RAM ROBOTICS AND MECHATRONICS

ESTIMATION OF LIVER RESPIRATORY MOTION USING A SURROGATE SIGNAL BASED ON A DEEP LEARNING APPROACH

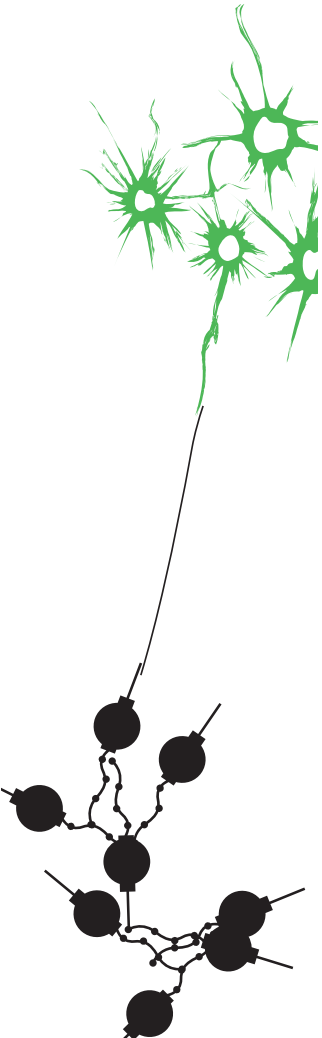
G. (Georgios) Lappas

MSC ASSIGNMENT

Committee:
prof. dr. ir. C.H. Slump
dr. ir. H. Naghibi Beidokhti
dr. ir. M. Abayazid
dr. B. Sirmaçek
dr. ir. B.J.F. van Beijnum

January, 2020

002RaM2020
Robotics and Mechatronics
EEMathCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands



UNIVERSITY OF TWENTE. | TECHMED CENTRE UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

Estimation of liver respiratory motion using a surrogate signal based on a Deep Learning approach

Georgios Lappas[1]

Abstract
Liver intervention can become a challenging task due to the respiration induced motion. The latter causes misalignment between the interventional mapping obtaining pre-treatment and the changed anatomical parameters during application phase (liver biopsy or radiotherapy) leading to increased damaged of healthy tissue as well as inaccurate targeting of hepatic tumors. In the presented work, Respiratory Motion Estimation is exploited where using external signals (surrogates), it is possible to estimate the liver actual motion. The proposed work has been evaluated in several breathing patterns in comparison with previous studies making usage of ultrasound (US) sensor as surrogate, placed on the human's abdominal region. Next, three regression models (simple linear regression, polynomial fitting, single layer perceptron) were utilized to correlate the liver motion with the US signal and consequent trained to estimate the superior-inferior (SI) motion of the liver upper border available in 2D Magnetic Resonance Imaging (MRI) sagittal images. Additionally, extending the conventional framework and taking advantage of Deep Learning and more specifically Long Short-Term Memory (LSTM) networks, it is feasible to predict the liver motion in a short future state combined with a classifier that can detect the performed respiration type. The proposed DL approach has been validated in MRI on ten healthy human subjects when the findings revealing an estimation of the liver motion in SI direction with a Root Mean Square Error (RMSE) accuracy below 1.2 ± 0.2 mm (95% CI) and a capability of liver motion prediction for 6 sec ahead enabling a safer examination decreasing the likelihood for potential risk during an image-guided intervention.

Keywords
Respiratory Motion Estimation (RME) — Liver — Deep Learning (ML) — Long Short-Term Memory (LSTM) — Regression — k-nearest neighbors (k-NN) — Surrogate Signals — Magnetic Resonance Imaging (MRI) - Ultrasound (US)

[1] *Department of Robotics and Mechatronics (RaM): Medical Robotics Group, University of Twente, Enschede, The Netherlands*

Contents

1. Introduction

1.1 Liver Cancer

Liver cancer is defined as a substantial and costly healthcare issue from the World Health Organization. In 2018, it was included 1st in the list of most common cause of cancer death counting almost 780000 deaths worldwide [1]. Research and study over the last decades, open the way up for improved diagnosis, treatment planning and prevention. Advanced imaging techniques have been exploited utilising Machine Learning, image processing and signal analysis methods to enhance the physicians' vision and subsequently leading to better healthcare services. Although the continuous development in that field, abdominal and thoracic region-based interventions remain still a drawback for the healthcare stakeholders as problems showing off during image acquisition such as one used for hepatic lesions scanning and image-guided interventions like biopsy, tumor's ablation and radiotherapy [2, 3, 4]. To step into the studied problem, it is important firstly to address the effects of an inaccurate clinical evaluation of a liver-diseased patient. Starting with the image acquisition, Magnetic Resonance(MR), Ultrasound (US), Computed Tomography (CT) or X-ray are the most common imaging techniques used in such cases while organ's induced motion may hide blurring or ghosting artefacts in the generated images [5, 6].

While some of the available imaging techniques provide high temporal resolution which is translated to high frame rate and capturing better the motion and properties of an organ, others have higher spatial resolution that is connected with a higher image quality. The needs of physicians, for an enhanced diagnosis, are high image quality and real-time acquisition frame rate but unfortunately, those two requirements cannot be present simultaneously within one imaging technique yet. For instance, MRI is preferred due to high visibility advantage of soft tissue based on high spatial resolution while US is commonly used for its high temporal resolution [7, 8, 9].

Additionally, inadequately targeting for image-guided interventions has an adverse risk as it may create misalignment and discontinuities between the fixed guidance information and the anatomical adaptations of the target organ when a lot of approaches use the assumption of following the same respiratory motion at every breathing cycle [10, 11]. For instance, during a radiotherapy beams are hitting the human body to treat the cancerous cells of a tumor but on the other hand due to the anatomical changes the beam gets misaligned at each step causing increased damage of healthy tissue. In some cases, the physicians circumscribed a larger area around the actual tumor to ensure that the target will be covered at each treatment session independent the anatomical adaptations. This can create risk for increased radiation dose or insufficient targeting. When there are no clear guidelines regarding the doctors' requirements for a sufficient liver biopsy or an accurate radiotherapy, in the presented work the following parameters have been taken into account: real-time respiratory liver motion prediction, minimization of the error rate independent of the breathing pattern [11, 12].

In most cases, a liver biopsy takes approximately five minutes when the doctor prefer an imaging machine to guide the needle with high update rate that ideally can decrease the number of penetrations and the duration for collecting the tumour sample. Minimizing the procedure duration and the potential error of the doctor due to liver respiratory motion will guarantee less likelihood for internal bleeding or haemorrhage which are mentioned as one of the most common causes for morbidity after a percutaneous liver biopsy. In previous studies, the error was varying from 0.7mm till 2.5mm depending on the direction and evaluated only in free-breathing patterns on healthy subjects and on liver phantoms. In the presented work, the proposed approach shows promising results assuming a minimization of the error about 15-20% based on several breathing patterns. [10, 13]. In the following sections of the introduction part, the fundamental parameters that lead to decreased accuracy diagnosis will be explained from the theoretical perspective along with the current solutions and limitations.

1.2 Problem Statement

Starting with the simplest method to minimize the organ motion because of respiration, breath-holding can be useful whereas its disadvantage is the limited application time, restricted to no more than 30 sec that has been proved to be an insufficient time interval for real-time examinations/interventions. Moreover, this technique gives an uncomfortable feeling in many situations resulting in an inadequate solution. Secondly, gating has been utilised to deal with that issue. This method involves image acquisition using a fixed window, capturing the end-inhalation (EI) or the end-exhalation (EE), relying on external signals. Although gating seems to be appropriate enough to solve the problem, it introduces a higher time acquisition duration in order to capture complete breathing patterns based on different respiration phases. Motion tracking is an alternative proposed technique that uses markers to cope with the respiratory motion. It can be either invasive or non-invasive based on the nature of the markers (fiducial markers or externally placed onto the targeted area). After markers placement, the motion is tracked using an imaging device compatible with US, CT or X-ray. Except of the invasiveness nature of the method, motion information can be only available for the restricted region of markers rather than the region of interest (ROI) [10, 14, 15].

An inaccurate target detection will subsequently cause increased damage of the healthy cells and tissues, insufficient treatment planning and possibly higher likelihood of recurrences [16, 17]. To dive deeper into the problem, as stated previously, most organs in the human body are susceptible to changes in their structure as well as alterations in the motion based on the breathing pattern that is performed at each time period. This is caused because of two physiological human body's functions. Firstly, the diaphragmatic muscle located

in the thorax gets contracted on the inhalation phase and the rib cage muscles start moving which secondly create an increase in thoracic volume inhaling air into the lungs [18]. Of course, variations are possible depending on the following parameters:

- subject's posture
- performed breathing pattern
- variations between individuals
- motion of other organs
- changes in the relative contributions and magnitudes

Taking into account those adjustments, a brief description for the variations regarding the respiratory pattern is given in the following section. Starting with the intra-cycle variations or better explained as variations of motion within a single breathing cycle, this refers to the different motion's paths that are followed during inspiration and expiration. Similarly, there are also inter-cycle variations and are related to variations of motion between different respiratory cycles where the motion's paths varies between one breathing cycle and another cycle [19, 20, 21, 22, 23, 24, 25]. Respiratory motion affects mainly the thoracic and abdominal areas' organs and based on previous work, can have on average a displacement of 16.5 mm in one direction following a shallow breathing pattern [4, 26, 27]. An example can be found in Figure1 where the liver motion in the vertical direction is presented as displacement of the upper liver surface indicated with the blue grid. This difference in position has been created based on the liver motion between inhalation and exhalation. Moreover, further studies findings indicate a liver Superior-Inferior (SI) motion to vary between 10.0-21.3 mm and based on the outcome using a robotic phantom simulation, the respiratory liver motion has been measured for a displacement of 10-40 mm in SI direction, 1-12 mm in Anterior-Posterior (AP) direction and 1-5 mm in Medial-Lateral direction [4, 10, 26, 28, 29].

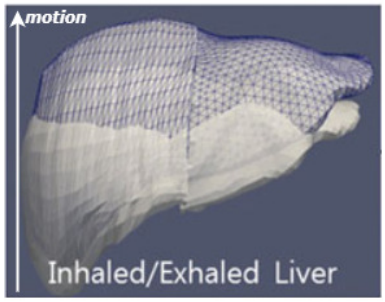


Figure 1. Liver motion representation due to different phases of breathing. Liver upper surface during inhalation is presented with blue grid while the corresponding position during exhalation is shown with white color [30].

1.3 Respiratory Motion Estimation (RME)

Due to the aforementioned limitations and downsides, the healthcare stakeholders investigate further techniques to compensate with the RME. This can be done by modeling the relationship between the motion of interest (e.g organ actual motion over time) with a surrogate signal (e.g displacement of a marker over time). The correlation between the two data will generate a model that can estimate the internal organ motion using only the surrogate signal.

Deepen to the problem, the correlation between the two signals is determined by a set of parameters generated on the training section where the internal motion and surrogate signals are fed simultaneously to a fitting method. This training will be performed offline while on the test phase or during the intervention as better described, an estimation of the internal motion data will be performed based only on the surrogate signal as depicted in Figure 2. The corresponding figure splits into two sections, the training and prediction phases. In training phase which is performed offline, the liver motion data are acquired simultaneously with the surrogate data presented as A-mode-US wave [31] measurements and fed to the regression models. In the training phase the regression will learn some parameters that represent the correlation between the two signals and the regressors will create the so-called motion model which simulates the correlation of the two signals. During the prediction phase, the trained motion model will be loaded and fed only with the surrogate signal. This will attempt to estimate the internal liver motion based on the learned parameters of the training phase.

As shown in Figure 2, every RME approach consists of four sub-processes:

1. Internal motion selection: the targeted internal motion, commonly with a high spatial resolution but also including low temporal resolution [32].

2. Choice of surrogate(s): external signal having a strong correlation with the internal motion data but it is not possible to directly be measured. Usually, the surrogates have high temporal resolution, thus is an advantage for their choice in those applications.

3. Motion model: it is often called correspondence model and it is a mathematical formula that can describe the correlation between the internal motion data and the surrogates through some parameters. A further explanation is given in 1.4.

4. Fitting method: this is related to the method that the correspondence model utilizes to optimize the fitting process of surrogate to the training data.

When all the aforementioned are filled out, the motion estimates can be calculated and the prediction of the internal motion can be performed based solely on the surrogate data and the learning parameters of the motion model found in the training phase.

1.4 Correspondence model

To do this, a correspondence model needs to be generated that represents mathematically a strong relationship between the

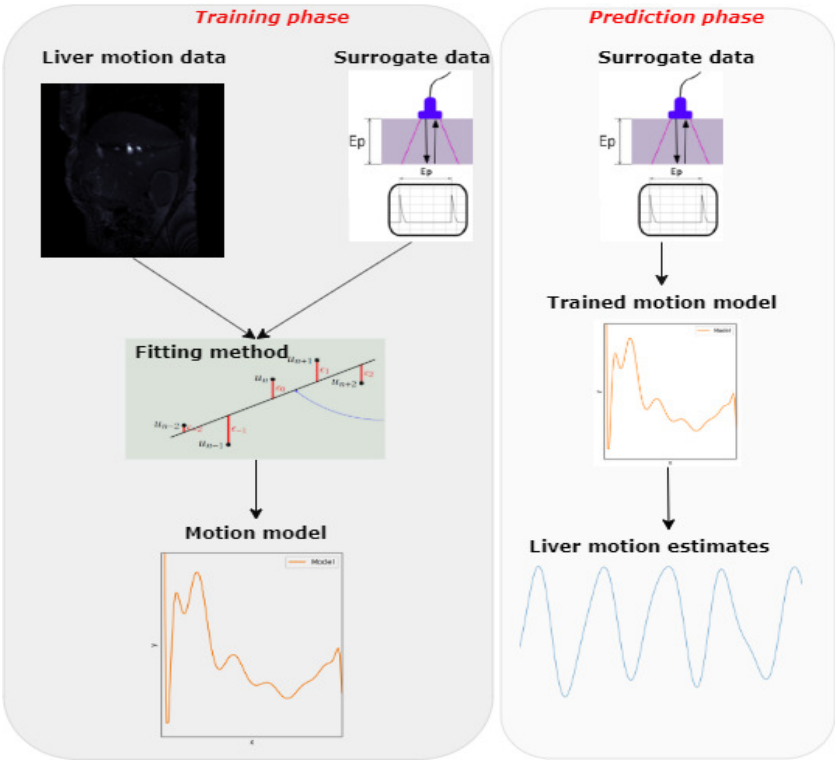


Figure 2. Overview of the RME process. In training phase the liver actual data are acquired simultaneously with the surrogate data (A-mode US waves) and fed to the fitting algorithm/regression model. The algorithm will calculate some parameters standing for the correlation of the two signal and will create the so-called motion model. Next, in the prediction phase, the optimal parameters found in previous step will be used combined only with the surrogate signal this time to estimate the liver motion.

internal motion data or the target location and the surrogate data. This relationship can be approximated either 'directly' (see Figure 3) or 'indirectly' [11]. Since the direct model is utilised in the presented work and due to limitation space, the indirect model will no be depicted but it will be explained briefly in the following lines. As shown in Figure 3, using the direct correspondence model, the learning parameters of the motion model will have a direct linear or non-linear behavior regarding the two signals. Briefly, for a direct correspondence model, the relation between the internal motion and the surrogate can be formed as:

$$\boldsymbol{M}(t) = \phi(\boldsymbol{s}(t)) \tag{1}$$

where **s(t)** is the surrogate signal, ϕ stands for the direct correspondence model and **M(t)** is the estimate of the motion (a vector of the target position at a specific timestamp). Note that the amount of degrees of freedom (DoF) corresponds to the model is subjected to the amount and type of the surrogate signal(s) and that in direct correspondence, the values of the surrogate data directly schematize the target motion estimation parameters [33].

Except of direct correspondence, there is also the indirect correspondence models that schematizes the target motion based on a number of internal parameters determining the DoF of the motion model [34]. In this case, there is no direct measurement of the internal variables during the motion model adjustment to approximate the internal motion but the surrogate is a subgroup of the motion estimates done by the motion model and in order to figure out the best approximation between the measured surrogate(s) and the estimates of the surrogate(s), the internal parameters are optimized by the motion model. For indirect correspondence model, the formula that describe this function can be written as:

$$\boldsymbol{M}(x(\hat{t})) = \phi(s(\boldsymbol{x}(\hat{t}))) \tag{2}$$

where $x(\hat{t})$ is the vector of internal variables, such as position in respiratory phase, $\phi(x)$ is a vector of motion variables determined by the internal variables. The idea behind it is the generation of a reference image followed a transformation based on the motion variables and subsequently a function handles the output to reproduce the surrogate(s) [35, 36, 37].

1.5 Previous work

As previously stated, most of the studies focused on the changing respiration breathing patterns within one breathing cycle (intra-cycle variations) but also between different respiratory phases (inter-cycle variations) when to a greater extend, some researchers emphasize on the alteration of the breathing patterns between patients as well [11]. To tackle the respiratory

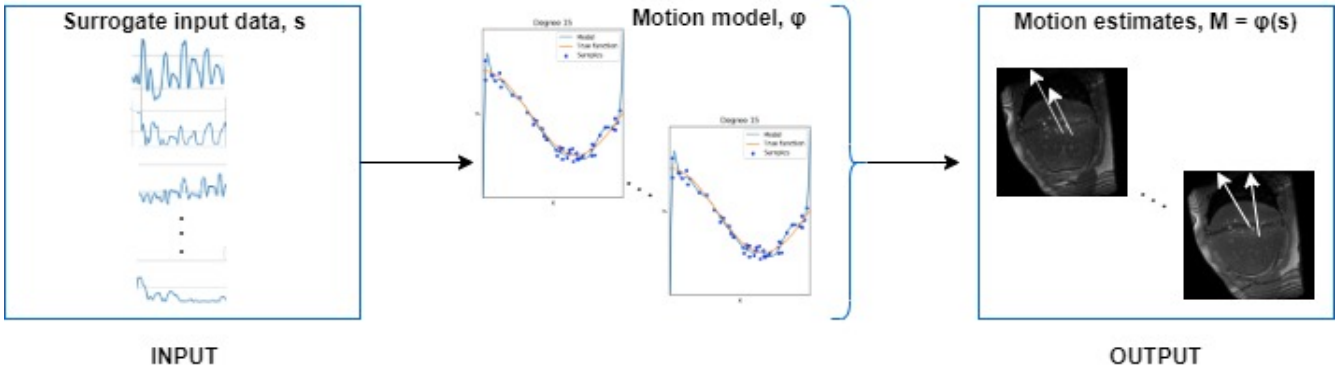


Figure 3. Direct correspondence model function. This model estimates the internal motion directly using the surrogate signal resulting in a linear or non-linear relation between the two signals.

motion problem, several solutions have been proposed. A summary of the related studies about the motion data, the surrogate data and the validation experiments conducted, is presented in Table 1.

To start with the most recent work, Berijanian et al. used a robotic phantom to simulate the liver motion in SI and AP direction using actuators. As a surrogate data, the authors exploited optical tracking via a camera detecting skin markers and additionally an inertial motor unit (IMU) placed on the needle hub inserted to the liver phantom. In the corresponding work, linear fitting methods have been used and more specifically linear regression as well quadratic polynomials. On the downsides, the model validations have been applied to a liver phantom showing discontinuities from inhalation to exhalation and vice versa. To continue with Abayazid et al. study where the authors developed a motion model using data from an Electromagnetic (EM) tracker as internal motion, IMU sensor for acquiring the surrogates and the Random k-Labelset as fitting method. For model evaluation, a motion phantom has been used and the limitations of this work are related to the needle bending that used integrating the IMU sensor. The needle bending can affect the relation between the measured external motion of the needle hub and the actual motion of the tip, leading to misalignment.

Fahmi et al. in the same year, exploited MRI data as actual motion combined with camera-tracked external markers data. After validation on human subjects, the authors estimated the liver motion with a Mean Absolute Error (MAE) of 2 mm but this approach may fail with real data as the dataset for training/testing was consisted only of 3 subjects resulting in an non-generalizable and robust motion model to unseen data. Additionally, Chen et al. presented two motion models applying linear and ridge regressions feeding accelerometers with bellow data as surrogates and MRI images as the actual motion. Moreover, Shin et al. research follows testing a correspondence model of CT-scans and digital protractors with calipers on a motion phantom when Preiswerk et al. took advantage of the same data type for internal and surrogate motion as in the presented work. For the former's and latter's work, the main limitation was that the breathing patterns adaptations which have not been taken into account, thus including only one motion model for all the different breathing patterns proved to be insufficient for accurate liver motion estimation. A study on animal subjects conducted by Lei et al. in 2012 expanded further the research in the corresponding field while the findings cannot be considered adequately enough to perform interventions on human subjects due to respiration discontinuities and other factors. Concluding with the work of Buerger et al. in the same year, the researchers used MRI as the internal motion representation and also as the surrogate data while the general framework has been tested on human subjects [9, 10, 26, 13, 38, 39, 40, 41, 42].

1.6 Objectives and Research Question

This study attempts to continue the work for the liver motion estimation using surrogate signals based on a machine learning model on that basis. [10, 26]. The findings will contribute into further improvement in image-guided interventions by minimizing the damage of healthy tissues as the model framework can be adaptive to different breathing patterns. In addition, the framework of the corresponding work has also the capability of liver estimation on a future state and not only on the present leading to a better treatment planning and open the way for further research regarding the independent MRI compatibility intervention as the patient's liver motion state can be estimated for a short time interval after his/her removal from the MRI-room. Furthermore, the model will be evaluated on patients data and not in phantoms or animal subjects that presented in previous research. This will give an intrinsic evaluation of this method. The objectives of this work are to investigate if an ultrasound transducer signal can be used as surrogate data that has a strong correspondence with the actual liver motion. In the same context, three different regression models are evaluated on experimental data of healthy subjects along with Deep Learning approach for surrogate prediction and finally a classifier to test for its capability on detection of different breathing patterns.

As a result, the main research question is the following: "What is the performance of a Deep Learning approach in liver motion estimation for different types of respiratory motion?".

Table 1. Table of related work on liver Respiratory Motion Estimation (RME) presenting the internal motion and surrogate data representation along with the validation applications.

Reference	Internal Motion Data	Surrogate Data	Validation on
Berijanian et al., 2019	IMU	Optical Tracking & IMU	Motion Phantom
Abayazid et. al., 2018	EM tracker	IMU[1]	Motion Phantom
Fahmi et. al., 2018	MRI	External Markers	Human Subjects
Chen et al., 2017	MRI	Accelerometers & Bellows	Motion Phantom & Human Subjects
Shin et al., 2017	CT-scan	Digital Protractor & Calipers	Motion Phantom
Preiswerk et al., 2017	MRI	US	Human Subjects
Durichen et al., 2013	Ultrasound	Multi-modal sensors	Human Subjects
Lei et al., 2012	CT-scan	EM tracker	Animal Subjects
Buerger et al., 2012	MRI	MRI	Human Subjects

There are also some sub-questions that were arisen:
i) "Does an ultrasound sensor signal have strong correlation with the liver internal motion data? If yes, can its data describe sufficiently enough the liver internal motion due to respiration?"
ii) "To what extent the performed breathing pattern can affect the surrogate signal prediction performance in terms of accuracy?"
iii) "What is the performance of a classifier in breathing pattern detection?"
At the end of Chapter 1, the respiratory motion problem has been analyzed along with the conventional approaches, detailing the suggested framework of the corresponding work. Moreover, the goals of the corresponding work have been analyzed along with the contribution and objectives of this thesis. The outline of the rest work is organized into five main chapters. In Chapter 2, all the necessary information for the suggested surrogate data and prediction sections are provided. Moreover, in the same chapter, an brief explanation for the classification step, placed into the activity recognition systems field, is given along with all the previous steps that are needed. Next, in Chapter 3, the design of the suggested research approach, the data acquisition and the workflow are presented. In the same section, all the applied processing techniques are shown followed by the temporal synchronization methods and an explanation about the evaluation metrics used in the motion model for the correlation of surrogate and internal motion data. The second part focuses on the classification work that has been done, presenting all the intermediate steps and the evaluation framework that has been followed. In Chapter 4, quantitative results, interpretations and analysis of the two model's outputs are provided followed by the final main sections, Chapter 5 including the recommendations for future work and Chapter 6 including the final findings. At the end, in the Appendices, a summary of the measurement protocol, with the description of the MRI and ultrasound devices, is given. Additional information regarding the equipment and the designed protocol can be found in G.Veenstra study [43].

2. Theoretical Background

In this Chapter, the available surrogate data will be explained and the choice of the selected surrogate signal will be justified. Moreover, the theoretical background for the surrogate prediction compartments will be provided followed by an in depth model explanation.

2.1 Surrogate Data

It is important to emphasize that although the organ actual motion can be acquired with a low frame rate (low temporal resolution), the surrogate can be proved helpful because of its high temporal resolution. This can be significantly beneficial in many applications where the physicians cannot acquire the organ's actual motion using an acceptable high temporal resolution. For instance, imagine that a patient can be removed from the MRI-bore and using only prediction for the surrogates, the doctors are able to estimate the internal motion of the targeted region [16]. The following criteria have been taken into account for selecting the appropriate surrogate signal in the presented work: imaging modality independent, non-discomfort causes to the participants and drift-free signal [44].

2.1.1 Available Surrogates

Starting with the available surrogate data, there are either scalar surrogate data or higher dimensional data. Since the focus of this research is based on one surrogate signal, there will be no analysis for higher dimensional surrogates.

In most cases where MRI data represent the actual motion, the physicians need a MRI-compatible device to measure the surrogate. All non-invasive methods that have inference of the target position relying on the respiratory surrogate(s) data belong to this category, Thus, it is common to utilise either MR echos or respiratory bellows. MR echo makes use of the human body property containing water and a relatively small area is magnetised in order to measure the position of the targeted region over time [45]. Regarding the respiratory bellow, it measures the inhaled and exhaled air flow using an air filled bag placed between subject's abdomen and a rigid surface that circumscribes subject's body [46].

Furthermore, in the same philosophy, spirometers are used but mostly to correct the motion applied in radiotherapy field

[47, 48]. Optical tracking technology, such as infrared camera usage or laser based tracking tracking systems are alternatives for surrogate's acquisition, providing the option to acquire more than one points on the abdominal or thoracic regions [49, 50, 51, 52, 53, 54] . In addition, accelerometers are alternatives surrogates used during percutaneous interventions and MRI-compatible, making them a sufficient candidate for surrogate signal [9, 39, 55].

Opposed to the previous methods, fiducial markers are used implanted (invasive) in the target area or close to that providing a more robust vision when the target cannot be visualized using imaging techniques while previous studies on phantom robot showed the potentiality of electromagnetic (EM) tracking systems using electrical signals. In this case, the author of the corresponding work used a phantom robot simulating the liver and integrated a 5 DoF EM sensor into the needle's hub to measure the displacement between the reference needle tip and the target as well as the needle insertion angle [13].

2.1.2 Surrogate Selection

All the aforementioned surrogates raised problems and created drawbacks, either related to MRI-compatibility or for their low correlation to respiration motion while some patients have stated discomforts during the examination and in many cases, the approach's nature and its materials are cost-effective for the healthcare system. In addition, some of them although they have been designed for respiratory gating of MR image acquisition, their application was restricted after extensive studies proving technical issues related to lack of information about breathing amplitude [32, 39, 42, 49, 56, 57, 58, 59, 60]. As a result, in the presented work, an US sensor has been chosen to be studied and used. The reasons are clear as this approach offers high frame rate, it is relatively cheap, MRI-compatible and non-invasive. Moreover, as the goal was to find a surrogate signal that can also predict the liver motion in a state future, US sensor has an ability of real-time outcome that can potentially allow image-guided procedures outside of the MRI-bore.

2.1.3 Artificial Neural Networks

In the corresponding thesis, neural networks have been utilised for the surrogate prediction and while for the the surrogate classification and fitting methods between the liver motion and surrogate signal machine learning algorithms have been exploited. The former that are analyzed in this section belong to the category of Artificial Intelligence and specifically to Artificial Neural Networks.

A neural network consists of connected units which called artificial neurons or perceptrons and are likened to the brain's neurons. These neurons are grouped in layers and are the main computational blocks. Every neuron is composed of an axon which produces the output of the neuron, the dendrites which transfer the input signals to the neuron and the synapses which is located between the axon of the previous neuron and the dendrites of the next one and it is responsible for the communication between the neurons. For an artificial neural network, the synapses can be translated to weights that change during training to approach the best solution for the given problem, the axon is represented by the bias which can decide whether to activate the neuron or not and the dendrites can be seen as the connections between the inputs and the weights that will be forwarded to the main body of the neuron. Then all the processed input data are added together and the value is passed through the decision procedure of the axon. If the sum which can be shown in Figure 4 is above a specific threshold value the neuron or perceptron is activated, otherwise it not.

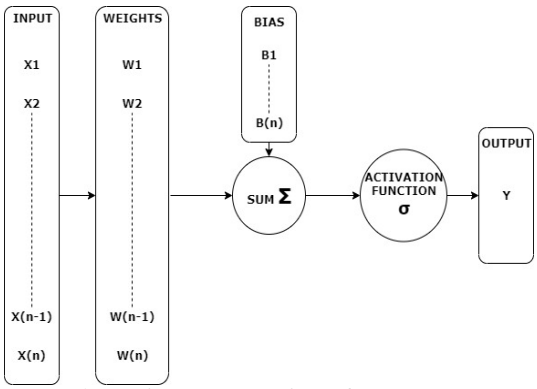


Figure 4. Schematic representation of a perceptron. A series of inputs $(X_1, ..., X_n)$ are multiplied with the corresponding weights $(W_1, ..., W_n)$ and are summed up after the bias $(B_1, ..., B_n)$ parameter addition. This sum $(\sum)$ will be fed to the activation function (σ) of that neuron and then it will be determined if the output neuron y will be fired or not based on a threshold value.

In the same content, a neural network is an application of ML field which focuses on deployment of models that have the capability of automatic learning via self training based on a given dataset without user-intervention or programming it explicitly [61]. ML focuses on deployment of models that learn by themselves through feeding them data. In comparison with the rules-based systems, such as one system that requires human intervention to code the knowledge into it, ML algorithms learn how to take decisions based on the given data. ML can be categorized in four algorithm groups: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. In the present work, the supervised ML algorithm is used which means that during training the correct answer is also provide to the computer. In addition, in ML lot of different NN architectures are available, such as convolutional neural networks [62, 63], deep belief networks, fully convolutional networks [64] or a Recurrent Neural Network which is used in the presented work.

2.1.4 Recurrent Neural Network

These models are called Recurrent Neural Networks (RNN) and take up to maximize relevant information of the input data to the output while this happens based on previous states information. A famous type of RNN which performs better than

the conventional model is called Long Short-Term Memory (LSTM) and its advantage over the other is the memory cells inside the LSTM units which allow longer time lags memorization [65]. In order to store long memory time stamps, a stack of LSTMs is used as shown in Figure 5.

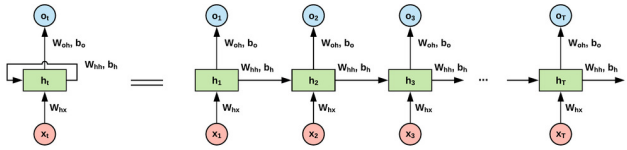


Figure 5. Stack of Recurrent Neural Network, folded and unfolded. The inputs x, stands for a type of sequential data while the outputs o at each section of the network is the prediction of the next state based on the previous one. The main blocks, called h and contain the weights and the activation functions of the RNN model while the black arrows between them represents the communication among them, from one step to the next one. Notice that although the unfolded stack of RNNs shows many different blocks, the h block is the same and returns its output back to its input [66].

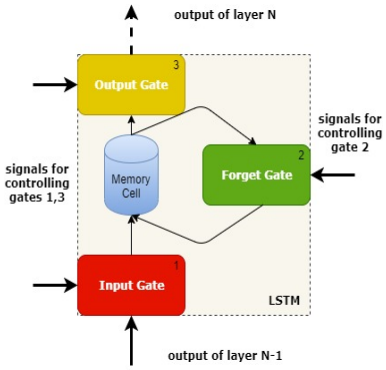


Figure 6. Schematic representation of an LSTM unit, called h-block in Figure 5. Every LSTM block has four inputs: previous state output, signal for controlling the input and output, plus the forget gate control signal. The input gate is used to store the new information, the output gate is used as a controller for the affection of stored value to the output activation. Additionally, the forget gate decides how important is a new information stored on the block determing its removal or not at each step.

2.1.5 Network's Layers

A neural network consists of different types of layers and compartments. For instance, the presented model has LSTM units, dropout layers and also in the last part a fully connected or dense layer with a linear activation function.

- **Input layer**: the input layer is the first connecting point of the dataset with the network. In the presented work, the dataset of the surrogate is consisted of one dimensional vectors containing floating positive values.
- **Normalization layer**: this part belongs to data preparation framework and cannot be skipped since it normalizes the data to follow a uniform distribution with mean value of one and zero standard deviation. The goal behind this idea is to scale all values so the complete dataset will lie down on a common scale without high variations in the ranges of the values.
- **LSTM layer**: every LSTM unit has three gate blocks. The first one or the "input gate" controls if the input of the new information can be memorized. The second one is called "forget gate" and determines how long certain values can be held in memory. Final the "output gate" controls how much the value stored in memory affects the output activation of the block. A graphical representation is given in Figure 6 and focusing on the black arrows outside of the LSTM unit, it is easily observable that there are four input signals and one output, all corresponding to the special compartments of the LSTM and the previous/next part of the complete network.
- **Dropout layer**: is a process which selects randomly neurons from the network and deactivates them during training to provide a more robust approach. For example, applying a 50% dropout in a specific hidden layer will cause a 50% probability of deactivation to all neurons of that layer as shown in Figure 2.11. The two main reasons for applying dropout after the LSTMs units are related to overfitting probability reduction, plus that it helps to build a more robust model as it creates a more stable relation between the neurons removing any correlation that might be created between the neuron in one layer [67].
- **Dense layer**: in the last layer of the model, there are connections between the nodes of that layer with all activation functions of the previous layer.
- **Activation layer**: this layer composes an essential part of the network as it contains functions that determine whether a neuron will generate an output or not. In simpler words, the activation functions are linear/non-linear mathematical functions that regulate the output of a neural network. There are several available while the most commonly used are the Sigmoid, the Rectified Linear Unit or ReLU. In the presented work, a linear activation function has been used for the mapping requirements of the project and it can be derived from the following formula.

$$\boldsymbol{A} = c \cdot \boldsymbol{X} \tag{3}$$

where X stands for the input, c is a constant that the function multiply with the input and A is the output. Using a linear function (Figure 7), the last part of the network or the dense layer will be fed with the output of the dropout layer and it will multiply the inputs with

some weights for each neuron resulting in an output signal proportional to the input. The advantage of the utilized activation function against another linear is the wider range of available values that can be generated. For instance, in case of applying the step function, the output will be restricted to Yes or No.

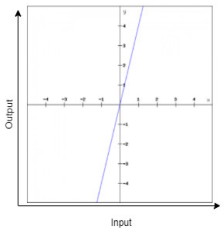


Figure 7. Behavior of linear activation function. In y-axis, the output values are presented while the input is depicted in horizontal axis values.

2.1.6 Loss Function

A loss function is a metric to evaluate the network performance and stands as a metric for the dissimilarity of the ground truth and the prediction of the model. The loss is computed when the weights and biases are set while in the presented thesis, the Mean Square Error (MSE) has been utilized to evaluate the accuracy of the model during training and test phases. This is a well-known loss function and can be calculated using the following equation:

$$\boldsymbol{MSE} = \frac{1}{N} \cdot \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \tag{4}$$

where N is the number of data points, Y_i is variables for the actual values or observed valued when $\hat{Y}_i$ holds for the predicted/estimated values. MSE is an estimator which measures the average of the squares of the errors or the average square of the difference between the predicted and the actual values of the surrogate and can take values ranging from $-\infty$ to $+\infty$ depending on the difference between the calculated units.

2.1.7 Optimizer

Adam has been chosen as optimizer for the prediction model of the surrogate signal and its main function is to improve the weights and biases in terms of quickest to solution adjustments. The aim is to enhance the network prediction resulting in the possible lowest value of the loss function. In 2015, Adam optimizer has been introduced in the Machine Learning community while it started quickly to overcome other proposed optimizers like stochastic gradient descent (SDG) due to its fast learning response combined with less computational load.

In Figure 8, a comparison between Adam and different optimizers is presented indicating the lower training load of Adam based on the literature researcher conducted in 2016. Moreover, the behavior of the optimizer depends on the learning rate which can be translated to a value that determines how big will be the changes in the network parameters from one step to the next on in the training phase. For example, a larger learning rate is connected with larger updates from epoch to epoch on each batch leading to higher impact changes at the model's performance [68, 69].

2.1.8 Model's Hyper-parameters

Tuning a neural network can be difficult due to higher or lower impact on performance. Accuracy values that have been generated on the training and external validation are used in order to get an insight on the behavior of the network hyper-parameters. The latter can be either the number of epochs or the number of layers while the ultimate goal is to achieve a high external validation accuracy along with a high training accuracy. Of course, the accuracy of the training, in most cases, overcome the testing phase corresponding values but when then the model is optimally tuned, the output will be the highest possible accuracy combined with the optimal generalization ability on unseen data.

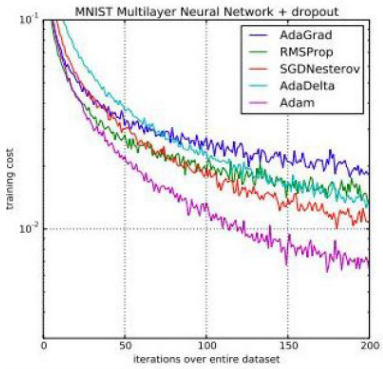


Figure 8. Comparison between the Adam and other optimization algorithms, training a multi-layer perceptron on the MNIST dataset using dropout optimization function [69]. As shown, in y-axis the training cost of each optimizer is presented along with the corresponding number of iterations that needs to make a complete scan of the dataset. Adam has the lowest workload cost during training compared to the rest.

But what is accuracy? Accuracy is a metric for the model performance and stands for the ratio of correctly classified instances divided by the total number of instances. Furthermore, it should be highlighted that the training accuracy is an indication about the learning capability of the network while the internal validation accuracy is used to find the optimal network parameters. In the test phase, the validation accuracy gives an insight of the model's performance on untrained data by classifying them correctly or not.

As it previously mentioned, there are hyper-parameters affecting the response of the neural network. The study in this field is quite broad but since it is not the main focus of the thesis, specific parameters will be analyzed due to space limitations.

- **Epochs**: is the time step when when all the batches of data are loaded and trained for at least once. In some

case, it is useful to randomize the data at every training epoch, but in the case of time-related prediction, this could be destructive for the learning process, thus is was chosen not to have randomization process.

- **Batch size**: is an additional factor that decides the number of fed data of the complete dataset that will be fed to the LSTM model at every epoch. The larger the batch size, the more features are available for the model, consequently the higher chance of generalization of the network. Secondly, a larger batch size offers a higher learning rate management resulting on faster learning response for the chosen optimizer. In the opposite side, difficulties in the learning response can be created applying a large batch size as the larger the batch is, the higher variations will be included in each batch, and the model should perform many changes to fit to all data.
- **LSTM units and neurons per unit**: stacking additional hidden layers or neurons per hidden layer, it increases the dimensionality. This fact can be essentially helpful to approximate higher complexity problems. In contrary, increasing the number of layers and neurons, result in a deeper network with higher chance of overfitting due to its higher capability on storing parameters, such as the weights and the biases which creates the question why to generate a deeper model. The answer related to the bias values. In case the model is highly biased, it is wise to increase the complexity and the number of parameters and can be observed when a model has relative low accuracy compared to some achievable baseline. On the other hand, this may lead to high variance but a solution would be to acquire more data and increasing the training set or regulating this with dropout or batch normalization operations.

2.2 Classifier

During the surrogate prediction phase, a classifier is needed to select the appropriate trained LSTM model based on the different breathing pattern presented mostly on the dataset. Thus except of the model training, a classifier training is an important requirement factor for a successful complete application. To reach the final step for the surrogate prediction which is the classification, it is mandatory to pass by some data processing steps before which include noise removal, signal segmentation and feature extraction methods. The complete framework can be called activity recognition system and a schematic representation is given in Figure 9. Further explanation about the detailed processes involved into the classification problem will be provided in Chapter 3 with the different evaluation metrics utilised for that part.

3. Methods and Materials

In Chapters 1 and 2, all the necessary information about the the general approach of the presented work has been presented while in this Chapter, the reader will dive into the methodology in more technical terms with a detail analysis of each step. As previously stated in Chapter 1.3, the corresponding work has two compartments running in parallel, the liver-surrogate data motion model creation and the surrogate prediction model. Thus, the framework till reaching the split part will be presented first followed by the different methods for each of the two compartments that have been applied. At the end of each part the evaluation methods will be presented to be clear and distinct which steps at every process were followed.

3.1 Overview

Firstly, the available types of surrogates are analyzed in Chapter 2.1 and the US sensor is chosen out for the independent imaging modality nature and the strong correlation with the liver actual motion. Secondly, the motion representation has two compartments, the internal motion which is represented by the upper border liver displacement in the SI-direction using MRI saggital images while the surrogate motion is represented by the magnitude of the received signal of the deepest to the liver measurement of the emitted US field. At this point, two different paths are followed, one for the liver-surrogate motion model and one for the surrogate prediction model. Regarding the former model, next, the relation between the surrogate and internal motion is depicted as the correspondence motion model that has been chosen to be linear based on the literature review, since the complexity has been kept low and the performance was acceptably enough. Finally, the last factor that needs to be regulated is the fitting methods. For the presented case, linear regression methods (simple linear regression, polynomial fitting, single layer perceptron) have been utilized based on the literature review that has been done where researchers assessed the consequences of different factors to the correlation between the surrogate signal and the internal motion [49, 50, 71]. As far as concerned the surrogate prediction model, firstly the type of neural network called LSTM and presented in Chapter 2 is utilised along with a k-NN classification algorithm [72]. Note that intermediate steps are taken place between the last two aforementioned steps which will be explained in the following sections.

3.2 Workflow

A schematic representation of the workflow is given in Figure 10. The workflow can be split into two phases, the training and the prediction steps. Focusing first on the training step, the process starts with the simultaneous acquisition of sagittal MRI liver images and the received emitted pulses from the US sensor as surrogate signal. Both signals are acquired for a specific period. Afterwards, pre-processing steps follow which are diffrent for each data type. For the MRI data, the liver upper border displacement in mm in SI-direction is segmented at each frame while for the US data, a Hilbert transform function is applied on the raw data followed by data selection for choosing only the deepest to the liver wave presenting the magnitude of the wave in mm over time.

Next, a process for dataset split and signal segmentation is taking place. Data split is performed for selecting the data

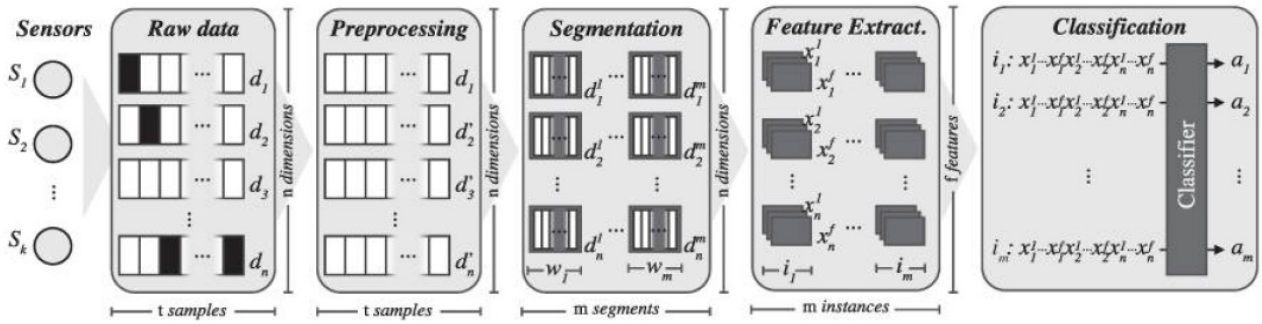


Figure 9. Steps involved in activity recognition system. Starting with the corresponding sensor which generate the raw data followed by some pre-processing steps. Signal segmentation is the needed to cut into smaller segments the complete dataset for the feature extraction process which follows afterwards. All the extracted features will be used at the end to predict the different N classes[70].

that will be used for training of the motion model and the surrogate prediction, data that will be used for the models' optimization and finally the dataset that will be used for the models' evaluation in test phase. The amount of data, the splitting process and further details are provided in the following section. After having excluding the test data, a signal segmentation process comes next where the dataset which will be used for the liver motion-surrogate data motion model are the complete experiments(called per experiment data) of the MRI when the data that will be used for the surrogate prediction model are segmented per activity (per activity data) to be fed later on the LSTM model and the classifier. At this point, two different pathways will be followed for the rest of the work since one corresponds to the motion model framework utilised on the estimation of the liver using the surrogate while the other is related to the surrogate prediction that can be used to estimate the liver motion in short future states.

Regarding the motion model, the per experiment data, are filtered for noise cancellation, temporal synchronized and aligned. Afterwards the data are fitted to the regression models and the learning parameters are created leading to the creation of the motion model. Using the trained motion model in the test phase, it is possible to obtain the liver motion based solely on the surrogate data.

As far as concerned the surrogate prediction model the US data per activity are utilised followed by filtering method to remove the unwanted noise. Next, the surrogate model has two leafs: one for the surrogate prediction and one for the classification of the surrogate data into classes of known breathing types. The left leaf is used to find out the learning parameters for the surrogate data and later on, in test phase to predict it when the right leaf with the classification step will be used to train a model for classifying segments of the surrogate and based on that, selects the most suitable LSTM trained model to make the surrogate prediction.

For the LSTM models training the internal processes and functions have been analyzed in Chapter 2 while for the classification training will be analyzed in the following section. Briefly, the US data is processed and after feature extraction and dimensionality reduction algorithms, the transformed data are fed into the classifier. At the end of the training framework, five LSTM models and one classifier trained on five different breathing types are available along with the motion model parameters for the liver motion and the surrogate signal.

Next, in the prediction phase acquiring only the surrogate data are acquired followed by Hilbert transformation processing and selection of the deepest to the liver measurement. The two leafs again exist as in the training process but the difference between them is related whether or not is desired to predict the surrogate signal and consequently make predictions on the estimation of the liver or just to estimate the liver based on the available surrogate data.

3.3 Experimental setup

In the previous sections so far, a complete overview of the conventional approaches has been provided to mitigate the problem of the respiratory liver motion. Additional to that, all the requirements for the different parts of the motion model and its parameters has been given while in this section, the data acquisition with the experiments protocol and the suggested fitting methods will be analyzed and validated on human subjects. A schematic representation of all the processes taken place in Figure 10 is presented but from a distinct point in Figure 11. In the given experimental setup overview, at the training part, note that the patient's data are acquired from the MRI bore applying a regular MRI coil around the participant thorax and the surrogate by sending pulses and receiving the corresponding echoes. Both signals are acquired simultaneously, but with different acquisition rates. At the prediction phase, the participant can leave the MRI-bore and depending only on the surrogate acquisition, the motion data are estimated.

3.4 Data acquisition and Dataset Split

The data collection has been performed using a limited inter-subject variability while the general characteristic of every

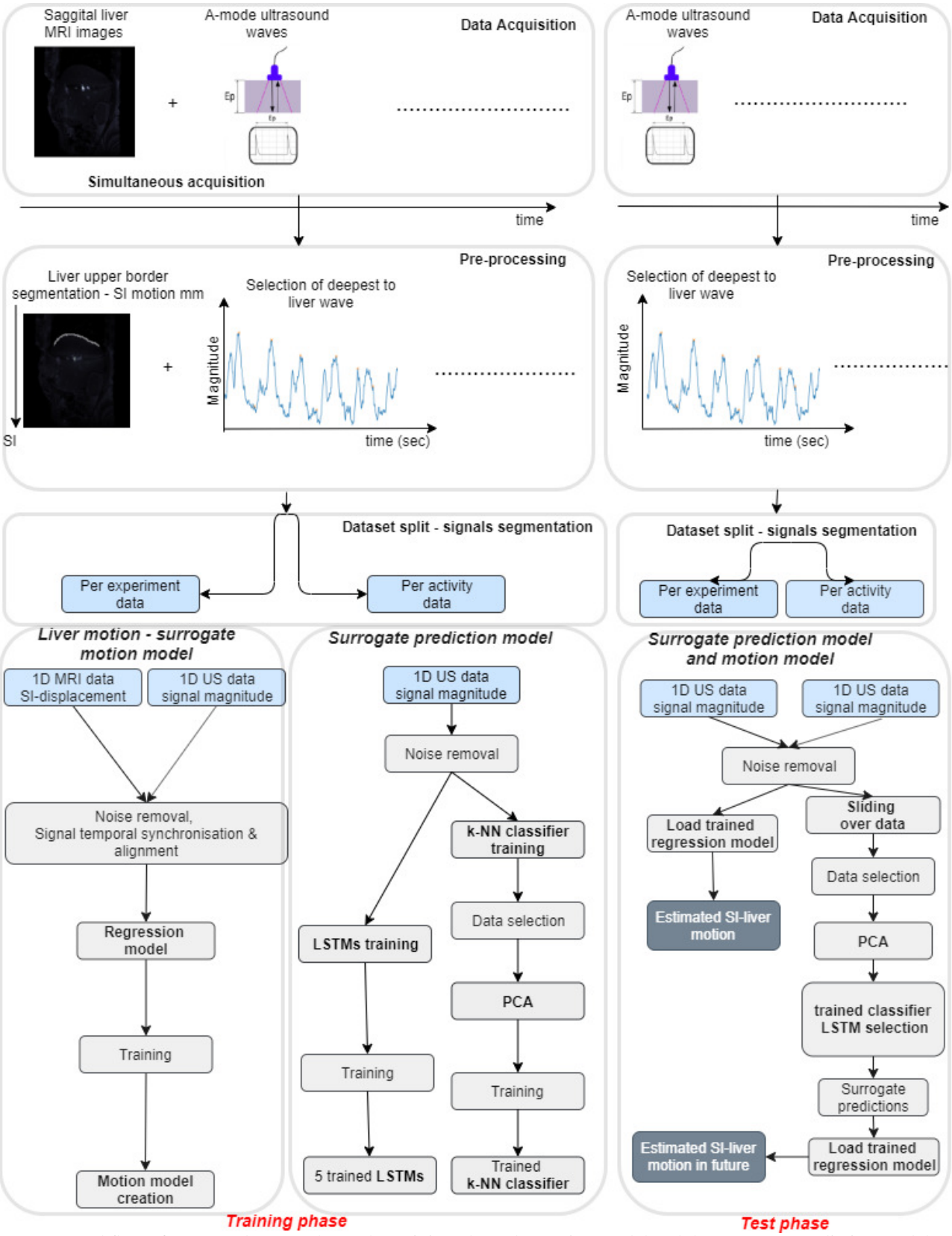


Figure 10. Workflow of suggested approach. At the training phase the motion model and the surrogate prediction model are taking place with different frameworks while in the test phase a common framework is followed using only the surrogate signal to estimate the liver actual motion and make prediction for a future state.

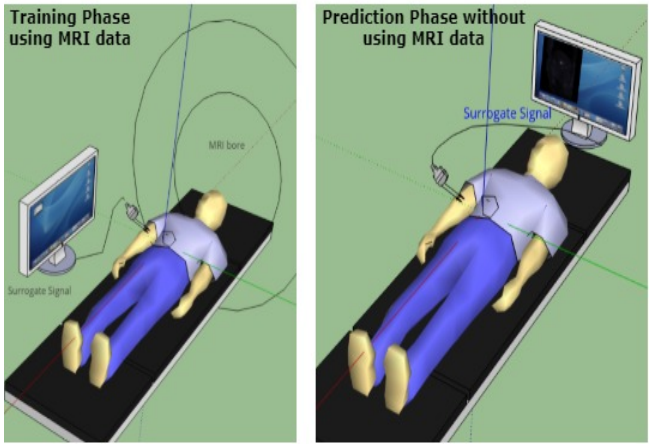


Figure 11. Overview of the experimental setup in training and prediction phase. While in the training phase, the model exploits MRI and surrogate data, in prediction phase only the surrogate signals are used to predict te motion estimate.

subject is given in table 2. The participants have been placed either in the MRI table or in a table of the RaM lab with the surrogate sensor attached to his/her skin close to the liver area and the experiments were handled by an instructor in order to reassure that the subjects will perform the desired breathing patterns listed below:

- 2 minutes of regular breathing
- 2 minutes of intermittent breath holding
- 30 seconds of regular breathing for recovery
- 2 minutes of short and shallow breathing
- 30 seconds of regular breathing for recovery
- 2 minutes of deep and heavy breathing
- 30 seconds of regular breathing for recovery
- 30 seconds minute of intermittent coughing

The chosen duration was either due to feasibility issues (couching, shallow, breath-holding for longer periods could be exhausted and feel discomfort) or due to requirements for the MRI acquisition (e.g at least 50 images of 10 breathing cycles including different breathing patterns). Note that the small recovery sections have been used to recover every subject back on the regular breathing state after the performed exercise.

Table 2. Participants Table including the characteristics of age, details, BMI and which examinations have been through.

ID	AGE	GENDER	BMI	MRI/US
A	33	M	27.4	Y/Y
B	26	F	20	Y/Y
C	26	M	22.1	Y/Y
D	33	M	24.4	Y/Y
E	25	F	21.1	Y/Y
F	26	M	26.9	N/Y
G	26	M	23.3	N/Y
H	24	M	21.7	N/Y
I	27	M	27.2	N/Y
J	27	M	19.7	N/Y

The final goal of this thesis is to predict the internal motion of the liver in a future state given a training data acquired only by the surrogate signal. At this point, it is important to have an indication about the performance of the model in new unseen data or the so called external validation (test) phase. Consequently, it is important to have an overview of the created algorithm using both training and testing error values. Note the difference between the training phase and the test phase datasets: the training set is used for learning the model's parameters while the test set is utilized to estimate the performance of the best model into new unseen data. Moreover, in the training part, there is also the internal validation set which is used in model selection (tuning, hyper-parameter choice etc). In case this step will be skipped, the model may be selected to perform well in a particular training set, the so-called overfitting and the performance of a model in a test set that has been used in model selection will be an optimistic approximation of the real-life performance. Since in this case, the available surrogate was only one, this results in uni-variate feature and no need for shrinkage method as suggested for multiple features. Based on the aforementioned steps, the acquired dataset needs to be splitted into three parts: training - data to fit the models, internal validation - data to find optimal parameters and external validation compartments - data to evaluate the model accuracy. As there no close formula for the ration among these dataset, 80% has been used for training phase (70% training/30% internal validation) and the rest 20% for the test phase. For the evaluation of the models, 10-fold cross-validation and leave-one out validation have been used [73].

3.5 Surrogate Prediciton - Classifier

3.5.1 Signal segmentation

For the surrogate classification, as in every classification problem, it is substantial to divide the data into smaller segments than the initial one in order to extract valuable information from every segment of the acquired data. This presented approach is based on a fixed window of 300 samples that slides over the raw US data and segment it into multiple samples using a 50% overlapping as shown in Figure 12. The window size has been chosen to be 300 samples or 6 secs in order to include at least two complete breathing cycles at each segment (inter- & intra-variability). This finding came up after literature research. The breathes of an average adult person varies from 12 till 18 per minutes if he/she belongs to the age group under 65 y.o. Moreover, according to literature, many researchers commonly use a 50% overlapping fixed window when working with activity recognition systems. On the other hand, variations on those two parameters may lead to faster detection with low computational cost but with a trade-off in accuracy since sometimes based on the available data, it is not possible to include a complete cycle at each segment. In the drawbacks, a relative small window size, can potentially lead to higher accuracy rates but on the other hand, it will be high computationally cost effective and it is not recommended or

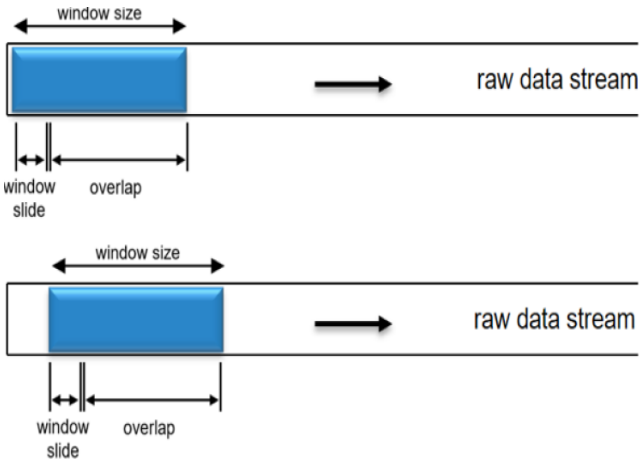


Figure 12. Signal segmentation based on a fixed size window with overlap [74].

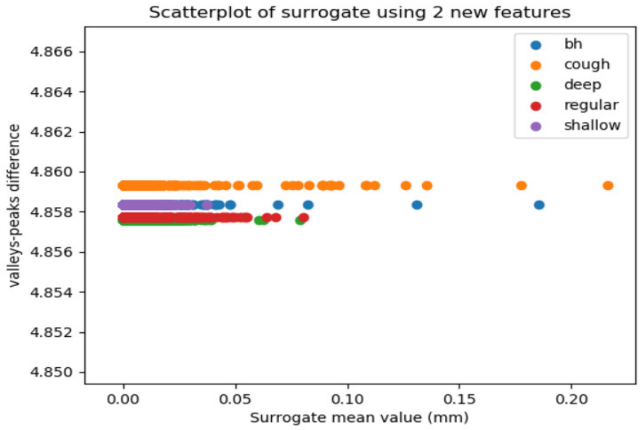


Figure 13. Scatterplot of mean values along with the difference between peaks and troughs features for the five classes.

by increasing a lot the window size, it may lead to detection of more complex activities increasing again the computational cost.

3.5.2 Feature Extraction

This step will play a major role on the model performance later on since feature extraction is the process of transforming large input data into a reduced set of weighted features. The goal of this step is to extract the most valuable information (features) of a window segment (output of the previous step) representing on the other hand adequately the data characteristics (generalizable features). Most commonly used types of feature extraction involve either time domain or frequency domain features. On the other hand, time domain features such as mean, median or variance are simple in calculations in comparison with frequency domain features including Fourier Transformation calculations increasing the computational complexity. Thus, it has been decided that the features to be tested should belong to time domain field and more specifically the chosen one are mean, median and difference between the peaks and troughs of the data.

After trial and error, it has been found that mean values as new feature were not as valuable as median, thus the final choice included only median values and the difference between peaks and troughs of the data at every different breathing pattern. The comparison for the two features can be seen in Figures 13, 14. The distinction between the classes is better using the median values but not enough using the 2D space, thus inserting the class feature as well and projecting the data into 3D space, Figure 15, the outcome is much more clear.

3.5.3 Dimensionality Reduction

Principal Component Analysis or PCA has been exploited for dimensionality reduction from 2D space (median values, peaks-troughs values) to one feature. PCA technique combines the inputs in a specific way to remove unwanted/least important information (input) while still retain the most valuables parts of all the variables. The dimensionality reduction

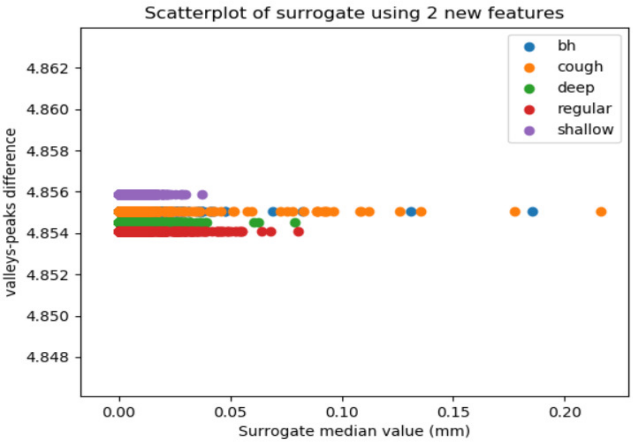


Figure 14. Scatterplot of median values along with the difference between peaks and troughs features for the five classes.

idea is based on simple linear projection. An amount of independent vectors is chosen and the data is projected on those. This can happen by projecting a vector (datapoint) on another vector (projection direction) using the inner product of the two vectors and the output is a scalar. When having available several datapoints and project them on many linearly independent vectors as the dimensionality of the original datapoint, it is feasible to reconstruct the datapoints by using the obtained scalars and the chosen projection directions. In order to find the vector that maximizes the variance of the projected data (feature that describes the most the data), eigenvector-eigenvalue decomposition of the covariance matrix of the original data should be performed. The outcome after performing PCA on the selecting two features can be shown via a scatterplot of the new feature (principal component) and the class feature shown in Figure 16.

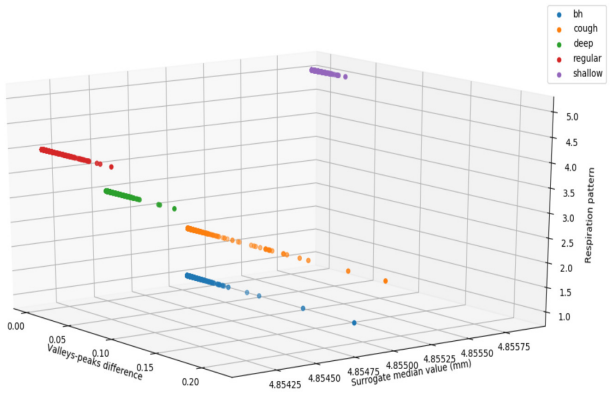


Figure 15. 3D scatterplot construction using the features of Figure 14, plus the label features of the class that belongs to.

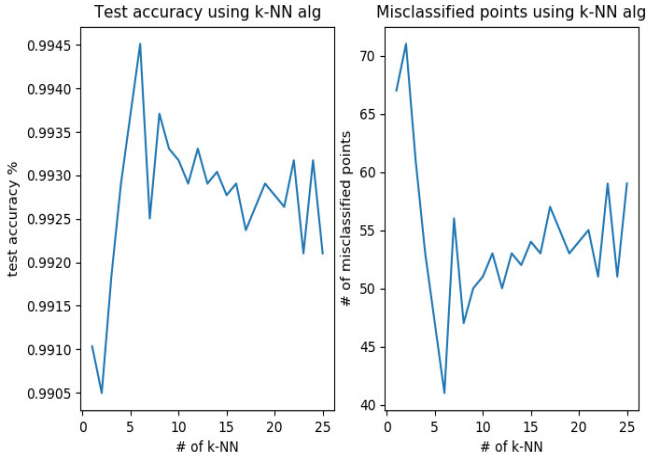


Figure 17. Accuracy and number of misclassified points for different k-nearest neighbors values evaluated on the test set.

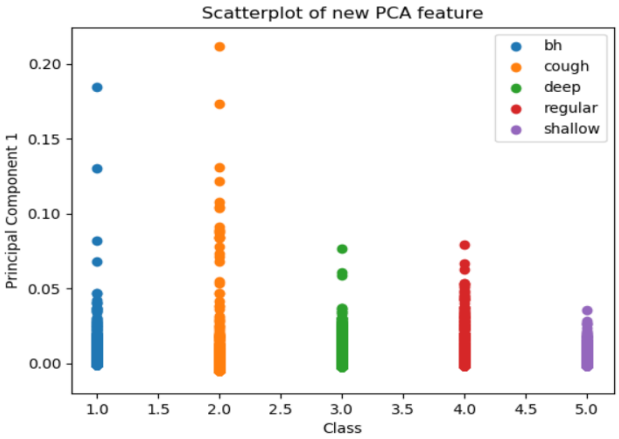


Figure 16. Scatterplot of the output of PCA, first principal component and the class feature. Using the new feature providing the ability to distinguish the five classes in 2D space.

3.5.4 Classification

After the data processing methods, namely: signal segmentation, feature extraction and dimensionality reduction, the data is divided into training, internal validation and external validation set or test set. Supervised machine learning models are utilised to learn parameters of the training set and detect the breathing activities on unseen data. Different methods have been tested, among them Random Forest Classifier (decision-tree based approach), Multilayer Percetron and k-NN algorithm (distance-based approach). The latter gave the best performance in terms of computational time, accuracy and misclassified samples, thus it has been chosen as the most appropriate one. k-NN or called k-nearest neighbors algorithm main assumption is that the dataset can be classified into different groups, based on their similarities and geometric properties. The algorithm's approach to measure the similarity of each datapoint is related to measuring the distance from a new instance to the instances that has been trained to fit. The new instance will be assigned to the class with the closest k neighboring instances.

As the data have been processed utilizing programming language Python 3.x / MATLAB and taking advantage of the package 'Scikit-learn', a classification model of the k-NN with the default hyperparameters has been created. However, this can not guarantee that the hyperparameters used as inputs to the classifier will generate the optimal outcome. Thus, To find the optimal parameter for the number of the closest k-neighbors, several models have been deployed and evaluated in terms of accuracy and misclassified points as well, as it can be seen in Figure 17. The highest accuracy and the least amount of misclassified points have been given from model using 6 nearest neighbors, thus this was the choice of the final model hyper-parameter.

The final model hyperparameters are defined as follow: model = KNeighborsClassifier($n_{neighbors} = 6$, weights='uniform', algorithm='auto', $leaf_{size} = 30$, p=2, metric='minkowski')

3.5.5 Evaluation Metrics

At this step, for measuring the performance of the classifier, different metrics have been exploited. As the most commonly used in activity recognition system, classification accuracy, precision and recall, F-score as well the confusion matrix have been used in this section. Moreover the ROC curves have been calculated along with the AUC for each of the five classes (breath-holding, coughing, deep breathing, regular breathing, shallow breathing).

Starting with the explanation of the used metrics and more specifically with the confusion matrix which is a summary of the correct classified and misclassified predictions for each class compared to the actual labels. As shown in Figure 18, in the vertical axis, the actual labels of the values for each class are lying down while in the horizontal axis, the predicted values for each class are assigned. The elements along the main diagonal represent the correct classifications when the elements outside of it stand for the misclassified points. Furthermore, looking at the confusion matrix, information

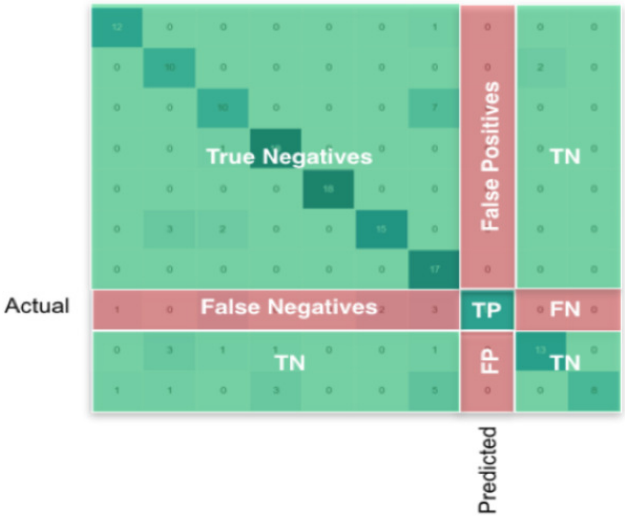


Figure 18. Example of confusion matrix [75].

such as true positive (TP) values (correct classification for positive instances), true negative (TN) values (correct classification for negative instances), as well as false positive (FP) and false negative (FN) which represent the incorrect classifications of negative examples to positive class and vice versa, respectively, are visible.

Moving to classification accuracy metric is the most straightforward way to measure the performance and is defined as the number of correctly classified data points over the total number of datapoints (complete dataset).

$$\textbf{\textit{Accuracy}} = \frac{\#correctclassifications}{\#datapoints} \tag{5}$$

or

$$\textbf{\textit{Accuracy}} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Regarding the precision and recall values can also be obtained from the confusion matrix and can be defined as:

$$\textbf{\textit{Precision}} = \frac{TP}{TP+FP} \tag{7}$$

$$\textbf{\textit{Recall}} = \frac{TP}{TP+FN} \tag{8}$$

Similarly, the F1-score is also added as a metric for the accuracy and is defined as:

$$\textbf{\textit{F1-score}} = \frac{2\cdot(precision*recall)}{(precision+recall)} \tag{9}$$

and is a metric for precision and recall of the classification system. A value of 1 for that score can be translated to perfect precision and recall while in the worst case scenario of a system that misclassifies all the samples, a value of zero is the possible value. In addition, the ROC curve which stands for Receiver Operating Characteristic, has been used to have a more informative aspect of the classifier behaviour. The ROC curve shows the relationship between the number of true positive and false positive classifications and is an intrinsic image of the classification performance since the accuracy metric does not provide information such as which misclassifications are worst than others. In contrary, this is visible from the ROC curve. For example, consider a problem where a machine learning classifier needs to perform blood analysis determining if the subject belong to cancer-diseased patients or non. The impact of misclassified samples as false positive is much less compared to an incorrect prediction of false negative which means that a diseased patient cannot be detected correctly. Finally, the area under the curve or AUC in the ROC curve is also a valuable information to obtain as the larger this area is, the better the classifier performs under different threshold values for the true positive and false positive samples.

As stated previously, in Chapter 1.4, in order to find the correlations between the two imaging modalities, the choice of the surrogate, the internal and the surrogate motion representation, as well as the motion correspondence model along with the fitting method need to be taken into account.

3.6 Processing methods

Starting with the acquired surrogate signal, there were several available measurements from the different penetration depth (0.7 to 7cm) of the US echo, but after careful research, it has been found that the most outer or the closest to the skin measurements involve noise or disturbance of the detection of the desired signal to measure [76]. Thus, the closest to the liver penetration depth has been used and it has been preprocessed, using the Hilbert transformation that maps the x(t) to $x(t)+i\cdot\tilde{x}(t)$, resulting in an output U which has been produced by the following formula:

$$\textbf{\textit{U}} = log(abs(Hilbert(U_{raw}))) \tag{10}$$

The goal was to enhance the deepest to the liver measurements which were the most valuable information as they will not get involved with other factors such as noise.

2D MRI DICOM series was the raw version of the liver motion representation which consequently analyzed using MicroDICOM viewer to extract the frame images per time step based on the predefined frame acquisition of the machine. The generated images were processed using Python 3.x including image enhancement methods, such as contrast and brightness enhancement to increase the pixel ins tensity and make it easier for the further steps. Afterwards, masks along with thresholding have been applied to reduce the region of interest for segmenting the contour. The corresponding masks have oval shape and include in all images the outer area of liver. This is a tricky part because the liver borders interfere with other organs or walls (lung-diaphragm) and edge detection algorithm cannot find real distinct edges in the image.

Moreover, morphological operations have been applied such as erosion and dilation using different kernels depending on the liver representation and its scale within the image to obtain the outer contour of the liver. Additionally, a mask have been applied to select only the upper border of this contour and track the highest pixel position in y-axis standing for the internal motion representation. A schematic representation of the corresponding operations can be found in Figure 19

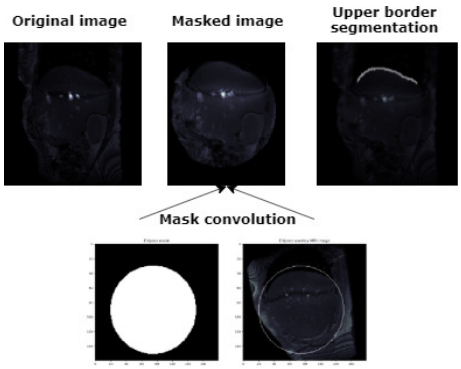


Figure 19. Signal processing operations to obtain the liver upper border segmentation at each frame.

Now, both signals are available with different lengths and magnitude scales. Next step includes signal smoothing and noise cancellation for both the surrogate and the internal motion data as in some cases the noise distortion was high resulting in motion detection by the fitting methods. Two methods have been tested for that reason, Low Pass Filtering (LPF) and Savitsky-Golay (SG) filtering which has a similar implementation as the LPF (see Figure 20).

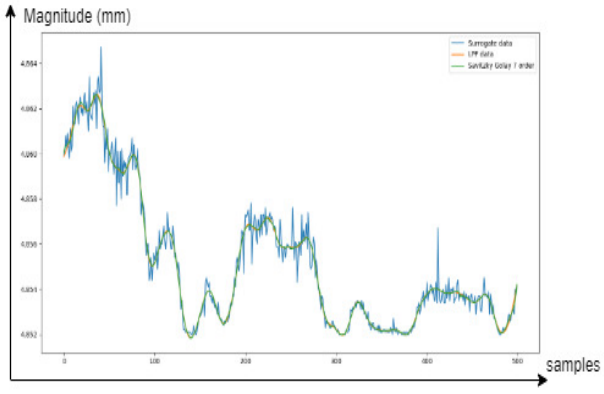


Figure 20. Signal smoothing on surrogate data using Savitsky-Golay and low pass filtering. Comparison of the results using ground truth data.

The SG filter belongs to the finite impulse response (FIR) filter family using least squares idea whose impulse response has finite duration. The least squares tries to find the optimal (minima) distance between the actual and fitted data by

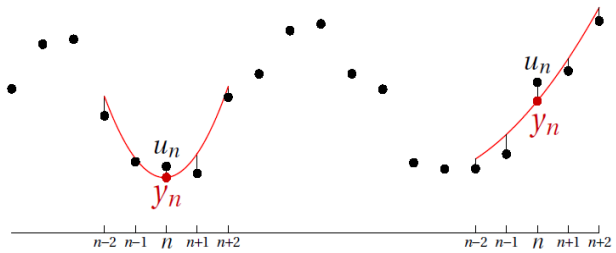


Figure 21. Savitsky-Golay filter function [77].

solving the equation:

$$c_* = \left|\left| argmin_c \begin{bmatrix} e_2 \\ e_1 \\ \vdots \\ e_{-2} \end{bmatrix} = \begin{bmatrix} u_{n+2} \\ u_{n+1} \\ \vdots \\ u_{n-2} \end{bmatrix} - W \cdot \begin{bmatrix} c_{1n} \\ c_{2n} \end{bmatrix} \right|\right| \quad (11)$$

The vector of ε stands for the errors, while the vector U represented the actual data points and the c_{1n} and c_{2n} are the linear regression coefficients and the vector W is the weights. A schematic representation can be found in Figure 21 where it is easily visible the approximation of SG filter to the data. U_n is the input or actual data and Y_n presented with red lines is the fitting method. The idea is to approximate around each signal instance U_n locally by a polynomial. Note that the local time frame is normally symmetric

3.6.1 Temporal Synchronization and Alignment

Subsequently, the signals have been resampled using FFT space and linear interpolation in order to obtain similar sample rate. The method performs upsampling to a real-valued signal x of length N by transforming to the frequency domain and adding N/2 zeros at the end and then transforming it back to the time-domain. All the functions for the temporal synchronization have been utilized by the Scientific Kit-learn package of Python for signal processing. The upsampling factor has been determined each time taking into account the available number of surrogate data divided by the number of the available MRI data using an adaptive method. Subsequently after resampling, the two sequences gained the same length by cropping the larger one using two different methods and applying temporal synchronization through temporal time stamps. Before conducting the latter step, it was important to normalize the two signals and to scale them obtaining similar value for their magnitudes. The internal motion has been converted firstly from pixel-based to mm-based scaling using the information for pixel size generated by the MRI machine and secondly at the surrogate by summing the mean of the internal motion and multiplying with the standard deviation of the same signal, the goal was achieved. The final step for both signal was now temporal synchronization and since they have the same starting point, two methods have been used for that reason (see Figure 22, 23).

The first operation works as following:

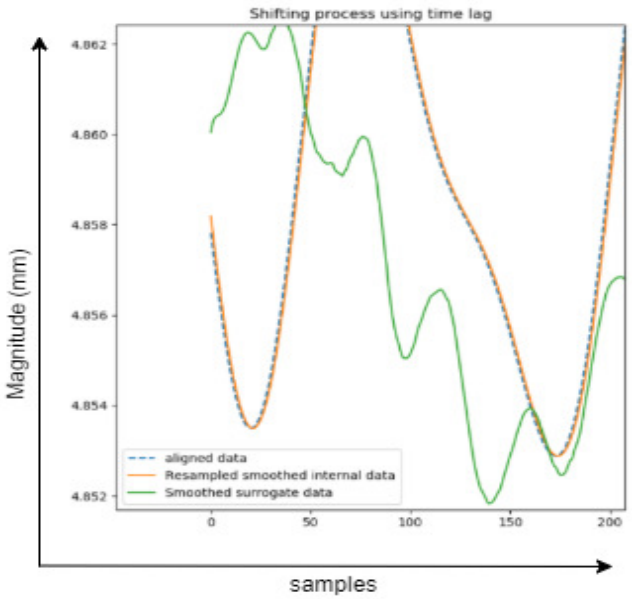


Figure 22. Signal alignment using time lag shift. Green data are the de-noised surrogate data, orange colored data stand for the internal the resampled liver data and the dashed blue line corresponds to the aligned data after time-lag shift.

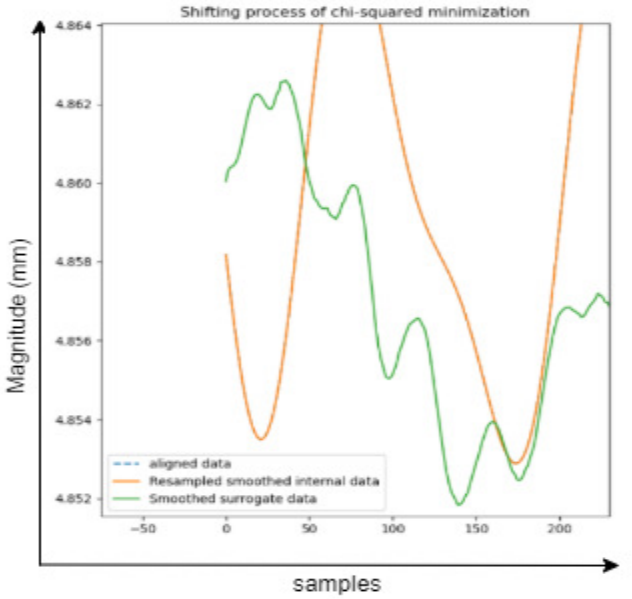


Figure 23. Signal alignment using χ^2 minimization. Green data are the de-noised surrogate data, orange colored data stand for the internal the resampled liver data and the dashed blue line corresponds to the aligned data after chi-squared minimization process.

- calculate length difference in sequences (N)
- crop (N) sample at end point of the larger signal
- find time lag (k) through cross-correlation analysis
- shift the desired signal by k-times (offset minimization)

while the second process seemed to be better

- calculate length difference in sequences (N)
- crop (N) sample at end point of the larger signal
- find time lag (k) through cross-correlation analysis
- minimize the offset between the two signals using χ^2 value metric.

Since both methods worked sufficiently fine, 2nd method has been chosen to work with due to faster computational time and because of keep fixed the initial point for both signals.

3.7 Linear Fitting Methods

Since the input is a scalar variable, it means that there are two-dimensional sample points corresponding to the surrogate data and the internal motion which stands for the dependent variable.

Conventionally, the correlation between the internal motion and the surrogate signal can be performed using linear fitting algorithms such as ordinary multivariate linear regression, lasso/ridge regression regularization analysis or quadratic polynomial regression but these algorithms have some drawbacks. For instance, the low complexity models are more biased while the high complexity model cannot be utilised in real-time applications. Thus, a fully automatic method using Machine Learning (ML) algorithms would be significantly beneficial in the corresponding problem. Since, the prediction of the surrogate data is a sequence pattern problem, this can be translated to sequence to sequence prediction using a supervised learning approach which looks for long term dependencies.

3.7.1 Simple Linear Regression

Conventionally, the algorithms creates a linear line which tries to predict optimally the dependent variable as a linear function of the independent one. The method functionality is based on the ordinary least squares (OLS) behavior where the accuracy of the predicted values is measures by calculating the squared residual error. The latter is the vertical distance between the point of the given dataset and the fitted line. The goal is to minimize that distance and the slope of the calculated fitted line is determined by the correlation between the dependent and independent variables, and the corresponding standard deviations. The fitted line can be described in:

$$\boldsymbol{y} = f(x_i) = \alpha + \beta \cdot x \tag{12}$$

where x are the inputs, f is the output, β is the corresponding slope between input and output and a is the point on y-axis where x=0. Output is depicted as the internal motion data, α as a bias, β represents the parameters of the motion model and x is the vector of the model features of the surrogate

3.7.2 Polynomial Regression

The polynomial regression can be imagined as a more complicated version of the linear regression. The main difference with the linear regression has to do with the transformation of the original features to higher order terms. In most cases, the polynomials are quadratic fit, means that the features will be

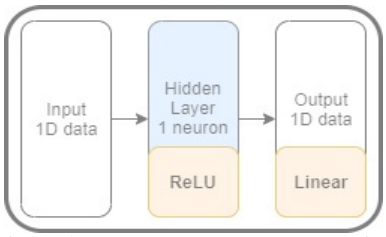


Figure 24. Single layer perceptron representation. The input is processed through a layer based on the ReLU activation function and its output is used as input to the last layer where a linear function is applyed

transformed to 2nd order terms remaining as a linear model as the parameters are associated with the features x and x^2 are still linear. For example, the linear regression formula will become

$$y = f(x_i) = \alpha_N x^N + \alpha_{N-1} x^N - 1 + ... + \alpha_1 x + \alpha_0 \quad (13)$$

3.7.3 Single layer perceptron

For this fitting method, again a regressor was utilized but in that case a simple neural network has been created as depicted in Figure 24. The network architecture is simple, the input is 1D data of the surrogate and in the hidden layer there is only one neuron where batches of data of input are fed and multiplied with the Rectified Linear Unit (ReLU) followed by the output with a linear function. In the presented case, the ReLU is a linear function thus the output will be a linear representation of the input. The choice of this activation function has to do with the promising results as it has been used for the first time in the AlexNet architecture and it turned out that can accelerate the gradient convergence 6 times faster compare to the Sigmoid function while the second advantage is the limited output from 0 till the maximum value of the input x and the low computational cost [64, 78]. The formula of ReLU can be analyzed as:

$$\boldsymbol{f(x)} = max(0, x) \quad (14)$$

3.8 Performance Metrics

Two evaluation metrics have been used for the test phase on unseen data and those are composed of Root Mean Square Error or RMSE and Mean Square Error (MSE). Since the latter has been analyzed in Chapter 2.2.5, its functionality will not be emphasized at this section. Regarding the RMSE is the standard deviation of the residuals and it is commonly applied as a metric in regression analysis to verify the outcome. The residuals can be seen as the red points distant from the fitted line, in Figure 25.

The residuals is a measurement for the distance between the regression line and the actual data points while RMSE is a valuable metric that shows the variance of these residuals or in simpler words it gives an indication about the variance of the actual data around the fitted line [79]. The black fitted

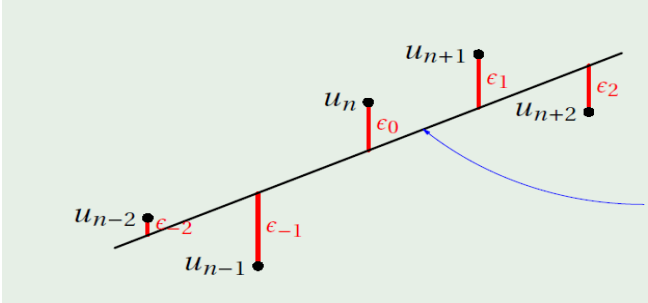


Figure 25. Residuals and linear regression line [77]. the black line represents the actual data while the dots stand for the motion model data. The red lines connecting the dots with the actual data are the errors or the so called residuals used to calculate RMSE.

line represents the regression function or $x_{*n}(t) := c_{1n} + c_{2n} \cdot t$.

Regarding the surrogate prediction, two metrics have been utilized the cosine similarity which in an indication about the signals' divergence (actual & ground truth). Secondly, the Euclidean distance has been exploited for the calculation of the difference in terms of amplitude. The cosine similarity is a metric of similarity between two vectors and measures the cosine angle between them and is unitless. It is clearly a metric indication for the magnitude difference and two vectors with the same orientation have a cosine similarity of 0 while two vectors with an angle of 90°(perpendicular) among them have a similarity of 0 indicating 50% de-correlation. Note that this value is independent of the their magnitude and the formula that describes this metric can be found below:

$$\boldsymbol{similarity} = cos(\theta) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{N}(A_i \cdot B_i)}{\sqrt{\sum_{i=1}^{N}(A_i^2) \cdot \sum_{i=1}^{N}(B_i^2)}} \quad (15)$$

The second metric that has been exploited to measure the similarity between the ground truth and prediction of the surrogate is called Euclidean Distance and is the distance of a straight line that connect two points in the Euclidean space. Using this metric, the distance between a fixed vector (ground truth) and an additional one is measured point-by-point and the sum of the length of all points stands for the Euclidean distance between the two vectors [80].

4. Experimental Results and Discussion

As stated previously in Chapter 3, in order to give an answer to the research question, the problem has been splitted into two parts: the first one has to do with the experimental part of the surrogate prediction while the second one with the correlation between the internal motion and surrogate signal. In this chapter, the outcome for the evaluation of the proposed respiratory motion estimation is presented and more specifically the evaluation of the surrogate prediction model comes first.

4.1 Surrogate Prediction

The predictive performance of the model has been tested using most of the available data to find out the optimal input parameters such as the choice of the appropriate depth. This is highly important as those measurements should have the highest possible correlation with the liver motion that can be translated to the most valuable information. Additionally, some received echoes contain different levels of noise or in other cases motion of adjacent organs and tissues. To explore the optimal depth values, the processed (traces) measurements of the the closest to the skin waves and the deepest to the liver penetration depth waves have been analyzed as the two extreme ends of the data after were considered more valuable to work with [26]. Starting with a visualization part of the processed data generated by the closest to the skin, from now on called depth-1 and the deepest to the liver penetration depth, from now on called depth-2668, is presented in Figure 26. In the relative Figure, on the first row a small part of breath-holding respiration pattern is exploited to show the artefacts and other motions presented while in the second row is the same breathing pattern using exactly the same time interval. In the vertical axis, the magnitude of the received pulse echoes is denoted and in the horizontal axis, the number of samples or the relative time stamp has been chosen to be shown. Note the differences in magnitude between the two depths with a continuous positive shift difference for all time steps in depth-1 compared to depth-2668. Moreover, in most cases the magnitude changes are diametrically opposed indicating on one graph the inhalation and on the other, the exhalation phase [10, 26].

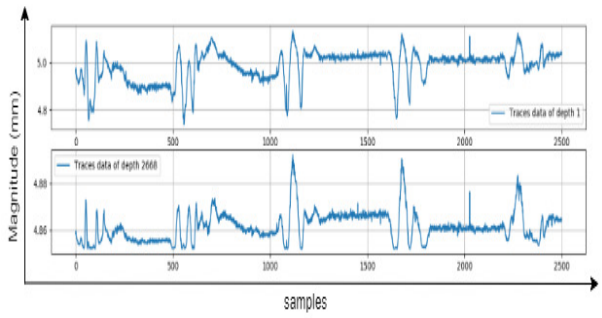


Figure 26. Different depth penetration measurements. First row, processed US wave signal of the closest to the skin layer while below it the US wave from the deepest to the liver measurement is presented.

After some data analysis, a 2nd step of verification for the aforementioned statement has been examined to ensure the findings that the depth-2668 was more robust and proving the assumption about containing the actual liver motion information. Assuming that there might be a correlation between the patterns of the two depths, the goal was to prove that a trained model on depth-2668 values and tested on depth-1 dataset could work better because depth-2668 contained the actual liver motion values and thus the model could handle better the small variations and the presented features. As a result, a test has been performed using different depth for training and prediction to validate the assumption. For limitation space reasons, the outcome after finding the average Euclidean distance and cosine similarity of all breathing pattern is presented in Table 3. As a result, this is an overall score including five respiration types.

Table 3. Different depth models exploring the correlation and differences.

Train	Test	Cos. Similarity	Euclidean Dist.
depth		$(\mu \pm \sigma)$	$(\mu \pm \sigma)$(mm)
2668	1	0.95 $\pm$0.22	0.04 $\pm$0.02
1	2668	0.47 $\pm$0.34	0.17 $\pm$0.07

Looking on the result of the trained model on depth-2668, both metrics have a better behavior on the dataset of depth-1 compared to the reversed experiment. Note that a perfect value for Cosine similarity is 1 and for the Euclidean distance is 0. This outcome comes in accordance with the findings of a data analysis experimental on subject A where using the values of the complete US experiment, it has been found that the magnitude measurements of depth-1 were ranged between [-1.72 , 5.64] with $\mu \pm \sigma$: 4.84 ± 0.20 while the corresponding values for depth-2668 have a range of [4.85 , 5.20] and $\mu \pm \sigma$: 4.87 ± 0.05. Checking the sigma values, the higher variations are obvious and more specifically the variations of the closest to the skin values are 4 times higher than those of the deepest to the liver measurement. The impact of those parameters is substantially crucial for the correct model prediction and consequently, depth-2668 measurements have been only included in the usable data. The models that have been developed for this step have been trained/tested on the same pattern while in the upcoming paragraphs, some models have trained and tested on different breathing patterns to find out potential correlations.

Now, as the problem of the usable data has been solved, it is time to step into the first main problem which is the prediction of the surrogate. Different models have been built and trained while all of them are sharing some common parameters. Namely, training has been done for 1 epoch using a batch size of 32, 3 layers of LSTM with 100 hidden neurons per layer, Adam optimizer with the default learning rate and α and β parameters while the length of time lags of the input and to be predicted signals were equal to 300 samples. Before starting with the results of the prediction models for the respiration patterns, take a look on the schematic representation of the surrogate signal presented in Figure 27. This depicts a small segment of 2500 samples of the complete experiment of subject A presenting the regular breathing pattern to have a better understanding on how the surrogate ground truth looks like.

Now, the reader should have a clear aspect of the visualization part of a surrogate and can have a further look on the corresponding visual results in Figure 28 where the five main

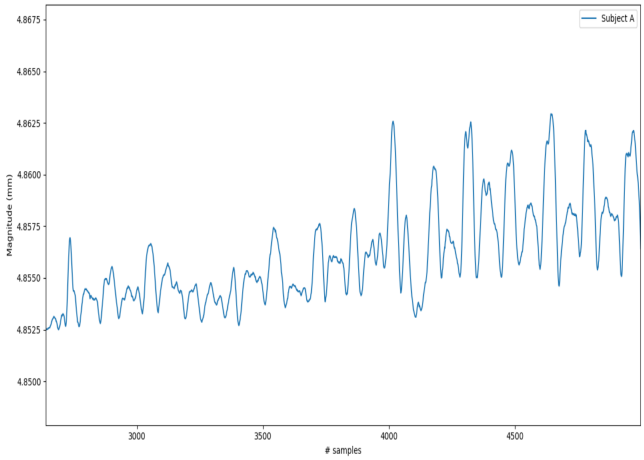


Figure 27. Surrogate signal magnitude of the deepest to the liver measurement acquired by subject A.

Table 4. Summary of prediction models using different respiration data.

Pattern		Euclidean dist.	Cos. similarity
train	test	$(\mu \pm \sigma)$(mm)	$(\mu \pm \sigma)$(radians)
regular		0.70 ±1.38	0.92 ±0.04
regular	others	0.01 ±0.001	0.87 ±0.01
deep		0.41 ±0.45	0.82 ±0.05
deep	others	0.01 ±0.001	0.77 ±0.10
cough		0.55 ±0.52	0.64 ±0.34
cough	others	0.01 ±0.001	0.87 ±0.01
breath-hold		0.68 ±1.28	0.68 ±0.35
breath-hold	others	0.01 ±0.001	0.43 ±0.18
shallow		0.68 ±1.42	0.77 ±0.31
shallow	others	0.01 ±0.001	0.57 ±0.11

respiration patterns are depicted. In all figures, the vertical axis represents the magnitude of the signals and the x-axis stands for the number of samples or the relative time step. The ground truth in each graph is denoted as blue while the corresponding prediction is set to orange and plotted on top of the ground truth to be easier to see the similarities and differences at each time step.

The best fit was given on the deep and coughing breathing types regarding the lowest Euclidean distance which have a value of 0.41 ± 0.95 mm and 0.55 ± 0.52 mm while on the other hand the highest values for Cosine similarity indicating a good fit for the regular and deep breathing patterns. Next breath-holding and shallow breathing share a similar value for the Euclidean distance with an increased σ value which can be translated extreme values due to the type of breathing resulting to high variations in the data. As far as concerned the σ, overall in the Cosine similarity values have a tendency not to observe high values resulting in a correctly aligned prediction in most cases. In contrary, an important aspect of those findings is that the prediction model in all cases can predict correctly the signal, but there are inconsistencies and misalignment on the peaks and valleys of the signal, coming in accordance with the previous studies where in that cases misalignment has been found between the actual and the estimated liver motion [10, 13, 11]. However, the discontinuities have a lower presence rate compared to previous findings due to the adaptivity of the model.

Next, in order to explore the correlation and the revealed features at each respiration pattern, five additional models have been trained in one specific respiration motion and tested on the rest breathing patterns. The results are summarized in Table 4 including the findings of the prediction using the same breathing pattern for train/test as well as the results for the models trained and tested in different respiration patterns. To begin with the 2nd group of models trained on breath-holding and shallow breathing data have achieved the two lowest Cosine similarity values. It is highly importance that the model of regular and coughing have performed remarkably prediction of the rest of the breathing patterns revealing that those two patterns' features are also contained in the rest of the patterns. For instance, those features can be a combination of a specific amplitude using a similar frequency rate. Secondly, some features as inhalation peaks may be repeated after a specific sequence in all signals. However, there was a difference on the amount of the available trained data for some respiration segments due to the acquisition method that has followed and also due to physical limitations mentioned in Chapter 3.1 and 3.2.

Finally, it should be highlighted that leave one out cross-validation (LOOCV) method has been utilized for the surrogate prediction findings. At the beginning, 10-fold cross-validation has been exploited but after research and discussion with signal processing experts from University of Twente, it has been chosen to replace that operation with the LOOCV method. The essential reason for that decision is related to reduction of overfitting on a specific dataset using neural networks. It is a common issue that observed when training a neural network, specifically using small number of training data as in the presented case. The issue arises because similar pattern of features are included into the train and test set coming from the same subject. As a result, in that case the algorithm will have a more "optimistic" performance on the test set than in real-life. Thus, having available 10 subjects data, it was considered wise to exclude 2 patients each time for the test phase and train the rest available data. No randomization process has been used on this validation step resulting to user dependent selection of the appropriate sets for the models construction.

Before moving to the 2nd main task which is the correlation of the two signals in the next section, a remark should be done. After careful data analysis on the surrogate signal and based on the visualization part in Figure 29 focusing on subject B, it has been found that the values did not come in accordance with the rest of the subjects. In the aforementioned graph, all the surrogates have been plotted together to obtain a more intrinsic aspect of the data distribution and as it can

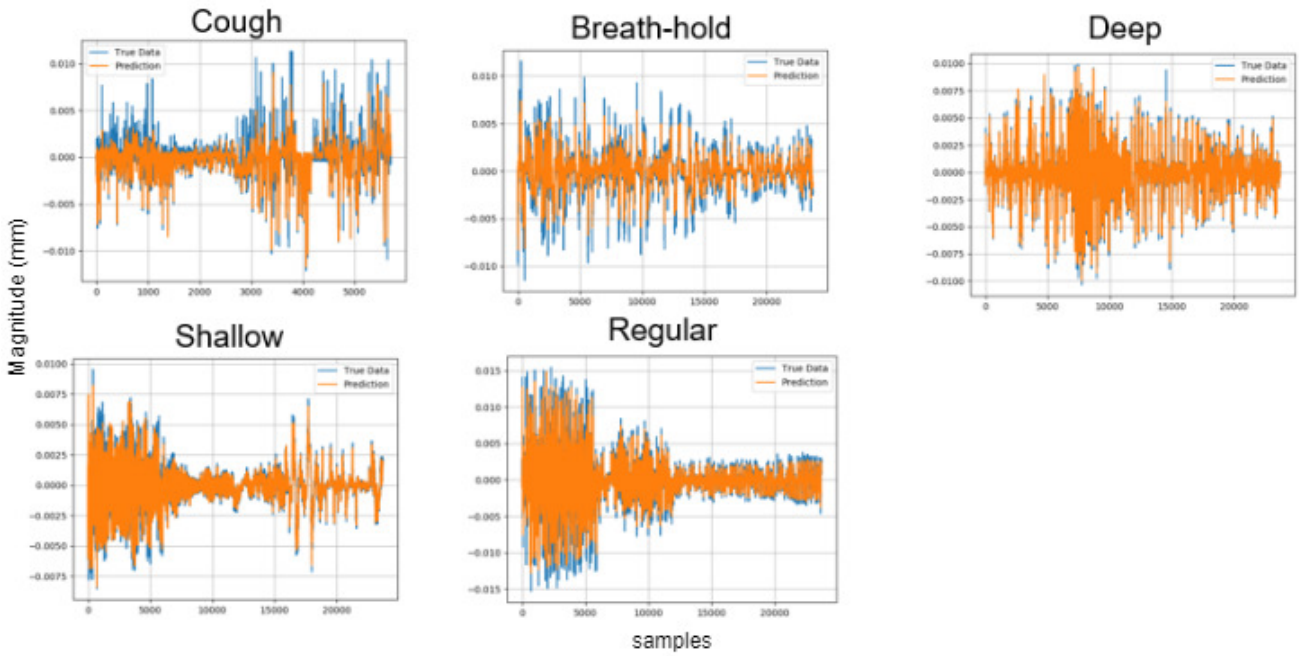


Figure 28. Visual outcome of the five different breathing patterns. Comparison of the ground truth and the prediction of the surrogate signal segmented per respiration type regarding the magnitude over time.

be seen when most subjects follow a range of values where their signal can lie down, data of subject B, indicated with dashed orange line, has an unusual pattern with high peaks and valleys diverging from the general data tendency. As a result, it was considered wise to exclude subject B from the rest of the experiments. An assumption to give an answer to that issue, it could be that subject B due to the low BMI according to Table 2, the liver motion has not been measured. On the other hand, the pulse due to high penetration ability might ”hit” a bone rather than the desired organ and this can result to such high values. A further assumption has been made for the object where the US transducer send the pulses and that could be a large bone such as the spinal cord. Another assumption could be a wrong positioning of the probe in the conducted experiment leading to measurement errors.

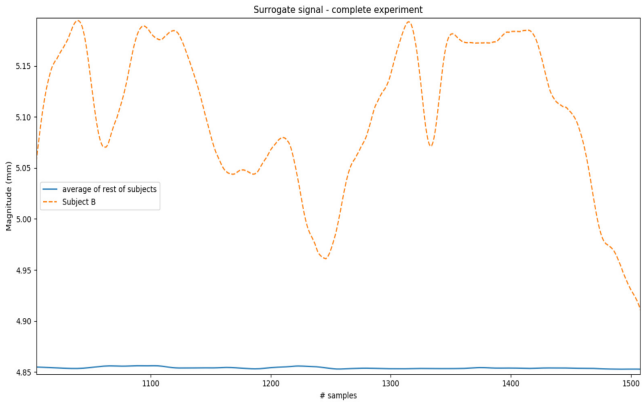


Figure 29. Unusual pattern for the magnitude of the surrogate signal of subject B compared to the rest subject data.

4.2 Classification Performance

Starting with the results of classification performance and the Table 5. As it is visible the ability of the classifier to distinguish the five class is 99% achieving classification of 7429 out of 7470 samples correctly. Specifically for the different classes: breath-holding, regular and shallow breathing samples can be classified correctly, as indicated from the precision, recall and F1-score values equal to one. Moreover, the deep breathing has slightly worst results than the aforementioned and this is mainly because of the samples of this class share some properties with the class of regular and shallow breathing at some segments. Thus, an assumption could be that some samples have been classified incorrectly due to that reason while this can be solved using additional data, adjusting the number of the searching nearest neighbors or using a smaller window segment. In addition, due to its lower number of available data coughing achieved the lowest score in classification performance for all different metrics but the findings are nothing else than positive since the precision and recall have a value of 0.98 and 0.97 which indicate a high classification performance as well. The overall accuracy of the activity recognition system can be depicted from the Table 5 where the area under the curve diverges slightly from the value of one for each class. Looking at this table, the interpretation is that the True positive rates are almost 1 for every class while the misclassified points or FPR is 0.

Similarly, as shown in Figure 30, four out of five class ROC's curves follow the same scheme while for the coughing

Table 5. Summary Table for the proposed activity recognition system including the precision, recall, F1-score, area under curve (AUC) and the number of instances per respiration type.

Class	Precision	Recall	F1-score	AUC	# instances
breath-hold	1 .00	1.00	1.00	0.998	2230
cough	0.98	0.97	0.98	0.986	441
deep	0.99	0.99	0.99	0.996	1465
regular	1.00	1.00	1.00	0.998	1917
shallow	1.00	1.00	1.00	0.998	1388
ACCURACY			0.99		7470

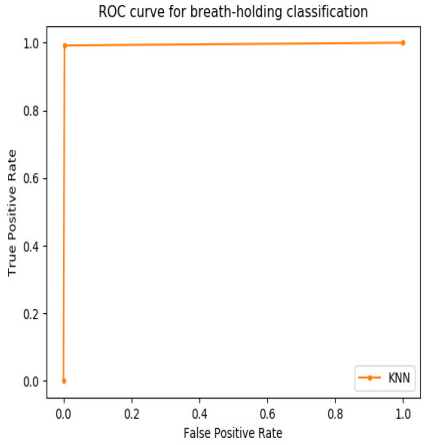


Figure 30. ROC curve schematic representation for breath holding.

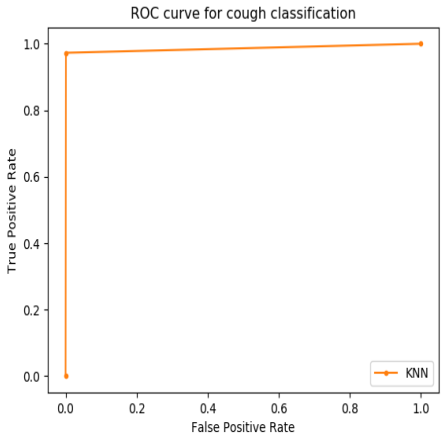


Figure 31. ROC curve schematic representation for coughing.

pattern with 4 times less available data, the ROC curve has a small divergence from the overall performance shown in Figure 31. An interesting point is that there is no imbalanced classes except of the coughing part of course and thus the classification algorithm does not need further metric explanatory.

The previous findings come in accordance with the confusion matrix of the proposed classifier shown in Figure 32. Starting with the dummy variable explanation, this is a class used when splitting the data into segments in order to create a fixed window length and iterate through the complete data. The process contained zero padding at the first segment in order to be feasible to create segments of the same size. Thus it shouldn't be considered as a class, although the 7 misclassified samples. Now focusing on the rest classes as it is visible, due to some common characteristics which are included at segments of different classes, the classifier cannot give an accuracy of 100%. Looking at the horizontal line which indicates the actual labels for the different classes, it is easily observed that the highest misclassifications performed at the breath-holding and coughing classes with 6 and 5 false positive samples, respectively but considering the amount of available test data, this findings are more than good enough. Similarly, the highest values for false negative observed between the classes cough-breath holding, deep-cough, shallow-regular and regular-deep with 26 samples in total misclassified. A possible solution to that problem can be either the additional data acquisition and retraining of the classifier or the usage of smaller window segments to avoid features of two classes be presented into one segments. On the other hand, this fact may decrease the intra-variability learning of the algorithm since with shorter segments, less breathing cycles will be included at each window. As a result, this result should be considered acceptable based on the application nature and requirements. The barplot next to the graph shows also the number of samples being classified at each class. Note that the complete data consisted of 7470 samples and a train/test split of 85%/15% has been performed for each class.

4.3 Internal motion and surrogate correlation

In this section, the aforementioned steps for the results of the post-processing methods, temporal synchronization and alignment, along with the fitting methods will be analyzed using the stated validation metrics. Afterwards, a discussion part will follow with interpretation of the results, explaining some assumptions.

4.3.1 Internal Motion and surrogate data analysis

Firstly, a schematic representation of the internal motion of subject is given in Figure 33 to obtain a clear aspect of how the data looks like. The presented graphs has pre-processed data after noise removal and signal smoothing. The depicted data are constructed after a complete MRI experiment of subject A and the motion has been generated based on the

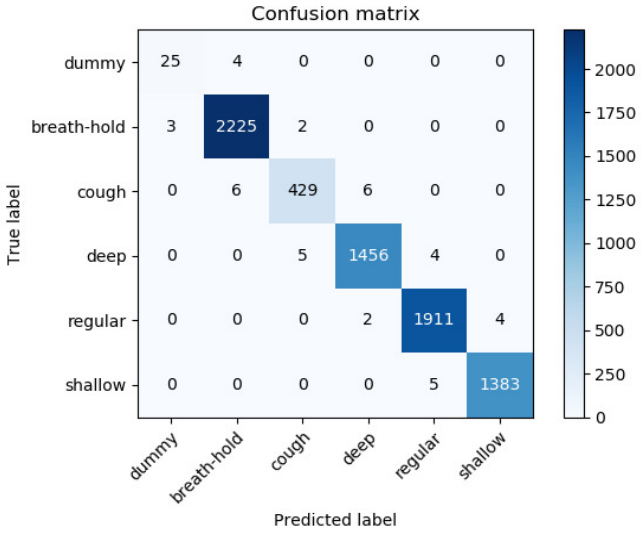


Figure 32. Confusion matrix for the different classes along with a barplot indicating the number od instances per class. The diagonal shows the correct classified samples.

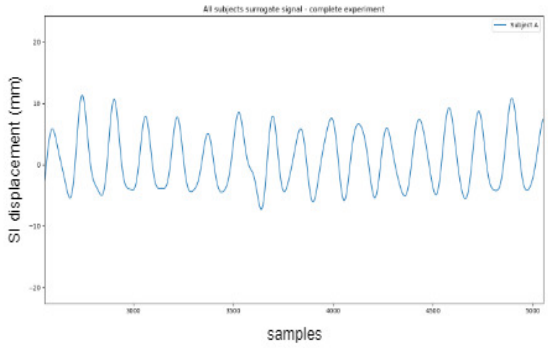


Figure 33. Liver displacement in SI direction over a time period.

SI-direction of the liver upper border's highest point over time. The vertical axis stands for the SI-displacement in mm while the x-axis gives the number of samples or the relative time step. In contrast with the surrogate motion's magnitude, the initial number of samples in this case are much lower since the acquisition rate of the surrogate was approximately 40 times higher than the one used for the MRI data. Furthermore, it has been noticed that the scale of observed magnitude of MRI scan was about 40 times higher compared to the magnitude of the surrogate and the motion was more periodic in some segments. Additionally, further data analysis has been performed revealing valuable information for the processed data. For that reason, the mean and standard deviation of the corresponding measurements for the surrogates and internal motion signals have been calculated and summarized in Tables 6, 7, 8, 9, 10.

Starting with the similarities of the liver motion data, the observed mean values showed a slightly convergences to the negative sign which comes in accordance with the assumption of larger or many more displacements in the lower area with

Table 6. Data analysis Table of the surrogate signal.

Subject	Surrogate magnitude
	$(\mu \pm \sigma)$ mm
A	4.86 $\pm$0.01
C	4.85 $\pm$0.01
D	4.85 $\pm$0.01
E	4.85 $\pm$0.01

the reference value the origin of the liver at the first frame. Of course this is subjected to the performed respiration type as regular breathing and coughing, most subjects have increased instances of negatively large displacements while the rest breathing patterns showed in most cases a longer motion of the liver in the superior direction. Add to that the subjects A and E had more or larger scale troughs than peaks for most upwards or downwards motion over time. This can be linked to the respiratory phase and the amount of inhaled and exhaled air varying from subject to subject and from fraction to fraction including all the variances in the proposed model. It is high importance to notice the high σ values for all subjects raising the need for further data analysis such as box plot to find out potential outliers or experimental errors. Thus, a small section will be dedicated to further analysis in the coming subsection. Regarding the surrogate measurements, all subjects have similar mean and low σ values leading to follow the same pattern. Of course, this is expected since the dataset is small and there is less inter-subject variability. When looking at the surrogate data and excluding subject B (as stated previously, experimental measurement error for this subject), the data contain lots of outliers diverging, although that the min,max values for all participants are close to the median. For all subjects (excluding B), the variance was really low about 0.01 mm, with magnitude values ranging between 4.85 and 4.90mm for a corresponding approximately 22.0-25.0 mm liver SI-displacement.

4.3.2 Power Spectral Density & Boxplot

As stated previously, further data analysis need to be performed to investigate the data distribution, the extreme values and how they affect the final outcome in the data. To start with Figure 34, the normalized power spectral density (PSD) functions of the surrogate and the liver motion data are presented for the different subjects. The PSD shows how strong is the variance at each frequency or simpler, reveals information for which frequencies have the most and the least variations. The PSD can be described in units of energy per frequency and can be directly calculated using the Fast Fourier Transformation. Looking at the same rows, it gives an indication about the intrafractional variances between the subjects which in that case, are really small or obscure translated into a non-patient specific model which can be easily generalizable. Similarly, this means that the model will include most variations and does not necessarily need to be updated at each session. Additionally, the clearness of the signal shows no noise in the data after applying the Savitsky-Golay filter and the pre-processing

Table 7. Data analysis of the internal motion using the regular breathing data.

Subject		A	C	D	E
[min, max]	mm	[-15.52, 7.95]	[-12.69, 7.96]	[-3.14, 17.46]	[-15.75, 6.26]
Superior displacement (S)	$(\mu \pm \sigma)$ mm	-1.67 ±4.64	-0.33 ±2.86	6.29 ±5.83	-7.90 ±4.35
Inferior displacement (I)	$(\mu \pm \sigma)$ mm	-12.94 ±2.15	-11.23 ±0.95	-1.05 ±0.89	-13.69 ±0.86
S - I displacement	$(\mu \pm \sigma)$ mm	11.27 ±5.12	10.91 ±3.01	7.34 ±5.90	5.79 ±4.43

Table 8. Data analysis of the internal motion using the breath-holding pattern data.

Subject		A	C	D	E
[min, max]	mm	[-19.28, 9.37]	[-10.92, 17.24]	[-4.76, 25.87]	[-15.90, 19.42]
Superior displacement (S)	$(\mu \pm \sigma)$ mm	-9.48 ±8.46	10.04 ±4.19	5.09 ±6.75	7.99 ±6.71
Inferior displacement (I)	$(\mu \pm \sigma)$ mm	-16.49 ±1.30	3.89 ±8.22	-0.63 ±2.07	2.05 ±11.07
S - I displacement	$(\mu \pm \sigma)$ mm	7.01 ±8.56	6.14 ±9.23	5.73 ±7.06	5.95 ±12.94

Table 9. Data analysis of the internal motion using the shallow breathing data.

Subject		A	C	D	E
[min, max]	mm	[-18.49, 7.91]	[-9.70, 15.08]	[-4.93, 30.54]	[-14.71, 18.18]
Superior displacement (S)	$(\mu \pm \sigma)$ mm	1.70 ±3.39	8.78 ±3.20	4.74 ±6.10	3.44 ±5.50
Inferior displacement (I)	$(\mu \pm \sigma)$ mm	-4.28 ±5.14	3.58 ±4.69	-0.33 ±3.37	-1.34 ±6.65
S - I displacement	$(\mu \pm \sigma)$ mm	5.97 ±6.16	5.20 ±5.68	5.07 ±6.97	4.77 ±8.63

Table 10. Data analysis of the internal motion using the deep breathing data.

Subject		A	C	D	E
[min, max]	mm	[-18.32, 39.93]	[-13.81, 19.73]	[-17.36, 43.95]	[-19.13, 28.22]
Superior displacement (S)	$(\mu \pm \sigma)$ mm	9.75 ±17.46	7.31 ±11.59	22.96 ±11.40	2.32 ±14.40
Inferior displacement (I)	$(\mu \pm \sigma)$ mm	-10.62 ±8.71	-4.84 ±10.63	0.90 ±7.57	-10.35 ±8.54
S - I displacement	$(\mu \pm \sigma)$ mm	20.37 ±19.51	12.15 ±15.73	22.05 ±13.69	12.67 ±16.75

Table 11. Data analysis of the internal motion using the coughing pattern data.

Subject		A	C	D	E
[min, max]	mm	[-14.55, 5.99]	[-14.76, 3.83]	[-11.81, 32.01]	[-14.01, -0.44]
Superior displacement (S)	$(\mu \pm \sigma)$ mm	-1.35 ±3.23	-7.57 ±4.99	18.04 ±7.96	-13.02 ±0.05
Inferior displacement (I)	$(\mu \pm \sigma)$ mm	-12.44 ±1.24	-12.12 ±1.63	3.76 ±7.49	-13.28 ±0.21
S - I displacement	$(\mu \pm \sigma)$ mm	11.09 ±3.46	4.55 ±5.25	14.27 ±10.93	0.27 ±0.22

steps for both signals. At the end, it is important to mention that the signals have not been distorted by any other organ motion such the heart beats. For the liver motion data, this can be surely avoided due to the choice of the temporal resolution that cannot ”see” such high frequencies. On the other hand, regarding the surrogate signal, the clear obtained results, highly depends on the choice of the deepest to the liver measurement. The penetration depth was chosen to be the deepest available, about 7cm that resulted in the shown outcome, plus the fact that the US transducer was placed away from the heart, avoiding undesired signal detection generated by the heart frequency [81]. For a better understanding, boxplots for the liver SI-displacement data and for the US magnitude data were plotted together, shown in Figure 35. A boxplot can describe the data distribution using five parameters, namely: minimum, Q1, median, Q3 and maximum. It explains also the potential outliers along with their values and gives a further understanding for the data, such as about symmetry, skewness or how the data are grouped. Looking at the left part of Figure 35, where the liver motion PSD is shown, it is visible that the median (middle value at this dataset) of subject A and C is similar when this is also true for the subject B and D. Subject B has the largest range of motion followed by participants A, C and D range of motion. Furthermore, the stated groups of participants shares also similar minimum and maximum values but there are some extreme values located for participant C and many more and larger for participant E. This can be related to some high peaks and valleys for most respiration patterns that have been observed for this specific participant but due to limitations space are not presented here. These extreme values are fully connected with the breathing type and/or combined with this specific subject. The visual information for this subject showed a shifted signal compared to rest of the subjects (A, C) shared common properties(e.g median etc). General the liver motion range followed a range as expected from the previous work [27].

4.3.3 Polynomial Order Analysis

Reaching the second main core of the presented work which is the fitting methods, it has been found that for the polynomial

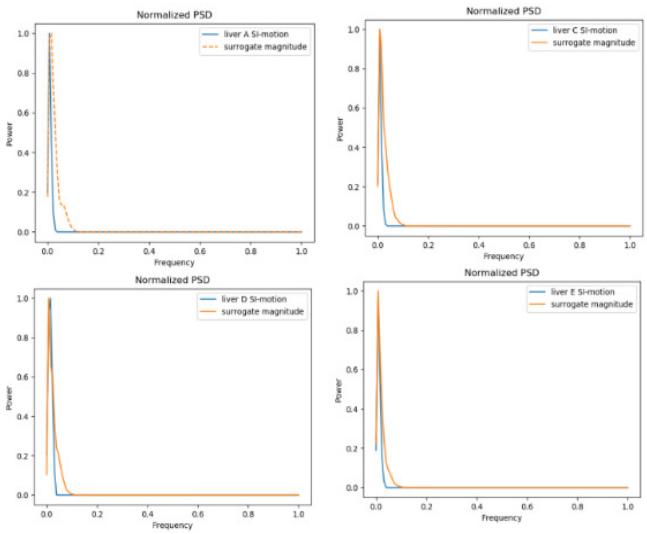


Figure 34. Normalized power spectral density (PSD) functions for subjects A,C,D and E. The liver SI-motion is shown with orange for every subject and a corresponding surrogate signal's PSD is presented along with blue color. The signals store most of the energy in the frequency range of [0.05, 0.1Hz].

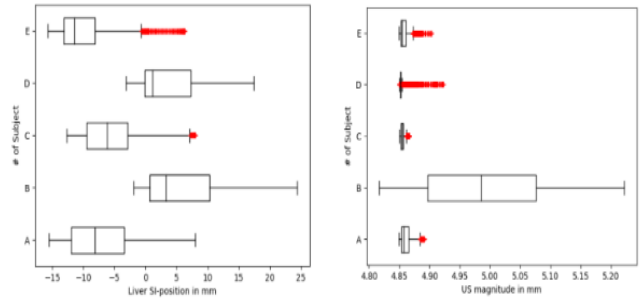


Figure 35. Boxplots of the liver motion and surrogate signal depicted per subject data.

fitting further analysis was needed to determine the optimal parameters for this and the single layer perceptron model, in contrast with the simple linear regression without any regularization method that seemed to be easy in construction.

Starting with the polynomial fit, to find the optimal degree for the polynomial different models have been developed. The final goal for this method was to evaluate the train and test error in order to find out the best fit on the data depending solely on the polynomial degree. For that reason, a wide range of different degree values has been tested and its summary can be found in Figure 36. In this Figure, the vertical axis contains the error in terms of RMSE while the evaluated polynomial orders range between 1 and 20. The train error (blue line) is noticed to be higher than the test set error(red line), this is probably due to the selected small test set. This means that the model has optimal values for this (overfitting). On the other hand, if there is a reduction in the order such as going back on 2nd or 3rd order the error will get increased. As a

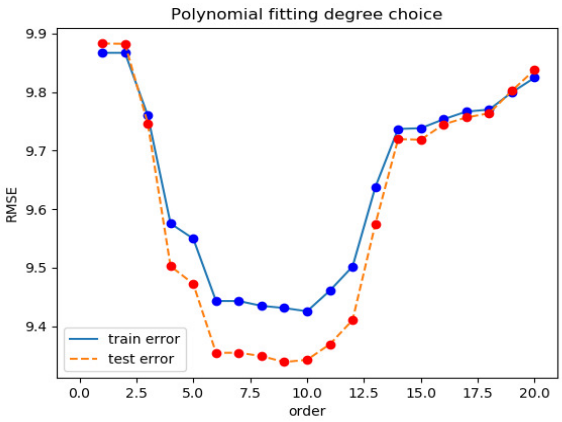


Figure 36. Selecting the polynomial degree using train and test set.

result, it has been decided to kept in the minima error value of test set although be aware for the aforementioned information and choosing the final polynomial to have an order of nine.

4.4 Estimation Accuracy

In this subsection, firstly a full description on the followed validation methods will be given, namely k-fold cross validation. Next, the findings of the three different motion models will be presented along with the selection of the best one and comparison with the previous research work will follow in the next Chapter.

The first evaluation method is called K-fold cross validation and the number of suitable folds has either to do with the amount of available data or with the level of confidence interval and how "optimistic" wants to be the result. For the selected work, 10 folds have been chosen as proposed from the most studies related to sequence analysis. Using a 10-fold cross validation, the dataset is split into 10 parts and subsequently the training of the model starts using k-1 as training set and the k-th split for the test. Note that there is no interaction between train and test set to avoid representation of the same data into the two sets. Although it is possible due to randomization process before the splitting process to have patterns of the same patient into train and test set but statistically the validation method is correct.

To start with the MSE metric, the findings are summarized in Table 12. In the second column, simple linear regression's (SLR) results are presented and focusing on the differences between the subjects, the best fit has been performed using data of subject D achieving a MSE of 0.70 ± 0.12 mm. Note that the σ value is also on of the lowest among the others. Moving to the 9-th order polynomial fit, the results are similar with the SLR and in all cases achieved a similar or higher value for the MSE. On the other hand, the number of parameters used for the construction of that fitting models is higher compared to the one of simple linear regression model and by taking into account the computational load, it does not add a really high value to the results. Now, looking on the data of the sin-

Table 12. Table with the normalized MSE values for the three different models. All the values are given in mm.

Subject	SLR	Polynomial	Perceptron
	$(MSE \pm \sigma)$	$(MSE \pm \sigma)$	$(MSE \pm \sigma)$
A	0.94 ±0.07	0.86 ±0.12	0.76 ±0.64
C	0.92 ±0.11	0.97 ±0.08	0.85 ±0.52
D	0.70 ±0.12	0.83 ±0.06	0.50 ±0.86
E	0.88 ±0.12	0.97 ±0.08	0.90 ±0.49

Table 13. Table with the normalized RMSE values for the three different models. All the values are given in mm.

Subject	SLR	Polynomial	Perceptron
	$(RMSE \pm \sigma)$	$(RMSE \pm \sigma)$	$(RMSE \pm \sigma)$
A	1.15 ±0.13	1.15 ±0.14	0.73 ±0.68
C	1.16 ±0.14	1.15 ±0.14	0.88 ±0.48
D	1.16 ±0.12	1.16 ±0.14	0.76 ±0.64
E	1.15 ±0.14	1.16 ±0.15	0.89 ±0.48

gle layer perceptron trained for 100 epochs with a batch size of 100, the mean values have significantly decreased about 20-25% while the σ value got increased, introducing higher variance on that model. All the results have been calculated with a 95% confidence interval and the values are presented normalized in mm. The average model performance for the MSE was 0.88 ± 0.09 taking into account all different fitting models.

Moving on the RMSE metric-Table13, three additional models have been developed to find out the behavior and measure the accuracy using the second evaluation metric. To begin with the SLR, the values were higher, about 0.20-0.30 mm than the corresponding ones using MSE. In this case, all subjects using the SLR fitting method had the same mean value with slightly higher standard deviation values. In the same manner, the polynomial fitting generated similar to the SLR and slightly decreased. On the other hand, the perceptron network performed much better also utilizing that metric, as previously done with the MSE as well, decreasing the overall error. The error values in that case were approximately 30-35% lower, resulting in a much more accurate estimated signal. Of course, again the *sigma* values were higher, about 5 times which can introduce some uncertainties in the final result. At the end, it should be highlighted that using the RMSE, subject D did not differ from the rest participants results excluding the perceptron performance and the overall mean error value for the RMSE is 1.15 ± 0.002.

5. Limitations and Future Work

For the presented study, a 3T field MRI machine with a temporal resolution of 0.8s and a spatial resolution of (1.875,1.875) mm has been used to acquire the motion data. The spatial resolution was good enough overall for detecting the liver motion in most directions while on the other hand, the machine could provide higher frame rate resulting in longer scanning sessions. As a result, a machine with higher temporal resolution could increase the model performance providing additional images which can capture inhalation and exhalation variations. Secondly, the experiments have been conducted using ten healthy subjects without having any lesion or tumor compared to previous studies. [32, 49] Finding the most relevant clinically region of interest was crucial and challenging. Different segments of the liver have been explored to find the optimal region of interest. Firstly, anatomical features like the vessels have been exploited in combination with a convolutional neural network called U-net. The idea behind it was an automated segmentation of a point which could not provide intrinsic information about the liver motion and located relatively in the middle of the liver. It was based on image enhancements methods, morphological operations and image region-based segmentation method using global thresholding [82, 83]. As shown later, inconsistencies of the vessels from frame to frame combined with the relatively low spatial resolution leading to blurring in the image pixels, thus it has been decided to exclude this method from the presented approach. In addition vessel selection was needed in case of multiple detection when the performance of this model was highly dependent on the image contrast and brightness levels and consequently to the MRI machine settings. A schematic representation is shown in Figure 37. In the first row an annotated image of the liver with a vessel is presented and the masked image used as a label-ground truth for the neural network presented below.

Next to that a method using Canny edge detection algorithm have been exploited to find the complete contour of the liver and the select the central point on that but unfortunately, the method proved to be insufficient due to incorrect liver contour segmentation [84]. For both cases, higher spatial resolution was needed in order to enhance the distinctive regions of the liver and its walls around it. Based on the previous findings, further improvements includes an increased number of subjects for an enhanced inter- and intra-cycle variations as well as intrafractional variations. Furthermore, experiments with patients who have a real tumor should be conducted since the anatomical features will be present in the images and there is no need for detection of other region of interests. On the other hand, using solely the vessels can be helpful as stated and a recommendation for future work would be to create first a dataset of ground truth with the position of the vessels in the image using the aforementioned method but taking into account a higher temporal and spatial resolution as well as delineation of the vessels from a medical expert and then training a neural network to perform automatically the process of the vessel detection and consequently generating the liver motion. Additionally, in the presented study, only the SI motion has been used since it has the largest amount of displacement but for a more accurate and higher model performance, the AP motion should also be explored. After having available both motions, PCA can be applied to create 1D weights that can be fed to the linear fitting methods later on. Another improvement would be to use different models for modelling the

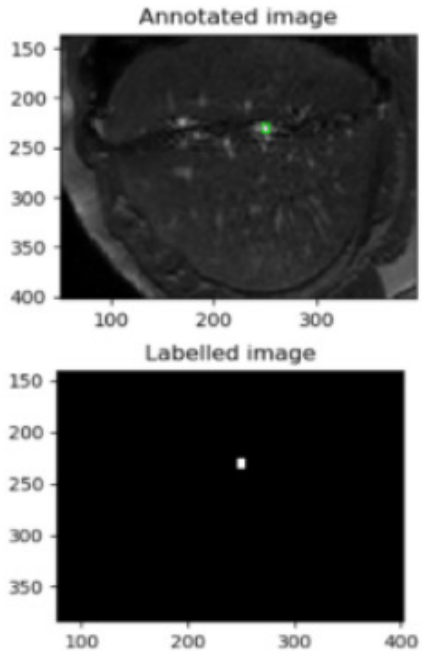


Figure 37. Vessel segmentation method. Upper: 2D MRI liver image rescaled and highlighted in the location of a vessel. Lower: Mask image or label corresponding to the upper image. In this case the pixel belonging to the vessel have value of 1 while the pixels of background get a value of 0. A set of those images has been fed to the automated segmentation algorithm.

relation of the surrogate and the liver motion in inhalation and exhalation phases. Of course this would probably cost in the computational time of the final model since a classifier should be utilised to assign to the two different classes the new data at each session.

Moreover, it is also important to provide an in-box temporal alignment for the two signals making use of a shared clock or signal triggering box for generating enhanced results in terms of error estimation. This could also remove some workload from the author since some post-processing methods would not be needed. Looking at Tables 12, 13, the outcome between the first two models (simple linear regression and polynomial fitting) and polynomial fitting are similar while there is a significant decrease of the error using the single layer perceptron model. Note that it is possible to obtain a lower estimation error using a more complex regression model combined with an increased subject data and multiple surrogate signals [32, 85, 39]. At the end, it would be interested to mention that the presented model outperformed the previous studies findings in terms of error minimization and available data [10, 86]. On the other hand, evaluation of the model should be performed on a real clinical problem and additional test using metrics such R square should be done. Similarly, the duration of 6sec prediction for the surrogate signal and consequent for the liver motion, it is a promising result but taking into account the high probabilities for overfitting due to the small amount of available trained data, further research should be done. Compared to the recent study of Lin et al. where the available dataset includes more than 1000 subjects, it is feasible to predict correctly the liver motion only within 300ms when the available data for the presented thesis were 10 participants. As a result, the findings of this work arises further analysis to validate the final outcome.

6. Conclusions

In the presented study, the author utilised an RME approach for the liver SI-motion estimation and consequently the framework has been evaluated. The framework consists of two parts, the motion model between the liver motion and the surrogate data and secondly the surrogate prediction combined with its classifier. Surrogate signals presented as the deepest to the liver emitted wave from an ultrasound sensor have been used as the data that has strong correlation with the liver actual motion and due to its MRI compatibility, non-imaging dependent and because of its high temporal and spatial resolution. On the next phase, both data, surrogate and liver are fed into the regression models and the model learns the parameters in order to create the so-called motion model. Afterwards, it is possible to estimate the motion data depending only on surrogate data. By training the surrogate model to make predictions for the upcoming state, it is also feasible to estimate liver motion in a future state.

Three different regression models have been exploited and evaluated on its performance comparing the motion estimated with the actual liver values. Data generated by healthy subjects have been used acquired using MRI modality simultaneously with the US data. Firstly, the data have been segmented on the liver upper border in SI-direction while secondly have been processed and aligned with the US signal and consequently split for model training and evaluation. The regression models have been trained using the training set, optimally tuning of their parameters performed using the internal validation set and evaluated on the test data. As a second step, the surrogate data have been fed to a neural network in order to obtain a model that can predict its motion on a future state and through its high correlation with the liver motion, the latter can be predicted also in the future providing an advantage to the clinical world. The offered advantage has application to radiotherapy or liver biopsy predicting th liver motion resulting in a minimized damage of healthy tissue during a potential procedure. The training phase contain training of five different models for the five breathing patterns (regular, deep, shallow, coughing and breath-holding) while in the prediction phase, there is a classifier sliding over the data with a fixed window length and selecting one of the several trained models based the performed respiration pattern. Summarizing, the objective of this thesis was to choose suitable surrogate signals along with regression methods for the liver motion estimation and prediction on a future state. The final outcome showed that using an US sensor as surrogate signal can give an estimation error

less than 1.2 ± 0.1 mm using RMSE metric and the surrogate signal can be predicted 6 seconds ahead with an accuracy of 2.5mm using Euclidean Distance using a classifier for activity recognition of 99.5%. Compared to previous findings, the results are promising and further studies are needed for building more complex models combined with higher dimensionality data and increased liver motion data. [10]

Acknowledgements

Firstly, I would like to thank Dr. ir. M. Abayazid for his continuous willingness and supervision which helped me to overcome difficult working situations and stressful periods. Moreover, I would like to thanks Dr. B. Sirmaçek and Dr. ir. H. Naghibi Beidokhti for the provided programming, medical nature problems and research supervision, over the whole thesis period. Next, I would like to thank you Prof. Dr. ir. C.H.Slump and Dr. ir. B.J.F van Beijnum for their guidance and supervision. At the end, I would like to thank you my family and my friends who continuously supporting me and motivating me.

References

[1] W. H. Organization, "Cancer, 2018."

[2] A. C. Society, "Treatment of liver cancer by stage, 2018."

[3] C. R. UK, "Treatment, 2018."

[4] P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, *et al.*, "The management of respiratory motion in radiation oncology report of aapm task group 76 a," *Medical physics*, vol. 33, no. 10, pp. 3874–3900, 2006.

[5] M. L. Wood and R. M. Henkelman, "Mr image artifacts from periodic motion," *Medical physics*, vol. 12, no. 2, pp. 143–151, 1985.

[6] C. R. UK, "Tests to diagnose, 2018."

[7] D.-H. Wu, L. Liu, and L.-H. Chen, "Therapeutic effects and prognostic factors in three-dimensional conformal radiotherapy combined with transcatheter arterial chemoembolization for hepatocellular carcinoma," *World journal of gastroenterology*, vol. 10, no. 15, p. 2184, 2004.

[8] W. health organization, "Cancer key facts, 2017."

[9] M. Abayazid, T. Kato, S. G. Silverman, and N. Hata, "Using needle orientation sensing as surrogate signal for respiratory motion estimation in percutaneous interventions," *International journal of computer assisted radiology and surgery*, vol. 13, no. 1, pp. 125–133, 2018.

[10] S. Fahmi, F. F. Simonis, and M. Abayazid, "Respiratory motion estimation of the liver with abdominal motion as a surrogate," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 6, p. e1940, 2018.

[11] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King, "Respiratory motion models: a review," *Medical image analysis*, vol. 17, no. 1, pp. 19–42, 2013.

[12] A. Grant and J. Neuberger, "Guidelines on the use of liver biopsy in clinical practice," *Gut*, vol. 45, no. suppl 4, pp. IV1–IV11, 1999.

[13] B. Sirmaçek, M. Abayazid, H. N. Beidokhti, and M. Berijanian, "Toward patient-specific estimation of hepatic tumors respiratory motion: A finite element-based machine learning approach," in *33rd International Conference on Computer Assisted Radiology and Surgery 2019*, 2019.

[14] C. Saw, E. Brandner, R. Selvaraj, H. Chen, M. S. Huq, and D. Heron, "A review on the clinical implementation of respiratory-gated radiation therapy," *Biomedical imaging and intervention journal*, vol. 3, no. 1, 2007.

[15] Y. Zhou, K. Thiruvalluvan, L. Krzeminski, W. H. Moore, Z. Xu, and Z. Liang, "Ct-guided robotic needle biopsy of lung nodules with respiratory motion–experimental system and preliminary test," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 3, pp. 317–330, 2013.

[16] S. Shimizu, H. Shirato, B. Xo, K. Kagei, T. Nishioka, S. Hashimoto, K. Tsuchiya, H. Aoyama, and K. Miyasaka, "Three-dimensional movement of a liver tumor detected by high-speed magnetic resonance imaging," *Radiotherapy and oncology*, vol. 50, no. 3, pp. 367–370, 1999.

[17] S. R. Abdel-Misih and M. Bloomston, "Liver anatomy," *Surgical Clinics*, vol. 90, no. 4, pp. 643–653, 2010.

[18] J. B. West, *Respiratory physiology: the essentials*. Lippincott Williams & Wilkins, 2012.

[19] J. R. McClelland, S. Hughes, M. Modat, A. Qureshi, S. Ahmad, D. Landau, S. Ourselin, and D. Hawkes, "Inter-fraction variations in respiratory motion models," *Physics in Medicine & Biology*, vol. 56, no. 1, p. 251, 2010.

[20] S. Hughes, J. McClelland, S. Tarte, J. Blackall, J. Liong, S. Ahmad, D. Hawkes, and D. Landau, "Assessment of respiratory cycle variability with and without training using a visual guide.," *Cancer Therapy*, vol. 6, no. 2, 2008.

[21] J. Blackall, S. Ahmad, M. Miquel, J. McClelland, D. Landau, and D. Hawkes, "Mri-based measurements of respiratory motion variability and assessment of imaging strategies for radiotherapy planning," *Physics in Medicine & Biology*, vol. 51, no. 17, p. 4147, 2006.

[22] G. Benchetrit, "Breathing pattern in humans: diversity and individuality," *Respiration physiology*, vol. 122, no. 2-3, pp. 123–129, 2000.

[23] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing," *Journal of applied physiology*, vol. 22, no. 3, pp. 407–422, 1967.

[24] J. T. Sharp, N. B. Goldberg, W. S. Druz, and J. Danon, "Relative contributions of rib cage and abdomen to breathing in normal subjects," *Journal of Applied Physiology*, vol. 39, no. 4, pp. 608–618, 1975.

[25] A. De Troyer and M. Estenne, "Coordination between rib cage muscles and diaphragm during quiet breathing in humans," *Journal of Applied Physiology*, vol. 57, no. 3, pp. 899–906, 1984.

[26] F. Preiswerk, M. Toews, C.-C. Cheng, J.-y. G. Chiou, C.-S. Mei, L. F. Schaefer, W. S. Hoge, B. M. Schwartz, L. P. Panych, and B. Madore, "Hybrid mri-ultrasound acquisitions, and scannerless real-time imaging," *Magnetic resonance in medicine*, vol. 78, no. 3, pp. 897–908, 2017.

[27] K. Langen and D. Jones, "Organ motion and its management," *International Journal of Radiation Oncology* Biology* Physics*, vol. 50, no. 1, pp. 265–278, 2001.

[28] J. Ehrhardt, C. Lorenz, *et al.*, *4D modeling and estimation of respiratory motion for radiation therapy*, vol. 10. Springer, 2013.

[29] M. A. Clifford, F. Banovac, E. Levy, and K. Cleary, "Assessment of hepatic motion secondary to respiration for computer assisted interventions," *Computer Aided Surgery*, vol. 7, no. 5, pp. 291–299, 2002.

[30] T. W. Kang, M. W. Lee, D. I. Cha, H. J. Park, J. S. Park, W.-C. Bang, and S. W. Kim, "Usefulness of virtual expiratory ct images to compensate for respiratory liver motion in ultrasound/ct image fusion: A prospective study in patients with focal hepatic lesions," *Korean journal of radiology*, vol. 20, no. 2, pp. 225–235, 2019.

[31] L. Jonveaux, "Arduino-like development kit for single-element ultrasound imaging," *Journal of Open Hardware*, vol. 1, no. 1, 2017.

[32] F. Ernst, V. Martens, S. Schlichting, A. Beširević, M. Kleemann, C. Koch, D. Petersen, and A. Schweikard, "Correlating chest surface motion to motion of the liver using ε-svr–a porcine study," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 356–364, Springer, 2009.

[33] A. P. King, R. Boubertakh, K. S. Rhode, Y. Ma, P. Chinchapatnam, G. Gao, T. Tangcharoen, M. Ginks, M. Cooklin, J. S. Gill, *et al.*, "A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures," *Medical image analysis*, vol. 13, no. 3, pp. 419–431, 2009.

[34] J. M. Blackall, G. P. Penney, A. P. King, and D. J. Hawkes, "Alignment of sparse freehand 3-d ultrasound with preoperative images of the liver using models of respiratory motion and deformation," *IEEE transactions on medical imaging*, vol. 24, no. 11, pp. 1405–1416, 2005.

[35] A. P. King, C. Jansen, K. S. Rhode, D. Caulfield, R. Razavi, and G. P. Penney, "Respiratory motion correction for image-guided cardiac interventions using 3-d echocardiography," *Medical Image Analysis*, vol. 14, no. 1, pp. 21–29, 2010.

[36] A. P. King, C. Jansen, R. Boubertakh, K. S. Rhode, R. Razavi, and G. P. Penney, "Model-based respiratory motion correction using 3-d echocardiography," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1465–1468, IEEE, 2008.

[37] D. Peressutti, E.-J. Rijkhorst, D. C. Barratt, G. P. Penney, and A. P. King, "Estimating and resolving uncertainty in cardiac respiratory motion modelling," in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 262–265, IEEE, 2012.

[38] D.-S. Shin, S.-H. Kang, D.-S. Kim, T.-H. Kim, K.-H. Kim, H.-J. Koo, M.-S. Cho, J.-S. Ha, D.-K. Yoon, and T. S. Suh, "Development of an advanced deformable phantom to analyze dose differences due to respiratory motion," *Progress in Medical Physics*, vol. 28, no. 1, pp. 1–10, 2017.

[39] R. Dürichen, L. Davenport, R. Bruder, T. Wissel, A. Schweikard, and F. Ernst, "Evaluation of the potential of multi-modal sensors for respiratory motion prediction and correlation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5678–5681, IEEE, 2013.

[40] P. Lei, F. Moeslein, B. J. Wood, and R. Shekhar, "Real-time tracking of liver motion and deformation using a flexible needle," *International journal of computer assisted radiology and surgery*, vol. 6, no. 3, pp. 435–446, 2011.

[41] C. Buerger, R. E. Clough, A. P. King, T. Schaeffter, and C. Prieto, "Nonrigid motion modeling of the liver from 3-d undersampled self-gated golden-radial phase encoded mri," *IEEE transactions on medical imaging*, vol. 31, no. 3, pp. 805–815, 2012.

[42] B. Chen, N. Weber, F. Odille, C. Large-Dessale, A. Delmas, L. Bonnemains, and J. Felblinger, "Design and validation of a novel mr-compatible sensor for respiratory motion modeling and correction," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 123–133, 2016.

[43] G. Veenstra, "Generating high frame rate mr images using surrogate signals," Master's thesis, University of Twente, 2019.

[44] W. Lu, D. A. Low, P. J. Parikh, M. M. Nystrom, I. M. El Naqa, S. H. Wahab, M. Handoko, D. Fooshee, and J. D. Bradley, "Comparison of spirometry and abdominal height as four-dimensional computed tomography metrics in lung," *Medical physics*, vol. 32, no. 7Part1, pp. 2351–2357, 2005.

[45] P. G. Danias, M. V. McConnell, V. C. Khasgiwala, M. L. Chuang, R. R. Edelman, and W. J. Manning, "Prospective navigator correction of image position for coronary mr

angiography.," *Radiology*, vol. 203, no. 3, pp. 733–736, 1997.

[46] C. Santelli, R. Nezafat, B. Goddu, W. J. Manning, J. Smink, S. Kozerke, and D. C. Peters, "Respiratory bellows revisited for motion compensation: preliminary experience for cardiovascular mr," *Magnetic resonance in medicine*, vol. 65, no. 4, pp. 1097–1102, 2011.

[47] D. A. Low, T. Zhao, B. White, D. Yang, S. Mutic, C. E. Noel, J. D. Bradley, P. J. Parikh, and W. Lu, "Application of the continuity equation to a breathing motion model," *Medical physics*, vol. 37, no. 3, pp. 1360–1364, 2010.

[48] D. A. Low, P. J. Parikh, W. Lu, J. F. Dempsey, S. H. Wahab, J. P. Hubenschmidt, M. M. Nystrom, M. Handoko, and J. D. Bradley, "Novel breathing motion model for radiotherapy," *International Journal of Radiation Oncology* Biology* Physics*, vol. 63, no. 3, pp. 921–929, 2005.

[49] A. S. Beddar, K. Kainz, T. M. Briere, Y. Tsunashima, T. Pan, K. Prado, R. Mohan, M. Gillin, and S. Krishnan, "Correlation between internal fiducial tumor motion and external marker motion for liver tumors imaged with 4d-ct," *International Journal of Radiation Oncology* Biology* Physics*, vol. 67, no. 2, pp. 630–638, 2007.

[50] P.-C. M. Chi, P. Balter, D. Luo, R. Mohan, and T. Pan, "Relation of external surface to internal tumor motion studied with cine ct," *Medical physics*, vol. 33, no. 9, pp. 3116–3123, 2006.

[51] H. Fayad, J.-F. Clément, T. Pan, C. Roux, C. C. Le Rest, O. Pradier, and D. Visvikis, "Towards a generic respiratory motion model for 4d ct imaging of the thorax," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pp. 3975–3979, IEEE, 2009.

[52] S. E. Geneser, J. Hinkle, R. M. Kirby, B. Wang, B. Salter, and S. Joshi, "Quantifying variability in radiation dose due to respiratory-induced tumor motion," *Medical image analysis*, vol. 15, no. 4, pp. 640–649, 2011.

[53] J. Hinkle, P. T. Fletcher, B. Wang, B. Salter, and S. Joshi, "4d map image reconstruction incorporating organ motion," in *International Conference on Information Processing in Medical Imaging*, pp. 676–687, Springer, 2009.

[54] S. Vedam, V. Kini, P. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan, "Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker," *Medical physics*, vol. 30, no. 4, pp. 505–513, 2003.

[55] B. Stemkens, R. H. Tijssen, B. D. de Senneville, H. D. Heerkens, M. Van Vulpen, J. J. Lagendijk, and C. A. van den Berg, "Optimizing 4-dimensional magnetic resonance imaging data sampling for respiratory motion analysis of pancreatic tumors," *International Journal of Radiation Oncology* Biology* Physics*, vol. 91, no. 3, pp. 571–578, 2015.

[56] F. Odille, N. Cîndea, D. Mandry, C. Pasquier, P.-A. Vuissoz, and J. Felblinger, "Generalized mri reconstruction including elastic physiological motion and coil sensitivity encoding," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 59, no. 6, pp. 1401–1411, 2008.

[57] E.-J. Rijkhorst, D. Heanes, F. Odille, D. Hawkes, and D. Barratt, "Simulating dynamic ultrasound using mr-derived motion models to assess respiratory synchronisation for image-guided liver interventions," in *International Conference on Information Processing in Computer-Assisted Interventions*, pp. 113–123, Springer, 2010.

[58] M. Filipovic, P.-A. Vuissoz, A. Codreanu, M. Claudon, and J. Felblinger, "Motion compensated generalized reconstruction for free-breathing dynamic contrast-enhanced mri," *Magnetic resonance in medicine*, vol. 65, no. 3, pp. 812–822, 2011.

[59] J. D. Hoisak, K. E. Sixel, R. Tirona, P. C. Cheung, and J.-P. Pignol, "Correlation of lung tumor motion with external surrogate indicators of respiration," *International Journal of Radiation Oncology* Biology* Physics*, vol. 60, no. 4, pp. 1298–1306, 2004.

[60] A. Schweikard, G. Glosser, M. Bodduluri, M. J. Murphy, and J. R. Adler, "Robotic motion compensation for respiratory movement during radiosurgery," *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery (ISCAS)*, vol. 5, no. 4, pp. 263–277, 2000.

[61] A. Samuel, "Machine learning definition."

[62] C. Hsin and C. Danner, "Convolutional neural networks for left ventricle volume estimation," 2016.

[63] A. Karpathy, "Cs231n convolutional neural networks for visual recognition," *Neural networks*, vol. 1, 2016.

[64] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks. in advances in neural information processing systems (pp. 1097-1105).," 2012.

[65] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *arXiv preprint arXiv:1808.03314*, 2018.

[66] S. Banerjee, "Recurrent neural network, medium, 2018."

[67] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[68] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[70] Y. Saez, A. Baldominos, and P. Isasi, "A comparison study of classifier algorithms for cross-person physical activity recognition," *Sensors*, vol. 17, no. 1, p. 66, 2017.

[71] H. Fayad, T. Pan, J. François Clement, and D. Visvikis, "Correlation of respiratory motion between external patient surface and internal anatomical landmarks," *Medical physics*, vol. 38, no. 6Part1, pp. 3157–3164, 2011.

[72] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[73] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.

[74] K. Konsolakis, "Physical activity recognition using wearable accelerometers in controlled and free-living environments," 2018.

[75] Stackoverflow, "Confusion matrix explanation."

[76] T. L. Szabo and P. A. Lewin, "Ultrasound transducer selection in clinical imaging practice," *Journal of Ultrasound in Medicine*, vol. 32, no. 4, pp. 573–582, 2013.

[77] *Advanced Techniques for Signal Analysis*.

[78] F.-F. Li, A. Karpathy, and J. Johnson, "Cs231n: Convolutional neural networks for visual recognition," *University Lecture*, 2015.

[79] Statisticshowto, "Root mean square explanation."

[80] A. Singhal *et al.*, "Modern information retrieval: A brief overview," *IEEE Data Eng. Bull.*, vol. 24, no. 4, pp. 35–43, 2001.

[81] K. E. Barrett, "Ganong's review of medical physiology," 2012.

[82] J.-J. Ding, "The class of" advanced digital signal processing", the department of electrical engineering," *National Taiwan University (NTU), Taipei, Taiwan*, 2008.

[83] P. Bosdogianni and M. Petrou, *Image processing: the fundamentals*. Wiley, 1999.

[84] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[85] M. Wilms, R. Werner, J. Ehrhardt, A. Schmidt-Richberg, H.-P. Schlemmer, and H. Handels, "Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy," *Physics in Medicine & Biology*, vol. 59, no. 5, p. 1147, 2014.

[86] H. Lin, C. Shi, B. Wang, M. F. Chan, X. Tang, and W. Ji, "Towards real-time respiratory motion prediction based on long short-term memory neural networks," *Physics in Medicine & Biology*, vol. 64, no. 8, p. 085010, 2019.

Appendices

Appendix A - Measurement protocol and breathing types

In this part, it is important to describe the measurement protocol to have a spherical knowledge of the experiment and the data acquisition process where the collection of the two signals have been performed.

Starting with the measurement protocol, two rounds of measurements have been performed. The first one where five healthy subjects (two females and three males) above 18 y.o have taken part in the human subject experiments conducted in the University hospital of Münster (Münster , Germany) and the second one in University of Twente Robotics and Mechatronics (RaM) lab (Enschede, The Netherlands). Each subject was informed about the presented study, receiving a description of the project, the goals and the procedure of the experiment and after the ethical committee of University of Twente approved the proposal. The principal investigators have taken the responsibility to be sure about the MRI safety regulations and informed the participants as well. More in depths, the experiments conducted in Germany included simultaneous acquisition of MRI and US data used in the fitting methods dataset and in the surrogate prediction as well when the acquired data from the Netherlands have been used solely for the surrogate prediction model.

The participants have been asked to follow a sequence of different breathing patterns including regular, normal calm breathing, breath-holding, intermittent holds at the end of exhalation, shallow, short in- and exhalations, deep, large in- and exhalations and coughing. The idea behind the chosen breathing pattern was to include a variety of all possible respiration types while the main limitation of the study was that the participants had to be continuously apply the given instructions for breathing. The aforementioned breathing types have been chosen since they meet some requirements that they are applied in real-time application such interventions or biopsies. Furthermore, the participants were free to breathe using either the lungs or the abdominal breathing resulting in as natural as possible breathing. Some of the included breathing patterns properties that used are developed in the list below to explain why they have been chosen:

- reduce the respiratory motion during image acquisition
- introduce rapid short motion and simulates the hyperventilation
- introduce slow but larger motion
- introduce artefacts in MRI scans

Appendix B - Description of the MRI and Ultrasound devices

For the experiments, a Philips 3T Intera MRI has been used from the Radiology Department of University hospital of Münster combined with a 3.5MHz A-mode ultrasound sensor(Optel Opbox V2.1 Mini Ultrasonic Box with Integrated Pulser and Receiver provided by the Optel Ultrasound Technology, Poland, Wrocław) mounted to the skin at the right

area of the abdomen near the liver. The MRI machine was set according to the following settings:

- 3T field strength
- Sagittal imaging plane
- Fast Field echo pulse sequence
- Cartesian trajectory
- Temporal resolution: 0.79s
- Spatial resolution: (1.875,1.875) mm
- Slice thickness: 2mm
- Echo time: 2.63ms
- Repetition time: 5.27ms
- Image output: (160,160) pixels

while the setting for the US sensor are consisted in the list below:

- Pulse voltage: 240V (level 10)
- Pulse width: 2.8ms
- Sampling frequency: 33.3MHz
- Gain: 24 (pre-amplifier) + 15 (constant) dB
- Delay: 10ms
- Trigger: timer PRF with frequency of 50Hz