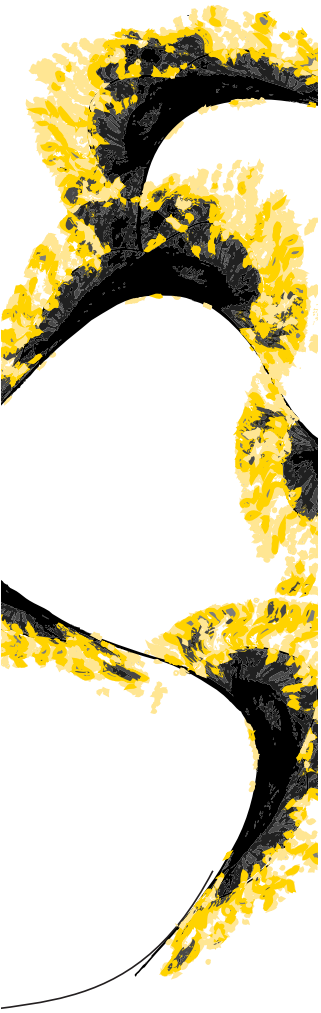


RAM ROBOTICS AND MECHATRONICS

ESTIMATION OF LIVER RESPIRATORY-INDUCED MOTION USING AN INFRARED CAMERA FOR MINIMALLY INVASIVE SURGERY

M. (Mostafa) Elgebily

MSC ASSIGNMENT

Committee:
dr. ir. M. Abayazid
A. Cordon, MSc
dr. I.S.M. Khalil

August, 2024

057RaM2024
Robotics and Mechatronics
EEMCS
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-induced motion using an infrared camera for minimally invasive surgery

Mostafa Elgebily
EEMCS University of Twente
Department of Robotics and Mechatronics (RaM)
Enschede, Netherlands
mostafaahmedgalalelgebily@student.utwente.nl

***Abstract*—Liver respiratory induced motion (RIM) presents a significant challenge during biopsies and minimally invasive hepatic surgeries. A method of battling this RIM is the use of respiratory motion estimation (RME). RME utilises external signals called surrogates to relay internal liver RIM noninvasively. This study presents a novel RME method for liver Superior-Inferior (SI) motion tracking using Infrared (IR) imaging as a surrogate signal and deep learning using convolutional neural networks (CNNs) to directly derive the displacement of the liver from the IR images of the nose/mouth area, as a ROI, in patients. This method utilises the raw image data as an input to the fitting method along with MRI liver dome displacement values as ground truth training values. The training and validation splits were used to train the model on the motion patterns while the test split was fed to the trained model to predict motion data from the thermal images. The model was then validated on three healthy human subjects, each participating in three sessions to acquire displacement and thermal ground truth data using an MRI machine and a thermal camera. This paper, to the best of the author's knowledge, presented a new method for tracking the liver and proved that it can be a viable option. The R^2 scores ranged from 0.75-0.98 throughout the 9 sessions and the MAE ranged from 0.91 mm-3.8 mm. The authors also suggested the use of higher resolution (temporal and spatial) MRI and ultrasound (US) scanners to significantly improve the results and allow the exploration of 3D motion detection instead of only SI.**

***Index Terms*—Respiratory Motion Estimation (RME), Respiratory Induced Motion (RIM), Magnetic Resonance Imaging (MRI), Thermal Infrared Imaging (IR), Convolutional Neural Networks (CNN), Surrogate Signals, Computer Vision, Motion Tracking.**

I. Introduction

A. Liver Cancer

The liver is the largest internal organ in the body [1] and is located in the upper right abdomen. About 9.5 of 100,000 people in the world are affected by liver cancer [2]. A large number of these cancer cases could be treated with early diagnosis and medical interference. The type of treatment and interference differs depending on the type of cancer present and the stage of the cancerous tumour, which can be identified through needle biopsies. These treatments are placed in two main categories, namely, *local treatments* and *systematic treatments* [3]. Local treatments include tumour ablation using percutaneous needle insertion, invasive surgery and radiotherapy as they only target the cancerous cells. In contrast, systematic treatments include drug treatments, such as chemotherapy, immunotherapy, or targeted therapy, as they can affect the entire body and not just the cancerous cells.

Multiple imaging modalities are used alongside these local treatments to aid surgeons during diagnoses and operations by showing the positions of the liver, the motion of the organs or the position of the surgical tools. The importance of these modalities during the diagnosis or detection of cancerous cells is very crucial in some cases to detect the tumours early, as this increases the chance of their treatment and termination. They can also be used to detect which stage this cancer is at as well as the location of these tumours. Some other modalities include X-Ray imaging and Molecular and Nuclear Imaging (PET and SPECT) [5].

Computed Tomography (CT) scans, Magnetic Resonance Imaging (MRI) and ultrasound (US) imaging are the most commonly used modalities in image-guided interventions. The type of modality used differs according to which organ is being operated on (type of tissue layer and its visibility in the imaging modality) and the type of surgical instruments used in the operation (MRI cannot work with ferromagnetic materials). It also depends on whether or not the modality needed should have a high image resolution or a high frame rate, such as the case with MRI and US, respectively [4], for real-time applications during surgery.

B. Respiratory Induced Motion (RIM)

Organs move in the body, especially in the abdominal area, mainly due to the motion of the lungs during inhalation and exhalation. This is caused by the lungs pushing on the diaphragm and rib cage which are in turn pulling/pushing on other organs. This motion is called Respiratory Induced Motion (RIM) and it is the main reason for organ motion during interventional procedures. This motion makes it very hard for the operator to carry out the surgery because the organ location becomes hard to pinpoint while the patient is breathing. The liver displacement due to this RIM can range between 0.8-2.5 cm [7] in a single direction and is mainly dominant in the Superior-Inferior (SI) direction rather than the Anterior-Posterior (AP) direction. The SI displacement is seen to be typically around 25 mm in normal breathing conditions, less than 2 mm in the AP direction [8] and around 1.5 mm in the Left-Right (LR) direction [9]. The SI-AP directions can

be seen in Fig. 1. The LR direction is along the axis going into the page, normal to the SI-AP plane.

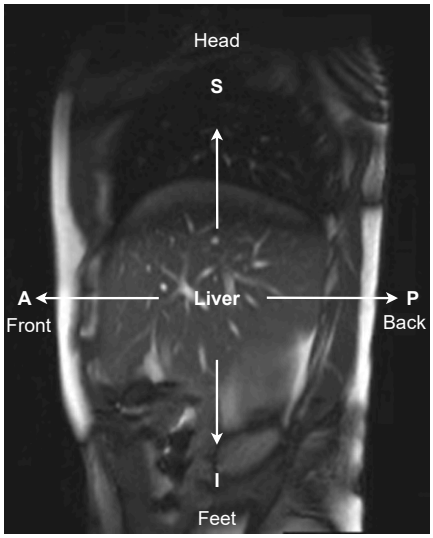


Fig. 1: Superior-Inferior and Anterior-Posterior directions are shown for liver motion. Left-Right motion is normal to the SI-AP plane, going into the figure.

Because of the significance of RIM, imaging modalities must be used to provide visual feedback to the clinician, which can then be further processed to measure and track the liver/tumour location throughout the whole surgery. The average breathing rate of a person is 12-15 breaths per minute at rest and if the imaging modality does not capture at this rate, the captured images of the liver location will be inaccurate. This will cause great uncertainties in locating the liver tumour during the operation, especially in the cases of coughing and gasping [4], where these will cause image artefacts, namely in the case of MRI. This problem could also cause unnecessary exposure of healthy cells to radiation, in radiotherapy, which may damage them and further put the patient at risk.

A conventional method to mitigate such abnormal motions is to ask patients to carry out a "breath-hold" [10]. In this breath-hold, patients are asked to hold their breaths for a certain amount of time and the operation would only be carried out during these windows of breath-holds. However, this proves to be somewhat unreliable in some cases such as in elderly patients, babies/infants, patients with breathing illnesses or patients under heavy sedation (or oversensitive to anaesthesia) where it is hard to carry out a sufficiently long breath-hold or reproduce the same position of that breath-hold, in the case of sedation [7]. Furthermore, the operation time increases drastically because the surgeon is only allowed to operate during these short windows of breath-holds [7].

C. Respiratory Motion Estimation (RME) using External Surrogate Signals

Respiratory Motion Estimation (RME) is used to estimate the motion of the liver rather than measure it. It works by utilising an external signal that does not measure the actual motion of the liver but rather another Region of Interest (ROI) that has a strong correlation with the actual motion and can be easily measured with a high frame rate/frequency [7]. A brief overview of the steps that occur during RME can be seen in Fig. 2. This signal is called *Surrogate Signal* (SS) and it is fed into a correspondence model. A correspondence model uses the correlation between the surrogate and the ground truth object (liver in this case) and is trained using a fitting method along with ground truth data from one of the common imaging modalities previously mentioned. This correspondence model estimates the values of the ground truth without the need for the actual ground truth. In other words, after training, the correspondence model estimates the liver motion and location only using data from the surrogate signal with no ground truths. This method allows for a more versatile setup when common imaging modalities cannot be used.

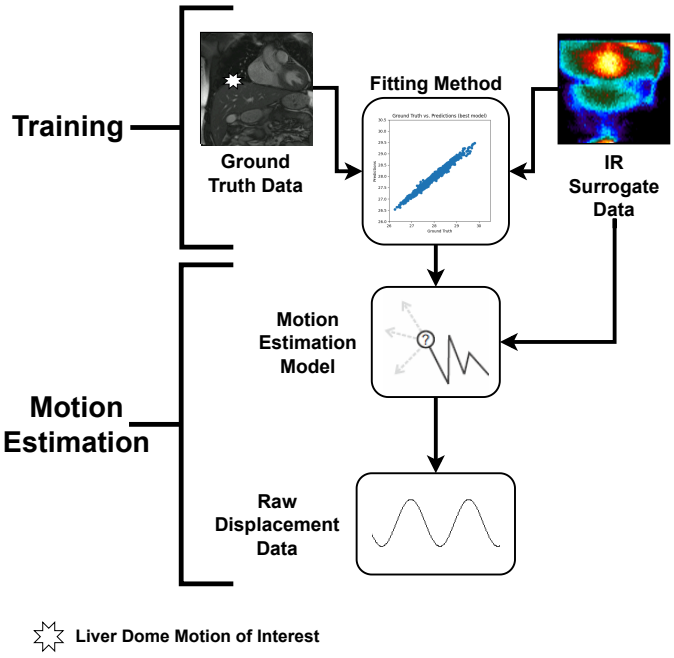


Fig. 2: Flowchart of respiratory motion estimation steps: The ground truth data (actual liver motion) and the IR surrogate data (thermal breathing images) are both acquired simultaneously. The motion estimation model is then trained using the fitting method chosen to independently predict the internal motion without the need for ground truth motion data. The trained model can then output raw displacement data of the liver motion when given a thermal image as an input.

D. Goals of this study

This study will focus on RME of the liver using Infrared (IR) imaging as a surrogate signal. The ground truth data was supplied from MRI. The goals of this study are listed as:

- Validating the feasibility of using thermal imaging as an independent (doesn't depend on other modalities during testing phase) external surrogate signal by comparing them to MRI ground truth values

- Develop a neural network (NN) correspondence model that is fed processed IR images to estimate the RIM of the liver (per patient)
- Validating the accuracy of the model by conducting a clinical study with healthy human subjects as participants.

E. Contribution and Novelty

IR is chosen as a surrogate signal and an IR camera is used to acquire the breathing of a patient. Using the correlation between temperature magnitude change, we can estimate the displacement magnitude change of the liver. To the authors' knowledge, IR imaging of the nasal area temperature and relating it to internal organ motion has not been thoroughly studied in literature. Thus, the main goal of this study is to investigate whether it can estimate internal liver motion. Furthermore, a correspondence model that uses an input of raw IR images into an NN to directly estimate liver motion has never been presented before, based on the author's literature review, presented in the following section. Thus, the authors have decided to implement this idea to test the validity of this new surrogate signal.

F. Paper Outline

The paper follows a chronological, step-wise structure. Firstly, the types of surrogate data are mentioned in section II. In the same section, thermal imaging of the nose/mouth ROI will be chosen and justified as the optimal surrogate choice during MRI acquisition and different thermal cameras will be evaluated to choose the final model used in the experiments. Additionally, the designed motion models are presented in section III-C and preliminary experiments that were performed to decide the parameters and metrics for the model training are explained in the same section. The tracking algorithm and pre-processing algorithms are presented in section III-D. The conducted human subjects experiments are presented in detail in section IV followed by the steps taken during post-processing in section IV-D. Finally, the results are presented in section V and discussed in section VI followed by the conclusion in section VII. Limitations, improvements and future suggestions will also be explored in section VII

II. Literature

A. Surrogate Signal

1) Surrogate System Requirements: As briefly mentioned before in section I-D, our surrogate data must be MR safe as we are using MRI for the main imaging modality for our ground truth data. A few requirements are set in place for our surrogate signal to fulfil and so this will help in the selection of the signal type.

- The signal selected must work seamlessly with MRI without causing artefacts in the signal data or the ground truth data during image acquisition
- The signal selected must work safely inside the MR environment and not put the patient or the operator at risk
- The signal must have a high enough capture rate to capture changes in breathing as well as sudden changes like coughing and hyperventilating. Normal breathing rate is about 12 to 20 breaths per minute, while breathing rate under load (exercise) goes to 35-40 breaths per minute and it is said that above 40 breaths is considered hyperventilation [11]. This means that our signal must capture breathing at least 40 breaths per minute, which is 0.667Hz. This means that our capture rate must be greater than double the breathing rate, so at 1.3Hz to capture all breathing patterns [7]
- The selected signal must not expose the patient to health risks (such as radiation) or cause additional discomfort or pain during the operation

2) Surrogate Signal Comparisons: Some common candidates for surrogate signals used in RME and their issues are as follows:

- **MR navigators** are the most common surrogates for MRI modalities used for RME [7]. The main issue with MR navigators is that they increase acquisition time due to them essentially being extra RF pulses emitted during the main scan. Next to that is the issue of being totally MRI dependent and so cannot work unless MRI acquisition is the chosen imaging modality.
- **Spirometers** are also used along with **respiratory bellows** to measure the airflow in and out of the lungs [7], [12], [14]. The main difference between each is that spirometers measure the airflow directly in and out of the lungs while respiratory bellows are airbags that measure the airflow coming in and out of them when the throat or abdomen induces pressure on them. The main issue with both methods is that they cause a lot of discomfort to the patient, especially if it is a trauma victim (burns, sensitive skin, lacerations or abrasions) [12]. Additionally, they also face an issue with air leakage, causing a drift in the results, and they restrict the operating area due to their placement position, respectively [7].
- **External markers and RGBD cameras** are also common surrogate signals used for their high accuracy relative to their cost and MRI compatibility [7]. However, a main limitation is that the external markers are placed on the abdomen, near the liver, and so near the operation area of the surgery. RGB-D cameras also share a similar limitation where they rely on the visualisation of external abdominal changes. This could induce external noise while the surgeon is operating on the patient near the markers or the abdominal ROI, which would reduce the reliability of the estimated data. It could also cause signal losses or lost frames if the markers or ROI were blocked from the camera's field of view (FOV) while the surgeon is operating.
- **Infrared imaging** has been used before but not to determine liver displacement from temperature readings but rather to estimate breathing rate (BR) [12]–[14]. It has proven successful with high accuracy in estimating the

BR of patients at rest, under load [13] and with different breathing patterns and techniques [12]. No prior study, to the best of our knowledge, has tried to correlate it to liver displacement, except with the use of external markers rather than temperature readings, which presents the shortcomings mentioned above. However, with the correct fitting model, it seems plausible, given the high accuracy of BR estimation. Another point that helps IR is that it is MR compatible and does not affect image quality during acquisition for either the IR images or the MRI images [16].

For these reasons, we chose IR imaging as a surrogate signal for this study. According to our goals from subsection II-A1, our system has the following features:

- The signal is MR compatible with no artefacts created for either image acquisition process. IR imaging is also fast enough to capture at higher rates than MRI, giving it an advantage over the conventional modality.
- IR cameras can work within the MR environment, especially if shielding is implemented as an extra safety precaution. This can be evident from the fact that MRI rooms contain cameras inside them as well.
- IR cameras do not interact with the patient and are a non-contact spirometry method so they cannot induce extra pain or discomfort to the patient as well as not subject them to extra radiation.
- IR cameras can work with extreme case patients due to not requiring any belts, bellows or measuring instruments to be connected to patients that are too small, fragile or traumatised [12].
- IR cameras usually have frame rates ranging from 9Hz and up to 200Hz [17], [18]. Thus, this means that IR cameras are sufficiently fast to capture random and extreme BRs seeing that we need at least 1.3Hz

B. Motion model

1) Correspondence model: In this study, we are examining the displacement of the liver due to the RIM. Due to our choice of IR imaging as a surrogate signal (with our ROI being the nose and mouth areas), we need to explicitly address the correlation between our measured signal and the actual motion data. A few papers have discussed using BR to estimate liver displacement [4], [6], [7] using different regression techniques. This works due to the linear correlation between the lungs pushing on the rib cage and the diaphragm, during inhalation, which then pushes on the liver and vice versa during exhalation [7]. This linearity allows regression models to estimate the internal motion.

We can determine the cyclic displacement of the liver due to respiration with this knowledge. Our IR method presents a new research opportunity that has not been studied enough yet. This is namely the magnitude of temperature change captured and visualised by the IR camera. The visualisation is the observations of the thermal camera, that show the magnitude of temperature change due to breathing, i.e. the strength at which the subject is breathing. The nose/mouth area becomes hotter and colder during exhalation and inhalation, respectively. This can then be tied to the strength that the lungs push on the organs. Along with the ground truth data, this can be used to train an estimation model. The trained model can estimate the internal displacement of the liver. The intensity and hue of the thermal image correspond to the thermal signal (temperature) of the nose/mouth ROI, which is then translated using our fitting method into liver displacement.

2) Fitting method: Following our qualitative analysis in section II-B1, we decide to implement a novel fitting method to estimate our results in our methodology. This method utilises the raw image data and deep learning using convolutional neural networks (CNNs) to directly derive the displacement of the liver from the IR images of the ROI in patients. The filter and image processing tools available in the FLIR Research Studio (To be further discussed in section II-C), allow us to define our ROI and pre-process the video file into frames of appropriate resolution and size for input to the model.

C. IR Camera

To choose the most suitable camera for our purposes, we created a market search of the most common models used in other papers [12]–[14]. Of the cameras we found, three main makes (Teledyne FLIR, VarioCam and SEEK) were found to be sufficient for our testing and experimentation purposes. We then researched the different models that suited our needs. These models are discussed in Table I.

In [12], they were estimated breathing rate using IR imaging rather than with bellows and spirometers. The VarioCam 800 series camera was used and got very accurate results, claiming about 97.64% of the errors were smaller than one breath per minute. In [13], they recorded the breathing frequency (breaths per minute) as well as the temperature change frequency ($^{\circ}C$ per minute)of the ROI. They got mean squared errors (MSEs) in four experiments ranging from 0.055-0.508 (min^{-2}) and 0.001-0.003 $(^{\circ}C/min)^2$, respectively. They used the SEEK compact camera and proved the high accuracy of results. Finally, in [14], they measured the difference between breath volume (in litres) estimated using IR imaging and conventional contact spirometry using the FLIR ONE camera. They obtained errors as low as 0.0023 (L^2) MSE and a high accuracy of 0.9998 mean R-square value. These cameras, though very varied in their accuracies and their temporal and spatial resolutions, gave accurate results, based on the data analysis performed in their respective studies. This made us conclude that the main criteria of choice for our cameras would be based on the following:

- **Full-window resolution:** This would be the average spatial resolution of the camera needed to capture a bigger ROI to account for head movements so that temperature differences would not be too large when accidental motion occurs. A higher resolution would also provide a cleaner image with smoother features for training the model.
- **Maximum and average frame rates:** This is to test the faster capabilities of IR imaging under normal cir-

TABLE I: IR camera makes and models with their specs

Make	Model	Price (€)	Spatial resolution (pixel^2)	full-window frame rate (Hz)	max frame rate (Hz)	accuracy ($^\circ C$)
FLIR	A655sc	30,000	640×480	50	200 @ 640×120	$\pm 2\ ^\circ C$
	ONE Pro	450	160×120	$< 9 \approx 8.7$	-	$\pm 3\ ^\circ C$
VarioCam	HD head 800	33,000	1024×768	30	240 @ 1024×96	$\pm 1\ ^\circ C$
	HDx head S	12,900	640×480	30	60 @ 384×288	$\pm 2\ ^\circ C$
SEEK	Compact	235	200×150	$< 9 \approx 8.7$	-	$\pm 2\ ^\circ C$
	Compact PRO	470	320×240	> 15	-	$\pm 2\ ^\circ C$

cumstances. Faster frame rates can be used to capture faster breathing patterns and extreme instances such as coughing and gasping to test the limits of the camera usage in a real-life scenario.

- **Accuracy drop-off:** This is how accurately the camera is still able to capture minor changes in temperature at larger distances. The setup in the MRI room will require the camera to be farther from the subject due to the magnetic field of the MRI. Thus we require a camera than can still perform well at a larger distance from the subject. The setup will be further discussed in detail in section IV

Based on these findings, we chose the FLIR A655sc (Teledyne FLIR©, Oregon, United States of America) as the camera used for MRI testing. The high temporal rate and accuracy of $\pm 2\ ^\circ C$ would be valuable for NN pattern and feature training with MRI ground truth data. The accuracy drop-off at higher distances is also low for this camera so it can accurately capture the small temperature differences in the MRI setup. The non-ferromagnetic material used for the camera and the casing also allows it to be MR-safe.

As mentioned previously in section II-B2, we used the FLIR Research Studio program to process our images and perform data acquisition. The ease of use of the program as well as the powerful pre/post-processing options, made it easier to put the FLIR camera above the others.

III. Methodology

A. Motion Model: Analysis

It was outlined in [7], [15] that a motion model is made up of a few analytical steps. These are:

- **Choice of surrogate data**: The chosen type of surrogate signal used for this study is infrared (IR) as explained in section II-A2.
- **Choice of motion representation**: The physical motion represented in this study is the SI motion of the liver under the influence of respiration.
- **Motion model correspondence**: The relation between the actual data and the surrogate can be seen as a linear relation between the temperature magnitude and the strength of breathing to the displacement of the liver due to respiration. Furthermore, because the data is patient and session-specific (intra-fractional) a combination of these linear signals can be used to estimate the motion model [7]. Additionally, the use of a NN should also follow the same ideology where the input thermal images should have a corresponding liver displacement. The NN is thought to overcome intra-fractional errors as well due to the ability of the model to learn multiple breathing patterns and estimate the results on different test sets.
- **Fitting method**: Since the correspondence of the motion data and the surrogate data is linear, a CNN is used to determine the estimated displacement depending on the temperature in the ROI. The linearity of the data allows for easy pattern recognition by the CNN which should allow for higher precision and repeatability with different sessions and subjects.

B. Surrogate signal: IR imaging

The input for our model is IR images. They required some pre-processing before being used. The FLIR Research Studio was used for thermal image pre-processing. It allowed us to create and select specific ROI(s) and get the mean temperature measurements within them. Another feature was the ability to post-process the images after the acquisition, straight from inside the program, so that they can be prepared for NN training. The ROI overlay was removed and the image was cropped to only show the ROI. The programmed processed this simultaneously for all recorded frames of the video during acquisition which made the process of testing and tuning the NN using the camera data with different parameters much more efficient.

Another feature was the camera override parameters tool which allowed us to set the emissivity/reflectivity of the objects as well as the ambient temperature and the distance from the object. This was very useful as we were trying to capture the temperature of the nose/mouth ROI off a mirror in the MRI machine. The setup will be further discussed later in the experimentation section.

After the specific orientation and transformations were applied to the images, they were exported as individual frames to be fed into the CNN as ground truth thermal input for training, validation and testing.

C. Correspondence model: Convolutional Neural Network

As previously discussed in section II-B2, the use of a CNN has been chosen for this study. Using a CNN, we can use the thermal images from the IR camera as a direct input to the model and take advantage of the direct relationship between breath temperature and liver displacement. This would also allow for more versatility to the given breathing pattern inputs compared to conventional regression methods.

The premise that the model will be able to adapt to different breathing patterns means that a single model can be trained

and used for multiple test datasets. In other words, one model might be enough to estimate the liver displacement for multiple test subjects and sessions, completely removing the problem of intra-fractional error.

After we attain the frames and choose the appropriate image filter to use, normally grayscale (single-channel image), we must then define our CNN architecture as well as the parameters and metrics used for training. Firstly, we define our problem as a regression rather than a classification. This is due to having a range of observations (displacements in our case) correlating to the range of breathing frames (magnitudes) rather than a set number of labels to be assigned to our frames. For this reason, we use mean squared error, which is the most common method for this type of problem [21], [23], as our loss function for training and testing. We also utilise the AdamW optimiser to allow for adaptive model weights, momentum optimisation and weight decay regularisation [23]. This is used to mitigate the chance of overfitting the model to our training data.

We also use two model architectures of different complexities to test how model complexity affects the predictions. A simpler model with only two convolutional and max pooling layers and one dropout layer is used for the simpler model while another with four convolutional and max pooling layers and two dropout layers is used for the more complex architecture. Both architectures are shown below in tables II and III.

TABLE II: Simple Model

Layer Type	Output Shape	Parameters
Input	(458, 614, 1)	
Convolutional	(229, 307, 16)	160
Activation(ReLu)	(229, 307, 16)	
Max Pooling	(114, 153, 16)	
Convolutional	(114, 153, 32)	4,640
Activation(ReLu)	(114, 153, 32)	
Max Pooling	(57, 76, 32)	
Flatten	(138624)	
Fully Connected	(512)	70,976,000
Activation(ReLu)	(512)	
Dropout	(512)	
Fully Connected	(1)	513
Output	(1)	

The data is split into 80% training, 10% validation and 10% testing, where the training data is used for fitting the models, the validation data is used to make sure the model parameters are correctly set during the training and the testing data is used to evaluate the performance of the trained models on new unseen data.

D. Ground truth: liver edge tracking

To extract the ground truth displacements, we utilise a series of MRI scans of the liver and perform a motion detection process to track the displacement throughout the series. The MRI experiment and details about the parameters will be explained further in this paper's experimentation section (section IV).

TABLE III: Complex Model

Layer Type	Output Shape	Parameters
Input	(458, 614, 1)	
Convolutional	(458, 614, 32)	320
Max Pooling	(229, 307, 32)	
Convolutional	(229, 307, 64)	18,496
Max Pooling	(114, 153, 64)	
Convolutional	(114, 153, 128)	73,856
Max Pooling	(57, 76, 128)	
Convolutional	(57, 76, 256)	295,168
Max Pooling	(28, 38, 256)	
Flatten	(272384)	
Fully Connected	(512)	139,461,120
Dropout	(512)	
Fully Connected	(128)	65,664
Dropout	(128)	
Fully Connected	(1)	129
Output	(1)	

1) Edge detection pre-processing: The MRI series contains grayscale, coronal images of the abdomen with the liver being in the centre of the inspection window. Some pre-processing is applied to the image series to be able to start the tracking of the liver dome. The set of processes applied are as follows, in this order:

1) Gaussian filter
2) Contrast filter
3) Morphological dilation
4) Canny edge detection

In the case of line detection, it is better to have a blurred Gaussian image to start with so that there are fewer sharp features that would create noise artefacts. Thus, the Gaussian filter used had a 9×9 $pixel^2$ kernel size and a standard deviation in X and Y of 5 units. This ensured a smooth image, ready for contrast filtration, to only show the borders between the liver and surrounding organs without the added internal details of the organs. A contrast filter is applied to reduce the contrast of the surrounding organs/tissue details around the liver so that the liver is the main clear subject in the images. Thirdly, we apply a morphological transformation on the image, namely dilation, to fill the holes left by the aggressive contrast filter. This returned a filled, smooth, high-contrast image of the liver inside the abdomen. Finally, Canny edge detection is applied to the resulting image to create a binary image with the edges of the liver dome clearly visible and easy to track.

2) Motion tracking: The motion detection of the edges uses a simple "white point tracking" condition. Other methods that employ the use of OpenCV's contour function (which detects full unbroken lines or filled elements in an image) were attempted. The accuracy however, was not as high as preferred due to the liver dome being more of a "broken line" rather than a straight line. Additionally, the liver image was

not a filled object in our "Canny edge image" but rather an open contour connecting to the rest of the organs around it, with just the dome being the clean and free edge required for tracking. This also meant that using a "centroid detection" approach [24] was not viable due to the "blob" not being a closed contour as required and the liver also undergoing deformations next to the SI motion which made the RIM tracking unreliable.

This presented a simple approach that performed well enough to track the cyclic breathing motion of the liver dome. This approach was a white point tracker, where we have a binary image with a black background and a white edge (the liver dome) comprised of white pixels moving. Thus, the simple approach was to track the white pixel on the liver dome in a specified window, as can be seen in Fig. 3. This gave us a smooth motion tracking of the pixel's displacement in the breathing range of the subjects. For increased accuracy, there was a grace/tolerance window of four pixels in the x-axis direction so that if the tracked pixel was lost, another pixel from the same line would be tracked, to ensure smooth and precise tracking.

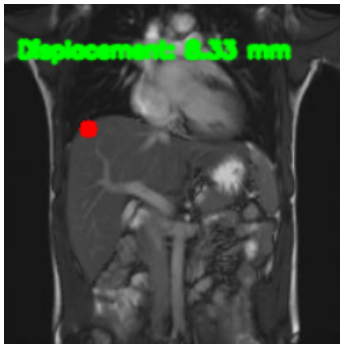


Fig. 3: Red point is the tracked pixel. Displacement indicator written in green as a title. The tracking was done in the background on the binary edge image but the original MRI video was left as an overlay for easier visualisation

The displacement of the points was calculated using a "frame subtraction" approach where the displacement of the pixel would be compared to the previous frame pixel location. A simple equation was used for this calculation:

$$Y = -x_{current} + x_{previous} \tag{1}$$

Where Y is the displacement of the tracked pixel, $x_{current}$ is the current SI position of the pixel and $x_{previous}$ is the previous SI position. This equation allows for the displacement calculation of the liver dome with the relative direction of "up/superior" being positive and "down/inferior" being negative. Finally, to get the correct displacements in mm rather than pixels, the pixel displacement values were multiplied by 1.667 which was the pixel-to-mm conversion factor of the MRI sequence used. The total displacement of the tracked point is then the sum of the individual displacement changes through the frames and plotted over the video duration as seen in Fig. 4.

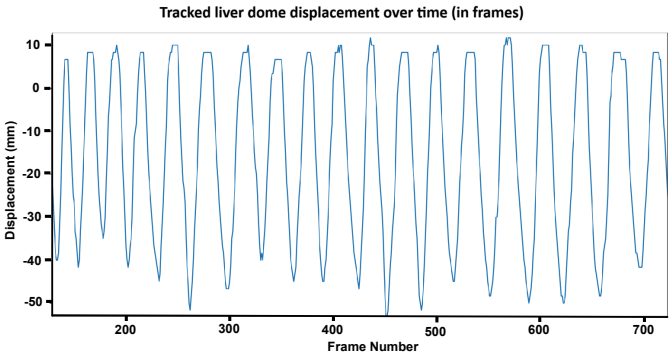


Fig. 4: Displacement of liver dome edge plotted over the duration of the video to show RMI of the liver dome

IV. Validation

A. Overview

As previously mentioned in the above sections, a RME approach was proposed with the surrogate signal being the Infrared signal, the ground truth being the MRI image series and the motion model being the CNN as a novel method we implement. In this section, we will discuss how we validated these proposals with experiments on human test subjects. Fig. 5 shows the experimental setup with the MRI and thermal camera in place. The subject lies down inside the MRI scanner with a mirror at a $45°$ angle above their head facing away from their abdomen. This was required to train the models offline on

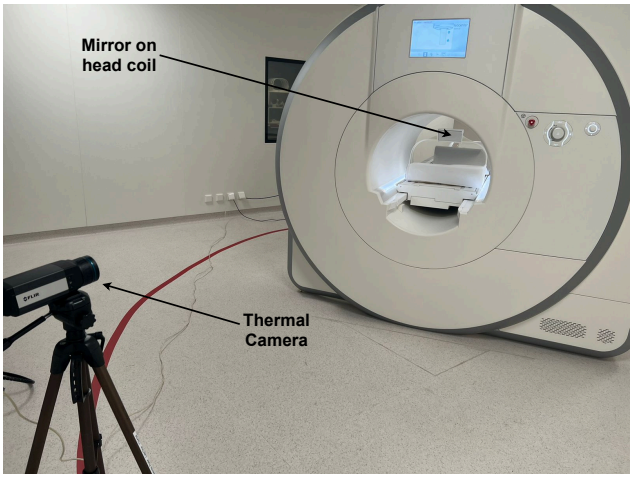


Fig. 5: Thermal camera set behind the MRI machine and the subject. The camera is aimed at a mirror fixed on the head-coil which allows the heat signal of the ROI to bounce back into the camera while the subject is lying down.

the ground truth results supplied by the MRI machine for the liver displacement. However, once the model is fully trained, this approach should permit the subject to be only lying in front of a thermal camera and based on the breathing exhibited, the liver displacement would be predicted by our models.

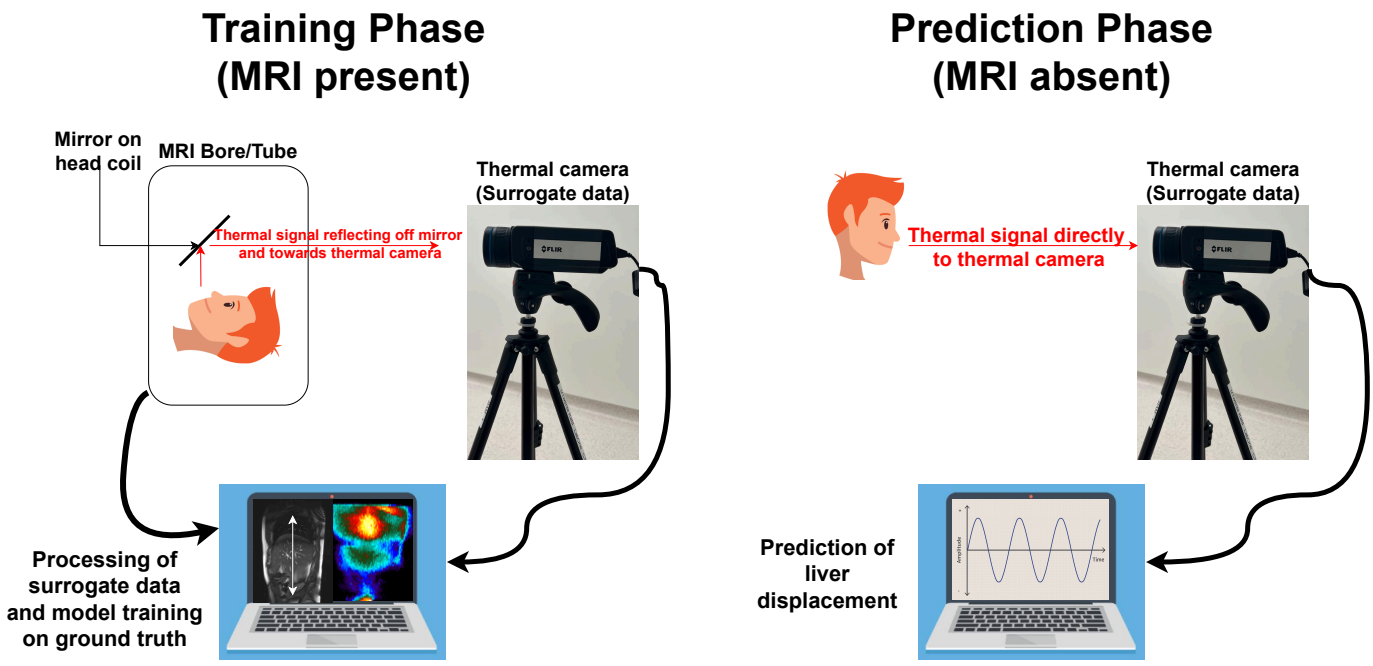


Fig. 6: Overview of the experimentation setup with the training phase (left) using an MRI as ground truth while the subject lies in a supine position inside the bore-tunnel and the camera is placed behind them along the sagittal plane. The setup is then changed during the prediction phase (right) to work directly with the subject and camera without the use of an MRI machine. The thermal camera captures the temperature change in the nose/mouth area and the model predicts the liver displacement directly using the input thermal images.

B. Workflow

As shown in Fig. 6, the data acquired from the experiments were the ground truth MRI displacements, surrogate Infrared images and the temperatures. These were then split into three sets of training, testing and validation with the splits being 80%, 10% and 10% respectively. The thermal datasets ranged in size between the subjects from 17k to 25k data points per session for three sessions per subject while the MRI dataset was about 2000 frames. The training and validation splits were used to train the model on the motion patterns while the test split was fed to the trained model to predict motion data from the thermal images.

C. Human Subject Experiments

1) Measurement Protocol: Three healthy male subjects, ages between 21 and 27, participated in the experiments. Each of the subjects was timely informed of the study goals, received a detailed experimentation protocol and filled in a consent form as well as an MRI safety checklist to ensure that there were no risks to them during the experiments and familiarise them with MRI safety regulations.
Each subject participated in three sessions of 11 minutes of scan-time each. The total investment time per subject was about 60 minutes including the three scans, the pre-scan debriefing and the MRI safety instruction. The experiments were conducted in the University of Twente (Enschede, Netherlands) and an ethical approval for the experiments was received from the NES ethical committee of the university. During the scans, each subject was asked to wear a disposable medical face mask and breathe normally while looking in the mirror during the acquisition. The use of the face mask was for the ease of thermal data acquisition. The heat would disperse on the mask more evenly (bigger surface area) than the skin on the mouth and nose area, allowing for a bigger view of the ROI to capture on the IR camera.

2) Experimental setup: The experimental setup consisted of two main modalities, namely MRI and thermal imaging. The MRI system used was the Siemens 1.5T MRI scanner (Magnetom Aera, Siemens Healthineers, Erlangen, Germany) shown in Fig. 5. This system is available at the University of Twente. As explained in Fig. 6, the subject was placed in a supine position with their face secured in a head-coil inside the tunnel bore. Additionally, the subjects would lie normally with their arms inside the machine and lying by their sides. The machine bed was then moved so that the scanner was positioned with the liver in clear view for acquisition of the full range of motion. Once the bed was in position, the subject was then asked to move their head into position inside the coil so that the mask ROI was perfectly centred in the mirror. This can be seen in Fig. 7.
The thermal camera used to capture the subject's ROI, as previously mentioned in section II-C, was a FLIR A655sc, shown in Fig. 5, from Teledyne FLIR© (Oregon, United States of America) with a 25° field of view (FOV) and 1.0 mm

Fig. 7: Subject lying inside MRI tunnel bore with the mirror attached to the head coil. The nose/mouth region of interest can be seen in the mirror. The subject would be wearing a mask and the thermal data would be picked up by the camera.

focal length, 24.6 mm-lens. The camera was placed behind the subject's head (facing the subject's transverse plane), outside the MRI's immediate magnetic field, approximately 3 m away from the mirror. This had to be done as a radius of $\pm$ 4 m from the centre of the MRI machine is marked as a boundary for safety, which can be seen in Fig. 5 as the red line, where no ferromagnetic objects are allowed inside, to avoid safety risks for the operators and the machine. To avoid potential damage to the camera, even though it was mostly aluminium and plastic, we adhered to the 3 m distance from the head coil.

3) Data acquisition: The MRI images are acquired using MRI sequence is a 'TRUFI interactive realtime' in the coronal plane, slice thickness = 10 mm, repetition time = 355.6 ms, echo time 1.17 ms, reconstructed resolution = 1.88×1.88 mm^2, field of view = 360×360 mm^2, temporal resolution approximately 0.4 fps. The thermal camera was set to record the mask ROI at 50 fps with a camera window of 640×480 mm^2. Temporal synchronisation and alignment were done using two breath-holds and feature matching.

D. Post-processing

1) Image processing: After the MRI and thermal acquisition, the MRI frames were segmented. As previously mentioned in section III-D, the MRI DICOM files were processed by implementing a series of morphological transformations and filters to isolate and track the upper liver dome. During preliminary testing, the AP motion was shown to be very small (around 2 mm [8]) and hard to track given our tracking algorithm and lower MRI spatial resolution used during the experiments. For these reasons, we decided to only focus on the SI motion of the liver. Thus, the final experimental frames were in the coronal plane to get a clear view of the liver dome and track away from the disturbances caused by the heart beating.

For the thermal images, the FLIR Research Studio program was used to apply a high-contrast filter on the images to increase the features (intensity due to temperature change) seen by the NN. The program also allowed for the exporting of the temperature values detected in the thermal images, within the mask ROI, as a CSV file without the need for further tracking or processing.

2) Signal synchronisation: Finally, after every MRI frame was transformed into liver dome displacement values in mm and thermal images into temperature values in $^{\circ}C$, the outliers were manually removed from the datasets. The MRI data was 2000 frames per session so it was interpolated to match the length of the thermal images. Finally, the breath-holds and unique patterns in the data were matched together to allow for correct time synchronisation. Additionally, a low-pass Gaussian filter was applied to remove the noise from the thermal images and allow only the cyclic breathing values to be visible.

V. Results

This section presents the results yielded for our RME approach with IR imaging as a surrogate signal. We conducted a few preliminary experiments to test our system's limits and choose the best model architecture and parameters for training. This will be presented in section V-A. We then show the final choice of the model and parameters in section V-B and the reason behind the choices made. Finally, the comparison between the estimated SI motion and ground truth MRI SI motion is presented in section V-C and their discussion in section VI.

A. Model and parameter choice

To choose the most viable model/parameters and test the feasibility of IR as a surrogate with CNNs, we conducted some preliminary tests to benchmark the system. Firstly, we had the camera set up without an MRI and directly facing the ROI to record the temperature changes and their corresponding frames. The conducted experiment was to see how accurately the CNN can estimate the temperature values based on the IR frames and in a varied breathing pattern.

One subject was used for this experiment and they were sitting directly in front of the IR camera and were breathing for 3 minutes at different rates while wearing a mask. Four types of breathing were done, each for 45 seconds. Firstly, normal breathing through the nose was done for the first time interval, followed by fast and shallow breathing through the nose for the following 45 seconds. The last minute and a half was the same pattern but through the mouth, namely 45 seconds of normal breathing through the mouth followed by 45 seconds of fast and shallow breathing through the mouth. The main reason behind this experiment was to extract a trend for the

parameters required for the real experiments and to test the limits of our architecture's prediction abilities.

The experiment showed us which model to use according to the estimation ability of the model on an unseen test set. In Fig. 9, we show the predictions of the models vs ground truth data (shown in blue), in a staggered view for clarity. After a few tests, we decided to continue with the simpler model because the more complex architecture was more prone to overfitting and getting stuck in local maxima/minima. The simpler model was better equipped to predict the different breathing patterns because it was less prone to overfitting due to less layers having an effect on the output [20]. We experimented with various loss functions, number of epochs and data shuffles. Due to this being a regressive problem, MSE was the best and most common loss function to use [21], [22] and that was also proven by Fig. 8a and further clarified in the isolated graph in Fig. 8b. The ground truth data is the blue data and the purple data is the fully trained MSE model, which performed better than the fully trained Mean Absolute Error (MAE) model (green).

The final model we chose was of the simple architecture shown in Table II. MSE was selected as a loss function and the optimizer used was AdamW, which reduces the complications of overfitting due to it offering adaptive learning rates, momentum optimization and weight decay regularisation [23]. Finally, The model was trained with a checkpointing callback and early stopping was implemented on the validation set. Fig. 8 shows the preliminary model estimating the temperature values and patterns of the input data with different breathing patterns. The MAE and MSE were both 0.999 $^{\circ}C$ and 1.01 $^{\circ}C^2$, respectively.

B. Final model training

The results from the preliminary test, mentioned in section V-A, showed that the system was feasible. Thus, we trained a similar model of the same architecture to one of the subject's sessions. Fig. 10 shows the scatter plot of the predictions vs. ground truth (Fig. 10a) and the validation loss vs. training loss of the model during training (Fig. 10b). The scatter plot was used as an indicator to see if the model was successfully improving its regressive abilities. Initially, we attempted to retune the previously trained temperature CNN, from section V-A, on the new displacement labels instead of completely retraining it, however, the results were not desirable. Thus, a new model was trained from scratch for one session and fine-tuned for the rest of the sessions instead. The fine-tuning involved keeping the model weights as they were and only retraining the last four layers of the model to change the displacement labels to match the session's specific displacements. Because of the high errors and the models not converging to a desirable loss value, we concluded that one single model cannot work on multiple subjects and cannot work on multiple sessions of the same subject (due to intrafractional error) as not only did different subjects have different breathing patterns and displacements, but also a single subject would breathe differently from one session to the other. When we applied the retuning strategy, model performance increased and predictions were more in-line with ground truth.

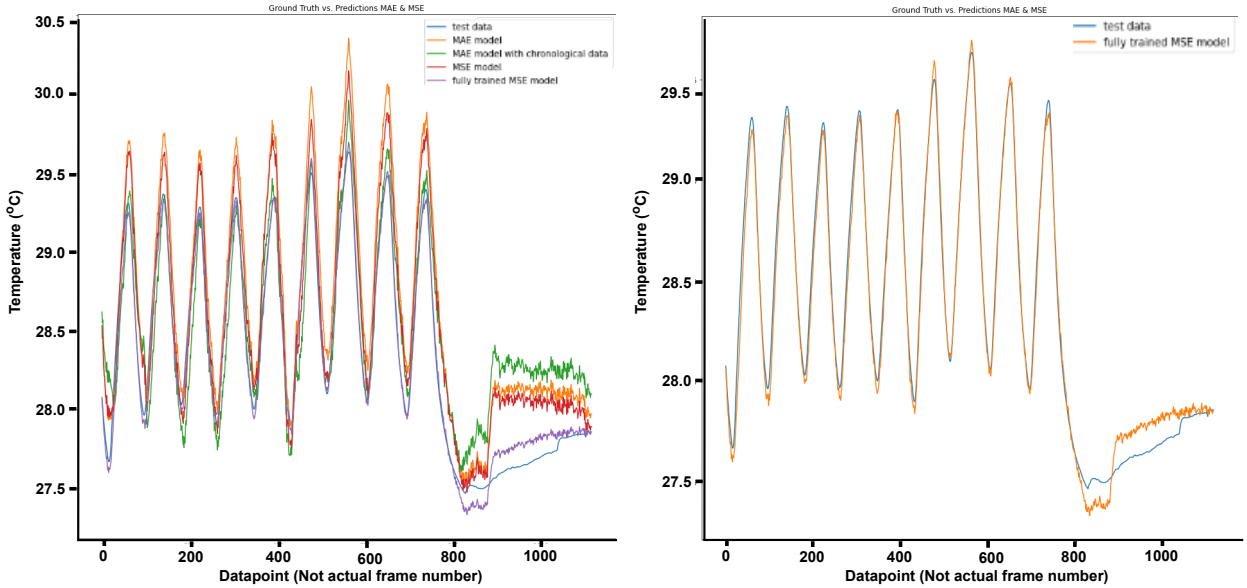
C. MRI ground truth vs IR imaging estimation

In this section, the results of the trained models on three subjects are shown. This includes three sessions of unseen data points from the 10% test split. Fig. 11 shows the model's performance on the test dataset for sessions 1-3 for every subject. The predictions were noisy because the IR image inputs were noisy themselves. Thus, a Gaussian filter was applied to the final predictions to allow only the low breathing frequency to pass and cut off the higher noise frequencies.

Data analysis was performed on the test results to see the quantitative performance of the model. Table IV shows the mean values μ and standard deviations (std) σ of all three sessions for Subjects 1, 2 and 3. The MAE, MSE and R^2 scores were also calculated and shown in the table. The R^2 score is a measure of the accuracy of the regressive estimation of the model is, relative to a baseline.

Subject 1 was a rather special case due to their different breathing method. Subject 1 exhibited diaphragmatic breathing [25], which is breathing by drawing in air through the stomach while the chest remains still. It is mainly exhibited in athletes [26] as it usually allows for the lungs to intake more air as the subject actively pulls the diaphragm downwards to fill up their lungs. As a result, the liver does not move as much as normal breathing but rather is deformed in place. Very little SI motion is exhibited however, the results were used as a case study to benchmark the model. We wanted to validate the model's accuracy on smaller displacements and jitters. As seen in Fig. 11 and Table IV, the CNN can predict the liver displacement values with sufficient accuracy (R^2 score higher than 0.75). Subject 1 had a lower accuracy due to the smaller displacements and the occasional deep breaths that they would take, which caused inconsistencies in the mean calculations. Subject 2 had a few limitations faced during the sessions, which will be discussed in section VI along with a breakdown of the data analysis.

Finally, Fig. 12 shows the box plot for the means and standard deviations ($\mu \pm \sigma$) for the Peak-To-Trough (PTT) displacements for all subjects' sessions. The blue data is the ground truth while the orange data is the predicted data. Fig. 12 shows that the model predictions are smaller than the ground truth, which is expected due to the Gaussian filter attenuation and the randomness factor of the test data compared to the training data. The figure shows that subject 3 exhibited deeper breathing relative to other subjects with the highest mean PTT displacement being their third session, at 45.2 mm. The results' variations allow us to test the performance of the model on multiple scenarios ranging from shallow breathing (subject 1's diaphragmatic breathing) to normal breathing (subject 2's sessions 1 and 2) and deep breathing (subject 3's sessions 1 and 3).



(a) Testing the CNN with different MSE and MAE configurations vs ground truth (blue)

(b) Isolated MSE graph to show model performance on the unseen ground truth test split

Fig. 8: Comparison of different loss functions and training configurations

TABLE IV: Motion Data Analysis: MSE, MAE, R^2 scores and Mean μ (mm) and standard deviation σ (mm) of the Liver SI motion for Peak-To-Trough (PTT) displacement both for MRI (Ground Truth) and IR surrogate (Predictions) for three sessions for subjects 1,2 and 3

	Session	MSE(mm^2)	MAE(mm)	R^2 score	$\mu \pm \sigma$ (Ground Truth PTT in mm)	$\mu \pm \sigma$ (Predictions PTT in mm)
Subject 1	1	2.46	1.22 ± 0.987	0.763	7.01 ± 3.26	4.98 ± 2.58
	2	5.51	1.75 ± 1.57	0.877	11.8 ± 6.27	9.30 ± 4.97
	3	5.13	1.24 ± 1.89	0.903	7.64 ± 6.68	5.78 ± 5.01
Subject 2	1	4.69	1.82 ± 1.38	0.856	15.0 ± 4.69	10.0 ± 3.80
	2	20.9	3.75 ± 3.00	0.745	24.5 ± 5.17	13.9 ± 4.54
	3	4.64	1.67 ± 1.76	0.969	31.0 ± 9.93	26.2 ± 10.1
Subject 3	1	9.44	2.59 ± 1.65	0.941	33.0 ± 4.63	26.7 ± 4.67
	2	1.52	0.906 ± 0.835	0.983	14.3 ± 10.4	13.0 ± 9.73
	3	9.51	2.57 ± 1.70	0.968	45.2 ± 7.99	39.2 ± 7.35
Overall Performance		7.09	1.94 ± 1.64	0.889	21.0 ± 6.56	16.6 ± 5.86

VI. Discussion

Looking at the previous section, we can see that the model performed better on larger breathing magnitudes, as in the third session for subjects 2 and 3. This can be seen in Table IV where the R^2 score is closer to 1, at 0.969 and 0.968 respectively. Subject 2's second session has a relatively lower accuracy of 0.745 even though the graph in Fig. 11 shows good tracking of the breathing pattern. However, due to the filtration added, the displacement magnitudes were attenuated. The small sacrifice of the full magnitudes was required as the data acquired was very noisy and only the low frequency of breathing was needed. Additionally, the setup's limitations caused a lot of trending and moving average errors in the thermal input data so it was hard for the model to learn features and patterns reliably.

Due to the way the experiment was set up, the mirror on the head coil would heat up due to the subject's radiated breath throughout the experiment. This would cause the camera to pick up the heat around the ROI and the gradual increase in the mirror's temperature, which caused a moving average error in most datasets. Furthermore, the subjects started breathing more aggressively halfway through the session due to the mirror fogging up from their breaths and them trying to actively cool it down which caused temperature fluctuations, specifically in the second session of the second subject. Thus, this caused the data to have a nonlinear trend in that session and the model had issues learning from the thermal images. Through these issues, however, the model was still able to strongly predict the pattern of breathing and estimate the displacement of the liver to an MAE of 3.75 mm. Overall, the second subject's second session and the third subject's first session had the highest trending in their data and it proved slightly harder for the model to learn from them. However, with some rigorous post-processing, the input data was still useful enough for the CNN to follow the general trend and predict the SI values with sufficient R^2 scores.

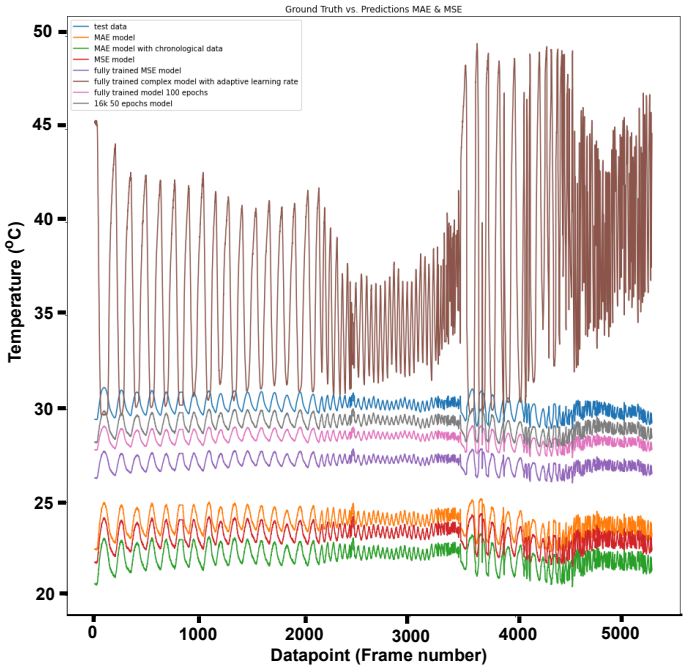
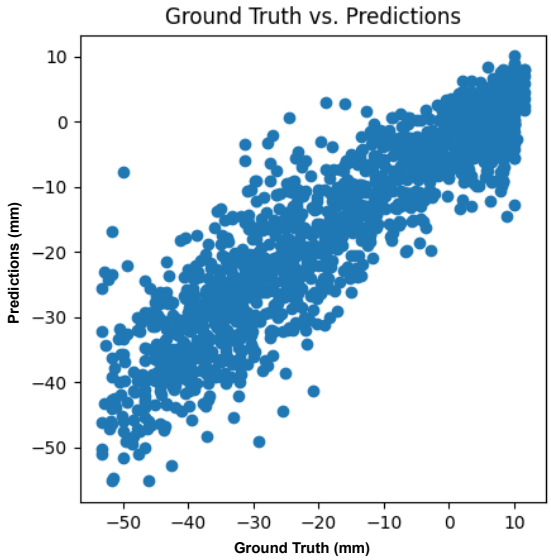
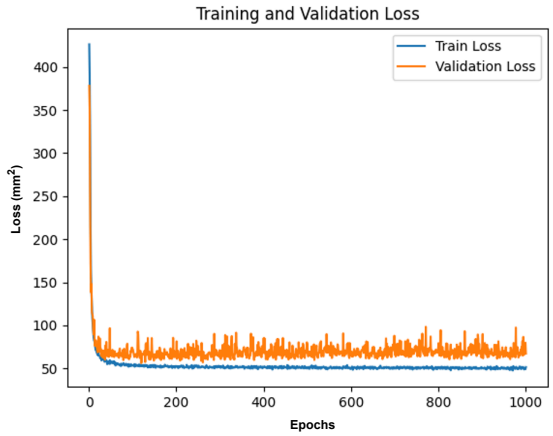


Fig. 9: Testing different loss functions, parameters, optimisers and model architectures in preliminary experiments to outline an optimal combination for the final testing. Temperature predictions in $^{\circ}C$ are made for every thermal frame input to the model.



(a) Scatter plot showing the trained model's predictions vs ground truth values on a subject's session test split



(b) Line plot showing the convergence of the validation and training losses for one subject session

Fig. 10: The model's performance during training on one session for one of the subjects

The results also showed that all subjects were breathing normally with means ranging from 7.01-24.5 mm. On the other hand, subject 2 was deep breathing during session 3 and subject 3 in sessions 1 and 3. That can be seen by the mean liver SI displacement of 31.0 mm, 33.0 mm and 45.2 mm, respectively. These values are all consistent with previous studies on liver displacement [7]–[9], [27], [28].

With a higher resolution camera or better camera lens, the input IR images would have a higher resolution and less noise. Shielding on the camera would also allow the camera to be placed closer to the mirror and less signal dissipation and accuracy drop-off. Finally, a cooling system, like a fan, to keep the mirror's temperature constant throughout the experiment, would reduce the trending and moving averages in the input IR data.

The case of subject 1 was still well-predicted by the CNN and even though the breathing pattern is not as cyclic as a normal chest-breathing subject, the CNN was still able to follow the breathing pattern based on the input thermal frames and predict the SI motion exhibited on the liver. The MAE values show that the average error was less than 2 mm and the R^2 scores show that the model had a high regression correlation for sessions 2 and 3 and a fair correlation for session 1.

Overall, the results from Table IV prove that our system works well to estimate the liver's SI motion using IR as a surrogate. The setup allows for the IR camera to capture the thermal ROI away from the abdomen, to allow freedom to the surgeon on the abdomen and be independent of surgical-induced noise. Furthermore, using the power of deep learning, the direct correlation between thermal images and liver RIM can be utilised in a CNN to get the predictions of the liver displacement. In a real-life application, The system could be even more accurate due to the sole use of the thermal camera without the MRI in the testing phase. This means that no mirror is needed to transfer the heat signal to the camera and that the direct heat from the ROI can be picked up with no moving averages or external noise. Furthermore, in clinical practice, an added value is the non-contact measurement allowed by the thermal camera. This means that the system would work with patients of all builds, cases and needs, ranging from trauma patients to infants to claustrophobic patients who cannot stay for long times in the MRI. A real-time version can also be implemented

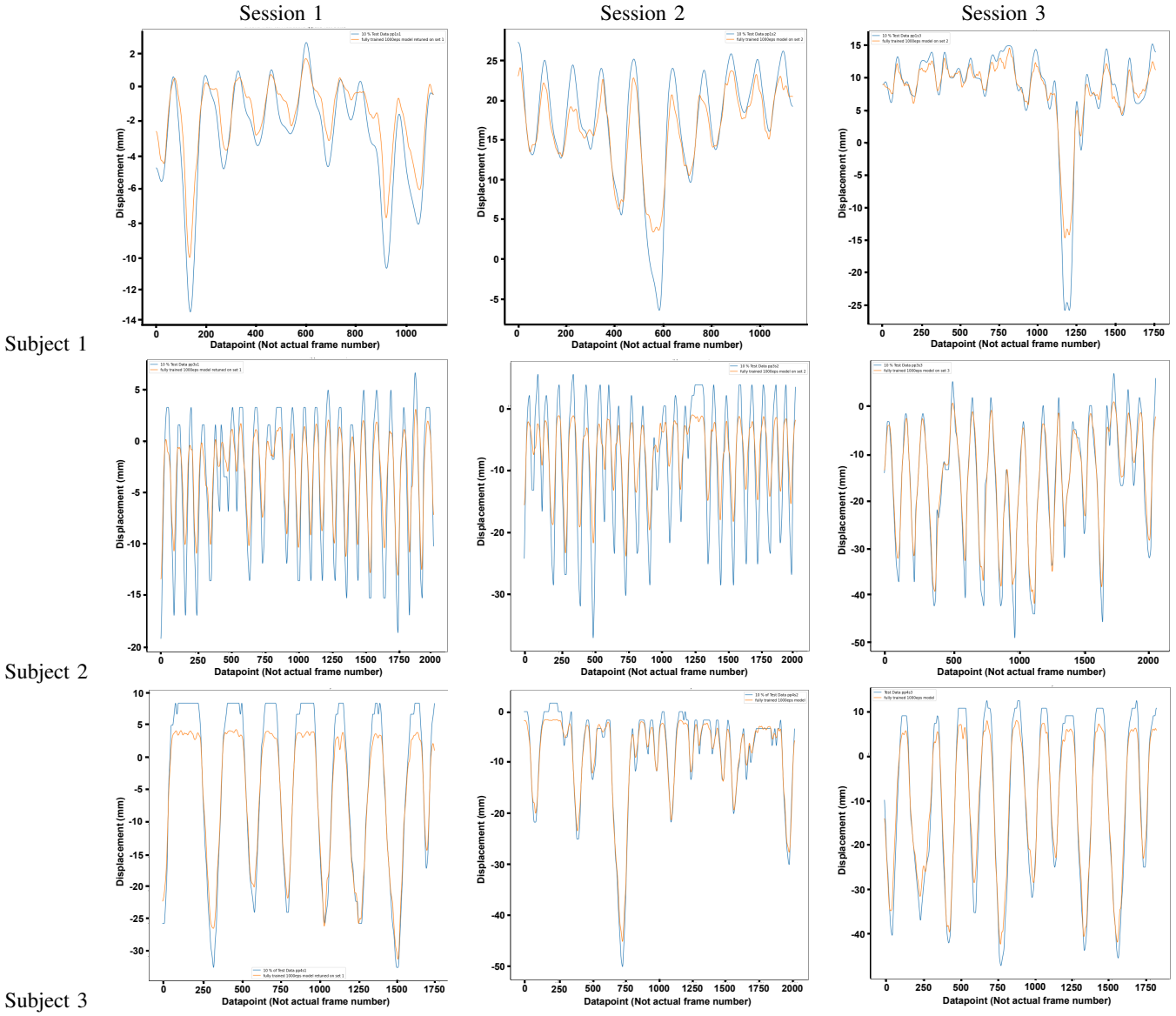
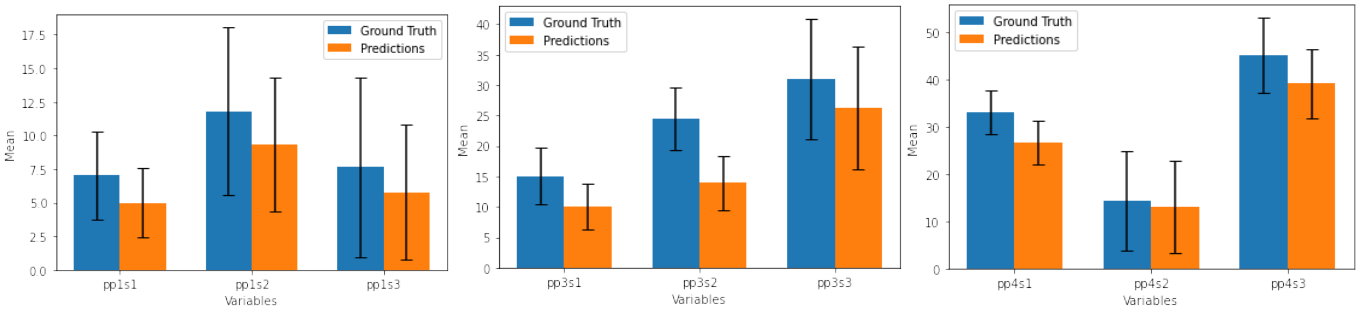


Fig. 11: The model's performance on subjects 1,2 and 3's test sets for each session. A Gaussian filter was applied to these results to show the model's prediction ability clearly and only allow the breathing frequencies to be visible.

so that a stronger AI can learn the thermal values as they are recorded and simulate the liver motion based on them.

This can also be a powerful tool in collaboration with augmented reality (AR), where the operator can see the liver in realtime, moving on the patient's body, for a clearer visualisation of the liver position and motion. Additionally, with the use of virtual reality (VR), the operator can even control a robotic arm, far away from the patient, and still be able to see the real-time visualisation of the liver motion. The compact size of the IR camera and the simplicity of thermal waves makes it the perfect surrogate for most environments, as no external radiation or electromagnetic waves can affect the thermal values and there is no need for a large space to house the thermal camera while in use.

Finally, a higher resolution MRI for the ground truth data would allow for 3D motion detection of the liver (SI-AP-LR) as more data can be extracted from the smaller motions. Additionally, the use of mode-A (Amplitude) ultrasound (US), in combination with MRI, would allow for a smoother ground truth training sample because of the high temporal resolution of the US. Mode-A ultrasound does not visualise the organs as mode-B (Brightness) US does, however, it has a higher accuracy when measuring internal organ dimensions [29]. A smoother ground truth means more data points within a breathing cycle, which also means more data to train the model from which could significantly reduce the acquisition time for patients as more data points can be acquired in fewer breathing cycles. The use of US could also allow the exploration of extreme breathing patterns such as coughs, gasps and hyperventilation, which usually have a higher frequency than



(a) Box plot of mean PTT displacements for subject 1 for all 3 sessions

(b) Box plot of mean PTT displacements for subject 2 for all 3 sessions

(c) Box plot of mean PTT displacements for subject 3 for all 3 sessions

Fig. 12: Box plots of the means μ and standard deviations σ (in mm) for the PTT displacements (ground truth in blue and predictions in orange). The data is further shown and analysed in Table IV

conventional MRI machines.

VII. Conclusion

In this paper, we developed and evaluated a novel RME approach to estimate the liver's SI RIM. An infrared camera was used to measure and record the temperature of the nose/mouth area as a ROI. IR was chosen as a surrogate due to the high correlation between breathing temperature and RIM on the liver, the MRI compatibility and the camera's high spatial and temporal resolution. The thermal images of the ROI temperature were used as an input to the fitting method along with MRI liver dome displacement values as ground truth training values. The fitting model was a CNN with a regression-like loss function and optimisers. Multiple loss functions and parameters were tested in preliminary exams to find the optimal combination and evaluate their accuracy and performance. The model was validated on three healthy human subjects, each participating in three sessions to acquire displacement and thermal ground truth data. The MRI and thermal images acquired were segmented and processed and then split into training, validation and testing segments. The trained model was further evaluated on the test segment of the split data for every subject session. The results showed that the model predicted the liver SI displacement values with good accuracy (overall MAE less than 2 mm and R^2 score of about 0.89, including the high error session of subject 2 session 2). Multiple factors affected the results of this paper. An MRI machine was used for the ground truth liver displacement values. The sequence used for acquiring the data had a low-resolution set, which made it harder to measure the AP and LR motion of the liver, on account of their small motions. A higher resolution would have allowed for 3D liver motion evaluation. Additionally, sharper displacement values would have been obtained allowing us to get smoother results from the tracking algorithm we used. Clearer MR images would have also allowed us to use a more accurate segmentation process, like the blob or centroid method, for more precise liver tracking rather than the simple "white pixel tracking" method used. Furthermore, the experiments were conducted on three healthy male subjects at the University of Twente. A future improvement to test the full potential of our approach is to use unhealthy patients as well as not only male subjects. This would allow us to track a specific region in the liver and benchmark the surrogate's ability to predict said ROI's displacement. Finally, temporal alignment was done manually by matching features in the data graphs and breath holds due to the MRI machine having a different time-stamp relative to the IR camera system. If a trigger is used for both systems to start and stop acquiring simultaneously, it would improve estimation accuracy and significantly reduce post-processing errors.

To sum up, this paper's objectives were to validate the feasibility and accuracy of IR as a surrogate in an MR environment, develop a neural network model to directly relate thermal images to liver motion and validate the approach on human subjects by conducting experiments. The authors fulfilled these goals by implementing a CNN-based model that utilised IR images and successfully predicted liver displacement values. In fact, to the best of the author's knowledge, no other study presented the use of CNNs to predict liver motion with the use of IR thermal imaging as a surrogate. This paper, to the best of the author's knowledge, presented a novel method for liver tracking and a novel surrogate signal for RME and proved that it can work as a viable option.

References

[1] Columbia Surgery - Center for Liver Disease and Transplantation. Accessed 2023 Sep https://columbiasurgery.org/liver/liver-and-its-functions#:~:text=The%20liver%20is%20the%20largest,hundreds%20of%20other%20vital%20functions.

[2] Li Q, Cao M, Lei L, Yang F, Li H, Yan X, He S, Zhang S, Teng Y, Xia C, Chen W. "Burden of liver cancer: From epidemiology to prevention." *Chin J Cancer Res.* 2022 Dec 30;34(6):554-566. doi: 10.21147/j.issn.1000-9604.2022.06.02. PMID: 36714347; PMCID: PMC9829497.

[3] American Cancer Society - Treatment types. Accessed 2023 Sep https://www.cancer.org/cancer/managing-cancer/treatment-types.html

[4] Preiswerk F, Toews M, Cheng CC, Chiou JG, Mei CS, Schaefer LF, Hoge WS, Schwartz BM, Panych LP, Madore B. "Hybrid MRI-Ultrasound acquisitions, and scannerless real-time imaging." *Magn Reson Med.* 2017 Sep;78(3):897-908. doi: 10.1002/mrm.26467. Epub 2016 Oct 13. PMID: 27739101; PMCID: PMC5391319.

[5] National Cancer Institute - Division of Cancer Treatment & Diagnosis (Cancer Imaging Basics). Accessed 2023 Sep. https://imaging.cancer.gov/imaging_basics/cancer_imaging.htm

[6] Abayazid M, Kato T, Silverman S. G., Hata N. "Using needle orientation sensing as surrogate signal for respiratory motion estimation in percutaneous interventions." *International Journal of Computer Assisted Radiology and Surgery*, 2017.

[7] Fahmi S, Simonis FFJ, Abayazid M. "Respiratory motion estimation of the liver with abdominal motion as a surrogate." *Int J Med Robot.* 2018 Dec;14(6):e1940. doi: 10.1002/rcs.1940. Epub 2018 Aug 15. PMID: 30112864; PMCID: PMC6282606.

[8] Langen KM, Jones DT. "Organ motion and its management." *Int J Radiat Oncol Biol Phys.* 2001 May 1;50(1):265-78. doi: 10.1016/s0360-3016(01)01453-5. PMID: 11316572.

[9] Kato T, Takemasa K, Ikeda T, Sakagami H, Kato R, Narita Y, Oyama S, Komori S, Yamaguchi H, Murakami M. "Analysis of respiratory-induced motion trajectories of individual liver segments in patients with hepatocellular carcinoma." *J Appl Clin Med Phys.* 2024 Apr;25(4):e14257. doi: 10.1002/acm2.14257. Epub 2024 Feb 1. PMID: 38303539; PMCID: PMC11005968.

[10] Lal H, Neyaz Z, Nath A, Borah S. "CT-guided percutaneous biopsy of intrathoracic lesions". *Korean J Radiol.* 2012 Mar-Apr;13(2):210-26. doi: 10.3348/kjr.2012.13.2.210. Epub 2012 Mar 7. PMID: 22438689; PMCID: PMC3303905.

[11] Basu A, Routray A, Mukherjee R, Shit S. "Infrared imaging based hyperventilation monitoring through respiration rate estimation." *Infrared Physics & Technology, Volume 77*, 2016, Pages 382-390, ISSN 1350-4495, https://doi.org/10.1016/j.infrared.2016.06.014.

[12] Barbosa Pereira C, Yu X, Czaplik M, Blazek V, Venema B, Leonhardt S. "Estimation of breathing rate in thermal imaging videos: a pilot study on healthy human subjects." *J Clin Monit Comput.* 2017 Dec;31(6):1241-1254. doi: 10.1007/s10877-016-9949-y. Epub 2016 Oct 24. PMID: 27778207.

[13] Procházka A, Charvátová H, Vyšata O, Kopal J, Chambers J. "Breathing Analysis Using Thermal and Depth Imaging Camera Video Records." *Sensors*. 2017; 17(6):1408. https://doi.org/10.3390/s17061408

[14] Fraiwan L, Khasawneh N, Lweesy K, Elbalki M, Almarzooqi A, Abu Hamra N. "Non-Contact Spirometry Using a Mobile Thermal Camera and AI Regression." *Sensors.* 2021; 21(22):7574. https://doi.org/10.3390/s21227574

[15] McClelland, Jamie & Hawkes, D.J. & Schaeffter, Tobias & King, Andrew. "Respiratory motion models: A review." *Medical image analysis.* 2012. 17. 10.1016/j.media.2012.09.005.

[16] Testagrossa B, Ruello E, Gurgone S, et al. "Radio Frequency MRI coils and safety: how infrared thermography can support quality assurance." *Egypt J Radiol Nucl Med 52, 277* (2021). https://doi.org/10.1186/s43055-021-00659-y

[17] Will Frame Rate Make a Difference for Thermal? Published 2010 Feb. https://ipvm.com/reports/will-frame-rate-make-a-difference-for-thermal

[18] Guide to Choosing Frame Rate for Infrared Camera Core. Accessed 2024 Mar. https://www.gst-ir.net/news-events/infrared-knowledge/473.html#:~:text=Evaluate%20the%20frame%20rate%20capabilities,and%20reaching%20up%20to%20200Hz.

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

[20] Training and Validation Loss in Deep Learning. Published 2024 Jul. https://www.baeldung.com/cs/training-validation-loss-deep-learning

[21] Machine Learning 103: Loss Functions - *Why machine learning is an optimization problem* by Y.Natsume. Published 2022 Feb. https://towardsdatascience.com/machine-learning-103-loss-functions-37b84f3e9b73

[22] The 7 Most Common Machine Learning Loss Functions - *Every machine learning engineer should know about these common loss functions and when to use them* by Sparsh Gupta. Published 2023 Jun. https://builtin.com/machine-learning/common-loss-functions

[23] Hong Y, Sørensen S, Savarimuthu T, Cheng Z. "Needle Tip Tracking by Scattering Imaging." *New Technologies for Computer/Robot Assisted Surgery (CRAS).* 2023

[24] Find the Center of a Blob (Centroid) using OpenCV (C++/Python). Published 2018 Jul. https://learnopencv.com/find-center-of-blob-centroid-using-opencv-cpp-python/

[25] What to know about diaphragmatic breathing. Published 2023 Feb. https://www.medicalnewstoday.com/articles/diaphragmatic-breathing

[26] Science of Breath Work: *Breathing for Athletes*. Published 2022 Nov. https://blayze.io/blog/general/the-science-of-breathing-for-athletes

[27] Shimizu S, Shirato H, Xo B, Kagei K, Nishioka T, Hashimoto S, Tsuchiya K, Aoyama H, Miyasaka K. "Three-dimensional movement of a liver tumor detected by high-speed magnetic resonance imaging." *Radiother Oncol.* 1999 Mar;50(3):367-70. doi: 10.1016/s0167-8140(98)00140-6. PMID: 10392824.

[28] Beddar AS, Kainz K, Briere TM, Tsunashima Y, Pan T, Prado K, Mohan R, Gillin M, Krishnan S. "Correlation between internal fiducial tumor motion and external marker motion for liver tumors imaged with 4D-CT." *Int J Radiat Oncol Biol Phys.* 2007 Feb 1;67(2):630-8. doi: 10.1016/j.ijrobp.2006.10.007. PMID: 17236980.

[29] Hamidzada WA, Osuobeni EP. "Agreement between A-mode and B-mode ultrasonography in the measurement of ocular distances." *Vet Radiol Ultrasound.* 1999 Sep-Oct;40(5):502-7. doi: 10.1111/j.1740-8261.1999.tb00382.x. PMID: 10528845.