LUND UNIVERSITY

Master of Science
Thesis
VT2023

Correlation between Surface and Tumour Motion in Lung Cancer

- including Deep Learning Perspectives

Caisa Kjellström

Supervisors

Sofie Ceberg, Annika Mannerberg, Mustafa Kahdhim, Emilia Persson, and Jens Engleson

Medical Radiation Physics, Lund
Faculty of Science
Lund University
www.msf.lu.se

Att Hantera Andningsrörelse Vid Strålbehandling av Lungcancer

Populärvetenskaplig sammanfattning

Av de 70 000 personerna som blev diagnoserade med cancer i Sverige 2021, fick runt 4000 lungcancer. Lungcancer är vanligen elakartade celler som utgår från lungans slemhinna. Behandling innefattar oftast operation eller strålterapi, ibland i kombination med cellgifter eller immunterapi.

Inför strålbehandling gör patienten en röntgenundersökning som ger bilder i tre dimensioner, en så kallad datortomografiundersökning. Läkarna använder sedan dessa bilder för att markera ut området som man vill bestråla, och organ runt om som man vill undvika att bestråla. Efter detta görs en dosplan där behandlingen simuleras i ett datorprogram för att göra en så optimerad behandling som möjligt. Där kan parametrar som strålfältstorlek, strålslag och ingångsvinkel bestämmas för att ge en hög och jämn fördelning av strålningen till tumören och så lite som möjligt till intilliggande frisk vävnad.

Vid behandling ligger patienten på en brits under en linjäraccelerator som producerar joniserande strålning. För lungcancer används ibland en speciell teknik som ger mycket hög dos till liten volym med mycket snabbt dosfall i intilliggande vävnad. Detta kallas stereotaktisk strålbehandling. Vid denna typ av behandling är det mycket viktigt att patienten ligger helt still.

Om tumören ligger i bröstkorgen, vilket är fallet för lungcancer, kan den röra sig under behandling på grund av andning. Risken är då att strålningen missar tumören om den rör sig med andningen. Läkarna kompenserar ofta för detta genom att göra det område som ska behandlas något större än den faktiskta tumören baserat på datortomografiundersökningen. Detta säkerställer att tumören bestrålas trots rörelsen. En svårighet som kan uppstå är att vi ofta andas olika och rörelsen som uppstår kan variera mycket för samma person vid olika tillfällen. Detta innebär att det finns en risk att patienten andas annorlunda under behandling jämfört med under datortomografiundersökningen.

Vi behöver därför verktyg för att kontrollera hur andningen ser ut före och under behandling. Ett verktyg för att göra detta är optiska ytskanningssystem. Dessa system mäter hur mycket bröstkorgen rör sig och kan eventuellt ge en indikation på tumörrörelsen. Eftersom man inte ser själva tumören blir detta alltid en uppskattning, och för att bli säkrare på att tumören rör sig som förväntat med bröstkorgen eller magen kan en statistisk korrelation göras mellan dessa.

Detta arbete har undersökt denna faktiska statistiska korrelationen mellan bröstkorgens rörelse och tumörens rörelse i kroppen på lungcancerpatienter. Som hjälp har jag använt data från optiska ytskanningssystem samt datortomografibilder och undersökt hur många millimeter som de båda rör sig. Som tillägg till detta har jag även gjort en datormodell som ska hjälpa att förutspå var tumören kommer att ligga utifrån bröstkorgsrörelsen.

Abstract

Purpose: The purpose of this master thesis was to retrospectively investigate correlation between surface and tumour motion in lung cancer patients, alongside deep learning applications of the results. Additional correlations such as age, tumour volume and anatomical placement of the tumour were also investigated.

Materials and Methods: 48 lung cancer patients treated with Stereotactic Body Radiation Therapy (SBRT) were included in this study. Delineation of the tumours was made on 4-Dimensional Computed Tomography (4DCT) images where each tumour was delineated in each of the eight respiration phases. Tumour volume and centre of mass coordinates in Left-Right (LR), Anterior-Posterior (AP) and Superior-Inferior (SI) directions were retrieved in all respiration phases. The total translational shift from maximum exhale phase was computed. Surface motion data was acquired from a surface imaging system which was recorded during 4DCT simulation of the patient.

The Spearman Correlation Coefficient (SCC) between tumour and surface motion was calculated. An additional Spearman correlation was calculated between the SCC and patient features (age, tumour volume, smallest distance to thoracic wall, distance to thoracic spine and to chest surface, lung volume, and use of abdominal compression belt). Wilcoxon's rank sum test was performed to determine statistical significance between the groups with and without abdominal compression belt.

An artificial neural network (ANN) model was created to investigate the possibility to predict tumour motion given the surface motion as an input to the model. The model was built like a feed-forward ANN with two hidden layers using Rectified Linear Unit (ReLU) activation functions. Training, validation, and test data set split were 36/45, 6/45 and 3/45 respectively.

Results: Strong correlation ($0.70 < SCC < 1.0$) was found for 36%, 49%, 67% and 73% of the patients in the LR, AP, SI, and total 3D direction respectively. The correlations between SCC and patient features investigated were weak ($SCC < 0.30$). There was a statistically significant difference in correlation coefficient values ($p < 0.05$) between groups with and without abdominal compression belt in SI direction. A stronger correlation was seen in the patient cohort with abdominal compression belt ($SCC = 0.79$) than in the patient cohort without ($SCC = 0.70$).

When taking the ANN model uncertainty in terms of the mean squared error, 87.5% of the ground truth values were within the predicted interval.

Conclusion: The correlation between surface and tumour motion was strong in majority of the patients in SI and total 3D direction. However, SCC varied amongst the patient data set and the result should therefore be interpreted as an indication of possible further individualisation of treatment and treatment planning. If correlation is to be used for treatment planning, an action plan could be applied for when the patient breathing pattern is irregular or different.

Abbrevations

4DCT - 4-Dimensional Computed Tomography
Adam - Adaptive Moment Estimation
AP - Anterior-Posterior
ANN - Artificial Neural Network
COM - Center of Mass
CT - Computed Tomography -
CTV - Clinical Target Volume
FANN - Feed-forward Artificial Neural Network
GTV - Gross Target Volume
IGRT - Image Guided Radiotherapy
LR - Left-Right
MRI - Magnetic Resonance Imaging
MSE - Mean Squared Error
PTV - Planning Target Volume
ReLU - Rectified Linear Unit
ROI - Region of Interest
RT - Radiotherapy
SBRT - Stereotactic Body Radiotherapy
SCC - Spearman's Correlation Coefficient
SGD - Stochastic Gradient Decent
SGRT - Surface Guided Radiotherapy
SI - Superior-Inferior
SUS - Skånes Universitetssjukhus

Contents

1 Introduction and aim

Out of the approximately 70 000 people diagnosed with cancer in Sweden in 2021[1], close to 4000 were diagnosed with lung cancer[2]. The predominant form of lung cancer involves the proliferation of malignant cells that grow on the mucous membrane in the lung and can both be primary or secondary tumours. In the latter case it is called metastasis. Metastases often grow in the lungs since the blood circulates through the lungs first after coming back to the heart[3]. Treatment is usually surgery or radiotherapy (RT), and it is common to combine RT with chemotherapy which are medicines that target the cancer cells[3].

Stereotactic body radiation therapy (SBRT) is an RT technique defined by high fractional radiation dose (> 5 Gy) delivered to a small volume (< 7 cm diameter)[4] with steep dose gradients. The total dose is delivered in a small number of fractions. At Skåne University Hospital (SUS), Lund, lung cancer can be treated with SBRT delivering 45 Gy over 3 fractions (15 Gy per fraction) over a week for peripheral targets. Depending on the location of the tumour, an abdominal compression belt can be used to reduce the intrafractional motion according to the local clinical routine.

SBRT is sensitive to patient motion due to the high fraction dose and low number of fractions delivered. During a small fraction dose the effect of missing the target is smaller than for a high fraction dose which SBRT uses. The effects on both potential reduction in total dose to the target and the damage on healthy tissue increases in this case. The dosimetric consequences are therefore larger for SBRT if the target is affected by motion due to respiration.

Respiration causes intrafractional motion, i.e. motion during treatment delivery. Intrafractional motion is caused primarily by anatomical motion, e.g. motion of skeletal muscular, cardiac, and gastrointestinal systems. In contrast, interfractional motion occurs between treatment sessions and includes setup uncertainties[5].

Respiratory motion affects the thoracic and abdominal regions and contribute to artefacts in computed tomography (CT) images, affects the need for increased planning target volume (PTV) and deviation in planned and delivered dose[5]. Motion artefacts give less optimal basis for the planning stage with regards to incorrect density in artefact volume as well as visually identifying the tumour. The respiratory motion alone requires an increased PTV margin to account for positions of the target in the different respiratory phases. This in return increases the radiation field size and introduces an increased risk of harming healthy tissue around the tumour[5]. Regarding the radiation delivery there is an introduced deviation in planned and delivered dose to target due to the multi-leaf collimator (MLC) component and target moving differently in relation to each other i.e. interplay effects[5].

To minimise the radiation dose to primarily organs at risk while maintaining the high target dose, several motion compensation techniques can be used. Keall et al.[5] mentions five main methods of motion compensation in RT: (1) motion-encompassing methods (such as slow CT or 4-Dimensional Computed Tomography (4DCT), to capture the entire respiratory motion), (2) respiratory gated techniques, (3) breath-hold techniques, (4) forced shallow-breathing methods, and (5) real-time tumour tracking with adaptive beam repositioning.

Currently at SUS a 4DCT examination is performed on patients for SBRT for lung cancer in the preparatory stage of the workflow. This is to encompass the whole respiratory motion of the patient and to retrieve an average position of the tumour. This is used, in combination with the average position of the tumour in different respiratory phases, to ensure that the PTV is covering the full motion range.

An abdominal compression belt (forced shallow-breathing method) is used when the tumour is located in the lower area of the lungs according to local routines. This decreases superior-inferior (SI) motion and allows usage of smaller target volume margins.[6]

4DCT imagery at SUS is used together with the optical surface system Sentinel (C-Rad Positioning AB, Uppsala, Sweden) to monitor patient motion and respiration. This method is called Surface Guided RT (SGRT) and is non-ionising[7] and non-invasive[8] in comparison to x-ray imagery and surgically implanted markers. SGRT can also be used for several of the aforementioned motion compensations such as gating and breath hold[8], but the application this study will explore is real-time tumour motion monitoring.

Respiratory motion of the patient can, however, vary both during and between treatment fractions[5]. This can affect the treatment if the planning was based on one breathing pattern and during treatment the patient breathes according to another pattern. This means that the motion-encompassing methods might not be enough to ensure high precision treatment delivery. The respiration needs to be monitored throughout the treatment delivery to make sure the patient is breathing approximately the same way as during the 4DCT examination. Even then it can not be guaranteed that the tumour will follow the same breathing pattern. Therefore, it would be of greatest interest if we could find a correlation.

To use SGRT for real-time tumour motion monitoring, there must be a strong correlation between tumour motion and surface motion. Correlation between patient surface and internal anatomical motion has been studied in several ways before[9; 10; 11; 12]. These studies show that there is correlation between both surface motion and internal landmarks[12], and strong correlation (correlation coefficient between 0.70 and 1.0) between surface motion and tumour motion[9]. In addition, if there is correlation there might be a possibility and potential for a deep learning model to predict the motion of the tumour based on the surface motion. Artificial neural networks (ANN), a subfield of deep learning, can be used for many fields such as regression and pattern recognition[13].

1.1 Aim

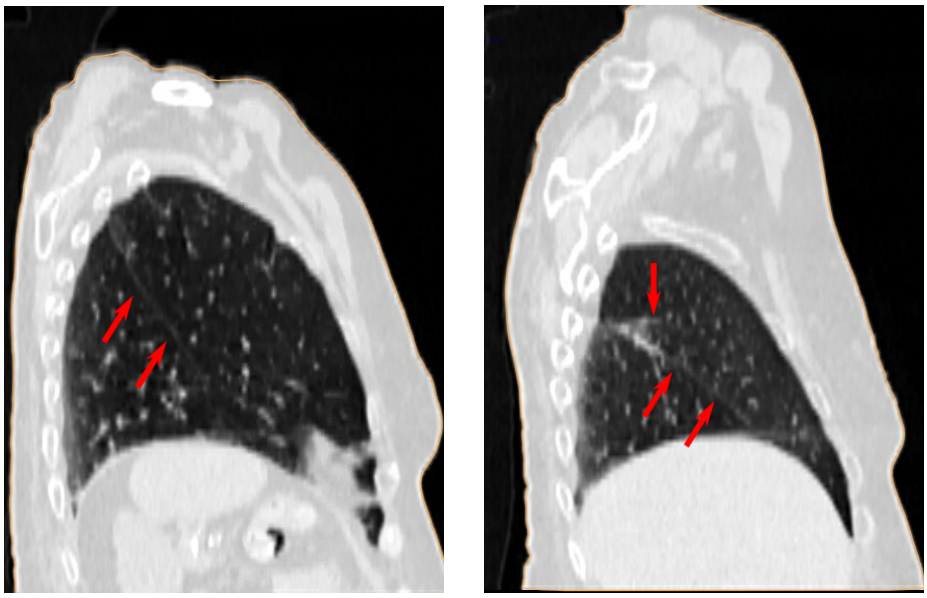
The aim of this study was to investigate the correlation between the tumour motion and the surface respiratory motion for lung cancer patients. Other variables such as anatomical placement of tumour, tumour volume, use of abdominal compression belt and their correlation to the internal/external motion were also studied. Further, the possibility to predict tumour motion based on surface data was investigated using an in-house developed ANN model.

2 Background

2.1 Lung anatomy and respiration

Respiration operates as gas exchange of oxygen and carbon dioxide in the arterial blood and is an involuntary function of the body, i.e. we perform it even unconscious[14]. The lungs are located in the thoracic cavity and are surrounded by the rib cage. Inspiration requires respiration muscles that expands the lung volume and thus pulling in air. These include the diaphragm and intercostal muscles between the ribs. The diaphragm pulls the lungs downward and pushes the abdomen inferiorly and anteriorly. The intercostal muscles contract and pulls the ribs superiorly and anteriorly. During expiration the muscles return to pre-inspiration position due to their elasticity.[14]

The lungs are divided by mediastinum and the heart. The lungs themselves are separated into structures called lobes. These can be identified by the fissures between the lobes (figure 1). The right lung has one oblique and one horizontal fissure, giving it three lobes: upper, middle and lower lobe. The left lung, however, only has one oblique fissure dividing it into an upper and lower lobe.[15]



(a) Lung fissures in left lung. (b) Lung fissures in right lung.

Figure 1: Lung fissures in both left and right lung, marked with arrows for clarification. Images are from the same patient.

2.2 CT and 4DCT

4DCT can be used when tumour or anatomical motion needs to be taken into account during treatment planning of RT. 4DCT was first mentioned back in 2003[16] and is a 3D volume CT acquired over multiple phases in a periodic motion[17], such as the respiratory cycle. CT images are acquired simultaneously as a respiratory monitoring system, such as a surface imaging system, which acquires respiratory amplitude data. The acquisition is usually performed over several respiratory cycles when the CT couch moves in the gantry and are sorted either retrospectively or prospectively into bins. Each bin represents a phase in the respiratory cycle and is then 3D reconstructed into volumes. These bins can be shown in a loop over the phases to visualise the motion during the respiratory cycle[16].

Identifying the phases of the respiratory cycle can be performed either through measuring respiration amplitude (amplitude binning) or via temporal allocation (phase binning) (figure 2). For measuring respiratory motion for the purpose of treatment planning at SUS, retrospective acquisition and phase binning are used.

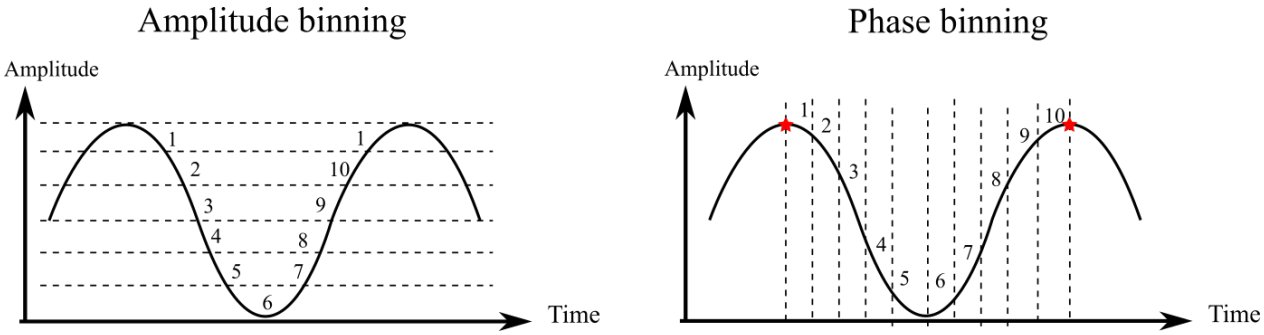


Figure 2: Schematic presentation of amplitude binning (left) and phase binning (right). In amplitude binning the phases are determined by the fraction of maximum recorded amplitude[16]. In phase binning the phases are determined by the temporal relationship to the respiratory cycle, i.e. the cycles are separated in fractions of time between defined occurrences, most commonly peak inhalations[18]. Stars in the phase binning image indicates defined peak inhalations.

2.3 Surface imaging systems in RT

Surface imaging systems for RT uses a non-ionising method of monitoring patient motion and is used in SGRT. SGRT have high temporal and spatial accuracy and the use of these types of systems have increased rapidly the last decade[7]. Al Hallaq et al.[7] describes in the AAPM TGR 302 that modern surface imaging systems use several different principles to detect patient motion such as laser scanning, time-of-flight, stereovision, and structured light imaging. Phase and amplitude can be extracted from the respiratory motion data and can be used for monitoring and gating.

An example of an optical surface scanning system is the Sentinel System (C-Rad Positioning AB, Uppsala, Sweden). Sentinel uses laser scanning technology with a laser divided into three lines that sweeps over the patient to collect data for a 3D surface image[8]. The laser scanning system is based on the report by Brahme et al.[19], where the unit includes a laser emitter, an optical scanner and a camera. The laser emitter emits a narrow-line laser which is manipulated by the optical scanner, consisting of a mirror mounted on a galvanometer (figure 3). This in turn can sweep the laser over the patient. The camera is made of a metal-oxide semiconductor detector that captures the laser lines. The location of contour data in 3D space is calculated with optical triangulation, which uses the angle of incidence on the detector to relate the captured image with the distance of reflection[20]. The pixels are then back projected, and a 3D surface can be reconstructed per camera image.[19].

During 4DCT, the purpose of the surface imaging system is to record the patient surface motion in a defined region of interest (ROI) to acquire a respiration curve during 4DCT (figure 4).

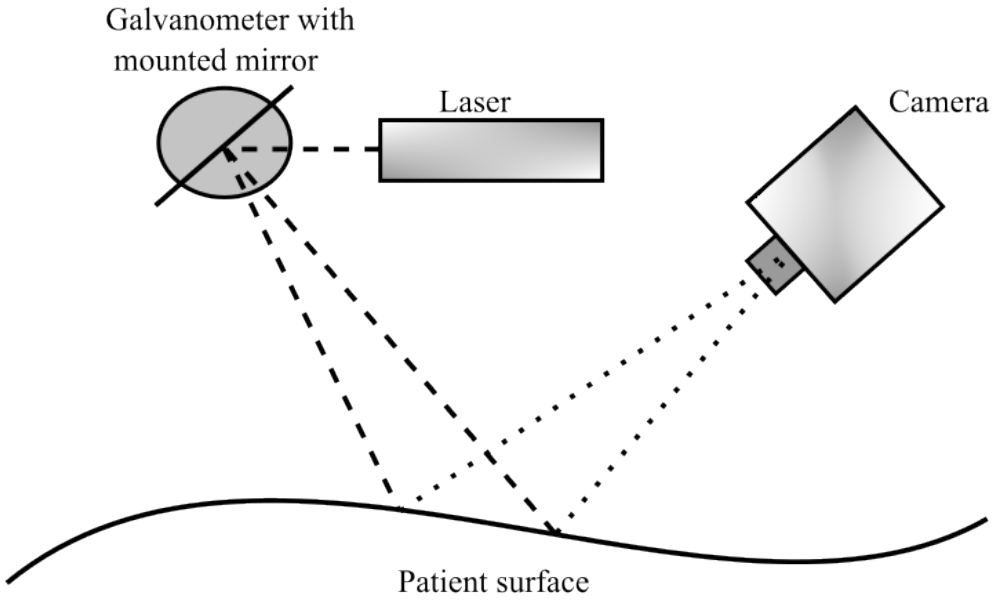


Figure 3: Schematics of a laser scanning optical system. The narrow-line laser is emitted on the mirror mounted on the galvanometer. This reflects onto the patient surface which in turn is picked up by the camera. Image inspired by Brahme et al.[19].

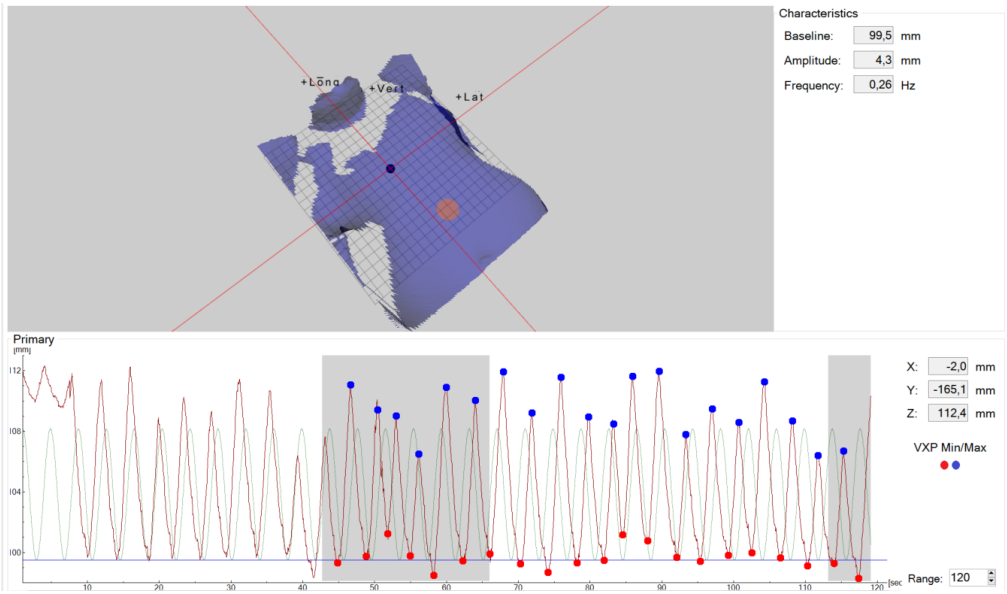


Figure 4: An example of a monitored breathing curve (bottom graph) and surface scan (top image) from the Sentinel system. The red dot on the patient surface (top image) marks the region of interest (ROI) that acquires the breathing motion. Blue and red markers in the breathing motion (bottom graph) marks inhale and exhale maxima respectively. Gray area represents where recording of the breathing motion is enabled, i.e. connection of the Sentinel system and the CT-scanner, and the white area between the gray intervals represents the breathing motion that was recorded during 4-dimensional computed tomography (4DCT) acquisition.

2.4 Artificial neural network

An Artificial Neural Network (ANN) is a series of functional transformations that is structured into layers[13; 21]. The name comes from the resemblance of the axon structures in the nerve cells, where each layer or node in the ANN represents a neuron in the axon[21]. The layers, also called hidden layers, consist of non-linear functions that transform activations, which are linear combination of input data, weight and bias parameters[13].

The layers can be fully connected between themselves, meaning every layer node has input from all previous layer nodes[21; 22]. Fully connected layers are also called dense layers (figure 5).

There are several types of ANN, including feed-forward, recurrent, competitive and probabilistic. In this study the feed-forward ANN (FANN) will be discussed since the FANN model architecture can accept inputs and process them in a forward direction without the need for any feedback loops, making them less complex than other neural network architectures and more efficient and easier to train[23].

The data used for an ANN model is traditionally divided into training, validation and test data sets. The training set makes up the majority of the data and is solely used for

training and optimising the model parameters. The validation set is used to monitor the training progress of the trained model performance. The test set is used for final assessment of the model and should not be intersecting the other sets[21].

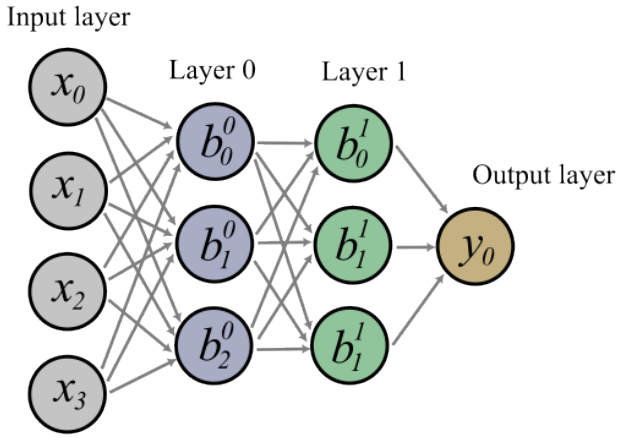


Figure 5: Visualisation of a typical fully connected feed-forward artificial neural network (FANN) with input data x, weight and bias b, output y. In this visualisation, the input has a shape of four units and the output has a shape of one unit. The layers between the input and output layers are usually called hidden layers.

2.4.1 Activation function

An activation function, sometimes called transfer function, is the function in each layer that transforms the activations to a fitting format for output. It can be viewed as a function that determines decision boundary, i.e. when the node is being activated. The choice of activation function is determined by the data set's appearance and distribution, but it is required to be non-linear. Otherwise, there would be a stacking of linear function layers resulting in a overall linear transformation independent of input, dimensionality, or repetition.[24; 13]

The Rectified Linear Unit (ReLU) is a common activation function and a truncation of a linear function (figure 6).[22]

$$ReLU(z) = \max(0, z) \tag{1}$$

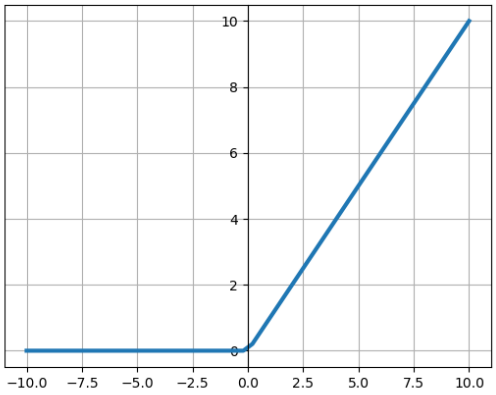


Figure 6: Plot of the rectified linear unit (ReLU) function.

The ReLU function does not change the size of the input and it solves the problem other activation functions face, namely vanishing gradients. During common training protocols of the ANN, error backpropagation is used. This method uses gradients of the error propagated backwards through the model to train it. These gradients are usually fractional and with many layers to pass through, the gradients become negligibly small. This will limit the model ability to learn from the training data set. The ReLU function surpasses this by keeping its gradient to 1.[24]

For the output layer, a different activation function must be chosen. It is determined by the purpose of the output, e.g. binary classification or scalar values. For regressions, a linear function is recommended[13].

2.4.2 Loss function

Loss functions are used in ANN model training and are used to measure discrepancy between the output predictions and the corresponding targets. The goal is to minimise the loss function and therefore the discrepancy, where the model parameters are optimised accordingly[13; 21].

For regression problems, a standard mean squared error (MSE) function is used as loss function. If there are n predictions ($\hat{y}$) and n targets (y), the MSE is computed as

$$\text{MSE} = \frac{1}{n}\sum_{i=0}^{n-1}(\hat{y}_i - y_i)^2. \tag{2}$$

2.4.3 Optimisation Algorithm

To minimise the loss function, there are several methods of optimisation. A common method is Stochastic Gradient Decent (SGD) which is built on the classic gradient decent method[21]. Gradient decent uses the derivative of the function to find the local minimum, in this case

the loss function. It uses a learning rate term η to determine the "step size" of the decent and therefore the convergence time. Each gradient decent iteration needs to iterate through all of the training set, which can be a workload for large training sets. To ease this workload, SGD approximates the loss function with only one target-prediction pair per iteration[21].

The Adam (Adaptive moment estimation) optimisation algorithm is a further development on the SGD method. Where the classic SGD use one learning rate for all parameters that does not change throughout training, the Adam optimiser use individual learning rates for all parameters that updates during training. It combines two other SGD extensions; Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp)[25].

2.4.4 Regularisation

Overfitting is a challenge in some ANN models and means that the model can perform well on predicting the training data but fails to achieve the same performance on the validation and test data, i.e. it can not generalise. To avoid overfitting, the operator can use certain approaches such as 1) use a larger dataset, 2) make the model simpler, or 3) introduce a regularisation term in the layers[21].

There are several regularisation methods to choose from such as L1 and early stopping. The L1 regularisation term is introduced in the hidden layers and forces the parameter elements to zero. This changes the impact of each parameter and levels out the prioritisation. The amount of levelling is decided by a weighting factor λ, where lower λ gives heavier levelling[21].

Early stopping is a type of regularisation where the training is stopped before the validation loss starts increasing, i.e. the amount of training epochs is restricted. An advantage of early stopping is that it does not modify the loss function or parameter vector like L1 regularisation does[21].

3 Materials and Methods

3.1 Patient inclusion

A total of 48 patients with lung cancer treated with SBRT, 45 Gy/3 fractions at SUS were retrospectively enrolled in this study. Inclusion criteria were lung cancer diagnosis, fractionation, and patients treated 2019 (3 patients, used for uncertainty evaluation) and 2021 and after (45 patients, used for tumour delineation). Of the 45 patients used in delineation, 14 patients were treated using abdominal compression belt. Anatomical tumour placement varied across the population of 45 patients where 16 patients had their tumour in left upper lobe, 3 in left middle lobe, 9 in left lower lobe, 11 in right upper lobe, and 6 in right lower lobe. All patients were previously scanned with 4DCT using a Siemens Somatom Definition AS (Siemens Medical Solutions, Erlangen, Germany) prior to treatment combined with respiratory motion monitoring, Sentinel.

The 4DCT protocol for all patients had 3 mm slice thickness in SI direction, 0.09 pitch, 500 mm FOV, 120 kVp and 40 mAs/rotation (dose reduction software used). The reconstruction software used was Siemens' SAFIRE (I30s medium smooth kernel), and the software for binning the dataset of approximately 800 projections into eight phases utilised phase binning.

The Regional Ethical Review Board in Lund has approved retrospective research of radiotherapy data (No. 2013/42).

3.2 Delineation and center of mass retrieval

A total of 45 patients from 2021 and after were delineated. Delineations were made in the contouring module of Eclipse treatment planning system (v.15.6, Varian Medical Systems, Palo Alto, CA, USA). Delineations of the tumour volume were made on 4DCT images. The lungs, trachea and bronchial tube volumes were automatically segmented with the automatic tool using predefined Hounsfield unit values. To minimise errors, the bronchial tube and trachea volumes were not removed. They were assumed to maintain their volume throughout the eight respiration phases.

For each patient, the tumour was delineated in each phase and the 3D center of mass (COM) coordinates in X, Y and Z corresponding to left-right (LR), anterior-posterior (AP), and SI respectively were retrieved alongside lung volume and tumour volume. Tumour volume was delineated to replicate a gross target volume (GTV). The maximum exhalation phase was used as starting point and the COM motion was relative to this phase. The maximum exhalation phase was identified as the phase with the smallest lung volume. Maximum exhalation phase was also used for surface motion data analysis. One person made all delineations.

For calculation of the total 3D tumour motion between each phase the vector was computed with origin in the maximum exhalation phase COM. The vector length D was calculated using equation 3,

$$D = \sqrt{(q_x - r_x)^2 + (q_y - r_y)^2 + (q_z - r_z)^2} \tag{3}$$

where q is the tumour coordinate in the maximum exhalation phase, and r is the tumour COM in other phases. Similarly, the standard combined uncertainty of D (u_c) was calculated as the positive square root of the combined variance u_c^2 which in turn is given by equation 4[26],

$$u_c^2 = \sum_{i=1}^{N} \left(\frac{\partial D}{\partial x_i} \right) u^2(x_i) \tag{4}$$

where each $u(x_i)$ is a standard approximated uncertainty and x_i is the i:th variable in function D.

3.3 Evaluation of tumour delineation uncertainties

The three patients from 2019 were used for an uncertainty evaluation of tumour delineation. Two of the investigated patients had a spherical shaped tumour (figure 7,8) whereas the

third patient had an irregular tumour shape (figure 9). For each of the three patients, the tumour volumes (the GTV) were repeatedly delineated 11 times by one person, performing the delineations over two days to minimise effect of habit. The tumours were delineated in all CT phase images, and volume and COM-coordinates in X, Y and Z direction were acquired.

For each delineation i, positions on the axes x_i^p, y_i^p, z_i^p were acquired for each phase $p = \{1, 2, 3, 4, 5, 6, 7, 8\}$, and the mean value of the axis position was calculated $(x_{mean}^p, y_{mean}^p, z_{mean}^p)$. The absolute deviation of each position from the respective mean values was then calculated as

$$\begin{aligned} x_{dev,i}^p &= |x_i^p - x_{mean}^p| \\ y_{dev,i}^p &= |y_i^p - y_{mean}^p| \\ z_{dev,i}^p &= |z_i^p - z_{mean}^p|. \end{aligned} \tag{5}$$

A mean value of these was in turn calculated to get a mean deviation per phase,

$$\begin{aligned} x_{dev,mean}^p &= \tfrac{1}{11} \textstyle\sum_{i=1} x_{dev,i}^p \\ y_{dev,mean}^p &= \tfrac{1}{11} \textstyle\sum_{i=1} y_{dev,i}^p \\ z_{dev,mean}^p &= \tfrac{1}{11} \textstyle\sum_{i=1} z_{dev,i}^p. \end{aligned} \tag{6}$$

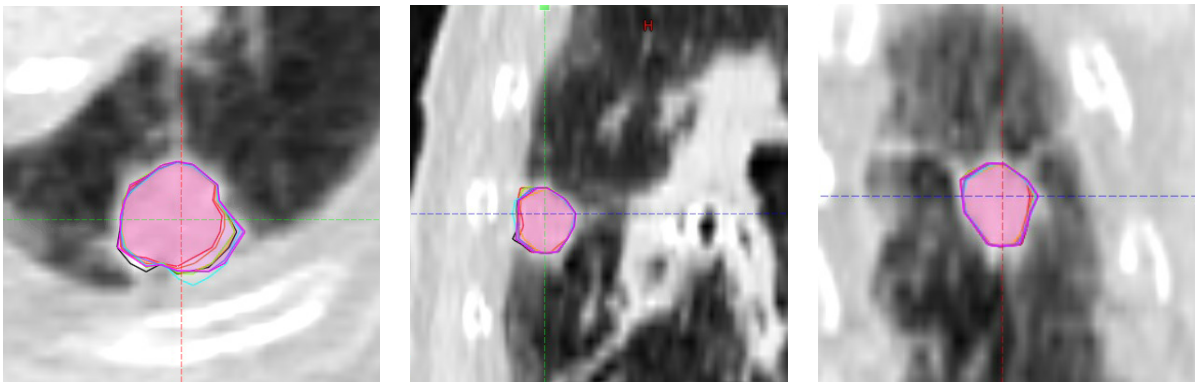
Since there were eight phases, each set included eight values (one for each phase). A mean value was calculated over each set in turn as

$$\begin{aligned} x_{dev,mean}^n &= \tfrac{1}{8} \textstyle\sum_p x_{dev,mean}^p \\ y_{dev,mean}^n &= \tfrac{1}{8} \textstyle\sum_p y_{dev,mean}^p \\ z_{dev,mean}^n &= \tfrac{1}{8} \textstyle\sum_p z_{dev,mean}^p \end{aligned} \tag{7}$$

where n denotes patient. Lastly, a mean value over the patients thus a standard approximated uncertainty of the COM position in each direction u_x, u_y, u_z was calculated:

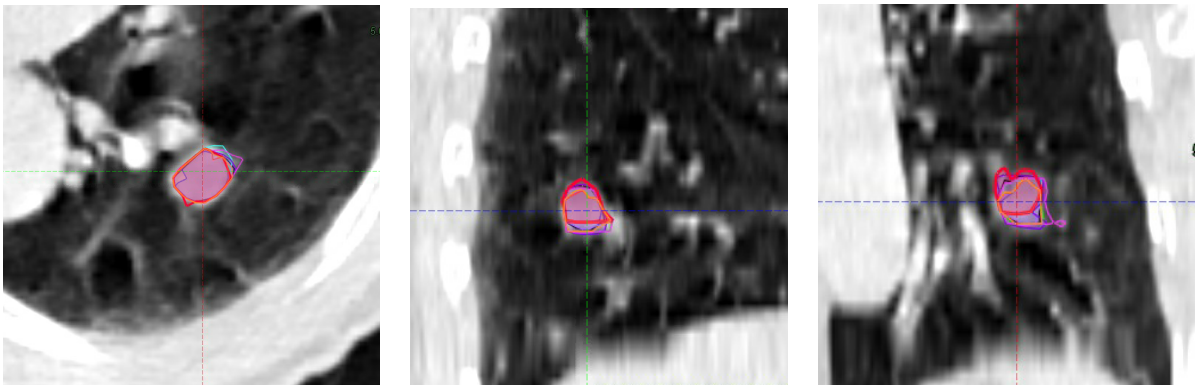
$$\begin{aligned} u_x &= \tfrac{1}{3} \textstyle\sum_{n=1} x_{dev,mean}^n \\ u_y &= \tfrac{1}{3} \textstyle\sum_{n=1} y_{dev,mean}^n \\ u_z &= \tfrac{1}{3} \textstyle\sum_{n=1} z_{dev,mean}^n. \end{aligned} \tag{8}$$

To qualitatively investigate how delineation errors affected the deviation from the mean COM coordinate, systematic errors were investigated as part of the uncertainty evaluation. Five systematical error delineations were introduced for each patient. These were a combination of over- and underestimated volume delineation and missing the tumour (figure 10).



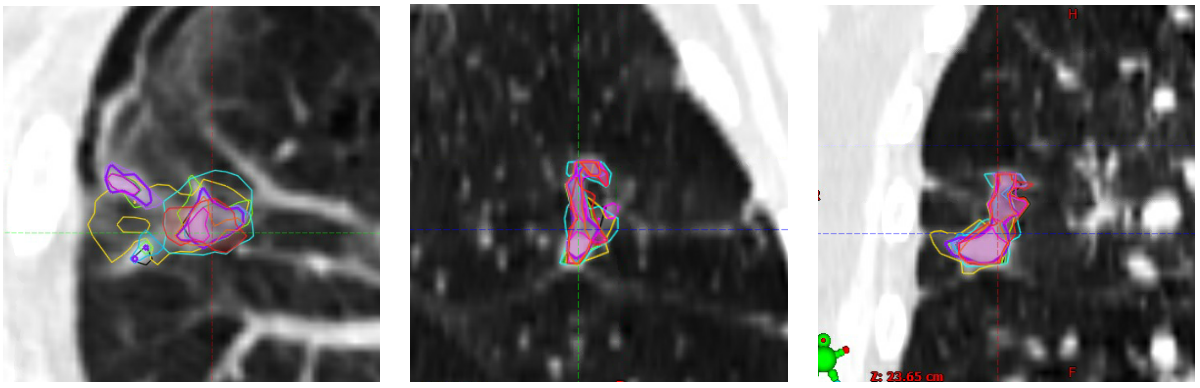
(a) Transversal projection. (b) Sagittal projection. (c) Frontal projection.

Figure 7: Overlapping delineation volumes for the first uncertainty patient in all projections.



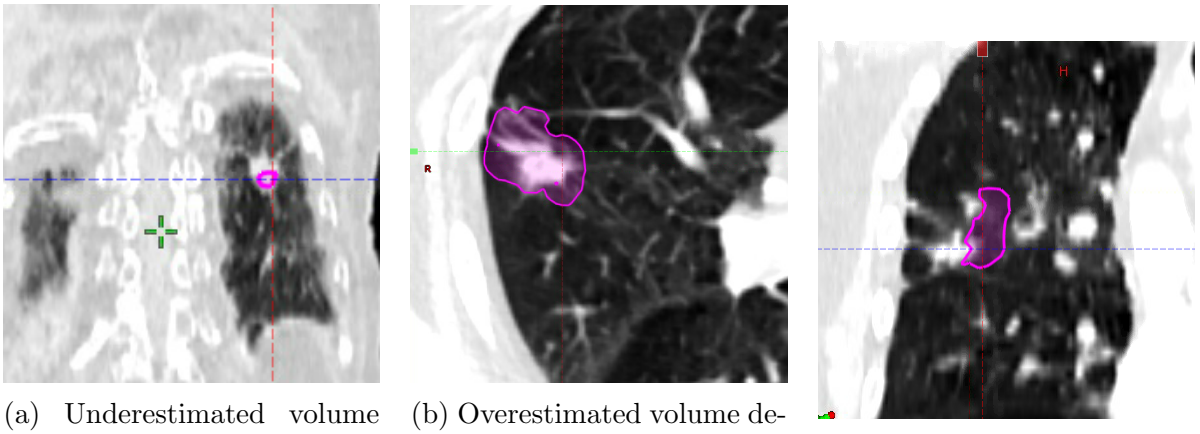
(a) Transversal projection. (b) Sagittal projection. (c) Frontal projection.

Figure 8: Overlapping delineation volumes for the second uncertainty patient for in all projections.



(a) Transversal projection. (b) Sagittal projection. (c) Frontal projection.

Figure 9: Overlapping delineation volumes for the third uncertainty patient in all projections.



(a) Underestimated volume delineation of the tumour. (b) Overestimated volume delineation of the tumour. (c) Missing the tumour.

Figure 10: Examples of systematic errors used to qualitative study the delineation uncertainty.

3.4 Surface motion data from Sentinel system

To monitor surface respiratory motion, the Sentinel system uses a 2 cm diameter ROI placed manually on an anatomical reference location on the patient, e.g. between the navel and solar plexus. The included patients were free breathing during 4DCT acquisition, meaning that no visual or audio aid nor breathing instructions were given to the patient. The recorded breathing amplitude was connected to the 3D space and isocenter of the room.

Surface motion data for the patients were retrospectively retrieved from Sentinel system log files. It consisted of respiratory amplitude (only in vertical direction, i.e. AP direction anatomically) and phase division as well as markers for the software's identified inhalation and exhalation points. Additional information such as time and timestamps of when scan and study started was also available in the log files (figure 4). The respiratory motion data

was sampled based on the 4DCT phase data to retrieve which respiratory amplitude data points belonged to which phases. The mean respiratory amplitude for each phase was then calculated using these data points.

The surface motion relative to maximum exhalation phase was calculated by subtracting the amplitude of exhalation phase with the amplitude of the other phases.

3.5 Motion data analysis

Spearman's correlation coefficient (SCC) was used to evaluate the internal/external correlation. The Spearman's correlation is a non-parametric evaluation technique that operates on the rank of the data set rather than the raw data[27] and detects monotonic trends but not linearity. Each variable, in this case tumour and surface motion, is ranked separately from lowest to highest. A sum of the squares of the difference of ranks between the variables is normalised to an interval $SCC = [-1, 1]$. SCC can then be calculated according to equation 9,

$$SCC = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{9}$$

where d denotes the difference between ranks for each data pair and n denotes the number of data pairs[27].

The SCC is insensitive to data outliers as well as non-normal distribution since it operates on the ranks rather than the raw data. For this study, cut off values of SCC defined as weak ($SSC < 0.30$), moderate ($SCC = [0.30, 0.69]$) and strong ($SCC = [0.70, 1.0]$) were established to interpret the external/internal correlation in accordance with Wang et al[9]. The SCC was calculated in Excel (Microsoft Corporation, Redmond, WA, USA) by converting the tumour motion and surface motion for each patient into ranks separately with the RANK.AVG function. The CORREL function was then applied on these ranked sets to retrieve the SCC. This was done in LR, AP, SI and total 3D direction.

Spearman's correlation analysis was made between the computed SCC and patient age, tumour volume, smallest distance to thoracic wall, distance to thoracic spine and to chest surface, lung volume, and with/without abdominal belt compression. The data set of all patients' SCC and patient features were separately ranked with the RANK.AVG function, and the CORREL function was then applied to the SCC data set and one feature at a time to acquire the correlation coefficient between the SCC and patient feature.

The linearity of the internal/external correlation was studied by completing a linear fit between tumour motion and surface motion in Excel and acquiring the coefficient of determination R^2. R^2 is the proportion of variation of data points explained by the regression line or model[28], and the closer R^2 is to 1 the better the regression fit.

A Wilcoxon's Rank Sum test was performed on the groups with and without belt compression to investigate statistically significant differences between the groups. This is similarly to Spearman's correlation a non-parametric statistical test appropriate for data sets where a normal distribution can not be assumed. It is designed for independent samples with the null hypothesis that the compared groups come from the same distribution. The data is

ranked like in Spearman's correlation, and the lowest of the two rank sums are used as comparative test statistics[28]. The Scipy package in Python was used for Wilcoxon's Rank Sum test evaluation, computing a p-value. In this case $p < 0.05$ was used to indicate statistical significance[29].

3.6 ANN modelling

The goal of the ANN model was to predict the tumour motion by using surface motion as input.

3.6.1 Training, validation and test data

There were in total 20 input attributes used for the model; surface motion (eight attributes, one for each phase), surface mean motion, surface median motion, use of compression belt (one-hot encoded), and if the tumour was positioned in upper, lower or middle lobe (one-hot encoded). The tumour motion (eight attributes, one for each phase) were the output for the model to predict based on the input attributes.

The data used to build the model was drawn from the patient data collected from the previous data analysis (45 patients from 2021). The dataset was randomly divided into training, validation and test sets. 36 patients were used for training the model, 6 patients were used for validation, and 3 patients were used for test.

3.6.2 Model architecture

Link to full code can be found at GitHub at https://github.com/caisakj/ANN-model-tumour-prediction.git. The ANN model was built like a FANN with two dense hidden layers with 128 units in each utilising ReLU activation functions. An L1 kernel regulariser ($\lambda = 0.0001$) was used in both layers to minimise overfitting. The output layer had eight units to obtain eight data points resembling the tumour motion attributes and used a linear activation function. For model compilation, the Adam optimiser was used. Chosen loss function was MSE and 400 epochs was used during training.

An uncertainty interval was created around the predicted tumour motion data points from the model. The interval was defined as $[+\sqrt{MSE}, -\sqrt{MSE}]$ where MSE is the retrieved MSE value from the model computed according to equation 2. This represented the interval which the model could predict the tumour motion to be within.

Predictions were made with the model on the test data, where in total 3x8 data points were acquired (three patients with eight data points each). The uncertainty interval was applied and centered around the predicted data points and plotted together with the true values for each test patient respectively.

4 Results

4.1 Motion data analysis

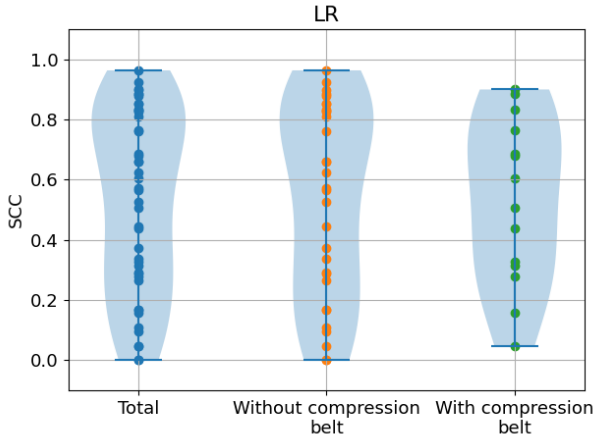
Acquired absolute mean values of the correlation coefficient SCC were 0.53 (LR), 0.58 (AP), 0.73 (SI) and 0.74 (total 3D motion), all ranging from moderate to strongly correlated. Medians were 0.57 (LR), 0.66 (AP), 0.77 (SI) and 0.80 (total 3D motion). The values varied between and within the directions (figures 11a-11f). Distribution of patients between the correlation classifications is presented in table 1.

Table 1: Table of patient percentage proportion within the correlation classifications.

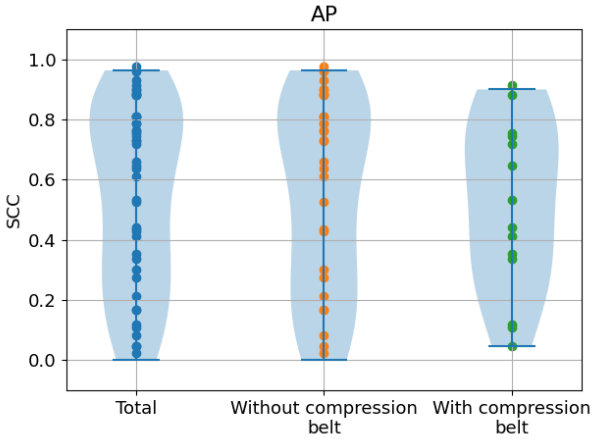
	LR	AP	SI	Total 3D
N.o patients	45	45	45	45
N.o strongly correlated (0.70-1.0)	16 (36%)	22 (49%)	30 (67%)	33 (73%)
N.o moderately correlated (0.30-0.69)	17 (38%)	13 (29%)	12 (27%)	10 (22%)
N.o weakly correlated (<0.30)	12 (27%)	10 (22%)	3 (7%)	2 (4%)

A statistically significant difference ($p = 0.004$) was found between the groups with and without compression belt in the SI direction. The group with compression belt showed a higher SCC (0.79) compared to the group without compression belt (0.70). Correlation between SCC for total 3D motion and patient features showed weak correlation (< 0.30) over all features.

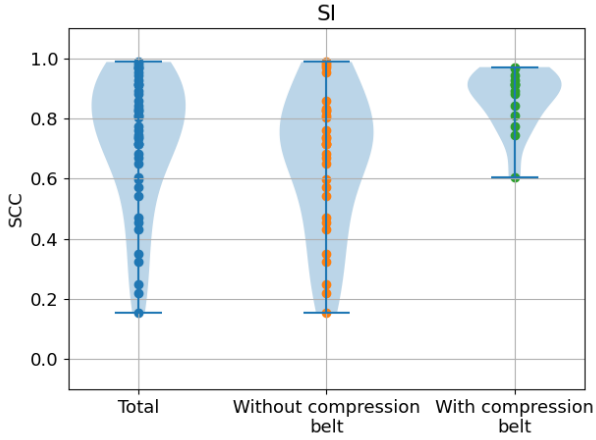
The mean R^2 value was 0.63 and the median was 0.68. The R^2 varied between patients (figure 11e). There was no statistical difference of R^2 between groups with and without compression belt in neither direction ($p = 0.05$).



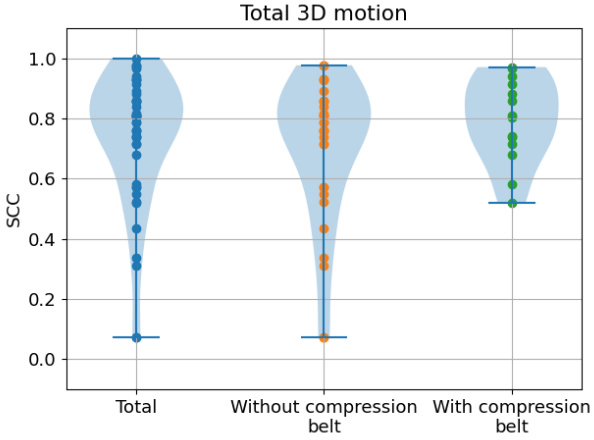
(a) Left-right (LR) correlation.



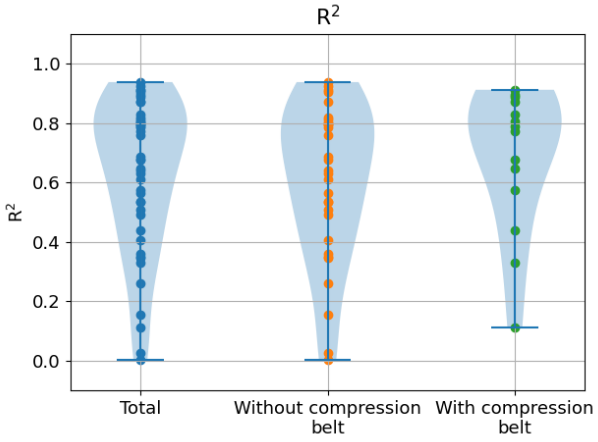
(b) Anterior-posterior (AP) correlation.



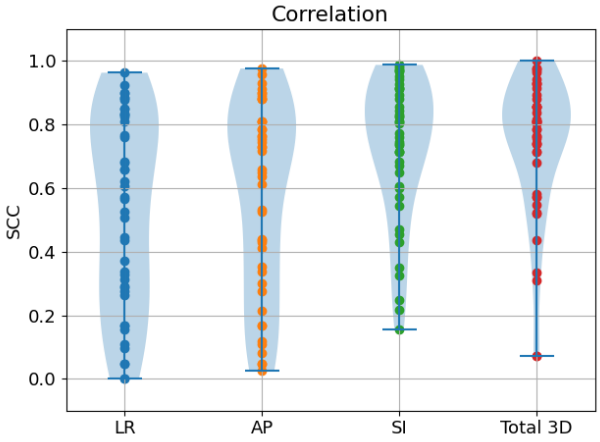
(c) Superior-inferior (SI) correlation.



(d) Total 3D motion correlation.



(e) R^2 values.



(f) Correlations (total group) in all directions.

Figure 11: Violin plots of Spearman's correlation coefficients (SCC) in a) LR, b) AP, c) SI and d) total 3D directions separated into groups of with and without compression belt and both combined (total). E) Similar violin plots of R^2 from fit between surface and tumour motion separating the patient group into with and without compression belt and total patient group. F) Violin plots of the SCC (total patient group) in all directions.

4.2 Tumour delineation uncertainties

The ratio between largest and smallest delineated target volume V_{max} and V_{min} respectively in all phases and patients is presented in table 2. Generally, $V_{max}/V_{min} < 2$ for the spherical tumours (patient one and two) except for one outlier in the first CT-phase. $V_{max}/V_{min} < 3$ for the non-spherical tumour (patient three).

Table 2: Table of V_{max}/V_{min} over the phases in the three patients. The value is unitless.

	V_{max}/V_{min}		
CT phase	Patient 1	Patient 2	Patient 3
1	1.23	2.50	2.06
2	1.70	1.44	1.46
3	1.36	1.29	2.47
4	1.64	1.43	1.45
5	1.42	1.29	1.70
6	1.33	1.38	2.21
7	1.55	1.71	2.05
8	1.21	1.63	2.50

Figures 12-14 show how the mean deviation from mean COM position in each phase varied in all directions. A common trend was that the deviation in SI was larger than the other directions. All deviations were under 1 mm except for patient three (non-spherical tumour).

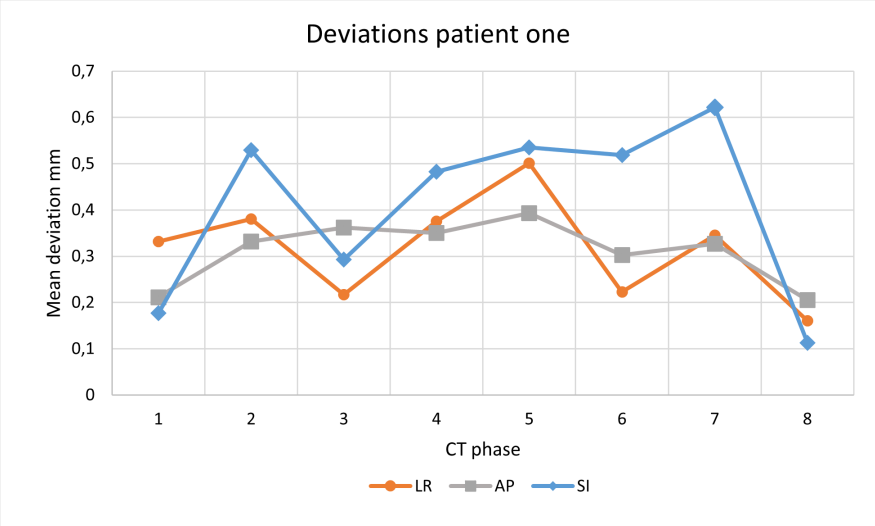


Figure 12: Plot of mean center of mass (COM) deviations from mean COM position for patient one. X-axis is the CT-phase, y-axis is the mean deviation in mm. Circle points are left-right (LR) direction, square points are anterior-posterior (AP) direction, and diamond points are superior-inferior (SI) direction.

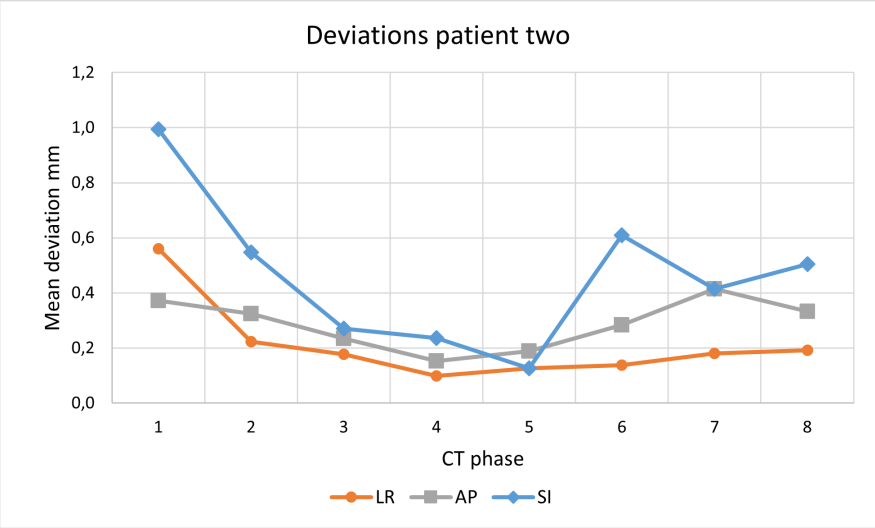


Figure 13: Plot of mean center of mass (COM) deviations from mean COM position for patient two. X-axis is the CT-phase, y-axis is the mean deviation in mm. Circle points are left-right (LR) direction, square points are anterior-posterior (AP) direction, and diamond points are superior-inferior (SI) direction.

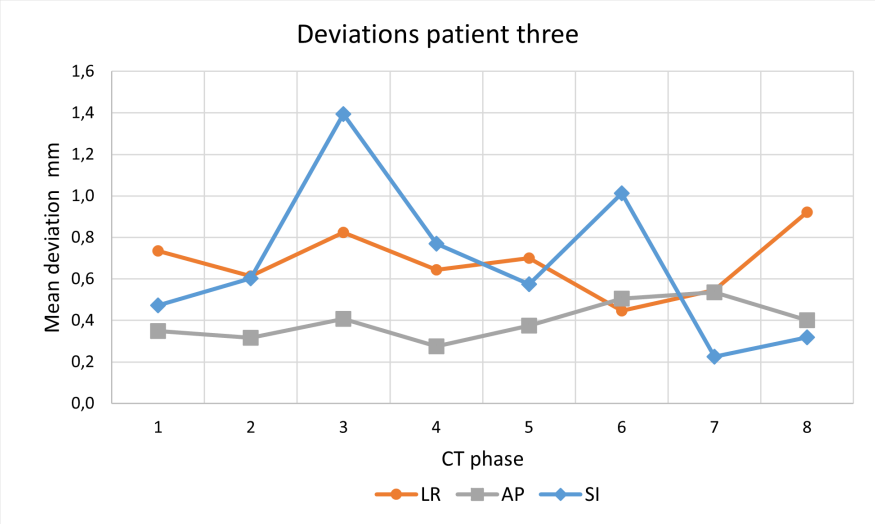


Figure 14: Plot of mean center of mass (COM) deviations from mean COM position for patient three. X-axis is the CT-phase, y-axis is the mean deviation in mm. Circle points are left-right (LR) direction, square points are anterior-posterior (AP) direction, and diamond points are superior-inferior (SI) direction.

When evaluating the systematic errors, the general result was that the deviation from mean COM position within the dataset increased (table 3).

Table 3: Systematic error evaluation. Position deviation denotes the general trend of deviation from the mean center of mass (COM) position over all directions (LR, AP, SI) in number of data points. If no specific amount of data points is written, it implies all data points. Total amount of data points is 24 (eight per direction).

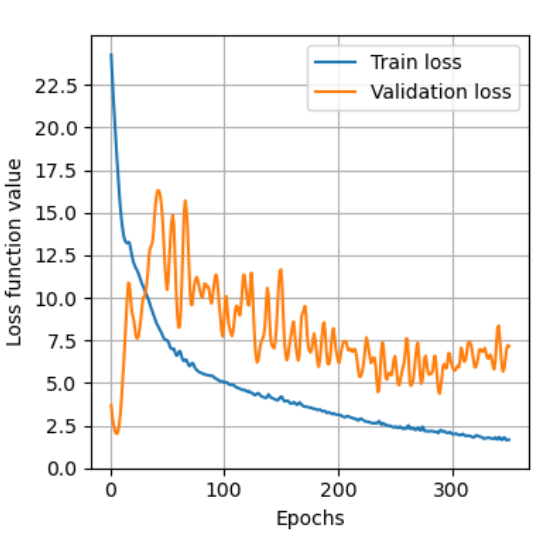
		Position deviation
Patient 1	Without error	<1 mm
	With error	12/24 <1 mm 12/24 >1 mm and <2 mm
Patient 2	Without error	<1 mm
	With error	11/24 <1 mm 13/24 >1 mm and <2 mm
Patient 3	Without error	22/24 <1 mm 2/24 >1 mm and <2 mm
	With error	8/24 <1 mm 9/24 >1 mm and <2 mm 7/24 >2 mm

The calculated mean deviation in each direction, which in turn becomes the estimated standard uncertainty for COM position, was computed with equations 5-8 as $u_x = 0.4$ mm,

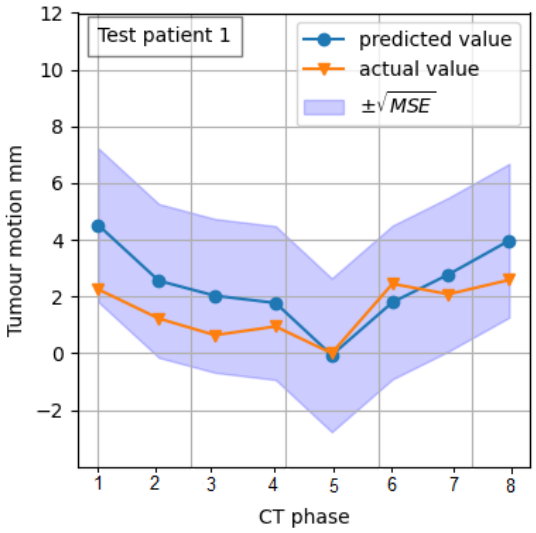
$u_y = 0.3$ mm and $u_z = 0.6$ mm, which is smaller than the voxel size of the image. With evaluation with the voxel size, it was decided that the voxel size was going to be used as the COM position uncertainty.

4.3 ANN model

The training and validation loss is presented in figure 15a. The square root of MSE was $\sqrt{MSE} = 2.7$ mm, acting as basis for the prediction interval as shown in figure 15b. Regarding the accuracy, 87.5% of the test data was within the predicted interval ± 2.7 mm. The plots of the remaining test patients are presented in Appendix A.



(a) Plot of the model performance, how the loss function value progressed for the training and validation. X-axis is epoch number and y-axis is loss function value. The two data sets are separated by colour, but the validation data is less stable than the training data.



(b) One of the three patients from the test data set. X-axis is CT-phase number and y-axis is tumour motion in mm. Model predicted value has circle points and actual values have triangle points. Shaded area marks the square root of mean squared error $\sqrt{MSE}$ interval and it is centered around the predicted values.

Figure 15: The performance of the artificial neural network (ANN) model on the a) validation and b) test data sets.

5 Discussion and Conclusion

5.1 Correlation and data analysis

In many cases of the patient group, assumption of correlation between tumour motion and surface motion is valid as is needed for use of SGRT. Even though this might sound promising, it is important to consider the variation of SCC over the patient group.

The value of R^2 is interpreted as how well the data fits the linear regression model. With an R^2 closer to 1, the data can be well approximated with a linear model. The data is fairly linear but there was a wide spread of these values for these patients (figure 11e) which needs to be considered. The patients with $R^2 < 0.5$ ($n = 12$) all had SCC< 0.6 except for one (SCC$= 0.68$), which intuitively says that a smaller R^2 does not necessarily mean small SCC. However, more data is needed to be able to carry out statistical tests.

The statistically significant difference between the groups with and without abdominal compression belt in SI direction can be caused by respiratory motion affecting the anatomy a lot in SI direction with the diaphragm expanding and contracting. The compression belt restricts motion of the abdomen and diaphragm thus actively affecting the SI motion. This could mean that if the patient's internal/external correlation is used during treatment planning and delivery, the abdominal compression belt could improve the correlation and whatever parameters are changed with it. It is difficult to say whether it is possible or not with this small of a sample size, the fact that it is a secondary finding, and that there is no paired data in this study.

The higher correlation for the group with compression belt can not tell, however, if the tumour motion itself is smaller or equally large as without compression belt. There can still be a correlation between a small-motion tumour and a high-motion stomach and vice versa, as long as their motion patterns coincide.

Performing a statistic test on the groups with different lung lobe placements was considered. It was never done because of the uneven and small sample sizes in each group. If this analysis is to be performed in the future, more patients with a larger spread of tumour positioning need to be collected such that a statistically representative number of patients are in each group. There are studies[5], however, that have shown that anatomical placement can make a difference in tumour motion. There is therefore a possibility that further inclusion of patients would show different correlation between surface and tumour motion, depending on anatomical placement, in this study as well.

There might be an issue of using the two groups of with and without compression belt together in this study - from the beginning it was not thought to be a problem but since there was a statistical significance, it should be taken into consideration in further studies to avoid bias. In hindsight, it could be anticipated that the group of patients utilising an abdominal compression belt would have less tumour and surface motion[6], potentially affecting the correlation between surface and tumour motion, thus constituting a potential source of bias. This selection bias stems from the fact that abdominal compression belt as a criterion was not included in the patient selection, thus patients with abdominal compression belt were

randomly included.

The correlation between surface and internal landmarks and tumours have been studied previously. In a study by Vedam et al.[10], they used surrogate points on the apex of the diaphragm for internal motion during respiration. Their data was acquired with fluoroscopy studies of patients for internal motion, and with infrared camera-based systems for external motion. Despite various breathing instruction such as audio instructions, visual feedback, and no instructions at all, results showed a strong relationship ($p < 0.0001$) between external motion and diaphragm motion during respiration.

A lung tumour is however not guaranteed to follow the diaphragm's motion. In Fayad's et al.[11] study they investigated correlation between surface respiratory motion and anatomical landmarks, as well as points of interest on the tumour. They used 4DCT images for internal marking and a calculated distance map from 4DCT volumes was used for surface motion. They found higher correlation with landmarks and surface regions of interest (ROI) placed centrally on the chest and abdomen compared to those placed more peripherally. The use of a real-time monitoring system was recommended for this type of use. The highest correlations that were calculated had a value of approximately 0.8.

In another study by Wang et al.[9] they studied similar correlation to Fayad et al. They used surface data from a real-time signal system (Varian RPM), and coordinates of outlined tumours in the liver and lung from 4DCT images. The surface data was correlated to tumour motion as well as other parameters such as tumour volume and patient age. Results showed correlation with external and internal markers motion (values approximately 0.9), but only moderate or weak correlation to the patient features age, gender, tumour volume, internal tumour motion magnitude, and external marker motion magnitude (values < 0.5). The authors emphasised on individual differences in external and internal motion, and that individual assessment of correlation should be considered.

5.2 Uncertainty analysis

The reason the patients used for the uncertainty analysis are from a different year, and are not included in further analysis, is that it was not discovered until after the uncertainty analysis was done that there was not sufficient surface motion data on these patients to continue the correlation analysis. The surface motion data was not used in the uncertainty analysis. After the discovery, the patent selection was changed to the year 2021 and after, where sufficient surface motion data could be found. The remaining patient inclusion criteria are the same, however.

The tumour volume does not change with breathing. It can stretch and deform but the volume does not change with it. The values in table 2 should therefore ideally be 1. Spherical tumours seemed easier to reproduce with respect to target volume than irregularly shaped tumours. Despite the volumetric variance, the deviations in the acquired COM were in many cases kept below 1 mm in all directions (figure 12,13,14).

In an article by Weiss and Hess[30] they compiled various studies of interobserver variability for several diagnoses, lung cancer among them. They also addressed the difficulty that

is the medical concept of GTV and CTV, and how important experience is when delineating target volumes. For lung tumours, the ratio between the largest target volume V_{max} and the smallest target volume V_{min} differed from 1.6 to up to 3.74 in different studies compiled in the article. The modalities used for these studies were either CT or CT combined with Magnetic Resonance Imaging (MRI).

However, it can look different for intraobserver studies. Louie et al.[31] performed both an inter- and intraobserver study and found that intraobserver variability was lower than that for interobserver. This means that the uncertainties in delineations were reduced in this study by only having one person to delineate, which might be the difference that is seen in V_{max}/V_{min} values from this study and Weiss and Hess' results.

In the current study, when systematic errors were introduced the deviation in COM increased. There were in total 5 delineations out of 16 that were introduced systematic errors in each phase in all patients. The order of magnitude of COM position deviation was the same for both normal delineation and delineation including systematic errors in all three patients. The values were higher, however, with the inclusion of systematic errors leading to the trend observed in table 3. It is important to keep in mind that the mean COM position changes with inclusion and exclusion of systematic error slightly, and this of course might affect the deviations. All of this should be taken into consideration as a form of margin, however this discussion is only qualitative and a set number for this margin is not possible to acquire from the results of this study.

The larger errors occured for the irregularly shaped tumour (patient 3), which could be an effect of the more complex position of the COM point. Each patient had a mix of different systematic errors (figure 10), and these could together cause an error that is larger than if only one of them was present.

Nonetheless, the larger dimension of the voxel size sets the limit of what we can observe, and we need to consider the effects of eventual systematic errors. Thus, it was decided after the uncertainty analysis to use voxel size as approximated uncertainty. These uncertainties were used as $u(x_i)$ in computation of the propagated error with equation 4 for total 3D motion, which resulted in propagated errors between 1-3 mm.

The approximated uncertainty and error were not used for correlation analysis. This is because the uncertainty of the correlation coefficient is dependent on sample distribution and coefficient value, rather than the error of the data[32]. Regarding the ANN model, the approximated error was used as a frame of reference when assessing $\sqrt{MSE}$ and comparing the model's uncertainty with the tumour motion uncertainty.

It is important to know the uncertainties that come with target delineation. A well-known uncertainty is needed to apply not only correct PTV, but also correct GTV and CTV. Delineation is dependent on the image quality as well as the one performing the delineation and is an important part of the RT workflow. Image modality and data acquisition affect the quality of delineation through e.g. introduction of artefacts. The clinical experience will in turn introduce another uncertainty and variation between observers. The PTV is based on other uncertainties such as intra- and interfractional motion as well as CT uncertainties, in contrast to the CTV and the GTV[33].

5.3 Surface motion data

According to local clinical routine, the ROI should be placed on a surface that moves with respiration for the respiration study with Sentinel. According to the results by Fayad et al.[11], the highest correlation coefficients between internal and external motion on a patient is achieved on the central abdominal area. Taking this into account, it was justified placing it between the navel and solar plexus. Such placement assured that the derived SCC was maximised to the fullest extent.

An issue with the dataset acquired from the Sentinel system was that the data did not exclude outliers (figure 17a,17c in Appendix A) in the respiratory motion data. There were 26 out of the 45 patients that had visible outliers or uneven breathing pattern. These outliers can change the real mean position during a phase, as much as an uneven breathing pattern (figure 17b in Appendix A) during the 4DCT scan. If these are not excluded, they can also contribute to CT artefacts by being included in the image reconstruction thus affecting delineation process.

Furthermore, it can not be retrieved from the Sentinel data whether these outliers have been manually excluded before reconstruction or not. It can be retrieved from the reconstruction program, but the reconstruction data is regularly deleted due to hard drive space. This means that for the patient group used (patients from 2021) this can not be done. This can introduce a systematic error between the surface and tumour motion correlation with the method used in this study, since no outliers or uneven breathing patterns were excluded. To make the surface and tumour motion data more representable, the manual exclusion points should be extracted and excluded from the surface data.

These outliers can affect the ANN model in turn, by giving partially erroneous input data to train on. However, training on erroneous data might have the same effect as an image recognition model training on blurry or noisy images. It could, with enough data, maybe even see through the outliers on its own.

5.4 ANN modelling

The goal of the model is to reduce the loss (function) to as close to zero as possible. It is difficult to know if 7.5 (approximately where the validation loss ends in figure 15a) is a low enough number to be considered low since loss is arbitrary between models. However, it can be compared to the size of the output. The output should be in the order of 1 mm to 10 mm, and for the loss to be in the same order might indicate a relatively high loss function value for the validation loss. It is important to keep in mind, however, that only six patients were used in the validation set which is a very small data set, compared to the training data set that had 36 patients and could achieve a lower loss function value.

The stability of the model loss should therefore be prioritised due to the acknowledged small data sets. It shows how the training is progressing, and both the validation and train loss seem to stabilise (figure 15a) which is important to avoid over- or underfitting.

As previously mentioned, overfitting is a common issue that can occur for ANN models. Here, regularisation parameters, early stopping and changing the architecture was combined

to reduce the overfitting as much as possible. However, the main issue with the fluctuating validation loss (figure 5) is probably the too small dataset used. Comparing the train loss that had 36 patients to train on, with a more stable curve, with the validation loss with 6 patients used, it is a clear difference.

At some point, it did not help adding heavier regulariser terms or less layers or units per layer. In the end, the tumour motion pattern can be complex, and the model can not be too simple either. That is where more patients would be needed to both validate and train the model. Another method of avoiding overfitting is "dropout", where dropout layers are added to exclude a certain number of neurons for training.

There are many ways of constructing an ANN model, and the model designed for this study was made with a simple architecture. This was to reduce complexity and in turn reduce overfitting as much as possible, but a similar effect could probably have been achieved with various manipulations of the regularisation term L1, combining with other regularising methods or adding more patients. This is just one method of optimising ANN models with the data available.

For model accuracy, the validation nor model accuracy was not plotted. This is because the accuracy in this case is the accuracy of the prediction for each point and not an interval, which was not the intention with the model. It is very difficult to exactly pinpoint the prediction and it would not be "fair" to the model to try to predict on only 6 patients during validation. The accuracy, for the purpose of this study, is instead the percentage of data points within the interval which is calculated after model predictions. Even then, the accuracy might be misleading and the MSE should be focused on instead. A large MSE means large loss function and can be interpreted as an uncertainty of the model. The goal is rather to decrease the MSE.

The uncertainty of the model, the MSE, can be compared to the uncertainty of 3D total tumour motion acquired from error propagation. These errors varied between approximately $1 - 3$ mm for all patients, which is the same order of magnitude as the $\sqrt{MSE}$ (2.7 mm). As of now they have comparable errors but hopefully the MSE can be reduced by optimising the model further and introduce a larger training data set.

With all of these results, there is a possibility to apply an action plan for when the patient respiration pattern is irregular or different at the time for treatment compared to the planned situation. An example could be the use of a model similar to the one developed in this study to predict the deviation of the tumour location from the assumed location of treatment, in which an action in term of taking extra images or repositioning might be required. Depending on how well the tumour and surface actually correlates, it would be interesting to see if this could be used further for gating, or even live tumour tracking and adaptive beam repositioning.

6 Conclusion

The correlation between surface and tumour motion was strong ($0.70 < SCC < 1.0$) in a large portion of the studied patients, especially in SI (67% of the patients) and total 3D (73% of the patients) directions. The SCC were not correlated to any patient feature or anatomical position of tumour, but the use of an abdominal compression belt showed statistically significant higher correlation ($SCC = 0.79, p < 0.05$). The correlation coefficient value is varying, and the result should rather be interpreted as an indication of possible further individualisation of treatment and treatment planning.

For those patients that have a strong correlation, surface imaging systems are appropriate to help monitor the patient respiration. With further development of the used model, there is a possibility to apply an action plan for when the patient respiration pattern is irregular or different from planned pattern.

7 Future prospects

In the future, correlation between surface and tumour motion could be used for live tumour monitoring during treatment or gating. Based on an understanding of the surface and tumour motions relative to each other, the respiration pattern can be closely monitored with reduced margins during treatment, and if an altered breathing pattern is detected, treatment can be halted accordingly. This could open opportunities for more conform treatment delivery to spare more healthy tissue surrounding the tumour, thus decreasing eventual treatment side effects.

To further investigate if these steps can be made, a study with a larger patient data set should be made. There need to be larger sample sizes of tumour position in lung lobes, and larger sample sizes of groups with and without compression belt (and paired data as well), to increase statistical power. The impact of abdominal compression should be more investigated as well, and not just be a secondary finding.

Further, the model could be improved with a larger patient data set to both train and validate on more data. With this, the internal uncertainty that is Mean Squared Error could possibly decrease thus minimising the uncertainty interval. This might provide a more robust prediction. The model could further be used and trained on surface data acquired during treatment where x-ray images are taken to confirm tumour position and in return update the model to produce more accurate predictions based on changed breathing pattern.

Acknowledgements

I want to first thank my supervisors - you have been so encouraging and helped me down to earth when I derailed. For all the practical and clinical advice, and for rereading my report more times than I can count. Thank you for helping me channel all my crazy ideas into a master thesis.

I would also like to thank the other physicists at SUS - for always having your doors open and for answering my questions. Thank you for making me feel welcomed.

To my fellow office neighbours Ivan, Sara and Thi - thank you for being the best company, for the encouragement, and for bouncing ideas with me.

To the Medical Radiation Physics' Class of 2023 - these three last years at MSF have been fantastic with you all, and I am endlessly thankful for the support we have shown each other. We did it!

Lastly, to my family and friends - thank you for the immense support and encouragement throughout these five years of studies.

References

[1] Socialstyrelsen. Cancerregistret; 2022. Available from: https://www.socialstyrelsen.se/statistik-och-data/register/cancerregistret/.

[2] Cancerfonden. Statistik lungcancer – dödlighet & överlevnad;. Available from: https://www.cancerfonden.se/om-cancer/statistik/lungcancer.

[3] Cancerfonden. Lungcancer – Symtom, orsaker och behandling | Cancerfonden; 2021. Available from: https://www.cancerfonden.se/om-cancer/cancersjukdomar/lungcancer.

[4] Dieterich S, Ford E, Pavord D, Zeng J. Chapter 17 - SRS and SBRT. In: Dieterich S, Ford E, Pavord D, Zeng J, editors. Practical Radiation Oncology Physics. Philadelphia: Elsevier; 2016. p. 228-40. Available from: https://www.sciencedirect.com/science/article/pii/B9780323262095000171.

[5] Keall PJ, Mageras GS, Balter JM, Emery RS, Forster KM, Jiang SB, et al. The management of respiratory motion in radiation oncology report of AAPM Task Group 76a): Respiratory motion in radiation oncology. Medical Physics. 2006 Sep;33(10):3874-900. Available from: http://doi.wiley.com/10.1118/1.2349696.

[6] Bouilhol G, Ayadi M, Rit S, Thengumpallil S, Schaerer J, Vandemeulebroucke J, et al. Is abdominal compression useful in lung stereotactic body radiation therapy? A 4DCT and dosimetric lobe-dependent study. Physica Medica. 2013 Jun;29(4):333-40. Available from: https://linkinghub.elsevier.com/retrieve/pii/S112017971200035X.

[7] Al-Hallaq HA, Cerviño L, Gutierrez AN, Havnen-Smith A, Higgins SA, Kügele M, et al. AAPM task group report 302: Surface-guided radiotherapy. Medical Physics. 2022 Apr;49(4). Available from: https://onlinelibrary.wiley.com/doi/10.1002/mp.15532.

[8] Kügele M. Surface guided radiotherapy [Doctoral Thesis (compilation)]. Lund: Lund University, Faculty of Science; 2021. ISBN: 9789178959563.

[9] Wang G, Song X, Li G, Duan L, Li Z, Dai G, et al. Correlation of Optical Surface Respiratory Motion Signal and Internal Lung and Liver Tumor Motion: A Retrospective Single-Center Observational Study. Technology in Cancer Research & Treatment. 2022 Jan;21:153303382211122. Available from: http://journals.sagepub.com/doi/10.1177/15330338221112280.

[10] Vedam SS, Kini VR, Keall PJ, Ramakrishnan V, Mostafavi H, Mohan R. Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. Medical Physics. 2003 Apr;30(4):505-13.

[11] Fayad H, Pan T, Clément JF, Visvikis D. Technical note: Correlation of respiratory motion between external patient surface and internal anatomical landmarks. Medical Physics. 2011 Jun;38(6):3157-64. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3379968/.

[12] Paolani G, Strolin S, Santoro M, Della Gala G, Tolento G, Guido A, et al. A novel tool for assessing the correlation of internal/external markers during SGRT guided stereotactic ablative radiotherapy treatments. Physica Medica. 2021 Dec;92:40-51. Available from: https://www.sciencedirect.com/science/article/pii/S1120179721003379.

[13] Bishop CM. Pattern recognition and machine learning. Information science and statistics. New York: Springer; 2006.

[14] Peters RM. The mechanical basis of respiration: an approach to respiratory pathophysiology. 1st ed. Boston: Little, Brown; 1969.

[15] Jones J. Lung | Radiology Reference Article | Radiopaedia.org; 2011. Available from: https://radiopaedia.org/articles/lung?lang=us.

[16] Dieterich S, Ford E, Pavord D, Zeng J. Respiratory Motion Management for External Beam Radiotherapy. In: Practical Radiation Oncology Physics. Elsevier; 2016. p. 252-63. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780323262095000195.

[17] Cox JD, Chang JY, Komaki R. Image-guided radiotherapy of lung cancer. New York: Informa Healthcare; 2008. OCLC: 183179117.

[18] Li H, Noel C, Garcia-Ramirez J, Low D, Bradley J, Robinson C, et al. Clinical evaluations of an amplitude-based binning algorithm for 4DCT reconstruction in radiation therapy: Amplitude-based binning algorithm for 4DCT reconstruction. Medical Physics. 2012 Jan;39(2):922-32. Available from: http://doi.wiley.com/10.1118/1.3679015.

[19] Brahme A, Nyman P, Skatt B. 4D laser camera for accurate patient positioning, collision avoidance, image fusion and adaptive approaches during diagnostic and therapeutic procedures: Laser camera image fusion and adaptive therapy. Medical Physics. 2008 Apr;35(5):1670-81. Available from: http://doi.wiley.com/10.1118/1.2889720.

[20] Blais F. Review of 20 years of range sensor development. Journal of Electronic Imaging. 2004 Jan;13(1):231. Available from: http://electronicimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.1631921.

[21] Bonetto R, Latzko V. Machine learning. In: Computing in Communication Networks. Elsevier; 2020. p. 135-67. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780128204887000219.

[22] Jianxin W. Introduction to convolutional neural networks.. National Key Lab for Novel Software Technology. Nanjing University. China; 2017.

[23] Giurgiutiu V. SIGNAL PROCESSING AND PATTERN RECOGNITION FOR PWAS-BASED STRUCTURAL HEALTH MONITORING. In: Structural Health Monitoring. Elsevier; 2008. p. 589-656. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780120887606500131.

[24] Santosh K, Das N, Ghosh S. Deep learning: a review. In: Deep Learning Models for Medical Imaging. Elsevier; 2022. p. 29-63. Available from: https://linkinghub.elsevier.com/retrieve/pii/B978012823504100012X.

[25] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014. Publisher: arXiv Version Number: 9. Available from: https://arxiv.org/abs/1412.6980.

[26] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, et al.. Evaluation of measurement data — Guide to the expression of uncertainty in measurement;. Published: Joint Committee for Guides in Metrology, JCGM 100:2008. Available from: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6.

[27] Griffiths D. A Pragmatic Approach to Spearman's Rank Correlation Coefficient. Teaching Statistics. 1980 Jan;2(1):10-3. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9639.1980.tb00369.x.

[28] Gauthier TD, Hawley ME. STATISTICAL METHODS. In: Introduction to Environmental Forensics. Elsevier; 2007. p. 129-83. Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780123695222500063.

[29] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020 Mar;17(3):261-72. Available from: http://www.nature.com/articles/s41592-019-0686-2.

[30] Weiss E, Hess CF. The Impact of Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) Definition on the Total Accuracy in Radiotherapy: Theoretical Aspects and Practical Experiences. Strahlentherapie und Onkologie. 2003 Jan;179(1):21-30. Available from: http://link.springer.com/10.1007/s00066-003-0976-5.

[31] Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, et al. Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. Radiotherapy and Oncology. 2010 May;95(2):166-71. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167814010000058.

[32] Niven EB, Deutsch CV. Calculating a robust correlation coefficient and quantifying its uncertainty. Computers & Geosciences. 2012 Mar;40:1-9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0098300411002251.

[33] Njeh CF. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. Journal of Medical Physics / Association of Medical Physicists of India. 2008;33(4):136-40. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2772050/.

8 Appendix A - Additional figures

In this section, figures not included in the report but still used as a basis of discussion are included. Figures included are more plots of patient data used in the ANN model, and examples of how the respiratory data from the Sentinel looks like when plotted to illustrate the potential irregularity of the individual patient.

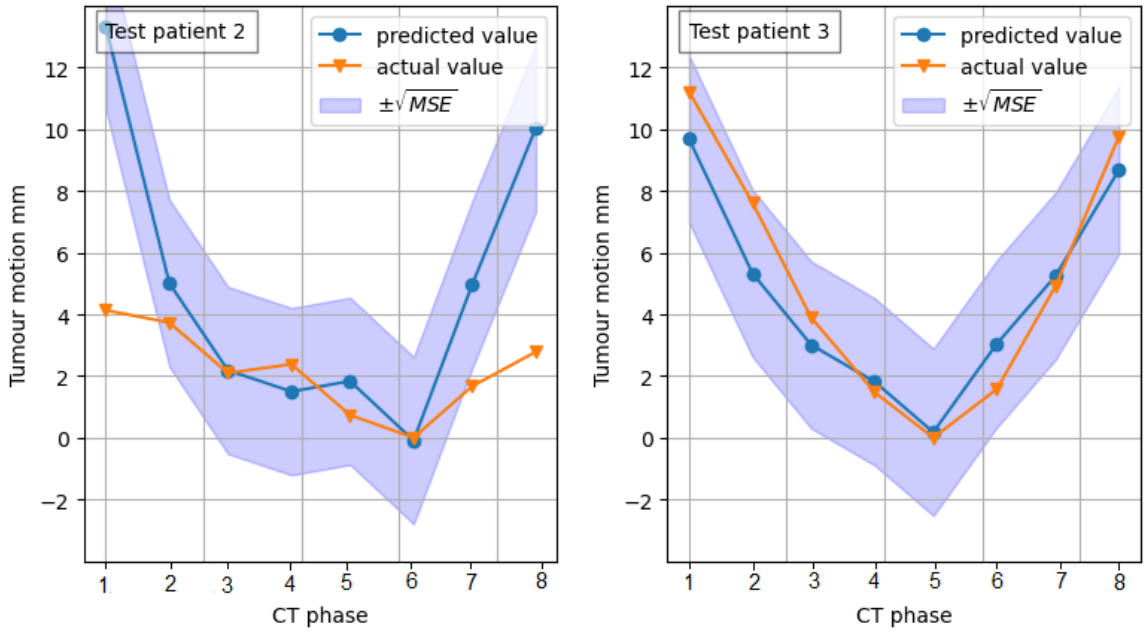
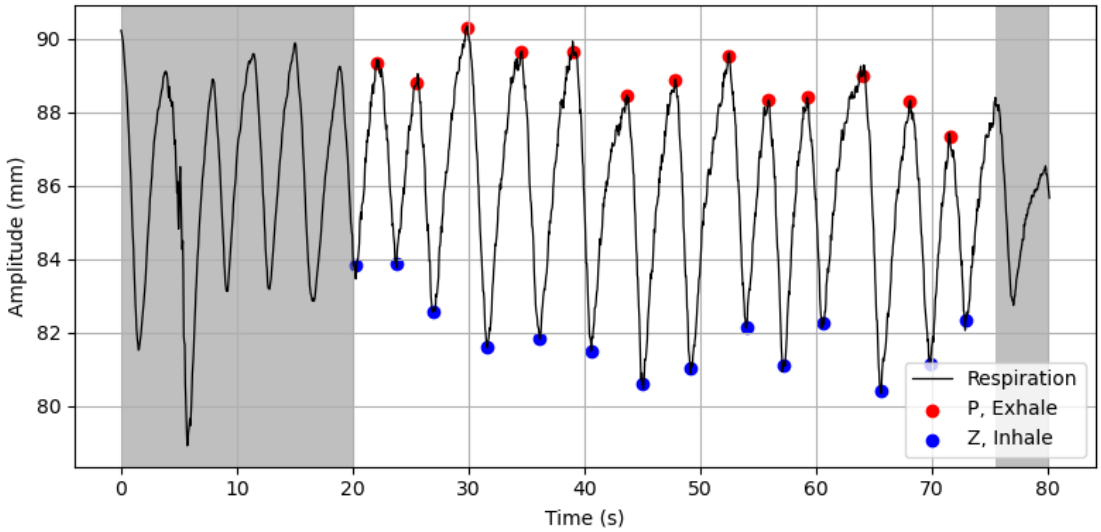
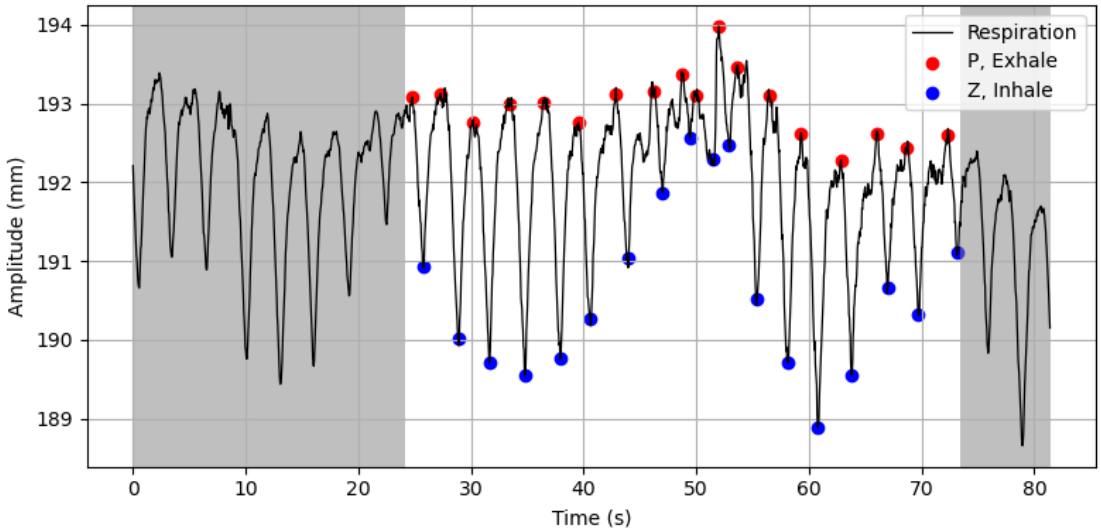


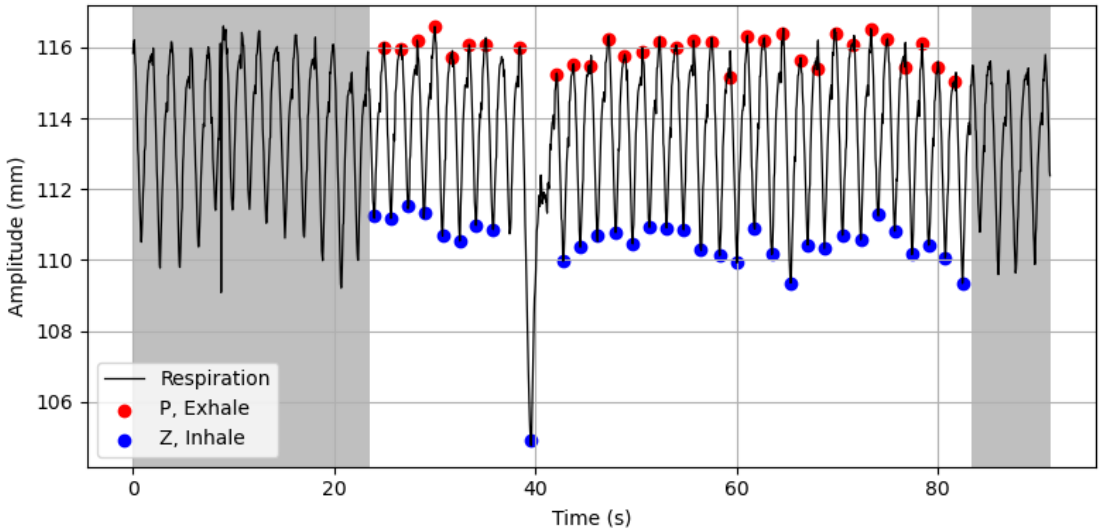
Figure 16: Plot of predicted and actual values alongside the predicted interval of tumour placement. Test patient 2 and 3 used for validation of ANN model. X-axis is CT-phase number and y-axis is tumour motion in mm. ANN-model predicted value has circle points and actual values have triangle points. Shaded area marks the square root of mean squared error $\sqrt{MSE}$ interval and it is centered around the predicted values.



(a) Sentinel breathing motion data for a patient with even breath pattern and no outliers.



(b) Sentinel breathing motion data for a patient with uneven breath pattern.



(c) Sentinel breathing motion data for patient with outliers

Figure 17: Sentinel breathing motion data for different patients. This is a similar plot to what is shown in figure 4. Blue and red markers marks inhale and exhale maxima, gray area represents where breath motion recording starts outside 4DCT acquisition. White area between represents breath motion recording during 4DCT acquisition. Important to note is that the data is upside down.