Manuscript Title: **A Machine Learning Enabled Wireless Intracranial Brain Deformation Sensing System**

Dear Editor,

We thank the reviewers for their critical evaluations of our manuscript. We have addressed the comments in detail as described below.

We hope that our responses fulfill the criteria for publication in IEEE TBME.

Sincerely,

Albert Kim, Ph.D.

Assistant Professor, Electrical and Computer Engineering

Temple University

**Reviewer: 1**

1. If some basic and existing machine learning approaches (not newly developed) were used, the authors are suggested to mention in the title, e.g. “A basic machine learning approach enabled……”

**Answer:** We thank the reviewer for the comment on the title of the manuscript. The main objective of this paper is to explore three different machine learning algorithms to apply to a novel brain deformation sensing principle. Thus, we still believe the title is appropriate.

1. Some figures are too small to read, the authors can use separate figures to make the size larger;

**Answer:** We greatly regret about the small figures and fonts. To make all figure more readable, we increased the font size in Fig. 1-3 and 7-9 (formerly 8-10).

1. Since the machine learning approaches are available, detailed instruction is not needed;

**Answer:** We appreciate for the suggestion. We have removed detailed instructions related to the machine learning algorithms. Instead, we provided rationales and configuration information related to our brain deformation sensing system.

* Pg. 2, left column, line 34-37: *“As a proof of concept, we used three machine learning algorithms, including random forests (RF), k-nearest neighbors (KNN), and a multi-layer perceptron-based neural network (MLP-NN).”*
* Pg. 4, left column, line 35 – Pg. 5, right column, line 9: “*We implemented RF, KNN, and MLP-NN as standard baseline approaches in this experiment.*

*RF is a popular machine learning algorithm due to its simplicity and versatility [36]* *while being one of the most powerful statistical learning approaches [37].* *RF algorithm has demonstrated robustness for regression problems with larger datasets [36].* *As our calibration data also has strong regression characteristics (especially in the xy plane), we employed RF as the first ML algorithm. We experimentally set the training parameters for the RF algorithm (maximum depth = 28, number of estimators = 116) by cross-validation process in which we minimized the prediction error by adjusting the parameter values.*

*KNN, on the other hand, is one of the least complicated non-parametric methods for pattern classification. This is another baseline approach that was selected because KNN can generally achieve the best performance possible and yield extremely low error rates.”*

* Pg. 4, right column, line 6-11: *“The number of neighbors, k, was determined carefully to minimize misclassification error by monitoring the performance indicators (e.g., RMS error and R2 value). In our case, the value of k = 9 minimized the error rate during the cross-validation process and our model was able to capture adequate context for the block of features used for training the KNN.”*
* Pg. 4, right column, line 13-25: *“MLP-NN [44] was also used to train the brain deformation sensing system. MLP-NN mimics the basic learning framework of a biological neural network. NNs are inherently non-linear and tend to perform better when dealing with non-linearities [45], [46] (i.e. viscoelastic behavior of brain deformation). It consists of artificial neurons (nodes) designated as the input layer, single or multiple hidden layers, and the output layer. In this study, our MLP NN network consisted of five layers including an input layer, three hidden layers, and an output layer. The input layer comprises 3 input nodes (B1, B2, B3), the hidden layers are sequentially cascaded with 1024, 512 and 256 nodes respectively. The output layer has 3 nodes which produce the predicted coordinate (x, y, z).”*

1. The authors are suggested to stress the novelty of their work and compare other methods regarding Intracranial Brain Deformation Sensing;

**Answer:** We thank the reviewer for the critical suggestion. We initially mentioned the novelty is the improvement from our previous works: (pg. 2, left column, line27-34) *“Notably, we address the two major shortcomings of our previous work by using (1) a magnetic tunnel junction (MTJ) sensor that offers higher sensitivity and longer operation range (improved sensing volume by three orders of magnitude), and (2) a machine learning model that predicts brain deformation with better accuracy than previously reported measurements using a standard Gaussian model [27].”* However, we agree with the reviewer and revised the manuscript to highlight the novelty. We also compared our work with other similar technologies and summarize them in Table IV.

* Pg. 1, right column, line13 – Pg. 2, left column, line 25: *“Although research into TBI has spanned several decades, the manner in which mechanical impact (e.g., blast wave exposure) injures the brain remains highly disputed. This is attributed to poor understanding in the relationship between initial mechanical injury and downstream biological and behavioral alteration. Cadaveric/surrogate human and rodent head models and computer simulations have been useful in characterizing mechanical injury profiles to date [14]–[17]. Several computational approaches using wearable accelerometer/gyroscope data to find a relationship between external stimuli and the resulting TBI [18]. However, they suffer from lack of experimental validation.*

*In parallel, direct measurements of intracranial brain biomechanics were also attempted. For example, M. Chavko et al. [20] measured the intracranial pressure during blast-TBI events using an implantable fiber optic sensor in the rat brain. More recently, wireless technologies were utilized to measure intracranial brain deformation (as well as general physiological sensing) [20]–[25]. Notably, we previously reported a wireless brain deformation sensing system that is capable of measuring brain deformation under blast waves [27]. We employed an implantable soft magnet and paired a three-dimensional array of giant magnetoresistance sensor (GMR). The soft magnet was implanted on the dura of live and dead rat brain and the mimicked blast waves were exposed to the head. The relative magnetic strength gradient due to the position change could be translated into spatiotemporal information by an empirical model (a sum of three-variable double Gaussian equations).*

*In this paper, we present a machine learning (ML)-enabled wireless sensing system that is capable of measuring intracranial brain deformation in real-time.”*

* Pg. 2, left column, line 41-49: *“The proposed sensing system is low cost compare to MRI/CT scan, smaller in size and is a modular device. With a validated rodent model of TBI combined with the power of ML modeling and cutting-edge sensor technology, we anticipate the ML-enabled brain deformation sensing system may provide a new experimental tool and model to gain greater insight into the link between the mechanics of TBI, structural damage in the brain, and future subsequent biological mediators of TBI pathology.”*
* Pg. 7, right column, line 9-16: *“Table V summarizes this work compare to other similar technologies. Overall, machine learning-enabled brain deformation sensing system improved our previous work by order of magnitude for the sensing range and the accuracy. This work also comparable to the simulation or computational approaches. We have to mention that the presented sensing system is the only experimental approach that could model the brain deformations based on the real-time measurement.”*

TABLE IV: A COMPARISON WITH OTHER METHODS FOR MEASURING BRAIN DEFORMATION

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Authors | Method | Parameter | Type | Resolution | Error | Accuracy |
| This work | Soft Magnet + MTJ + ML | Deformation | Experimental | 5µs/1µm | ~ 140 µm | 92-97% |
| Song et al., 2015 [25] | Soft Magnet + GMR | Deformation | Experimental | 5µs/10µm | - | 85% |
| M. Chavko et al., | Fiber optic | Pressure | Experimental | 100 kHz/4kPa | - | - |
| Alshareef et al., 2020 [54] | Sonomicrometry | Deformation | Simulation | ±0.024 mm | - | 94-98% |
| Tonutti et al., 2017 [55] | FEM + ML (ANN/SVR) | Deformation | Simulation | - | 0.18±0.13mm | ±3 mm |
| Knutsen et al., 2014 [56] | Tagged MRI | Deformation | Computational | 18.4 ms/160×24 voxels | 0.141 (NRMSE) | - |
| Schiavone et al., 2009 [58] | Light Aspiration Device + FEA | Elasticity | Computational | - | ±0.01 mm | - |

1. Some related works are missing, authors are suggested to comment on them in the introduction part: (1) “Overview of Recent Development on Wireless Sensing Circuits and Systems for Healthcare and Biomedical Applications,” IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Volume: 8, Issue: 2, 2018.

(2) “Wandering Pattern Sensing at S-Band,” IEEE Journal of Biomedical and Health Informatics, Vol. 22, Iss. 6, pp. 1863-1870. November, 2018.

(3) “Implantable Wireless Intracranial Pressure Monitoring Based on Air Pressure Sensing,” IEEE Transactions on Biomedical Circuits and Systems, Volume: 12, Issue: 5, 2018;

(4) “High-Sensitivity Wireless Displacement Sensing Enabled by PT-Symmetric Telemetry,” IEEE Transactions on Antennas and Propagation, Volume: 67, Issue: 5, 2019;

(5) “5G Based User-centric Sensing at C Band,” IEEE Transactions on Industrial Informatics, Vol. 15, No. 5, pp. 3040-3047, May, 2019.

**Answer:** We thank the reviewer for pointing the important literatures. We cited these works in the introduction.

* Pg. 2, left column, line 3-5: *“More recently, wireless technologies were utilized to measure intracranial brain deformation (as well as general physiological sensing) [20]–[25].”*

1. Please add some quantitative results in the conclusion part;

**Answer:** Authors thank the reviewer for the suggestion. We included the quantitative results in the conclusion.

* Pg. 8, left column, line 25-33: *“The sensing system is capable of measuring brain deformation in real-time over a large range of sensing volume (up to 12 × 12 × 4 mm3). The ML models (e.g., NN) generate the three-dimensional live animal brain deformation with the high accuracy: as low as 97.9 µm absolute error, 0.074 normalized RMSE, 0.933 Pearson’s R, 208.58 µm Fréchet distance, or 0.13 Procrustes disparity. This accuracy would contribute to finding intracranial brain deformation more closely to the actual value.”*

1. Other machine learning approaches can be considered in the future work.

**Answer:** We agree with the reviewer. There are more advanced machine learning approaches that can be considered. As future works, we will investigate convolutional neural network (CNN) and Long short-term memory (LSTM). CNNs are designed to learn spatial/temporal correlations of the data. Since, it offers less complexity than MLPs, they are easier and more efficient to train on correlated or context dependent data. Similarly, LSTMs are recurrent networks which are designed to learn temporal behavior of the sequence. The data used in this study is a sequential. When these temporal segments are used by sliding window approach, LSTMs are expected to perform better. Considering these, we plan to use them in the future to further increase the accuracy of the outputs and robustness of the system. We have added the followings in the discussion.

* Pg. 7, right column line 30 – Pg. 8, left column, line 5: *“**The ML model can also be further improved with more advanced algorithms, such as convolutional neural network (CNN) or Long short-term memory (LSTM)). CNNs are designed to learn spatial/temporal correlation of the data and LSTMs learn the temporal behavior of the sequence. Since the brain deformation is a time-series data in three dimension (spatiotemporal), these algorithms may enhance the accuracy and robustness.”*

**Reviewer: 2**

1. Your only conclusive validation data is from the camera. if you look at these results (Table III) there is no significant difference between these three methods. so how we can say that which one is better?

**Answer:** We appreciate the reviewer for the critical comment. We added one more comparison of the camera data and Gaussian model in needle insertion experiment (Table III). The new results indicate that all MLs performed better than the Gaussian model *in vitro*. We also then separated *in vitro* (Table III) and *in vivo* results (Table IV) to minimize the confusion. In a new Table III, we compared the ML model with the Gaussian model against the camera data and Table IV summarized the comparison of the ML and Gaussian models each other. In our new Table III, it is unambiguously clear that ML models outperform the Gaussian model. We extrapolate this finding to the interpretation of position data based on magnetic strength for in vivo results as well. In fact, the two are virtually the same; we measure the magnetic strength to infer the position of the soft magnet. However, the issue is that there is no real-time brain deformation measurement technique for *in situ* and *in vivo* animal experiments. Thus, we chose the Gaussian model as the closest alternative for the ground truth; although we expect that ML models to be superior to the Gaussian model. We have added the following explanation in the manuscript.

Additionally, as per feedback from the other respective reviewers, we concluded that R2 and RMSE values may not sufficient to validate the 3D trajectories. Thus, we added Fréchet distance and Procrustes disparity to validate the ML models as well as how well the predicted trajectory fit to the empirically derived trajectories (Updated Table III). The following is added to address our rationale for additional performance indicators. We added the followings to highlight our points.

* Pg. 4, left column, line 26-29*: “For the validation, we considered the Gaussian model as the closet alternative for the ground truth since there is no real-time brain deformation measurement technique for in situ and in vivo.*
* Pg. 5, left column, line 25-42: “*In order to find similarities between the predicted deformation trajectory compared to the reference trajectory (i.e., camera measurements for in vitro and empirical model for in vivo), we used two similarity metrics, which are Fréchet distance and Procrustes disparity. Fréchet distance considers the location and the sequence of the points along the curve during distance calculation. This metric is often considered superior to Hausdorff distance, which is another popular distance metric due to its sensitivity to the outliers in the datasets [47], [48]. The Procrustes disparity is a part of the Procrustes analysis in which a shape is optimally superimposed on a reference shape by translating, rotating, and uniformly scaling the shape object [49]. Procrustes analysis performs the optimal transformation to minimize the disparity between two shapes. By calculating the Procrustes disparity, we can observe how well the predicted dataset fits the reference dataset. The minimum value (zero) indicates the datasets are identical in shape.”*
* Pg. 6, left column, line 22-30: *“the Fréchet distance was calculated 371.22 µm, 339.93 µm, and 288.96 µm whereas, the Procrustes disparity score was 0.0325, 0.0184 and 0.0138 for RF, KNN, and MLP-NN, respectively. The Fréchet distance and the Procrustes disparity score indicate that the predicted trajectories were well fitted with the reference trajectory during the needle insertion test where the trajectory predicted by the MLP-NN was closer to the reference trajectory with minimum Fréchet distance.”*
* Pg. 6, right column, line 5-7: *“Table III also unambiguously indicates that ML models outperformed the Gaussian model. Overall agreements between the ML models and ground truth camera measurements showed low error, higher correlation, and less disparity.”*
* Pg. 6, right column, line 21-23: *“We extrapolated the in vitro results (Table III) to the interpretation of position data based on magnetic strength for in vivo experiment.”*
* Pg. 6, right column, line 42-47: “*The Fréchet distance was calculated 94.28 µm, 85.71 µm and 57.26 µm for dead rat and 288.78 µm, 257.12 µm and 208.58 µm for live rat while the Procrustes disparity scores were 0.054, 0.0402 and 0.0395 for dead rat and 0.1926, 0.154, 0.1304 for live rat using RF, KNN, and MLP-NN, respectively.”*
* Pg. 7, left column, line 18 – Pg. 7, right column, line 8*: “While both the Gaussian and ML models are virtually the same as both interpret the magnetic strength to infer the position of the implanted soft magnet, there is no evidence to show which approach is more accurate due to the lack of real-time brain deformation measurement technique for in situ and in vivo (imaging cannot sufficiently capture deformation with high temporal resolution in real-time). Thus, we chose the Gaussian model as the closest alternative for the ground truth, although we expect that the ML models to be superior to the Gaussian model.*

*Combining the in vitro and in vivo results, our experimental results support the hypothesis that the ML models represented the brain deformation more faithfully than the Gaussian model The ML model could adopt each data point and its neighboring data relations to create an accurate brain deformation trajectory.* *Additionally, NN is an ideal algorithm for estimating the dynamic and nonlinear functions [45], [46]; thus, it is expected to model the change in the physiological properties of the brain tissue. Simple linear regression algorithms, such as RF and KNN, could be more suitable for linear dependencies. However, RF and KNN are the common baseline ML algorithm, allowed us to demonstrate the ML usability with the sensing system.”*

TABLE III.   
COMPARISON OF IN VITRO NEEDLE INSERTION

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | ML Algorithm | Mean Abs. Err (µm) | NRMSE | Pearson Corr., R | Fréchet Distance  (µm) | Procrustes  Disparity |
| PVC gel (Camera  vs. Gaussian) | - | 330.1 | 0.061 | 0.886 | 510.6 | 0.215 |
| PVC gel (Camera  vs. ML) | RF | 137.4 | 0.069 | 0.984 | 371.22 | 0.0325 |
| KNN | 128.9 | 0.069 | 0.991 | 339.93 | 0.0184 |
| NN | 90.0 | 0.046 | 0.993 | 288.96 | 0.0138 |

1. Using correlation to show the result maybe is not a good choice. because correlation heavily depends on range. I like to see more strong statistical tests (such as t-test) to see the difference. As you can see from your results, all the methods are very close.

**Answer:** We thank the reviewer for bringing this to our attention. As suggested, we have performed t-test to validate the similarities among Gaussian model and ML models. We observed that the results are not statistically different (high p-value) for all comparison groups (Fig. S1). We added the following statistical analysis in the manuscript.

* Pg. 7, left column, line 13-17: *“**We have also performed t-test for each predicted result with references (i.e., camera data from in vitro needle insertion and Gaussian model from in vivo experiments). We observed high p-values (> 0.1), which indicate the predicted data was not statistically different.”*

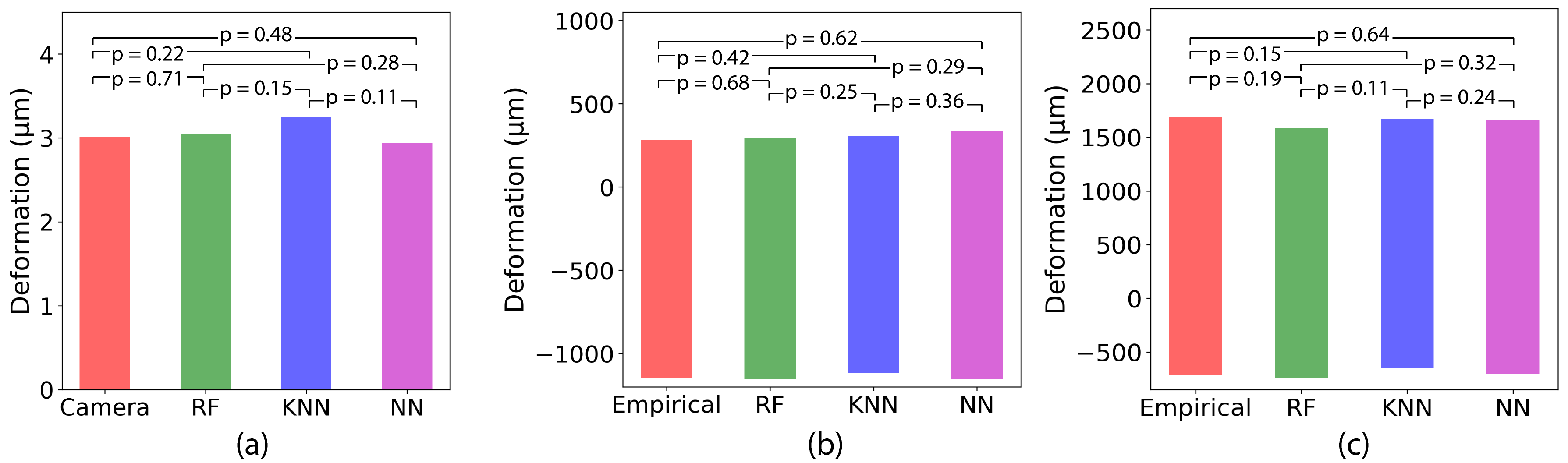


Fig. S1. (a) t-test for needle insertion test results, (b) t-test for brain (dead rat) deformation test results, (c) t-test for brain (live rat) deformation test results

1. Also, TBI is heavily depend on blood pressure and you can improve your device by putting some feature to use that.  you can cite some paper to enrich your introduction.

[1] Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice   
[2] An iPhone application for blood pressure monitoring via the oscillometric finger pressing method   
[3] Formulas to Explain Popular Oscillometric Blood Pressure Estimation Algorithms

**Answer:** We appreciate for the insight. We agree that blood pressure is important parameters in TBI, and it is expected to improve our model. In addition, we also recognized other parameters, such as cerebrospinal fluid, and intracranial pressure. As suggested, we have added the followings to enrich the introduction and discussion (as the future work).

* Pg. 1, right column, line 5-8: *“**Depending on the location and blood pressure, the acceleration and direction of deformation may also vary [8-10]”*
* Pg. 7, right column, line 27-30: *“The future research is also planned. The TBI also heavily depends on blood pressure [9], [10], cerebrospinal fluid (CSF), or intracranial pressure (ICP), therefore, the ML algorithms will be trained on these additional parameters.”*

**Reviewer: 3**

1. N=2 animals, one dead and one alive. I would recommend a larger dataset considering the power and danger of ML based approaches.

**Answer:** We regret for missing an important information of the sample size. We repeated *in vivo* experiments four times for each dead and live animals (n = 4). The MLs were trained with all data (no exclusion). The machine learning algorithms can also be over trained for the specified subjects/samples as we have seen from the calibration dataset (Table 1); close loop training accuracy was higher than open loop test accuracy. Thus, we believe that dataset was sufficient for this work. We added the correct information in the manuscript.

* Pg. 4, left column, line 13-14: *“Total eight animals (n = 4 for each group) were exposed to a positive peak overpressure blast wave...”*

1. In determining number of k-nearest neighbors; a more thorough explanation of k=9 is needed.

**Answer:** We thank to the reviewer for this. The k = 9 was the optimal estimated value during the training process. Nine nearest neighbors for the KNN captures enough context for the window/block of features. This information is updated in the edited manuscript. Please refer to Question #3 for the manuscript revision.

1. R-squared is not a typical metric used to assess the quality of KNN or decision trees; explain further this comparative choice; recommended ROC curve.

**Answer:** In this experiment, ROC curve was not possible to generate for KNN or RF models since they do not generate posteriors or log-likelihood measures for the classes. It is still possible to generate ROC based on number decision trees or neighbors; however, comparison between the models will not be appropriate. Instead, we added two more performance indicators and updated the results. Please refer to Question #8 for the detail.

1. The machine learning algorithms need larger train/test set to claim robustness.

**Answer:** We agree with the reviewer. In this paper, we have implemented a deterministic model to learn and predict the dynamic behavior of the brain deformation. The system could be over trained using the calibration data that had 72,000 datasets. Each set contains data from three sensors along with three dimensional cartesian coordinates. Our models showed the reliable results when we validated with 20% of the calibration data (closed loop test), *in vitro,* and *in vivo* experiments. For the clarification, we removed our claim over the robustness.

1. Inserting magnetic sensors into brain tissue should confound the outcome of TBI based pathogenesis.

**Answer:** The reviewer is right that implanting a magnetic sensor in the brain confound the TBI outcomes. However, our focus in this paper is to explore the machine learning approach to understand the initial TBI mechanical injury, and it is not to understand the biological and behavioral alteration. We anticipate this work may provide a new experimental tool or model for future TBI research (please refer our answer for Question #4).

However, we considered the possible complication due to the implantable soft magnet. The soft magnet was to minimize the mechanical effect to the brain. It was engineered to move with the brain during the deformation and not to injure the tissue; thus, the Young’s modulus of the soft magnet was set similar to that of the brain (1–40 kPa depending on rate and magnitude of the impact) [27-28]. We also fabricated it using biocompatible materials, which are iron oxide nano particles and silicone elastomer (Ecoflex). The *in vivo* experiments were also conducted to minimize the complication. The animal receives a soft magnet just before the experiment. However, we also agree with the reviewer that we cannot be conclusive regarding the long-term effect. To acknowledge such limitation, we added the following in the manuscript.

* Pg. 2, right column, line 25-30: *“The soft magnet was to minimize the mechanical effect to the brain. It was engineered to move with the brain during the deformation and not to injure the tissue; thus, the Young’s modulus of the soft magnet was set similar to that of the brain (1–40 kPa depending on rate and magnitude of the impact) [27], [28]”*
* Pg. 7, right column, line 22-26: *“Lastly, the presented sensing system may not appropriate for the long-term study. Although our soft magnet was fabricated using biocompatible materials, its long-term effects to the biological system may confound the outcome of TBI-based pathogenesis.”*

**Reviewer: 4**

1. Authors have previous work in the same area. I recommend adding a related work section to present the previous work and its limitation (provide more details) and a comparison with others works. It is interesting also to see if already there are other implementations use machine learning or dictionary learning approach to predict the brain deformation.

**Answer:** We thank the reviewer for the suggestion. We have revised our introduction to discuss our previous work more in detail and some previous works. We also have added a comparison table with related work (Table IV). Please refer our answer for Question#4 for detail answer.

1. The magnet sensor outputs are filtered using an average filter to reduce the noise and are applied to the machine learning algorithms. I suggest integrating a feature extraction bloc to extract statistical features and offload the work of the classifiers. The use of feature extraction bloc can enhance the coefficient of determination scores.

**Answer:** We thank the reviewer for the critical comments. It is true that the signal representation from feature extraction block could provide more specific information for the problem at hand. But finding the best features is a separate engineering problem. It is expected that the underlying features are inherently learned by the models themselves. However, in the machine learning communities, such as speech and image recognition, the raw signal features are used for the classification. Additionally, adding a feature extraction block from the sensors would reduce the amount training/evaluation samples significantly since the feature extraction would capture features on a window of samples and not for individual samples.

1. I suggest also to remove the filter stage and see if it affects the prediction scores or no.

**Answer:** We thank the reviewer for the suggestion. As suggested, we validated the necessity of the filter stage. We have removed the filter before training the model. The data shows that the accuracy of prediction results was deteriorated without a filter. The NRMSE values were increased by 29.54 % average without the filter. Similarly, the R2 values also under performed by 4.03 % without the filter. The absence of the filter greatest affected KNN in terms of both NRMSE (64.29% increase for close loop test) and R2 (5.25% decrease for close loop test) performance.

**Results without using the filter:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ML algorithm** | **Training parameters** | **Close loop** | | | | **Open loop** | | | |
| **Training time (sec)** | **Prediction time (sec)** | **NRMSE** | **R2 (%)** | **Training time (sec)** | **Prediction time (sec)** | **NRSME** | **R2 (%)** |
| **RF** | Depth = 28  Estimator = 116 | 18.001 | 0.301 | 0.036 | 92.468 | 14.105 | 0.302 | 0.097 | 87.546 |
| **KNN** | k = 9 | 0.051 | 0.115 | 0.069 | 89.721 | 0.039 | 0.095 | 0.113 | 86.894 |
| **NN** | 3 hidden layers  Epoch = 3000 | 421.16 | 0.434 | 0.078 | 88.367 | 325.41 | 0.358 | 0.119 | 87.538 |

**Results with the filter (Table 1, Pg. 5):**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ML algorithm** | **Training parameters** | **Close loop** | | | | **Open loop** | | | |
| **Training time (sec)** | **Prediction time (sec)** | **NRMSE** | **R2 (%)** | **Training time (sec)** | **Prediction time (sec)** | **NRSME** | **R2 (%)** |
| **RF** | Depth = 28  Estimator = 116 | 18.467 | 0.306 | 0.024 | 96.91 | 14.873 | 0.304 | 0.088 | 91.3 |
| **KNN** | k = 9 | 0.050 | 0.107 | 0.042 | 94.69 | 0.037 | 0.097 | 0.097 | 90.4 |
| **NN** | 3 hidden layers  Epoch = 3000 | 412.78 | 0.425 | 0.059 | 92.41 | 313.69 | 0.391 | 0.107 | 89.3 |

1. I suggest considering some recent and relevant studies in the comparisons

**Answer:** We thank the reviewer for the suggestion. We have added recent and related work in the introduction and Table IV (pg. 7). Please see Question #4 and #8 for the detail.

**Reviewer: 5**

This paper discusses a sensing method for brain deformation through a wireless magnetic sensor. The sensor has been described in reference [17]. One of the contributions of this manuscript is in the use of three magnetic tunnel junction (MTJ) sensors instead of giant magnetoresistive sensors (as in ref. [17]) to detect the magnetic field of a soft magnet. Using MTJ sensors, the authors increase the sensing range of the brain deformation sensor by a factor of 8 compared to their previous work (from 0.5 mm to 4 mm). Another contribution is in the use of machine learning (ML) algorithms to predict the position of the soft sensor, which is the main content of this manuscript.

1. However, there are some issues with the papers, namely the prediction accuracy of the trained machine learning models and in the clarity of Section III D. In Vivo Validation. In the paper, the authors compared the performance of the trained models and the Gaussian model with several measures: absolute error, normalized RMSE (NRME), R and R2. However, the author states that the machine learning models are superior to the Gaussian model only based on R2 value (ignoring the NRMSE and average absolute error values). Whereas, the NRMSE of the gaussian model is the lowest. Besides, the average absolute error of the Gaussian model also better than the Neural Network Model.

**Answer:** We thank the reviewer for this critical comment. Table II summarizes measured calibration data and reconstructed by Gaussian model and ML algorithms. However, the performance indicators, such as the mean of absolute error or normalized RMSE are greatly affected by a couple of outlier data, which were prominent towards the margin of the sensing volume. In the meantime, the experimental results (i.e., *in vitro* and *in vivo)* were a series of data that progress with certain patterns; thus, the ML models could outperform. We added one more comparison of the camera data and Gaussian model in needle insertion experiment (Table III). The new results indicate that all MLs performed better than the Gaussian model *in vitro*. In the manuscript, we have acknowledged the Gaussian model’s higher score in the calibration results and updated the new data (camera vs. Gaussian) in Table III to highlight our points. Please also see our answer to Question #8.

* Pg. 6, left column, line 2-5: *“However, we observed a couple of outliers in ML data towards the margin of the sensing volume, which increased the mean absolute errors and NRMSE. Thus, we further compared the performances using in vitro needle insertion”*
* Pg. 6, right column, line 5-8: *“Table III also unambiguously indicates that ML models outperformed the Gaussian model. Overall agreements between the ML models and ground truth camera measurements showed low error, higher correlation, and less disparity.”*

TABLE III.   
COMPARISON OF IN VITRO NEEDLE INSERTION

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | ML Algorithm | Mean Abs. Err (µm) | NRMSE | Pearson Corr., R | Fréchet Distance  (µm) | Procrustes  Disparity |
| PVC gel (Camera  vs. Gaussian) | - | 330.1 | 0.061 | 0.886 | 510.6 | 0.215 |
| PVC gel (Camera  vs. ML) | RF | 137.4 | 0.069 | 0.984 | 371.22 | 0.0325 |
| KNN | 128.9 | 0.069 | 0.991 | 339.93 | 0.0184 |
| NN | 90.0 | 0.046 | 0.993 | 288.96 | 0.0138 |

1. In Section III.D, the authors used previously measured in vivo brain deformation data from dead and live rats. However, there is not enough clarity about the source of the data and how they used these data. Are these data from reference [17]? If these data are from reference [17], did the authors directly fetch the magnetic data to the trained machine learning models? If this is the case, the authors could explain more about how the ML models can be transferred directly between the new MTJ based sensor to the old GMR based sensor.

**Answer:** We thank the reviewer for this critical comment. The *in vivo* data was from our previous work and we directly used the magnetic strength measured with the GMR sensors. Since our ML model is trained to predict the position of the soft magnet based on the magnetic strength rather than the sensor output voltage, the same ML model can be used for both GMR and MTJ, independent of the sensor type and variations in sensor parameters.

By introducing MTJ sensors, we have a better resolution and lower detection limit compared to GMR and thus extending our sensing range. Thus, the measured magnetic strength should not vary with the sensor type. To illustrate this, we compared the GMR and MTJ sensors outputs that measure the same soft magnet (Fig. S2). As expected, both sensors read identical magnetic strength (R2 = 0.99) except the MTJ sensor could detect the magnetic field at further distance (1 mm vs. 4 mm in z-direction). We have added the following in the manuscript to further clarify this point.

* Pg 4, left column, line 19-29: “*Since both GMR and MTJ sensors measured the same magnetic strength from the soft magnet (R2 = 0.99), except the MTJ sensor has higher sensitivity and extended sensing range (1 mm vs. 4 mm in z-direction), the MTJ data-trained ML models could transfer the GMR data without any discrepancy. However, we should acknowledge that the possible source of error could be due to the soft magnet variability during fabrication.”*

A close up of a map

Description automatically generated

Fig. S2: GMR vs. MTJ sensors: both sensors read the same magnetic strength from the soft magnet

1. Also, in Table III, the accuracy of the measurements are compared between the Gaussian and ML prediction. It becomes confusing since the Gaussian model is used as the ground truth. Then the authors stated that “the ML models represented the brain deformation more faithfully than the Gaussian model”. If the authors could clarify this section (e.g. adding more discussion about how to interpret the data, the reasons why the Gaussian model is used as the ground truth, the source of the differences, etc.), it will greatly improve the quality of the paper.

**Answer:** Again, we thank the reviewer for suggestion that could improve the quality of the paper. To minimize the confusion, we separated *in vitro* (Table III) and *in vivo* results (Table IV). In Table III, we compared the ML model with the Gaussian model against the camera data. Table IV summarizes the comparison of the ML and Gaussian models each other.

In our new Table III, it is unambiguously clear that ML models outperform the Gaussian model. We extrapolate this finding to the interpretation of position data based on magnetic strength for in vivo results as well. In fact, the two are virtually the same; we measure the magnetic strength to infer the position of the soft magnet.

However, the issue is that there is no real-time brain deformation measurement technique for *in situ* and *in vivo* animal experiments. Thus, we chose the Gaussian model as the closest alternative for the ground truth; although we expect that ML models to be superior to the Gaussian model. We have added the following explanation in the manuscript.

* Pg. 4, left column, line 26-29*: “For the validation, we considered the Gaussian model as the closet alternative for the ground truth since there is no real-time brain deformation measurement technique for in situ and in vivo.*
* Pg. 6, right column, line 5-7: *“Table III also unambiguously indicates that ML models outperformed the Gaussian model. Overall agreements between the ML models and ground truth camera measurements showed low error, higher correlation, and less disparity.”*
* Pg. 6, right column, line 21-23: *“We extrapolated the in vitro results (Table III) to the interpretation of position data based on magnetic strength for in vivo experiment.”*
* Pg. 7, left column, line 18-30*: “While both the Gaussian and ML models are virtually the same as both interpret the magnetic strength to infer the position of the implanted soft magnet, there is no evidence to show which approach is more accurate due to the lack of real-time brain deformation measurement technique for in situ and in vivo (imaging cannot sufficiently capture deformation with high temporal resolution in real-time). Thus, we chose the Gaussian model as the closest alternative for the ground truth, although we expect that the ML models to be superior to the Gaussian model.*

*Combining the in vitro and in vivo results, our experimental results support the hypothesis that the ML models represented the brain deformation more faithfully than the Gaussian model.”*

1. Section II.C Calibration (Training Data). Does the presence of ferromagnetic materials affect the measurement of MTJ? If it affects the measurement, is there any particular method to avoid it?

**Answer:** Thank you for raising an important concern. The presence of ferromagnet materials may affect the training dataset due to altered magnetic strength of the soft magnet. Thus, the calibration was performed in a ferromagnetic material free environment. Furthermore, we performed an additional experiment to examine the safe distance to the ferromagnetic material (i.e., magnet). We observed the closest source of magnetic field was 400 mm away and no significant trace of magnetic field was detected using the MTJ sensors. To provide this additional information, we included the following.

* Pg. 3, right column, line 39-42: *“During the calibration (in fact, all experiments), the presence of other ferromagnet materials may affect the sensor reading; thus, it was done in a magnetically isolated environment.”*

1. What cost function is used to train the ML models?

**Answer:** In this experiment, during training, we have used Mean Squared Error (MSE) as a cost function to find the optimum parameters for the algorithms (such as estimator for RF, k for KNN, epoch for NN etc.).

1. In Section III.E.3, why the input of the Neural Network has six nodes (B1, B2, B3, x, y, z)? Is the position reconstruction not only based on the magnetic field measurements? What do x, y, z in the input represent?

**Answer:** We thank the reviewer for bring this to our attention. It was our mistake. The input of NN has three input nodes: B1, B2, and B3. We have corrected the statement.

* Pg. 4, right column, line 21-24: *“The input layer comprises 3 input nodes (B1, B2, B3), the hidden layers are sequentially cascaded with 1024, 512 and 256 nodes respectively.”*

1. In Section II.F, can the authors explain more about the closed-loop test? Does it mean the authors use all of the calibration data to train the network?

**Answer:** The reviewer interpreted it correctly. All of the calibration data were used to train the ML models during the closed-loop test and 20% randomly selected data were used to test the models. We clarify this information as following.

* Pg. 4, right column, line 40-42: *“We also conducted closed-loop tests using all calibration data to train the ML algorithm, then use randomly selected 20% of the calibration data was used for validation.”*

1. Adding the reconstruction results from the Gaussian model in Fig. 8 could help clarify the comparison between ML models and the Gaussian model.

**Answer:** We appreciate for the suggestion. We have modified Fig. 7 (previously Fig. 8) where we added two more plots: (a) the measured calibration and (b) the Gaussian model. The ML models were compared against both references (Table II - IV).

1. Section III.D. The authors stated “… the brain has a nonlinear viscoelastic deformation characteristic, which is difficult to model using simple linear regressions models such as RF and KNN”. However, RF and KNN model are used to predict the position of the soft magnet. The position measurements should be able to be performed regardless of how the brain deforms. Could the authors explain more about the statement?

**Answer:** We agree with the reviewer that the position measurement/prediction will be performed regardless how the brain deforms. However, in this statement, we described that our results showed NN produced a better result in the animal test since the brain tissue deformed more dynamic manner, in other words, non-linearly. The general regression models (i.e., RF and KNN) were more suitable for linear dependencies. Additionally, RF and KNN are also considered as the common baseline ML algorithm, allowing us to demonstrate the ML usability with the sensing system. We have added the following explanations.

* Pg. 7, left column, line 31 – Pg. 7, right column, line 8: *“The ML model could adopt each data point and its relations to the neighboring data to create an accuracy brain deformation trajectory. Additionally, NN is ideal algorithm for estimating the dynamic and nonlinear functions [45], [46]; thus, it is expected to model the change in physiological properties of the brain tissue more faithfully. Simple linear regression algorithms, such as RF and KNN, could be more suitable for linear dependencies. However, RF and KNN are the common baseline ML algorithm, allowed us to demonstrate the ML usability with the sensing system.”*

1. Is there any consideration if the soft magnet is not parallel with the sensor during the brain deformation?

**Answer:** We thank the reviewer for the critical question. As the reviewer pointed out, we agree that the orientation of the soft magnet respect to the sensor may affect the readings. To ameliorate this potential error, we chose a small dimension (3 mm in diameter) for the soft magnet not only for minimizing the shape-dependent magnetic strength change but also to be considered as a point magnetic source. We previously examined the change in the soft magnets’ position in our *in vivo* experiments after exposure to the blast wave, and we observed almost no hysteresis in terms of position and orientation. The potential misorientation effect during deformation is an interesting question we can pursue in our future work.

* Pg. 7, right column, line 20-21: *“There was also a potential misorientation effect of the soft magnet during brain deformation, which we will pursue in our future work.”*

1. Table III is hard to understand, mainly due to the title of the plot is measurement accuracy. To measure the accuracy, we need to compare the measurement with a ground truth, which the authors do in the previous section (comparing camera with the magnetic sensor). What is the ground truth for live and dead rats’ experiments?

**Answer:** We thank the reviewer for this critical point. We have revised the title of Table III to “*Comparison of brain deformation measurements*.” Here, we compared the ML-predicted on dead and live rats’ brain deformation data with the Gaussian model. Because there are no real-time brain deformation measurement technique, we considered the Gaussian model as the best alternative. We have added an explanation in the manuscript (Please see our answer in Question #4 and #22 for the detail).

**Reviewer: 6**

1. It is clear that brain tissue deformation due to external forces can lead to TBI. However, it is not clear how the proposed approach can help or prevent TBI. Also, it is not clear if the model developed based on tissue from rat's brain is valid for human brain, which is much bigger and so more vulnerable to shock. More, it is not clear how the implantation of magnets changes brain tissue properties and how this influences the measurements.

**Answer:** We appreciate for the comment. Although research into blast-induced injuries has spanned several decades, the manner in which primary blast wave exposure mechanically injures the brain remains highly disputed. Cadaveric/surrogate human and rodent head models and computer simulations have been useful in characterizing blast mechanical injury profiles to date, but they suffer from lack of experimental validation. Post-TBI biological changes are scantly documented at best in both humans (ethical constraints) and rodents (low study volume) and have yet to be explored with relation to blast mechanical injury profiles. For these reasons, the relationship between initial blast mechanical injury and downstream biological and behavioral alteration remains poorly understood. Thus, in this paper, we have demonstrated a novel method to measure intracranial brain deformation in real-time, providing a 3D trajectory. With a validated rodent model of TBI combined with the power of ML modeling and cutting-edge sensor technology, it is possible to gain greater insight into the link between the mechanics of TBI, structural damage in the brain, and future subsequent biological mediators of TBI pathology. Please refer our answers to Question#4 for the novelty of this work (how the proposed approach can help TBI) and Question#15 for the concern of implanting the soft magnet.

*Detailed comments:*

1. P2.L46 "The output magnetic strength data was then processed using a moving average filter to reduce random noise" Please provide the cutoff frequency of the filter and the sampling frequency.  
   MLP-NN can act as a moving average filter, so 5-tap filter mentioned above is not necessary or even possibly not optimal. Please comment on it.

**Answer:** Ideally, neural networks are able to learn any underlying linear or non-linear functions given that it has enough data and the appropriate number of parameters. Regardless, research communities use various preprocessing steps to clean up the dataset prior to training such networks [S1]. Preconditioning steps could include normalization, filtering, etc. of the signals. Even though, the neural networks are inherently able to learn underlying transformations itself, in practice, the preconditioning always help model perform better. Various studies such as artifact rejection of EEGs, blind source separation in speech show significant improvements in the performance after the preprocessing steps [S2-S3].

The 5-tap moving average filter is an optimal value calculated from experiments. The filter used during the calibrationprocess had cut off frequency 42.9 Hz where sampling rate was 350 samples per second. The parameter was set to eliminate the noise from nearby instruments, powerline noise and any high frequency interference that might appear during the calibration data measurement. During the test phase of this experiment, the sampling rate was 25k samples/second, and the cut off frequency was 3 kHz. We applied this filtering process to all data to maintain similarity between the inputs for each algorithm (RF, k-NN and MLP-NN).

* Pg. 3, left column, line 28 – Pg. 3, right column, line 17: “*During the calibration, we employed a 5 tap low pass filter with cut-off frequency 42.9Hz (sampling rate: 350 samples/sec) to eliminate the noise from nearby instruments, powerline noise, and any high frequency interference. During the in vitro and in vivo test, the cut-off frequency was 3 kHz (sampling rate: 25k samples/sec). After filtering, the data is prepared to train the ML models or to test the ML models. We performed the same pre-processing method for all ML models to keep the consistency of the inputs.”*

[S1] Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Transactions on nuclear science, 44(3), 1464-1468.

[S2] Visser, E., Otsuka, M., & Lee, T. W. (2003). A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments. Speech Communication, 41(2-3), 393-407.

[S3] Wilson, S. B., M. L. Scheuer, R. G. Emerson, and A. J. Gabor. “Seizure Detection: Evaluation of the Reveal Algorithm.” Clin Neurophysiol 115. 10 (Oct 2004): 2280-91

1. P4.L25 "The RMSE value is a measurement of the spread or variance of the predicted values."  
   I would expect it to be root of mean square error of the distance between each predicted value and the corresponding reference value in 3D, as it is stated a few lines above the cited sentence.

**Answer:** We thank the reviewer for the comment, and we apologize for the mistake. We have fixed the cited sentence.

* Pg. 5, left column, line 6-9: *“The RMSE was calculated in terms of root mean square of the absolute errors of the distance between the predicted values and corresponding reference values in three-dimensional space.*

1. P4.L57 "Randomly selected k - 1 datasets were used for training the model, and one remaining dataset was used for *validating the model evaluation*)." If the reference is known, as it is stated in the description of RMSE (P4.L18), cross-validation seems to be not necessary, since we know the true. Please clarify.

**Answer:** We believe that even though all the sequences are known in prior, the cross-validation for k-fold sets would validate the models to learn each section/block/cluster of the experiment, and not just beginning or end of the sequence.

1. I would like to ask for more comments on calibration process. The reason is: if I know the property of any single magnet to be implemented (which can be done) and known the position of the magnet in the brain. I should be able to calculate the strength of magnetic field in any point.  
   Therefore more explanation is necessary why machine learning approach provides much higher accuracy.

**Answer:** We thank the reviewer for the valuable comment. We may be able to determine the position based on the property of the implanted magnet. However, the magnetic property is available as a single number without knowing the relationship to next position. In other words, the magnetic sensor produces a scalar value, and this value can be same for many different positions. In this regard, we applied two techniques: 1) we positioned three magnetic sensors in triangular orientation to obtain the directional information and 2) ML approach to determine the relationship among data. We previously modeled in empirical approach using Gaussian models. However, it was poorly estimated. As the ML is well explained throughout the manuscript, we elaborated the sensors placements in the manuscript to clarify our points.

* Pg. 2, right column, line 33-37: *“Since each sensor output was scalar (resistance value) corresponding to the magnetic field strength and did not represent the direction of the soft magnet movements, we employed three magnetic tunnel junction (MTJ) sensors (Micro Magnetics STJ-240) positioned in a triangulation.”*

**Reviewer: 7**

1. Page 2, second column, line 6: “This configuration resulted in an overall sensing volume of 12 × 12 × 4 mm3 with a sensitivity of 0.12 Ω/μT.”: why this sensor placement is optimal and give you better sensing volume? Please explain the reason.

**Answer:** The sensor placement was determined to cover the maximum volume consists of sagittal, frontal and transverse planes over the head. Triangulated placement of three sensors, therefore, helped to obtain the most possible sensing volume. This orientation also optimized with the soft magnet that has limited magnetic strength. The sensitivity is already determined from the manufacturer. Our answer to Question #35 also provides the rationale for this configuration.

1. Page 2, first column, line 14: “Overall, the ML-enabled model exhibited excellent agreement with the experimentally measured calibration data….”: what does overall mean here? Please rephrase to make it more accurate.

**Answer:** We regret for this. We have removed “overall,” to minimize the confusion.

1. Page 2, second column, line 19: “Three different machine learning algorithms were evaluated in this application: random forests (RF), k-nearest neighbors (KNN), and a multilayer perceptron neural network (MLP-NN)”: why did you choose these three ML algorithms, among many other ML algorithms and why three?

**Answer:** Authors appreciate for the reviewer’s critical question. We have answered this in Question #3.

1. Page 2, first column, line 34: “The soft magnet had a dimension of 3 mm in diameter and 2 mm in height”: What is the soft magnet dimensions important for? E.g., for placement on the surface if cortex? Is 2mm thin enough for this purpose? Please clarify this.

**Answer:** The dimension of soft magnets dictates 1) feasible magnetic strength for MTJ sensor to detect and 2) implantation procedure. The reported soft magnet dimension was optimized based on these two parameters.

1. Page 2, second column, line 37: “We informally optimized the averaging frame length, which serves as a low-pass filter [22], to be 5. ”: what does informally mean here? you should justify using of this length for time window and unit is missing for 5.

**Answer:** We thank the reviewer for this critical comment. This was a pre-processing step for cleaning up the signals and the 5-tap filter is an optimal value calculated from experimentation. Please refer our answer to Question #2 for the revision in the manuscript.

1. Fig 2: “soft magnet displacement is proportional to brain deformation”: Do you consider cases, in which there is a soft magnetic displacement, but there is not any brain deformation, e.g., displacement due to the normal CSF/brain movement?

**Answer:** We thank the reviewer for the critical question. We believe that normal movement due to cerebrospinal fluid may be the slow movement. Thus, additional decision-making algorithm would easily differentiate TBI and non-TBI brain movements. As this work focuses the development of the brain mechanical deformation sensing system, we will need to investigate more rigorous biological parameters as the future work. Please also refer to our answer to Question #10. We stated the following as the future work.

* Pg. 7, right column, line 27 - 30: *“The future research is also planned. The TBI also heavily depends on blood pressure [9], [10], cerebrospinal fluid (CSF), or intracranial pressure (ICP), therefore, the ML algorithms will be trained on these additional parameters.”*

1. Fig 2(d): What is DAQ in d? data acquisition? You should include the full form in the text.

**Answer:** The reviewer is correct. It is data acquisition system. We had mentioned in the text (Pg. 2, right column, line 49: “*“The sensor array was connected to a data acquisition system (National Instruments PCI-6040E) for real-time data processing.”*). We have also made change in Fig. 2.

1. Page 3, first column, line 1: “…and 1 mm step-size in the z-direction by the MTJ sensor array”: why larger step size along z axis?

**Answer:** We thank the reviewer for bring it to our attention. The larger step size in z-axis was due to the experimental limitation. We acknowledge the limitation in the discussion section.

* Pg. 7, right column, line 17-20: *“We acknowledge that our study has some limitations.* *Due to experimental limitations, we could not achieve a finer step size along the z-axis (1 mm). However, ML algorithms could easily interpolate the missing data in the gaps.”*

1. Page 3, second column, line 2:” A needle-like applicator (diameter = 3 mm) was slowly inserted next to the soft magnet at a velocity of 0.3 mm/s to create z-direction deformations.”: Why did you choose this velocity? How do you make sure what you record is not simply the needle insertion artifact, if there is any?

**Answer:** We selected this velocity to create a slow deformation to accurately capture using the video camera. Multiple insertion data during our experiment confirm that the measured deformation was not the needle artifact.

1. Page 3, section D1 and D2: It would be nice if you include an image of the experimental setup.

**Answer:** We thank the reviewer for the suggestion. We have included Fig. 6 to show the experimental setups.



Fig. 6. Experimental setup: (a) *in vitro* needle insertion test in PVC gel, (b) *in vivo* blast tube setup with dead and live rat. [27]

1. Page 3, second column, line 16: “The reference data were interpolated from an empirical model (i.e., a sum of two 3-variable double Gaussian distributions)”: What do you mean by interpolation of the reference data?

**Answer:** We regret for this confusion. We intent to say we used the empirical model (i.e., Gaussian model) as the validating reference. We revised it for the further clarification.

* Pg. 4, left column, line 26-29: *“For the validation, we considered the Gaussian model as the closet alternative for the ground truth since there is no real-time brain deformation measurement technique for in situ and in vivo.”*

1. Page 4, section E3, line 9: why do you consider coordinates as inputs? Is it because of the closed loop test? Then you should mention it.

**Answer:** The authors would like to thank the reviewer for the correction. The text has been replaced with the following:

* Pg. 4, right column, line 21-24: *“The input layer comprises 3 input nodes (B1, B2, B3), the hidden layers are sequentially cascaded with 1024, 512 and 256 nodes respectively. The output layer has 3 nodes which produce the predicted coordinate (x, y, z).”*

1. Page 4, second column, last line: “The ML enabled reconstruction of the calibration map (Fig. 8) addresses one of the two major shortcomings of our previous”: what are these shortcomings? You should name them here.

**Answer:** The shortcomings were the limited sensing range and the poorly estimated calibration map. We have previously mentioned it in the introduction (Pg. 2, left column, line 17-22). For further clarification, we have dictated the shortcomings.

* Pg. 5, right column, line 27-yy: *“The ML enabled reconstruction of the calibration map (Fig. 7) could enhance the sensing range in better esimation, two major shortcomings of our previous work [26].”*

1. Page 5, first column, line 6: “All three ML models resulted in R2 scores over 92% with average absolute values less than 500 μm”: do you mean absolute error?

**Answer:** We regret for such mistakes. We have added “error” in the text (Pg. 5, right column, line 36)

1. An important application of the proposed brain deformation sensing method, in addition to what the authors have already mentioned in the paper, *is detection of origin for cortical spreading depolarization (CSD) waves in traumatic brain injuries (TBIs) and concussion* [Chamanzar, et. al., 2019; Hofmeijer et al., 2018; Hartings, et al., 2014]. CSD is an important possible outcome of TBI, which usually starts to propagate from the site of injury and/or brain deformation, and it can cause permanent secondary brain damages. There have been recent efforts, e.g., EEG silent localization [Chamanzar, et al., 2019] to detect the origin of CSD, and this brain deformation sensing method has the potential to be used to enhance this detection. These applications/examples can be mentioned in the introduction and/or discussion section to emphasize the strength/impact of the proposed method in this paper. Citations are included below:

[1] J. A. Hartings et al., “Spreading depression in continuous electroencephalography of brain trauma,” Ann. Neurol., vol. 76, no. 5, pp. 681–694, 2014.

[2] J. Hofmeijer et al., “Detecting cortical spreading depolarization with full band scalp electroencephalography: An illusion,” Frontiers Neurol., vol. 9, no. 17, pp. 1–9, 2018.

[3] A. Chamanzar et al., "An Algorithm for Automated, Noninvasive Detection of Cortical Spreading Depolarizations Based on EEG Simulations," in IEEE Transactions on Biomedical Engineering, vol. 66, no. 4, pp. 1115-1126, April 2019.

[4] A. Chamanzar and P. Grover, "Silence Localization," 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, 2019, pp. 1155-1158.

**Answer:** We thank the reviewer for such a great insight. We have mentioned it in the discussion as the future work and we are thrilled to work on the suggested application.

* Pg. 8, left column, line 6-13: “*Furthermore, we envision that the proposed method can be utilized in the* *detection of cortical spreading depolarizations (CSD), which is important possible outcome of TBI [51]–[53]. CSD usually starts to propagate from the site of injury (i.e., brain deformation), and it can cause permanent secondary brain damage. The recent effort to detect the origin of CSD is through EEG silent localization [54]; however, our method has potential to be used to enhance this detection.”*

**Reviewer: 8**

1. Since the final objective is to measure TBI, following the results in Fig.10, what would it be expected to see in the results of patients with TBI? Would patients with TBI have results closer to dead animal (in compared to healthy)?

**Answer:** We thank the reviewer for the raising the question. The objective of this paper is to investigate a novel sensing mechanism and the use of machine learning algorithms to predict intracranial brain deformation. Since the human brain is much bigger and more vulnerable to shock, the model we presented may not directly applied to the human subject. However, we anticipate this work may provide a new experimental tool or model for future TBI research. Please refer to our answer to Question #4 that address the novelty of this work.

We also answered the discrepancies between animal (both dead and live) and human in Question #31.

1. The authors present a method that is capable of measuring brain deformation with range of sensing distance up to 4 mm. In the current literature: is 4 mm is reasonable for sensing TBI injuries (is this physiological)?

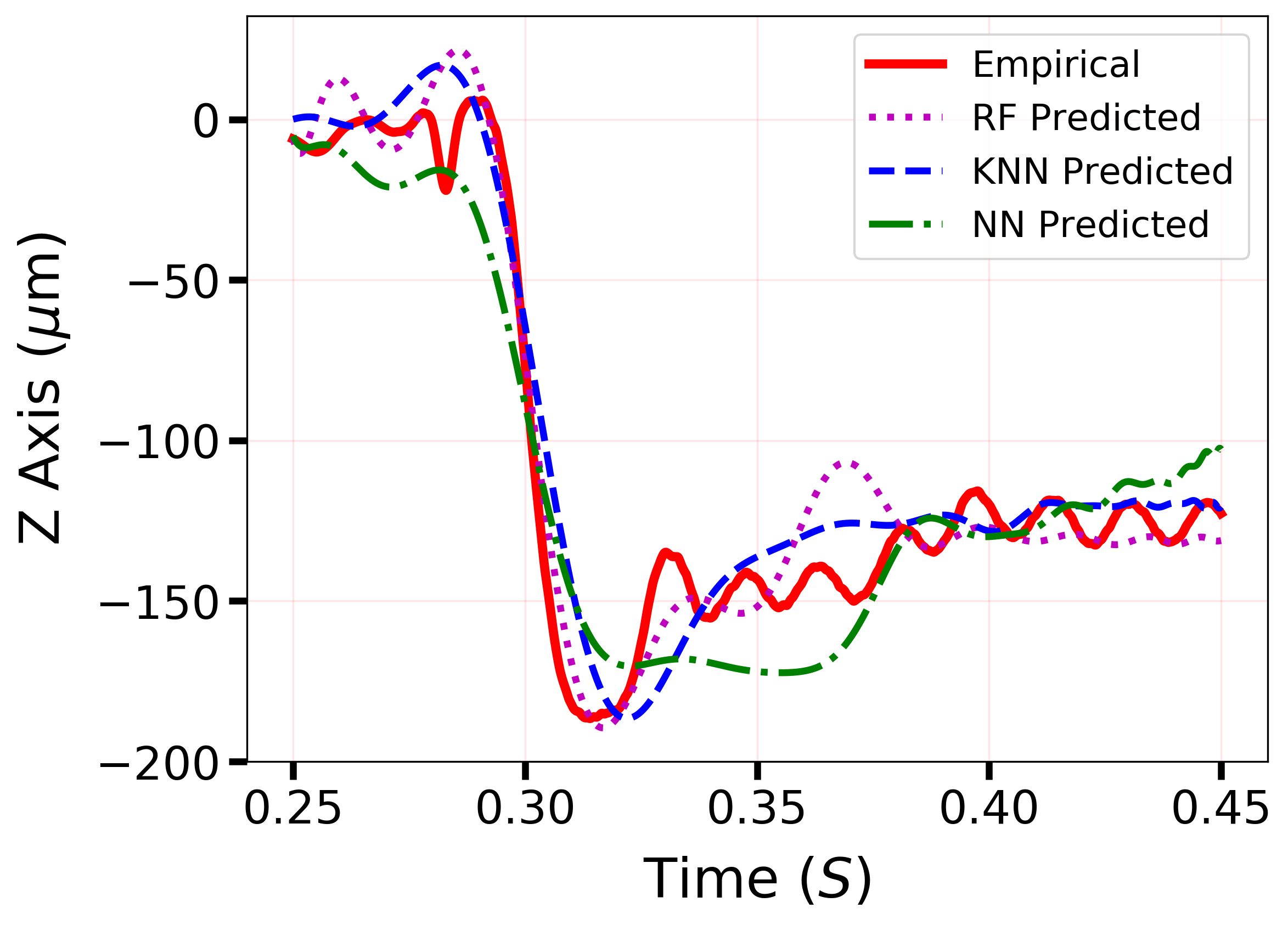
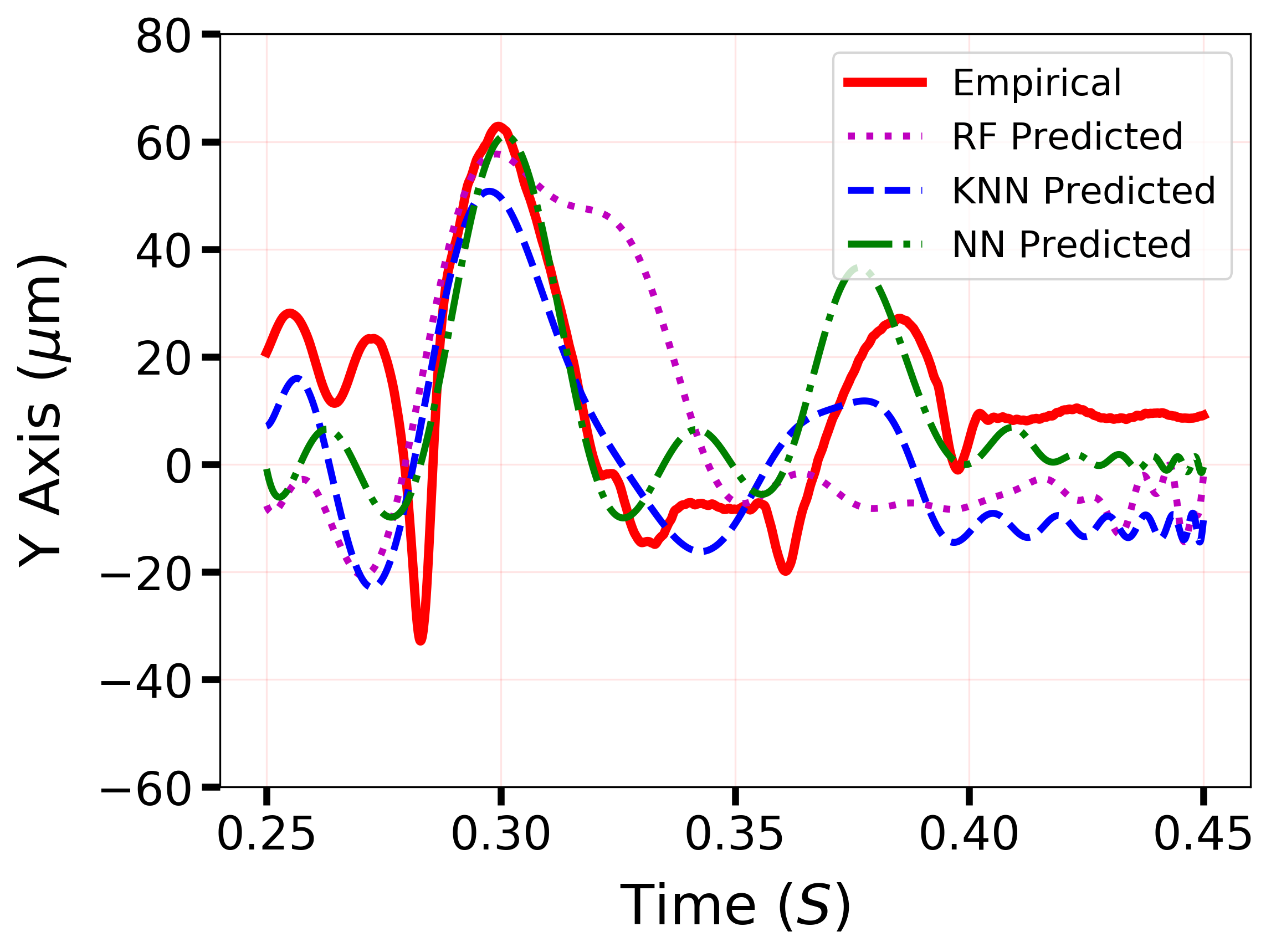
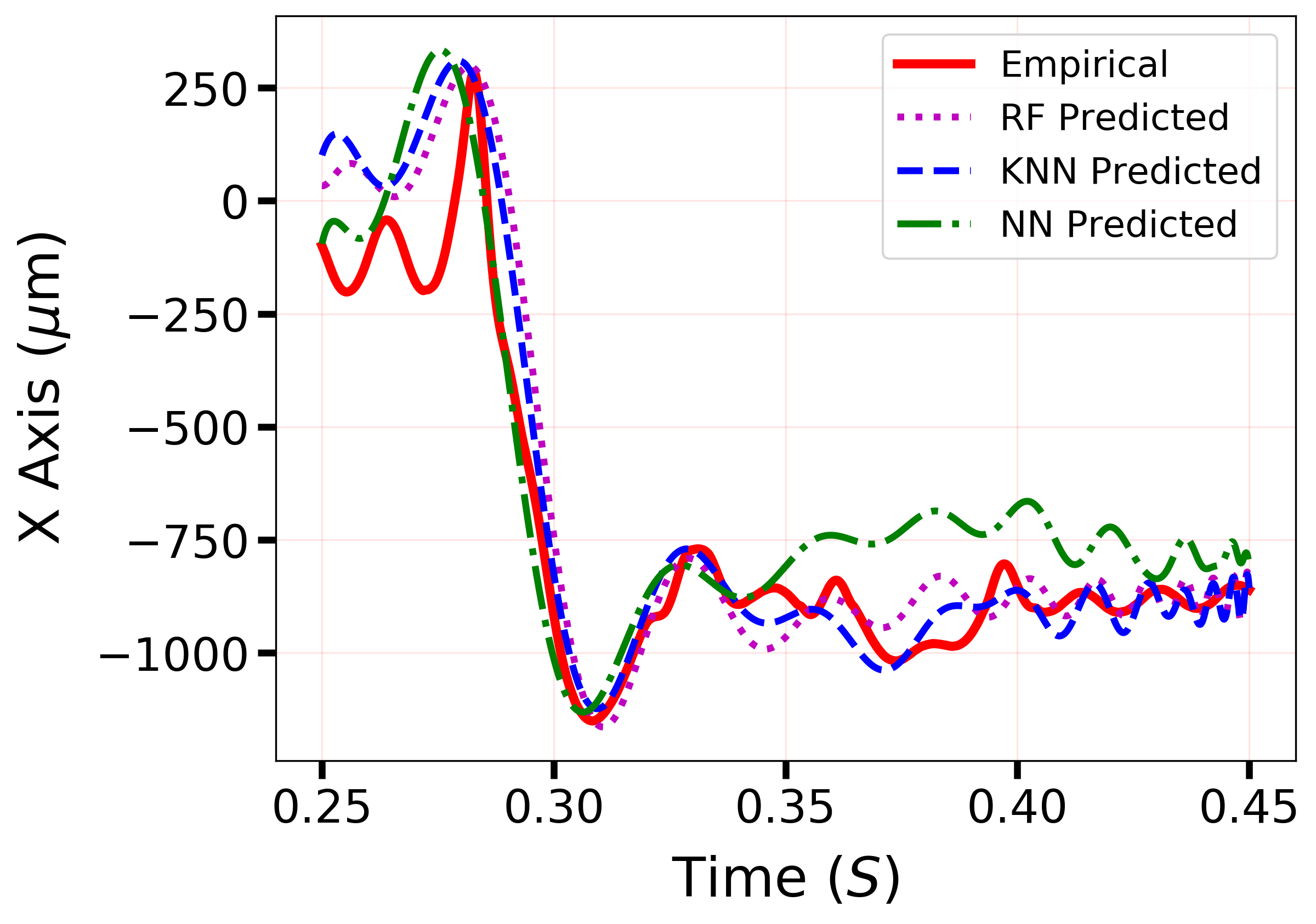
**Answer:** We thank the reviewer for allowing us to think more practical cases. 4 mm sensing distance was determined during the development. The localized brain deformation due to mild acceleration can reach up to 5% of total dimension [S4]. Considering the average length of the human brain is 15 cm, the deformation is expected to be 7.5 mm, which is greater than the limit of our sensing system. However, the objective of this work is to develop a brain deformation sensing system to be used as a new experimental tool. As such, 4 mm sensing range is sufficient for the rodent stud as also demonstrated in the paper.

[S4] Bayly PV, Cohen TS, Leister EP, Ajo D, Leuthardt EC, Genin GM. Deformation of the human brain induced by mild acceleration. J Neurotrauma. 2005;22(8):845–856. doi:10.1089/neu.2005.22.845

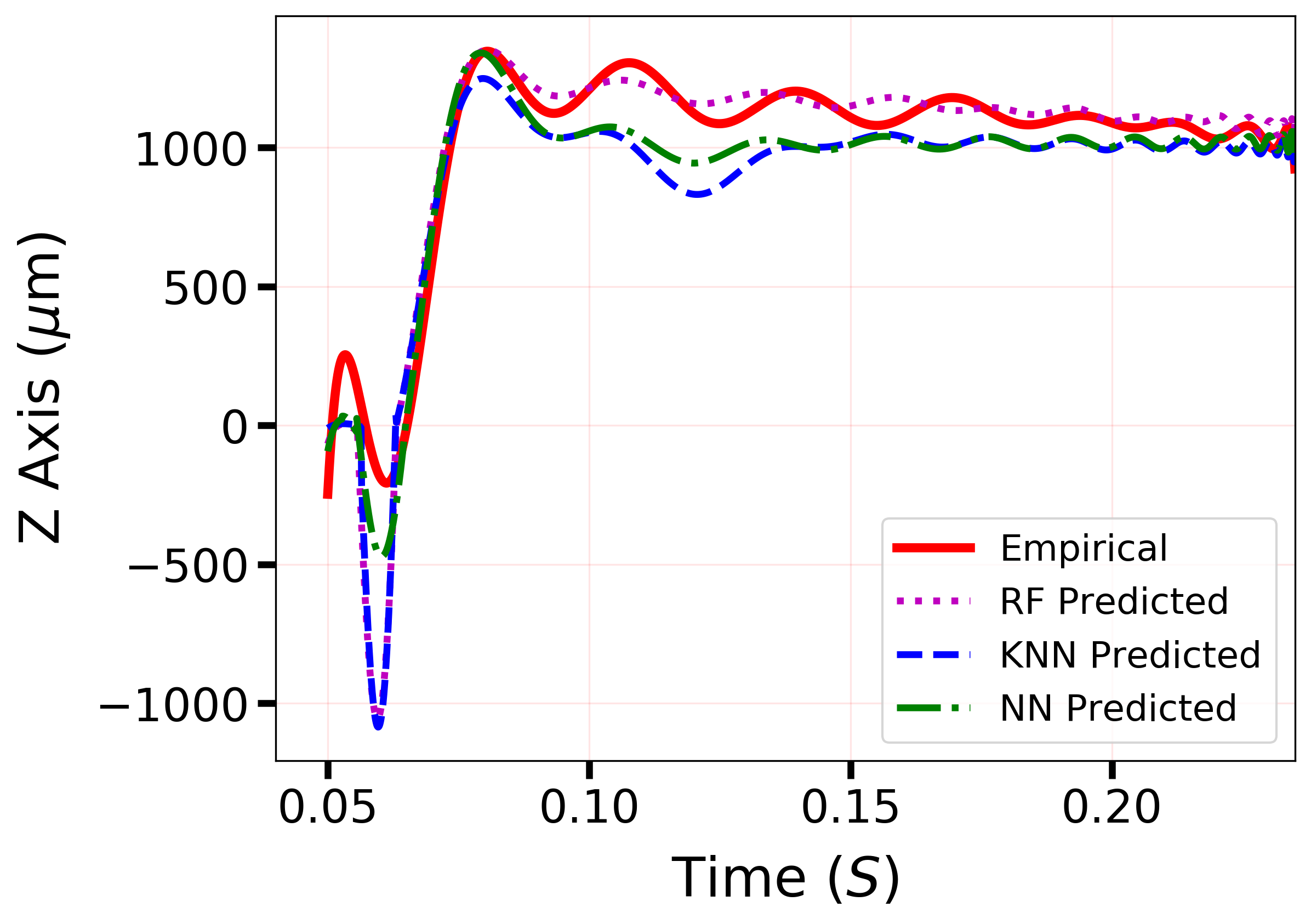
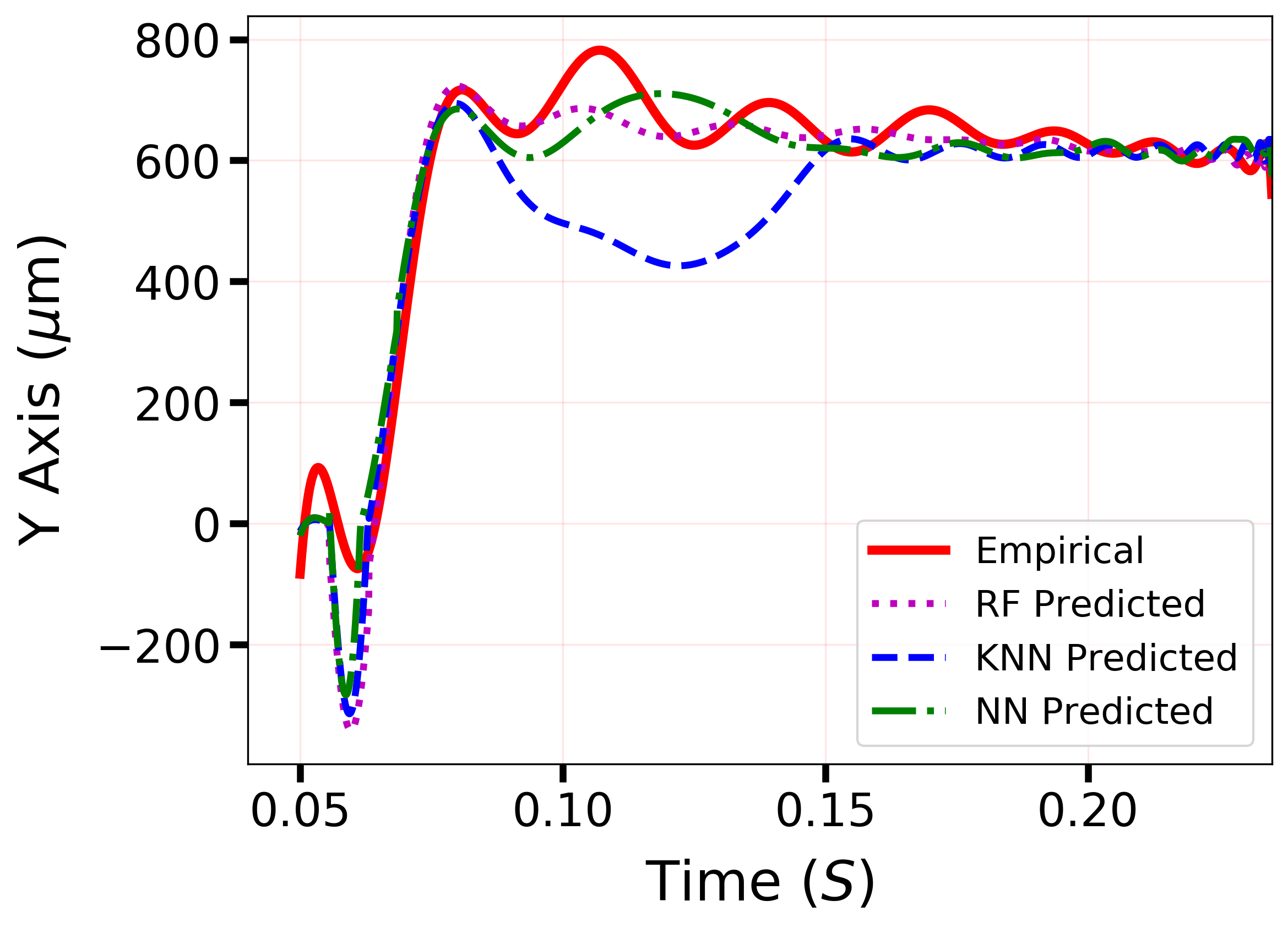
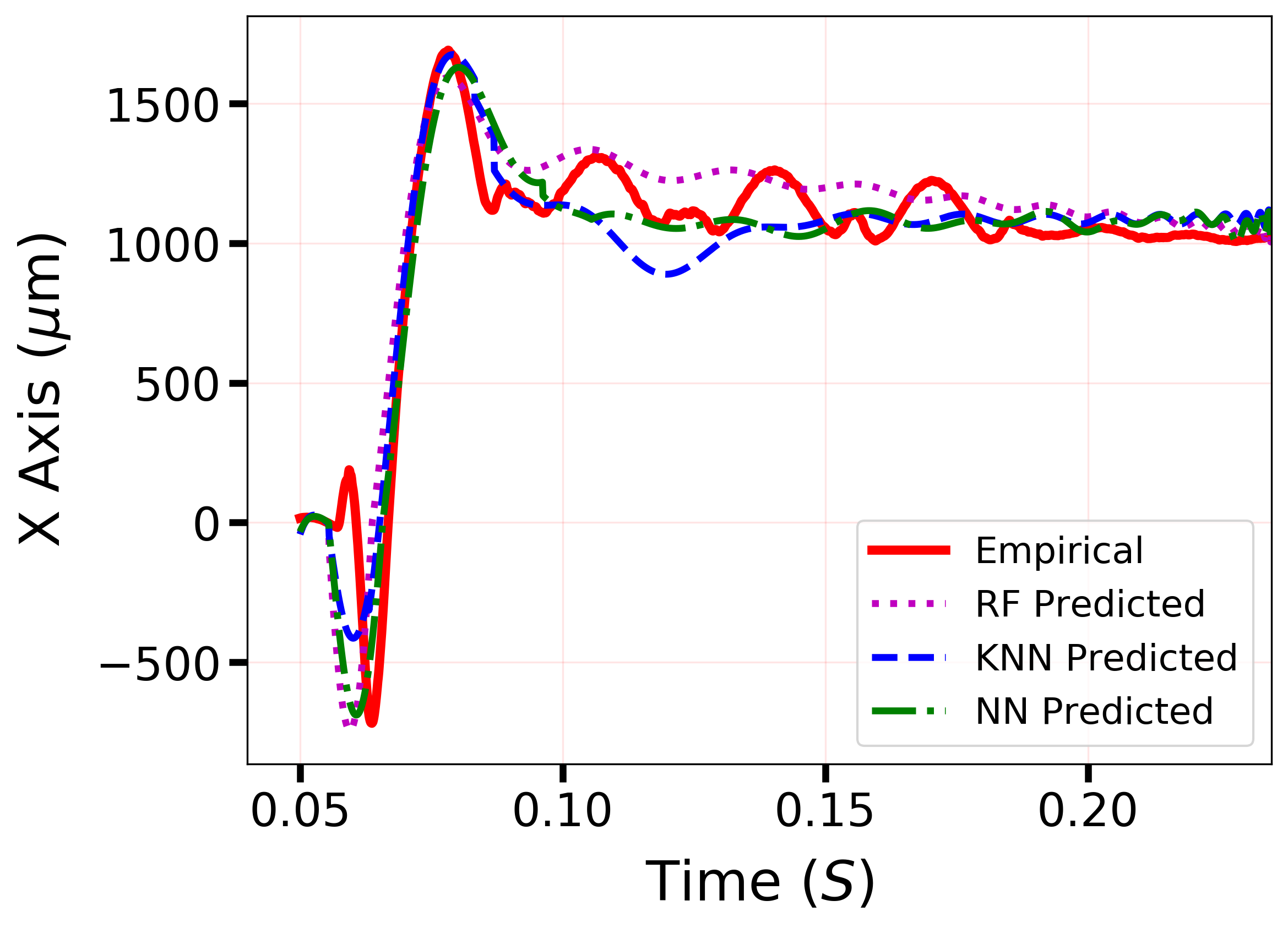
1. Fig.10: Instead of one trajectory line as in Fig.10c&d - discuss how to convert it to a 2D figure, i.e. to a full figure close to that obtained from the camera in the 2D Fig.9a.

**Answer:** We thank the reviewer for the suggested different view. We have plotted the 2D figures as suggested (attached below). We observed a few minor discrepancies among models, which were reported throughout the manuscript. As we can see from the plots, the predicted x, y and z values followed the empirical x, y, z values closely when predicted using RF (R2 > 0.68), KNN (R2 > 0.79), and MLP-NN (R2 > 0.78) for dead rat and RF (R2 > 0.86), KNN (R2 > 0.78), and MLP-NN (R2 > 0.87) for live rat. However, we believe that it is better to present our results grouped by the ML algorithms to make clear transition to 3D trajectory in Fig. 9(c) and (d).

*Dead rat:*



*Live rat:*



1. How will the magnets arrangement will have to adjust - what will be the challenges?

**Answer:** We thank the reviewer for the valuable question. The soft magnet was implanted on to the surface of the dura mater. Once it was implanted, it was fixed, and no adjustment was needed. However, it was only true for our experiments, which was the sensing system validation. If the long-term biological study is designed, the arrangement could be necessary. In such case, the arrangement of the soft magnet can be done by external magnet that would steer the soft magnet. However, it may cause additional injury. This will be considered as the future study. We have summarized this and added the following in the manuscript.

* Pg. 7, right column, line 20-21: *“There was also a potential misorientation effect of the soft magnet during brain deformation, which we will pursue in our future work.”*

1. Fig. 10 - negative values: In Fig.10b for NN ML-model the Z reaches up to Z=-1000 while in Fig.10d NN Z-axis reaches to no farther than Z=-250. Why?

**Answer:** We thank the reviewer for the question. In fact, the z-direction deformation reached up to -1000 µm in both Fig. 10(b) and 10(d) (currently Fig. 9(b) and 9(d)). If the reviewer meant the z-direction deformation was -1000 µm for live animal and -250 µm for dead animal, the z-axis deformation for dead animal was lower due to the additional post-mortem stiffening created in the brain which may have changed the viscoelastic behavior.

1. In Fig.10a&c vs Fig.10b&d: dead animal NN ML-model has only negative Z while live animal has mostly positive Z (except from the beginning) - why? could it be due to different measurement time (dead animal: 250-450ms, live animal: 50-250ms)? - discuss.

**Answer:** Authors thank the reviewer for the insight. The live rat brain was more pliable and softer compared to the dead rat brain. Also, the sustained displacement was higher for the live rat (1.6 mm) compared to dead rat (1 mm) [26]. The value of the measurement time is respect to the moment of the blast wave exposure; thus, it is independent. Although, the time duration would matter. Therefore, we hypothesis that the difference between the brain deformation direction is due to the post-mortem stiffening. The post-mortem lack of blood perfusion and consequential decreased water constant, resulting solid-like characteristics when compared to the live rat [26].

1. It was indicated (page 6 lines 36-37) that NN was found to be the best ML model (highest correlation). There are three other findings that might support that NN-ML is superior (discuss):

i) Smoothness of the trajectory - Fig 10d: For live animal the ML lines in Fig.10d doesn't look smooth at short times whereas [17] looks smother. Nevertheless, the NN looks the most smoothest in relation to other ML models and therefore have a more physiological representation. There are methods to calculate the smoothness of a trajectory - consider calculating (or discussing) trajectory smoothness for Fig 10d.

Note, however, that in Fig. 10d the NN trajectory looks exceptional in sense that it has different shape (particularly for short times) compared to all others (other ML & [17]).

ii) Camera - Fig. 9c: Indeed NN output looks the closest to the camera output - in particularly at the beginning and at the end of the signal.

iii) Minimum - Fig.10d: In Fig.10d NN reaches to a minimum of Z=-250 while others reach to a minimum of Z=-1000. Could this indicates that NN represents a more physiological representation of a live animal?

**Answer:** We thank the reviewer for the comments. It is true that we have found MLP-NN to produce most smooth data compared to other methods. As suggested, we have calculated trajectory smoothness for Fig. 9 (c) and (d) (formally Fig. 10(c) and (d)) in the discussion of the manuscript. We also add two more performance indicators, a few more comparisons to address the superiority of the MLP-NN (please see our answers in Question #8, #20, and #22).

* Pg. 6, right column, line 15-27: *“We also calculated the line root mean square roughness average (Rq) of the deformation trajectories to find the smoothest prediction. Rq can be defined as the arithmetic mean of the RMS value of the roughness profile determined by the squared deviation from the centerline for the length of the profile. For in vitro the needle insertion test, the values were 583 µm, 518 µm and 469 µm for RF, KNN and MLP-NN, respectively. For in vivo blast wave experiment, the values were 136.69 µm, 106.79 µm and 94.87 µm for the dead rat, and 154.33 µm, 147.07 µm and 126.03 µm for live rat using RF, KNN and MLP-NN, respectively. The higher smoothness of trajectory using NN may indicates a more physiological representation of actual tissue deformation, which requires further investigation.”*

1. Discuss the benefits of MTJ: MTJ is not expensive (?) and/or is more accessible than MRI/CT. MTJ can be used, for example, for the TBI from the shaken baby syndrome (SBS): In SBS, sometimes CT exams are inconclusive, perhaps MTJ could add additional information? MTJ with supplementary measurement of neuronal noise (*Dvir et al. 2018 Sci. Adv. 4 eaar6277*) can be used jointly to rule out sudden infant death syndrome (SIDS) in SBS cases.

**Answer:** We thank the reviewer for the insight and suggestion. MTJ sensor is indeed low cost than MRI/CT scan. It is smaller in size and is a modular device. Our proposed system uses the MTJ sensor array, which can also be beneficial to the shaken baby syndrome. We have added this in the manuscript.

* Pg. 2, left column, line 41-43: *“The proposed sensing system is low cost compare to MRI/CT scan, smaller in size and is a modular device.”*
* Pg. 8, left column, line 13-18: *“**Another application could be integrating our sensing scheme with existing methods (neuronal noise measurement) in diagnosing TBI from shaken baby syndrome (SBS) to assist in finding conclusive results to rule out infant death syndrome (SIDS) in SBS cases [54].”*

1. It is needed to further emphasize the implementations challenges that were overcame in order to improve the previous sensors (study [17]) to the current MTJ sensor.

**Answer:** By using MTJ sensors there was 3 orders of magnitude increase in sensing volume as well as ten times higher sensing range on z-direction compared to our previous study. By using this sensor, we could capture deformations which was out of range for detection with the GMR sensors. Therefore, it provides access to sensing brain deformation inside thicker skull (larger mammals vs. rat). Also, the soft magnet size could be minimized in the future to reduce the surgical footprint and change in physiological properties of the brain tissue. The implementation challenges were in fact with the ML, which we fully explored in the manuscript. To emphasize our point, we point out the following. Additionally, we discussed the novelty of our work (please see our answer for Question #4).

* Pg. 2, right column, line 43-49: *“The MTJ sensors provide a sensitivity that is ten times greater than that provided by the giant magnetoresistive (GMR) sensors which were used in our previous study [25]–[27]. Extended range and sensitivity help to minimize soft magnet dimension to reduce surgical footprint and change in physiological properties of the brain tissue.”*

1. In Fig. 2b: What represent the black dots and what the red line?

**Answer:** We thank the reviewer for the comments. Black dots are the data points. Red line is the connecting line. Each data point represents the magnitude of magnetic strength at the given time. However, we have removed the dots to simplify the representation of magnetic sensor data.