MRI RF-induced Heating Prediction for Complex-shaped Passive Implantable Devices using Neural Network Methods

by
Jiajun Chang

A dissertation submitted to the Department of Electrical and Computer Engineering,
Cullen College of Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Electrical Engineering

Chair of Committee: Dr. Ji Chen

Committee Member: Dr. Driss Benhaddou

Committee Member: Dr. Jiefu Chen

Committee Member: Dr. David Jackson

Committee Member: Dr. Wolfgang Kainz

University of Houston
May 2022

ACKNOWLEDGMENTS

I would like to thank the following people, without whom I would not have been able to complete this research, and without whom I would not have made it through my Ph.D. degree.

I would like to firstly thank Dr. Ji Chen for providing tremendous help to me and gave me an opportunity to join his lab when I first arrived in University of Houston and met some difficulties in life and academia. He also provided great guidance on how to conduct research independently and how to be a successful Ph.D. student throughout the past five years. Without him, I wouldn't be the person I am today.

I would like to thank my dissertation committee: Dr. Driss Benhaddou, Dr. Jiefu Chen, Dr. David Jackson and Dr. Wolfgang Kainz for giving me great help and insightful advice through my dissertation proposal and defense. Their questions and thoughts are precise and important which always addresses on some places that I was missing. I would also like to thank Dr. Stuart Long for helping me through my Ph.D. qualification exam.

I would like to thank my senior colleague and friend Dr. Ran Guo for helping me tackle the problems in the research projects and encouraging me all the time during the difficult times. I would like to thank Dr. Qianlong Lan and Dr. Jianfeng Zheng for the insights on my research direction. I would also like to thank my friends Wei Hu, Shuo Song, Yu Wang and Rui Yang for the stimulating discussions about the future and happy distractions to rest my mind outside my academic path, especially table tennis matches we played together. I would like to thank Yongqi Xiao for accompanying me and provide me great support on life during my highs and lows.

Last but not least, I would like to thank my parents and my families for the wise counsel and sympathetic ear. As a primary school English teacher and a network technician back in China, none of my parents can provide me professional advice. However, they listened to me patiently and provided me with life wisdoms instead to help me manage my life in US better. It is them who sacrifice a lot of time to shape me from a kid from a small city in middle China to a Ph.D. candidate today. I would like to give my utmost love to them.

ABSTRACT

Magnetic Resonance Imaging (MRI) Radiofrequency (RF) -induced heating are one of the major safety concerns for patients with Passive Implantable Medical Devices (PIMDs) implanted inside the body. Evaluation of RF-induced heating includes experimental measurements and full-wave Electromagnetic (EM) simulations which will cost a significant amount of time and computational resources. Neural Network (NN) methods are introduced as a data-driven regression model which trains the parameters using device features to implement the predictions of RF-induced heating. While the previous NN models cannot predict the RF-induced heating of complex-shaped PIMDs as the device structure cannot be characterized by several parameters. Also, no discussions have been made on the strategy of training dataset selection.

In this study, mesh-based Convolutional Neural Network (CNN) models are introduced to implement the heating prediction for complex-shaped PIMDs. Tibia Plating System and Spinal Fixation System device models are developed with variations on geometrical features. In-vitro and in-vivo EM simulations are performed with device mesh information and peak SAR values extracted. Incident Electric field information and mesh information are combined to form the input of the CNN models. After training and testing, CNN model convergence and data correlations are observed as a metric of CNN general prediction efficacy. The distribution of absolute errors and absolute percentage errors are shown to further investigate the prediction performance for different data. For the selection of training dataset, naïve strategy is initially introduced which uses different sizes of training dataset to find the training dataset with good prediction performance and least data possible. Principal Component Analysis (PCA) is performed on the input matrices of CNN

model, which provides a threshold as the best prediction performance a training dataset can achieve with the dataset size same with the number of top-ranking Principle Components (PCs). Overall, mesh-based CNN models can predict the RF-induced heating of some complex-shaped PIMDs with acceptable performance. With the guidance of PCA analysis, the optimal size of training dataset can be determined before the simulations are performed which can save a lot of time used by obtaining excessive simulation results.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Implantable medical devices are devices that are placed inside or on the surface of the body which has a wide variety of applications on clinical treatment [1]. Among them, Passive Medical Devices (PMDs) are medical devices that serves its function without supply of electrical energy or any source of power other than that directly generated by the human body or gravity. The implantable PMDs, namely Passive Implantable Medical Devices (PIMDs), are passive devices that are totally or partially introduced, surgically or medically, into the human body [2]. One of the biggest features PIMDs have is that they don't have the external power supply. Commonly used PIMD types includes orthopedic implants which deal with the bone fractures, stents which are used for opening space in vessel, catheters which are used for the transmission of liquid inside the body and other types.

Magnetic Resonance Imaging (MRI) is one of the most prominent medical imaging techniques. Generally, MRI system uses large magnets which aligns all protons when a patient lies into the MRI coil. Different brightness is expected when protons in different tissue re-align which forms the MRI image. MRI is widely used and has the advantage of no ionizing radiation which is the principle of computed tomography (CT) scan or X-ray.

However, MRI is not risk-free for everyone. Patients with active or passive implantable medical devices inside the body will encounter safety hazards during the MRI examinations. Those safety hazards include the heating, torque or device malfunction caused by the gradient coil, RF coil or the static magnetic field (B_0) of MRI system [3]. For patients with PIMDs implanted, Radiofrequency (RF)-induced heating is one of the major concerns. The RF field in MRI system generates a magnetic field which is

perpendicular to the main magnetic field near the Larmor Frequency to create a change in proton alignment. Generally, the metallic part of the PIMDs will interact with the RF fields which will cause the power deposition near the edge of the devices [4]. This power deposition will cause temperature rise and will probably cause irreversible body tissue damage to the patients.

Thus, the evaluation of the RF-induced heating is a crucial work for every PIMD. Normally, a significant amount of experiment tests is expected in an in-vitro environment to obtain the temperature rise or Specific Absorption Rate (SAR) of a PIMD during a period under MRI RF field exposure [2]. Since PIMD has many types with different geometrical shapes and configurations, it is unrealistic to test every device experimentally as such is a great consumption of time and labor. To solve this, Electromagnetic (EM) simulations are performed which can obtain the SAR value without the experimental setup for various device configurations [5]. While EM simulations can ease the complexity of experimental setup, large amount of time is still expected for running all the simulations.

Recently, with the development of High Performance Computing (HPC), deep learning techniques are becoming more popular and applicable for classification and regression applications. Among all the deep learning models, neural network is one of the most famous models which utilizes the analogy of human brain neuron cells [6]. The network is composed of many neurons that apply arithmetic operations to the input. Prior knowledge is needed for the update of the network parameters until low loss is achieved. Later, Convolutional Neural Network (CNN) is invented with backpropagations as network update which has the unparalleled performance on image classification [7]. For RF-induced heating evaluation, since a lot of PIMDs devices have similar structures, a

potential idea is that a neural network model can be used with part of the simulated RF-induced heating data as training dataset to predict the rest RF-heating data.

In this research, a convolutional neural network model is proposed to predict the RF-induced heating of complex-shaped PIMDs. Computational models of PIMDs are developed in simulation software and the geometrical shape features are kept as much as possible. Since CNN architecture normally has 2-D images as input, the meshes of the device models are extracted and sliced to multiple 2-D input images. Incident Electric field information is extracted as well and is combined with the device mesh information. Peak spatial-averaged SAR values are selected to represent the RF-induced heating and are treated as the output of the network. With training of network, the convergence and the overall error level of the proposed CNN model are evaluated. The distribution of the error function for training dataset and testing dataset are also investigated.

2. Problem Statement

For all PIMDs, there exists possibilities that the devices may be used near the MRI environment. Thus, MR safety labeling is needed for every PIMDs [8]. Devices that cannot enter the MRI scanner room should be labelled as MR Unsafe and patients with these devices implanted cannot be under MRI scan. Devices that have no safety hazards under MRI environment should be labelled as MR Safe. The third label is MR Conditional which most of the PIMDs are labelled as. Multiple conditions are included for these PIMDs when they intend to enter the bore of the MRI system. So, the evaluation of RF-induced heating is one of the required procedures during the labelling of MR Conditional. During the evaluation of RF-induced heating of PIMDs, a comprehensive evaluation method is applied which includes EM simulations and experimental measurements. Multiple EM simulations is one of the important steps in the evaluation method which consists of in-vitro simulations and in-vivo simulations. In in-vitro simulations, PIMDs are placed inside an ASTM phantom which is defined in standard [2]. Computational human models are used instead in in-vivo simulations to hold the PIMDs. Due to the large variations on evaluation configurations including device types, landmark positions, device positions and so on, the amount of in-vitro and in-vivo simulations needed is large, which will cost a lot of time and computational resources.

Based on the circumstance, some prediction models are needed for predicting the RF induced heating with the knowledge of simulated heating data to reduce the number of simulations performed. Neural network is one of the prominent models used for PIMD RF-induced prediction. In some previous literatures, ANN models are firstly applied to predict the RF-induced heating of generic plates and stents. Besides, CNN models are applied to

implement the RF-induced heating prediction with the introduction of device 3-D meshes as the network input.

However, there still exists some restrictions on the current neural network prediction methods. For ANN models, the input layer is formed by numerical parameters that characterize the device features. Generic PIMDs can be easily described using several parameters including the length, width, depth and screw related parameters. While for commercially available PIMDs, the shape is more complicated which cannot be described using merely several parameters. Besides, previous literature with CNN prediction also uses generic PIMD to extract the device meshes, which cannot prove the robustness of the CNN prediction on complex-shaped PIMDs.

Moreover, previous literatures did not raise the discussion of how to select the training dataset size. This is an important consideration for Electromagnetic problems. For conventional CNN applications such as image classification, the acquisition of data and their corresponding label is easy since there a large number of online resources that can be easily reached and downloaded. So, the size of training dataset doesn't need to be considered as large number of images can be obtained with little effort for training the network. While for EM problems, the situation is different. The RF-induced heating cannot be collected but to finish the associated EM simulation, which means that the number of data contained in training dataset is equal to the number of EM simulations performed. The optimal training dataset for RF-induced heating should contain as less data as possible while still receive good prediction result when used for network training.

Based on the illustrations above, the RF-induced heating prediction of complex-shaped PIMDs using CNN needs to be validated:

i. Can CNN successfully predict the RF-induced heating of the complex-shaped PIMDs? If so, how will the network convergence and error level be?

ii. What will be the optimum architecture of the CNN model? What will be the appropriate network input and output pattern?

iii. What is the philosophy of data selection for CNN training and testing? What is the appropriate data amount for the training dataset?

3. Preliminary Literature Review

According to ASTM F2182-19^{e2} standard, the RF-induced heating of PIMD is measured with the PIMD placed inside the ASTM phantom under MRI environment. The ΔT during certain time period is captured. Either ΔT or SAR value converted from ΔT is considered as the RF-induced heating value. Davis et al. investigated the RF-induced heating of small metallic implants using experimental measurements with an RF coil [9]. Later, clinical MRI system is used as the RF excitation to measure the heating of PIMDs [10]-[13]. With the introduction of EM simulations, numerical investigations can be conducted with experimental measurements to add more possibility on the RF-induced heating evaluation with more variations. Liu et al. investigated the RF-induced heating of orthopedic implants under 1.5T and 3T system [14]. Implants with different sizes are modelled using Finite-Difference Time-Domain (FDTD) EM simulation software. Experimental measurements of one implant is also performed and temperature rises are recorded to validate the SAR values are evaluated from the simulations.

With EM simulations, it is possible to explore the trend of RF-induced heating based on variations. The RF-induced heating of PIMD is dependent on many factors, which is divided in two aspects: the device intrinsic geometric parameters and the device loading conditions, which includes the variations on the excitations, device loading positions, etc. Ran et al. investigate the effect of device length on RF-induced heating of different plate systems under 1.5T and 3T environment [15]. Other PIMD types, such as external fixation devices [16]-[18] and stents [19]-[20] are also addressed based on variations of device shape. Other literatures focus on investigating the effect of loading conditions on RF-induced heating. Lucano et al. discussed the EM field pattern generated by different RF

coil types [21]. Yang et al. investigated RF-induced heating differences of multiple PIMDs with generic birdcage coil and TEM coil [22]-[24]. Beside the coil structure, Xia et al. focus on investigating the effect of in-vivo device loading conditions which including human model orientations, landmark positions and device locations on the RF-induced heating of the external fixation devices [25]. Other literatures also discussed the RF-induced heating scenario with different PIMD types [26]-[27].

With all the previous literature based on exhausting all the possible configurations with in-vitro or in-vivo simulations, Zheng et al. firstly used a fully connected artificial neural network (ANN) model to predict the peak spatial 1g-averaged SAR value of the generic plate devices [28]. Optimization algorithm is used to improve the initial network weights and biases to reach better network convergence [29]. For predicting RF- induced heating of complex-shaped PIMD models, Lan et al. propose a ANN model which takes 20 device dimension and surface-based parameters as input [30]. The model reaches a root mean square of 6.21W/kg for 1-gram averaged SAR with an averaged value of 119.6W/kg, which is a low prediction error level. However, the number of data used is limited making the network easier to converge. The mesh-based CNN model was firstly designed which extracted the mesh of the generic plate device systems that have variations on device and screw sizes to predict the peak SAR values [31]. The previous literature proved that the meshed-based CNN model can be used to predict the RF-induced heating of the generic PIMDs, whereas the robustness of the model on the complex-shaped clinical-based PIMD models needs to be validated.

4. Objectives

In this research, the ability of CNN models to predict the RF-induced heating of complex-shaped PIMDs are investigated. Computational models based on the Tibia Plating System and Spinal Fixation System are developed as the complex-shaped models. Variations on the device configurations are considered such as device length, screw length, screw positions, etc. In-vitro and in-vivo simulations with all models are performed with commercial FDTD simulation software. CNN models based on typical CNN architectures are used in this research. The mesh information and incident electric field information are extracted from simulations and are combined as the input of the CNN model. The prediction efficacy of the proposed CNN model is investigated. After that, the strategies of selecting optimal training dataset is compared and discussed.

5. Mechanism behind Neural Network, RF-induced Heating and Principal Component Analysis

5.1. Mechanism behind Neural Network

Artificial Neural Network (ANN) is one of the widely used deep learning models. It is invented based on the idea of human neuron cells. A neuron cell includes several key components: the cell body containing nucleus and other structures, many branched extensions called dendrites and one long extension called axon. When a neuron is activated, the electric impulse will be generated from the neuron cell body. The signal transmits through the axon to reach the edge where the dendrites of other neurons are connected. The biological neural network in human brain contains many layers with neurons in every layer connecting with other neurons with previous and next layer. Neurons in each layer receives signal from the previous layer, process the signal and transmit the signal to the next layer.

In ANN models, the basic components are artificial neurons which process numerical values instead of electric signals. The output of the neuron is calculated using [32]

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W}^T + b) \tag{1}$$

where

$$\mathbf{X} = \left[x_1, x_2, ..., x_n\right] \tag{2}$$

and

$$\mathbf{W} = \left[w_1, w_2, ..., w_n\right]. \tag{3}$$

X is the input vector containing all the input to the neuron, **W** is the weight vector containing all the weights that are associated with each input and *b* is the neuron bias term. ϕ is the activation which provides the non-linearity to the neuron.

Normally, ANN models consist of fully-connected layers in which every neuron in two adjacent layers are connected with each other, which is shown in Figure 5.1. However, the fully-connected structure has some restrictions. The neurons cannot capture the correlation between different inputs as the connection between inputs and neurons are independent. Also, since every neuron weight is for every connection, there is a large number of weights used and those weights cannot be reused for similar features from input.

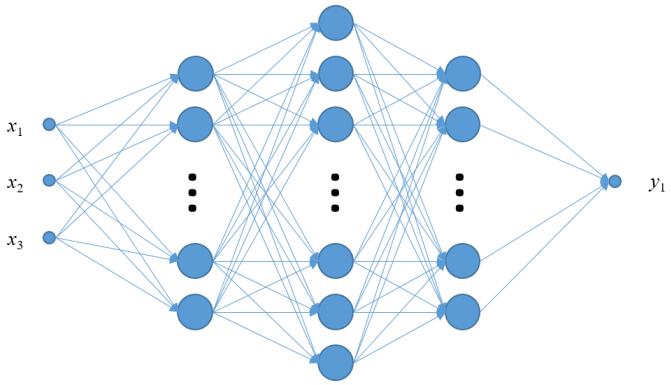


Figure 5.1 Three layer ANN model example consisting of fully-connected layers

In order to solve these restrictions, Convolutional Neural Network (CNN) is invented. The network architecture was initially used in image recognition applications in which the features are extracted based on small areas on the input images. CNN utilizes the advantage of shared filters through the operation of convolution which can also capture the geological domain features of the input image.

The convolutional layer is the core structure in a CNN model. For 2-Dimension (2-D) convolution, input pattern is a 3-D matrix containing multiple 2-D images representing different channels. For color images, there will be three channels which are red, green and blue channel. The small-size 3-D matrix used for extracting regional features is called filter. The depth of filters is the same with the number of input channels. During the convolution operation, each filter slides on the input along the first two dimensions. The overlapped

filter values and input values are multiplied and summed together. The multiplication sum is added with filter-specific bias value to form the corresponding value of output layer. As a filter is moving on the input layer, a 2-D matrix is obtained as a sublayer of output which is called feature map. After the convolution, the number of feature maps generated is the same with the number of filters used. These 2-D feature maps forms the 3-D input layer for the next operation. The illustration of the convolutional operation is shown in Figure 5.2.

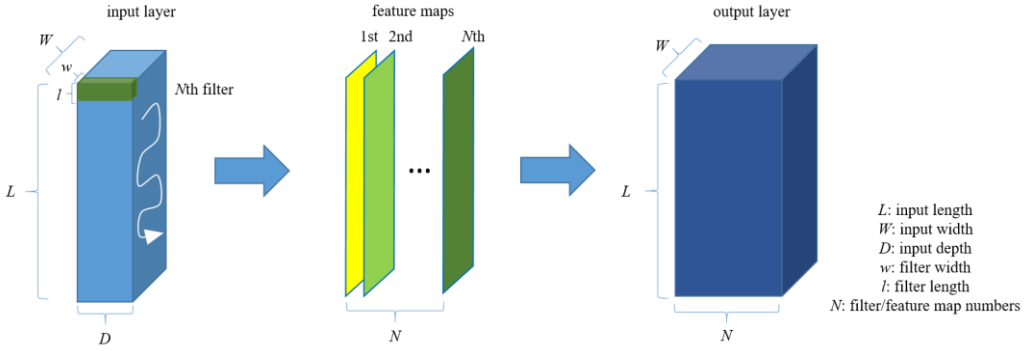


Figure 5.2 Convolutional layer operation

In default, the length and width of the output layer are smaller than those of input layer as the multiplication sum will only produce one result for all the values with the range of filter length and width. While in some cases, the output layer is required to keep the same shape with input layer in order to capture the features at edge. To solve that, paddings are applied which are extra columns and rows outside of the input layer. Usually, paddings contain only zero values. The illustration for padding is shown in Figure 5.3.

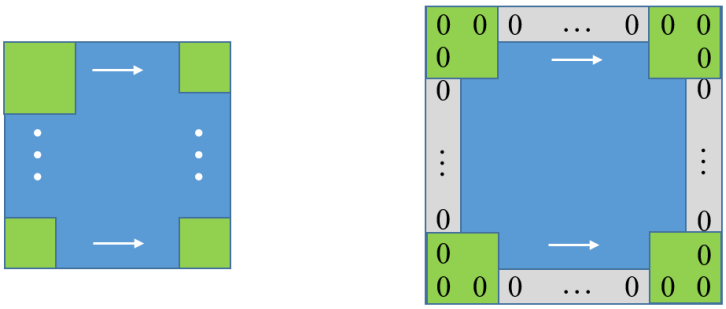


Figure 5.3 Padding operation

Another core structure in CNN is max pooling layer. The principle is simple: filters with similar shape of convolutional filters are applied on the input layer while max pooling filters don't have any weights. The filters pick and keep the maximum value of input layer within the filter area. With this maximum pick, it can reduce the size of the 2-D image to release the redundant computational resources while retain the most dominant local feature of the input. The illustration of the max pooling operation is shown in Figure 5.4.

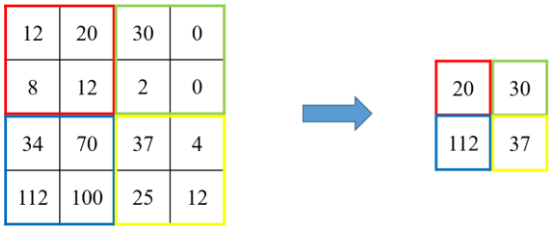


Figure 5.4 Max pooling layer operation

After several convolutional layers and max pooling layers, the data size is reduced and significant features are obtained. After this, a flatten layer is needed. The flatten layer reshape the 3-D output matrix from previous layer to 1-D vector. Since the features extracted don't include geological information, it is not necessary to keep the 3-D shape and no convolution operations are needed. Thus, after the flatten layer, fully-connected layers are applied to create the mapping between features extracted from image and output.

The CNN model is a data-driven model, which means that it needs some data to update the filter weights to decrease the prediction error of the model. The procedure is as follows: Firstly, the training data are divided into small portions which is named batches. A batch of data is fed to the input layer. Every layer receives its input, conducts all the operations, calculates the output and passes it to the next layer until reaching the output layer. This is called forward pass. Then, the averaged value of loss function is evaluated after predicted output values are obtained for the batch to find the error level of the network prediction. After that, the contributions of each output connection to the error are calculated using chain rule. The contributions of connections between layers are also evaluated reversely from output layer to input layer, which is called backpropagation. Lastly, the parameters and the filter weights are tweaked based on the contribution of each connection. When this full procedure is finished for one batch, all the network parameters are updated and same steps are repeated for a new batch to calculate the updated upper level. The time period that all the training data have gone through the procedure once is called an epoch.

Among the training procedure, backpropagation is the most important step as it provides the right direction for weights updates to keep reduce the prediction error. After backpropagation, gradient descent algorithm is normally used for updating filter weights. Based on the algorithm, the updated weight is evaluated as

$$w = w - lr * \frac{dLF}{dw} \tag{4}$$

where w is the filter weight, lr is the learning rate which controls the step to update the weight and LF is the loss function. During the training procedure, the learning rate can be variant. The strategy of designing the learning rate value along training is called the

optimization. CNN model with different optimization algorithms will have different network convergence which means the error level will decrease differently through epochs.

At the beginning of training, the initial filter weights are not zeros since it will cause the failure on the evaluation on gradients. In this research, the Glorot Uniform Initialization [33] is used to initialize all the filters. With this method, all weights are randomly picked from the uniform distribution of interval $\left[-\sqrt{\frac{6}{fan_{in}+fan_{out}}}, \sqrt{\frac{6}{fan_{in}, fan_{out}}}\right]$, where fan_{in} is the number of input units and fan_{out} is the number of output units.

The whole dataset needs to be divided as training dataset and testing dataset. The training dataset is for updating the filter weights to reach lower loss function values while the testing dataset is for evaluating the convergence and prediction performance of the network. Also, within training dataset, a validation dataset is needed as the behavior of predicting on new data during network training should be monitored.

5.2. Mechanism behind RF-induced heating

The RF-induced heating is associated with RF power deposition near the PIMD. To quantify the power deposition, temperature rise (ΔT) or Specific Absorption Rate (SAR) is used [2] . Normally, these two parameters can convert to each other with proportional relationship. SAR is used to represent the RF-induced heating in this study. SAR is defined by the energy absorbed per unit mass, which is expressed using electric field as

$$SAR(\mathbf{r}) = \frac{\sigma(\mathbf{r})}{2\rho(\mathbf{r})}\left|\mathbf{E}(\mathbf{r})\right|^2 \tag{5}$$

where $\mathbf{E}(\mathbf{r})$ is the local electric field, $\sigma(\mathbf{r})$ is the electrical conductivity of the unit mass and $\rho(\mathbf{r})$ is the density of the unit mass. The unit of SAR is W/kg. This formula defines the local SAR, which is the SAR value at one single point. However, this value has little

significance for representing the power deposition, since it is too sensitive and may have drastic change along different locations. Moreover, the energy absorbed will be conducted to the nearby tissue, making the single point SAR unable to estimate the energy absorbed by surrounding tissue. To solve that, the averaged SAR value is introduced as the power deposition standard. The averaged SAR value inside a certain mass or certain volume is evaluated as

$$\langle SAR \rangle_M = \frac{1}{M} \int_{R(M)} SAR(\mathbf{r}) dm \tag{6}$$

and

$$\langle SAR \rangle_V = \frac{1}{V} \int_{R(V)} SAR(\mathbf{r}) dv . \tag{7}$$

The averaged SAR is evaluated in a Region of Interest (RoI) of mass or volume. With different RoI used, the averaged SAR values have different significance. When the hole human body region is set as RoI, whole body SAR (*wb*SAR) is obtained which is used to estimate the power level of the total RF exposure. According to IEC-60601-2-33 [34] , the maximum limit for *wb*SAR is 2W/kg. This limit is also helpful for the power normalization for RF coil in EM simulations.

If the Finite-Difference Time Domain (FDTD) EM simulations are used for PIMD, the cubic RoI is used to evaluate the averaged SAR as the cubic mesh grid is preset for the simulations. The RoI will be one-gram or 10-gram cube containing lossy materials whose edges are aligned to the mesh grid [35]. Normally, the peak value of the averaged SAR is the object of interest which is used to represent RF-induced heating.

5.3. Mechanism behind Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction algorithm which extracts the components that contain the highest variance of the source matrix [36].

It is an unsupervised algorithm with no training data needed and is applied for many applications such as feature selections, noise filtering, etc. For mesh-based CNN models, device mesh matrices are one of the crucial information to form the model input. While the mesh information of different configurations under same PIMD are similar geometrically, PCA analysis can be performed on the device mesh to obtain the Principal Components (PCs) which provides implications on which mesh regions contain most of the device geometry information.

Assume that the 3-dimensional (3-D) device mesh matrix is reshaped to 1-D $1\times p$ row vector. Then, mesh information of every model is concatenated is a $n\times p$ matrix, where n is the number of PIMD models included in the study and p is the total number of pixels. The total data matrix is expressed as

$$\mathbf{X}=[x_1,x_2,...x_n]^T \tag{8}$$

where $x_1,x_2,...x_n$ are the 1-D mesh vectors of the device models. The PCA is conducted as follows [37]: Firstly, every element inside the matrix is subtracted with the column mean in order to every column with a mean of zero. This step ensures the effect of original value magnitude is eliminated. Secondly, the covariance matrix of the data matrix is evaluated. The covariance matrix is defined as

$$Cov=\begin{bmatrix} \mathrm{cov}(x_1,x_1) & \mathrm{cov}(x_1,x_2) & ... & \mathrm{cov}(x_1,x_n) \\ \mathrm{cov}(x_2,x_1) & \mathrm{cov}(x_2,x_2) & ... & \mathrm{cov}(x_2,x_n) \\ ... & ... & ... & ... \\ \mathrm{cov}(x_1,x_n) & \mathrm{cov}(x_n,x_2) & ... & \mathrm{cov}(x_n,x_n) \end{bmatrix} \tag{9}$$

where $\mathrm{cov}(x_p,x_p)$ is the variance square of vector x_p and $\mathrm{cov}(x_p,x_q)$ is the covariance between vector x_p and x_q. After the $n\times n$ covariance matrix is obtained, the eigenvectors

and eigenvalues of covariance matrix are calculated. Normally, Singular Vector Decomposition (SVD) algorithm is used to implement the calculation. After calculation, *n* eigenvectors and eigenvalues are obtained as

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n] \tag{10}$$

and

$$\boldsymbol{\lambda} = [\lambda_1, \lambda_2, ..., \lambda_n]. \tag{11}$$

Assume the eigenvectors and eigenvalues are ranked in descending order based on the magnitude of eigenvalues, then the eigenvector with the highest eigenvalue is the first PC that contains the most information. The PCs obtained are linear combinations of the original variables and every original 1-D vector can be express as the weighted sum of the PCs [38]. For example, the mesh of a device model is built by the combinations of pixels, which is express as

$$image(x_1) = w_1 \cdot (\text{pixel } 1) + w_2 \cdot (\text{pixel } 2) + ... + w_p \cdot (\text{pixel } p) \tag{12}$$

where x_1 is the first row vector containing the mesh information, which is

$$x_1 = [w_1, w_2, ..., w_p]. \tag{13}$$

After the PCA, the device mesh can be reconstructed by the linear combinations of highest-rank PCs, which is expressed as

$$image'(x_1) = \text{mean} + w_1' \cdot (\text{PC } 1) + w_2' \cdot (\text{PC } 2) + ... + w_m' \cdot (\text{PC } m). \tag{14}$$

Note that, the reconstructed mesh using PCs is not exactly the same with the original mesh but contained most of the original information. Originally, every pixel in the mesh is independent which means that almost all pixels have to be included to include the device geometry. With PCA conducted, only a few number of PCs needed to cover the essential device mesh information. When top-ranked PCs are reshaped back to 3-D matrix, each element value is proportional to the general information this element contains about

the feature of this PIMD, which can be used as a guidance for the training dataset selection for CNN models.

6. RF-induced Heating Prediction of Tibia Plating System

6.1. Model development

In this research, the computational models are developed from the distal tibia plating system from the commercial PIMD manufacturer, as shown in Figure 6.1 [39]. The distal tibia plating system provide the treatment for the distal tibia bone fracture. When the devices are implanted around the tibia bone, they add the fixation to the bone. Additionally, locking screws will be applied at certain positions on the plate to provide extra stability and rigidity. The computational models of the plate system are developed directly in SEMCAD X (14.8.1, SPEAG, Zurich, Switzerland), which is a commercial EM FDTD software. The computational models are formed by primitive 3-D shapes provided in the software. In order to make the computational models close to the real devices, some edges of the models are blended and the whole models are bent with different angles on three axes. The front, side and top view of the computational models are shown in Figure 6.2.

Figure 6.1 Four types of tibia plate system: Anterolateral, Medial, Anterior and Posterior

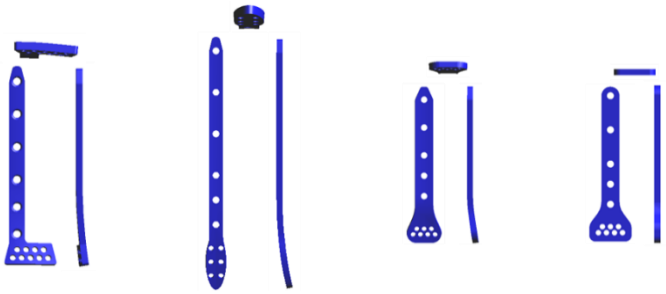


Figure 6.2 Computational models developed in SEMCAD

Among all the plate types, four of them are chosen as the template to develop computational models in this research: Anterolateral, Medial, Anterior and Posterior, which is shown in Figure 6.2. Four plate types are implanted to the tibia bones at different sides based on the bone fracture occurrence. The top prolonged part of the plates is defined as top frame as it provides most fixation and the bottom part is defined as bottom frame which usually are anatomically contoured around the bone to decrease the soft tissue irritation caused by the top frame. For the Anterolateral plate, there are two plate subtypes: left and right, which the top frame will be along the tibia bone and lean slightly to the left or right. Medial plates also consist of left and right plates which will be placed on the inner side of left or right bones. Anterior and posterior plates are normally located at the front and back side of the tibia bone and do not have left and right subtypes.

Due to different bone fracture circumstances, every tibia plate type has multiple length settings. Anterolateral plates have the biggest variations which ranges from 66mm to 244mm. Medial plates have the minimum length of 103mm and the maximum length of 278mm which is the largest maximum through all four plate types. Anterior and posterior plates have smaller length, ranges from 62mm to 130mm. Different lengths correspond to different number of screw holes on the top frame. Since the screw hole numbers are predefined, the available plate lengths are also fixed, which is tabulated in Table 6-1. The screws that are applied on the plates have two types: Cortical screws are applied on the top frame screws holes which have a diameter of 3.5mm and Cancellous screws are applied on the bottom frame screw holes which have a diameter of 2.7mm. The location of screws holes, top frame and bottom frame are shown in Figure 6.3. The length of screws ranges from 10mm to 60mm to provide flexibility on stabilization. Clinically, the screws are

applied at different positions depending on the need, while some common rules still need to be satisfied. On top frame, one screw must be applied on the positioning hole, which is for determining the implantation position of the plate device. Apart from the screw on positioning hole, one or two more screws are applied on the top frame. On bottom frame, more than half of the screw holes are filled with screws in order to guarantee the stability. In this research, 2 screws are applied on the top frame, one is located at the positioning hole and the other is located at either the highest screw hole or the screw hole at the middle of top hole and positioning hole. The bottom screw configurations are designed while the number of screws ranges from N/2 to N (N is the total number of holes at bottom frame).

Table 6-1 Lengths of tibia plates (Unit: mm)

Number of Cortical screw holes	Anterolateral	Medial	Number of Cortical screw holes	Anterior	Posterior
4	66	103	3	62	57
6	91	128			
8	117	154	5	84	77
10	142	179			
12	168	204	7	107	98
14	193	230			
16	218	278	9	130	118
18	244	/			

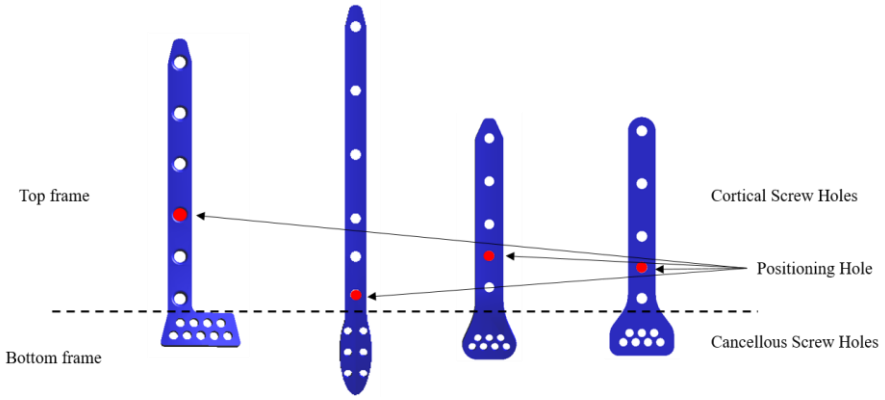


Figure 6.3 Screw holes of tibia plate models

Technically, every screw that are applied on the plate can have different lengths. In this research, 4 kinds of screw lengths configurations are selected. Three of them have all screw lengths the same of 10mm, 30mm and 60mm which corresponds to the minimum length, median length and the maximum length. For the rest configuration, every screw randomly selects a length from 10mm, 30mm and 60mm, which is defined as “pseudo-random” configuration, as shown in Figure 6.4.

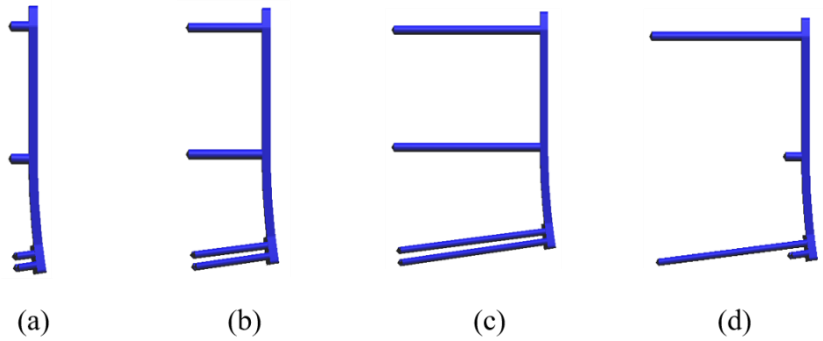
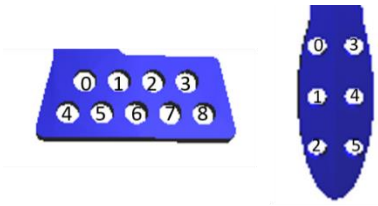


Figure 6.4 Screw length configurations: (a) 10mm, (b) 30mm, (c) 60mm, (d) pseudo-random

Based on all the variations on plate types, plate length and screw configurations, a total number of 1416 plate configurations are created, which is tabulated in Table 6-2. Bottom frame screw holes are indexed to be better expressed for all screw configurations used in this research, as shown in Figure 6.5. In next part, each plate is placed inside the ASTM phantom with RF coil and in-vitro EM simulations are performed. The meshes of the devices and the peak averaged SAR value will be extracted, which forms the dataset for training and testing of the CNN model.

Table 6-2 Total plate configurations

Plate type	Plate length(mm)	Screw length(mm)	Main frame screw configuration	Bottom frame screw configuration
Anterolateral	[66,91,117, 142,168,193, 218, 244]	[10,30,60, pseudo-random]	[[top hole, positioning hole], [mid hole, positioning hole]]	[[0,1,3,4,8],[0,2,3,4,8],[0,3,4,5, 8],[0,3,4,6,8],[0,3,4,7,8], [0,1,2,3,4,5,6,7,8]]
Medial	[103,128,154, 179, 204,230,278]			[[1,2,4,5],[0,2,3,5],[0,1,3,4],[0, 1,2,3,4,5]]
Anterior	[62,84,107, 130]			[[0,2,3,6],[0,1,2,3,6],[0,2,3,4,6],[0,2,3,5,6],[0,1,2,3,4,5,6]]
Posterior	[57,77,98, 118]			[[0,2,3,6],[0,1,2,3,6],[0,2,3,4,6],[0,2,3,5,6],[0,1,2,3,4,5,6]]



(a) (b)

Figure 6.5 Low frame indexing strategy: Anterolateral, Anterior and Posterior(a), Medial(b)

6.2. Simulation Settings

After developing all the plate configurations, in-vitro simulations are designed and performed for collecting the dataset. All simulations are performed in the FDTD EM full wave simulation software SEMCAD. A generic birdcage coil is used as the RF excitation in the simulation, which is generally used in the in-vitro simulations to replace the physical coil model. The coil consists of eight current sources that serves as eight rungs of the coil and sixteen lumped elements to serve as the end rings at the top and bottom. The working frequency of the coil is 64MHz, which corresponds to the Larmor Frequency of the 1.5T main field strength. An ASTM phantom that is designed following ASTM F2182 standard [2] is placed inside the RF coil to load the devices. The phantom is filled with standard

conductive gel that has a depth of 90mm with relative permittivity of 80.38 and electrical conductivity of 0.47 S/m. The devices are placed inside the phantom and are immersed by conductive gel such that they are located at the center depth of the gel. All devices will obey the same rule for positioning inside the phantom that the devices are located at zero landmark position which stands for center z axis position and the devices are close to the right inner wall of the phantom with a fixed distance of 20mm, where the largest E field exposure is expected. Devices have the orientation that all screws point to the negative x direction. The device position inside the phantom is shown in Figure 6.6.

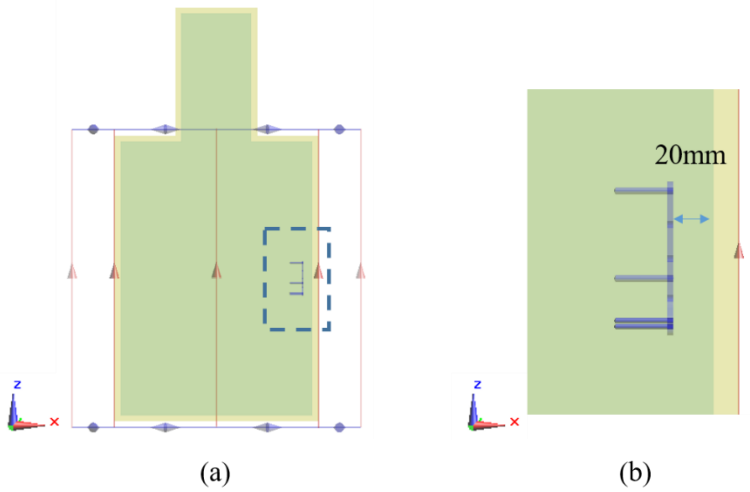
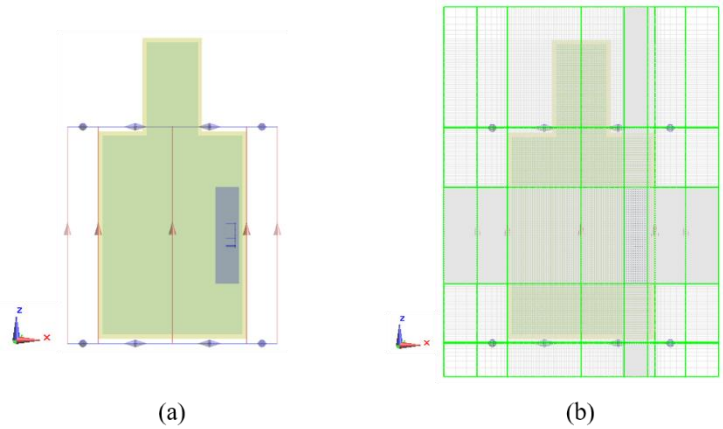


Figure 6.6 Device position inside phantom (a), details (b)

An FDTD harmonic simulation at 64MHz is set for every device inside the phantom. Total period is set to 20 as the current source generates sinusoidal wave. The ASTM phantom uses the electrical property of the acrylic material (ε=3.7, σ=0) and the devices will be defined as Perfect Electric Conductor (PEC). At outside of the coil, Absorbing Boundary Condition (ABC) are set for all boundaries to prevent the significant effect of reflection wave. The global mesh grid setting is defined before the simulation. For background mesh, a padding of 100mm is added outside the coil. The phantom mesh has a

resolution of 10mm and the conductive gel has a resolution of 5mm. Due to the different length profiles of different plate types, mesh setting of the plates may cause the inconsistence of the global mesh size, which will cause problems during mesh extraction. In solution of that, a grid mask is applied to wrap outside of the devices. The grid mask has a size of 70mm*50mm*290mm which is large enough to cover all the devices in this research. The grid mask has a resolution of 0.8mm on x, y and z directions, which will produce the same mesh for all the devices. Totally, the global mesh has a size of 190*108*511= 10.48M, as shown in Figure 6.7.



(a) (b)

Figure 6.7 Grid mask (a), global mesh setting (b)

All the simulations will be performed on a computer with GPU hardware acceleration (NVIDIA® GeForce GTX 1080). CUDA software acceleration tool which is embedded into SEMCAD is also applied. With them, the time duration for one simulation is thirty-five minutes, which makes the total simulation duration 826 hours.

After all the simulations are completed, postprocessing work is needed to obtain the mesh and the RF-induced heating data for every device in preparation of CNN model training. For device mesh, the mesh grid of grid mask is extracted in every simulation. The original grid is a 3-D matrix with a size of 89*65*364≈2.1M cells. The value of the matrix

elements represents the index of the material inside the grid cell. Since only two materials appear inside the grid mask which are conductive gel and PEC, the values of matrix are redefined: the value for gel is 0 and the value for PEC is 1. One of the modified mesh matrices is shown in Figure 6.8 as x-z slices at different y values. Overall, 1416 mesh matrices are extracted from all simulations.

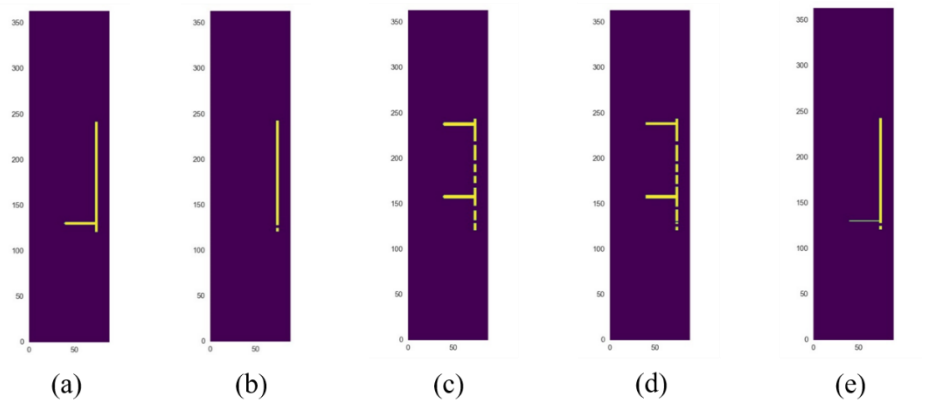


Figure 6.8 x-z slices of device mesh at y = -3.125mm(a), -2.34mm(b), -1.56mm(c), -0.78mm(d) and 2.35mm(e)

Meanwhile, the incident field information is extracted from one simulation as well. In this simulation, all the grid settings are the same with device simulations, except that there is no device model inside the phantom. The incident E field inside the grid mask box is exported. Since the largest dimension for place device models is z-direction, the incident E field component on z-direction is the dominating component on RF-induced heating. Thus, only z-direction incident E field is retained to represent the incident field. The simplified E field is express as

$$\mathbf{E}_{inc}(i,j,k) \approx \vec{z} \cdot E_{incz}(i,j,k) \tag{15}$$

where $\mathbf{E}_{inc}(i,j,k)$ is the discrete incident E field at location (i, j, k) and $E_{incz}(i,j,k)$ is the z-direction component of the discrete incident E field at location (i, j, k). $\mathbf{E}_{inc}(i,j,k)$ is also a 3-D matrix with a shape of (89,65,364) which is the same with material index matrix.

In order to combine the mesh information and incident E field information inside the mask box, elementwise multiplication is performed between $\mathbf{E}_{inc}(i,j,k)$ and each index matrix, which is shown as

$$E_{mesh}(i,j,k)=\mathbf{E}_{inc}(i,j,k)\times Mesh(i,j,k)=\begin{cases}E_{incz}(i,j,k), PEC\\ 0, Non-PEC\end{cases} \tag{16}$$

where $Mesh(i,j,k)$ is the material index matrix. After the operation, 1416 E_{mesh} matrices are generated.

To reduce the computation burden for CNN training and PCA analysis, sampling is conducted on each E_{mesh} matrix on all three axes. One element of every two are picked and retained which makes the new shape of E_{mesh} matrix (45,33,182). Also, since normal CNN model cannot deal with complex values, E_{mesh} matrices need to be converted to real-valued matrices so that they can be processed by the CNN model. Firstly, the real and imaginary part of each E_{mesh} matrix is split into two matrices E_{meshRe} and E_{meshIm}. Then, two matrices are concatenated on the first dimension, which is the x direction, which is expressed as

$$\begin{aligned}E_{mesh}(2*i,j,k)&=E_{mesh\text{Re}}(i,j,k), i=0,1,2,...,48\\ E_{mesh}(2*i+1,j,k)&=E_{mesh\text{Im}}(i,j,k), i=0,1,2,...,48\end{aligned}. \tag{17}$$

After that, the new E_{mesh} matrix has a shape of (90,33,182) and it is treated as the input of the CNN model.

In this research, only the peak value of the 1g-averaged SAR is picked and used as the RF-induced heating data, which is expressed as

$$psSAR_{1g}=\max\langle SAR\rangle_{1g}. \tag{18}$$

The sole $psSAR_{1g}$ value is extracted for each simulation using python scripts and postprocessing tool in simulation software.

When the raw $psSAR_{1g}$ values are extracted, the input power of the RF coil is set to 1W for every simulation. After that, power normalization is expected as coil in every simulation will have different input power. The power normalizations are performed such that the whole-body SAR (*wb*SAR) of each simulation is equal to 2W/kg.

With all the postprocessing work being done, 1416 3-D mesh matrices and normalized $psSAR_{1g}$ values are extracted from the in-vitro simulations. Then, these data will be treated as the input and the output of the CNN model to implement the training and testing of the network.

6.3. CNN architecture

A sequential Convolutional Neural Network (CNN) model is used for the prediction of RF-induced heating for the plate devices. The sequential model is derived from AlexNet [40], which is a typical CNN architecture. The architecture construction, network training and testing and the results evaluation are all implemented in Python with Keras [41], which is a high-level deep learning toolbox.

The visualization of the network structure is shown in Figure 6.9. The network consists of three 2-D convolutional layers, three max pooling layers, one flatten layer and three fully connected layers. Convolutional layers and max pooling layers operate on 2-D data which will be flattened to 1-D vector by flatten layer and then go through the fully connected layers. Since the mesh of the devices is a 3-D matrix, it is sliced along y axis to become multiple 2-D x-z matrix which corresponds to the multiple input channels for the

CNN input. The output layer only has one neuron whose output is the $psSAR_{1g}$ value of each plate device.

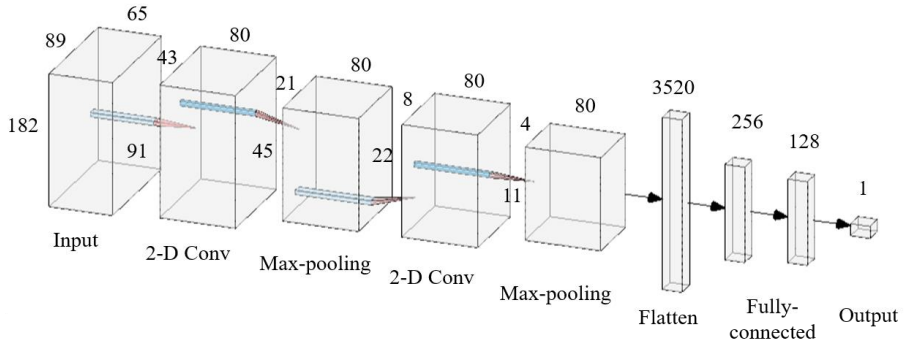


Figure 6.9 CNN model structure used in this research

Originally, the 3-D E_{mesh} matrix had a shape of (90,33,182). The matrix is then transposed to have its shape changed to (182,90,33). When the transposed matrix is considered as input layer in CNN model, the 3-D matrix is treated as 33 channels of 2-D input images with a shape of (182,90). The filters used for the convolutional layer have a width of 3 and a length of 3. Both convolutional layer has 80 filters. No paddings are applied outside all input layers so that the size of feature maps decreases to obtain the more important features. After each convolutional layer, 1 max pooling layer is applied. The filter has a width of 2 and a length of 2 so that the maximum of 4 values covered by the filter is retained.

The loss function determines which type of the error will be evaluated and how will the weights be updated. Mean absolute percentage error (MAPE) is selected as the loss function in this CNN model, which is defined by

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{19}$$

where n is a batch of data to be pushed into the network for one-time backpropagation. A_t is the actual value and F_t is the predicted value. The metrics of the model are the measures

that are observed through the network training which are good reflections of the network convergence and the prediction performance. Beside the MAPE, mean absolute error (MAE) is also selected as the metrics of the model which is defined by

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|A_t - F_t\right|. \tag{20}$$

For this CNN model, Adam optimizer is chosen to be the optimization which the learning rate value is decaying with the network training going [42].

Totally, 70% of the data, which is 991 data, is randomly selected as training data and the rest 30% of the data, which is 425 data, is selected as testing data. Within training dataset, 20% is selected as validation data whose loss function will be evaluated during network training. Both validation and testing data don't participate in the backpropagation and the filter weight update. Overall, the ratio of the training data, validation data and testing data are 56%:14%:30%. In this research, the epoch number is set to 100. Within the training dataset, all data will be divided into multiple batches which is the smallest unit for a one-time backpropagation. The batch size is set to 32 in this research.

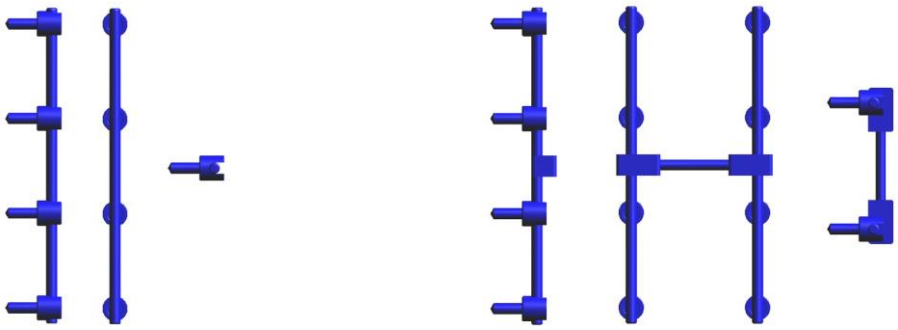
Also, PCA analysis is performed on the complex E_{mesh} matrix to find the optimum size of training set. Firstly, the 3-D E_{mesh} matrix is reshaped to 1-D vector with length of 540540. Then, 1416 vectors are concatenated to form the PCA matrix with the shape of (1416, 540540) which PCA analysis will perform on. After the analysis, the Cumulated Explained Variance (CEV) of the first n PCs are calculated. n- CEV(n) relationship is plotted afterwards and analyzed which shows the percentage of information covered by first n PCs. The number of PCs that have covered majority of the information will be the optimum size of the training set.

7. In-vitro RF-induced heating prediction of Spinal Fixation System

7.1. Model development

Spinal Fixation Systems (SFS) is a kind of PIMD that is designed to provide stability and rigidity to the spine that may be weaker than normal due to some diseases. Typical SFS consists of three components: rod, pedicle screw, and connector. In clinical procedure, pedicle screws are introduced at the pedicle area of different spinal levels. Rods are applied to connect the screws at the same side and connectors are placed between two rods to provide extra stability.

In this research, computational models are developed based on clinical SFS devices. Two types of SFS are selected to develop the models: Single Rod System (SRS) and Double Rod System (DRS). The front, side and top view of two SFS systems are shown in Figure 7.1.



(a) (b)

Figure 7.1 Single Rod System (a), Double Rod System (b)

SRS only has one rod and all the screws are distributed along the rod. DRS has two rods that are parallel each other with screws distributed on each rod with the sole connectors placed at the midpoint between two rods. Key parameters of SFS models are shown in Figure 7.2. Since clinical SFS devices have different length configurations, various parameter configurations are used to form SRS and DRS to ensure that more

clinical variations are covered. The configuration of the parameters are shown in Table 7-1. For connector length, 0mm means the model is a SRS device.

Table 7-1 SFS Component Configurations

Parameters	Range (unit: mm)
Screw Length	[10, 30, 50, 70]
Screw Diameter	[3.5, 4.5]
Rod Length	[80, 100, 120, 140, 160, 180, 200, 220]
Rod Diameter	[3.5, 5.5, 7.5]
Screw Spacing	[30, 50]
Connector Length	[0, 20, 40, 60]

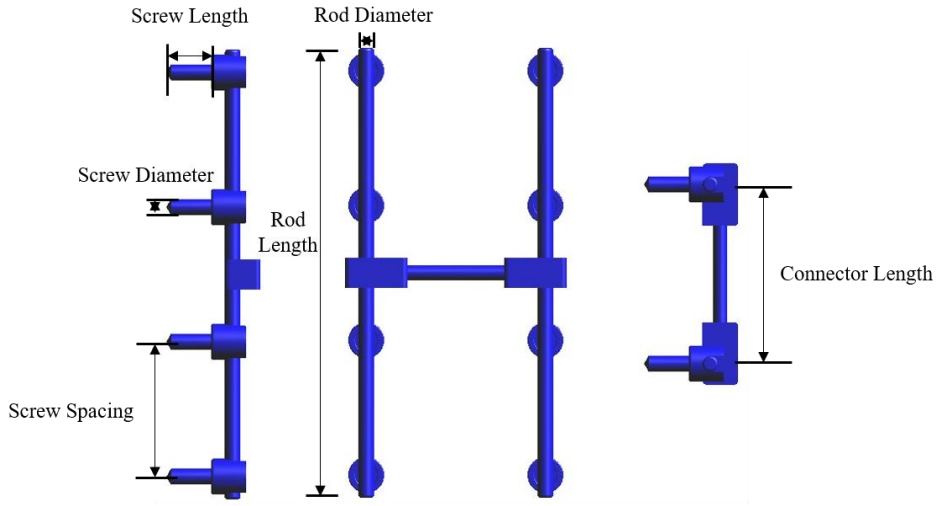


Figure 7.2 Parameters of SFS model

The number of screws is determined by the available parameter settings above, which can be calculated as

$$N = \begin{cases} RodLength \setminus ScrewSpacing, & if \ \ RodLength \bmod ScrewSpacing \neq 0 \\ RodLength \setminus ScrewSpacing - 1, & if \ \ RodLength \bmod ScrewSpacing = 0 \end{cases} \quad (21)$$

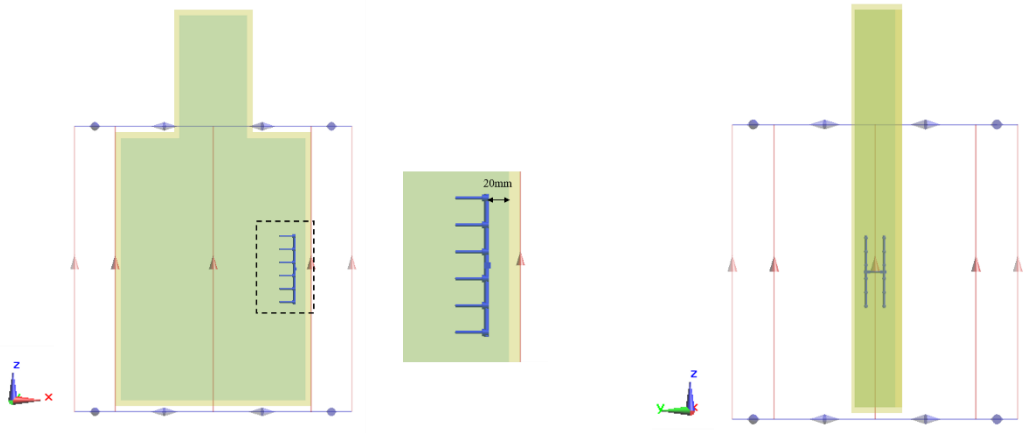
where N is the number of screws on each rod. The distance between the edge screws and the rod tip is also determined based on the parameters, which can be calculated as

$$d = (RodLength - (N-1) \times ScrewSpacing) \div 2. \quad (22)$$

Combining all the parameters listed above, 1536 computational models are developed totally.

7.2. Simulation Settings

All simulations are performed in SEMCAD X. A generic birdcage coil is used as the RF excitation which is composed of 8 current sources and 16 lumped elements. All current sources are working on 64MHz frequency and phase delays are set between each source so that the coil generates EM field with circular polarization. An ASTM phantom is used which is filled with conductive gel. The conductive gel has a relative permittivity of 80.38 and an electrical conductivity of 0.47S/m. Each of 1536 SFS model is placed in the phantom at a fixed location. On z direction, the rod midpoint is at z=0. On y direction, the rod is at the center depth of the gel. Also, the model has a distance of 20mm to the right inner wall of the phantom. The location of the device inside phantom is shown in Figure 7.3.

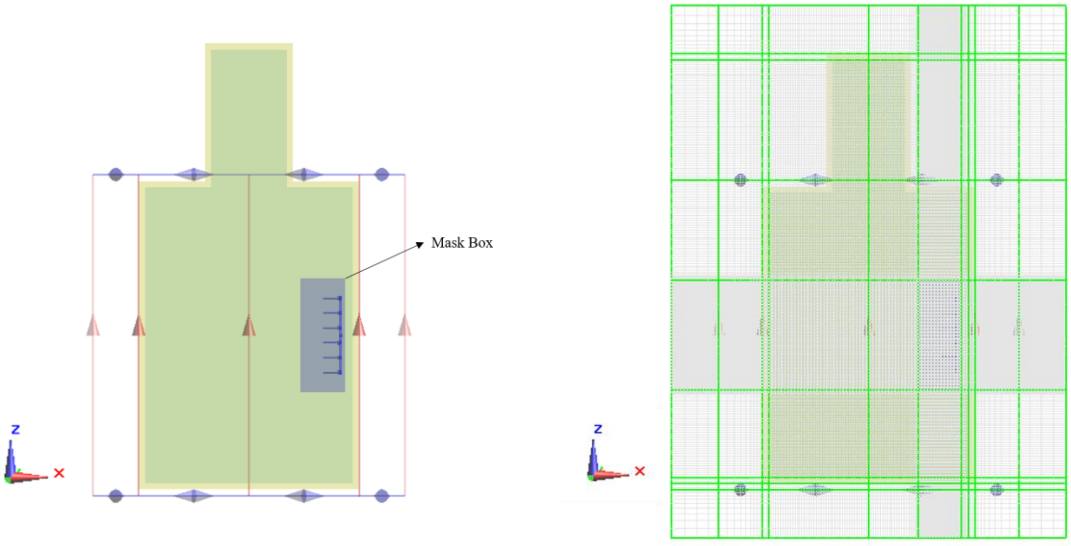


(a) (b)

Figure 7.3 Location of SFS device model inside ASTM

FDTD simulation is set for every simulation. The simulation duration is set to 20 periods in order to ensure the convergence. For material type, SFS model is set as PEC and

conductive gel is set as dielectric using the standard parameters (ε=80.38, σ=0.47). The phantom is also set as dielectric using the parameter of acrylic material (ε=3.7, σ=0). The boundary of the simulation is set to Absorbing Boundary Condition (ABC) with high strength to prevent the interference of probable reflection at the boundary. For grid setting, resolution of the conductive gel is set to 5mm and resolution of the phantom is set to 10mm. Since every SFS model has different size, to keep the same grid setting for every model, a grid mask box is applied outside of the model. The mask box has mesh resolution of 1mm and is not included in the simulation while its mesh resolution has higher priority than SFS models. Thus, All SFS models have 1mm resolution and the mesh coordinates are the same. Outside the RF coil, the 100mm padding is applied on x, y and z directions. The global mesh setting is shown in Figure 7.4.



(a) (b)

Figure 7.4 Mask box position and global mesh setting

After all settings, the total number of mesh cells is 197*131*397 ≈ 10.2454M. The number of cells inside mask box is 91*78*231 ≈ 1.64M. The original SFS model and discretized SFS model are shown in Figure 7.5.

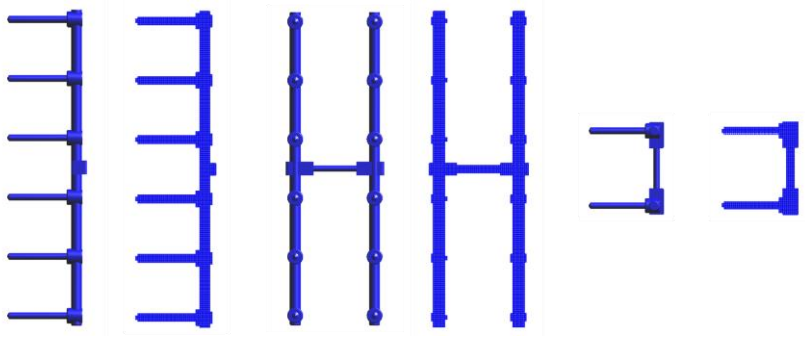


Figure 7.5 Front, side and top view of original SFS model (left) and discretized model (right)

After simulations are finished, the geometrical mesh information and incident E field information are extracted. The geometrical mesh extracted in the 3-D material index matrix of the material inside mask box area. Since there are only two materials inside the mask box, which are the device model and the conductive gel, the material indexes are redefined: the value for PEC is 1 and the value for gel is 0. Overall, 1536 material index matrices are obtained. Meanwhile, the incident E field information is exported from one simulation. In this simulation, grid settings are kept the same while no device models are placed inside the phantom. The E field inside the mask box is exported. Since for E field generated by RF coil, z-directional component is the dominant component, the incident E field can be approximated as

$$\mathbf{E}_{inc}(i,j,k) \approx \vec{z} \cdot E_{incz}(i,j,k) \tag{23}$$

where $\mathbf{E}_{inc}(i,j,k)$ is the discrete incident E field at location (i, j, k) and $E_{incz}(i,j,k)$ is the z-direction component of the discrete incident E field at location (i, j, k). $\mathbf{E}_{inc}(i,j,k)$ is also a 3-D matrix with a shape of (91,78,231) which is the same with material index matrix.

In order to combine the mesh information and incident E field information inside the mask box, elementwise multiplication is performed between $\mathbf{E}_{inc}(i,j,k)$ and each index matrix, which is shown as

$$E_{mesh}(i,j,k) = \mathbf{E}_{inc}(i,j,k) \times Mesh(i,j,k) = \begin{cases} E_{incz}(i,j,k), PEC \\ 0, Non-PEC \end{cases} \quad (24)$$

where $Mesh(i,j,k)$ is the material index matrix. After the operation, 1536 E_{mesh} matrices are generated and are treated as the input of the CNN model.

To reduce the computation burden for CNN training and PCA analysis, sampling is conducted on each E_{mesh} matrix on all three axes. One element of every two are picked and retained which makes the new shape of E_{mesh} matrix (46,39,116). Also, since normal CNN model cannot deal with complex values, E_{mesh} matrices need to be converted to real-valued matrices so that they can be processed by the CNN model. Firstly, the real and imaginary part of each E_{mesh} matrix is split into two matrices E_{meshRe} and E_{meshIm}. Then, two matrices are concatenated on the first dimension, which is the x direction, which is expressed as

$$\begin{aligned} E_{mesh}(2*i,j,k) &= E_{mesh\text{Re}}(i,j,k), i = 0,1,2,...,48 \\ E_{mesh}(2*i+1,j,k) &= E_{mesh\text{Im}}(i,j,k), i = 0,1,2,...,48 \end{aligned}. \quad (25)$$

After that the new E_{mesh} matrix has a shape of (92,39,116) and it is considered as the input of the CNN model.

For RF-induced heating data, one-gram averaged peak SAR (psSAR_{1g}) is evaluated from each simulation. The psSAR_{1g} distribution of one SFS model is shown as follows in Figure 7.6. It can be shown that, the psSAR_{1g} values at the tip of the rod and the tip of the screw are high. Also, due to the symmetry of the EM field in phantom, the psSAR_{1g} values at the top and bottom of the model have small difference. Thus, it is not enough to use only one psSAR_{1g} as the worst-case RF-induced heating since there might be other hotspots that produce high heating. Based on this circumstance, two psSAR_{1g} values are evaluated for each SFS model as the RF-induced heating data. After each simulation is finished, two

Regions of Interest (ROIs) are set for rod tip and the outmost screw, which is shown in Figure 7.7. Two maximum one-gram averaged SAR values are found inside each region and are selected as the rod peak SAR($psSAR_{rod}$) and screw peak SAR($psSAR_{screw}$).

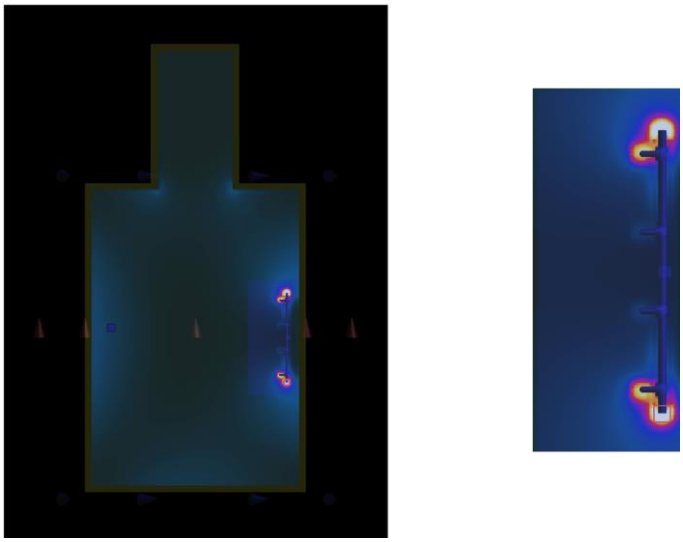


Figure 7.6 Example x-z slice of SAR_{1g} distribution

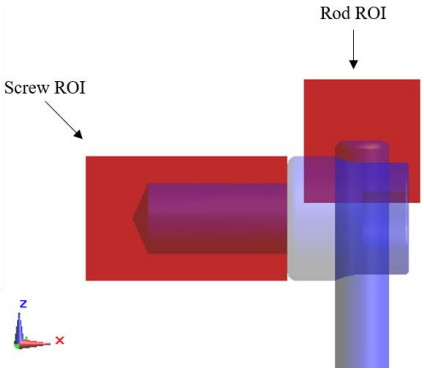


Figure 7.7 The ROI setting for rod tip and screw tip

In order to find the optimum training dataset size, PCA analysis is performed. All 1536 3-D complex E_{mesh} matrices are reshaped to 1-D vectors and are concatenated together, forming the new 2-D matrix with a shape of (1536, 208104). PCA analysis is performed on this 2-D matrix. After the analysis, the Cumulated Explained Variance (CEV) of the first n PCs are plotted and analyzed which represents the percentage of information covered by first n PCs.

7.3. CNN Architecture

In this research, all the Neural Network training is implemented by Python. The Keras toolbox, which is the high-level API to call the Tensorflow toolbox functions, is used for building the model architecture, training and testing of the CNN model. The CNN model used in this research is derived from AlexNet, which is a classical CNN architecture. The visualization of the model is shown in Figure 7.8. The model consists of one input layer, two 2-D convolutional layers, two max-pooling layers, one flatten layer, two fully connected layers and one output layer.

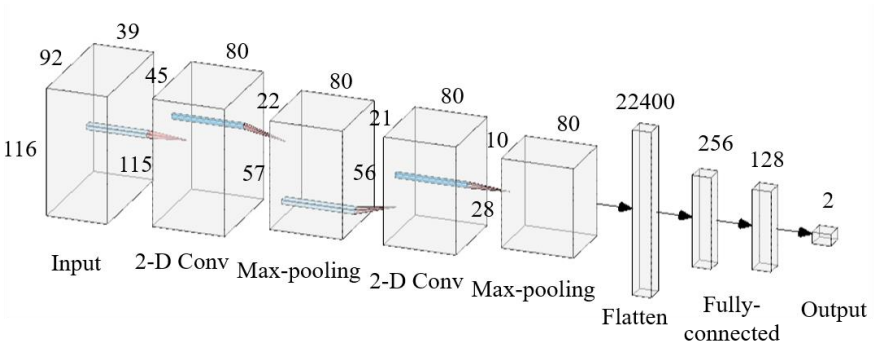


Figure 7.8 CNN model architecture

For input layer, the 1536 E_{mesh} matrices are concatenated to a 4-D matrix with a shape of (1536, 92, 39, 116). Then, the matrix is transposed to adjust the order of the dimensions. After transposing, the shape is changed to (1536, 116, 92, 39). Since the 2-D convolution is applied in this model, every 3-D E_{mesh} matrix can be considered as multi-channel 2-D images. Based on this perspective, 1536 is the number of input samples, 116 and 92 are the height and width of 2-D image and 39 are the number of channels.

For two convolutional layers, the number of filters is set to 80. Each filter has a height of 2 and width of 4. Also, the strides of the convolution are set to 1 on height and 2 on width. Since on the width direction, every incident E field information appears in a pair

of E_{meshRe} and E_{meshIm} values, the filter need to move every 2 steps on the width direction so that all real and imaginary values inside the filter are paired. For maxpooling layers, the filter shape is set to (2,2) which will extract and retain the local maximum value among 4 values inside the filter. After convolutional layers and maxpooling layers, the feature map shape is changed to (28,10,80). The flatten layer reshapes the 3-D feature maps to 1-D vector with a length of 22400. Then, two fully-connected layers are applied with 256 neurons and 128 neurons respectively. The output layer has two neurons, which corresponds to two peak SAR values extracted from simulations.

After the architecture of the CNN model is set, the model is compiled and ready for training. For loss function, mean absolute percentage error (MAPE) is used for network refinement and backpropagating process. MAPE is defined as

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{26}$$

where A_t is the true value and F_t is the predicted value. Comparing to other absolute error metrics, MAPE can better express the error with respect to the true values level. Apart from MAPE, mean absolute error (MAE) is selected as the error metric functions which is evaluated and monitored during every training epoch in order to check the convergence of the model. Adam optimization is chosen as the optimization of the model which combines the advantages of traditional algorithms to implement accelerated gradient descent procedure.

For CNN training settings, training epochs are set to 100, which means all training data are repeated 100 times for the backpropagation and network parameter update. The batch size that represents the number of data used in each gradient update is set to 32. Moreover, 20% of the training data is selected as the validation data. Validation data do

not participate the network update and the error metrics of validation data are calculated after each epoch in order to monitor the performance of network model on new data.

8. In-vivo RF-induced heating prediction of Spinal Fixation System

8.1. Model Development

In this research, a human computational model is used to replace the ASTM phantom to hold the SFS models and perform the EM simulation. The human model used is the Duke model which is a highly-detailed CAD model developed from a 34-year old adult male from Virtual Family project. The front, side and top view of the model is shown in Figure 8.1. The body tissue models in the Duke model are assigned with normal human EM parameters and thermal parameters. Thus, the in-vivo simulation with inhomogeneous Duke model can better mimic the real RF-induced heating scenario inside the human body.

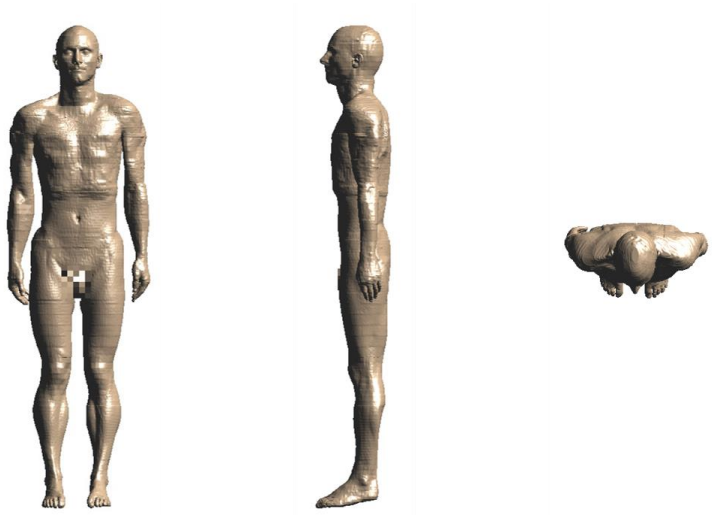
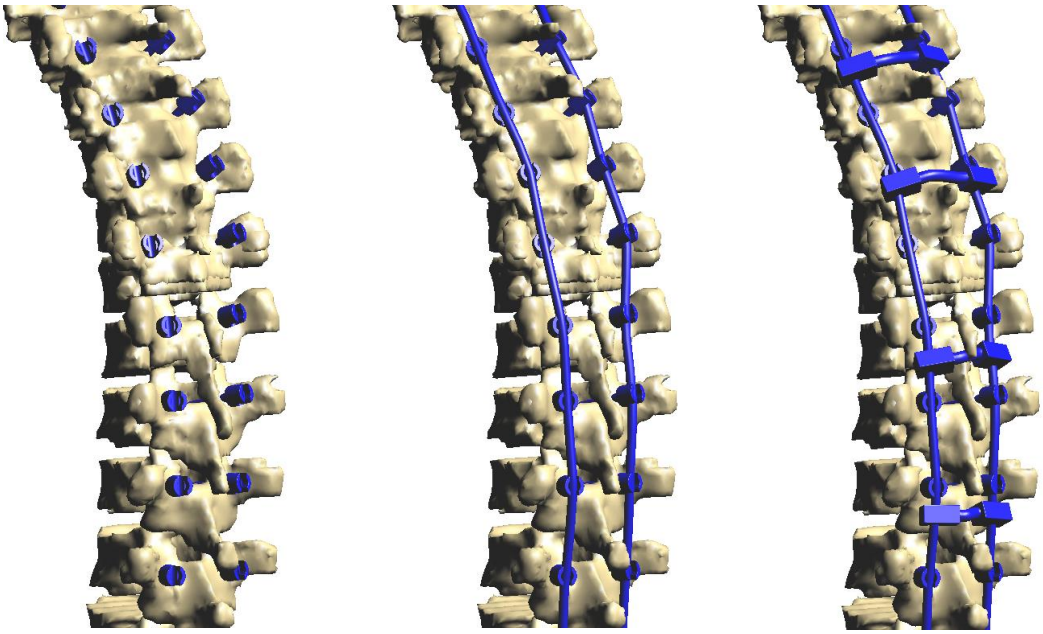


Figure 8.1 Front, side the top view of Duke model

Normally, DRS is used for the most clinical treatment cases. Thus, no SRS models are developed for this research. The development of the SFS sample models in Duke model follows several steps. Firstly, the uniform-sized screws are placed at the pedicle area of the

spine levels which ranges from C1 of Cervical level to L4 of Lumbar level. Then, two rod models are created to run through all the screws at left side and right side. Finally, connector models are applied between two rods with a distance of approximately 20mm on z-direction. The procedure of sample model development is shown in Figure. The overview of the sample model is shown in Figure 8.2.



(a) (b) (c)

Figure 8.2 The procedure of developing SFS model in Duke model

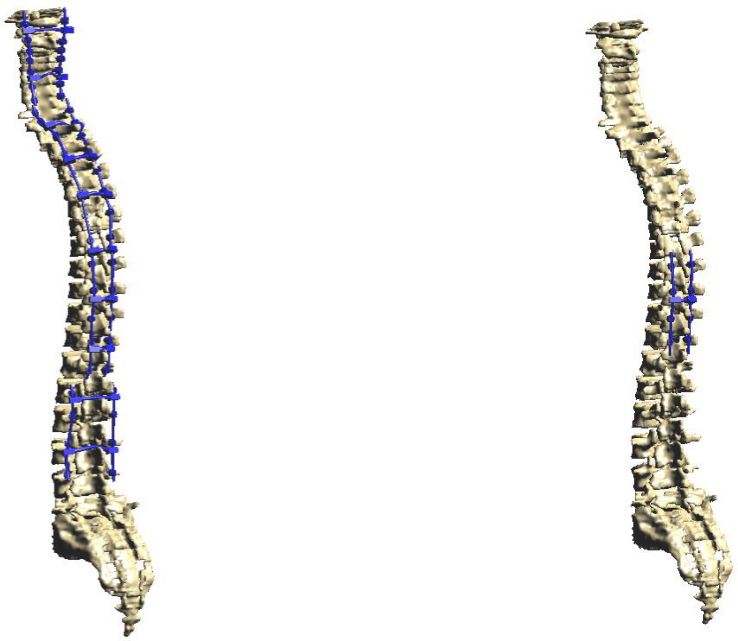
Since the SFS component used only have one geometric size, the SFS models with various rod length and position on the Vertebrae are generated in order to include as many as clinical variations. Firstly, two random numbers are created as the device center z coordinate z_0 and the rod z coordinate length l, which is

$$\begin{aligned} z_0 &\in \left[Z_{\min}, Z_{\max} \right] \\ l &\in \left[80mm, 220mm \right] \end{aligned} \tag{27}$$

where Z_{min} is the minimum z coordinate of the rod sample and Z_{max} is the maximum z coordinate of the rod sample. Then, the minimum z coordinate z_{min} and the maximum z coordinate z_{max} of the SFS model can be calculated as

$$\begin{aligned} z_{\min} &= z_0 - \frac{l}{2} \\ z_{\max} &= z_0 + \frac{l}{2} \end{aligned}. \tag{28}$$

Once the z coordinate range of SFS model is determined, the rest part of the sample rods will be cut and removed. All the screws and connectors whose center's z coordinate is out of the z coordinate range are also deleted. The rest components form the generated SFS model. All SFS components and the SFS model generated randomly are shown in Figure 8.3. Based on this method, 896 SFS models with different midpoint location and rod length are generated.



(a) (b)

Figure 8.3 The overview of sample components (a) and the randomly-generated SFS model (b)

Since the size of Duke model is too large, it is unfeasible to include mesh information of the whole Duke model. Thus, a fixed-size box is defined to sample the device and the nearby body tissue. The box has a size of 140mm*143.9mm*220mm. For different SFS models, the box is moving along z axis so that the center of the box has the same z coordinate with the center of SFS model. The box is shown in Figure 8.4.

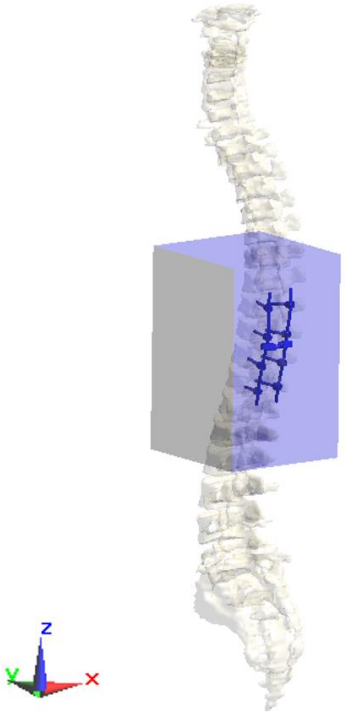


Figure 8.4 The sample box used to extract the device information

8.2. Simulation Settings

All in-vivo simulations are performed in SEMCAD X. For excitation, a generic birdcage coil is used as RF excitation. The diameter of the coil is 750mm and the length of the coil is 450mm. The coil is composed of 32 current sources which form two end rings and 16 lumped elements which form the rungs. The coil is working on 128MHz which is the Lamour Frequency of 3T MRI system. The z direction distance between the coil isocenter and the human is defined as landmark position. Zero landmark is defined by when

the z coordinate of the center of thalamus is zero. In this research, the landmark position is set to 350mm so that the middle part of Vertebrae is inside the coil with the maximum RF exposure, which is shown in Figure 8.5.

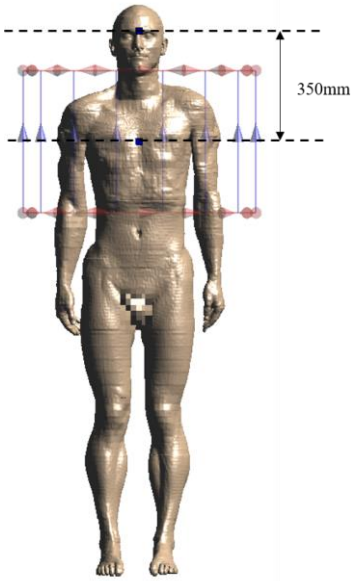


Figure 8.5 The landmark position of Duke model used in this study

The FDTD simulation period is set to 20 to ensure the convergence of the simulation. For material setting, all the human tissue models are defined as dielectric material and are assigned with the EM parameters of human tissue average values. The SFS model is defined as PEC. Absorbing Boundary Condition (ABC) is used as the boundary of the simulation. For the mesh setting, every tissue has a resolution setting of 2mm. No resolution is assigned for SFS model, so the grid mesh of the model will follow the surrounding tissue's grid. With this setting, The original model and the discretized model are shown in Figure 8.6. The total number of cells is 298*178*923 ≈ 49.96M.

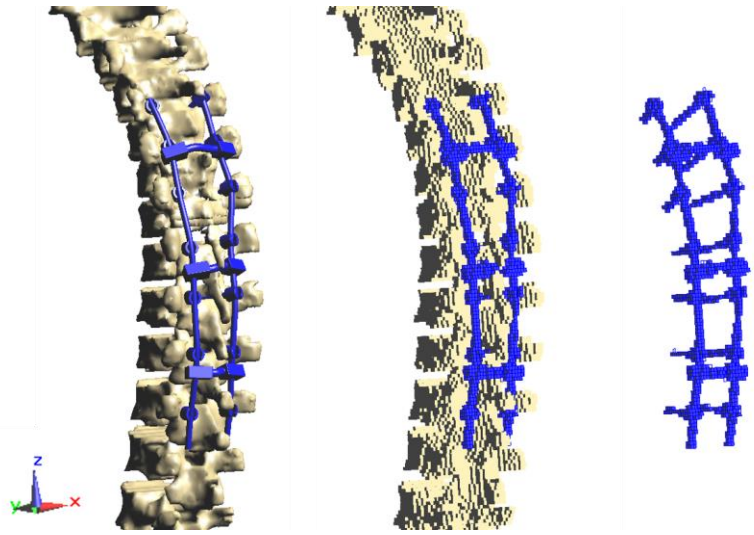


Figure 8.6 The original Vertebrae and SFS device models and discretized models

With the grid setting, the shape of the sample box is (71,73,110) and the number of cells are 71*73*110 ≈ 0.57M. In order to obtain the incident E field information inside the sample box, an extra simulation is conducted with the same Duke model and RF coil without the presence of the SFS model. After that. the incident E field inside every sample box is exported. For simplicity, only z-direction component of the E field is retained. The modified incident E field is

$$\mathbf{E}_{inc}(i,j,k) \approx \vec{z} \cdot E_{incz}(i,j,k)\,. \tag{29}$$

Thus, 896 complex-valued E_{inc} matrices are exported for all SFS models. Meanwhile, the electric conductivity (σ) of body tissue and device model inside the sample boxes are extracted. For human tissue, σ ranges from 0 S/m to 2.14 S/m. Since the assumption of PEC is that it has infinite large σ, which cannot be used for CNN model training, the σ of PEC is assign as 5 S/m. Moreover, the material index matrix inside the sample box is extracted as well. The indexes of human tissue ranges from 1 to 78 and the index of PEC is 79. In order to highlight the feature of the PEC in input information, all

the indexes are redefined: PEC's index is set to one and all the other tissues' index is set to zero. After that, 896 3-D σ distribution and redefined index matrices are obtained.

Postprocessing operations are conducted for the incident E field and sigma matrices to form the input format of CNN. Since normal CNN model is not able to process the complex-valued input, the real part and imaginary part of E_{inc} are split to form two matrices with the same shape of E_{inc}, which is denoted as E_{incRe} and E_{incIm}. Then, E_{incRe} and E_{incIm} and σ matrices are concatenated together on the first dimension, which is x direction. The new E_{mesh} matrix will be expressed as

$$\begin{aligned} E_{mesh}(2*i, j, k) &= E_{inc\mathrm{Re}}(i, j, k), i = 0,1,2,...,70 \\ E_{mesh}(2*i+1, j, k) &= E_{inc\mathrm{Im}}(i, j, k), i = 0,1,2,...,70 \end{aligned} . \tag{30}$$

The shape of three original matrices is (71,73,110), thus the shape of E_{mesh} matrix is (142,73,110). The new formed matrix is considered as the input of the CNN model.

In terms of RF-induced heating data, instead of using two peak SAR values which is used in in-vitro case, only one peak SAR value is extracted and used as the heating data. Because unlike in-vitro cases in which the high SAR values tend to appear at the tip of rod and outmost screw, the SAR distribution will be more complicated for in-vivo cases due to the inhomogeneity of body tissue surrounding the SFS model. Thus, the number and location of the hotspots may be uncertain. While the number of output is fixed for CNN model, only one worst case $psSAR_{1g}$ value is considered in simplicity.

8.3. CNN architecture

The CNN model used in this research is derived from AlexNet. For in-vivo cases, the electrical conductivity distribution is inhomogeneous in Duke model. Also, the distribution of incident E field is complex compared to in-vitro cases. Thus, the mesh information, incident E field information and electrical conductivity information are not

merged as one input matrix since the filter may not be able to capture all features from the input. Based on this situation, a new CNN architecture is introduced. The visualization of the CNN model used in this research is shown in Figure 8.7.

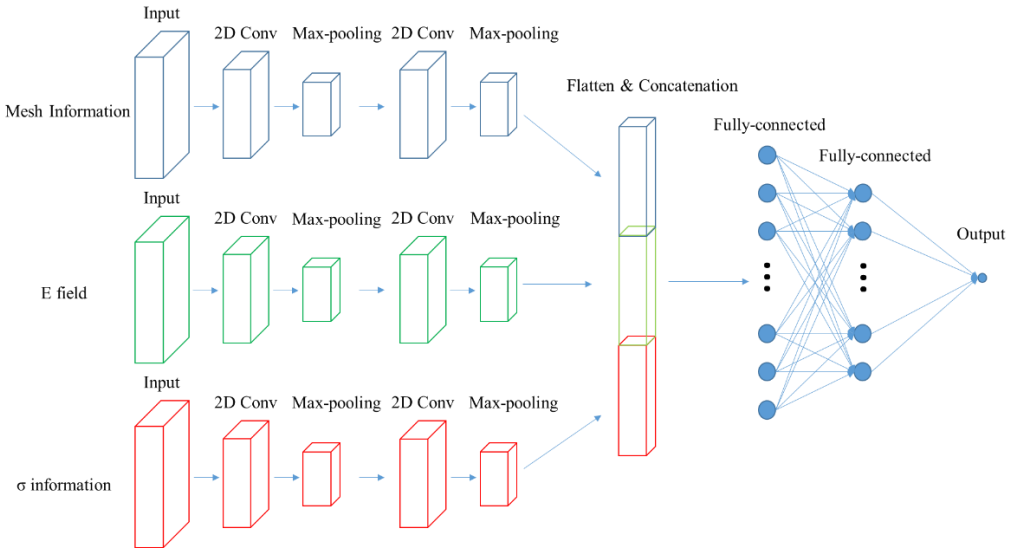


Figure 8.7 CNN architecture

The model consists of three parallel sub-CNN networks. Each network is derived from AlexNet and includes one input layer, two 2-D convolutional layers and two max-pooling layers. After that, one flatten layer, two fully connected layers and one output layer are included.

896 mesh, incident E field and electric conductivity matrices are concatenated separately to 4-D matrices. The mesh matrix and σ matrix have a shape of (896, 71, 73, 110) and the incident E field matrix has a shape of (896, 142, 73, 110). Then, three matrices are transposed to adjust the order of the dimensions so that the second dimension and the fourth dimension are swapped. Three 4-D matrices are treated as input layers of three sub-CNN networks.

For every convolutional layer, the number of filters is set to 40. Each filter has a height and a width of 10. For max-pooling layers, the filter shape is set to (2,2) which will

extract and retain the local maximum value among 4 values inside the filter. After all convolutional layers and max-pooling layers, the feature map shape is changed to (20,11,40). The flatten layer reshapes three 3-D feature maps to three 1-D vectors with a length of 8800. Then, three 1-D vectors are concatenated to one 1-D vector with a length of 26400. By the concatenation, the features extracted from three different inputs are merged. Then, two fully connected layers are applied with 256 neurons and 128 neurons respectively. The output layer has one neuron, which corresponds to the peak SAR values extracted from simulations.

After the architecture of the CNN model is set, the model is compiled and ready for training. mean absolute percentage error (MAPE) is used as loss function. Apart from MAPE, mean absolute error (MAE) is selected as the error metric function. Adam optimization is chosen as the optimization of the model.

For CNN training settings, training epochs are set to 100. The batch size that represents the number of data used in each gradient update is set to 32. Moreover, 20% of the training data is selected as the validation data.

9. Results

9.1. Step size for selecting training set

When selecting the training dataset for the network, the physical knowledge cannot be ignored. Empirically, the device length is one of the major factors to affect the peak spatial SAR value. If training dataset is formed by plate device data that have large step size on the device length, the prediction behavior will probably be significantly affected. Thus, the appropriate device length step size is investigated under 1.5T and 3T environment.

A metal cylinder rod model is used as the device model which is regulated in ASTM 2182 standard [1] with the same birdcage coil and ASTM phantom. The rod is parallel to the z axis and has a distance of 20mm to the right wall of phantom. The length of the rod ranges between 50mm to 370mm with a step size of 10mm. Overview of all the models is shown in Figure 9.1. A total of thirty-three in-vitro simulations are performed, mesh of the rods and $psSAR_{1g}$ values are extracted afterwards. For training dataset, rod data with different step size on rod length are chosen to form different datasets. Step size of 10mm, 20mm, 40mm and 80mm are chosen which makes training dataset consist of 33, 17, 9 and 5 data. The testing data will include all 33 data to evaluate the performance on predicting correct trend of RF-induced heating changing.

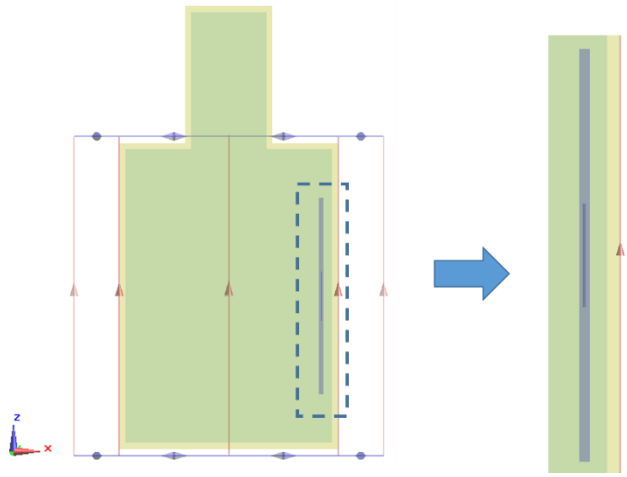


Figure 9.1 Overview of models with ASTM rod

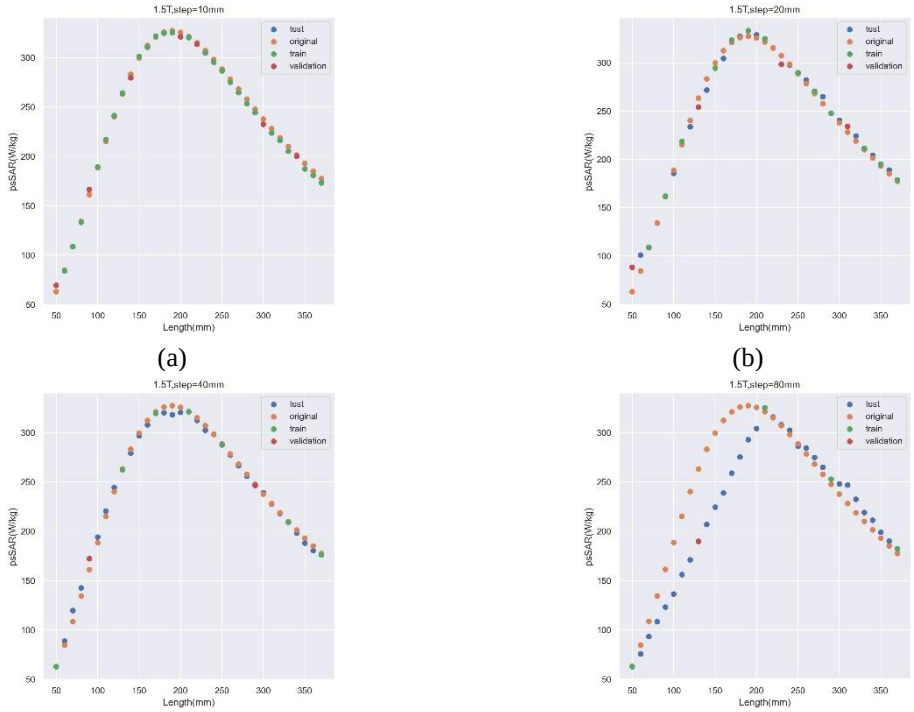
The scatter plot of predicted training data, predicted validation data, predicted testing data and all the original data under 1.5T and 3T environment are shown in Figure 9.2 and Figure 9.3. It can be shown that, for 1.5T environment, the network receives good prediction performance for step size of 10mm, 20mm and 40mm. When step size is set to 80mm, the prediction has significant error as the training data size is too small. While for 3T environment, the prediction already gets worse when step size is increased to 40mm. These results correlate with the RF field distribution features that the step size of the rod length for training data should be less than 1/10 of the wavelength. The wavelength formula inside certain medium which is expressed as

$$\lambda = \frac{2\pi}{k'}, \tag{31}$$

$$k' = \mathrm{Re}(k) = \mathrm{Re}(\omega\sqrt{\mu\varepsilon_c}), \tag{32}$$

and

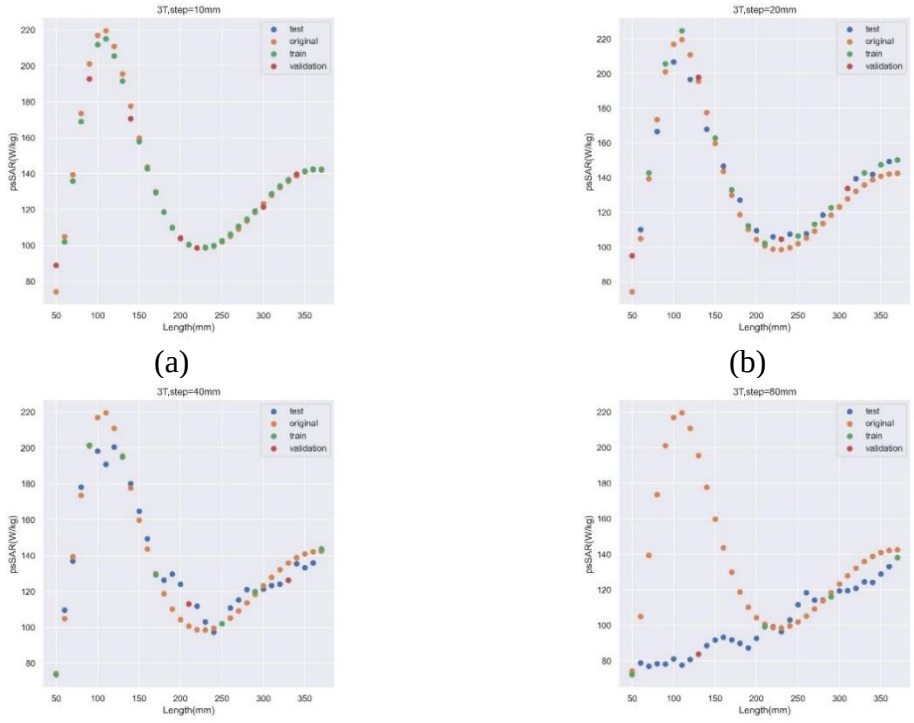
$$\varepsilon_c = \varepsilon_0\varepsilon_r - j\frac{\sigma}{\omega}. \tag{33}$$



(a) (b)

(c) (d)

Figure 9.2 Scatter plot of data under 1.5T. Step size =10mm(a), 20mm(b), 40mm(c) and 80mm(d)



(a) (b)

(c) (d)

Figure 9.3 Scatter plot of data under 3T. Step size =10mm(a), 20mm(b), 40mm(c) and 80mm(d)

The wavelength inside the conductive gel is 432.2mm for 1.5T and 243.93mm for 3T. Thus, for 1.5T system, the step size should be less than 43.22mm and for 3T system, the step size should be less than 24.39mm.

9.2. Results of RF-induced heating prediction of Tibia Plating System

9.2.1. Convergence and Correlation

The whole training duration is around 15 minutes to finish. During the training, the MAP and MAPE of training dataset and validation dataset is evaluated at the end of every epoch. The training and validation MAPE are shown in Figure 9.4. It can be seen that the error level decreases significantly in the first ten epochs. The error level has decreased to 5% at epoch 10. After that, the error level fluctuates below the 5% mark except the validation MAPE where the over 5% error appear in 1 epoch. Also, the validation MAPE are close to the training MAPE, which means that the network is robust to the new data and there is no overfit for the CNN model which might cause a significantly higher error level for validation dataset than training dataset.

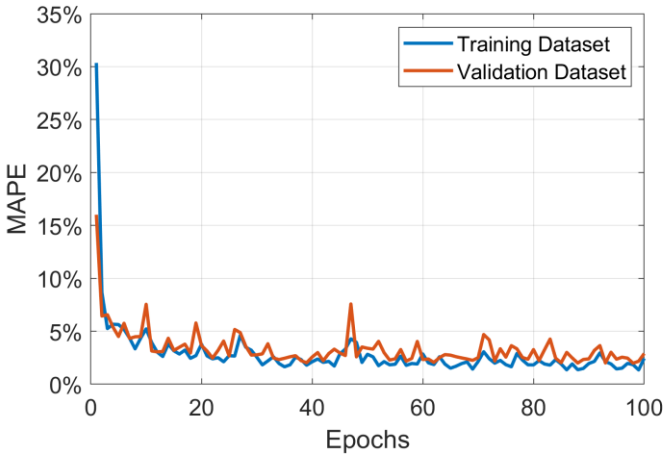


Figure 9.4 The MAPE level of training dataset and validation dataset during training

Once the training is finished, the testing dataset is used for evaluating the performance of predicting new data. The average peak SAR values, MAE and MAPE

values of training dataset (including validation data) and testing dataset are tabulated in Figure 9.1. It shows that, the averaged SAR value for both training and testing dataset is around 200W/kg. Testing dataset MAE is slightly higher than training dataset, which is reasonable as testing dataset consists of new data that did not participate the network training. Both MAPE are below 2% and testing dataset has a higher error level than training dataset. Overall, the absolute error and percentage error level is very low.

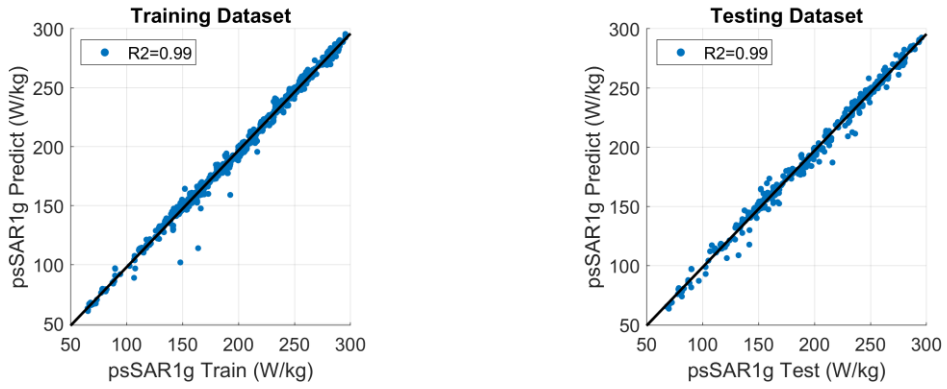
Table 9-1 Averaged SAR level and error level of training and testing dataset

	Training	Testing
Averaged SAR value (W/kg)	203.90	196.82
MAE	2.52	2.93
MAPE	1.46%	1.80%

The correlation between predicted peak SAR value and true peak SAR value of training dataset and testing dataset is also investigated. R^2 score is used as the measure of the correlation, which is defined by

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{34}$$

where y_i is the true value, f_i is the predicted value and $\bar{y}$ is the average value of true values. The scatter plot of training dataset and testing dataset correlation is shown in Figure 9.5. Both training dataset and testing dataset have a R^2 score of 0.99. It can be shown from the scatter plot that, majority of the data points have a very limited distance to the line x=y, which means that the predicted value and true value are highly correlated. Only a few of data points are observed in both datasets to deviate the line. Thus, the CNN model achieves high correlation between predicted value and true value.

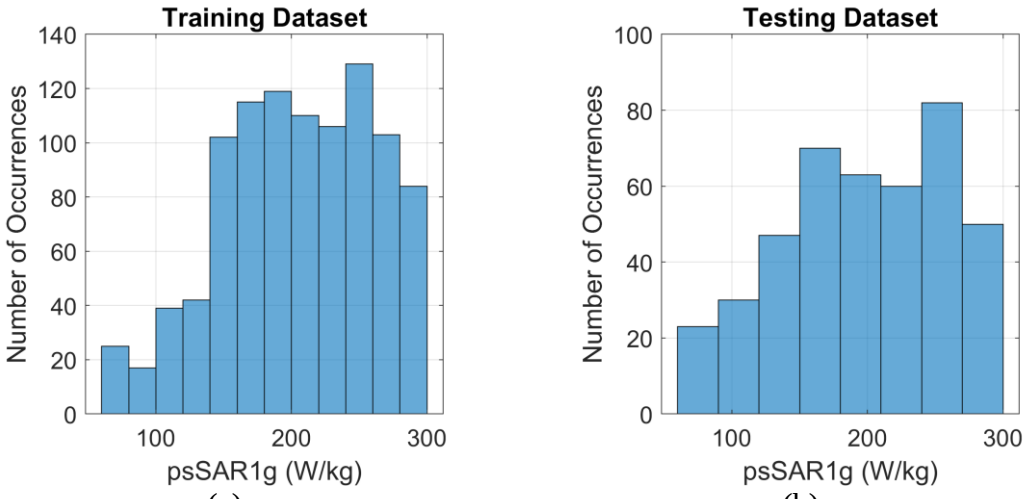


(a) (b)

Figure 9.5 Correlation between predicted and true value for both datasets

9.2.2. Error Level Distribution

After the network convergence and the total error level are investigated, the distribution of the error is also observed as the averaged value cannot reveal all the information sometimes. Firstly, the distribution of the peak SAR values for both datasets are investigated and the histograms of distribution are shown in Figure 9.6. It can be obtained from the figure that, the peak SAR values of the plate devices are concentrated with the range of [90W/kg, 300W/kg] while the range of [150W/kg, 200 W/kg] has the largest number of occurrences. There are also no outliers for peak SAR values.



(a) (b)

Figure 9.6 Distribution of peak SAR values for training dataset (a) and testing dataset (b)

The distribution histograms of MAE for training dataset and testing dataset are shown in Figure 9.7. From the figure, majority of the training and testing data has an MAE value of less than 10W/kg with only a limited number of large error level are observed. The maximum MAE value for training dataset is 50.04W/kg and the maximum MAE value for testing dataset is 29.03W/kg.

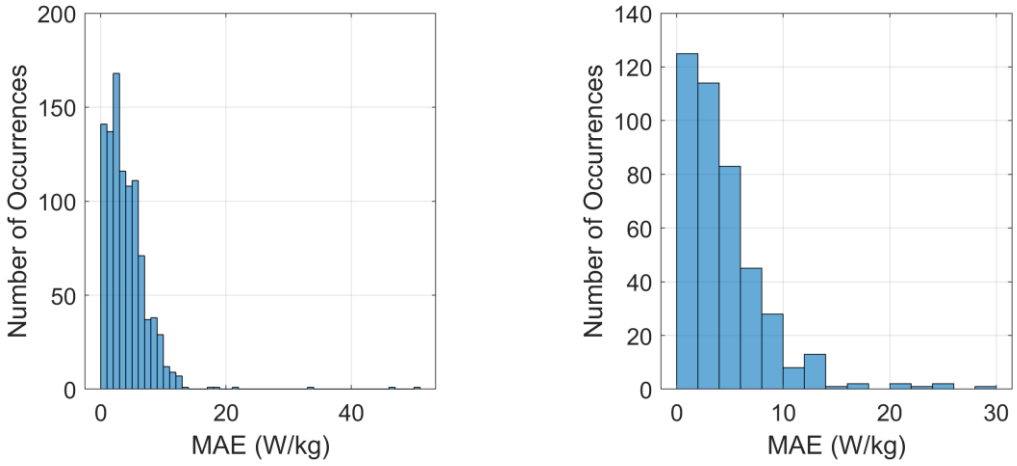
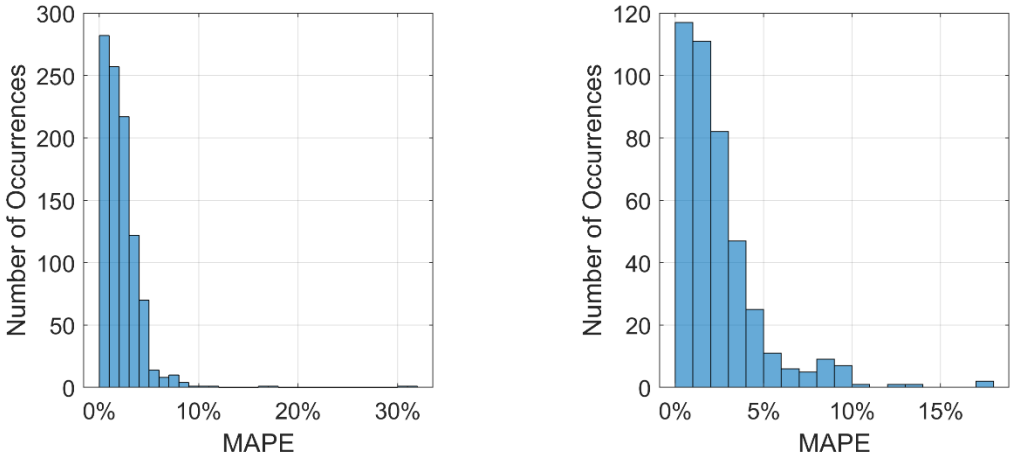


Figure 9.7 Distribution of absolute error for training dataset (a) and testing dataset (b)

Distribution histograms of MAPE for training dataset and testing dataset are also investigated and the shown in Figure 9.8. It can be observed that, majority of data has a MAPE level of less than 5%. For testing dataset, the portion of data that have MAPE values larger than 5% is more than training data. The maximum MAPE of training dataset is 31% and the maximum MAPE of testing data is 17.77%.



(a) (b)

Figure 9.8 Distribution of absolute percentage error for training dataset (a) and testing dataset (b)

Breaking down the error level distribution, the following information can be obtained: over 90% of the training data and testing data has small MAE and MAPE level. Outliers are observed as the maximum MAE is 50.04W/kg, while it is still acceptable as the SAR value is larger and the MAPE is 31% which is the more applicable metric to describe the error level, which means that the CNN model in this research can predict the peak SAR value for tibia plate devices with low error level in most of the circumstances.

9.2.3. Selection of Training Dataset

Initially in this research, 70% of the total data is used to form the training dataset, which is 991 data. This is an empirical choice from the general CNN applications where all the image data are easily acquired with little difficulty and large amount of data are picked to form the training dataset in order to ensure the thorough training of the CNN model. While in EM related problems, the philosophy is different. Since every peak SAR value data is obtained from a finished full-wave simulation, it means that the same amount of the simulations with training dataset size needs to be performed to obtain the dataset, which will still be time-consuming. In this case, if less training data are used for network

training while the network can achieve the similar acceptable performance, time and computational resources can be saved. Thus, it is necessary to investigate the least possible amount of training data that can have similar prediction performance with more training data.

Based on the new philosophy, the naïve strategy for finding optimal training data amount is to sweep the training data size and find the smallest possible number of training data that has the significant prediction performance. The strategy is as follows: 20% of the whole data is randomly selected as testing data which will not change later. For the rest of the data, different portion of data are selected as training dataset, starting from 10% to 80% of the whole data with a step of 10%. The validation data amount is always the 20% of the training data. Totally, there will be eight different training data amounts. Since every time the network is under training, the initial weights are different, different correlation and error levels are expected. Training data will consequentially be randomly selected 10 times for each dataset amount. The network will then be trained for 10 times as well for each data amount to have the most comprehensive description of the network performance with certain training data amount.

After training with all the datasets and evaluating all the error metrics and correlations, Table 9-2 shows the correlation, MAE and MAPE level with different training data amounts. It can be clearly shown that, for training data less than 30% of the whole data, although the averaged error level stays low with a maximum less than 8.2% of MAPE, both error levels are nearly two-fold of the rest error levels. Also, the correlation for 10% and 20% of training data reach less than 0.94 for both training and testing. In contrast, the error levels start to improve from 30% training data with training MAPE of 2.31% and

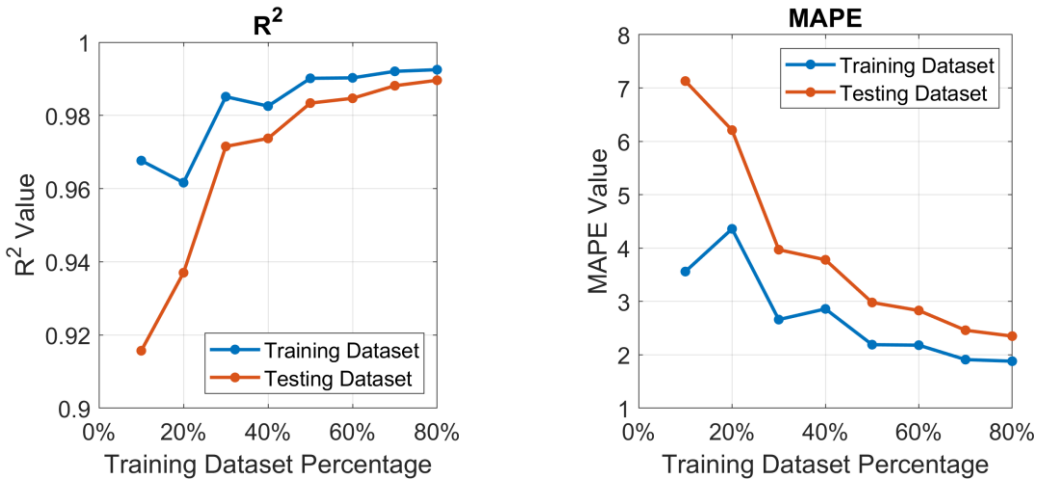
testing MAPE of 3.3%. The correlation also improves as the R^2 score of testing dataset for 30% training data nearly reaches 0.98.

Table 9-2 Correlation and error levels on different training data amounts

Training data percentage	Training data amount	Training Dataset			Testing Dataset		
		MAE	MAPE (%)	R^2	MAE	MAPE (%)	R^2
10%	142	11.52	5.94	0.9174	14.64	8.19	0.8811
20%	283	11.332	5.663	0.9356	12.437	6.796	0.9252
30%	425	4.43	2.31	0.9885	5.78	3.30	0.9791
40%	566	5.61	2.88	0.9841	6.26	3.52	0.9800
50%	708	4.98	2.57	0.9874	5.70	3.19	0.9831
60%	805	4.35	2.23	0.9894	4.89	2.70	0.9873
70%	991	4.94	2.51	0.9875	5.18	2.84	0.9859
80%	1133	4.87	2.42	0.9883	5.10	2.73	0.9872

In order to further investigate the trend of error levels and correlations with training data amount increasing, Figure 9.9 shows the change of MAPE and R^2 score with the increase of training data. MAE curve is not shown as it has the same trend with MAPE. It can be shown that, when the data amount is less than 425, the MAPE level increase and the R^2 score decreases significantly. When the data amount surpasses 425, which is 30% of the total data, the error level decreases sharply and R^2 score reaches over 0.98. For training data, the prediction performance sees a setback when training data increase to 20% from 10% of the total data. One probable reason is that the network is not converged yet with few training data. Thus, the prediction error level may be unpredictable. Also, including more training repetitions may be able to have more stable error levels and the setback may disappear. One needs to be noticed that, when the training data keep increasing from 425, the prediction performance does not monotonically improve. The averaged performance even deteriorates when data amount increases from 800 to 1000 where the network overfit may appear. Besides, the low threshold of training data amount that affect

the convergence of the network training is the data amount, rather than percentage portion of whole data, which is reasonable as the network will still converge with 400 data even the whole dataset size increases.



(a) (b)

Figure 9.9 Change of error level and correlation when training data amount increases

The naïve strategy provides a comprehensive analysis on the CNN model prediction performance with different sizes of training datasets. While this strategy still needs large amount of simulations to be performed to obtain the training data. Also, significant CNN model training repetitions takes another large amount of time. Thus, PCA is brought in which conducts analysis on the incident E field and mesh information. Multiple PCs are obtained and ranked by the information each PC covers. The number of highest-valued PCs can be considered as the training dataset size with the assumption that the training dataset covers as much information as the highest-ranked PCs.

After the analysis, 1415 PCs are obtained which are 1D vector with same length with 1-D vector input. These vectors are reshaped back to 3-D matrices with shape of (182,45,33) which is the same with complex E_{mesh} matrix. Then, the magnitude of every element inside matrix is evaluated for the visualization. The center slice of the magnitude

matrix for the 1st, 11th, 21st, 31st, and 41st PC are shown in Figure 9.10. It can be clearly seen that, the plate in the device model contributes the largest variances as the vertical component has higher value in the 1st PC. Meanwhile, the screws contribute the smaller variances as the horizontal component stands out in the subsequent PCs.

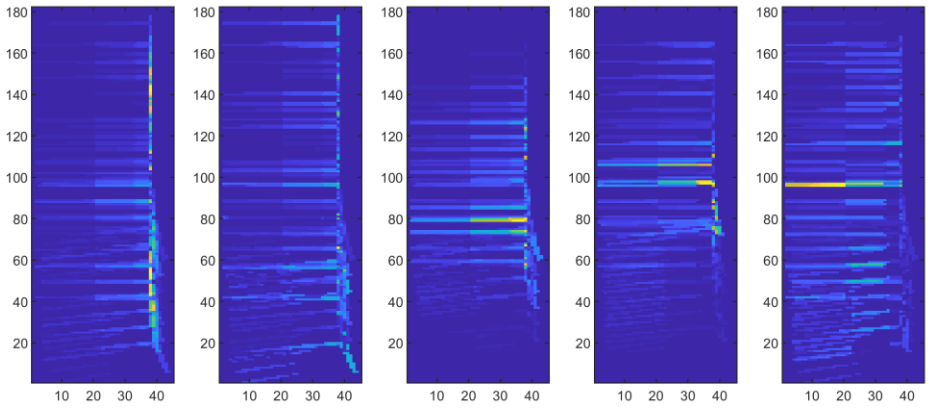


Figure 9.10 Center slice of magnitude of 1st, 11th, 21st, 31st and 41st PC

Afterwards, CEV of PCs are plotted and shown in Figure 9.11. . It can be seen that, the first 90 PCs have covered over 90% of the total variance and the first 400 PCs have covered over 99% of the total variance. If these 400 PCs form the training dataset, less than 1% of MAPE could be achieved as the input has covered the majority of the information.

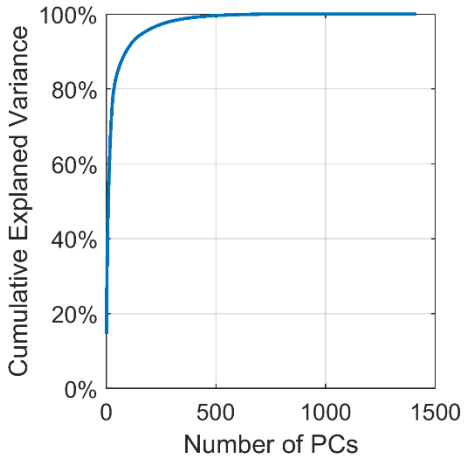


Figure 9.11 The relationship between CEV curve and number of PCs

Moreover, the 1-CEV(n) curve is plotted and overlaid with the scatter plot of

MAPE curve with different number of training data in Figure 9.12. The 1-CEV(*n*) indicates the percentage of the total energy that is not included in the datasets for the CNN model development. This would be considered as an indication of the lowest level that a CNN model can achieve. It can the implied from the figure that both the 1–CEV(*n*) and the MAPE of the CNN model decrease as the size of the training dataset increases. However, as expected, the MAPE would be still higher than the 1–CEV(*n*). For the RF-induced heating for the tibia plates used in this research, based on the 1–CEV(*n*), it appears that the size of 400 datasets is a good indication that all energy that has been included in the training sets. However, the error levels from the CNN model are still slightly higher than the 1–CEV(*n*).

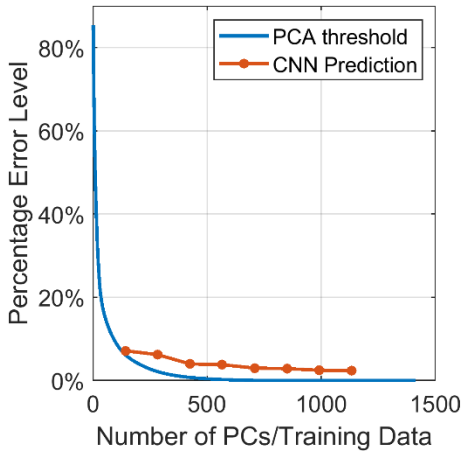


Figure 9.12 Comparison between information not covered by PCs and the percentage error level of CNN model with different training dataset size

Based on previous results, 1–CEV(*n*) can be used as a reference for choosing the appropriate training dataset size. The percentage error calculated from the PCA can be considered as the lower limit of the prediction error from the CNN model. For example, to achieve a MAPE level of 10%, the training dataset must contain at least 100 datasets, as 100 PCs will produce a 10% error level. Overall, the PCA can be used as an assistant tool

to estimate the best-fit size of the training datasets.

9.3. Results of In-vitro RF-induced heating prediction of Spinal Fixation System

9.3.1. Convergence and Correlation

During training, the error metric functions are evaluated at the end of every epoch. The MAPE level of training dataset and validation dataset are shown in Figure 9.13. It can be shown that, MAPE level decreases significantly in the first 20 epochs to around 10%. After that, both MAPE level decreases slowly and MAPE of training dataset keeps lower than validation MAPE until at the last epoch. This is reasonable as the validation data are new data for the trained CNN model. No model overfit is observed since the difference between 2 MAPE levels are small.

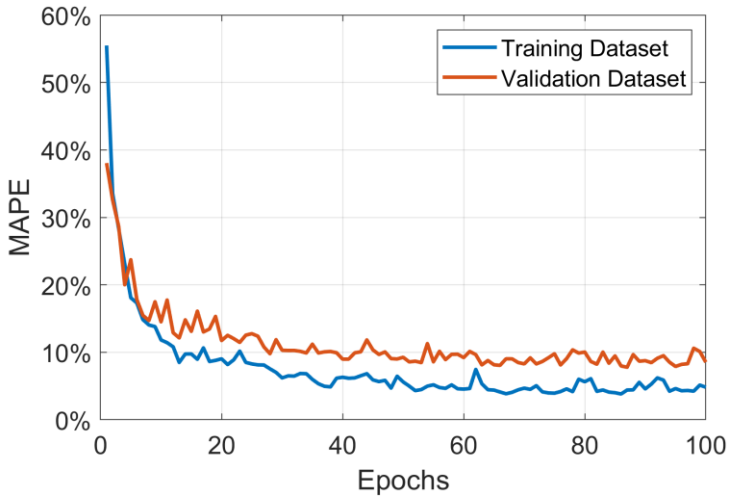


Figure 9.13 The MAPE level of training dataset and validation dataset during training

After training, the averaged value of training and testing dataset, MAE and MAPE of rod peak SAR and screw peak SAR is tabulated in Table 9-3. It can be shown that, the averaged rod peak SAR value is 25% higher than averaged screw peak SAR value, which is reasonable as the rod tip is closer to the top and bottom of the phantom where larger EM

field exposure is expected. However, the peak SAR value at the screw tip is also not negligible. Both prediction yields low MAE with the maximum value as 5.53W/kg. MAPE values are small as well with the maximum value less than 10%. Both error levels of testing dataset are higher than that of training dataset which is acceptable and expected.

Table 9-3 Averaged SAR level and error level of training and testing dataset

	Rod		Screw	
	Training	Testing	Training	Testing
Averaged SAR value (W/kg)	100.43	100.83	72.82	68.99
MAE(W/kg)	4.58	5.53	3.35	4.99
MAPE	4.99%	5.85%	5.41%	7.88%

For the correlation between true SAR data and predicted SAR data, R^2 score is used as the metrics of data correlation. The scatter plot of training and testing dataset between true value and predicted value is shown in Figure 9.14.

From the figure, both rod and screw predictions have good performance to receive R^2 scores over 0.98 for training dataset and R^2 scores over 0.95 for testing dataset. Note that, although rod and screw have similar training dataset prediction performance, rod prediction behaves better than screw prediction. It can be observed from the figure that, screw prediction has more deviated data points at the high peak SAR value. Based on this phenomenon, one probable reason is that the screw peak SAR are mainly concentrated at lower values which backpropagation are mostly based on. Thus, the CNN model will receive mediocre prediction performance when it tries to predict the high value screw peak SAR.

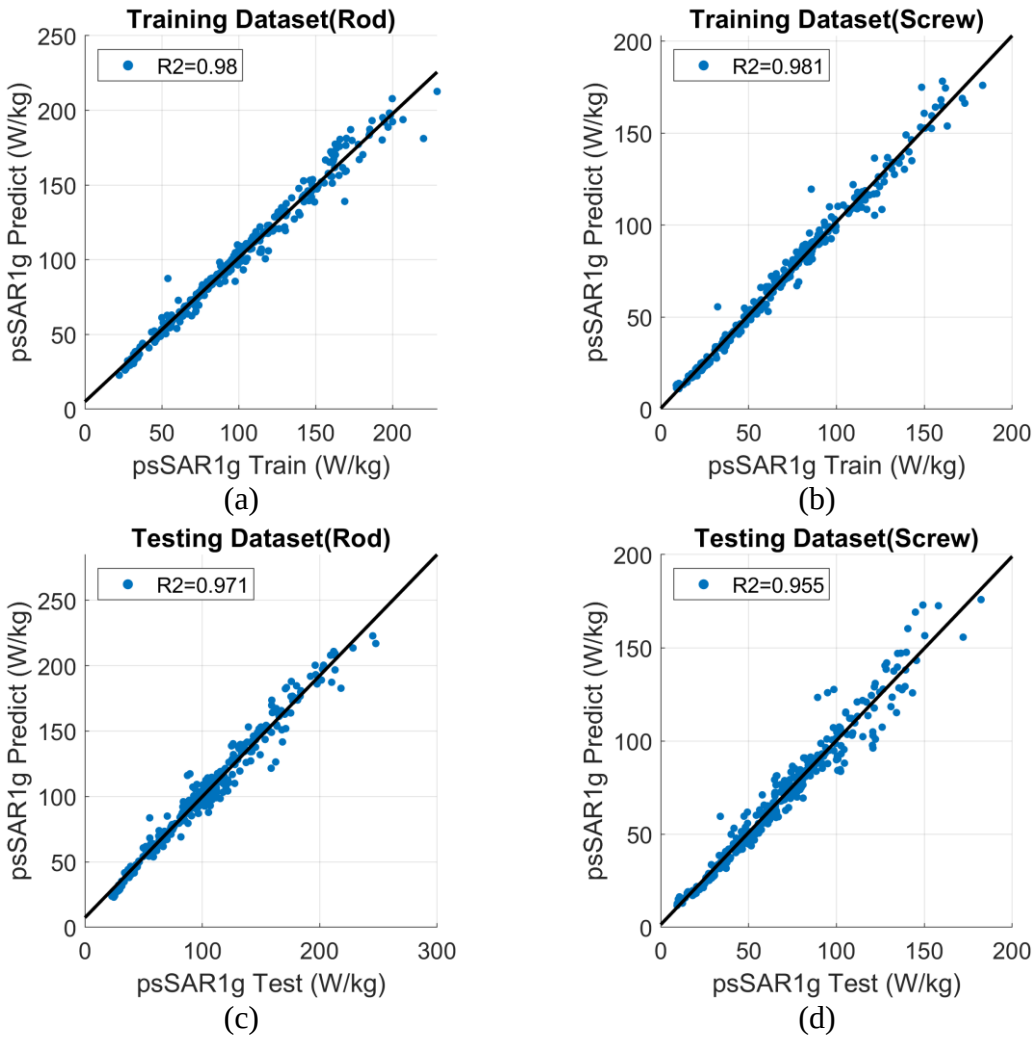


Figure 9.14 Correlation between predicted and true value for both datasets

9.3.2. Error Distribution

In order to further investigate the distribution of the prediction error, histograms of MAE and MAPE distribution are generated. Before that, it is necessary to investigate the distribution of the peak SAR values of training dataset and testing dataset which is shown in Figure 9.15. It can be shown that, all the peak SAR values are with the range of [0,250W/kg] while rod peak SAR has larger minimum and maximum values than screw peak SAR. Majority of the rod peak SAR values are near 100W/kg while majority of the screw peak SAR value locate at the 75W/kg for both training dataset and testing dataset,

which provides the evidence for the phenomenon mentioned before that the CNN model has inferior prediction performance for screw peak SAR.

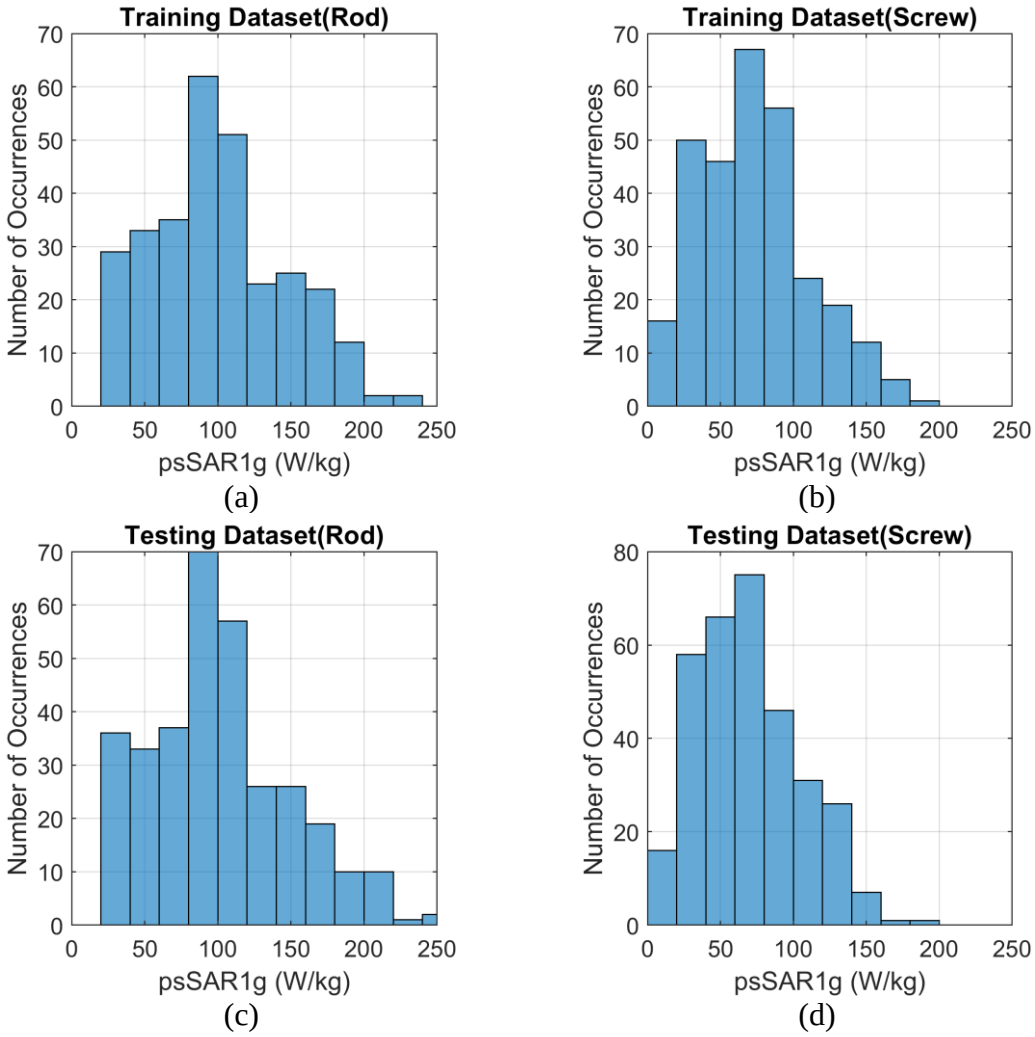


Figure 9.15 Distribution of rod peak SAR values for training dataset (a) and testing dataset (b) and screw peak SAR values for training dataset (c) and testing dataset (d)

The absolute error and absolute percentage error distribution of training dataset and testing dataset for rod and screw prediction are shown in Figure 9.16 and Figure 9.17. For absolute error, all the data are within the range of [0, 40W/kg]. While several outlier error levels up to 40W/kg appear, over 90% of the data have error levels that are less than 10W/kg.

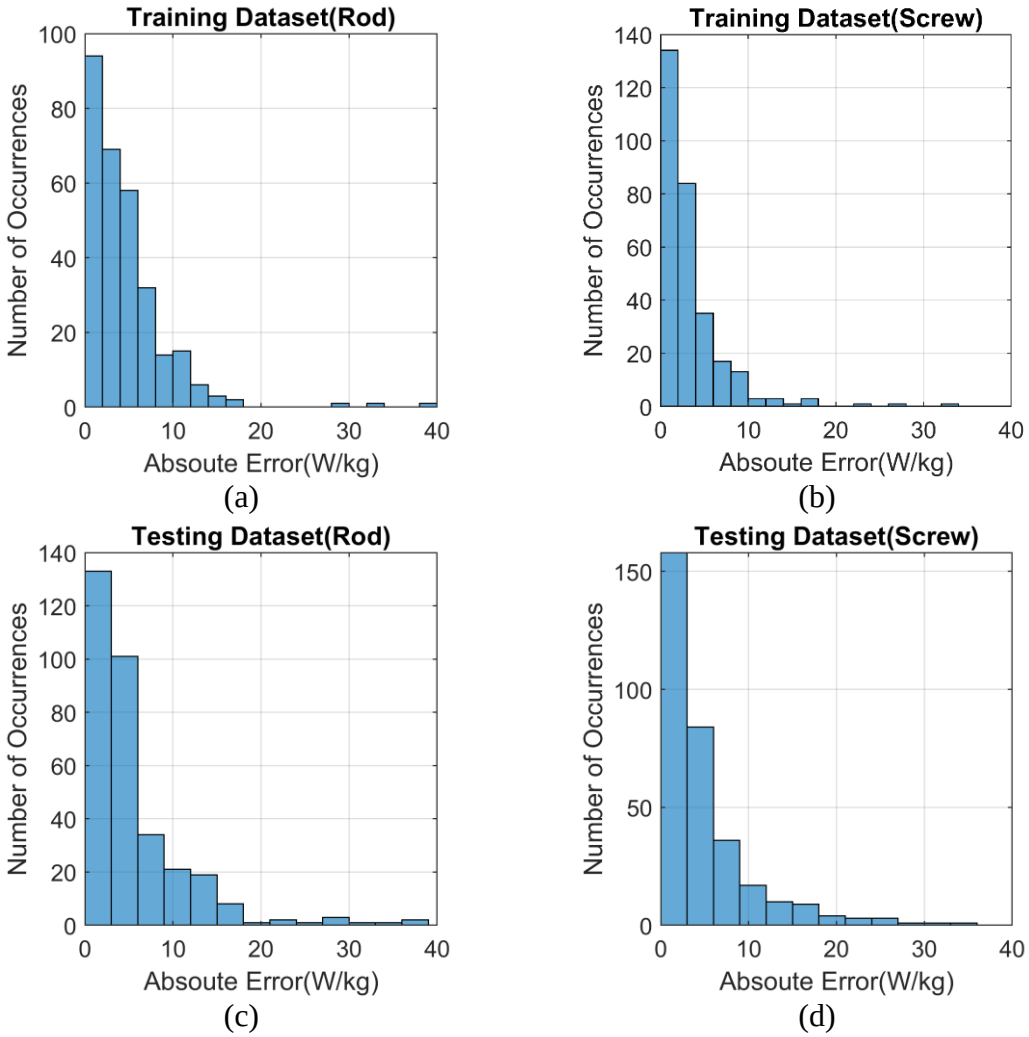


Figure 9.16 Distribution of rod absolute error for training dataset (a) and testing dataset (b) and screw absolute error for training dataset (c) and testing dataset (d)

For absolute percentage error, rod prediction and screws distribution have the similar error distribution. Over 90% of the training data have a less than 10% level and over 90% of the testing data have a less than 15% level. However, outlier values exist for both training dataset and testing dataset where maximum percentage error of 60% appears in training dataset and percentage error of 80% appears in testing dataset. Screw prediction has more outlier error values than rod prediction as multiple percentage error values are located in the range of [20%,80%].

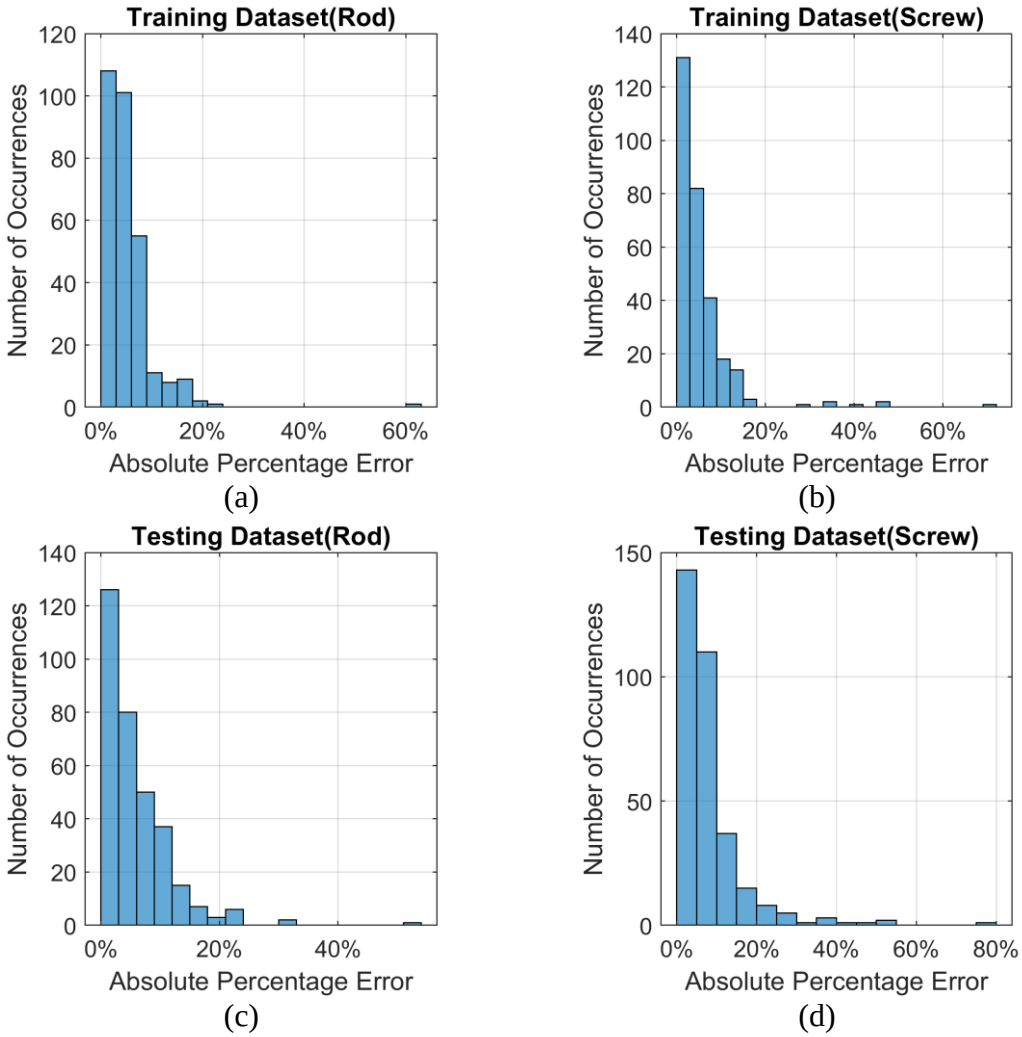


Figure 9.17 Distribution of rod peak SAR absolute percentage error for training dataset (a) and testing dataset (b) and screw peak SAR absolute percentage error for training dataset (c) and testing dataset (d)

9.3.3. Selection of Training Dataset

Since PCA analysis is performed for the Tibia Plating System RF-induced prediction and shows good indication on the appropriate size of the training dataset, naïve strategy for SFS RF-induced heating prediction is omitted in order to save time and computer resources. For PCA analysis, 1535 1-D vector PCs are obtained which have a common shape of (1,208104). These PCs are reshaped to 3-D matrix with the shape of

(116,46,39) in order to check the geometrical features of PCs. Also, the explained variance of every PC is evaluated with the rank to represent the information every PC covers.

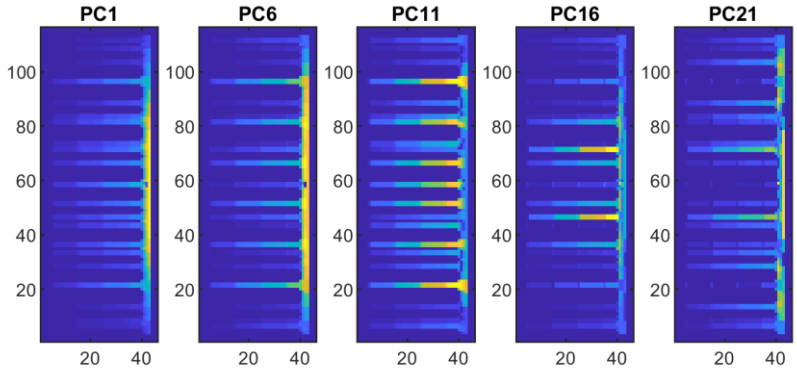
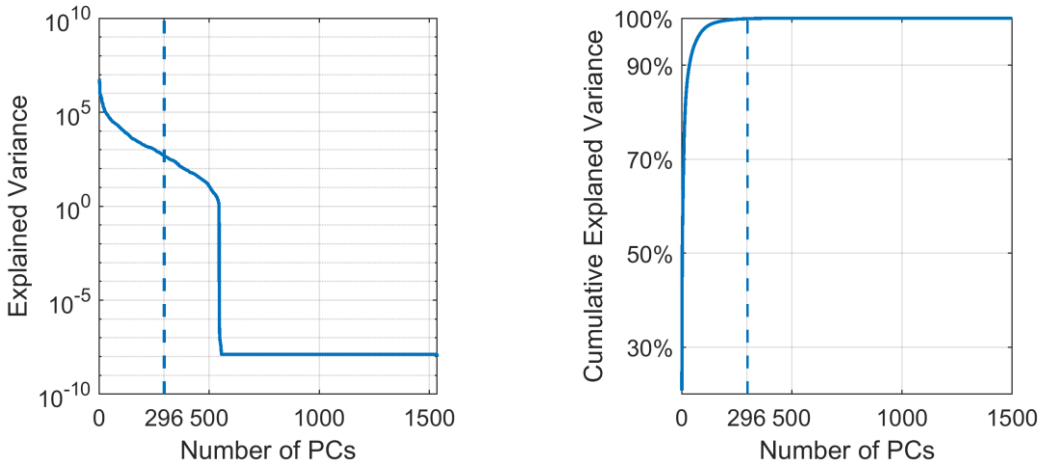


Figure 9.18 Center slice of magnitude of 1st, 6th, 11th , 16th and 21st PC

In order to visualize the variance each PC covers, the magnitude of every element is evaluated. The center slice of 1st, 6th, 11th, 16th and 21st PC magnitude is shown in Figure 9.18. It can be shown that, the rod feature is covered in the highest variance PCs and the screw features are covered in the subsequent PCs. This result correlates with the peak SAR distribution as rods are the dominant components of the peak SAR values which implies that PCA analysis can extract the geometric feature of the SFS models.

The Explained Variance value is evaluated for every PC and the CEV curve is evaluated afterwards, which is shown in Figure 9.19. From the figure, it can be observed that the variance at 296th PC has decreased by 100 times from 10^5 to 10^3. Also, from the CEV curve, the first 296 PCs accounts for 99.9% of the variance which means that only 0.1% of the information is not covered by the first 296PCs. According to the PCA analysis for RF-induced heating prediction for Tibia Plating System, if choosing 296 as the training dataset size, the best possible performance it can achieve is MAPE level of 0.1%. Thus, the training dataset size chosen in this research is 296.



(a) (b)

Figure 9.19 Explained Variance curve (a) and CEV curve (b)

9.4. Results of In-vivo RF-induced heating prediction of Spinal Fixation System

9.4.1. Convergence and Correlation

The MAPE level for training dataset and validation dataset after each epoch is evaluated. The trend of the MAPE along epochs are shown in Figure 9.20. It can be seen that, the error level significantly decreases to roughly the level of 20% in the first 20 epochs. From the 20th epoch to the 40th epoch, the error level keeps decreasing to 10%. After the 40th epoch, the error levels decrease slowly and reaches the level of below 10% at the end of training. The validation MAPE is lower than training MAPE in the first 25 epochs and surpasses the training MAPE level in the following epochs. No model overfit is observed since the difference of 2 error levels is negligible although MAPE of validation dataset is slightly higher than that of training dataset.

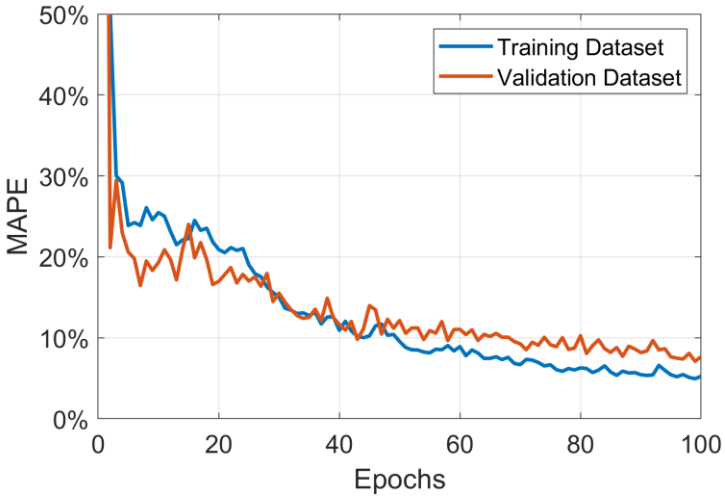


Figure 9.20 The MAPE level of training dataset and validation dataset during training

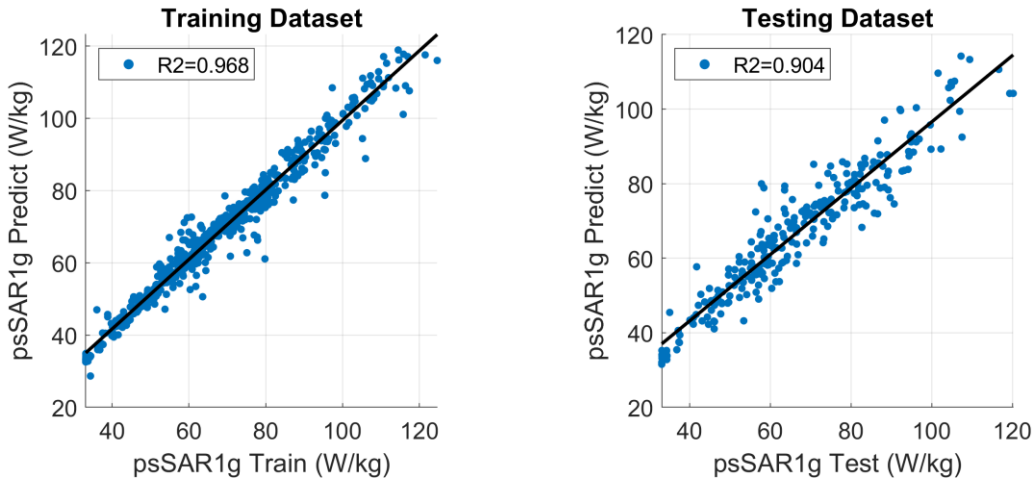
After training, the averaged value for peak SAR, MAE and MAPE is tabulated in Table 9-4. From the table, averaged peak SAR values for training dataset and validation dataset are approximately 67W/kg. Both MAE level are around 5W/kg while testing MAE is higher than training MAE for roughly 1W/kg. For MAPE, both levels are around 5% level while testing MAPE has a slight higher level over 5%.

Table 9-4 Averaged SAR level and error level of training and testing dataset

	Training	Testing
Averaged SAR value (W/kg)	67.61	66.28
MAE (W/kg)	4.58	5.53
MAPE	4.99%	5.85%

For training and testing data correlation, R^2 score is used as the metric. R^2 scores of training dataset and testing dataset are evaluated. The scatter plot of training and testing true and predicted value is shown in Figure 9.21. The R^2 score of training dataset is 0.968 and the R^2 score of training dataset is 0.904. It can be seen from the figure that, for training dataset, majority of the data points are near the x=y reference line with only small deviations. For testing dataset, the data points are more spread around the reference line. For both dataset, there exists some outliers that have significant distance to the line. Data

points that have higher peak SAR values in testing dataset have more outliers, which may be due to the lack of high peak SAR data in training dataset.

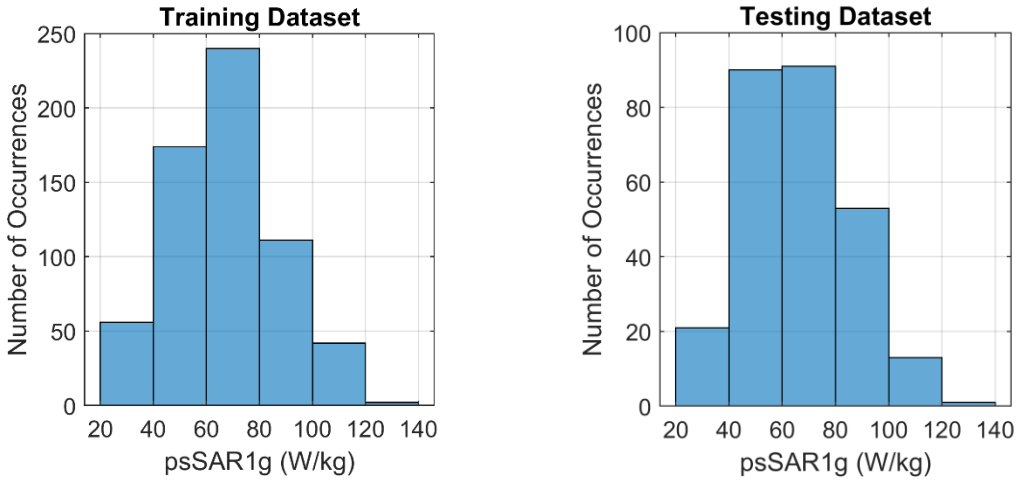


(a) (b)

Figure 9.21 Correlation between predicted and true value for training and testing dataset

9.4.2. Error distribution

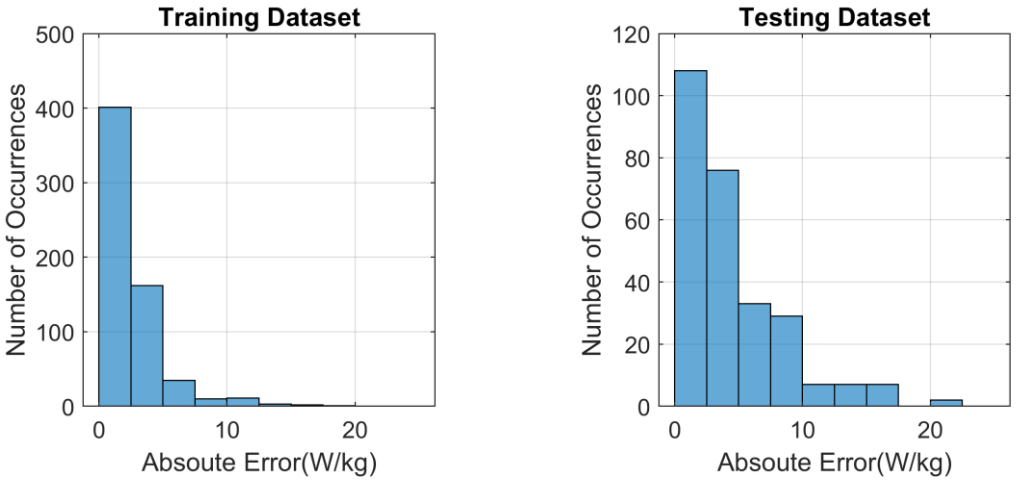
In order to better investigate the details of the error levels, distribution of the absolute error and absolute percentage error are evaluated. Before that, the distribution of peak SAR values of training and testing dataset are shown in Figure 9.22 in order to have a better understanding of data range. From the figure, it can be shown that, all the peak SAR values have a range of [20W/kg, 140W/kg]. For training datasets, majority of the data concentrate in the range of [60W/kg,80W/kg] while for testing dataset majority of data concentrate in the range of [40W/kg,80W/kg].



(a) (b)

Figure 9.22 Distribution of peak SAR values for training dataset (a) and testing dataset (b)

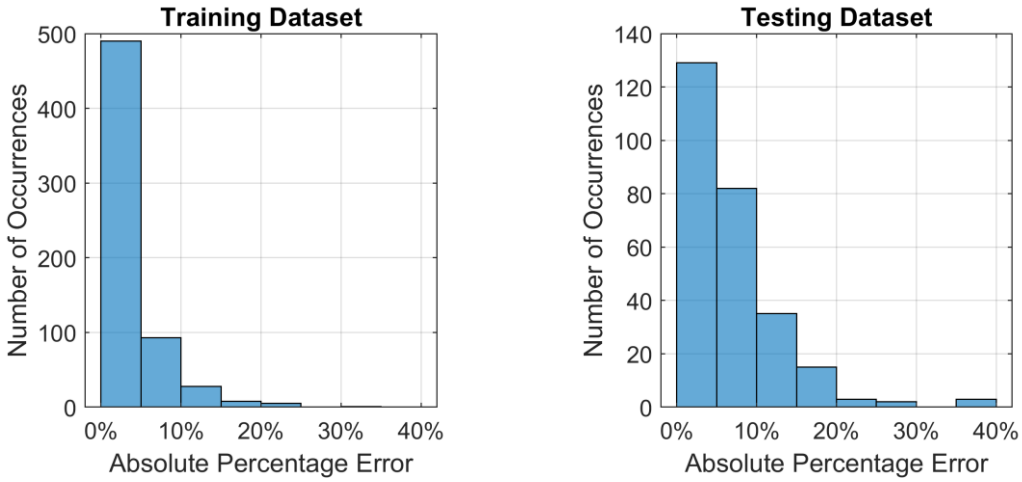
The distribution of absolute error are shown in Figure 9.23. For training dataset, 90% of the data have an absolute error level of less than 5W/kg and 68% of the testing data have an absolute error level of less than 5W/kg. The number of occurrences decreases sequentially for both datasets when the error level is increasing except few outliers at approximately 20W/kg for testing dataset.



(a) (b)

Figure 9.23 Distribution of absolute error for training dataset (a) and testing dataset (b)

The distribution of absolute percentage error are shown in Figure 9.24. For training dataset, 93% of the data have an absolute error level of less than 10% and 78% of the testing data have an absolute error level of less than 10%. A few outliers whose error level is as high as 40% are observed for testing dataset while all the percentage error are less than 25% for training dataset.



(a) (b)

Figure 9.24 Distribution of absolute percentage error for training dataset (a) and testing dataset (b)

10. Discussions

10.1. Prediction Performance

In this research, three CNN models are established for predicting in-vitro and in-vivo RF-induced heating for different PIMDs. All three CNN models receive good convergence with error level decreasing rapidly within the first five or ten epochs. Firstly, it proves the robustness of Adam optimizer. Different from the traditional gradient descent method such as Stochastic Gradient Descent (SGD), Adam optimizer will adjust the learning rate quickly after a quantity of backpropagation procedures. Thus, after several epochs, all training data have participated in the backpropagation procedures for multiple times. The learning rate is adjusted to the "smartest" value. Secondly, the error level has some oscillation after it has decreased to a low level, which is very reasonable since after every backpropagation procedure, the filter weights are slightly changed even if the network has converged. Thus, the output will have small changes while these changes will not bring significant change on the error level. Lastly, the error level of validation data is higher than training dataset, which is also reasonable as the data in validation dataset are not introduced into the backpropagation. So, these new data will have higher error level. It can be observed in some epoch that the validation error level is lower than training error level, which is reasonable as the model is not converged with the filter weights changing significantly. Some coincidence may appear that validation error is lower than training error and this phenomenon will disappear as more epochs are experienced during training.

In this research, the data correlation is all acceptable with all R^2 score larger than 0.90. It means that the proposed CNN models have captured the mapping between incident EM field information, device geometrical information and RF-induced heating. The

training dataset always have higher R^2 score, which correlates with the previous phenomenon that training dataset has lower error level than testing dataset. One needs to be noted that, the R^2 score can only depict the correlation of an ensemble data, which means that the bad correlation caused by outliers that have larger error on prediction result may be hidden by the majority of the data. Thus, the distribution of error levels needs to be further investigated to evaluate the amount of outlier data. This is important since if the CNN model is applied for RF-induced prediction as clinical reference, the possibility of significant overestimation and underestimation needs to be investigated to prevent the incorrect MR safety labeling.

The distribution of MAE and MAPE is investigated. Generally, majority of the training and testing data has low error level while several outliers with large error are observed. The outliers appear in both training dataset and testing dataset, which implies that the RF-induced heating pattern in complicated. Even if training dataset has always received high data correlation and low error level, there are still some device models whose RF-induced heating cannot be predicted by the CNN model with limited error, which shows the restriction of the CNN network that one model can only work in specific domain rather than implement the general prediction.

10.2. Training Dataset Selection

For the selection of training data, the naïve strategy is proposed firstly. This strategy sweeps the number of training data from as few as 10% of the total data to 80% of the total data. The result from the training data amount sweeping for Tibia Plating System heating prediction shows that, using large percentage of total data as training data may not be mandatory as 30% of the total data can achieve an R^2 score of 0.98. This illustrates that the

CNN applications in RF-induced heating prediction are different from traditional CNN applications. For image classification using CNN, large amount of labelled images can be obtained through online resources. Based on that, training dataset always contains large amount of data which are over 50% of total data, like 70%. Features on incident EM field and device geometric information can be extracted using 30% of the total data for Tibia Plating System. Note that, 30% may not be a universal threshold as different incident field and device profiles are expected for different PIMD types. For those PIMDs who have complicated shapes, 30% of total data may not suffice for CNN model to learn all the features. Also, this strategy requires the largest percentage of total data to be obtained, which means that the same number of simulations being performed. For Tibia Plating System application, at least 80% of the EM simulations needs to be finished which are 1133 simulations, which will still cost a significant amount of time. Only the time of 20% of total simulations are saved, which shows the low efficiency of the naïve strategy.

PCA analysis is introduced which will only be performed on the 3-D input matrices and doesn't need any peak SAR data, which means that it can be performed before the batch simulations to provide a guidance on choosing the optimum size of training dataset. Note that, PCA analysis only provides a "lower bound" which means the lowest error level a CNN model can achieve with a training dataset that has the same number of data with the number of PCs. This is based on the assumption that the training data have covered the same variances as what the highest-ranked PCs cover. In reality, the training data always cover less variances. Thus, the error level of selected training dataset is always higher than uncovered variances by highest-ranked PCs with the same size. Currently, there is no method to determine the worst error level a training dataset can achieve. In summary, to

secure a limited error level, the optimum size selection needs to be slightly aggressive while keeping within the smallest possible size.

Moreover, the PCA analysis is not applied on the in-vivo heating prediction. Different from two in-vitro heating prediction study where the mesh information and incident field information are multiplied together, input matrices for in-vivo case concatenates the real and imaginary part of incident E field and conductivity of the local material. So, both information types are not merged together. So, PCA analysis cannot provide the information covered by input when matrices with independent input are used for the analysis. In the future, the appropriate input pattern for PCA analysis that could provide the variances of incident field and mesh information needs to be further investigated.

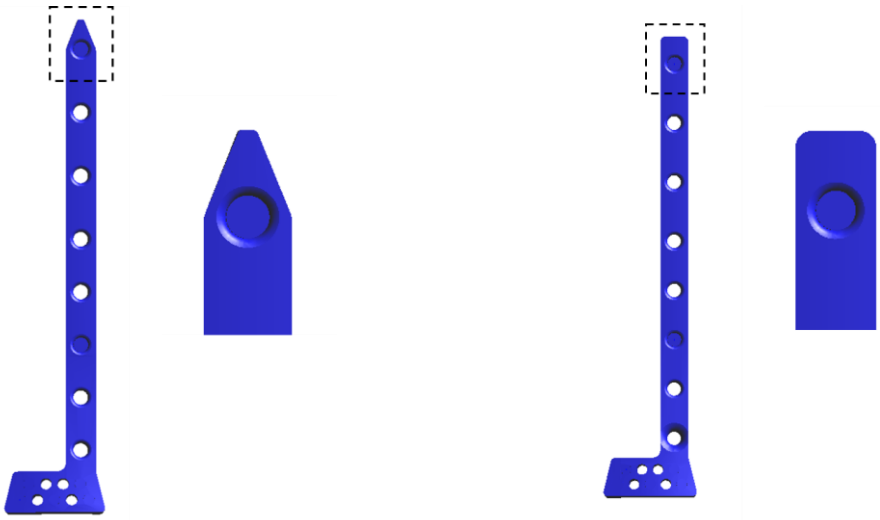
10.3. Sensitivity Analysis of CNN

In this study, CNN models are proved to be able to predict the RF-induced heating of complex-shaped PIMDs with limited error level. However, the robustness of the CNN models is not discussed. Thus, sensitivity analysis is conducted to investigate how sensitive the CNN model is to changes in the input configurations [43]. In this analysis, perturbation method is used [44] on the CNN model of the tibia plate system heating prediction. According to the method, small changes are applied on the input of the CNN model as perturbations and the output is observed to evaluate the error level which measures the sensitivity of the CNN model.

For the CNN model used, input is a combination of device mesh and incident E field. Adding perturbations on the incident E field is inappropriate since it is difficult to change the incident E field pattern in the phantom. Thus, perturbations can be added on the

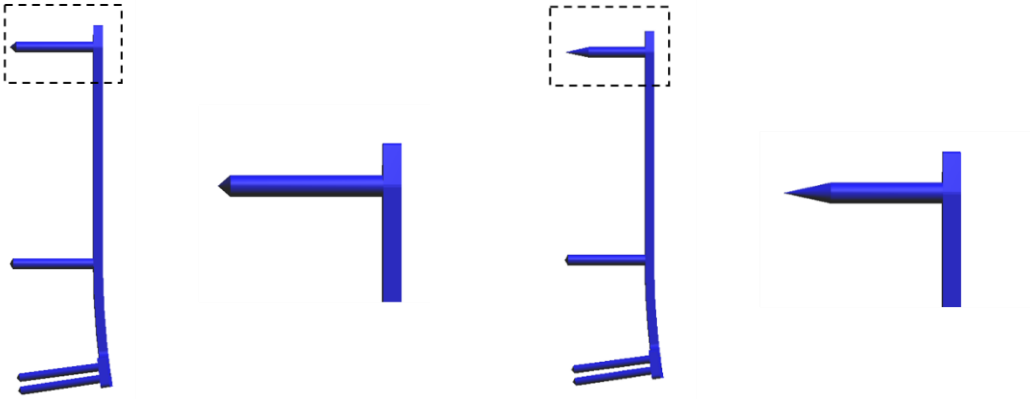
device mesh which means that slight changes can be applied on the device structure. This is more realistic as the clinical devices made from different manufacturers may have slight difference on the device shape. Three kinds of perturbations are used on four tibia plate types in this analysis: the shape of plate tip, the shape of screw tip and the existence of the top screw.

Firstly, the width of plate tip of all tibia plate devices used for training CNN model is half of the plate width. To create the perturbation, several plate device models are created so that the width of plate tip ranges from 20% of the plate width to 100% of the plate width. The shortest plate tip and the longest plate tip pattern are shown in Figure 10.1. Then, the tip length of the top screw is redesigned. Original plate device models used for CNN training have the tip length of 1mm. New models are designed so that the tip length of top screw ranges from 2mm to 8mm, which is shown in Figure 10.2. For new generated models, only one perturbation appears on each model. Screws with length of 10mm, 30mm and 60mm are used at fixed positions.



(a) (b)

Figure 10.1 Shape of plate tip: 20% of plate width (a) and 100% of plate width (b)



(a) (b)

Figure 10.2 Shape of top screw tip: 2mm (a) and 8mm (b)

Totally, 162 new models are developed and 162 EM simulations are performed using the same simulation configuration. Mesh information, incident E field and peak SAR values are extracted after simulation and postprocessing is conducted to form the input of CNN model. The input of perturbated model data is passed to the pre-trained CNN model to obtain the predicted SAR values. The sensitivities of SAR with respect to the perturbation size are investigated first. Two quantities are defined as the measure of the SAR sensitivity. The first quantity is defined as

$$s_1 = \frac{\left|\mathrm{SAR}_{originalTrue} - \mathrm{SAR}_{newPredicted}\right|}{\Delta l} \tag{35}$$

and the second quantity is defined as

$$s_2 = \frac{\left|\mathrm{SAR}_{newTrue} - \mathrm{SAR}_{newPredicted}\right|}{\Delta l}, \tag{36}$$

where $\mathrm{SAR}_{originalTrue}$ is the true SAR value of the original device model with no perturbation, $\mathrm{SAR}_{newPredicted}$ is the predicted SAR value of perturbated device model, $\mathrm{SAR}_{newTrue}$ is the true SAR value of perturbated device model and Δl is the length difference between the

original model size and the perturbated model size. The first quantity measures how sensitive the predicted SAR values are comparing to the original training data and the second quantity measures how sensitive the predicted values are comparing to their own true values. Besides, the MAPE levels are evaluated between predicted values and true values of the perturbation models to measure the robustness of the CNN model with the effect of each perturbation.

s_1 and s_2 with different plate tip width is tabulated in Table 10-1 and Table 10-2. It can be observed that, the sensitivity value is higher for the models with 40% and 60% of plate width as tip width. The reason is very straightforward: the original model's plate tip width is 50% of the plate width so the Δl is smaller comparing to other widths. With the predicted SAR values have small differences between models, a sub-millimeter Δl will cause significant change on the sensitivity values. The maximum sensitivity reaches to over 9 W/kg/mm, which is an acceptable amount since the averaged peak SAR value is around 200 W/kg.

Table 10-1 s_1 with plate tip width perturbation (unit: W/kg/mm)

Screw Length (mm)	Plate Tip Width (% of plate width)							
	20%	30%	40%	60%	70%	80%	90%	100%
10	0.81	1.39	3.98	3.52	2.56	1.79	1.29	1.17
30	2.21	3.29	6.87	6.34	2.95	1.75	1.34	2.07
60	2.45	3.69	7.65	7.25	3.80	2.62	1.98	2.08

Table 10-2 s_2 with plate tip width perturbation (unit: W/kg/mm)

Screw Length (mm)	Plate Tip Width (% of plate width)							
	20%	30%	40%	60%	70%	80%	90%	100%
10	1.30	1.51	6.72	5.21	2.83	2.77	2.64	2.33
30	3.24	4.47	8.54	6.66	3.00	1.64	1.17	1.88
60	3.11	4.53	9.21	9.07	4.77	3.30	2.51	2.14

s_1 and s_2 with different screw tip length is tabulated in Table 10-3 and Table 10-4. It can be observed that, the sensitivity value decreases when the screw tip length is increasing. The reason is similar with the plate width perturbation that the original model has a screw tip length of 1mm and Δl is larger for longer screw tip. Also, the sensitivity value is larger for models with longer screws. The peak SAR location is around the screw tip according to observations. Thus, the change of screw tip will have more significant sensitivity. The maximum sensitivity value reaches over 8.500 W/kg/mm.

Table 10-3 s_2 with screw tip length perturbation (unit: W/kg/mm)

Screw Length (mm)	Screw Tip Length (mm)						
	2	3	4	5	6	7	8
10	3.77	1.71	1.55	1.29	1.08	0.88	3.77
30	3.10	3.13	2.61	2.63	2.64	3.37	3.10
60	6.29	3.21	2.13	1.66	1.36	1.11	6.29

Table 10-4 s_2 with screw tip length perturbation (unit: W/kg/mm)

Screw Length (mm)	Screw Tip Length (mm)						
	2	3	4	5	6	7	8
10	3.46	2.83	2.60	2.75	2.68	2.37	3.46
30	6.46	2.77	2.05	2.46	2.24	1.87	6.46
60	8.55	3.99	2.53	2.21	1.45	0.91	8.55

In summary, the CNN model used in this study is robust and stable and is immune to the perturbation on slight changes on the device shape.

10.4. Stability Analysis of CNN

Beside the sensitivity, the stability of a CNN model is an important measure of a good model. The stability of a CNN model is defined by the convergence of the testing dataset error level when more testing data are introduced for a pre-trained model. If the prediction error stays at certain level when the testing data keep increasing, it means that the pre-trained CNN model is stable.

In this analysis, the CNN model of the tibia plate system heating prediction is used for error level investigation. It is obtained from previous study that the training dataset can have training result with 30% of the total data. Thus, in this analysis, a pre-trained CNN model using 30% of the total data as training data is selected, which is 425 data. The rest 70% data are used as testing data, which is 991 data. All data are shuffled so that the models with all geometrical variations are included in both training and testing dataset. After training, the testing data are applied to the CNN model one by one. Every time a predicted value is obtained, the mean, standard deviation and the maximum error are evaluated iteratively for all the previous-tested data.

The mean, standard deviation and the max value of the absolute error with respect to testing data amount are shown in Figure 10.3. It can be seen that, the mean and the standard deviation have fluctuation for the first 200 testing data. Then both value converge to certain level with the MAE reaches 4W/kg and standard deviation reaches 5W/kg. The maximum error stays stable during the introduction of 200th data to 800th data while finally increases to 45W/kg with more testing data are introduced.

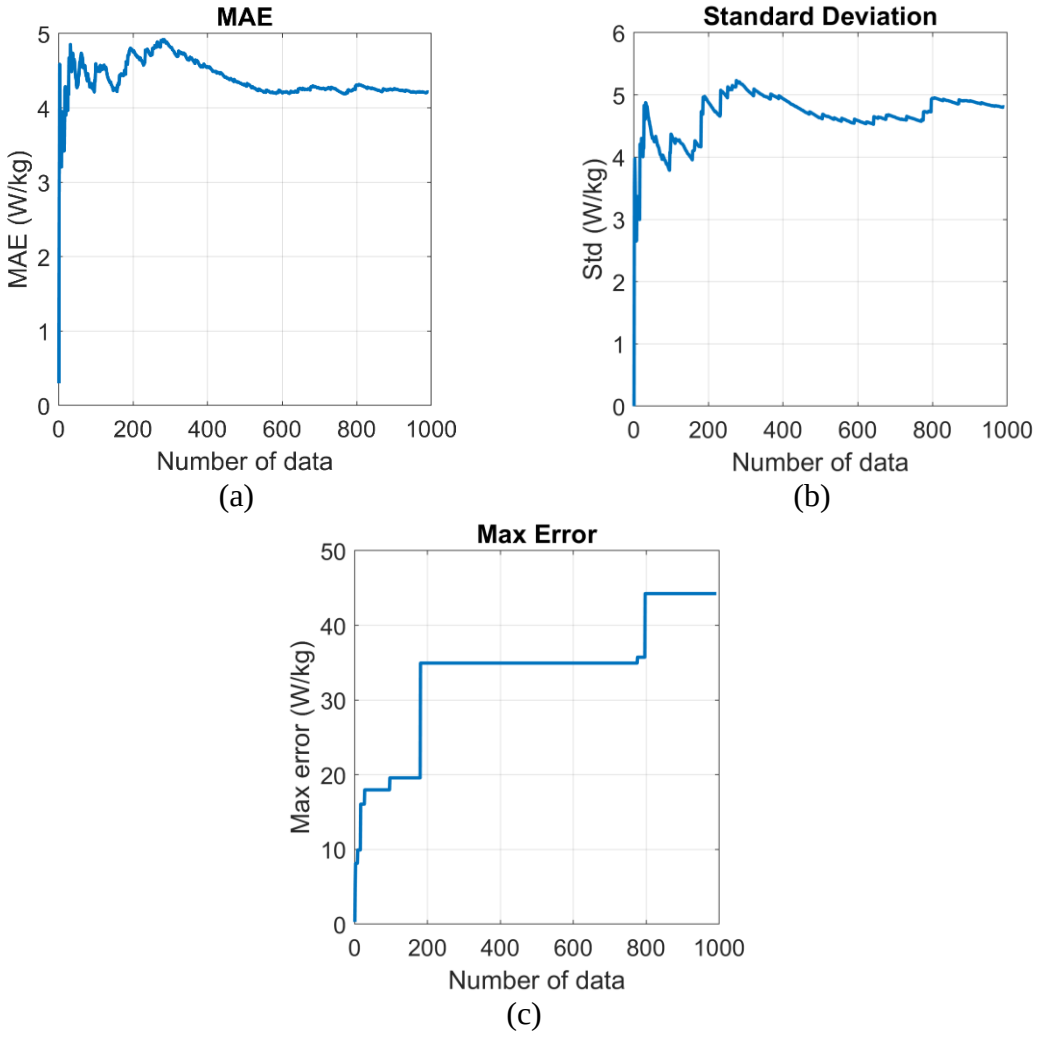


Figure 10.3 Mean (a), standard deviation (b) and max value (c) of testing data

From the figure, the mean and standard deviation are stable with the introduction of more testing data which can prove that the pre-trained CNN model are stable for tibia

plate models with geometrical variations. The maximum error keeps increasing since the order of testing data is tightly related to the appearance of maximum error. If the data with higher error is used for testing in early phase, the maximum error will stay stable for all the following testing. Also, if more testing data are introduced, the maximum error will eventually converge as the prediction loss is bounded.

Noted that, the geometrical variations of the device models used for training and testing CNN model are not boundless but have a range, which will ensure that the models with similar structures are used for training and testing. If device models whose geometrical configuration are totally random and are out of the training data range are used for testing, the error level may not converge as the CNN model are seeing new structures and the prediction may be unpredictable.

11.Conclusions

In this research, the problem of RF-induced heating of PIMDs are addressed and the possibility of predicting RF-induced heating of PIMDs are investigated. Computational models of Tibia Plating System and Spinal Fixation System are developed with geometrical variations. ASTM phantom is used for in-vitro scenario and Duke model from Virtual Family is used for in-vivo scenario. After that, the incident E field information is exported and device geometrical information is extracted as the pattern of material indexes or electrical conductivity. Both information types are merged through element multiplication or concatenation to form the input of CNN model. CNN models derived from AlexNet with different architectures are applied as the regression model. After training, network convergence, data correlation and error distribution are further investigated. PCA analysis is performed on the input matrices for assisting the selection of training dataset.

From the results, it can be summarized that the CNN model has the ability to capture the feature of the training data quickly as the network model receives rapid convergence. No overfit is observed as the difference between training dataset error and validation dataset error can be neglected. All data have acceptable correlations which shows that the overall prediction performance is acceptable. Although one can be seen from the error distribution analysis that outliers with significant error level appears for both training dataset and testing dataset, future stochastic analysis can be performed to infer the possibility of overestimation and underestimation to calibrate the CNN model.

In terms of training data selection, the naïve strategy uses an exhaustive method by sweeping different number of training data for CNN model training to find the optimal training dataset size. This method seems to be not practical as it still needs the heating

results from large number of simulations. PCA analysis is introduced to provide a standard of what best performance a training dataset can achieve with the same size as the highest-ranked PCs. One advantage of it is that no simulations are needed before the analysis so that the number of simulations can be determined by the analysis which will save a lot of time for running majority of simulations.

Overall, the implementation of MRI RF-induced heating prediction using Neural Network method is proposed. Acceptable network convergence and data correlation are observed for in-vitro and in-vivo scenario. PCA analysis is used for determining the appropriate training dataset size as an *a priori* method which can reduce the time and computational burden. In the future, potential study includes the PCA analysis for in-vivo heating prediction and cross human model heating prediction applications. Also, the prediction performance on worst-case heating needs to be investigated in the future.

REFERENCES

[1] J. A. Johnson, "FDA Regulation of Medical Devices," 25 June 2012. [Online].

[2] ASTM F2182-02a, "Standard test method for measurement of radio frequency induced heating on or near passive implants during magnetic resonance imaging," ASTM, 2011.

[3] F. G. Shellock and J. V. Cures, "MR Procedures: Biologic Effects, Safety, and Patient Care," Radiology, vol. 232, no. 3, pp. 635-652, 2004.

[4] W. Kainz, G. Neubauer, R. Uberbacher, F. Alesch and D. D. Chan, "Temperature measurement on neurological pulse generators during MR scans," BioMedical Engineering OnLine, vol. 1, no. 1, pp. 1-8, 2002.

[5] E. Mattei, M. Triventi, G. Calcagnini, F. Censi and P. Bartolini, "Radiofrequency Dosimetry in Subjects Implanted with Metallic Structures Undergoing MRI: a Numerical Study," American Journal of Biomedical Sciences, vol. 1, no. 4, pp. 373-384, 2009.

[6] W. McCulloch and W. Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biophysics, vol. 5, pp. 115-133, 1943.

[7] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[8] F. G. Shellock and A. Spinazzi, "MRI safety update 2008: part 2, screening patients for MRI," American Journal of Roentgenology, vol. 191, no. 4, p. 1140, 2008.

[9] P. L. Davis, L. Crooks, M. Arakawa, R. McRee, L. Kaufman and A. R. Margulis, "Potential Hazards in NMRImaging: Heating Effects of Changing Magnetic Fields and RF Fields on Small Metallic Implants," American journal of roentgenology, vol. 137, no. 4, pp. 857-860, 1981.

[10] R. Kumar, R. A. Lerski, S. Gandy, B. A. Clift and R. J. Abboud, "Safety of Orthopedic Implants in Magnetic Resonance Imaging: An Experimental Verification," *Journal of Orthopaedic Research,* vol. 24, no. 9, pp. 1799-1802, 2006.

[11] F. G. Shellock, "MR imaging and cervical fixation devices: evaluation of ferromagnetism, heating, and artifacts at 1.5 Tesla," *Magnetic resonance imaging,* vol. 14, no. 9, pp. 1093-1098, 1996.

[12] F. G. Shellock, M. Nogueira and S. Morisoli, "Hr Imaging and Vascular Access Ports: Ex Vivo Evaluation of Ferromagnetism, Heating, and Artifacts at 1.5 t," *Journal of Magnetic Resonance Imaging,* vol. 5, no. 4, pp. 481-484, 1995.

[13] F. G. Shellock and V. J. Shellock, "Vascular Access Ports and Catheters: Ex Vivo Testing of Ferromagnetism, Heating, and Artifacts Associated with MR Imaging," *Magnetic resonance imaging,* vol. 14, no. 4, pp. 443-447, 1996.

[14] Y. Liu, J. Chen, F. G. Shellock and W. Kainz, "Computational and Experimental Studies of An Orthopedic Implant: MRI-related Heating at 1.5-T/64-MHz and 3-T/128-MHz," *Journal of Magnetic Resonance Imaging,* vol. 37, no. 2, pp. 491-497, 2013.

[15] R. Guo, J. Zheng and J. Chen, "MRI RF-Induced Heating in Heterogeneous Human Body with Implantable Medical Device," in High-Resolution

Neuroimaging - Basic Physical Principles and Clinical Applications, Ahmet Mesrur Halefoğlu, IntechOpen, 2018.

[16] Y. Liu, J. Shen, W. Kainz, S. Qian, W. Wu and J. Chen, "Numerical Investigations of MRI RF Field Induced Heating for External Fixation Devices," *Biomedical engineering online,* vol. 12, no. 1, pp. 1-14, 2013.

[17] X. Huang, Z. Wang, J. Chen and J. Zheng, "Numerical Study on MRI RF Heating for Circular External Fixators under 1.5 T MRI," in *2018 IEEE Symposium on Electromagnetic Compatibility, Signal Integrity and Power Integrity*, Long Beach, 2018.

[18] Y. Liu, W. Kainz, S. Qian, W. Wu and J. Chen, "Effect of Insulating Layer Material on RF-induced Heating for External Fixation System in 1.5 T MRI System," *Electromagnetic biology and medicine,* vol. 33, no. 3, pp. 223-227, 2014.

[19] X. Ji, J. Zheng and J. Chen, "Numerical Evaluation of RF-induced Heating for Various Esophageal Stent Designs under MRI 1.5 Tesla System," *Electromagnetic Biology and Medicine,* vol. 36, no. 4, pp. 379-386, 2017.

[20] X. Ji, J. Zheng, R. Yang, K. Wolfgang and J. Chen, "Evaluations of the MRI RF-induced Heating for Helical Stents under a 1.5 T MRI System," *IEEE Transactions on Electromagnetic Compatibility,* vol. 61, no. 1, pp. 57-64, 2018.

[21] E. Lucano, M. Liberti, G. G. Mendoza, T. Lloyd, M. I. Iacono, F. Apollonio, S. Wedan, W. Kainz and L. M. Angelone, "Assessing the Electromagnetic Fields Generated by a Radiofrequency MRI Body Coil at 64 MHz: Defeaturing Versus

Accuracy," *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 8, pp. 1591-1601, 2015.

[22] R. Yang, J. Zheng, J. Chen and W. Kainz, "Comparison Study of RF-induced Heating in Leg Phantom with Circular External Fixator for TEM and Birdcage Coils at 3 T," in *2017 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity*, Washington, D.C, 2017.

[23] R. Yang, J. Zheng, W. Kainz and J. Chen, "Numerical Investigations of MRI RF-induced Heating for External Fixation Device in TEM and Birdcage Body Coils at 3 T," *IEEE Transactions on Electromagnetic Compatibility*, vol. 60, no. 3, pp. 598-604, 2017.

[24] R. Yang, J. Zheng, S. Song, R. Guo and J. Chen, "Numerical Investigations of MRI RF-induced Heating for Passive Implants in Birdcage and TEM Body Coils at 3 Tesla," in *2020 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity*, 2020.

[25] M. Xia, J. Zheng, R. Yang, S. Song, J. Xu, Q. Liu, W. Kainz, S. Long and J. Chen, "Effects of patient orientations, landmark positions, and device positions on the MRI RF-induced heating for modular external fixation devices," *Magnetic Resonance in Medicine,* vol. 85, pp. 1669-1680, 2021.

[26] X. Ji, J. Zheng, R. Yang, W. Kainz and J. Chen, "Evaluations of the MRI RF-Induced Heating for Helical Stents Under a 1.5T System," *IEEE Transactions on Electromagnetic Compatibility,* vol. 61, pp. 57-64, 2019.

[27] J. Zheng, D. Li, J. Chen and W. Kainz, "Numerical Study of SAR for Multi-Component Orthopaedic Hip Replacement System During MRI," in *IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 2016.

[28] J. Zheng, R. Yang and J. Chen, "Fast prediction of MRI RF-induced heating for implantable plate devices using neural network," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, 2017.

[29] Q. Lan, J. Zheng and J. Chen, "Predicting MRI RF Exposure for Complex-shaped Medical Implants Using Artificial Neural Network," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, 2019.

[30] J. Zheng, Q. Lan, X. Zhang, W. Kainz and J. Chen, "Prediction of MRI RF Exposure for Implantable Plate Devices Using Artificial Neural Network," *IEEE Transactions on Electromagnetic Compatibility,* vol. 62, no. 3, pp. 673-681, 2019.

[31] Q. Lan, J. Zheng, J. Chang, R. Guo, W. Kainz and J. Chen, "Predicting MRI RF Exposure for Passive Implantable Medical Devices Using a Mesh-based Convolutional Neural Network," in *IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, 2021.

[32] A. Geron, Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, O'Reilly Media, Inc., 2019.

[33] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010.

[34] International Electrotechnical Commission, "International standard IEC 60601 medical electrical equipment. Part 2–33: Particular requirements for the basic safety and essential perfor- mance of magnetic resonance equipment for medical diagnosis.," 2010.

[35] IEEE P62704-1, "Standard for Determining the Peak Spatial Average Specific Absorption Rate (SAR) in the Human Body from Wireless Communications Devices, 30 MHz to 6 GHz. Part 1," IEEE, 2020.

[36] K. Pearson, "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559-572, 1901.

[37] L. I. Smith, "A tutorial on Principal Components Analysis," 26 Febuary 2002. [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

[38] J. VanderPlas, Python Data Science Handbook, O'Reilly Media, Inc., 2016.

[39] Arthrex, "Distal Tibia Plating System," [Online]. Available: https://www.arthrex.com/foot-ankle/distal-tibia-plating-system.

[40] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.

[41] F. Chollet, "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras.

[42] D. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," *arXiv preprint arXiv*, vol. 1412, no. 6980, 2014.

[43] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification." *arXiv preprint arXiv*:1510.03820, 2015.

[44] M. Cao, N. F. Alkayem, L. Pan, D. Novak, J. L. G. Rosa, "Advanced methods in neural networks-based sensitivity analysis with their applications in civil engineering." *Artificial neural networks: models and applications, Rijeka, Croatia, IntechOpen* , pp. 335-353, 2016.