

Real-time Radiofrequency Ablation Lesion Depth Estimation Using Multi-frequency Impedance with a Deep Neural Network and Tree-based Ensembles

Emre Besler, *Student Member, IEEE*, Yearnchee Curtis Wang, *Student Member, IEEE*,
and Alan V. Sahakian, *Fellow, IEEE*

Abstract—Objective: Design and optimization of statistical models for use in methods for estimating radiofrequency ablation (RFA) lesion depths in soft real-time performance. Methods: Using tissue multi-frequency complex electrical impedance data collected from a low-cost embedded system, a deep neural network (NN) and tree-based ensembles (TEs) were trained for estimating the RFA lesion depth via regression. Results: Addition of frequency sweep data, previous depth data, and previous RF power state data boosted accuracy of the statistical models. The root mean square errors were 2 mm for NN and 0.5 mm for TEs for previous statistical models and the root mean square errors were 0.4 mm for NN and 0.04 mm for TEs for the statistical models presented in this paper. Simulation ablation performance showed a mean difference against physical measurements of 0.5 ± 0.2 mm for the NN-based depth estimation method and 0.7 ± 0.4 mm for the TE-based depth estimation method. Conclusion: The results show that multi-frequency data significantly improves the depth estimation performance of the statistical models. Significance: The RFA lesion depth estimation methods presented in this work achieve millimeter-resolution accuracy with soft real-time performance on an ARMv7-based embedded system for potential translation to clinical RFA technologies.

Index Terms—Radiofrequency ablation, tumor, cancer, control, monitoring, machine learning, ensemble, lesion, depth, deep network, random forest, adaptive boosting

I. INTRODUCTION

RADIOFREQUENCY ablation (RFA) is a minimally invasive, high temperature ablation method, that is applied by exposing the undesired tissue to high-frequency alternating current via a catheter or electrode, causing death by coagulative necrosis above a certain thermal threshold [1], [2]. It has found a wide range of applications in the medical field, one of which is the field of cardiology, where RFA is used to treat a wide range of cardiac diseases by creating a lesion on the arrhythmic part of myocardium through a catheter that is placed in the heart through a vein [3], [4]. RFA has become one of the most popular methods for atrial fibrillation, with the highest rate of success at paroxysmal type

[5], [6]. Recent studies also show that RFA for different types of ventricular tachycardia is a preferable option over other types of treatment, mainly drug therapy, with lower mortality rates on the patients who underwent ablation therapy [7]–[9]. Barrett's Esophagus with dysplasia has been another disease for which RFA treatment has been routinely used. Studies show that patients with RFA treatment have higher rates of eradication of abnormal esophageal tissue [10], [11]. RFA has been a preferable treatment for chronic obstructive sleep apnea as well, destroying the small part of soft tissue that collapses to block the airways of patients [12].

RFA has proven very useful in the field of cancer treatment, and it is used to remove tumors by thermal ablation, which will also be the main focus of this study. The tumors on different tissues can be removed by inducing coagulative necrosis due to high temperature. This treatment has become more popular especially on types of cancer for which open surgery carries certain risks with a large group of patients. Hepatocarcinoma (HCC) is an important example where open surgery is risky and RFA treatment has become particularly popular [13], [14], as well as lung cancer, where open surgery is not even a viable option for many patients [15]. Some studies show RFA being used together with open surgery to avoid cancer recurrence [16].

Since RFA destroys tissue non-selectively due to its uncontrollable nature, real-time monitoring of the ablation extent is essential for the success of the therapy [17]. Without an accurate method to track the size and the location of the ablated area, there is a considerable chance for the cancer to recur, due to residual tumors left unablated as well as the danger of overablation, that is the destruction of non-target tissues and/or critical structures [18].

As of 2019, low-cost, real-time monitoring of thermal ablation remains a large biotechnological problem to be solved. Since a monitoring scheme that is non-invasive, visually-intuitive, real-time, low-cost, and does not require significant capital equipment is difficult due to the opacity of the tissue, many methods utilize the changes in tissue properties; more specifically electrical, optical and acoustic behavior under the ablation treatment. However, in settings where feasible, magnetic resonance (MR) thermometry is known to provide an accurate assessment of the ablation depth for many clinical applications such as the removal of hepatic tumors, using the Proton Resonance Frequency shift (PRFS) method that produces temperature images used for Thermal Dose (TD)

E. Besler is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208.

Y.C. Wang is with the Department of Electrical and Computer Engineering, California State University, Los Angeles, CA 90032.

A.V. Sahakian is with the Department of Biomedical Engineering and the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208.

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

calculations in real-time [19]. In 2010s, compact hardware setups have been designed for MR thermometry to keep the monitoring equipment as minimally invasive as possible, while retaining the accuracy of the temperature map and increasing the speed of the imaging process [20], [21]. Most recently, studies on both echo-planar imaging and PRFS that produce images for TD calculations have been used on cardiac ablation in combination with image processing algorithms to remove motion and magnetic artifacts for better image quality [22], [23].

Electrical impedance tomography (EIT) uses surface electrodes surrounding the tissue under evaluation to measure impedance paths that are reconstructed into tissue electrical conductivity to provide lesion depth images that can be 90%+ accurate [17], [24]–[26]. Pairs of electrodes act as current drivers and others measure the voltage. The components of the voltage signals are recorded. All potential pairwise combinations of current pairs and voltmeter pairs can be measured.

Another method of imaging RFA progress is using acoustic and optoacoustic modalities. Namely, ultrasound-based imaging and laser-excitation ultrasound-sensing methods are used to image [24], [27], [28]. Both imaging methods rely on acoustic waves traveling from the target tissue to the acoustic sensor. These waves can be emitted from the sensor itself, in the case of ultrasound imaging, or from an optical source, in the case of optoacoustic imaging. Acoustic-based imaging requires minimizing air gaps within tissue, since air strongly attenuates the acoustic wave [24]. Since RFA can potentially reach boiling temperatures, gas bubbles can be generated that make imaging the ablation zone with B-mode pulse-echo ultrasound imaging difficult [24]. Also, acoustic imaging techniques have difficulties describing the thermal behaviour of a homogenous medium above 45 °C [29].

Among the more recent and advanced methods that use ultrasound, Nakagami-based imaging which is an operator-friendly software addition to conventional pulse-echo system, has been shown to provide 94% accuracy for monitoring RFA lesions in liver tissues in real-time [24]. However, Nakagami imaging cannot image muscular tissue as muscle fibers generate strong backscatter echoes. Another ultrasound-based technique is adaptive ultrasound, which adjusts the medium parameter in an automated manner to ameliorate the problems of acoustic imaging techniques above a certain temperature and beats the estimation error performance of conventional methods [30]. In addition, a single-phase CBE (change in backscattered energy) imaging method based on only positive values has been developed and has been shown to outperform the conventional CBE imaging [31], [32]. Most recently, echo decorrelation imaging with real-time ultrasound data has been used on microwave-induced ablation lesions in liver tissues and achieved 88.6% accuracy [33].

Optoacoustic imaging, a combination of optical and acoustic techniques, uses pulses of lasers to excite tissue and ultrasonic sensor arrays to record acoustic emissions from these light pulses [27], [28]. Optoacoustic techniques can image both temperature and ablation status, with the ablation lesion having a sharp signal change due to the coagulation.

While data collection is quick for the methods that utilize

the changes in tissue properties, reconstruction is complex. Absolute EIT imaging essentially requires solving an optimization problem overlaid onto the finite element problem. Essentially, one of the methods to solve for the EIT reconstruction solution (the tissue electrical conductivity of each node) is to operate on a finite element model mesh [25], [34]. The electrical conductivity at each node of the mesh is fitted until all of the conductivities of the tissue mesh reproduce the same impedances measured with the EIT electrodes. The nodes of the mesh that have different conductivities than expected are then the regions of interest. Thus, computing a single EIT-based lesion depth map requires time on the order of seconds and minutes (time increases with accuracy from 2+ seconds for 70% accuracy to 100+ seconds for 90%+ accuracy) [35]. Optoacoustic methods are 95%+ accurate in the mm scale up to at least 6 mm [27]. However, the actual construction of the three-dimensional lesion depth map from the sensed data requires computation on the order of 400+ seconds [27], [28]. This is due to the reconstruction algorithms being of a tomographic nature, similar to EIT, that requires the computation of the inverse solution to the model [28].

In this study, we are improving the Machine Learning (ML) approach that was first introduced in Wang et al. [36], where a pseudo-EIT method used the single-frequency impedance data to train a feed-forward neural network and posed the ablation depth estimation as a classification task. More recently, Besler et al. [37], which is the most related work to this study, showed that the estimation performance can be improved by posing the monitoring problem as a regression task, using different ML models and noise-free impedance data that is collected by an embedded system-based setup. Both of these prior studies have used single-frequency data. The major novelty in this study is using the ability of the embedded system design to take multi-frequency impedance measurements and using them both to enhance the data size and to use a two-step ML pipeline to improve the depth map reconstruction precision. The feature engineering in this study is more extensive as well, enabling more complex ML models to be used. Even though the types of regression models in [37] are retained in this study, increasing their complexity by tuning their hyperparameters is predicted to improve the overall ML performances due to the larger multi-frequency dataset with more features. Yet another contribution of this study is the application of an embedded system to compute and control the actuation of the machine learning inference-based radiofrequency ablation control in real time.

II. MATERIALS AND METHODS

A. Ablation Hardware and Data Collection

The model that simulates breast tissue and the RFA device that both delivers the ablation and collects the impedance data were the same as in [36], [37]. The breast tissue model consisted of pork loin and pork belly tissue that was butchered and packaged on the same day by slaughterhouses in Chicago, ground immediately before each experiment, and heated via waterbath to near-physiological temperature of 34 °C, and poured in a 80 mm x 80 mm x 80 mm acrylic fixture. The

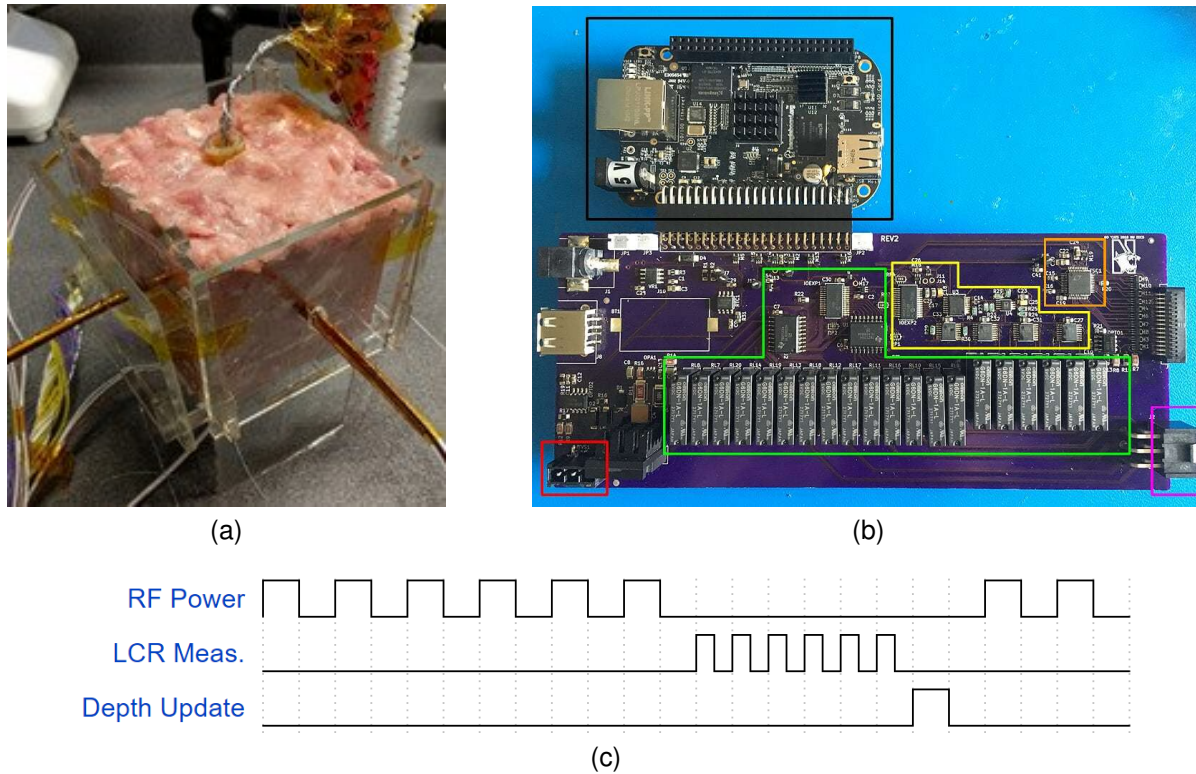


Fig. 1. a) The breast tissue model is shown with the RFA device inside and the thermal module inserted on each side. (Reprinted from [36]) b) The accessory board is shown with the parts of the embedded system indicated with different colors: The red box shows the RF generator connection socket, the green box shows the relay-based electrode switching subsystem, the yellow box shows impedance analyzer subsystem, the orange box shows an auxiliary temperature measurement subsystem, and the purple box shows the RFA device connection socket. (Reprinted from [37]) c) The timing diagram for the ablation, measurement and computation is shown. Each side of the device is activated for 7 seconds to deliver the AC pulse. After that, the impedance measurements at all 91 frequencies within a sweep are taken in 0.5 seconds per side. Lastly, the depths for all sides are calculated and the data is logged in another 7 seconds to complete 1 cycle. (Reprinted from [36])

RFA device was a sphere with a 40 mm diameter, fabricated on a stereolithography 3D printer using high-temperature tolerant resin. It was placed at the center of the model, leaving a 20 mm clearance at all directions. This entire setup is aimed to model a post-operation application of RFA, where the device would be inserted in the cavity opened after the tumor removal to ablate the remaining cancerous tissue. However, the principles of the individual-sided design in this study could also be used inside catheters or smaller RFA needles to perform the entire ablation process by itself. Perfusion or other tissue interfaces were not modeled within the tissue model at this time.

A resistance thermometer detector (RTD) input module was inserted through the clearances on each side, as shown in Figure 1a. The temperature data for each side after ablation was recorded using platinum 100 Ω resistance temperature detectors on this module. The detectors were placed at 0 mm, 5 mm, 10 mm and 15 mm depths from the side of the device. The temperature values for the depths in between were linearly interpolated. This interpolation was based on an assumption that the tissue between the temperature sensors is homogeneous such that linear interpolation yields a workable estimation [38]. Figure 2 shows temperature vs. time graph for two different sides at four different time points from the same ablation run. The general trend at all times shows an exponential decay, for which a linear approximation is

used. It is apparent from the figure that this approximation becomes more accurate for lower temperatures and deeper tissue levels. However, for smaller depth levels, an exponential interpolation in future studies might give a more accurate ground truth for ablation depth. Additionally, for a more complex model or in real-life, temperature-dependent tissue thermal properties (specific heat and thermal conductivity) can be utilized for spatial interpolation of tissue temperature. After the temperature was recorded for all depth values from 0.0 mm to 15.0 mm with a stepsize of 0.1 mm, tissue volumes at $\geq 43^\circ\text{C}$ for ≥ 10 minutes, $\geq 50^\circ\text{C}$ for ≥ 5 minutes or $\geq 57^\circ\text{C}$ for ≥ 2 seconds were considered ablated and the lesion depth was calculated. These threshold values were based on prior studies that tested different temperature and duration times for cell death under thermal ablation [39]–[41]. Assuming that the temperature at a given depth in the tissue has an upward trend, the continuous variations in temperature over time is handled as well by setting these thresholds.

The RFA device was divided into 6 sides (faces), each side having 4 stainless steel electrodes to deliver the AC pulses for ablation. These electrodes were connected to an embedded system that both delivered the AC pulses in low RF spectrum that realized the ablation and took the impedance measurements via the electrodes. Using the same electrodes both for ablation and for measurements eliminated the necessity to use

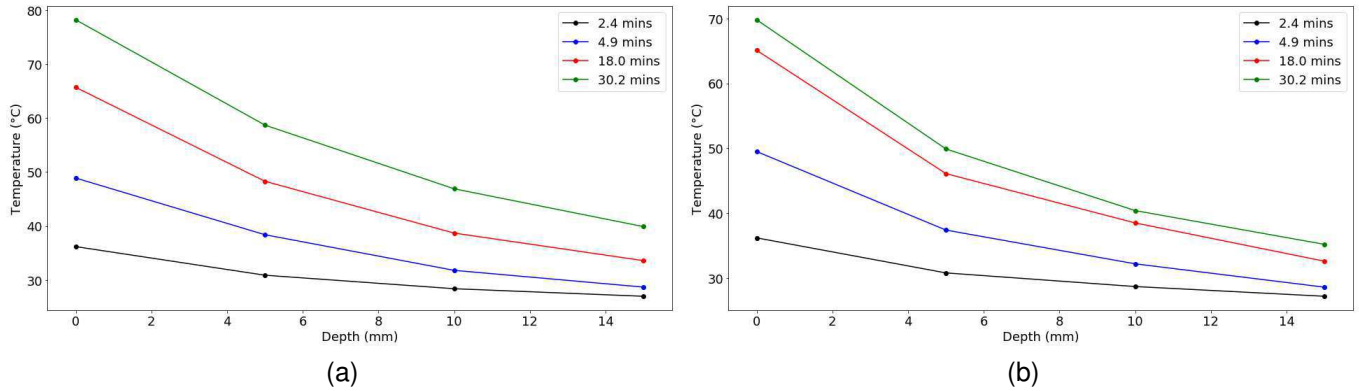


Fig. 2. The measured temperature values at 0, 5, 10 and 15 mm depth for different time points after the beginning of the ablation process. The temperature values in between are interpolated linearly. The measurements are from the same ablation run for a) superior face and b) lateral face of the ablation device.

any additional equipment that would affect the comfort of a patient in real life. Both the ablation device and a schematic that shows its sides are shown in Figure 3.

After one ablation cycle was complete for all 6 sides, the impedance data was collected for all the sides through the steel electrodes and measured at the impedance analyzer in the accessory board that controls the entire process. The ground truth for ablation depth was obtained after the ablation cycle by converting the temperature data to lesion depth in the data-logger software, using the mentioned thresholds. The side of the activated electrodes was also added to the data as another feature in the data logger. After the data was recorded, another ablation step started and the whole process was repeated. The entire timing diagram with the ablation, the measurements and the data logging is shown in Figure 1c.

The low-cost embedded system (Figure 1b) was first introduced in Besler et al. [37]. The embedded system consisted of the Beaglebone Black (Texas Instruments, Dallas, TX) and an accessory board that contains power switching, impedance, and temperature measurement circuits on-board. The system cost < \$250 for the parts, including the integrated circuits, accessory board printed circuit board and microcontroller board. The complex electrical impedance measurement subsystem is based on the AD5933 (Analog Devices, Norwood, MA) impedance analyzer integrated circuit, which was designed to measure the impedance magnitude and phase within 2% error range for a frequency range from 10 kHz to 100 kHz, following the low-impedance-ranged CN-0217 reference design from Analog Devices. For a given side and cycle, the impedance measurement circuit can make 91 impedance measurements within a frequency sweep from 10 kHz to 100 kHz, with a step size of 1 kHz. All the results in the related previous studies so far had been obtained with impedance measurements that were collected at a single frequency value, 100 kHz [36], [37]. This transition to multi-frequency measurements was the starting point of this study and predicted to be useful in a number of ways. Firstly, it would enable the setup to collect more data. For every sample in the single-frequency setup, there were now 91 samples each taken at a different frequency value within the sweep, all corresponding to the same ablation depth. Given that the data is not noisy and there is enough computational

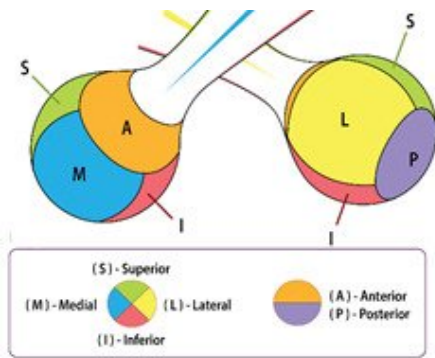
budget for a complex ML model to handle it, more data is always useful for a higher estimation performance. Secondly, the frequency became another feature of the dataset. From an ML perspective, this was interpreted as another dimension that could improve the predictive power of the models. From a medical perspective, it addressed the possibility that different frequencies can be more sensitive than others to the electrical properties of different types of tissue.

Along with the frequency, there were four other new features engineered in this study that were predicted to improve the ML performance as well. The first set was a binary feature that indicates whether the side of the RFA device from which the impedance measurement were taken was activated during the previous ablation cycle. The second set of new features were the second impedance magnitude and phase measurements. The final new feature was the ablation depth on that side in the previous cycle. These new features, respectively, were predicted to improve the ML performance in the following ways: (1) tissue not undergoing heating begins to cool, so while the impedance measurements of the tissue may look similar to a previous depth, the state of the tissue is that it was fully ablated, so this data should not be treated the same as the data collected when the tissue is undergoing active ablation; (2) there are potentially small gaps between the device electrodes and tissue, which may move after the first ablation cycle due to thermal expansion, so data from the second cycle is used to confirm the initial data; and (3) using the previous depth as an input is an attempt to prevent the models from producing output depths smaller than the previously-estimated depth, a problem that frequently occurred with the previous depth estimation models.

So, along with the original features of initial magnitude, initial phase, final magnitude, final phase and the 6 one-hot-encoded side features from [37], the frequency, previous state, previous depth, second initial magnitude and phase were used to create a 15-dimensional dataset. The ablation depth was the target value. There were 1,251,432 instances in the dataset, collected from 2,292 ablation cycles on 30 tissue model samples. The number of cycles performed per side per tissue sample were randomly distributed between 0 and 40 cycles, to account for all different ablation shapes possible.



(a)



(b)

Fig. 3. a) The RFA device design with four parallel electrodes on each side. There is a ring of electrodes at the anterior face. (Reprinted from [36]) b) The RFA device separates the electrodes into six faces, each face representing a side of a cube for direct geometric visualization within tissue. (Reprinted from [36])

Each cycle was 7 seconds long. This yielded different total ablation times per tissue sample, which is the summation of all the ablation cycles performed per side for that tissue sample. The total time duration of the training and simulation dataset ablations varied between 5-40 minutes depending on the target depths of each side of the tissue volume. This variation was due to generation of the full spectrum of training data for handling of cases up to 15 mm target ablation depths.

The estimation of the ablation depth was posed as a regression task. It was shown in the previous studies with the same setup that regression is a more direct approach than classification; especially with the enhanced data size and new features, the regression models in this study were predicted to outperform [37].

B. Predictive Analysis with Machine Learning

The ML part of this study contains the regression models from [37] whose utility for making depth estimations from impedance data has already been shown in that study. The ML models are used on the entire dataset to predict the ablation depth for all the instances. After all the estimations are obtained, those that pertain to the same frequency sweep at the same cycle and side are grouped together. Each group has 91 estimations and they belong to the instances created by the measurements of one full frequency sweep, having the same target depth value. These 91 estimations were merged to get the final ablation depth.

1) *Deep Neural Network*: The first ML model was a Deep Neural Network with a number of hidden layers and an output layer that contains only one single node, predicting the ablation depth as a regression task [42]. Since the dataset in this study had more instances and newly-engineered dimensions, a different architecture than the network in [37] was needed. The network was made more complex by increasing the number of nodes at each layer, so that it can handle the data without any underfitting. Furthermore, dropout layers are added after each fully connected layer to regulate the model during training [43]. The details for the final architecture is given in Section III.

For the optimization of the network, the ADAM algorithm was used as it is repeatedly shown to be most computationally-efficient algorithm for optimizing deep networks, combining the advantages of different extensions of Stochastic Gradient Descent [44]. The whole network was implemented on the Keras library running on top of the Tensorflow backend [45].

2) *Random Forest*: The second ML model is a Random Forest, an ensemble model with a Decision Tree as its base predictor [46]. Each tree is trained with the CART algorithm and given only a random subset of the dataset, making the entire ensemble model more robust to changes in the dataset [47]. After all trees are trained with different random subsets, the estimation for a new instance can be made by averaging the estimations of all the trees in the ensemble. Tree-based models are generally known to be useful for a classification task rather than regression because of the finite number of leaf nodes but the discrete target values in this study (depth levels between 0.0 and 15.0 mm, with a resolution of 0.1 mm) allows tree-based models to be used with a high regression performance.

3) *Adaptive Boosting*: The third and last ML model is Adaptive Boosting, another ensemble model that is based on decision trees, however, instead of training the trees in parallel on random subsets of the data, the trees are trained one after another, on the entire dataset [48]. A weight is assigned to each instance in the dataset, and these weights are updated after each tree is trained, increasing the weight of the mispredicted instances. This enables the next tree in the ensemble to pay more attention to the instances that the previous tree mispredicted. A predictor weight is assigned to each tree after it is trained, based on how many instances it predicts correctly, and all the trees add up to a complex model. After the training is complete, the estimation for a new instance is done with a weighted average of all the

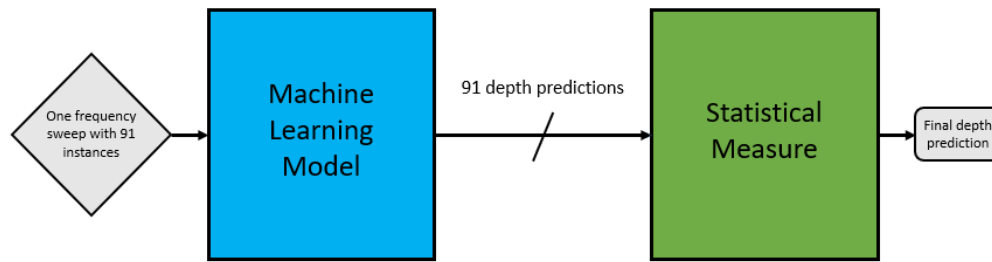


Fig. 4. The complete estimation flowchart after the ML models are trained. Each instance in the frequency sweep has 15 features, so a 91×15 matrix is given as the input to the ML model. The model predicts a depth value for each instance and gives a 91×1 vector as the output. The statistical measure merges these 91 depth values and returns the final depth estimation.

tree estimations, multiplying the predictor weights with the estimations, normalizing and taking the average.

4) *Merging the Estimations of a Frequency Sweep:* As briefly mentioned at the beginning of this section, the 91 instances of a full frequency sweep all pertain to the same depth value. The ML models were able to predict that value accurately for most of these instances in the sweep but for not all of them. So, these 91 estimations should be merged into a single and accurate depth estimation and while doing that, the correct estimations among the group should be taken advantage of, making up for the misestimations. This is implemented by using two basic statistical measures. The first measure was a truncated mean that keeps the 95% of the estimations, discards the rest on the high and low ends and takes the average. This was expected to result with an accurate final depth estimation for each sweep because it was assumed that most estimations among the group would be correctly made, pulling the average very close to the correct value. Secondly, mode operation is used to obtain the final depth estimation. It simply returns the most frequently occurring term in a given set and based on the assumption of most estimations being correct, it was predicted to almost always return the correct estimation because the misestimations within the frequency sweep would just be omitted.

After all the models are trained, Figure 4 has the block diagram that shows the estimation of the final depth for a given frequency sweep with 91 instances. The ML model estimates 91 depth values for 91 instances of one frequency sweep. The statistical measure merges these 91 depth estimations to a single depth value which is the final estimation.

C. Simulation Testing of Lesion Depth Estimation Models

To create test cases for the physical simulation testing, similar ablations were performed as with the collection of the training data. Since the biological properties would change from tissue to tissue in a real-life scenario, it is useful to assess the performance of a method on different tissue samples. Therefore, these ablation experiments were all performed on new tissue model samples in order to see how accurate the trained ML models are on a previously unseen model. These new ablation experiments were performed within lean pork loin tissue, the tissue that we previously determined to yield the worst estimation accuracy of the lesion depth estimation

models [36], [37]. However, 3 seconds of sleep time were inserted after the impedance measurement, to simulate the computation of the lesion depth map during the ablation procedure. Four simulation test ablations were created, yielding 24 ablation depths for testing the final physically-measured depth against the final predicted depth of the models. This data was set aside and not used for model training, validation, or statistical testing.

III. RESULTS

For both datasets and ML models, 80% of the data was used to train the models and tune the hyperparameters with a 10-fold cross-validation. (CV) The other 20% was held out to test how well the trained models generalize to new data. Since the ML task at hand was regression, R^2 and root mean squared error (RMSE) were used as estimation evaluation metrics, along with the residual plots to visualize the estimation performance.

The final neural network architecture was 4 layers with 300 nodes per layer, represented as a node graph of (10, 300, 300, 300, 300, 1). Dropout layers set to 30% dropout was used during the training of the neural network. Bias layers were initialized to normal random distributions and constrained to have a maximum norm of 20 to prevent saturation. ReLU was used as the activation function, following previous methods for regression with deep neural networks [36], [49].

As for the tree-based ensemble models, a regularization hyperparameter of the decision tree should be picked and tuned in order to avoid overfitting, which tree-based predictors are very prone to. As proven to work in the previous studies, maximum number of leaf nodes for each tree in the ensemble is picked to be tuned for both the Random Forest and the Adaptive Boosting model. Another hyperparameter to tune in both of them was the number of trees in the ensemble, on which the complexity of the model also depends.

The ensemble hyperparameters are tuned by a 2-dimensional grid search and the 10-fold cross-validation. The optimal Random Forest has 50 trees and 10000 leaf nodes maximum, whereas the Adaptive Boosting model has 20 trees with 20000 leaf nodes maximum.

Table I has the performance evaluation of all the models in this study, after the cross-validation and testing with the hold-out set. These results only show how well the models

TABLE I
OUTPUT PERFORMANCES OF ALL THE ML MODELS WHEN TRAINED, CROSS-VALIDATED AND TESTED ON THE ENTIRE DATASET

ML Model	Estimation Metrics for Regression			
	Test RMSE (mm)	Test R^2	CV-Average RMSE (mm)	CV-Average R^2
Deep Network	0.35	99.3%	0.35	99.3%
Random Forest	0.044	99.9%	0.034	99.9%
Adaptive Boosting	0.033	99.9%	0.021	99.9%

do before the estimations of the same frequency sweep are merged. So, it serves as a performance comparison between the ML models.

After the models were trained and their performance tested, they were run with the entire dataset to obtain all the depth estimations. As explained in Section II-A, the estimations for the instances of the same frequency sweep were grouped and merged into final depth estimations by taking their mode and truncated mean. This merging was done with the estimations of all three ML models separately. The RMSE values after the merging for all the frequency sweeps in the dataset is shown in Table II.

TABLE II
FINAL RMSES (IN MM) FOR ALL THE ML MODELS AFTER THEIR ESTIMATIONS ARE GROUPED AND MERGED

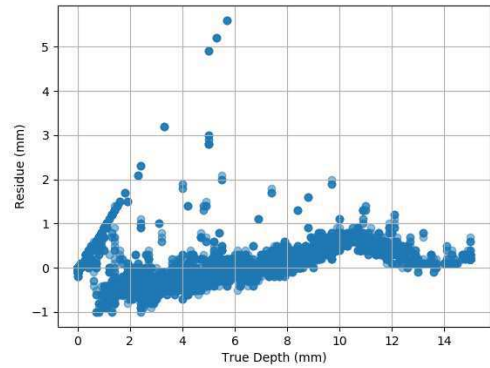
ML Model	Final RMSEs by Merging Function (mm)	
	Truncated Mean	Mode
Deep Network	0.40	0.35
Random Forest	0.01	0.01
Adaptive Boosting	0.01	0.00

On top of the numeric results, the residual plots of all the ML estimations, both before and after the merging, are shown in Figures 5 and 6. Figure 5 has the plots for the Deep Network and Figure 6 has the tree-based model plots. The plots for the Deep Network and the tree-based models were shown in separate figures for visualization purposes and because of the difference in the scales of the y-axes. In both figures, the first row has the residual plots of the entire dataset (all 1,251,432 instances) with their depth values predicted by the ML models. The second and third rows have the residual plots of the estimations for 13,752 frequency sweeps, merged by truncated mean and mode operations, respectively.

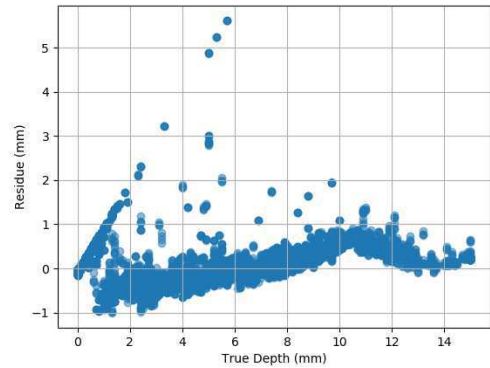
Lastly, the performance of the ML models against the physical measurements collected from the simulation testing ablations are in Table III.

IV. DISCUSSION

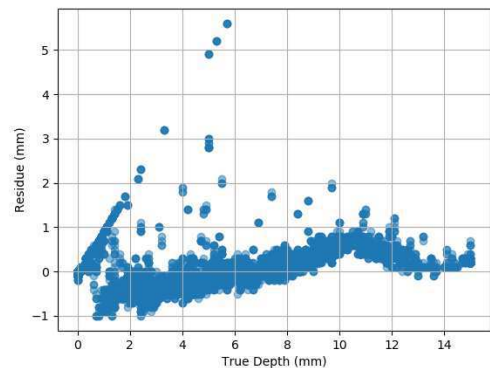
Before the results, the limitations of this study in comparison to a real-life scenario should be discussed. As mentioned in Section II-A, the tissue model presented within this study does not model perfusion or tissue interface issues. One of the reasons this decision was made was to focus on a proof-of-concept study for multi-frequency, low-cost embedded system-driven ablation lesion depth estimation. In future works, especially when considering other tissue models, the



(a) Raw data (All instances in the dataset)

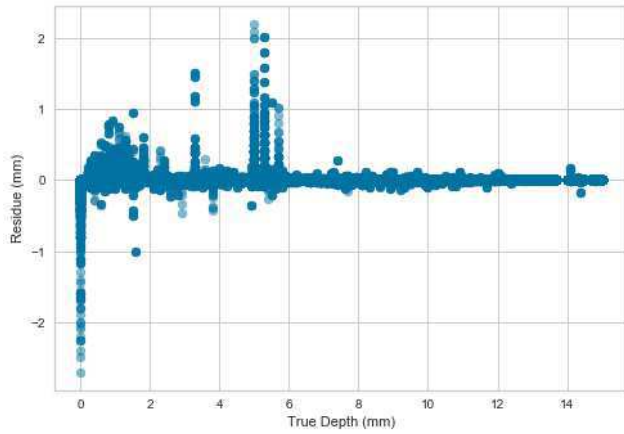


(b) Merged by truncated mean

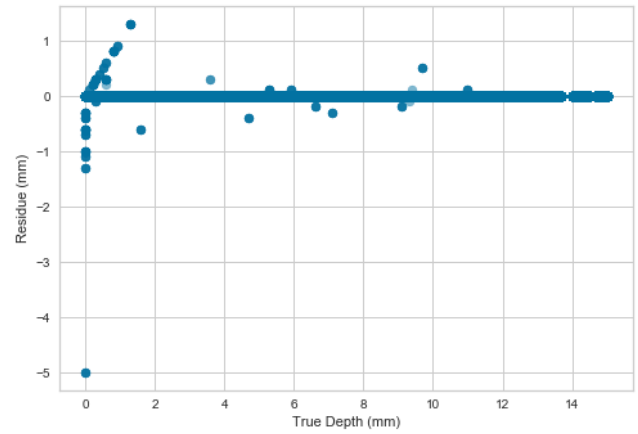


(c) Merged by mode

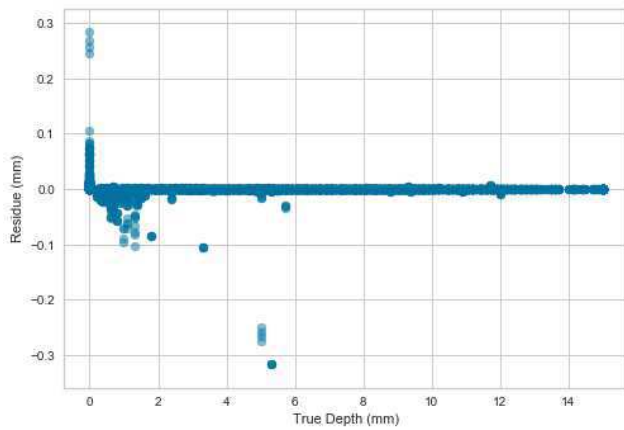
Fig. 5. The residual plots of the Deep Network and its merged estimations



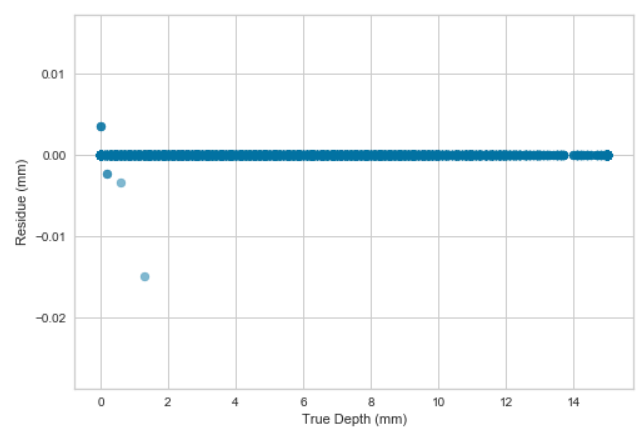
(a) Raw data with Random Forest



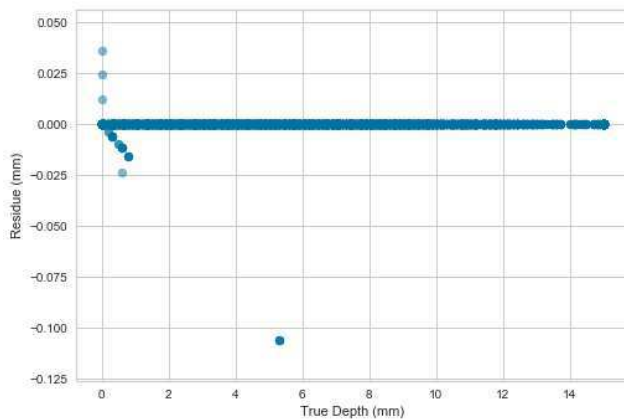
(b) Raw data with Adaptive Boosting



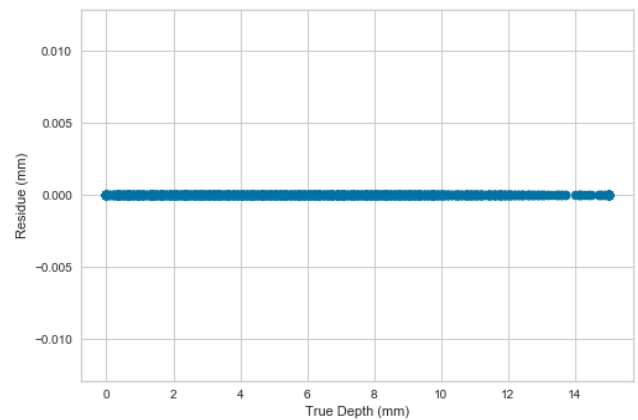
(c) Random Forest estimations merged by truncated mean



(d) Adaptive Boosting estimations merged by truncated mean



(e) Random Forest estimations merged by mode



(f) Adaptive Boosting estimations merged by mode

Fig. 6. The residual maps of both tree-based models and their merged estimations

physical modeling of perfusion will be very important due to the heat sink effect. We expect a few changes within the system are required to detect and handle perfusion as well as interfaces in heterogenous tissue: (1) detection of large blood vessels present within tissue in a subvolume similar to the classification of different tissue types performed by Laufer et

al. [50], [51]; (2) characterizing the change in tissue electrical properties as it undergoes simultaneous heating through the RFA applicator and cooling through the blood vessel heat sink; and (3) handling the change in tissue electrical properties within the machine learning model, either through additional training data for direct statistical handling or manual logic-

TABLE III
SIMULATION ABLATION LESION DEPTH MEASUREMENT PERFORMANCE OF ML MODELS AGAINST FINAL PHYSICAL LESION DEPTH MEASUREMENTS.

ML Model	Samples Measured	Mean Difference (mm)	Standard Error (mm)
Deep Network	24	0.5	0.2
Adaptive Boosting	24	0.7	0.4

based handling.

Additionally, the ex-vivo tissue model used within this study was processed by a slaughterhouse, such that the tissue electrical properties (especially with regards to capacitance) differ slightly from in-vivo tissues. Our study attempted to provide near-physiological simulation for the temperature-dependent conductivity magnitude by heating the tissue model to near-physiological temperature. However, it is likely that cell activity, particularly with the cell membrane, will affect the imaginary component of the complex electrical impedance measurements.

As for the results, when trained with the entire dataset in this study, both the Deep Network and the tree-based ensembles outperform the models in Besler et al. [37], which were based on training data collected with the impedance data at a single frequency. Therefore, it can be said that adding the new features introduced in Section II-A as well as enhancing the data size with multi-frequency measurements improved the performance of all three ML models. The RMSE of the tree-based models decreased by an order of 10. However, the biggest improvement belongs to the Deep Network when compared to the one in [37], which makes sense because even though it is generally the most powerful non-linear model, a network needs much more data than the tree-based non-linear models. The measurement method in this study maintains the big data that the network needs.

Even though there is considerable estimation improvement with the new features and data size, it is also apparent that there is still some room for improvement, especially after looking at the residual plots of both tree-based models in Figure 6. Most residues are close to zero, however there are still many instances where the inaccuracy is over 2 mm. This would be problematic for a real-time RFA scenario. Especially the buildup of inaccurate estimations for a true depth of 0 mm should be taken care of because if the estimation returns a high non-zero value when the true depth is 0 mm, it will leave a significantly large volume of the tumor unablated, increasing the possibility of cancer to recur. These considerations justify the need to merge the estimations of frequency sweeps.

After the 91 estimations of each frequency sweep is merged into one final depth estimation, there is considerable further improvement for the tree-based ensemble models. The RMSEs drop by yet another order of 10 when all the estimations of the frequency sweeps (in the entire dataset) are merged by taking their truncated mean. When merged by taking the mode, the Random Forest performance gets even better and the Adaptive Boosting algorithm reaches a perfect depth estimation performance for the entire dataset. This shows that the estimations for some frequencies are indeed better than others and furthermore, the lower RMSE after merging

means that the models in fact predict the depth correctly for the samples at most frequencies and this majority of correct estimations is able to make up for the false estimations within the frequency sweep. This pattern cannot be observed for the Deep Network results. As can be seen in Figure 5, neither the residual map, nor the RMSE change much with the merging of the estimations within a frequency sweep, meaning that the estimation error is about the same at all the frequencies.

Lastly, the performance of the ML models on the physical simulation data gives an idea on how they would perform in a real-time ablation procedure. The results on Table II show that some of the models experience the problem of generalizing to a real-world scenario. Even though the mean of the depth estimations for all the runs and all the sides are at an acceptable level, they are not as good as the final depth estimations that were obtained during model training.

The Deep Network model proves to be the most successful at generalizing to physical simulations with a 0.5 mean difference, outperforming the Adaptive Boosting algorithm even though the AB model reached a near-perfect estimation performance for both training and test data. So, the Deep Network proved itself to be robust under a different set of data, that is common for neural network models with a complex architecture and enough data to train on. The performance drop of Adaptive Boosting can be explained with the most common handicap of the tree-based models that is overfitting to the data they were trained on (even though they are regularized) and not performing as expected when tested on a separately collected data. Another example to this would be the performance of the Random Forest model on the physical simulation test, which is omitted in Table II because it had the worst estimation performance. So, after the first two simulation runs, it was not used anymore to predict the ablation depth for the simulations.

On the simulation data, all ML models performed more poorly than they did on the validation and test sets. While this can be explained with the ML concept of generalization to new data, another possibility is that the ground truth for ablation depth in the training data is not accurate enough. One major step to improve the accuracy would be to take the temperature measurements at more depth levels in the tissue with a better quality RTD module and more importantly, to use a more complex fit than linear interpolation for the temperature vs depth graph, before mapping the temperature data to the depth values.

As for temperature-depth mapping, another approach that can give a better ground truth would be thermal dose calculation, that takes the time integral over the recorded temperature values and thresholds the integral sum, which is the thermal dose, instead of the temperature. Its application to this setup is beyond the scope of this study, however it is certainly an

idea to explore for further studies [52], [53].

V. CONCLUSION

The results in this study show that a real-time monitoring scheme for Radiofrequency Ablation can be successfully implemented on a non-perfused breast tissue model with a pipeline of an non-linear ML model and a statistical merging operation. Taking advantage of the multi-frequency impedance measurements of the embedded system design helped the entire pipeline to achieve better results and it is also tested that the successful estimation performance can be generalized to a real-time ablation simulation. Future studies will focus on development of the multi-frequency depth estimation system with better ground truth for target depth values and also within a more realistic tissue model, taking perfusion and interface issues into account. Lastly, in-vivo experiments are a logical further step beyond the in-vitro measurements in this study.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under grants No. 1622842 and 1738541. The authors also would like to thank Terence Chan for his help with the processing of the datasets.

REFERENCES

- [1] D. R. Shah *et al.*, "Current oncologic applications of radiofrequency ablation therapies," *World journal of gastrointestinal oncology*, vol. 5, no. 4, p. 71, 2013.
- [2] C. L. Brace, "Radiofrequency and microwave ablation of the liver, lung, kidney, and bone: What are the differences?" *Current problems in diagnostic radiology*, vol. 38, no. 3, pp. 135–143, 2009.
- [3] N. F. Marrouche *et al.*, "Catheter ablation for atrial fibrillation with heart failure," *New England Journal of Medicine*, vol. 378, no. 5, pp. 417–427, 2018.
- [4] K.-H. Kuck *et al.*, "Cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation," *New England Journal of Medicine*, vol. 374, no. 23, pp. 2235–2245, 2016.
- [5] A. Verma *et al.*, "Approaches to catheter ablation for persistent atrial fibrillation," *New England Journal of Medicine*, vol. 372, no. 19, pp. 1812–1822, 2015.
- [6] H. Calkins *et al.*, "2017 hrs/ehra/ecas/aphrs/solacee expert consensus statement on catheter and surgical ablation of atrial fibrillation," *Ep Europace*, vol. 20, no. 1, e1–e160, 2017.
- [7] R. Tung *et al.*, "Freedom from recurrent ventricular tachycardia after catheter ablation is associated with improved survival in patients with structural heart disease: An international vt ablation center collaborative group study," *Heart rhythm*, vol. 12, no. 9, pp. 1997–2007, 2015.
- [8] J. L. Sapp *et al.*, "Ventricular tachycardia ablation versus escalation of antiarrhythmic drugs," *New England Journal of Medicine*, vol. 375, no. 2, pp. 111–121, 2016.
- [9] C. Zheng *et al.*, "Approach selection of radiofrequency catheter ablation for ventricular arrhythmias originating from the left ventricular summit: Potential relevance of pseudo delta wave, intrinsicoid deflection time, maximal deflection index," *BMC cardiovascular disorders*, vol. 17, no. 1, p. 140, 2017.
- [10] N. J. Shaheen *et al.*, "Radiofrequency ablation in barrett's esophagus with dysplasia," *New England Journal of Medicine*, vol. 360, no. 22, pp. 2277–2288, 2009.
- [11] B. J. Qumseya *et al.*, "Adverse events after radiofrequency ablation in patients with barrett's esophagus: A systematic review and meta-analysis," *Clinical Gastroenterology and Hepatology*, vol. 14, no. 8, pp. 1086–1095, 2016.
- [12] J. Farrar *et al.*, "Radiofrequency ablation for the treatment of obstructive sleep apnea: A meta-analysis," *The Laryngoscope*, vol. 118, no. 10, pp. 1878–1883, 2008.
- [13] S. Shiina *et al.*, "Radiofrequency ablation for hepatocellular carcinoma: 10-year outcome and prognostic factors," *The American journal of gastroenterology*, vol. 107, no. 4, p. 569, 2012.
- [14] R. Tateishi *et al.*, "Percutaneous radiofrequency ablation for hepatocellular carcinoma: An analysis of 1000 cases," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 103, no. 6, pp. 1201–1209, 2005.
- [15] T. De Baere *et al.*, "Radiofrequency ablation is a valid treatment option for lung metastases: Experience in 566 patients with 1037 metastases," *Annals of Oncology*, vol. 26, no. 5, pp. 987–991, 2015.
- [16] E. K. Abdalla *et al.*, "Recurrence and outcomes following hepatic resection, radiofrequency ablation, and combined resection/ablation for colorectal liver metastases," *Annals of surgery*, vol. 239, no. 6, p. 818, 2004.
- [17] H. Wi *et al.*, "Real-time conductivity imaging of temperature and tissue property changes during radiofrequency ablation: An ex vivo model using weighted frequency difference," *Bioelectromagnetics*, vol. 36, no. 4, pp. 277–286, 2015.
- [18] K. F. Chu and D. E. Dupuy, "Thermal ablation of tumours: Biological mechanisms and advances in therapy," *Nature Reviews Cancer*, vol. 14, no. 3, p. 199, 2014.
- [19] O. Seror *et al.*, "Real time monitoring of radiofrequency ablation based on mr thermometry and thermal dose in the pig liver in vivo," *European radiology*, vol. 18, no. 2, pp. 408–416, 2008.
- [20] M. A. Erturk *et al.*, "Radiofrequency ablation, mr thermometry, and high-spatial-resolution mr parametric imaging with a single, minimally invasive device," *Radiology*, vol. 281, no. 3, pp. 927–932, 2016.
- [21] N. A. Volland *et al.*, "Initial feasibility testing of limited field of view magnetic resonance thermometry using a local cardiac radiofrequency coil," *Magnetic Resonance in Medicine*, vol. 70, no. 4, pp. 994–1004, 2013. DOI: 10.1002/mrm.24534.
- [22] V. Ozenne *et al.*, "Improved cardiac magnetic resonance thermometry and dosimetry for monitoring lesion formation during catheter ablation," *Magnetic resonance in medicine*, vol. 77, no. 2, pp. 673–683, 2017.
- [23] S. Toupin *et al.*, "Feasibility of real-time mr thermal dose mapping for predicting radiofrequency ablation outcome in the myocardium in vivo," *Journal of Cardiovascular Magnetic Resonance*, vol. 19, no. 1, p. 14, 2017.
- [24] Z. Zhou *et al.*, "Monitoring radiofrequency ablation using real-time ultrasound nakagami imaging combined with frequency and temporal compounding techniques," *PLoS One*, vol. 10, no. 2, e0118030, 2015.
- [25] V. A. Cherepenin *et al.*, "Three-dimensional eit imaging of breast tissues: System design and clinical testing," *IEEE transactions on medical imaging*, vol. 21, no. 6, pp. 662–667, 2002.
- [26] A. Javaherian *et al.*, "A fast time-difference inverse solver for 3d eit with application to lung imaging," *Medical & biological engineering & computing*, vol. 54, no. 8, pp. 1243–1255, 2016.
- [27] G. A. Pang *et al.*, "Three-dimensional optoacoustic monitoring of lesion formation in real time during radiofrequency catheter ablation," *Journal of cardiovascular electrophysiology*, vol. 26, no. 3, pp. 339–345, 2015.
- [28] X. L. Dean-Ben *et al.*, "Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1922–1928, 2012.
- [29] X. Geng *et al.*, "Comparison of ultrasound temperature imaging with infrared thermometry during radio frequency

- ablation," *Japanese Journal of Applied Physics*, vol. 53, no. 4, p. 047 001, 2014.
- [30] Y.-D. Liu *et al.*, "Adaptive ultrasound temperature imaging for monitoring radiofrequency ablation," *PloS one*, vol. 12, no. 8, e0182457, 2017.
- [31] Z. Zhou *et al.*, "Monitoring microwave ablation using ultrasound echo decorrelation imaging: An ex vivo study," *Sensors*, vol. 19, no. 4, p. 977, 2019.
- [32] P.-H. Tsui *et al.*, "Using ultrasound cbe imaging without echo shift compensation for temperature estimation," *Ultrasonics*, vol. 52, no. 7, pp. 925–935, 2012.
- [33] L. Zhang *et al.*, "Ultrasound single-phase cbe imaging for monitoring radiofrequency ablation," *International Journal of Hyperthermia*, vol. 35, no. 1, pp. 548–558, 2018.
- [34] D. S. Holder, *Electrical impedance tomography: methods, history and applications*. CRC Press, 2004.
- [35] S. Martin and C. T. Choi, "A post-processing method for three-dimensional electrical impedance tomography," *Scientific reports*, vol. 7, no. 1, p. 7212, 2017.
- [36] Y. C. Wang *et al.*, "Real-time estimation of lesion depth and control of radiofrequency ablation within ex vivo animal tissues using a neural network," *International Journal of Hyperthermia*, vol. 34, no. 7, pp. 1104–1113, 2018.
- [37] E. Besler *et al.*, "Real-time monitoring radiofrequency ablation using tree-based ensemble learning models," *International Journal of Hyperthermia*, vol. 36, no. 1, pp. 428–437, 2019.
- [38] D. Haemmerich *et al.*, "In vitro measurements of temperature-dependent specific heat of liver tissue," *Medical engineering & physics*, vol. 28, no. 2, pp. 194–197, 2006.
- [39] S. N. Goldberg *et al.*, "Tissue ablation with radiofrequency: Effect of probe size, gauge, duration, and temperature on lesion volume," *Academic radiology*, vol. 2, no. 5, pp. 399–404, 1995.
- [40] I. V. Larina *et al.*, "Real-time optoacoustic monitoring of temperature in tissues," *Journal of Physics D: Applied Physics*, vol. 38, no. 15, p. 2633, 2005.
- [41] W. Malskat *et al.*, "Temperature profiles of 980-and 1,470-nm endovenous laser ablation, endovenous radiofrequency ablation and endovenous steam ablation," *Lasers in medical science*, vol. 29, no. 2, pp. 423–429, 2014.
- [42] Y. LeCun *et al.*, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [43] I. Goodfellow *et al.*, *Deep learning*. MIT press Cambridge, 2016, vol. 1.
- [44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [45] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [46] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, Springer, 2000, pp. 1–15.
- [47] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. Springer, 2012.
- [48] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear estimation and classification*, Springer, 2003, pp. 149–171.
- [49] M. A. Nielsen, *Neural networks and deep learning*. Determination press, San Francisco, CA, USA, 2015, vol. 25.
- [50] S. Laufer and B. Rubinsky, "Tissue characterization with an electrical spectroscopy svm classifier," *IEEE transactions on biomedical engineering*, vol. 56, no. 2, pp. 525–528, 2009.
- [51] S. Laufer *et al.*, "Electrical impedance characterization of normal and cancerous human hepatic tissue," *Physiological measurement*, vol. 31, no. 7, p. 995, 2010.
- [52] P. Mertyna *et al.*, "Radiofrequency ablation: The effect of distance and baseline temperature on thermal dose required for coagulation," *International Journal of Hyperthermia*, vol. 24, no. 7, pp. 550–559, 2008.
- [53] G. C. Van Rhooen *et al.*, "Cem43 c thermal dose thresholds: A potential guide for magnetic resonance radiofrequency exposure levels?" *European radiology*, vol. 23, no. 8, pp. 2215–2227, 2013.