Communication

Stroke Localization Using Multiple Ridge Regression Predictors Based on Electromagnetic Signals

Shang Gao [1] , **Guohun Zhu** [1,*] , **Alina Bialkowski** [1] **and Xujuan Zhou** [2]

1   School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia
2   School of Business, University of Southern Queensland, Springfield, Brisbane, QLD 4300, Australia
*   Correspondence: g.zhu@uq.edu.au

Abstract: Localizing stroke may be critical for elucidating underlying pathophysiology. This study proposes a ridge regression–meanshift (RRMS) framework using electromagnetic signals obtained from 16 antennas placed around the anthropomorphic head phantom. A total of 608 intracranial haemorrhage (ICH) and ischemic (IS) signals are collected and evaluated for RRMS, where each type of signal contains two different diameters of stroke phantoms. Subsequently, multiple ridge regression predictors then give the target distances from the antennas and mean shift is used to cluster the predicted stroke location based on these distances. The test results show that the training time and economic cost are significantly reduced as the average prediction time only takes 0.61 s to achieve an accurate result (average position error = 0.74 cm) using a conventional laptop. It has great potential to be used as an auxiliary standard medical method, or rapid diagnosis of stroke patients in underdeveloped areas, due to its rapidity, good deployability, and low hardware cost.

Keywords: antenna array; electromagnetic signal; machine learning; mean-shift; ridge regression; stroke; vector network analysis

MSC: 68T01

1. Introduction

Stroke is a medical emergency in which poor blood flow to the brain causes cell death and may result in permanent disability or death if not treated in a timely manner. According to the Australian Bureau of Statistics, stroke was recorded as the underlying cause of 8400 deaths in 2018, accounting for 5.3% of all deaths in Australia [1]. Apart from identifying the type of stroke, which can be either intracerebral hemorrhage (ICH) or ischemic stroke (IS), a critical task is to retrieve stroke position in order to implement the medical intervention as soon as possible. At present, clinical diagnosis of stroke mainly includes X-ray computed tomography (X-ray CT) and magnetic resonance imaging (MRI). However, both types of devices have poor portability, where the former has ionizing radiation and the latter has high cost and slow diagnosis.

In recent years, the feasibility of microwave radio frequency (RF) detecting brain stroke, breast tumors, and other internal lesions has been studied and demonstrated in many works, as dielectric properties of normal human tissue and diseased tissue vary greatly under microwave frequency [2–6]. Relatively, RF technology has not only the advantages of non-ionizing radiation, short scanning time, etc. but also great deployability in mobile clinics or ambulances [7]. Generally, the RF localization method can be divided into two categories: microwave imaging (MWI) and artificial intelligence (AI). In the MWI method, the antenna array transmits electromagnetic (EM) waves to penetrate the area to be measured and receive scattered signals to reconstruct the dielectric distribution of the area to be measured. A conventional delay and sum (DAS) with coherence weighting is reported to reconstruct hemorrhagic stroke images in a simulated 3D head model [3].

Another representative example of the MWI method is microwave tomography (MWT) based on the DBIM-TwIST algorithm [8]. However, the inherent non-linearity and ill-posedness of inverse scattering reconstructing images remain challenging; problems such as artifacts, low resolution, and long reconstructing time prevent their practical use in most clinical and non-clinical situations.

Reliable stroke location with immediate feedback without human visual discrimination is a challenge in the field. The application of AI methods to the classification of stroke subtypes has been explored [7,9,10]. Over the years, AI methods have also been increasingly applied to stroke location prediction and have shown great potential [9–12]. Many studies focus on the localization of hemorrhagic stroke and made progress [9,11,12]. For example, a framework for stroke detection on a deep-learned stochastic model was proposed recently, achieving mean and median IoU (intersection over union) of 0.84 and 0.94, respectively [9]. A few works have validated the localization of ischemic stroke. In 2021, Salucci et al. proposed a Multi-Step Learning-by-Examples Strategy for Real-Time Brain Stroke, where support vector regression (SVR) is used to estimate stroke coordinates [10]. However, only 25 possible stroke positions are considered in the experiment, and only stroke phantoms of the same size are used in simulation, lacking generality and scalability when applied in real situations. Therefore, the research gap in the development of localization algorithms that are not limited by stroke phantom size, sub-type, and locations is generated.

In this study, a ridge regression–meanshift (RRMS) framework is proposed to localize brain stroke regions using EM signals. Firstly, scatter parameters in 751 frequency points are collected through a 16-port antenna array vector network analysis (VNA). Secondly, multiple ridge regression (RR) predictors are constructed based on the scattering features from EM signals. Based on the predicted distances, the mean shift clustering method (MS) is used to find the densest point clusters, and take the cluster centroid as the center coordinate of the predicted target. The main contributions of this paper are as follows: Firstly, this paper uses experimental data rather than simulated data, covering more than 100 different stroke clot locations for modeling. Secondly, stroke classification is not required in advance in the training phase, but two different stroke subtypes, ICH and IS, and stroke clots of different sizes (20mm and 30mm) are mixed to improve the robustness of the model. Finally, it bypasses traditional image reconstruction, providing the location of the stroke blood clot in a shorter period of time without artifacts and low-resolution interference.

This paper's structure is organized as follows: The description of the experimental environment and methods are given in Section 2. Then the experiment procedures are provided in Section 3; Section 4 reports representative experiment results to verify the effectiveness of the proposed method. Finally, we summarize our findings and discuss comparisons with existing similar studies, as well as future work, in Section 5.

2. Experimental Environment

The data used in this experiment were obtained from phantom measurements, the anthropomorphic head model used is EMvision 3B headset, and the vector network analyzer (VNA) used was M9800A.

2.1. Anthropomorphic Brain Stroke Model

As shown in Figure 1a, the anthropomorphic head phantom was implemented by a skin–skull bucket (SSB), being filled with the average brain-emulating liquid which was used to mimic the average dielectric measurements of human brain tissue. Additionally, solid emulating cylinders of different sizes and specifications were used to mimic various stroke targets, which provided a more realistic design approach to represent stroke data in real-world situations. Figure 1b shows the brain stroke targets (phantoms) we used in our experiments, which were coded as HL, HM, IM, and IL from left to right, where H was hemorrhagic and I was ischemic, L referred to large size target and M referred to medium size target. A ground truth example can be seen in Figure 1c The descriptions and specifications of these stroke phantoms are shown in Table 1.
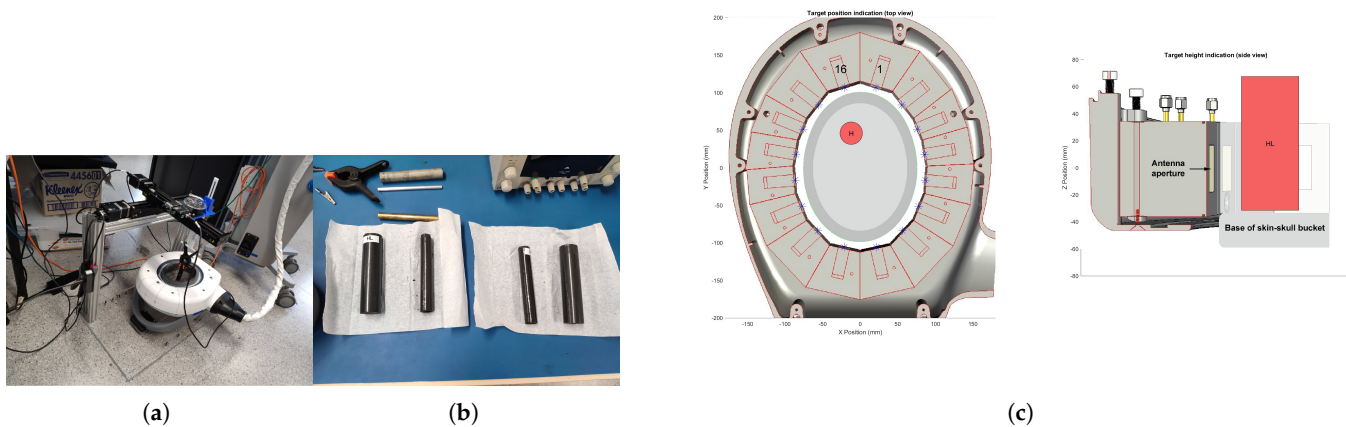
Figure 1. (**a**) Anthropomorphic head model within the array imaging domain upside-down, filled with average brain-emulating liquid. (**b**) solid cylinders HL, HM, IM, and IL from left to right. (**c**) An example of a ground truth image of an HL target positioned close to the periphery of the head phantom.

Table 1. The specifications of 608 brain stroke targets used in experiments.

Feature	Description	Sample	Diameter
HL	ICH Stroke(Target Type: L)	134	30 mm
HM	ICH Stroke(Target Type: M)	170	20 mm
IL	IS Stroke(Target Type: L)	134	30 mm
IM	IS Stroke(Target Type: M)	170	20 mm

2.2. RF Data Collection

All signals were generated and received using the VNA controlled by a laptop. To obtain the data, A 16-antenna array was designed and placed around the SSB. Among them, each antenna was both a transmitter and a receiver. The operating frequency of the antenna array was between 0.5–2 GHz, and a total of 751 frequency points were collected with 2 MHz steps. Therefore, a total of $751 \times 16 \times 16$ scattering parameters (data points) were captured for each case. In microwave RF signals, the dielectric properties of different stroke locations in the head can be reflected in the scattering parameters. Therefore, the scattering parameter matrix was collected in each movement and stored as a Touchstone format file by moving these targets along an 8 mm spaced grid pattern in X-coordinate and Y-coordinate.

2.3. Methodology

In this study, a novel two-stage stroke localization method is proposed that eschews traditional electromagnetic imaging methods.

Ridge regression (RR) is a biased estimation regression method proposed by Hoerl and Kennard [13], dedicated to linear data analysis. RR is essentially an improved least-squares estimation method. It prevents parameter overfitting by adding L2 regularization terms to standard linear regression. Assuming that dataset D has *N* samples, $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\}$, where $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ is the *i*th input variable (dimension = *p*). The solution of the ridge regression coefficients depends on the optimization:

$$\hat{\beta} = \text{argmin}_\beta \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}, \tag{1}$$

where y_i is defined as the corresponding distance between the stroke target from antenna port *i*. x_{ij} here refers to power transferred from other ports to port *i* when *i* is matched. β_i

and β_0 are parameters determined during learning process, where β_j is the coefficient term and β_0 is the regression error term. The penalty function of the L2-norm is represented by $\lambda \sum_{j=1}^{p} \beta_j^2$, where λ is the penalty coefficient, determined by grid search in the experiment.

2.4. MeanShift Clustering

MeanShift(MS) is a density-based nonparametric clustering algorithm, proposed by Comaniciu and Meer [14]. MS works by setting a kernel function to estimate the density of the samples. In this study, it is used to find the final centroid. Suppose there is a dataset with n samples and d dimensions defined as $D = \{x_1, x_2, x_3, \ldots x_n\}, D \in R^d$, then the bandwidth parameter h can be automatically estimated using *cluster.estimate_bandwidth* in *sklearn* package in python. The kernel density function is estimated as:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2}$$

The effect of clustering varies with bandwidth value h. The radially symmetric kernel is defined as:

$$K(\mathbf{x}) = c_{k,d} k\left(\|\mathbf{x}\|^2\right), \int c_{k,d} \cdot K(x)dx = 1 \tag{3}$$

where $c_{k,d}$ represents normalization constant which assures $K(\mathbf{x})$ integrates to 1. The goal of MS is to move the sample points in the direction of the local density increase, and the gradient is the direction in which the function increases the fastest.

3. Experimental Procedures

This experiment is mainly divided into three steps: data preprocessing, feature selection, and ridge regression–mean shift (RRMS) localization.

3.1. Data Preprocessing

In the experiment, four types of HL, HM, IM, and IL targets are mixed together to improve the generality and scalability of this framework and are divided into the training set and test set according to the ratio of 7:3. Table 2 shows the specifications of the training and test sets.

Table 2. Description of the training set and test set.

DataSet	Training Data (70%)	Testing Data (30%)
ICH + IS (Total: 608)	425	183

Before analyzing the scattered signal, according to the fact that most machine learning models do not support complex-number input, all experiments mentioned in this article extract the gain ratio(dB) of the original signal as features, where the gain ratio represents the power loss and gain ratio of antennas.

3.2. Feature Selection

Electromagnetic signals are a relatively sophisticated ultra-high-dimensional dataset. For each stroke data, there are $751 \times 16 \times 16$ S-parameter features. There may be a large number of irrelevant or redundant features in the dataset, and too many features may lead to problems such as model over-fitting and long running time.

Therefore, the s parameters corresponding to each antenna port are extracted separately before training. Afterward, the variance filtering method is used for feature selection. The frequency points with a variance less than the threshold mean the divergence of these features is too weak. Thus, the frequency points with low divergence and low correlation can be filtered out. The results of the five-fold cross-validation after feature selection are given in the next section.

3.3. RRMS Based Localization

This section presents a two-phase stroke localization method by combining ridge regression (RR) and meanshift clustering (MS) Specifically, RR is used to predict the distance between each antenna and the target in the modeling phase. For each antenna i, a predictor is built separately. The RR parameter alpha is found through grid search on the training set to find the best parameters. For all predictors, the distances from the target to each antenna are obtained after training, denoted as $d_1, d_2, d_3, \ldots, d_{16}$. In the localization phase, circular trajectories can be simulated with the center of the antenna i and the radius of d_i. MS is used to find the densest trajectories within the target area, which is the predicted target position. Figure 2 shows the process of the experiment.
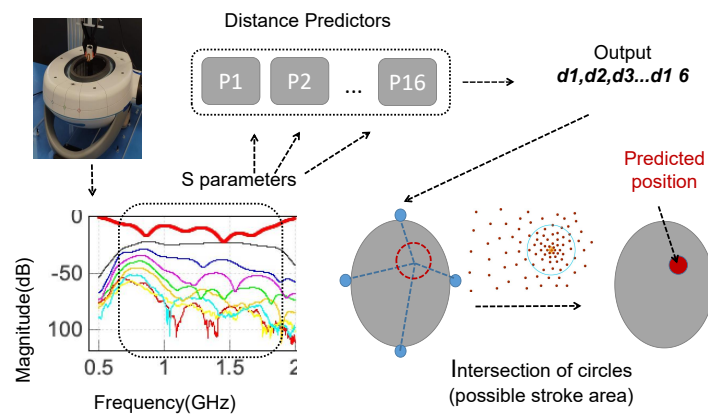

Figure 2. The two stages of proposed framework: prediction and localization.

4. Experimental Results

Across all reported results, the training phase takes 46.70s, and the average test time takes 0.61s, recorded by a standard laptop using a CPU with model AMD Ryzen$^{\text{TM}}$ with 16 GB RAM memory.

### 4.1. Feature Filtering Results

Figure 3 demonstrates the performances of evaluating predictor generalization using five-fold cross-validation (CV) after feature selection (threshold is set to 0.1). Each predictor performs differently but all exceed an average CV score of 0.9. Among them, predictor-16 performs the best on the training set, with an average CV score of 0.996.
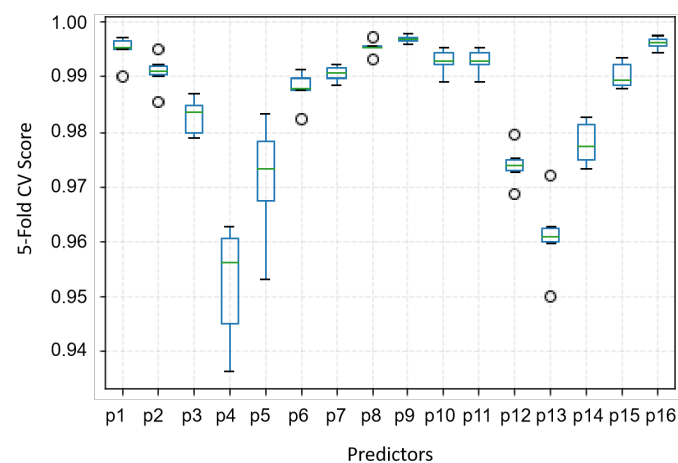
Figure 3. Five-fold cross validation score of d_1 to d_{16} on training data.

4.2. Predictor Performance on Test Set

To obtain predictors' performances on the test set, four predictors with the highest average scores on the training set are selected, namely port-0, port-7, port-8, and port-15. The corresponding RMSE (root mean squared error) is given in Figure 4. The results show that the RMSE of these four predictors on the test set is between 0.14 and 0.18.
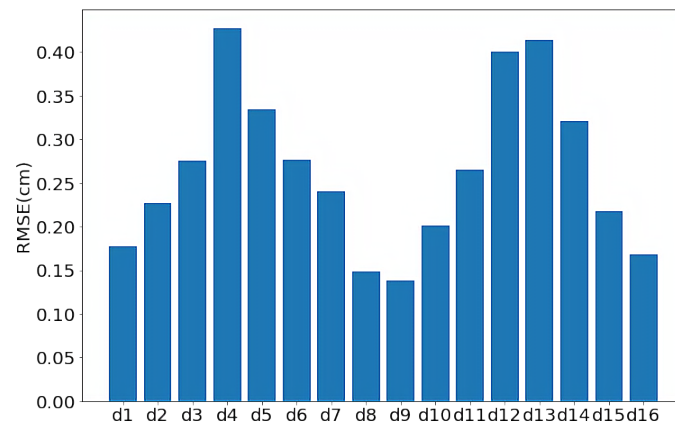
**Figure 4.** RMSE of predicted distance on test data.

4.3. Intersection of Localization Results

This section applies the intersection of circles to obtain the stroke targets. Line intersection methods have been applied in [12] to localize the stroke targets. The intersection of at least two circular trajectories can obtain a potential existing location of the target. Four circular trajectories based on the predicted distance d_0, d_7, d_8, d_{15} of the above four predictors are drawn, and MS is applied to estimate the target location with the highest density of scatter points on the plane (the simulated stroke location).

The range and distribution of localization errors are given in Figure 5. It can be observed that there is greater variability in the localizing error of the X coordinate system and larger outliers than the localizing error of the Y coordinate system. The Mean Absolute Error (MAE) of the test set of stroke localization in the X-coordinate and Y-coordinate is 0.69 cm and 0.14 cm respectively, and the MAE of the localization position is 0.74 cm. Figure 6 gives a random example of the target, where Figure 6a presents a circle intersection. Similarly, Figure 6b compares the prediction and ground truth, where the blue area represents the predicted stroke location and the orange area represents the true stroke location.

X _location MAE: 0.69
Y_location MAE: 0.14
Position MAE: 0.74
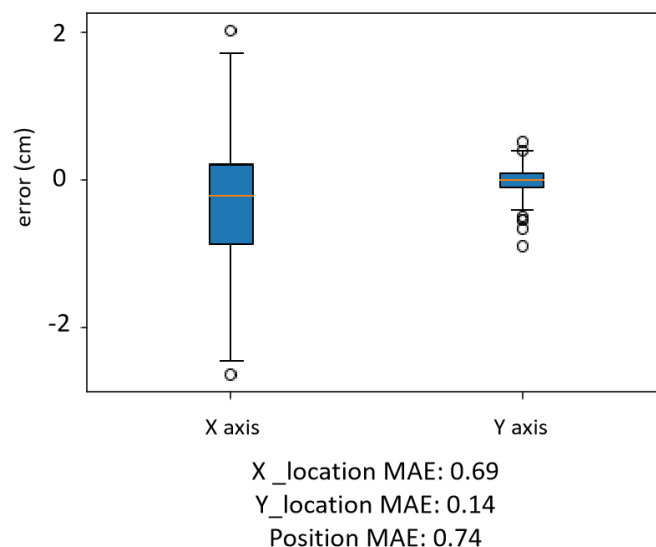
Figure 5. Localization errors of test set predicted by RRMS framework.

(**a**)   (**b**)
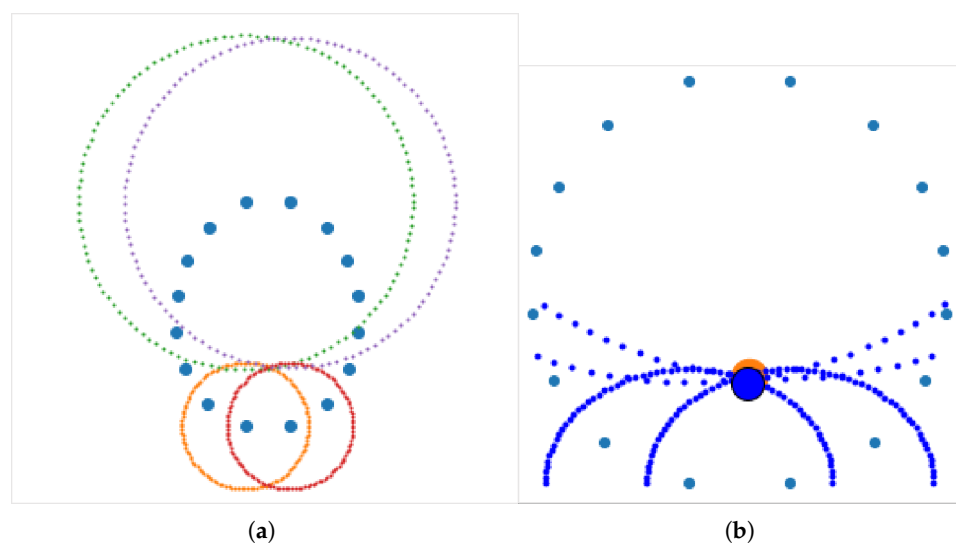
Figure 6. (**a**) Intersection result of circular trajectories based on the predicted distance d_0, d_7, d_8, d_{15}, and (**b**) comparison of ground truth (blue area) and predicted position (orange area).

5. Discussion

This work has presented a preliminary experimental evaluation of stroke localization based on the RRMS framework. The system utilizes S-parameters collected within a wide frequency range of 0.5 GHz-2 GHz, and the learning process of stroke location is independent of stroke subtype classification results and stroke phantom size. Specifically, stroke prediction results can be fed back in a short time without prior information. In addition, the microwave facility itself has the advantages of low hardware cost and good portability, which indicate that it is suitable for deployment in mobile clinics, or as a decision-support tool for rapid intervention and treatment in underdeveloped areas.

A comparison of similar studies localizing brain stroke based on EM signals is presented in Table 3. Wu et al. [11] convert it to classification by using the interconnection point of two lines that pass the stroke. Similarly, Zhu et al. [12] use geometric methods to localize ICH stroke which causes an asymmetric response to electromagnetic signals, based on the assumption that the healthy brain is approximately symmetric. In contrast, the proposed method can also be used to localize IS stroke while achieving higher prediction accuracy.

Table 3. Comparison with existing works.

Authors	Data Source Collection	Proposed Method	Observations	Results
Wu et al. [11]	Simulated data	SVM+PSO	280 negative signals 168 positive signals	82.7% accuracy on s11 78.3% accuracy on s12
Zhu et al. [12]	Phantom measurement	SCL	150	0.85 ± 0.08 Dice similarity on ICH
Tobon Vasquez et al. [15]	Phantom measurement	TSVD-Based Linear Imaging	3 cases, 12 reconstruction images for each case	Error ranges from 0.5 mm to 38 mm on ICH
Proposed work	Phantom measurement	RRMS	608	X axis error: 0.69 cm Y axis error: 0.14 cm

Tobon Vasquez et al. [15] use the captured scatter signal through a linear imaging algorithm to reconstruct stroke position and size, which is a conventional microwave technique to detect lesions (e.g., stroke, breast cancer). In fact, there is no reliable reconstruction method that can generate high-quality images so far. This is because the electromagnetic inverse scattering problem is ill-conditioned, nonlinear, and multimodal, i.e., it can have multiple local minima during iterative optimization. In addition to this, the localization of stroke should be obtained after image segmentation or dependent on the visual detection of experts. Conversely, the proposed method can directly obtain stroke position in the

Cartesian coordinate system, bypassing image reconstruction and other medical procedures that may be required.

It is worth noting that the results reported in this experiment are based on the four highest-scoring predictors of port-0, port-7, port-8, and port-1. This part of future work will further improve localization performance by including all 16 predictors. In addition, the antenna array used in the experiment was placed at the same height as the stroke target during data collection. We are currently exploring a 3D realistic experimental process, that is, introducing Z coordinates to observe whether the change of the scattering data affects this algorithm by changing the height of the stroke phantoms.

Author Contributions: Conceptualization, S.G. and G.Z.; methodology, S.G.; software, S.G.; validation, S.G., G.Z. and A.B.; formal analysis, S.G.; investigation, S.G. and G.Z.; resources, S.G.; data curation, G.Z.; writing–original draft preparation, S.G.; writing–review and editing, S.G., G.Z., A.B. and X.Z.; visualization, S.G.; supervision, G.Z., A.B. and X.Z.; project administration, G.Z.; funding acquisition, G.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is not publicly available according to the industry partner's restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Australian Bureau of Statistics. *Disability, Ageing and Carers*; Summary of Findings: Canberra, Australia, 2018.
2. Semenov, S.Y.; Corfield, D.R. Microwave tomography for brain imaging: Feasibility assessment for stroke detection. *Int. J. Antennas Propag.* **2008**, *2008*, 254830. [CrossRef]
3. Jalilvand, M.; Li, X.; Zwick, T.; Wiesbeck, W.; Pancera, E. Hemorrhagic stroke detection via UWB medical imaging. In Proceedings of the Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP), IEEE, Rome, Italy, 11–15 April 2011; pp. 2911–2914.
4. Lu, M.; Xiao, X.; Liu, G.; Lu, H. Microwave breast tumor localization using wavelet feature extraction and genetic algorithm-neural network. *Med. Phys.* **2021**, *48*, 6080–6093. [CrossRef] [PubMed]
5. Shokry, M.; Allam, A. UWB antenna for brain stroke and brain tumour detection. In Proceedings of the 2016 21st International Conference on Microwave, Radar and Wireless Communications (MIKON), IEEE, Krakow, Poland, 9–11 May 2016; pp. 1–3.
6. Jalilvand, M.; Zwick, T.; Wiesbeck, W.; Pancera, E. UWB synthetic aperture-based radar system for hemorrhagic head-stroke detection. In Proceedings of the 2011 IEEE RadarCon (RADAR), IEEE, Kansas City, MO, USA, 23–27 May 2011; pp. 956–959.
7. Zhu, G.; Bialkowski, A.; Guo, L.; Mohammed, B.; Abbosh, A. Stroke classification in simulated electromagnetic imaging using graph approaches. *IEEE J. Electromagn. RF Microwaves Med. Biol.* **2020**, *5*, 46–53. [CrossRef]
8. Karadima, O.; Rahman, M.; Sotiriou, I.; Ghavami, N.; Lu, P.; Ahsan, S.; Kosmas, P. Experimental validation of microwave tomography with the DBIM-TwIST algorithm for brain stroke detection and classification. *Sensors* **2020**, *20*, 840. [CrossRef] [PubMed]
9. Alon, L.; Dehkharghani, S. A stroke detection and discrimination framework using broadband microwave scattering on stochastic models with deep learning. *Sci. Rep.* **2021**, *11*, 1–9. [CrossRef] [PubMed]
10. Salucci, M.; Polo, A.; Vrba, J. Multi-step learning-by-examples strategy for real-time brain stroke microwave scattering data inversion. *Electronics* **2021**, *10*, 95. [CrossRef]
11. Wu, Y.; Zhu, M.; Li, D.; Zhang, Y.; Wang, Y. Brain stroke localization by using microwave-based signal classification. In Proceedings of the 2016 International Conference on Electromagnetics in Advanced Applications (ICEAA), Cairns, Australia, 19–23 September 2016; pp. 828–831.
12. Zhu, G.; Bialkowski, A.; Crozier, S.; Guo, L.; Nguyen, P.; Stancombe, A.; Abbosh, A. Brain Injury Localization in Electromagnetic Imaging using Symmetric Crossing Lines Method. *bioRxiv* **2022**. [CrossRef]
13. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67.

[CrossRef]

14.　Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]

15.　Tobon Vasquez, J.A.; Scapaticci, R.; Turvani, G.; Bellizzi, G.; Joachimowicz, N.; Duchêne, B.; Tedeschi, E.; Casu, M.R.; Crocco, L.; Vipiana, F. Design and experimental assessment of a 2D microwave imaging system for brain stroke monitoring. *Int. J. Antennas Propag.* **2019**, *2019*, 8065036. [CrossRef]