T2-weighted MR images while attempting to reduce prediction error by using information from MR images. The second approach employs multi-modal shape, texture and spatial feature extraction, reduction and fusion of T1-weighted and T2-weighted MR images. In addition, we designed an ensemble learning model with stacked generalisation that builds a mapping from these features to HU values over two levels of learning. Results showed a significant improvement achieved by the proposed methods.
Keywords: multi-modality; data fusion; machine learning; medical images; computed tomography; magnetic resonance imaging.

Résumé
De nos jours, des informations sur un phénomène peuvent être acquises à travers diverses modalités ; ceci a conduit à l'émergence du concept de la fusion de données multimodales. L'un des domaines d'application qui s'appuie fortement sur l'acquisition de données multimodales est la radiothérapie (RT), dans laquelle l'imagerie par la tomodensitométrie (TDM) et l'imagerie par résonance magnétique (IRM) sont les deux principales modalités utilisées. Cette dernière offre une visualisation supérieure des tissus mous, elle est donc utilisée pour la définition de tumeur, tandis que la première contient les informations nécessaires sur la densité d'électron pour le calcul de dose d'irradiation. Cependant, ce processus de planification du traitement de la RT présente de nombreux défauts, parmi lesquels l'exposition excessive aux rayonnements due aux acquisitions répétitives par TDM. Récemment, une idée pour remédier à cette limitation principale et d'autres a émergée ; elle consiste à synthétiser une image TDM appelée pseudo-tomodensitométrie (pTDM) à partir d'images RM. Dans cette thèse, nous cherchons à atteindre cet objectif en analysant les travaux existants et en proposant deux variantes des méthodes d'estimation d'image pTDM utilisant la fusion de données multimodale et l'apprentissage automatique. La première approche est dite hybride ; elle combine le recalage d'image, l'apprentissage non supervisé, la fusion d'image et une nouvelle technique de correction des valeurs Hounsfield Unit (HU). Cette approche vise à prédire une image pTDM à partir des images RM T2-pondérée tout en essayant de réduire les erreurs de prédiction en utilisant des informations provenant des images RM. La deuxième approche utilise la fusion des caractéristiques sur la forme, la texture et l'information spatiale multimodales d'images RM T1-pondérée et T2-pondérée. De plus, nous avons conçu un modèle d'apprentissage d'ensemble avec une généralisation empilée qui crée une correspondance de ces caractéristiques aux valeurs HU sur deux niveaux d'apprentissage. Les résultats ont montré une amélioration significative obtenue par les méthodes proposées.
Mots-clés: multi-modalité; la fusion des données; apprentissage automatique; images médicales; tomodensitométrie; imagerie par résonance magnétique.

ملخص
في الوقت الحاضر، يمكن الحصول على معلومات حول ظاهرة ما باستعمال أنماط مختلفة؛ مما ادى الى ظهور مفهوم دمج البيانات متعدد الأنماط. أحد المجالات التطبيقية التي ترتكز على تحصيل بيانات متعددة الأنماط هو العلاج بالأشعة (RT) حيث ، التصوير المقطعي (CT) والتصوير بالرنين المغناطيسي (MRI) هما الطريقتان الرئيسيتان المستخدمتان. هذه الأخيرة تزوّد تصوير مرئي عالي للأنسجة الرخوة، وبالتالي فهي تستخدم لتحديد الورم، في حين تحتوي الأولى على المعلومات اللازمة حول كثافة الإلكترونات اللازمة لحساب جرعة الإشعاع. ومع ذلك، العملية المتبعة لتخطيط العلاج بالأشعة تتخللها العديد من العيوب التي تتمثل في التعرض المفرط للإشعاع خلال عمليات التحصيل المتكرر على للأشعة المقطعية. في الآونة الأخيرة، ظهرت فكرة لتجنب هذا القيد الرئيسي وغيره. و المتمثلة في استخراج صورة CT ما تسمى شبه صورة مقطعية (pseudo-CT) من صور التصوير بالرنين المغناطيسي. في هذه الأطروحة، نهدف إلى تحقيق هذا الهدف من خلال تحليل العمل الموجود واقتراح نوعين من طرق استخراج صور pCT باستخدام دمج البيانات المتعددة الوسائط والتعلم الآلي. الطريقة الاولى هجينة تجمع بين مطابقة الصور، التعلم غير الخاضع للإشراف، دمج الصور، وتقنية جديدة لتصحيح قيم وحدة هاونسفيلد (HU) للصور المقطعية. تهدف هذه الطريقة إلى استخراج صور pCT من التصوير بالرنين المغناطيسي من نوع T2-weighted مع محاولة تقليل خطأ التنبؤ من خلال توظيف معلومات من صور الرنين المغناطيسي. الطريقة الثانية لاستخراج صورة pCT تستخدم دمج ميزات الشكل، الملمس والمعلومات الموضعية متعددة الأنماط لصور الرنين المغناطيسي من نوعي T1-weighted و T2-weighted. بالإضافة إلى ذلك، قمنا بتصميم نموذج تعلمي بالمجموعات بحيث يسمح بتكوين رابط بين هذه الميزات وقيم HU على مستويين من التعلم. أظهرت النتائج تحسنًا كبيرًا حققته الطرق المقترحة.
كلمات مفتاحية: تعدد الأنماط; دمج البيانات; تعلم الآلة; الصور الطبية; التصوير المقطعي; التصوير بالرنين المغناطيسي.