# Stacking Ensemble of Machine Learning Methods for Brain Tumor Size Prediction

***Abstract*—Predicting the size of human brain tumors accurately is a crucial step in early and effective treatment to improve patient outcomes. While available imaging techniques, such as MRI and CT scans, provide detailed insights into brain structures, but often involve high installation costs, complex installation procedures, lengthy scanning times and the need for extensive manual interpretation.**

**To address these challenges, we propose a stacking ensemble-based machine learning approach for predicting brain tumor diameter and volume using electromagnetic data obtained from a custom-designed patch antenna. A simulation of a multi-layered brain phantom using CST 2024 was designed to replicate the dielectric properties of human brain tissues, allowing us to extract essential features such as return loss, specific absorption rates, surface current, and electric and magnetic field values.**

**Five regression-based machine learning models, Linear Regression, Ridge Regression, Support Vector Regression (SVR), Random Forest (RF), and XGBoost, were trained on the simulated dataset. Then we combined the predictions through a stacking regressor to enhance overall accuracy. Experimental results show that the proposed model outperforms individual algorithms, achieving a high $R^2$ score of 0.9855 for tumor diameter prediction and 0.9749 for tumor volume estimation and the lowest RMSE and MAE. These results suggest that the combination of antenna-based clinical data collection and analysis with advanced machine learning techniques can serve as a next-generation diagnostic model, which will be a cost-effective, non-invasive, and reliable framework for brain tumor assessment.**

***Index Terms*—Machine learning, Stacking ensemble, CST simulation, Brain tumor, Regression model.**

## I. Introduction

A brain tumor is considered one of the most severe types of tumors, significantly impacting global health. Early detection with accurate estimation of tumor size is essential for appropriate treatment. Traditional diagnostic tools, such as MRI and CT scans, are widely used worldwide for identifying tumors and measuring the area where they have spread. While effective, these methods often suffer from limitations, including high operational costs, a time-consuming and complex installation process for the machine.

In recent years, machine learning (ML) has emerged as a powerful tool for medical diagnostics. ML can identify the complex interactional data patterns of biological and electromagnetic data. Our study introduces a stacking ensemble approach to predict brain tumor size and volume. Instead of relying on a single model, our stacking model considers multiple regression techniques, including Linear Regression, Ridge Regression, Support Vector Regression (SVR), Random Forest (RF), and XGBoost to create a robust meta-model capable of minimizing prediction errors.

We designed a microstrip patch antenna that was simulated within the CST Studio Suite 2024, along with a multi-layered brain phantom model mimicking the realistic tissue dielectric properties. The simulation was accomplished by varying the tumor diameter and volume multiple times. The antenna's electromagnetic responses, including return loss, electric and magnetic field strengths, surface currents, and specific absorption rates, were extracted and used as features for model training.

Our experimental findings reveal that the application of ML stacking regressor on our dataset outperforms individual baseline models and demonstrates high accuracy and low error rates for both tumor diameter and volume predictions. By integrating antenna-based simulation data with advanced machine learning techniques, this study contributes a promising, non-invasive, and cost-effective diagnostic framework that could complement existing imaging modalities and support improved clinical decision-making in brain tumor assessment.

## II. Related Works

Machine learning (ML) techniques are increasingly vital for interpreting the intricate datasets produced by antenna-based diagnostic systems in oncology. These algorithms are deployed to process antenna-derived signals, thereby improving the precision of cancer identification.

Research by Singh et al. [1] evaluated multiple ML models including Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron (MLP)—on S-parameter data, focusing on the S21 transmission coefficient's magnitude and phase. Their work utilized a dataset of 80 S21 variations representing different hemorrhage stages and one healthy model. The findings indicate that combining ML with a basic microwave antenna setup can substantially boost the accuracy of non-invasive brain hemorrhage diagnosis. In their study, the Decision Tree and Random Forest algorithms yielded the highest performance, with accuracies of 92% and 94%, respectively.

A separate investigation by Diao et al. [2] introduced the CondNet, a multi-track convolutional neural network (CNN) designed to deduce tissue dielectric properties from MRI-derived anatomical data. This architecture generates a continuous distribution of physical properties, overcoming the abrupt transitions typical of segmented voxel models.

Further comparative research [3] assessed Slot (XETS) and Vivaldi (BAVA) antennas for breast tumor detection, combining microwave imaging (MWI) with machine learning classifiers. Similarly, Aydin et al. [4] developed a method for

breast cancer detection using MWI and data mining. Their process involved creating a dataset from simulation data that included tumor characteristics, dielectric properties, frequency, and electric field values. Applying the KNN algorithm to 6,006 samples resulted in a 90% tumor detection accuracy rate.

In a study on tissue property prediction, Mattsson et al. [5] used a bandstop sensor to measure tissue permittivity and an ML algorithm to predict properties of skin, fat, and muscle. A Random Forest classifier achieved 82% accuracy for skin thickness prediction.

Senthil et al. [6] designed a flexible terahertz (THz) microstrip patch antenna for detecting superficial cancers like liposarcoma and basal cell carcinoma. The antenna, simulated in CST Microwave Studio, achieved a return loss of -38 dB and a gain of 9.6 dBi when tested on a tissue-mimicking phantom with a tumor. A hybrid ML model integrating K-means clustering with logistic regression reached a diagnostic accuracy of 95.06% by capitalizing on the dielectric contrast between healthy and malignant tissues. This THz sensing approach, enhanced by AI, showed superior classification accuracy and fewer false positives than standard MRI or X-ray methods, though challenges like limited penetration depth were noted.

Complementing this, Al-Gburi et al. [7] implemented an artificial neural network (ANN) to automatically classify breast tissue using microwave scattering data. Their system featured a high-gain Vivaldi antenna with a metasurface layer to improve sensitivity. The ANN was trained on features such as specific absorption rate (SAR) and return loss (S11) collected from a breast phantom with tumors of different sizes and locations. The model successfully learned to identify tumor-induced dielectric variations, achieving an 89% accuracy in classifying tissue as benign or malignant, showcasing ML's ability to decipher complex electromagnetic patterns for early diagnosis.

## III. Methodology

The methodology of our work comprises two segments. Where first segment consists of the antenna and phantom design, building a dataset from simulation. The second segment consists of data analysis using different machine learning models.

### A. Antenna and phantom Design

*1) Antenna specification:* For our simulation purposes, we designed a range of 2-4 GHz, a patch antenna designed on an FR-4 substrate (Fig. 1). We used lossy copper for the patch and ground material. The antenna resonates at 3.23 GHz. For impedance matching with the 50-ohm input, we used a quarter-wave transformer in the feedline.

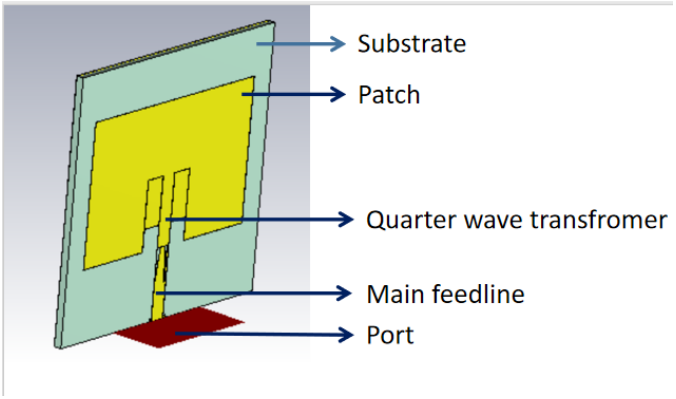


Fig. 1: Antenna design

TABLE I: Design Parameters of the Microstrip Patch Antenna

| Parameter | Value |
|---|---|
| Patch length ($L_p$) | 30 mm |
| Patch & ground thickness ($L_p$) | 30 mm |
| Patch Width ($W_p$) | 48 mm |
| Substrate length ($L_s$) | 55 mm |
| Substrate thickness ($h$) | 1.6 mm |
| Quarter-Wave transformer ($L_t \times W_t$) | 12.6 mm $\times$ 3.5 mm |
| Vertical feed length ($W_f$) | 15 mm |

*2) Phantom modeling:* Brain phantom refers to the simulation of the human brain with a specific tissue layer. We tried to mimic the exact human brain as much as possible by choosing a six-layer spherical shape and the dielectric properties of tissues. The specification of tissue and the six-layer listed in TABLE II. The phantom is placed between the transmitter and receiver antenna (see Fig.2). When the transmitter antenna transmits a signal to the receiver antenna, the phantom absorbs some portion of the signal, and we get different values in the receiver antenna for each antenna parameter.

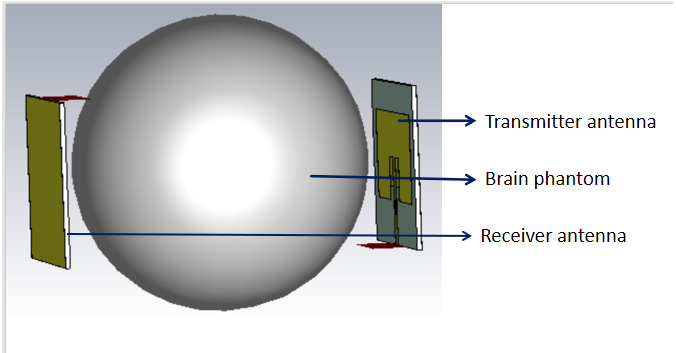


Fig. 2: Phantom position between the antennas

TABLE II: Dielectric Properties of Head Tissues

| Tissue | Thick. (mm) | $\varepsilon_r$ | Cond. [S/m] | tan$\delta$ | Ref. |
|---|---|---|---|---|---|
| Skin | 2 | 40.0 | 1.46 | 0.31 | [8] |
| Fat | 5 | 5.5 | 0.11 | 0.17 | [9] |
| Skull | 7 | 12.5 | 0.34 | 0.14 | [9], [10] |
| CSF | 2 | 69.0 | 2.41 | 0.22 | [9] |
| Brain (Grey) | 15 | 50.0 | 1.85 | 0.26 | [9], [11] |
| Tumor (Malignant) | 5–10 | 60.0 | 2.20 | 0.28 | [11], [12] |

*3) Dataset description:* From the simulation model, we construct the dataset using CST Studio Suite 2024. We collected 100 samples from the simulation results, focusing on the output parameter Cancer cell diameter, Cancer cell volume, $S_11$ (return loss), SAR (specific absorption ratio), $E_1$ (electric field at port 1), $E_2$ (electric field at port 2), $H_1$ (magnetic field at port 1), $H_1$ (magnetic field at port 2), $SC_1$ (surface current at port 1) and $SC_2$ (surface current at port 2). Where Cancer cell diameter, Cancer cell volume are the target parameters, and the rest of the parameters are feature parameters. However, the tumor diameter may help us to presume the depth of the cancer tissue, and tumor volume may help us to understand how much the cancer has spread.

### B. Application of machine learning

Machine learning (ML) is a branch of artificial intelligence that enables computers to learn patterns from data. The learning of data patterns helps computers to make predictions or decisions without being explicitly programmed [13]. There are several machine learning models, such as Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT), etc, that can be deployed to understand the data patterns and relationships of the variables from the dataset. However, our objective is to build an algorithm that can predict the tumor size and thickness (depth) successfully. To do so, we have chosen five ML regression models, such as Linear Regression, Ridge regression, SVR, RF and XGB.

*1) Data preprocessing:* Data preprocessing is the initial step in data analysis or machine learning, where raw data is cleaned, transformed, and prepared so that algorithms can work on it effectively. Raw data is often incomplete, noisy, inconsistent, or unstructured, so preprocessing ensures the data becomes usable, accurate, and meaningful. Using basic Python code, we performed data preprocessing to check for null values, verify data types, inspect data shape, and review statistical summaries.

*2) Pipelining:* In scikit-learn, a Pipeline is a way to chain multiple steps together so that they execute sequentially. Pipelines often include preprocessing steps like scaling or encoding. But the key idea is automation and safety. The pipeline ensures that the same preprocessing is applied consistently to both training and test data. It also prevents data leakage, because the transformations are fitted only on the training data and then applied to the test data.

*3) Evaluation parameter:* To know the predictive model performance, we selected Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Determination ($R^2$ score) evaluation parameters. To decide which model is best, we typically want the lowest errors (MAE, RMSE) and highest $R^2$ on the test set.

## IV. Results & Discussion

### A. Antenna output

First, we have to look at the antenna output to check whether it has an acceptable output or not. The antenna input frequency range we provided at 2-4GHz, where the antenna resonates at 3.23 GHz and gives an output of return loss $S_{11} = -26.65dB$ shown in Fig.3. The Voltage Standing Wave Ratio (VSWR) we got was 1.1, which is a very good value for a patch antenna ( see Fig.4). For patch antennas, a VSWR $\leq 2$ is acceptable, but designers usually aim for $\leq 1.5$ for better efficiency.

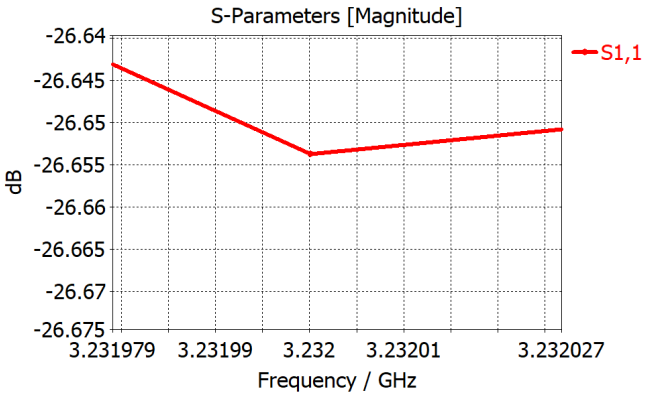


Fig. 3: Antenna return loss

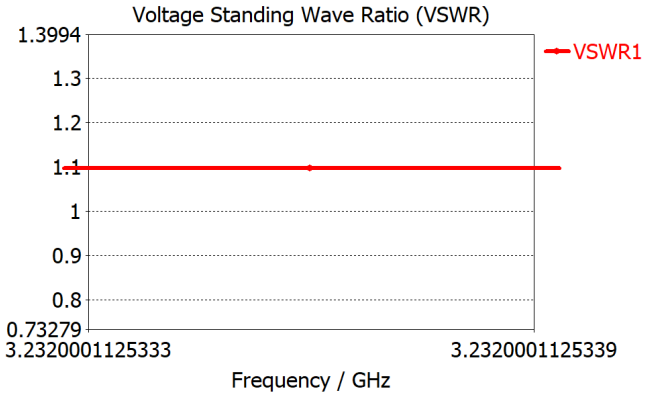


Fig. 4: Antenna VSWR

### B. Feature analysis

Feature analysis helps us to understand the rank of features that are correlated with the target variables. For predictive models, analyzing the features is very important. Because some features may be redundant or harmful in predicting the model. We have chosen ten feature variables and two target variables for our study (described in the dataset description section). To perform the feature analysis, we used a combination RF and XGB model. The importance score we got listed in the TABLE III demonstrates that no features are irrelevant.

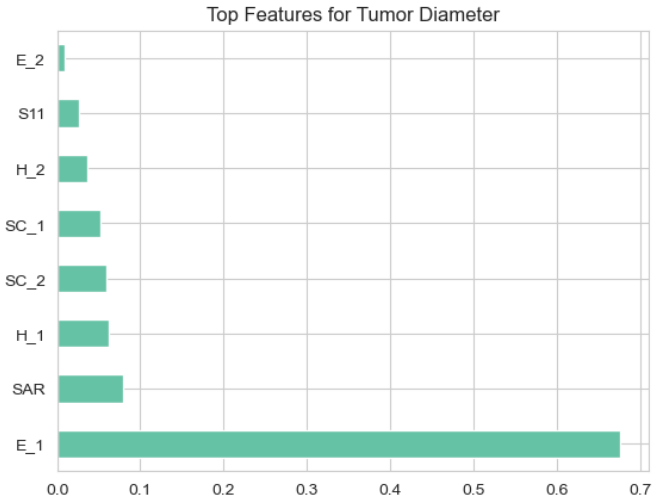


Fig. 5: Features correlated to tumor diameter

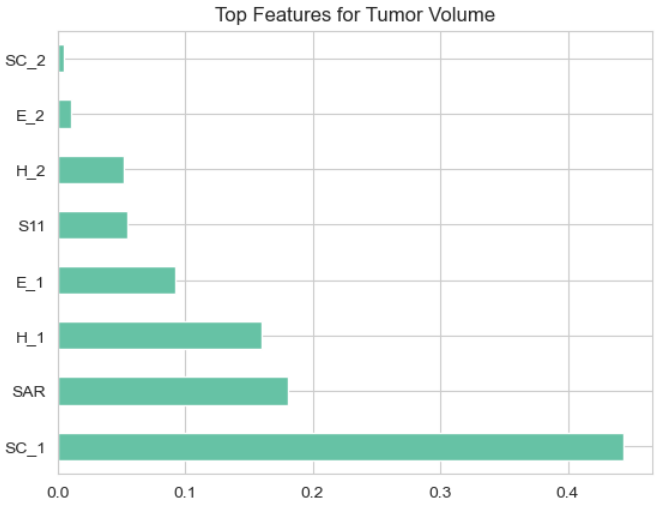


Fig. 6: Features correlated to tumor volume

But $E_1$ has the highest correlation with tumor diameter (Fig. 5), and $SC_1$ has the highest correlation with tumor volume (Fig. 6). So, analyzing the importance level, we decide not to exclude any variables or features for our future proceedings.

TABLE III: Feature Importances for Tumor Diameter and Tumor Volume (RF)

| Feature | Tumor Diameter | Tumor Volume |
|---|---|---|
| $E_1$ | 0.675780 | 0.092442 |
| SAR | 0.079222 | 0.181022 |
| $H_1$ | 0.061834 | 0.160288 |
| $SC_2$ | 0.058923 | 0.004667 |
| $SC_1$ | 0.052207 | 0.443637 |
| $H_2$ | 0.036889 | 0.052191 |
| S11 | 0.026635 | 0.055101 |
| $E_2$ | 0.008510 | 0.010652 |

### C. Tumor thickness and volume prediction

*1) Baseline model:* From TABLE IV & V, we can see that SVR is underperforming. On our tumor dataset, SVR (with an RBF kernel) had a negative $R^2$ for volume and a much higher MAE/RMSE than RF/XGB/Linear. The reason for underperformance may be due to improper data preprocessing, lack of feature scaling and unstable single-split evaluation. We trained our models directly on raw data that may have caused missing values to be handled inconsistently.

TABLE IV: Baseline performance of regression models for tumor diameter

| Model | Cross-Validation (CV) | | | Test Set | | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | $R^2$ | MAE | RMSE | $R^2$ |
| Linear | 5.68 | 7.36 | 0.715 | 5.64 | 6.75 | 0.785 |
| Ridge | 5.72 | 7.36 | 0.717 | 5.71 | 6.88 | 0.777 |
| SVR | 9.23 | 11.12 | 0.355 | 7.74 | 9.59 | 0.567 |
| RF | **4.10** | **6.08** | **0.799** | **3.64** | **5.69** | **0.848** |
| XGB | 4.95 | 6.88 | 0.751 | 4.87 | 8.20 | 0.683 |
| **Best Model for Tumor Diameter: Random Forest (RF)** | | | | | | |

TABLE V: Baseline performance of regression models for tumor volume

| Model | Cross-Validation (CV) | | | Test Set | | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | $R^2$ | MAE | RMSE | $R^2$ |
| Linear | 38901.32 | 52601.52 | 0.866 | 35717.83 | 46376.75 | 0.917 |
| Ridge | 39169.71 | 53603.36 | 0.862 | 37172.83 | 49449.45 | 0.905 |
| SVR | 117932.35 | 169548.08 | -0.337 | 139998.91 | 194396.86 | -0.466 |
| RF | **25465.99** | **36963.21** | **0.938** | **26906.06** | 50188.87 | **0.902** |
| XGB | 35064.10 | 53253.52 | 0.861 | 26605.38 | 52535.70 | 0.893 |
| **Best Model for Tumor Volume: Random Forest (RF)** | | | | | | |

*2) Regression model after data preprocessing and pipelining:* After the application of data preprocessing and proper scaling output result of SVR dramatically increased $R^2$ from 0.567 to 0.94131 in diameter prediction and from -0.466 to 0.94383 in volume prediction. It's not only SVR that improved, but almost all models show better performance (see TABLE VI). The Linear model performed best for cancer cell diameter prediction, and Rf worked best in volume prediction.

We used a pipeline with *SimpleImputer* to ensure no row or column losses from NaNs, and *StandardScaler* to normalize feature ranges. Later, integrated into CV, and the same preprocessing was applied consistently to each fold. This improvement is primarily due to three methodological refinements: (i) consistent preprocessing using imputation and feature scaling, (ii) integration of these steps into model pipelines to avoid data leakage, and (iii) evaluation with K-fold cross-validation rather than a single train-test split. Linear and Ridge regression benefited from standardized predictors, SVR performance dramatically improved due to kernel distance normalization, while tree-based models (RF, XGBoost) showed enhanced generalization from cross-validation.

*3) Stacking model:* A stacking model is an ensemble learning technique where we combine multiple different machine

TABLE VI: Upgraded baseline regression models' performance for tumor diameter and volume

| Target | Model | CV MAE | CV RMSE | CV $R^2$ | Test MAE | Test RMSE | Test $R^2$ |
|---|---|---|---|---|---|---|---|
| Diameter | **Linear** | **0.09576** | **0.11846** | **0.9833** | **0.07765** | **0.09879** | **0.98426** |
| Diameter | Ridge | 0.17589 | 0.21970 | 0.9537 | 0.11064 | 0.12647 | 0.97421 |
| Diameter | SVR | 0.22289 | 0.31401 | 0.9075 | 0.15522 | 0.19079 | 0.94131 |
| Diameter | RF | 0.22984 | 0.27978 | 0.9185 | 0.17290 | 0.23407 | 0.91166 |
| Diameter | XGB | 0.23905 | 0.29929 | 0.8950 | 0.16960 | 0.20634 | 0.93135 |
| Volume | Linear | 0.28118 | 0.34059 | 0.8939 | 0.31340 | 0.39482 | 0.90123 |
| Volume | Ridge | 0.30635 | 0.40125 | 0.8569 | 0.30504 | 0.42150 | 0.88743 |
| Volume | SVR | 0.30047 | 0.40028 | 0.8609 | 0.24524 | 0.29773 | 0.94383 |
| Volume | RF | **0.27957** | **0.36271** | **0.8835** | **0.16043** | **0.19904** | **0.97490** |
| Volume | XGB | 0.29685 | 0.37773 | 0.8690 | 0.18059 | 0.21411 | 0.97095 |

learning models and then use another model called the meta-learner to learn how to best combine their predictions [14]. The base models (level-0) we used in stacking are Linear, Ridge, SVR, RandomForest, and XGB models. We considered *LinearRegression()* as the final estimator (level-1 / meta-learner), hence it provides the best results. The performance of these models is discussed earlier. During training, stacking uses 5-fold cross-validation to generate predictions from base models so that the meta-learner is trained on out-of-fold predictions to avoid overfitting. The output value of the CV and test set we get from the Stacking model is listed in the TABLE IX

From the stacking experiment, we found that the Stacking model outperformed all other baseline models. For diameter, the Stacking model test MAE = 0.0780, test MSE = 0.0948. It provides a lower MES value than the best diameter predictor Linear model. Comparisons of the best diameter predictor models, Linear and Stacking, are shown in the TABLE VII.

TABLE VII: Comparison of Linear Regression and Stacking for Tumor Diameter Prediction

| Model | Test MAE | Test RMSE | Test $R^2$ |
|---|---|---|---|
| Linear | **0.0776** | 0.0988 | 0.9843 |
| Stacking | 0.0780 | **0.0948 (best)** | **0.9855 (best)** |

Similarly, in terms of volume prediction, Stacking also outperformed all other baseline models. Test RMSE and $R^2$ are almost the same, but Test MAE has a much lower value than RF model.

TABLE VIII: Comparison of Random Forest and Stacking for Tumor Volume Prediction

| Model | Test MAE | Test RMSE | Test $R^2$ |
|---|---|---|---|
| RF | 0.1604 | **0.1990** | **0.9749** |
| Stacking | **0.1570 (best)** | 0.1991 | **0.9749** |

TABLE IX: Performance of Stacking Regressor for tumor diameter and volume

| Target | Model | Cross-Validation (CV) | | | Test Set | | |
|---|---|---|---|---|---|---|---|
| | | MAE | RMSE | $R^2$ | MAE | RMSE | $R^2$ |
| Diameter | Stacking | 0.1035 | 0.1291 | 0.9811 | 0.0780 | 0.0948 | 0.9855 |
| Volume | Stacking | 0.2064 | 0.2551 | 0.9416 | 0.1570 | 0.1991 | 0.9749 |
| **Best Model across both targets: Stacking Regressor** | | | | | | | |

The scatter plot of the regression that predicts tumor diameter and volume are shown in Fig. 7 & 8. This visualizes how good the model is working. The dot point is much closer to the red line means the predicted value is very much close to the actual value for both diameter and volume.

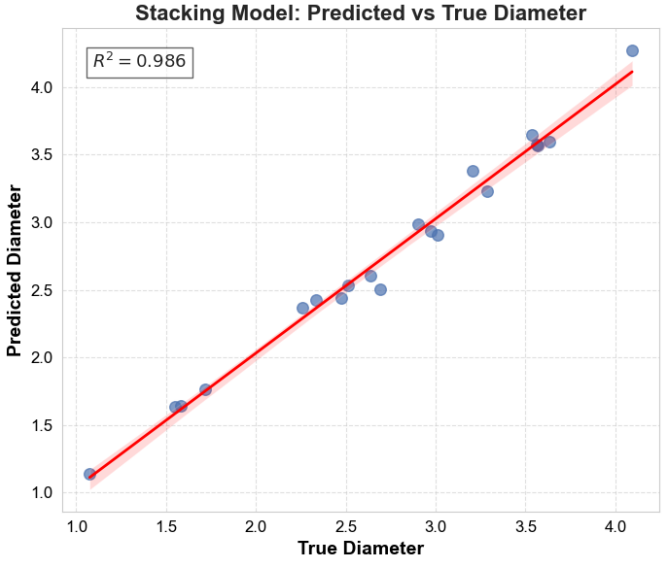


Fig. 7: Regression line for tumor diameter prediction

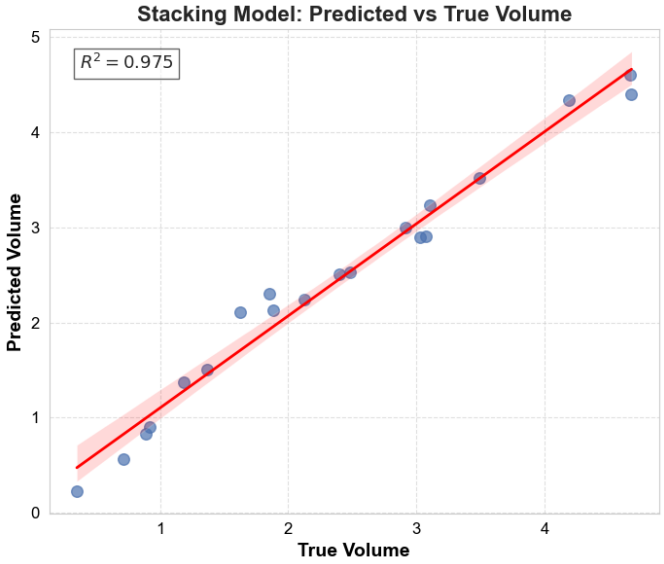


Fig. 8: Regression line for tumor volume prediction

## V. Conclusion

In our study, we worked on a simulated data sample, and the sample was intentionally kept small (100 samples). The reason behind keeping the dataset small is to make it suitable for real-world scenarios. Collecting clinical biomedical-related cancer data may be hard in the real world. So we tried to overcome this problem and build a model that is adaptive to a small data set. Our main objective was to study the feasibility of

whether brain size is related to the antenna data or not, and find a regression model that can predict the tumor size.

## VI. Future Works

Future work of our study centered on collecting clinical data from cancer patients with a real modeled antenna. If the dataset can be extended to a large sample, the feasibility of deep learning techniques such as Convolutional Neural Networks (CNNs), Transformers, or Graph Neural Networks (GNNs) can also be studied to predict the tumor.


## Acknowledgement

The authors acknowledge the use of AI-assisted tools (OpenAI ChatGPT, Deepseek and Grammarly) for grammar refinement, improving academic tone, and suggesting better phrasing. All ideas, experimental designs, and conclusions remain the sole responsibility of the authors.


## References


[1] A. Singh, B. Mandal, B. Biswas, S. Chatterjee, S. Banerjee, D. Mitra, and R. Augustine, "Microwave antenna-assisted machine learning: A paradigm shift in non-invasive brain hemorrhage detection," *IEEE Access*, 2024.

[2] Y. Diao, E. A. Rashed, and A. Hirata, "Large-scale analysis of the head proximity effects on antenna performance using machine learning based models," *IEEE Access*, vol. 8, pp. 154 060–154 071, 2020.

[3] R. A. Martins, J. M. Felício, J. R. Costa, and C. A. Fernandes, "Comparison of slot-based and vivaldi antennas for breast tumor detection using machine learning and microwave imaging algorithms," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2021, pp. 1–5.

[4] E. A. Aydın and M. Kaya Keleş, "Breast cancer detection using k-nearest neighbors data mining method obtained from the bow-tie antenna dataset," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 27, no. 6, p. e21098, 2017.

[5] V. Mattsson, M. D. Perez, L. Joseph, and R. Augustine, "Machine learning algorithm to extract properties of ate phantoms from microwave measurements," *International Journal of Microwave and Wireless Technologies*, pp. 1–8, 2024.

[6] M. Senthil Pandian, S. D. Nivethika, J. Idhikash, V. N. Yashwanth, A. Shaji, and P. Paulraj, "Cancer diagnosis optimization with a combination of flexible thz antennas and machine learning," *Engineering Reports*, vol. 7, no. 4, p. e70120, 2025.

[7] R. M. Al-Gburi, M. Alibakhshikenari, B. S. Virdee, T. M. Hameed, D. Mariyanayagam, S. Fernando, I. Lubangakene, Y. Tang, S. U. Khan, and T. A. Elwi, "Microwave-based breast cancer detection using a high-gain vivaldi antenna and metasurface neural network approach for medical diagnostics," *Frequenz*, 2025.

[8] V. Raicu, N. Kitagawa, and A. Irimajiri, "A quantitative approach to the dielectric properties of the skin," *Physics in Medicine & Biology*, vol. 45, no. 2, p. L1, 2000.

[9] S. Gabriel, R. Lau, and C. Gabriel, "The dielectric properties of biological tissues: Iii. parametric models for the dielectric spectrum of tissues," *Physics in medicine & biology*, vol. 41, no. 11, p. 2271, 1996.

[10] J. Wang and O. Fujiwara, "Fdtd computation of temperature rise in the human head for portable telephones," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1528–1534, 1999.

[11] M. Lazebnik, D. Popovic, L. McCartney, C. B. Watkins, M. J. Lindstrom, J. Harter, S. Sewall, T. Ogilvie, A. Magliocco, T. M. Breslin *et al.*, "A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries," *Physics in medicine & biology*, vol. 52, no. 20, p. 6093, 2007.

[12] W. T. Joines, Y. Zhang, C. Li, and R. L. Jirtle, "The measured electrical properties of normal and malignant human tissues from 50 to 900 mhz," *Medical physics*, vol. 21, no. 4, pp. 547–550, 1994.

[13] D. Pandey, K. Niwaria, and B. Chourasia, "Machine learning algorithms: a review," *Mach. Learn*, vol. 6, no. 2, 2019.

[14] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" *Machine learning*, vol. 54, no. 3, pp. 255–273, 2004.