

# TOWARDS UNCERTAINTY QUANTIFICATION FOR ELECTRODE BENDING PREDICTION IN STEREOTACTIC NEUROSURGERY

Alejandro Granados <sup>\*1</sup>, Oeslle Lucena <sup>\*1</sup>, Vejay Vakharia <sup>2,3</sup>, Anna Miserocchi <sup>2,3</sup>,  
Andrew W. McEvoy <sup>2,3</sup>, Sjoerd B. Vos <sup>2,3</sup>, Roman Rodionov <sup>2,3</sup>, John S. Duncan <sup>2,3</sup>,  
Rachel Sparks <sup>1</sup>, Sébastien Ourselin <sup>1</sup>

<sup>1</sup> School of Biomedical Engineering and Imaging Sciences, King’s College London, UK

<sup>2</sup> National Hospital of Neurology and Neurosurgery, London, UK

<sup>3</sup> Dept of Clin and Experim Epilepsy, Queen Square, UCL Inst of Neurol, UK

## ABSTRACT

Implantation accuracy of electrodes during stereotactic neurosurgery is necessary to ensure safety and efficacy. However, electrodes deflect from planned trajectories. Although mechanical models and data-driven approaches have been proposed for trajectory prediction, they lack to report uncertainty of the predictions. We propose to use Monte Carlo (MC) dropout on neural networks to quantify uncertainty of predicted electrode local displacement. We compute image features of 23 stereoelectroencephalography cases (241 electrodes) and use them as inputs to a neural network to regress electrode local displacement. We use MC dropout with 200 stochastic passes to quantify uncertainty of predictions. To validate our approach, we define a baseline model without dropout and compare it to a stochastic model using 10-fold cross-validation. Given a starting planned trajectory, we predicted electrode bending using inferred local displacement at the tip via simulation. We found MC dropout performed better than a non-stochastic baseline model and provided confidence intervals along the predicted trajectory of electrodes. We believe this approach facilitates better decision making for electrode bending prediction in surgical planning.

*Index Terms*— stereotactic neurosurgery, epilepsy, trajectory prediction, neural network, uncertainty quantification

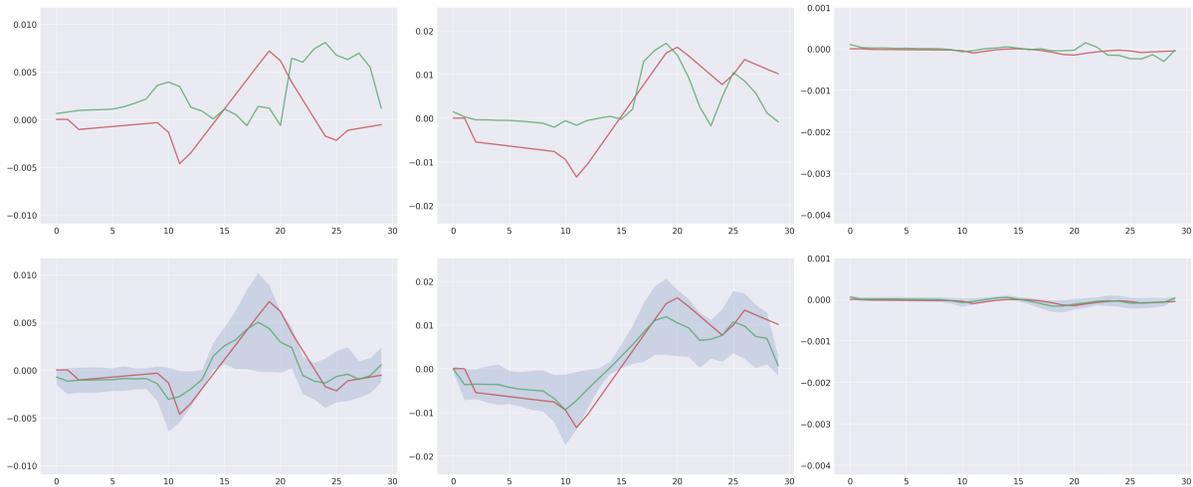
## 1. INTRODUCTION

Stereotactic neurosurgery, such as stereoelectroencephalography (SEEG) and Deep Brain Stimulation (DBS), relies on optimal surgical planning to target specific regions of the brain. During SEEG, for instance, between 8 and 14 electrodes are typically implanted on patients to identify the epileptogenic zone. Accurate placement is crucial for safety, diagnosis and treatment efficacy [1, 2, 3]. Neurosurgeons define planned trajectories by specifying entry and target points on pre-operative magnetic resonance (MR) T1-weighted (T1) images. Then, implantation accuracy is assessed on co-registered post-implantation computed tomography (CT) images [4]. However, electrodes may deflect from planned trajectories during insertion due to electrode design and tissue properties [5].

Previous approaches have proposed electrode trajectory prediction via mechanical modelling of tissue-needle interaction, whereby soft tissue and electrodes are discretised into finite elements that are numerically integrated [6, 7, 8]. More recently, prediction of electrode trajectories has been framed as a data-driven problem whereby

machine learning algorithms including random forests, neural networks and long short-term memory gates were used to regress 3 degree-of-freedom electrode bending computed as the difference between two material frames of electrodes modelled as elastic rods [5]. However, these methods do not provide a quantification of uncertainty for electrode bending prediction. Model simplifications and hand-crafted features may not be fully sufficient to take into account all the unknowns necessary to have an optimal model. Therefore, we believe that providing a metric of how reliable our output predictions are is essential for decision making.

Uncertainty can be divided into two types: epistemic (model parameters variability) and aleatoric (noise inherent in the data) [9]. One way to quantify uncertainty is through Bayesian inference [9]. However, conventional Bayesian neural networks (NNs) are time consuming. Bayesian approximation using dropout layers has been proposed to overcome training limitations by doing Monte Carlo sampling at inference stage and applied to medical imaging tasks [10, 11, 12].



**Figure 1.** Baseline (top) and MC dropout (bottom) 2D plots of mean (ground truth in red, prediction in green) and 95% confidence interval (light blue) of predicted local displacements:  $lux$  (left),  $luy$  (centre),  $luz$  (right).  $x$ -axis is mm from target point and  $y$ -axis is local displacement of tip in mm. Plots correspond to the electrodes shown in Fig. 2 top

In this paper, we present a quantification of epistemic uncertainty for the prediction of electrode bending. Compared to [5], we regress 3 degree-of-freedom local displacements instead of local bending since they are more clinically relevant. We focused on investigating epistemic uncertainty using the Monte Carlo (MC) dropout [13]. To the best of our knowledge, we are the first to investigate NN uncertainty in the context of stereotactic electrode trajectory prediction.

## 2. MATERIAL AND METHODS

### 2.1 Dataset

Pre-operative T1 and diffusion-weighted MR images, as well as post-operative CT images of 23 refractory epilepsy SEEG cases with a total of 241 electrodes (8761 interpolation points) were acquired at the National Hospital of Neurology and Neurosurgery (London, UK) and co-registered. The position of the electrode contacts was computed via automated electrode segmentation [4]. We manually identified anterior/posterior commensurate and interhemispheric point to transform electrodes into AC-PC coordinate space [14]. We

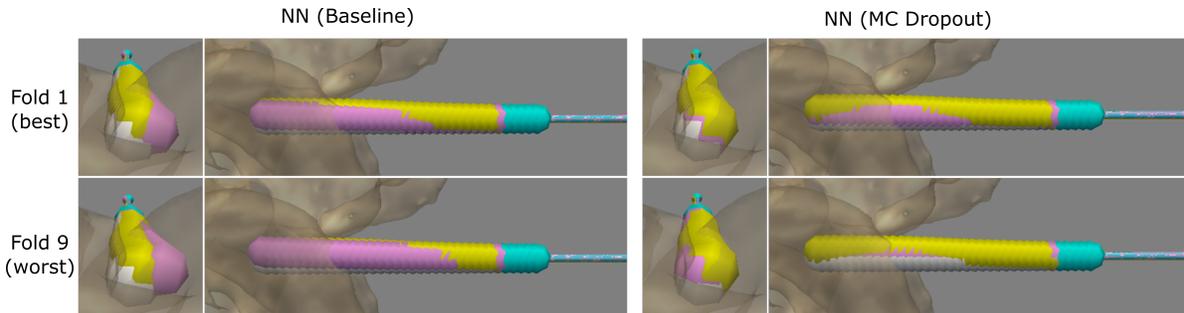
interpolate electrodes at 1 mm intervals and modelled them as elastic rods with orthogonal material frames between points. In this work, we confirm that the y-component of material frames points upwards, x-component points left and z-component points towards the entry point along the trajectory.

## 2.2 Feature Extraction

Similar to [5], we compute features related to: electrode trajectory (direction of tip, direction of bolt, curvature, use of stylet, local bending and displacement from rigid trajectory), structure (MR intensity, distances and angles with respect to scalp, cortex, white matter, and deep grey matter, intra-cranial length), parcellation (regions along trajectory, number of regions traversed, length of segments) and tractography (fibre direction, apparent fibre density and angles with respect to trajectory). In contrast to previous work [5], we normalise continuous variables and replace categorical variables with one-hot encoding vectors including stylet (yes, no, none), types of region (cortex, white, deep), entry points by type of lobe (frontal, central, temporal, parietal, occipital), and target points by type of region (frontal, central, temporal, parietal, occipital, insula, cingulum). For each point along the electrode trajectory, displacement from a rigid trajectory is computed. Then, scalar projection onto the local material frames components is computed, i.e. 3 degrees-of-freedom local displacement (referred to as  $lux$ ,  $luy$ ,  $luz$ ).

## 2.3 Uncertainty Quantification

MC dropout is a method to compute epistemic uncertainty and it is seen as a Bayesian approximation to a Gaussian Process [13]. Basically, for a given dataset  $X$  and annotations  $Y$  in a NN with weights  $W$ , MC dropout consists of finding a tractable parametrised distribution  $q^*(W)$  which minimises the Kullback-Leibler divergence to the true model posterior  $p(W|X, Y)$  [9]. In practice, without any further change in the model during the training, dropout is used at the inference stage to approximate a probability distribution for the model weights. That is done by computing  $T$  random forward passes. The final prediction is computed by averaging the  $T$  outputs from the model and the standard deviation is used to compute the 95% confidence interval. In this work, we define a total of  $T=200$  stochastic passes.



**Figure 2.** Results of electrode bending prediction (views from the tip and the side) for the best (top) and worst (bottom) folds of *Baseline* (left) and *MC Dropout* (right) neural network (NN) models. Example of an electrode (test dataset) with entry point on left middle temporal gyrus, targeting left hippocampus. Ground truth trajectory (yellow), rigid trajectory in the direction of the bolt (white), initial trajectory used for inference (blue) and predicted trajectory (pink). *NN (baseline)* RMSE=0.31 mm and *NN (MC dropout)* RMSE=0.22 mm of 30 interpolated points.

### 3. EXPERIMENTAL DESIGN AND VALIDATION

#### 3.1 Model Validation

In this work, we investigate uncertainty quantification from the model outputs rather than proposing a new NN architecture. We built as a baseline model a NN composed of: a parametric rectified linear unit (PReLU) activation [15], a hidden layer with a fully-connected layer followed by PReLU, and fully-connected layer to regress the local bending. PReLU activation is used as it adaptively learns the parameters of the rectifiers. From the baseline model we build a stochastic model by adding dropout layers, similar to [13], after the PReLU activations. Also, we empirically chose a dropout rate of 0.1 and one hidden layer with 1024 neurons through a random hyper parameter search on the training set. For all results reported, we performed 10-fold cross-validation, where for each fold we split our data into 80% of the cases for training and 10% of the cases for validation for NN training stage, and 10% of the cases as a hold-out test set for the inference stage.

The network was trained with an Adam optimizer to minimise the  $L_2$  loss function of the form  $loss(y, \hat{y}) = \frac{\|y - \hat{y}\|_2^2}{2}$  where  $y$  is the ground truth electrode local displacement and  $\hat{y}$  is the output inferred from the model. In the training stage, all the NNs were trained for 200 epochs, based on experimentally chosen convergence, with a weight decay of  $1 \times 10^{-6}$  and initial learning rate of  $1 \times 10^{-3}$ , which was reduced by half for every 50 epochs.

#### 3.2 Electrode Trajectory Validation

Trajectories were computed from local displacement predictions in which we computed an initial (planned) trajectory of 5 points (5 mm) from a rigid trajectory in the direction of the bolt. Our main loop consists of a) extending the tip of the electrode by one point (1 mm), b) model the electrode as an elastic rod, c) extract features from medical images, and d) predict and apply local displacement with respect to the material frame at the tip. Using cases in the test set, we generated trajectories for each electrode per each cross-validation fold. RMSE was computed between interpolated points of predicted and ground truth trajectories.

### 4. RESULTS AND DISCUSSION

We observe that *NN (MC dropout)* improved both model and trajectory performance compared to the baseline (Table 1). Through this model we quantified model uncertainty and obtained confidence intervals along the predicted trajectory of electrodes. Optimal fitting is found for the *lux* and *luy* displacements (Figure 1) for *NN (MC dropout)* while *NN (baseline)* fits poorly the data. Although *luz* displacement does not suffer significant changes (Figure 1, third column), we took it into account in our regression NN to provide stochasticity in the 3 degrees of freedom. During trajectory validation, errors are accumulated per each interpolation point during prediction as electrodes are inserted. Therefore, its RMSE is higher when compared to model validation errors (Table 1).

**Table 1.** 10-fold validation for the model and trajectory predictions. Mean (standard deviation) in mm.

Method	RMSE	
	Model ( $10^{-3}$ )	Trajectory
<i>NN (baseline)</i>	6.53 (0.17)	1.09 (2.46)
<i>NN (MC dropout)</i>	4.21 (0.11)	0.58 (0.72)

To demonstrate the performance *NN (MC dropout)* has on predicted trajectories, Figure 2 shows inferred electrodes (in pink) closer to the ground truth (in yellow) compared to *NN (baseline)*. In this example, we picked the folds that performed the best and the worst and selected an electrode that follows the direction of deflection optimally. However, we also noticed that it is not always the case and this might be caused by the features used or the amount of training data. Hence, further investigation needs to be carried out.

We only focus on epistemic uncertainty that can be explained by having large training datasets [9] whilst ignoring aleatoric uncertainty which takes into account noise in the observations. In the context of electrode trajectory prediction, more data is necessary to cover all possible planned trajectories and modelling the noise of inputs such as using heterocedastic uncertainty [9]. Hand-crafting features has an effect on model accuracy, i.e. they may not be enough to explain the electrode bending or they may be adding noise to NN regression. Possible solutions for this problem include an investigation of optimal input features and the use of end-to-end learning approaches.

## 5. CONCLUSIONS

We present a method to quantify uncertainty of electrode bending prediction for SEEG stereotactic neurosurgery using MC dropout. This approach improves electrode prediction accuracy compared to a non-stochastic baseline method. Moreover, uncertainty of local bending prediction provides an estimate of confidence interval that can be used in decision making during surgical planning by taking into account possible deviation of trajectories electrodes are likely to have in order to increase the likelihood of reaching the target whilst avoiding critical structures. A limitation of this approach is that we only considered uncertainty of the model limited to the amount of data. However, aleatoric uncertainty needs to be taken into account because the implantation of electrode could be better modelled as a stochastic process. This would improve predictions of displacements that are not commonly seen in the training data. Future work will investigate the feasibility of using uncertainty quantification to guide electrode insertion as part of surgical planning of stereotactic neurosurgery.

## Acknowledgement

This research was funded/supported by the Health Innovation Challenge Fund (WT106882), the Wellcome/EPSRC Centre for Medical Engineering [WT203148/Z/16/Z], EPSRC Research Council (EPSRC DTP EP/R513064/1), and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London and/or the NIHR Clinical Research Facility. We are grateful to the Wolfson Foundation and the Epilepsy Society for supporting the Epilepsy Society MR scanner. The views expressed in this publication are those of the authors and not necessarily those of the Wellcome Trust or NIHR. **Ethical Approval** All data were evaluated retrospectively. All studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. **Conflict of Interest** The authors declare that they have no conflict of interest. **Informed Consent** For this type of study formal consent is not required.

## References

- [1] N. Abolhassani, R. Patel, and M. Moallem, “Needle insertion into soft tissue: A survey,” *Medical engineering & physics*, vol. 29, no. 4, pp. 413–431, 2007.
- [2] T. De Jong, “Needle deflection in tissue,” Master’s thesis, Delft University of Technology, the Netherlands, 2015.
- [3] V. N. Vakharia, R. Sparks, A. G. O’Keeffe, R. Rodionov, A. Miserocchi, A. McEvoy, S. Ourselin, and J. Duncan, “Accuracy of intracranial electrode placement for stereoelectroencephalography: A systematic review and meta-analysis,” *Epilepsia*, vol. 58, no. 6, pp. 921–932, 2017.
- [4] A. Granados, V. Vakharia, R. Rodionov, M. Schweiger, S. B. Vos, A. G. O’Keeffe, K. Li, C. Wu, A. Miserocchi, A. W. McEvoy *et al.*, “Automatic segmentation of stereoelectroencephalography (SEEG) electrodes post-implantation considering bending,” *International journal of computer assisted radiology and surgery*, vol. 13, no. 6, pp. 935–946, 2018.
- [5] A. Granados, M. Mancini, S. B. Vos, O. Lucena, V. Vakharia, R. Rodionov, A. Miserocchi, A. W. McEvoy, J. S. Duncan, R. Sparks *et al.*, “A Machine Learning Approach to Predict Instrument Bending in Stereotactic Neurosurgery,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. plus 0.5em minus 0.4em Springer, 2018, pp. 238–246.
- [6] C. Duriez, C. Guébert, M. Marchal, S. Cotin, and L. Grisoni, “Interactive simulation of flexible needle insertions based on constraint models,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. plus 0.5em minus 0.4em Springer, 2009, pp. 291–299.
- [7] N. Hamzé, I. Peterlík, S. Cotin, and C. Essert, “Preoperative trajectory planning for percutaneous procedures in deformable environments,” *Computerized Medical Imaging and Graphics*, vol. 47, pp. 16–28, 2016.
- [8] R. J. Roesthuis, Y. R. Van Veen, A. Jahya, and S. Misra, “Mechanics of needle-tissue interaction,” in *2011 IEEE/RSJ international conference on intelligent robots and systems*. plus 0.5em minus 0.4em IEEE, 2011, pp. 2557–2563.
- [9] A. Kendall and Y. Gal, “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?” in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., 2017, pp. 5574–5584.
- [10] R. Tanno, D. E. Worrall, A. Ghosh, E. Kaden, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, “Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. plus 0.5em minus 0.4em Springer, 2017, pp. 611–619.
- [11] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, “Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation,” *Medical image analysis*, vol. 59, p. 101557, 2020.
- [12] Z. Eaton-Rosen, F. Bragman, S. Bisdas, S. Ourselin, and M. J. Cardoso, “Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network predictions,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. plus 0.5em minus 0.4em Springer, 2018, pp. 691–699.
- [13] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning,” in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 48. plus 0.5em minus 0.4em New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059.
- [14] S.-H. Choi, J.-G. Chi, Y.-B. Kim, and Z.-H. Cho, “Anterior commissure-posterior commissure revisited,” *Korean journal of radiology*, vol. 14, no. 4, pp. 653–661, 2013.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.