Machine Learning for Signal Modeling of GaN HEMTs

Prediction of S-parameters for GaN HEMTs using ANN, RF, GBR, AdaBoost ML models

Capstone Report

Serik Aitzhanov

Nazarbayev University
Department of Electrical and Computer Engineering
School of Engineering and Digital Sciences

This project report was created on TexStudio editing platform using LaTeX. All the figures were drawn using draw.io online software tool.

Electrical and Computer Engineering
Nazarbayev University
http://www.nu.edu.kz

Title:
Machine Learning for Signal Modeling of GaN HEMTs (Prediction of S-parameters for GaN HEMTs using ANN, RF, GBR, AdaBoost ML models)

Theme:
Machine Learning Based Transistor Device Modeling

Project Period:
Fall 2024 - Spring 2025

Project Group:
RF and Microwave Characterization Labaratory

Participant(s):
Aitzhanov Serik

Supervisor(s):
Mohammad Hashmi

Copies: 1

Page Numbers: 36

Date of Completion:
April 17, 2025

Abstract:

The main motivation of this work is to apply machine learning (ML) based techniques to predict S-parameters of Gallium Nitride High Electron Mobility Transistors (GaN HEMTs) in high power and high frequency RF and microwave systems. However, traditional simulation-based or experimental methods are accurate but costly and take lots of time. This work proposes ML as a potential alternative model for such complex nonlinear operating conditions to S parameter responses in GaN HEMTs. With five supervised machine learning models crafted based on Artificial Neural Network (ANN), Random Forest (RF), Gradient Boosting Regressor (GBR), AdaBoost based with Decision Tree (ADA-DT), and AdaBoost based with Random Forest (ADA-RF), a dataset of 17,280 samples was trained and tested based on bias voltages, temperatures, and with a constant frequency of 26 GHz. Metrics such as MAE, MSE, R^2, and training time were used to assess the performance, and visualizations, in the form of line plots, biased frequency plots, Smith Charts, etc., were used. Results showed that the RF and GBR models were the best in terms of accuracy and generalization, while the ANN model is the fastest in learning, but lags. Frequency was one of the input variables, and a feature importance analysis showed that frequency was the most important variable. This work, therefore, emphasizes that ML can prove an important contribution towards speeding up the design cycle of GaN HEMT RF systems and improving the RF system performance.

Contents

Preface

I would also expressly thank my supervisor Professor Mohammad Hashmi for the invaluable help he always extended, through the duration of the project. Here, I met with severe illness during the fall semester, missing about half of the semester with my severe allergy. Nevertheless, Professor Hashmi was very understanding and supportive. He recommended what to read, resources, advice, and encouraged me to at least keep progressing with the research even under very difficult circumstances. Patience, care and dedication were highly important in completing this project the way I have!

Also, I]want to thank my university for giving me the needed infrastructure and resources to do this project. Furthermore, I can also credit the data and the tools that made a big difference in conducting the research. Without the help and the insights provided by Professor Hashmi this project would not be as deep and as qualitey that it holds.

I would like to lastly thank all who assisted me during this research process, and to apply learned insights from this project to future work and applications within the field.

Nazarbayev University, April 17, 2025

Serik Aitzhanov
<serik.aitzhanov@nu.edu.kz>

Chapter 1

Introduction

High power and high frequency systems have been designed with better performance through the use of semiconductor devices called Gallium Nitride High Electron Mobility Transistors (GaN HEMTs)[1][2]. Hence, GaN HEMTs are used generally because they have large advantages over traditional semiconductors from Si and GaAs, especially in the areas of Radio Frequency (RF) and microwave applications[3]. As GaN-based devices have high thermal conductivity, a wide band gap, and can work at high voltages and temperatures, they are more preferable than the Si-based devices. GaN HEMTs have these characteristics, which make them very suitable for applications where high power and efficiency are required[4]. Hence, all this fact makes GaN HEMTs so common and used by so many engineers in the modern engineering communication systems (radar systems, satellite communications, high-performance amplifiers, wireless communication networks).[5][6]

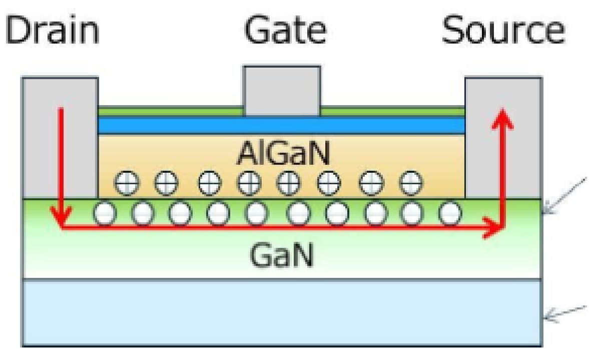


Figure 1.1: GaN HEMT semiconductor device.

Among the aspects given in the characterization of GaN HEMTs, one of the most important is the prediction and measurement of Scattering parameters, which are a collection of complex quantities to characterize how signals are propagated through the device[7]. However, we can regard S-parameters as the reflection, transmission, and loss characteristics of a device under different operating conditions, frequencies, and different bias points. The scattering parameters can solve the problem of performance assessment of the device, thus making the design of the RF circuits and systems more convenient. S-parameters are important in high-frequency applications such as communications and radar in order to understand how devices behave and how they will act in other components in the system[8].

However, previous experiments or simulation-based approaches were used for some of the measurements to predict the S parameters of GaN HEMTs. At the highest accuracy levels, the accuracy is high, but the computational cost and the time required for this accuracy is also high[9]. Due to the nonlinear behavior of GaN HEMTs, it requires a lot of resources for the simulation of GaN HEMTs under a wide range of conditions. Experimental approaches are usually expensive and time-consuming. A few of these challenges and the growing need for a better solution to model GaN HEMTs have encouraged the development of new alternative approaches to model these devices[10].

With more complex and performance-demanding RF systems, traditional methods of predicting S-parameters are outdated and insufficient[11]. A good solution to this problem may be implementing various Machine learning techniques to efficiently model the complex, nonlinear relationships between the operating conditions of GaN HEMTs and their performance characteristics, S-parameters[12]. Training ML algorithms with large datasets containing device measurements across a variety of conditions can provide quick and accurate predictions for S-parameters while excluding time-consuming simulations or expensive experimental measurements[22].

Machine learning techniques, such as Artificial Neural Networks, Random Forest, AdaBoost (Base Decision Tree), AdaBoost (Base Random Forest), and Gradient Boosting Regression, are well-known for solving such kind of tasks. These ML models accurately capture complex, nonlinear relationships within the data and successfully predict S-parameters across various operating frequencies and conditions. Also, ML models are easily modifiable with hyperparameter tuning and feature scaling, which can improve their accuracy in predicting, robustness, and ability to properly generalize unseen data[13].

The usage of Machine Learning models has several advantages, which are more efficient prediction of S-parameters and investigating the underlying behavior of GaN HEMTs. As an example, these models can show how different temperatures or bias voltages impact performance across different frequencies. These insights are helpful for engineers in creating better designs and more efficient and optimized RF circuits and high-performance GaN-based devices[14].

Eventually, ML models are considered to be a good alternative to traditional modeling and simulation methods, because they provide more rapid and accurate predictions of S-parameters of GaN HEMT devices. Implementation of Machine Learning techniques will help engineers to enhance not only the performance of GaN HEMTs but also better understand their behavior in high-frequency and high-power systems. This innovation will significantly improve RF and microwave technologies[15].

1.1 Related Works

In recent years, there have been a lot of users of machine learning techniques in RF and microwave signal modeling. Many researchers highlighted the potential of ML algorithms in predicting S-parameters and the behavior of GaN HEMTs[16]. For example, the Artificial Neural Networks (ANN) model is often used for predicting S-parameters and other performance metrics of GaN-based transistors. Moreover, this model is very useful for learning the nonlinear relationships between device parameters and output characteristics, making the ANN model an efficient alternative to outdated methods[17][18].

In addition to ANN, ML models such as Random Forest (RF) and Gradient Boosting Regression (GBR) have been increasingly demanded for device modeling. RF is well known for its ability to be robust in handling complex, high-dimensional data and its ability to avoid overfitting[19]. At the same time, the GBR model is good at making high-accuracy predictions and handling complex nonlinear relationships[20]. AdaBoost is another highly effective ML model that has also been popular among engineers in recent years. AdaBoost is used to improve GaN HEMTs devices' performance with base models such as DT and RF. The AdaBoost uses a series of base models to combine the predictions in order to minimize the probability of the bias and variance and improve the accuracy of the prediction[21].

Despite the fact that these ML techniques are widely used and researched, there is a need for comprehensive studies that effectively compare these machine learning techniques in predicting S-parameters across various frequencies and operating conditions of GaN HEMTs devices. This paper aims to provide a detailed analysis of the performance of five different machine learning models and their potential in predicting S-parameters and behaviour of GaN-based devices.

1.2 Ethical and Professional Responsibilities

- **Ethical Responsibility:**
  Ethical responsibilities should be used when applying ML for the signal modeling of GaN HEMTs. There is high importance to data integrity and privacy. Though the data set that this project is working with is technically rather than personally identifiable, it should not be plagued by data privacy issues when collaborating with other teams or if a team wants to make measurements in a house. Data is very strict and strict as they don't store the data publicly (or not only the public), they store data securely anonymized, and all that obeys the data protection regulation. The other thing is that the dataset is biased. However, if the operating conditions of the GaN HEMT are not trained with all the given data, predictions of GaN HEMTs are inaccurate, and the designed devices may be suboptimal or unsafe. However, in some cases, the ML model is really a black box, and the Engineer bases her/his confidence in the predictions of the models, and such predictions are not aware of the limitations of these models. For example, feature importance analysis and visualization are useful to enhance the model interpretability, such as to present model decisions in a more transparent way. However, this indicates that mainstream engineering roles can be automated in the design processes through the progression of ML. Replacing these technologies, or these tools, should help engineers do these things faster, iterate faster, get out of a single layer of abstraction, and so on. To go through this transition, however, we need to keep up to date on our professional development as well as get AI literacy. Finally, powerful modeling tools are equipped with a potentially dangerous tool in the hands of a user, willing or not. Within the scope of its socio-economy, such innovations are used with the help of it via strictly following professional codes of conduct like the IEEE code in all developmental phases from an ethical point of view.

- **Informed Judgments:**
  This project addressed one of the core needs, which was to make informed and responsible decisions for developing a correct, reliable predictive machine learning model for the S parameters of GaN HEMTs. Every methodological choice from model selection to and including the individual features of my algorithms was a balance between technical performance and practical engineering applicability as motivation. Based on the Decision Tree and Random Forest bases, the model selection has been escalated to a comparative experimentation of five supervised learning models, namely Artificial Neural Network, Random Forest, Gradient Boosting Regressor, and AdaBoost. The choices were not necessarily from a common model, but their results were justified with these evaluation metrics based on mean absolute error,

mean squared error, R^2, and training time. By design, I also paid particular attention to the update of splitting into train and test from a temperature perspective, that is, using conditions outside of the training range to test the models. Furthermore, this also served real-world engineering challenges by testing the models' extrapolation capability. Additionally, when I used RandomizedSearchCV, hyperparameter optimization, it made sure that models participated did at least fairly tune and that comparatively poorer test performance is real (and not arbitrary) deployment capacity. During the modeling process, the importance of interpreting/modeling results was stressed by feature importance and visualizations such as Smith Charts and frequency bias plots. On the way, I decided to include not only technical accuracy but also socio-technical factors of model transparency, generalizability, and handling of ethical data appropriately. The further involvement of these additional considerations guarantees the progress of the work not only in the state of the art but also with the meeting of the responsibilities of the responsible technological development and professional care, which will allow the project to be deployed into the actual world engineering environment.

- **Global Context:**
  ML is important to model GaN HEMT towards making the technology advancement, and always to keep the foundation standing from a multi-dimensional point of view. HEMTs based on GaN are indispensable for high-frequency and high-power applications for the 5G / 6G communication, communications, satellites, radars, electric vehicles, and conversion of renewable energy. This project will allow for the worldwide development of smarter, greener, cheaper technologies, and this project will create accurate scaling models. These models reduce development time and delay in the timeline between experimental validation and circuit simulation, facilitate faster and more efficient time to market, and deliver well in all applications of aerospace, defense, telecommunications, and more. In that sense, it allows democratizing access to the expensive simulation tools, so that ML can support smaller firms and universities, to have the RF system design and energy innovation capability without contributing to any part of the system design process in developing countries. It allows for the means of making widespread adoption of GaN technologies to guarantee global sustainability goals like increased power conversion efficiency, lowered thermal losses. A buffer for the environmental damage from the electronic systems and an aid for the international climate obligations. In addition, these models are also applicable to other conditions, as well as deployed anywhere in the region around the planet. As the broad (engineering and data science) movement strives to make border-crossing collaboration and tech more inclusive, there is unlikely to be a stop for work to integrate these ideas and the ideas involved. This helps in enabling global

efforts to import clean energy and tech for digital equity in the most efficient manner.

- **Economic Impact:**
Machine learning (ML) needs to be implemented to model the GaN HEMT signal due to both short-term and long-term economic reasons. In the short term, these models reduce the development costs by reducing the requirements for prototyping to be repeated and for circuit simulation. The reason for this is that traditional S parameter extraction requires complex equipment, costly measurement campaigns, and expert labor. This is, however, where ML-based models make a good trade-off as ML-based models are focused on shortening the design cycle and, at least, bring companies a much shorter timeline to bring a product to market. GaN HEMTs are long-term used and guarantee that GaN HEMTs and ML tools will be used to increase yields and productivity in industries such as telecommunications, defense, and renewable energy. Furthermore, this project promotes further Data Driven Design platforms development, supports the investment in the digital engineering infrastructure, helping in the acceleration of the trend of starting up organizations in the field of RF design tools. It will also help in job creation in the field of machine learning engineering, RF modeling, and AI-assisted hardware development. However, there are challenges. Firms smaller than some will be put off initial investments into ML frameworks, computing infrastructure, and workforce training. The traditional RF engineer should be trained to transition smoothly by using the data science tools of data and metrics to help those engineers upskill. But it has a long-term positive effect on the economy. Moving from functional to the behavioral level, ML will also help global competitiveness, reduce manufacturing costs, and open new markets for advanced RF technologies. Thus, this project is done to show how ML-based modeling can assist in economic innovation in the electronics industry by enhancing design efficiency and technology performance.

- **Environmental Impact:**
Furthermore, more eco-friendly engineering practices can be achieved due to the benefits of using ML to model GaN HEMTs to an even wider expression. In the construction of RF and microwave components, many physical prototypes are made, and the number of simulations is high, such that a considerable amount of energy, materials, and laboratory resources are expended. With ML models, S-parameters can be predicted well, which helps in minimizing repetitive prototyping, wastages of material, energy consumption, and thermal testing. While this is fast, it does not decrease the amount of emissions associated with manufacturing hardware or fabrication processes. Gan HEMTs feature high energy efficiency and thermal performance, which

makes them suitable for power converters, electric vehicles, 5G infrastructure, and renewable energy systems. The device performance is improved through ML-based modeling of the device, leading to the design of circuits with low power loss and good thermal management, which results in reducing energy consumption in large-scale systems. On any other account, they strengthen the efforts towards decreasing the amount of greenhouse gas emissions in the most relevant substances and industries. More importantly, ML reduces the reliance on expensive clusters to simulate heavy power consumers. What that means is that one can train an ML model once, and then it becomes more energy efficient with use since training a model requires fewer computational resources needed to make a prediction. In general, the data gathered during this project teaches people a green way to do high-frequency electronics. Thus, promoting energy-efficient design and thereby decreasing the environmental exposure, it enhances global sustainability in fields such as telecommunications, automotive, and renewable energy.

- **Societal Impact:**
  Implications of using machine learning within GaN HEMT modeling concern technological accessibility, improvement on energy efficiency, and an improvement on quality of life for society. GaN HEMTs are key to wireless communication, satellite, electric vehicle, and renewable energy, and thus have attracted a variety of research efforts. The modeling accuracy and speed optimization of this project increases potential modeling accuracy and speed on which engineers design more efficient technologies that consume less power, produce less heat, and generally work better—an eco, affordable technology for consumers. GaN HEMT models are accurate and can be used in making a reliable network infrastructure for global and digital inclusion in either the unconnected or underserved areas. Modeling optimization is useful in energy and will deploy efficient inverters for solar panels and electric vehicle chargers, and help obtain clean energy more broadly. Moreover, this project also allows more resource-conscious engineering through the reduction of the need for a trial-and-error prototype. It also extends the reach of what is considered conventional electrical engineering and provides the skill to work in interdisciplinary areas of AI-enabled electronics and means of career. This is especially helpful for young engineers entering a changing industry, but it also promotes workforce development and economic integration. As the ML tools are embedded into the engineering space, more collaboration among academia, industry, and public sectors will occur in order to have easier, lower availability of the future electronic systems. Finally, in this case, it shows how technological innovation is crucial, when appropriately thought through and in the view of both the society to which we belong and the planet on which we all live, but will be better connected,

more sustainable, and more equitable.

Chapter 2

Methodology

This section describes the full methodological framework used to predict S-parameters of the GaN High Electron Mobility Transistors (HEMTs) using machine learning (ML) techniques. The approach covers all steps of the modeling pipeline, from messy data availability processing through model evaluation, and seeks to offer a fast, accurate, and generalizable alternative to traditional RF device characterization techniques.

The main objective of this project is to find the ML model that is the best for predicting the real and imaginary components of all S-parameters ($S_{11}, S_{21}, S_{12}, S_{22}$) based on device input values (V_{GS}, V_{DS}, f, frequency, and temperature). The ability to predict S-parameters is important in RF and microwave engineering because S-parameters describe how signals propagate through and reflect from a device, making it easier for engineers to design GaN HEMTs if they know how input features will result in various S-parameters.

In this research, five different supervised regression models were implemented, assessed, and compared with each other:

1. **Artificial Neural Network (ANN)** – A model that can capture complex non-linear patterns. ANN is a machine learning technique that functions and looks like an artificial human brain.

2. **Random Forest (RF)** – Considered as an ensemble of robust decision trees, each decision tree is trained with a particular set of random noise.

3. **Gradient Boosting Regression (GBR)** – A sequential ensemble method that is optimized in order to gain high accuracy and low mistake percentage.

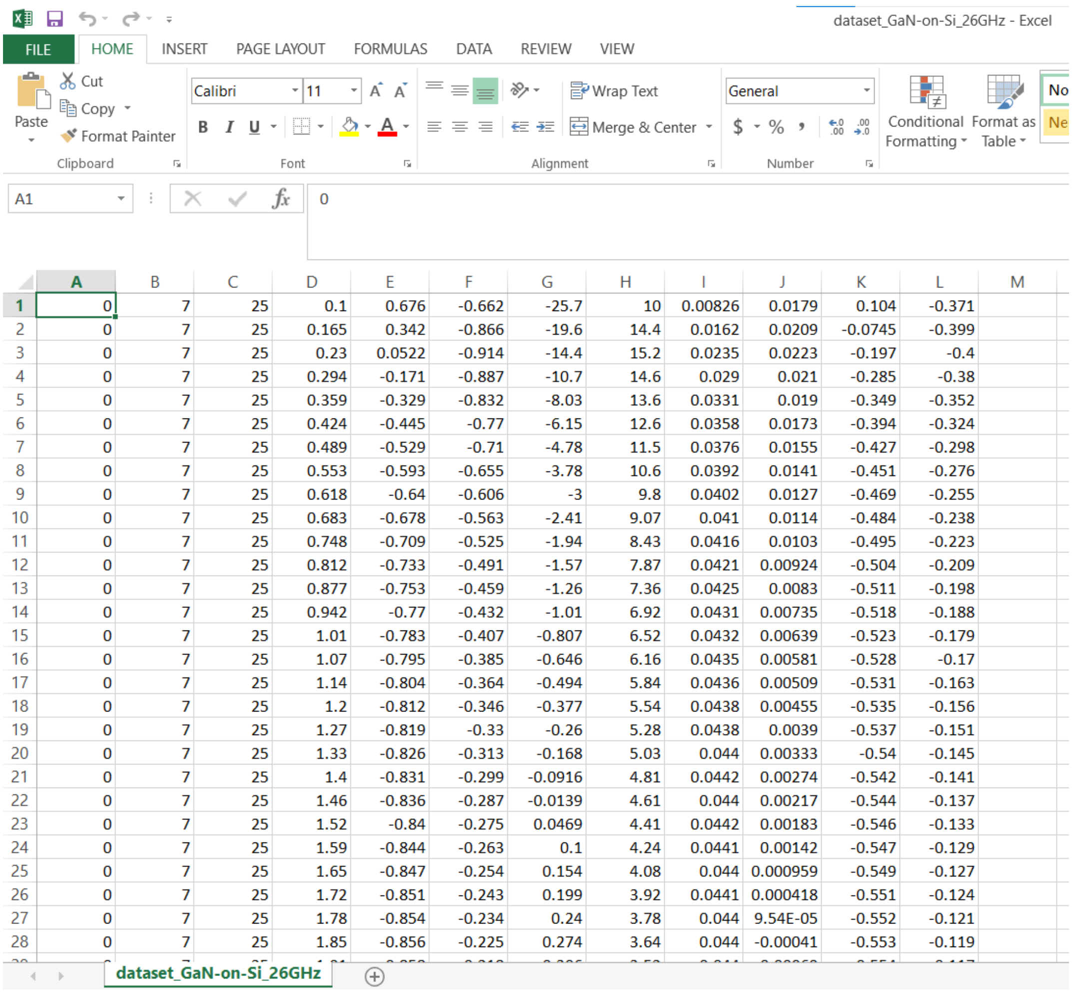
4. **AdaBoost with Decision Tree base (ADA-DT)** – A boosting algorithm which uses simple decision trees as weak learners, in this case, DT was used as a base.

5. **AdaBoost with Random Forest base (ADA-RF)** – A boosting algorithm uses simple decision trees as weak learners, in this case, RF was used as a base.

(AdaBoost creates a sequence of weighted decision trees, typically using shallow trees)

Each of these five models was implemented and trained independently using Python and then evaluated on various unseen temperatures in order to assess interpolation and extrapolation performance. To better understand the results and assess predictive capabilities, I used various visualizations such as Smith Charts, Line Plots, and Frequency bias points plots.

2.1 Dataset Description



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	7	25	0.1	0.676	-0.662	-25.7	10	0.00826	0.0179	0.104	-0.371	
2	0	7	25	0.165	0.342	-0.866	-19.6	14.4	0.0162	0.0209	-0.0745	-0.399	
3	0	7	25	0.23	0.0522	-0.914	-14.4	15.2	0.0235	0.0223	-0.197	-0.4	
4	0	7	25	0.294	-0.171	-0.887	-10.7	14.6	0.029	0.021	-0.285	-0.38	
5	0	7	25	0.359	-0.329	-0.832	-8.03	13.6	0.0331	0.019	-0.349	-0.352	
6	0	7	25	0.424	-0.445	-0.77	-6.15	12.6	0.0358	0.0173	-0.394	-0.324	
7	0	7	25	0.489	-0.529	-0.71	-4.78	11.5	0.0376	0.0155	-0.427	-0.298	
8	0	7	25	0.553	-0.593	-0.655	-3.78	10.6	0.0392	0.0141	-0.451	-0.276	
9	0	7	25	0.618	-0.64	-0.606	-3	9.8	0.0402	0.0127	-0.469	-0.255	
10	0	7	25	0.683	-0.678	-0.563	-2.41	9.07	0.041	0.0114	-0.484	-0.238	
11	0	7	25	0.748	-0.709	-0.525	-1.94	8.43	0.0416	0.0103	-0.495	-0.223	
12	0	7	25	0.812	-0.733	-0.491	-1.57	7.87	0.0421	0.00924	-0.504	-0.209	
13	0	7	25	0.877	-0.753	-0.459	-1.26	7.36	0.0425	0.0083	-0.511	-0.198	
14	0	7	25	0.942	-0.77	-0.432	-1.01	6.92	0.0431	0.00735	-0.518	-0.188	
15	0	7	25	1.01	-0.783	-0.407	-0.807	6.52	0.0432	0.00639	-0.523	-0.179	
16	0	7	25	1.07	-0.795	-0.385	-0.646	6.16	0.0435	0.00581	-0.528	-0.17	
17	0	7	25	1.14	-0.804	-0.364	-0.494	5.84	0.0436	0.00509	-0.531	-0.163	
18	0	7	25	1.2	-0.812	-0.346	-0.377	5.54	0.0438	0.00455	-0.535	-0.156	
19	0	7	25	1.27	-0.819	-0.33	-0.26	5.28	0.0438	0.0039	-0.537	-0.151	
20	0	7	25	1.33	-0.826	-0.313	-0.168	5.03	0.044	0.00333	-0.54	-0.145	
21	0	7	25	1.4	-0.831	-0.299	-0.0916	4.81	0.0442	0.00274	-0.542	-0.141	
22	0	7	25	1.46	-0.836	-0.287	-0.0139	4.61	0.044	0.00217	-0.544	-0.137	
23	0	7	25	1.52	-0.84	-0.275	0.0469	4.41	0.0442	0.00183	-0.546	-0.133	
24	0	7	25	1.59	-0.844	-0.263	0.1	4.24	0.0441	0.00142	-0.547	-0.129	
25	0	7	25	1.65	-0.847	-0.254	0.154	4.08	0.044	0.000959	-0.549	-0.127	
26	0	7	25	1.72	-0.851	-0.243	0.199	3.92	0.0441	0.000418	-0.551	-0.124	
27	0	7	25	1.78	-0.854	-0.234	0.24	3.78	0.044	9.54E-05	-0.552	-0.121	
28	0	7	25	1.85	-0.856	-0.225	0.274	3.64	0.044	-0.00041	-0.553	-0.119	

Figure 2.1: Used Dataset (dataset_GaN-on-Si_26GHz).

For this project, a dataset consisting of 17,280 samples of measurements of frequency, temperature, and various voltages for a GaN-on-Silicon High Electron Mobility Transistor that operates at a fixed 26 GHz frequency was used to train the ML models. The dataset was very important for training the models as it provided prepared actual values of S-parameters based on various input parameters. Each S-parameter's value of a device was a unique measurement that was done in practice using a combination of biasing conditions. While S-parameters themselves are complex values that describe the behavior of high-frequency signals that pass through or reflected by the transistor.

The used dataset contains four input features which describe the operating conditions: Gate-Source Voltage (V_{GS}), Drain-Source Voltage (V_{DS}), Temperature (T), Frequency (f). These features define the state under which the transistor's S-parameters are measured.

The target values in the dataset are the real and imaginary parts of the four complex S-parameters, which describe the behavior of the device: S_{11} (Input reflection coefficient), S_{21} (Forward transmission coefficient or gain), S_{12} (Reverse transmission coefficient or isolation), S_{22} (Output reflection coefficient). Each S-parameter is divided into its real and imaginary components, a total of eight outputs: Real & Imag S11, Real & Imag S21, Real & Imag S12, Real & Imag S22.

These eight output targets are the prediction objectives for the machine learning models. They predict both magnitude and phase information with the real and imaginary components, which helps in modeling the needed RF behavior of the transistor.

Also, the used dataset has a wide range of biasing conditions and temperatures (from 25°C to 175°C), which provide a robust and accurate training process for ML models to generalize to unseen data with the help of nonlinear dependencies.

2.2 Data Preprocessing

In order to create an accurate model training process and generalization to new conditions, several preprocessing steps were used. These steps were crucial to prepare the data to deal with high-dimensional, nonlinear values as S-parameters.

2.2.1 Missing Values Check

A full column-wise check of missing values was done using Pandas first. After doing that, the dataset is complete as the input and output variables do not have any missing values at all. It helped in giving us a smooth training procedure when we carried out our ML models.

Outliers were identified based on the Z-score approach (i.e., it computes the standard deviation of each data point from the mean). If the Z score (sample

Z/standard deviation) of a data point was greater than some threshold (Z > 3), such a data point was not used.

$$Z = \frac{x - \mu}{\sigma}$$

where:

- x: Data point.
- μ: The feature's mean value.
- σ: The feature's standard deviation value.

2.2.2 Temperature-Based Train/Test Split

Part of the data was split up into training and testing sets using temperature for real-world cases and to evaluate how well the models can generalize to new data. The ability of an ML model to generalize to unseen data is one of the most influential factors in predicting a GaN HEMTs behavior. The split also adds an extrapolation challenge to the regression problem, making it more robust after training.

After a lot of trials, the best split was to create a training set that included temperatures 25, 50, 75, 100, and 150 °C. These temperatures were used to train and tune ML models. The testing set included 125°C and 175°C. These temperatures were added to the testing set to implement the ability to extrapolate beyond known conditions.

This separation creates the practical conditions where a model trained in a range of laboratory temperatures must accurately predict performance under extreme field conditions.

2.2.3 Feature Scaling with MaxAbsScaler

Such ML models as ANN and GBR take advantage of normalized and scaled inputs. Therefore, to improve the training process, all features and target values were scaled using the MaxAbsScaler command. The input features (V_{GS}, V_{DS}, T, f) were scaled from -1 to 1 based on their maximum absolute values in the training set. While the eight real and imaginary S-parameter components were scaled independently using the same method.

A feature's scaled value, X_{scaled}, is calculated as follows:

$$X_{\text{scaled}} = 2 \times \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} - 1$$

MaxAbsScaler was chosen because it keeps values centered at zero and preserves relative physical quantities without distorting the shape of the data, which is important for some S-parameters. Also, this is helpful for some features, as it will ensure numerical stability during the training process in models.

To conclude, the preprocessing steps were done to create a clear and ML-friendly dataset that will allow for effective training and testing the performance of GaN HEMT devices with various ML models.

2.3 Machine Learning Models

The purpose of this thesis is to develop ML models that will be able to predict the S-parameters (real and imaginary components) of GaN HEMTs at different bias conditions and temperatures. Five various models were chosen for this work, each was trained independently for all eight Scattering parameters. To ensure successful performance comparison analysis, those five models were selected in terms of covering a wide range of learning criteria, including deep learning and ensemble-based techniques. The ability to predict S-parameters was assessed in terms of accuracy, robustness across various temperatures, and computational efficiency.

2.3.1 Artificial Neural Network (ANN)

A Multi-layer Perceptron regressor was successfully implemented as the ANN model that has been proven to be effective by its ability to approximate highly nonlinear functions. The config of this ML model used two hidden layers (128 neurons in the first layer and 64 in the second), activation function (Rectified Linear Unit which provides non linearity so that we can get the process of convergence fast) and optimizer (Adaptive Moment Estimation was chosen since having sparse gradients and adaptive learning rates), and the training stopped at 1000 epochs.

Both feature interactivity and complex, multivariate relationships that do not depend on feature interactivity were modeled by the ANN model due to its flexibility.

$$\hat{S}(V_{GS}, V_{DS}, f, T) = \text{ANN}([V_{GS}, V_{DS}, f, T])$$

Where:

- $\hat{S}$ is the predicted S-parameter.
- Artificial Neural Network (ANN) is a model of Artificial Neural Network trained to learn the relation between the input features and output S-parameters.

2.3.2 Random Forest (RF)

Random Forest is a tree-based ensemble model that employs bootstrapping aggregation or bagging, and would further increase generalization by using it. There is a random portion of the data and input features used for training each decision tree individually. The average of all trained trees is the final output. In the architecture of this model, I have used the number of trees(n_estimators), the maximum depth of trees (max_depth), the minimum samples per split and leaf, as well as the feature selection strategy (sqrt, log2, and None) as hyperparameters to be tuned by RandomizedSearchCV.

The RF model is especially useful and effective when capturing nonlinear dependencies and variable interactions in tabular data. The built-in feature importance score also contributes to interpretability, making this model robust and almost perfect for solving this regression task.

$$\hat{S}(x) = \frac{1}{N} \sum_{i=1}^{N} h_i(x)$$

Where:

- $\hat{S}$ is the predicted S-parameter.
- Input vector $x = [V_{GS}, V_{DS}, f, T]$.
- h_i is the prediction from the i-th decision tree.
- N is the number of trees in the forest.

2.3.3 Gradient Boosting Regressor (GBR)

Gradient Boosting Regressor is a tree-based model that creates trees sequentially. The new trees are made to correct the prediction errors of the old ones in order to make each subsequent tree more effective. However, in case the dataset is not well 'structured' the boosting technique makes the model effective and accurate. In the architecture of this model, I tuned the learning rate (feature selection), number of estimators (trees), max_depth (tree depth), as well as (subsample) ratio of examples to use each time a tree is created as the hyperparameters using RandomizedSearchCV.

The GBR model is frequently used in solving such regression problems, which involve complex relationships and tabular data, because the model has a strong regularization capability and resistance to overfitting when tuned correctly.

$$\hat{S}(x) = \sum_{m=1}^{M} \gamma_m \times h_m(x)$$

Where:

- $\hat{S}$ is the predicted S-parameter.

- Input vector $x = [V_{GS}, V_{DS}, f, T]$.
- γ_m is the learning rate.
- h_m is the prediction from the m-th weak learner (a regression tree).

2.3.4 AdaBoost with Decision Tree (ADA-DT)

The AdaBoost (Adaptive Boosting), from its name, combines several weak learners as bases to create a strong model for predictions. In this case Decision Trees model was used as a base weak learner. This model allows the new trees to focus on the data points that were previously mispredicted by old trees. In the architecture of this model, I used the number of estimators (n_estimators), the learning rate (to control the weight update magnitude), the tree depth of base learners (max_depth), and the minimum samples for split and leaf nodes hyperparameters that were tuned using RandomizedSearchCV.

The AdaBoost model based on Decision Trees was selected for its high efficiency and interpretability. The technique where each of the trees is a weak model itself, while the iterative reweighting and boosting enhance performance on difficult-to-fit patterns in the data.

$$\hat{S}(x) = \sum_{m=1}^{M} \alpha_m \times h_m(x)$$

Where:

- $\hat{S}$ is the predicted S-parameter.
- Input vector $x = [V_{GS}, V_{DS}, f, T]$.
- α_m reflects how well each tree performs (higher weight for better-performing trees).
- h_m is a weak regression tree trained in sequence.

2.3.5 AdaBoost with Random Forest (ADA-RF)

The AdaBoost (Adaptive Boosting) model in this case uses Random Forests as a weak learner or base estimator. This model was selected with the intention to combine the strengths of bagging from the RF model and boosting abilities from the AdaBoost model. The idea was to allow a base learner to be more powerful while still benefiting from AdaBoost's focus on correcting errors made by previous trees. In the architecture of this model, I used the number of boosting rounds, learning rate, and random Forest-specific parameters (depth, splits, and number of trees in the base RF)

The AdaBoost (base RF) model is more computationally intensive on the one hand, but on the other side is a very powerful ML model for solving problems where standard boosting may end up inefficient due to the weakness of its base learners.

$$\hat{S}(x) = \sum_{m=1}^{M} \alpha_m \times h_m(x)$$

Where:

- $\hat{S}$ is the predicted S-parameter.
- Input vector $x = [V_{GS}, V_{DS}, f, T]$.
- α_m weight for the m-th base learner (which is learned during boosting).
- h_m prediction from the m-th Random Forest.

2.4 Model Evaluation and Metric

I chose to use highly comprehensive evaluation metrics on 5 of the different ML models to ensure that I make a clear and fair comparison of model performance in regards to the used model itself.

1. **Mean Absolute Error (MAE)**: This is a metric that indicates the average prediction error, but not orientation. It directly represents the difference between the forecasted and actual values.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

2. **Mean Squared Error (MSE)**: It utilizes this value as a measure for the average squared difference between the measured and predicted values. MSE is a good way to examine large discrepancies between the actual and predicted values.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3. **R-squared (R^2)**: This metric represents how good the model is at predicting S parameters from frequency, temperature, and voltage values. It's a number from 0 to 1, such that 1 represents a perfect fit (100%) the model and the closer the result gets to 0 means the worse the fit of the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

4) **Training Time (seconds)**:

This metric shows how fast a model trains in seconds. The training time is a crucial aspect in ML modeling because it can show the effectiveness of the model (the faster it trains, the better).

Each model had training (25, 50, 75, 100, and 150°C) and testing (125°C and 175°C) sets. This tactic provided a robust extrapolation challenge and created real-world conditions for each model.

Next, scaling of the outputs was performed, and then all metrics for consistent training were computed. The predicted values were scaled back to visualize the results through Smith charts and Line plots in order to compare differences between the predicted and actual S parameter values.

These metrics, on the whole, served to better understand and assess the results of each model. Five models were compared in terms of accuracy, generalization capacity, and computational requirements and therefore MAE, MSE, R^2 and training time were computed. The models needed these metrics to assess their performance and compare them against each other.

2.5 Feature Importance Analysis

The feature importance values and plots were computed for every model except the ANN model (not a tree-based model). This metric helped in assessing how much importance every input feature has on the model's ability to predict S-parameters. After collecting all data, the frequency showed the greatest portion of importance on every model, while the importance of V_{GS} and V_{DS} values were not significant.

2.6 Visualization Techniques

The visualizations as Line plots, Biased frequency points plots, and Smith charts, were crucial in evaluating and interpreting the performance results of each ML model. Although numerical metrics give exact numbers, the visualization part provides visual analysis tools to better understand the results. They provided insights into the accuracy of predictions and the generalization behavior of predicted S-parameters.

2.6.1 Line Plots

The line plots were created in order to compare MAE, MSE, and R^2 metrics of all eight S-parameters of each model.

The plots were created for both the training and the testing sets. This visualization technique helped to evaluate the model's ability to solve the problems of overfitting and underfitting, showing the strength of each ML model. Then, with the collected data for every model, we can compare the effectiveness of each model in predicting S-parameters.

By plotting all models' numerical metrics results together, we can clearly compare them and highlight the best ones. These plots offer quick insight into where certain models outperformed other models.

2.6.2 Biased Frequency Points Plots

Frequency-dependent bias points plots of predicted vs. actual values were created for each S-parameter in every model to assess model performance under different biased conditions ($V_{GS} = 0V$ and $V_{GS} = 0.2V$).

The real and imaginary components for S-parameters were plotted against frequency at those biased points. This visualization technique helped to understand how accurately the models captured the inherent frequency-dependent behavior. Also, the plots were essential because frequency significantly impacts the behavior of the devices, especially in RF modeling, as we understood after feature importance analysis previously.

Eventually, these plots showed the predictive behavior of each model and allowed for checking possible phase alignments, amplitude variations, and frequency distortions of the model's training process.

2.6.3 Smith Chart Plots

Smith Charts were generated using the scikit-rf library, based on actual and predicted output values to fully understand the complex nature of S-parameter predictions. The plots helped to see the predicted and actual S-parameters values in the complex impedance domain. This technique is frequently used by engineers in RF modeling.

The Smith charts were generated for each model based on two operating conditions: High temperature ($V_{GS} = -0.6V, V_{DS} = 7V, T = 175°C$) and Low temperature ($V_{GS} = -1.6V, V_{DS} = 7V, T = 125°C$). Also, scaling of the output values was applied for each model to better visualize the results and avoid overlapping in the chart space.

The predicted and actual trajectories across frequencies were plotted across all S-parameters (S11, S21, S12, S22) for each model. Where solid lines represented actual values and dashed lines represented predicted values. The plots were especially useful to see an intuitive physical validation, which allows a quick detection of mismatches in reflection, transmission, and coupling behavior of the device.

Overall, the visualization techniques as line plots, bias point frequency plots, and the Smith charts, helped to create a comprehensive evaluation of models' performance and better compare the ability to predict the S-parameters of GaN HEMTs of each model.

Chapter 3

Results and Discussions

In this section, the performance results of five selected ML models (Random Forest, Gradient Boosting Regressor, AdaBoost (Base Decision Tree), AdaBoost (Base Random Forest), and Artificial Neural Network) used for predicting Scattering Parameters of GaN HEMT devices will be evaluated. The assessment was conducted with the help of such metrics as Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared values (R^2), and visualizations including line plots for the S-parameters, Smith Charts, and bias point analysis.

3.1 Model Performance Metrics.

Three numerical metrics were applied to a separate model performance evaluation (MAE, MSE, and R^2).

1. **Mean Absolute Error (MAE) and Mean Squared Error (MSE)** However, these metrics are important for the prediction power of each selected model for this work. The current MAE measure was written to be the average magnitude of the errors in the predictions, not the direction. The MSE is the average of the squares of the differences between actual and predicted values. The lower the values of both MAE and MSE represent the better ability of a model to effectively predict our output values, S parameters.

   Execution of training and parameter tuning on each model separately resulted in getting the results, and thus it can be concluded that the RF model can minimize Mean Absolute Error and Mean Squared Error for both testing and training sets. Clearly, the GBR model was the second best model and performed well too. The models did better, and there were few errors in predicting S parameters when compared to other models. However, the worst performance, according to the MAE and MSE results, was for the ANN model since it had poor performance on all S-parameters.

2. **R² Value (R^2)** The metric was helpful to see the ability of models to predict S-parameters and showed the percentage of efficiency in this prediction (from 0 to 1).

   Here, almost all models showed near-perfect performance, indicating more than 0.95 R^2 Value. However, comparing all models, RF and GBR models were the best models based on this metric again, indicating about 0.99 R^2 results. This shows that the RF and GBR models outperformed others when looking at our performance, assessing numerical metrics.

   Also, it is important to note that the ANN model showed the lowest score again, fluctuating from good to poor results across all S-parameters. This suggests that the ANN model can be stated as the weakest among all selected ML models.

3. **Training Time** These results were also important to measure in order to understand the ability of each model to effectively combine time for implementing the model on the data and efficiently performing predictions.

   The obtained results showed that the fastest learning ML model was the ANN model, ranging from 7.61s (for Imag S22) to 10.14s (for Real S21). While both the RF and GBR models' training time varied from 5s to 55s, when a bit faster training process was conducted while implementing the RF model. On the other hand, AdaBoosting models for both base estimators, RF and DT, showed a significant amount of time to train and evaluate the results. The training time varied from 25s to 290s.

3.2 Visualizations: Plots of Model Performance.

3.2.1 Line Plots of the Numerical Metrics for All Models.

These plots directly compared the MAE, MSE, and R^2 values results for both training and testing sets of the five selected models.

From the obtained plots, we can notice:

- RF and GBR models outperformed other models and showed the best results for each computed metric. Indicating that for our task to predict S-parameters, the models are the best to choose.
- AdaBoost (base RF) and AdaBoost (base DT), although not performing the best, presented good results and confirmed the ability to make predictions successfully.
- ANN model performed the worst, showing poor results in all metrics, especially in R-squared values. However, in some of MSE and MAE results the ANN model was at the same level of predicting as the other models.

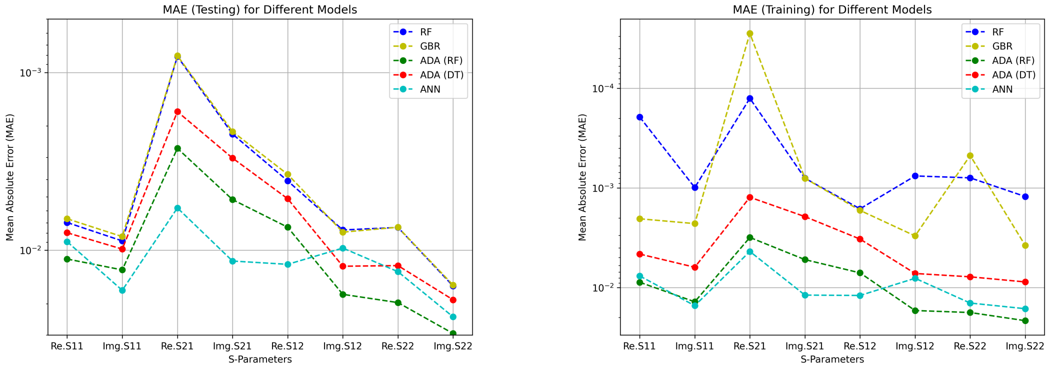


Figure 3.1: MAE results' comparison plots across five selected models.

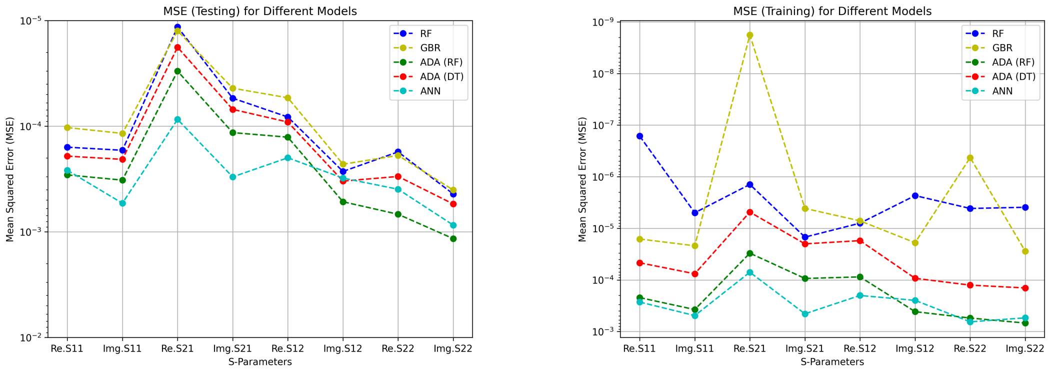


Figure 3.2: MSE results' comparison plots across five selected models.

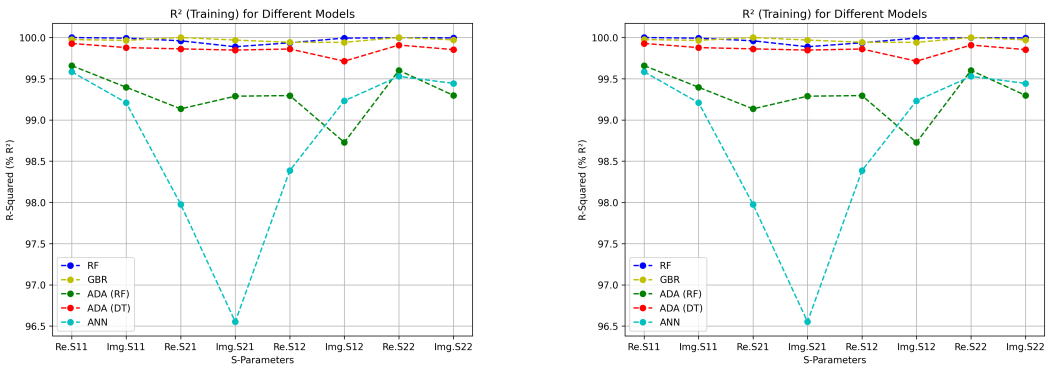


Figure 3.3: R^2 results' comparison plots across five selected models.

3.2.2 Smith Chart plots.

The Smith charts may be created for every model for the visualization of the impedance characteristic of models and for the analysis of the accuracy of the prediction, as the actual and predicted values of S-parameters are depicted in the chart space. It helped better assess the ML performance as all the deviations of the predicted S-parameters were seen.

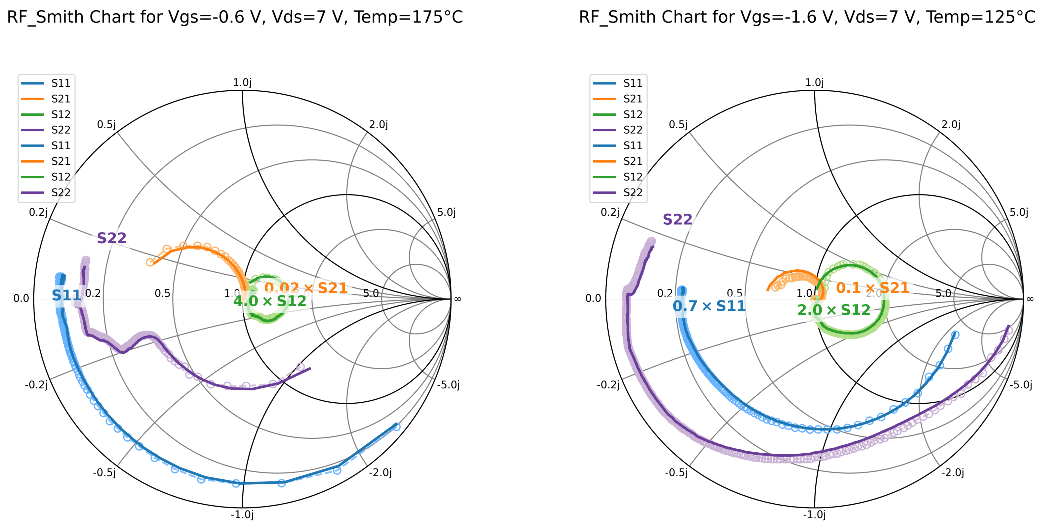


Figure 3.4: The RF model based on two different operating conditions, as Smith Charts.

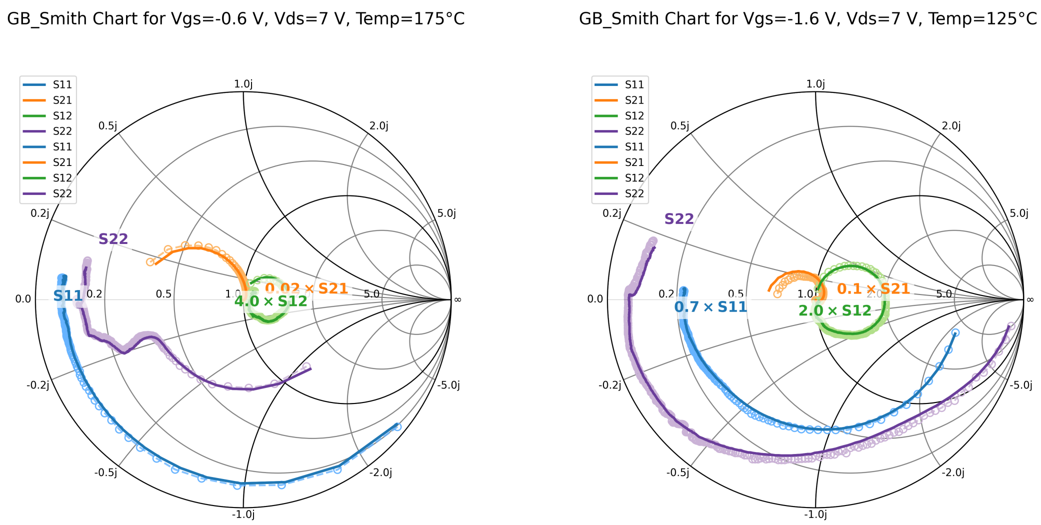


Figure 3.5: The GBR model based on two different operating conditions, as Smith Charts.

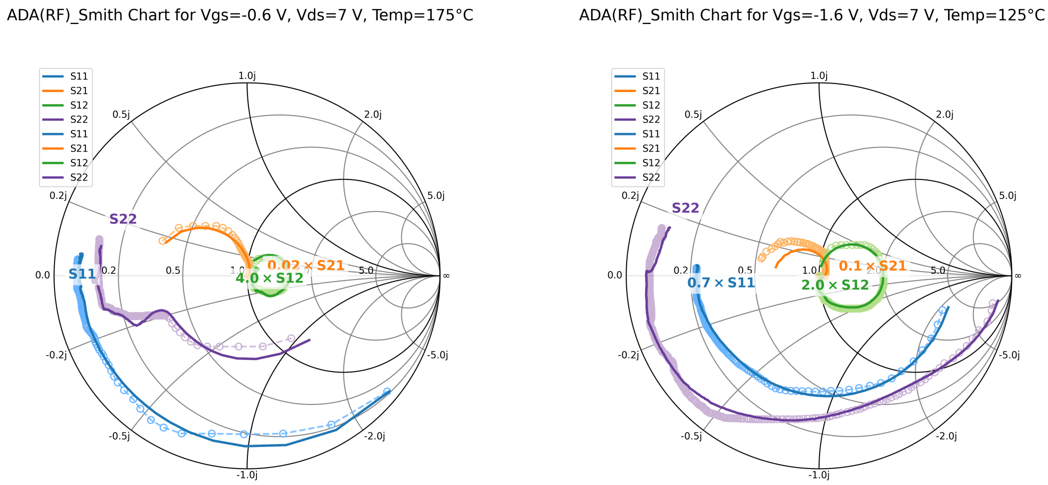


Figure 3.6: The AdaBoost (Base RF) model based on two different operating conditions, as Smith Charts.

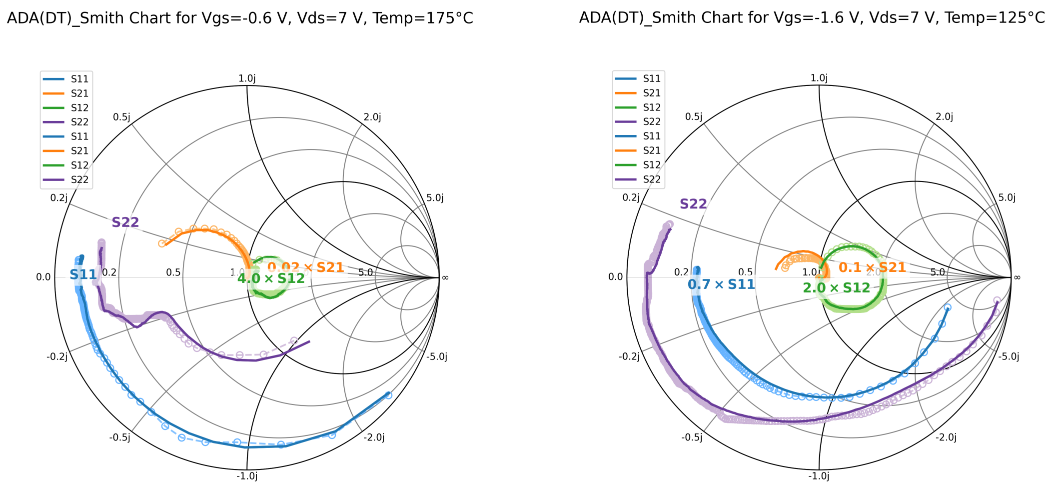


Figure 3.7: The AdaBoost (Base DT) model based on two different operating conditions, as Smith Charts.

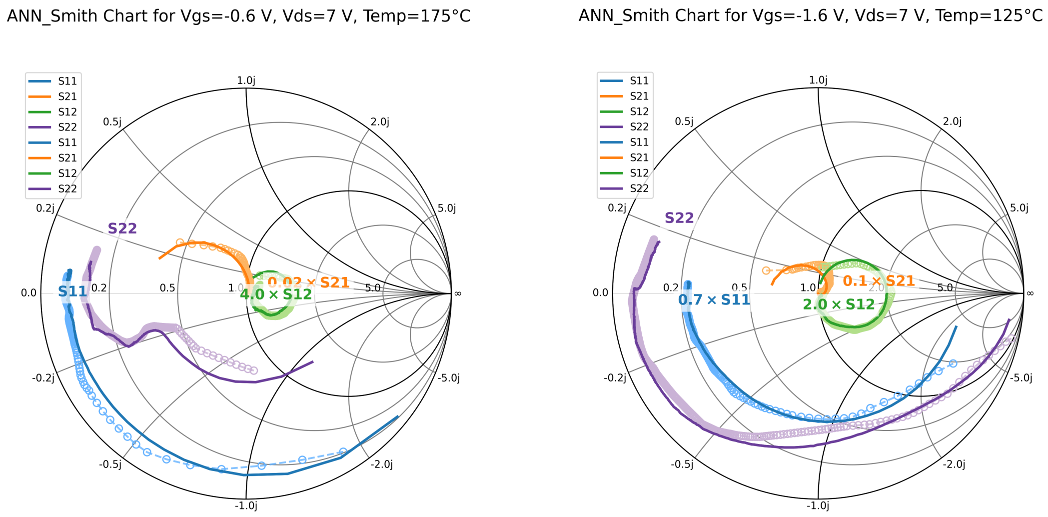


Figure 3.8: The ANN model based on two different operating conditions, as Smith Charts.

From the obtained plots, we can notice:

- The RF and GBR models' predicted values matched the trajectories of actual S-parameters almost ideally, especially for S11 and S12 (capturing the impedance behaviour).
- The AdaBoost (Base RF) and AdaBoost (Base DT) illustrated slightly more visible deviations from the actual values of the S-parameters trajectory, especially for S21 and S22. However, the deviations are not substantial and confirm that these models can be used in practice.
- The ANN model showed the weakest match, indicating the worst performance results across all S-parameters. Particularly, deviations from predicted values trajectory for high frequencies indicate the poor modeling of the impedance characteristics.

3.2.3 Biased Frequency Points Plots.

The purpose of biased frequency points plots is to present how good each model is at predicting output S-parameters at $V_{GS} = 0$ and $V_{GS} = 0.2$. Assessing how well a model can predict, in particular under biased conditions, can be done from the plots.

For every model, biased frequency points plots were generated for both imaginary and real S-parameters (8 parameters) at $V_{GS} = 0$ and $V_{GS} = 0.2$. The number of these plots is sixteen, but here, the plots for Imaginary S11 at $V_{GS} = 0.2$ will only be illustrated to show the deviations from the ideal match.

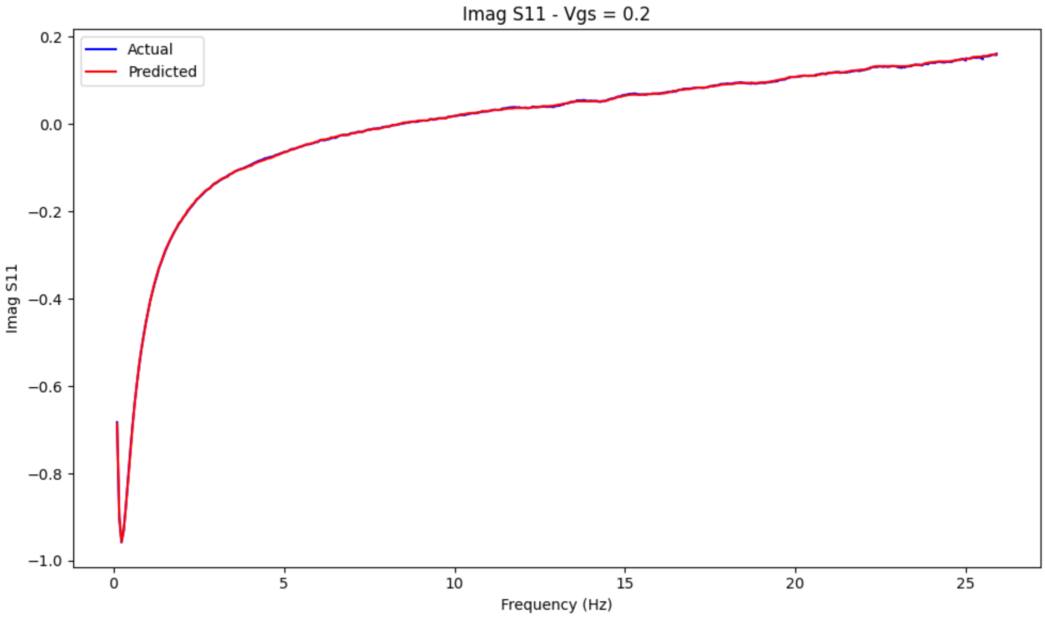


Figure 3.9: Biased Frequency Points Plot of Imaginary S11 for the GBR model at $V_{GS} = 0.2$.

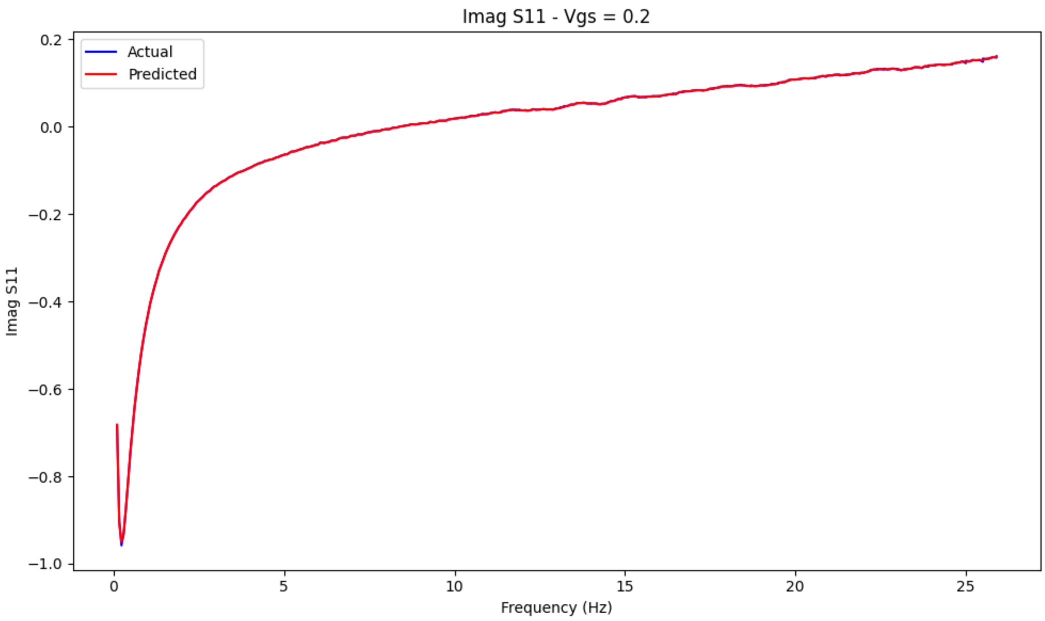


Figure 3.10: Biased Frequency Points Plot of Imaginary S11 for the RF model at $V_{GS} = 0.2$.

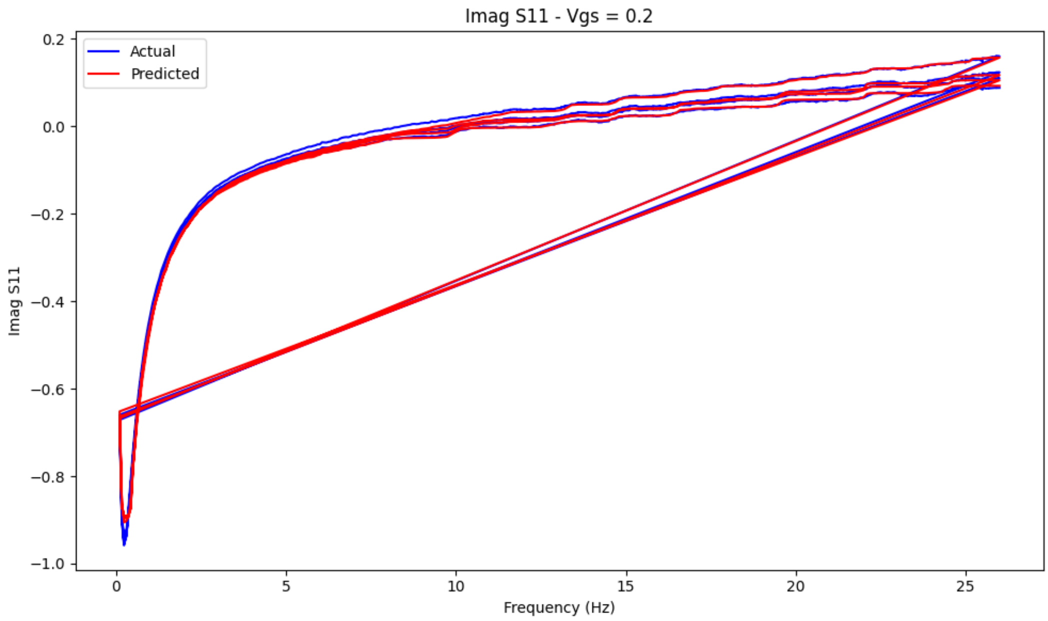


Figure 3.11: Biased Frequency Points Plot of Imaginary S11 for the AdaBoos (Base DT) model at $V_{GS} = 0.2$.

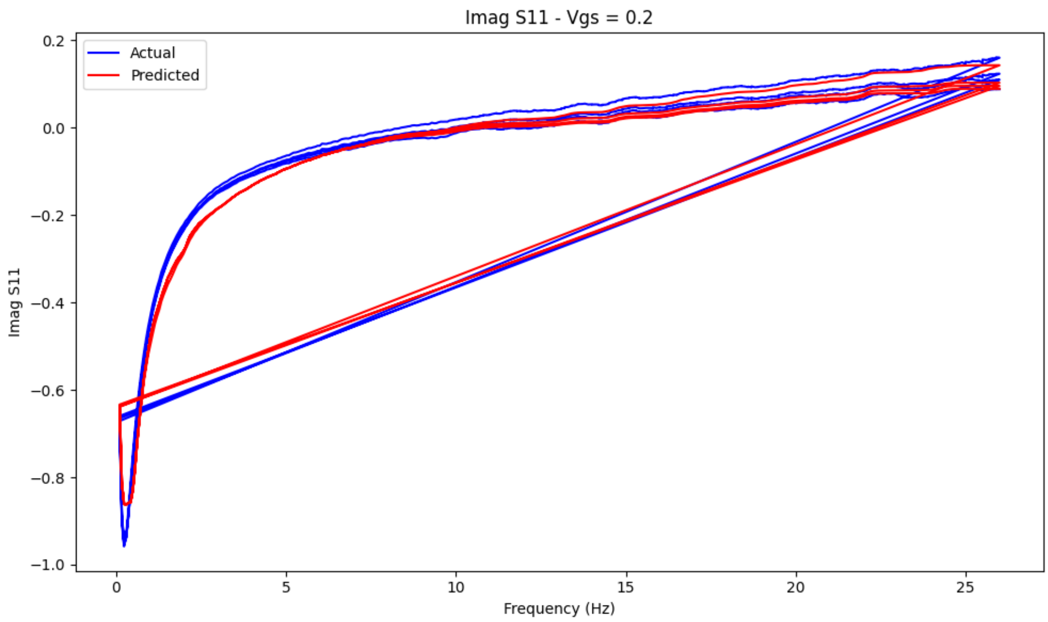


Figure 3.12: Biased Frequency Points Plot of Imaginary S11 for the AdaBoos (Base RF) model at $V_{GS} = 0.2$.

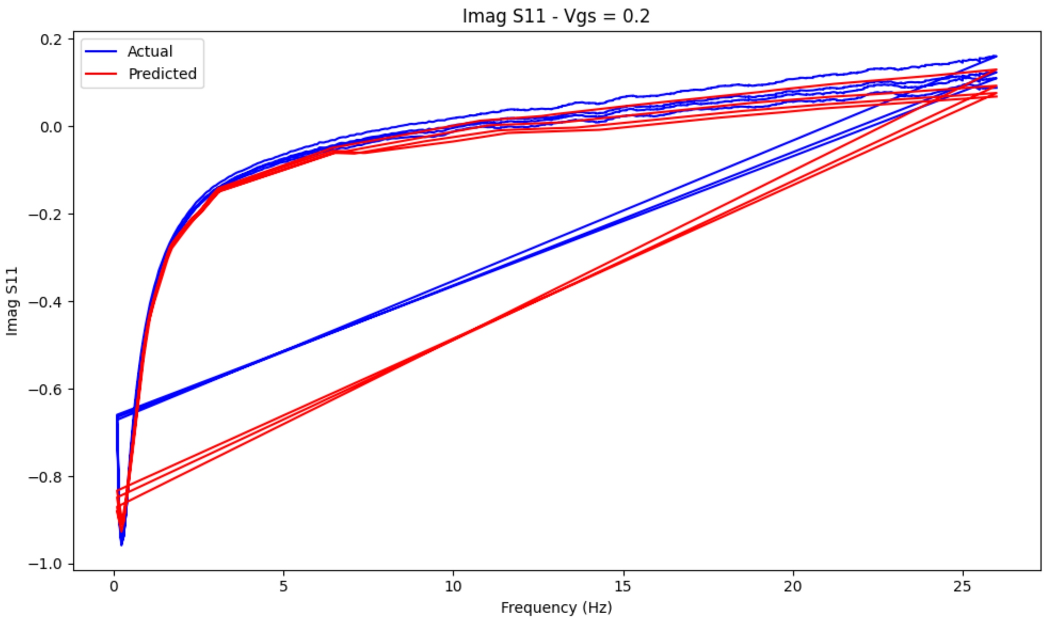


Figure 3.13: Biased Frequency Points Plot of Imaginary S11 for the ANN model at $V_{GS} = 0.2$.

From the obtained plots, we can notice:

- The RF and GBR models illustrated almost perfect performance, matching the actual S-parameter values under biased conditions. When comparing, the RF model showed slightly better performance.
- The AdaBoost (Base RF) and AdaBoost (Base DT) showed more errors in predicted vs actual S-parameter values lines, indicating less ability in predictions compared to the RF and GBR ML models.
- The ANN model was the worst one again, illustrating considerable deviations. That shows low precision in predicting the operational points of the GaN HEMTs.

3.3 Overall Model Performance

We assess the prediction accuracy of five machine learning algorithms over S-parameter data from GaN HEMT devices, including Random Forest and Gradient Boosting Regressor, as well as the AdaBoost system using Base Decision Tree and Base Random Forest in addition to Artificial Neural Network (ANN).

The developed models were tested under three evaluation metrics (Mean Absolute Error (MAE), Mean Squared Error (MSE) and R-squared (R^2)) and were found to combine them together. Preferably, success was achieved by the Random Forest and Gradient Boosting Regressor because they returned the lowest error rate with near-perfect R^2 values. These predictive models had few inconsistencies between forecasted and actual data points and provided better results than the rival models, which meant the generalization capability of these models was good.

With Decision Tree (ADA-DT) or Random Forest (ADA-RF) base estimators, AdaBoost gave good results with only additional errors in predicted S21 and S22 S-parameters. Despite slight differences in actual data points and the resulting predicted values, the predicted ones are accurate results of AdaBoost. This indicates that although AdaBoost is a good method for this evaluation, it is not as strong as RF or GBR.

The ANN model's training was very quick (7-10s), but its performance was the weakest of all the analyzed metrics. This model was difficult to execute because of the need for accurate nonlinear modeling relations to make accurate S-parameter forecasting. Use of high-frequency data by ANN yielded large deviations between simulated and measured S parameters and consequently gave reduced efficiency for predicting GaN HEMT impedance characteristics.

3.4 Best Application of Models

The most suitable models to predict the S-parameters of the GaN HEMT devices are the Random Forest and Gradient Boosting Regressor models. These models performed at diverse temperature and frequency ranges and resulted in error prediction errors remaining minimized and R^2 with a very good value. These models are essentially perfect for practical use in RF device models and optimization tasks and their exceptional execution makes them the choice of the day.

The RF model achieves high prediction accuracy and clear interpretability in which frequency measurement turns out to be the key predictor variable of S-parameters. The GBR model stands out because it does sequential ensemble in improving the accuracy by increasing the accuracy during each iteration, making it a great solution for demanding RF and microwave purposes.

However, in the other case, models based on AdaBoost (in particular, the base Random Forest variant) can be thought of as something more suitable in their application when training time or computational efficiency is slightly more important. Their model strikes a good balance between prediction accuracy and computational cost, though it is outperformed by RF and GBR models when predicting overall together.

Finally, although the ANN model was not intended to provide high prediction accuracy, it still has potential in some applications where fast training and a simple but less complex model is good enough, but it would need additional improvement for more demanding applications.

Chapter 4

Conclusion

Finally, this project investigated the potential of five machine learning models in the prediction of the S-parameters from the GaN HEMT devices to aid in the design and optimization of RF and microwave systems. In the process of rigorous testing and comparison, we found that Random Forest (RF) and Gradient Boosting Regressor (GBR) exhibited good results of accurate and reliable prediction in multiple operating conditions. The RF and GBR models successfully show that machine learning indeed could be a powerful tool in overcoming the limitations of traditional simulation-based methods, leading to a faster, more efficient, and cost-effective approach to characterize and optimize high-performance devices such as GaN HEMTs. Furthermore, feature importance analysis revealed the importance of frequency in explaining S parameters and implications for the engineer to improve the device design. However, in some cases, we can even replace these with ADABoost models or ANN, but not in the combination, as it was working poorly in comparison to RF and GBR. In order to improve these models, future works need to focus on the tuning process and find some hybrid techniques to improve the performance of the models and use them when facing complex conditions. Finally, in the context of this study, the potential of machine learning to disrupt the field of RF device characterization is the ability to provide accurate yet highly scalable tools for large-scale systems.

Bibliography

[1] Naeemul Islam et al. “Reliability, applications and challenges of GaN HEMT technology for modern power devices: A review”. In: *Crystals* 12.11 (2022), p. 1581.

[2] Chao-Tsung Ma and Zhen-Huang Gu. “Review of GaN HEMT applications in power converters over 500 W”. In: *Electronics* 8.12 (2019), p. 1401.

[3] Umesh K Mishra, Primit Parikh, and Yi-Feng Wu. “AlGaN/GaN HEMTs-an overview of device operation and applications”. In: *Proceedings of the IEEE* 90.6 (2002), pp. 1022–1031.

[4] B Mounika et al. “Recent developments in materials, architectures and processing of AlGaN/GaN HEMTs for future RF and power electronic applications: A critical review”. In: *Micro and Nanostructures* 168 (2022), p. 207317.

[5] Jan Kuzmik et al. “InAlN/GaN HEMTs: A first insight into technological optimization”. In: *IEEE transactions on electron devices* 53.3 (2006), pp. 422–426.

[6] Sadhana Singh, Tarun Chaudhary, and Gargi Khanna. “Recent advancements in wide band semiconductors (SiC and GaN) technology for future devices”. In: *Silicon* 14.11 (2022), pp. 5793–5800.

[7] Sharjeel Anjum, Hyeong-Yeon Lee, and Hong-Kyun Noh. “Rotation error detection of gallium nitride (GaN) substrate in MBE utilizing ensemble learning”. In: *Crystal Growth & Design* 23.6 (2023), pp. 4138–4146.

[8] Bo Liu and Jialin Cai. “Comparative analysis of nonlinear behavioral models for GaN HEMTs based on machine learning techniques”. In: *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* 37.2 (2024), e3177.

[9] Shivansh Awasthi et al. “Comparative Analysis of TCAD augmented ML Algorithms in modeling of AlGaN/GaN HEMTs”. In: (2023), pp. 1–5.

[10] Saddam Husain, Mohammad Hashmi, and Fadhel M Ghannouchi. “Comprehensive investigation and comparative analysis of machine learning-based small-signal modelling techniques for GaN HEMTs”. In: *IEEE Journal of the Electron Devices Society* 10 (2022), pp. 1015–1032.

[11] Ahmad Khusro et al. “Small signal behavioral modeling technique of GaN high electron mobility transistor using artificial neural network: An accurate, fast, and reliable approach”. In: *International Journal of RF and Microwave Computer-Aided Engineering* 30.4 (2020), e22112.

[12] Saddam Husain et al. “Comprehensive investigation of ANN algorithms implemented in MATLAB, python, and R for small-signal behavioral modeling of GaN HEMTs”. In: *IEEE Journal of the Electron Devices Society* 11 (2023), pp. 559–572.

[13] Saddam Husain et al. “Decision tree based small-signal modelling of GaN HEMT and CAD implementation”. In: (2022), pp. 1–6.

[14] Upma Singh et al. “A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments”. In: *Energies* 14.16 (2021), p. 5196.

[15] Germán González and Conor L Evans. “Biomedical Image Processing with Containers and Deep Learning: An Automated Analysis Pipeline: Data architecture, artificial intelligence, automated processing, containerization, and clusters orchestration ease the transition from data acquisition to insights in medium-to-large datasets”. In: *BioEssays* 41.6 (2019), p. 1900004.

[16] Michael I Jordan and Tom M Mitchell. “Machine learning: Trends, perspectives, and prospects”. In: *Science* 349.6245 (2015), pp. 255–260.

[17] Roheen Qamar and Baqar Ali Zardari. “Artificial neural networks: An overview”. In: *Mesopotamian Journal of Computer Science* 2023 (2023), pp. 124–133.

[18] Oludare Isaac Abiodun et al. “State-of-the-art in artificial neural network applications: A survey”. In: *Heliyon* 4.11 (2018).

[19] RMPA Meenal et al. “Weather prediction using random forest machine learning model”. In: *Indonesian Journal of Electrical Engineering and Computer Science* 22.2 (2021), pp. 1208–1215.

[20] Peter Prettenhofer and Gilles Louppe. “Gradient boosted regression trees in scikit-learn”. In: (2014).

[21] Amin Shahraki, Mahmoud Abbasi, and Øystein Haugen. “Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost”. In: *Engineering Applications of Artificial Intelligence* 94 (2020), p. 103770.

[22] Anjaneyulu Babu Shaik and Sujatha Srinivasan. “A brief survey on random forest ensembles in classification model”. In: (2019), pp. 253–260.

Appendix A

Appendix A name

In this section, feature importance plots for every tree-based model (only for one of the scattering parameters) are presented to confirm the importance of frequency across all input features

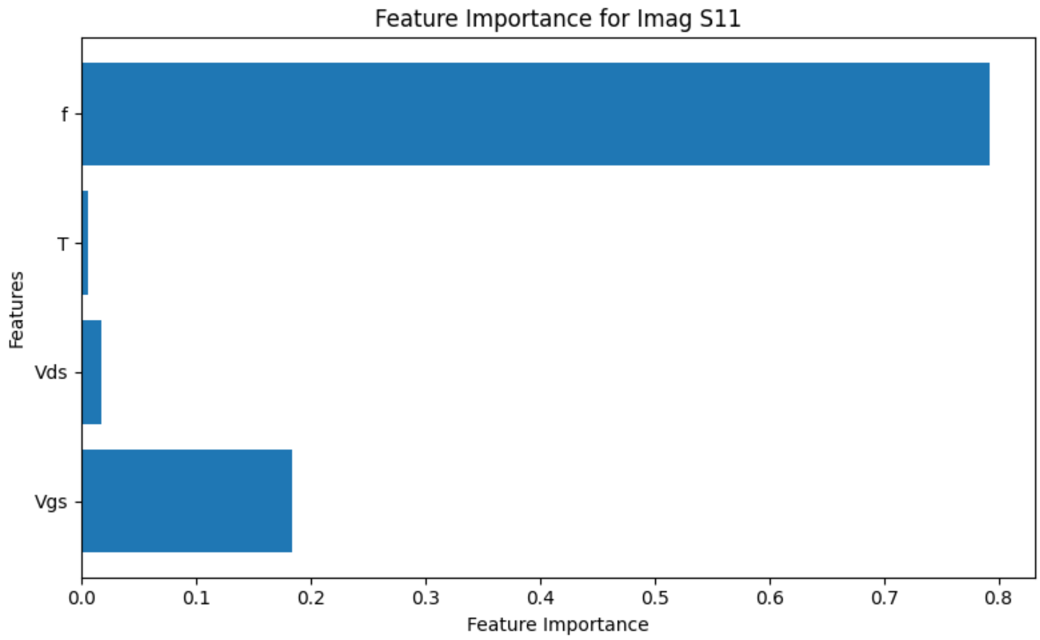


Figure A.1: Feature Importance for Imag S11 at GBR model.

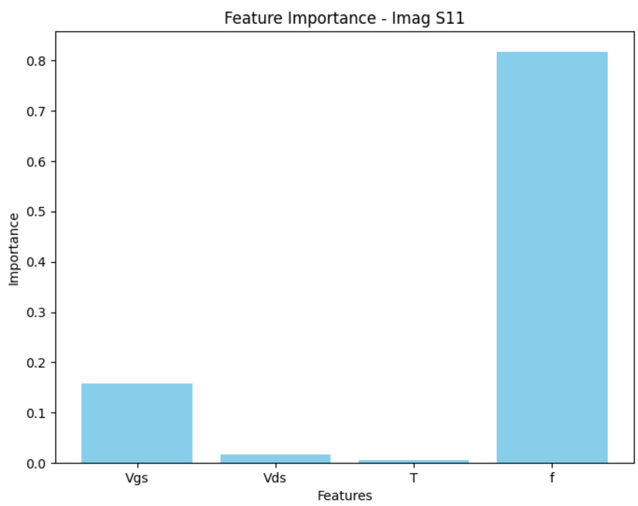


Figure A.2: Feature Importance for Imag S11 at RF model.

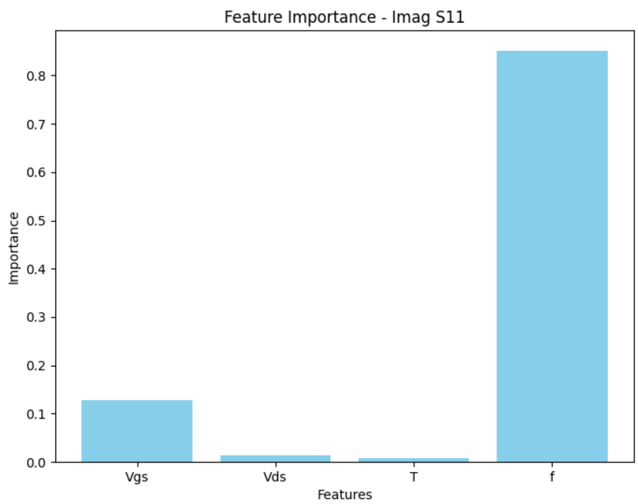


Figure A.3: Feature Importance for Imag S11 at AdaBoost (Base DT) model.

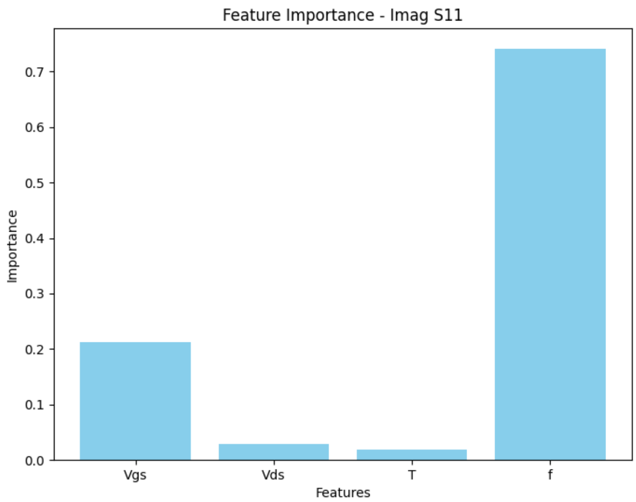


Figure A.4: Feature Importance for Imag S11 at AdaBoost (Base RF) model.