International Institute
Of Information
Technology Hyderabad

Analyzing Delay and Leakage Characteristics Across Technology Nodes: A ML Approach for Predictive Modeling

Mentor: Dr. Zia Abbas

TA: Mohammad Akhtar, Dheekshith Akula, Ritwik Goswami,
Chetan Mittal, Mehul Saraswat, Nouduru Venkata Raghavendra

Vedant Nipane, 2021102040
Pronoy Patra, 2021112019
Ashish Chokhani, 2021102016
Mayank Shivhare, 2021102026

07 May, 2024

Overview

Overview

By crafting datasets, we aim to develop accurate predictive models. Through the synthesis of Process, Voltage, and Temperature (PVT) combinations and subsequent simulation of circuits, we aim to create datasets that reflect real-world scenarios with precision. Which can be used to train ML models which can accurately predict leakage power for any given circuit

Introduction

The dataset generation process entails two primary components: the creation of Process, Voltage, and Temperature (PVT) combinations through sampling from predetermined distributions, and the utilization of these combinations to simulate circuits, with Leakage power serving as target variables.

Methodology

Dataset generation comprises of two main parts, generating Process, Voltage and Temperature (PVT) combinations by sampling these variables from pre-decided distributions and using these values to simulate circuits.

Data Analysis: Using various techniques to analayse the data to generate insights about the corelations present in the data.

ML Framework : Creating a novel method to accurately predict the dealy values.

Dataset Generation

Dataset generation comprises of two main parts, generating Process, Voltage and Temperature (PVT) combinations by sampling these variables from predecided distributions and using these values to simulate circuits.

Dataset Generation

Generating 10,000 samples using the different PTM file, ensuring the appropriate distribution of Process, Voltage, and Temperature (PVT) values to satisfy Monte-Carlo distribution criteria

The temperature range was set from -55 to 125 Celsius, following a uniform distribution to capture a broad spectrum of operating conditions. Nominal voltage was established at 1V, with uniform variations of ±10% to emulate practical voltage fluctuations

Dataset Generation

Parameters such as toxe, toxm, toxref, tox par, ndep, and xj for both PMOS and NMOS transistors were adjusted with a ±3σ variation, where σ represents the standard deviation calculated as the mean divided by 30.

For delay considerations, cqload variations ranging from 0.01f to 5f were incorporated using a uniform distribution.

Dataset Generation

9985	9983	24.55927	220.6396	2.719958	0.787381	2.15589	1.07E-09	1.02E-09	1.12E-09	8.14E-10	7.39E-09	5.49E+18	1.08E-09	1.13E-09	1.12E-09	7.85E-10	7.33E-09	4.36E+18
9986	9984	24.83281	234.1224	-47.1691	0.784295	3.462122	1.15E-09	1.05E-09	9.79E-10	7.71E-10	7.10E-09	5.52E+18	1.06E-09	1.03E-09	1.12E-09	8.14E-10	7.33E-09	4.27E+18
9987	9985	24.40074	209.8893	-8.46476	0.783937	0.91975	1.04E-09	1.11E-09	9.75E-10	7.64E-10	7.13E-09	5.31E+18	1.10E-09	1.08E-09	1.07E-09	8.16E-10	7.54E-09	4.28E+18
9988	9986	24.66666	241.9075	45.51774	0.816266	2.130853	1.03E-09	1.00E-09	1.02E-09	7.61E-10	7.18E-09	5.48E+18	1.05E-09	1.09E-09	1.08E-09	7.59E-10	7.10E-09	4.62E+18
9989	9987	24.41727	238.6659	-20.9619	0.874754	0.677772	1.06E-09	1.12E-09	1.02E-09	7.45E-10	6.94E-09	5.52E+18	1.13E-09	1.11E-09	1.12E-09	7.86E-10	6.79E-09	4.14E+18
9990	9988	23.61034	237.1344	-31.7759	0.785905	3.058087	1.04E-09	1.01E-09	1.08E-09	8.26E-10	7.16E-09	5.75E+18	1.13E-09	1.13E-09	1.10E-09	7.51E-10	7.17E-09	4.22E+18
9991	9989	24.09511	226.2576	45.51709	0.72642	4.80755	1.11E-09	1.04E-09	1.02E-09	7.57E-10	7.31E-09	5.50E+18	1.07E-09	1.12E-09	1.13E-09	7.87E-10	7.11E-09	4.49E+18
9992	9990	24.66098	228.7205	85.0884	0.765521	0.194509	1.04E-09	9.97E-10	1.11E-09	8.12E-10	7.34E-09	5.33E+18	1.04E-09	1.13E-09	1.11E-09	7.65E-10	7.09E-09	4.40E+18
9993	9991	24.14907	232.9779	-1.17267	0.752795	1.396864	1.05E-09	1.01E-09	1.07E-09	7.82E-10	7.62E-09	5.29E+18	1.12E-09	1.09E-09	1.10E-09	8.11E-10	7.12E-09	4.49E+18
9994	9992	24.33865	226.389	24.7488	0.721848	3.025667	1.06E-09	1.00E-09	1.06E-09	8.39E-10	8.17E-09	5.57E+18	1.14E-09	1.09E-09	1.06E-09	7.95E-10	6.69E-09	4.51E+18
9995	9993	22.99468	236.1828	88.52304	0.830153	1.503775	9.93E-10	1.07E-09	1.03E-09	7.93E-10	6.67E-09	5.51E+18	1.06E-09	1.08E-09	1.14E-09	8.07E-10	6.88E-09	4.29E+18
9996	9994	24.76455	227.2908	87.54471	0.785281	2.225158	1.09E-09	1.02E-09	1.01E-09	7.94E-10	7.40E-09	5.60E+18	1.08E-09	1.13E-09	1.13E-09	7.66E-10	7.52E-09	4.22E+18
9997	9995	24.96551	232.2298	55.11142	0.848873	4.652205	1.05E-09	1.05E-09	1.07E-09	8.26E-10	6.86E-09	5.50E+18	1.10E-09	1.11E-09	1.12E-09	8.29E-10	6.98E-09	4.43E+18
9998	9996	23.19845	227.4463	-41.4594	0.823329	0.069362	1.06E-09	1.14E-09	1.00E-09	8.08E-10	6.79E-09	5.36E+18	1.08E-09	1.18E-09	1.14E-09	7.86E-10	7.41E-09	4.44E+18
9999	9997	23.12403	233.4831	119.2275	0.829539	2.087758	1.08E-09	1.05E-09	1.07E-09	8.02E-10	7.24E-09	5.54E+18	1.09E-09	1.07E-09	1.18E-09	8.05E-10	7.01E-09	4.65E+18
10000	9998	24.49356	224.8856	113.7003	0.807239	2.117105	1.07E-09	1.05E-09	1.07E-09	7.83E-10	7.72E-09	5.63E+18	1.09E-09	1.12E-09	1.13E-09	7.85E-10	6.90E-09	4.32E+18
10001	9999	24.00054	238.0811	-48.822	0.773309	1.9156	1.05E-09	1.12E-09	9.75E-10	8.29E-10	7.25E-09	5.60E+18	1.07E-09	1.09E-09	1.11E-09	8.32E-10	7.58E-09	4.49E+18

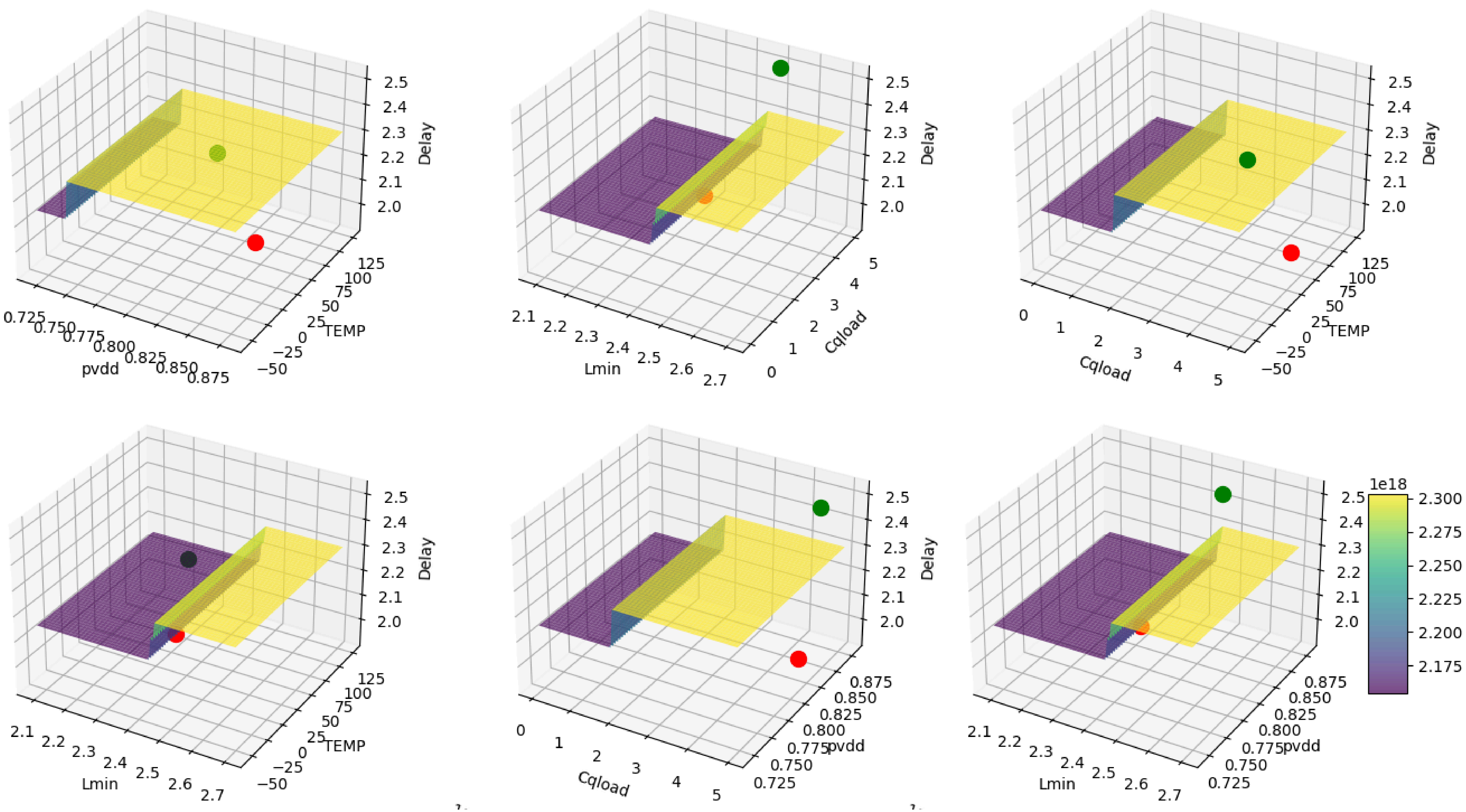
Data Analysis

Many data analysis techniques were used to better visualize the dataset and gain important insights of the data which can be exploited while using it for training the ML models such as:

- 3D plots visualization
- Dual Axis Plots
- Correlation
- GMM clustering
- PCA

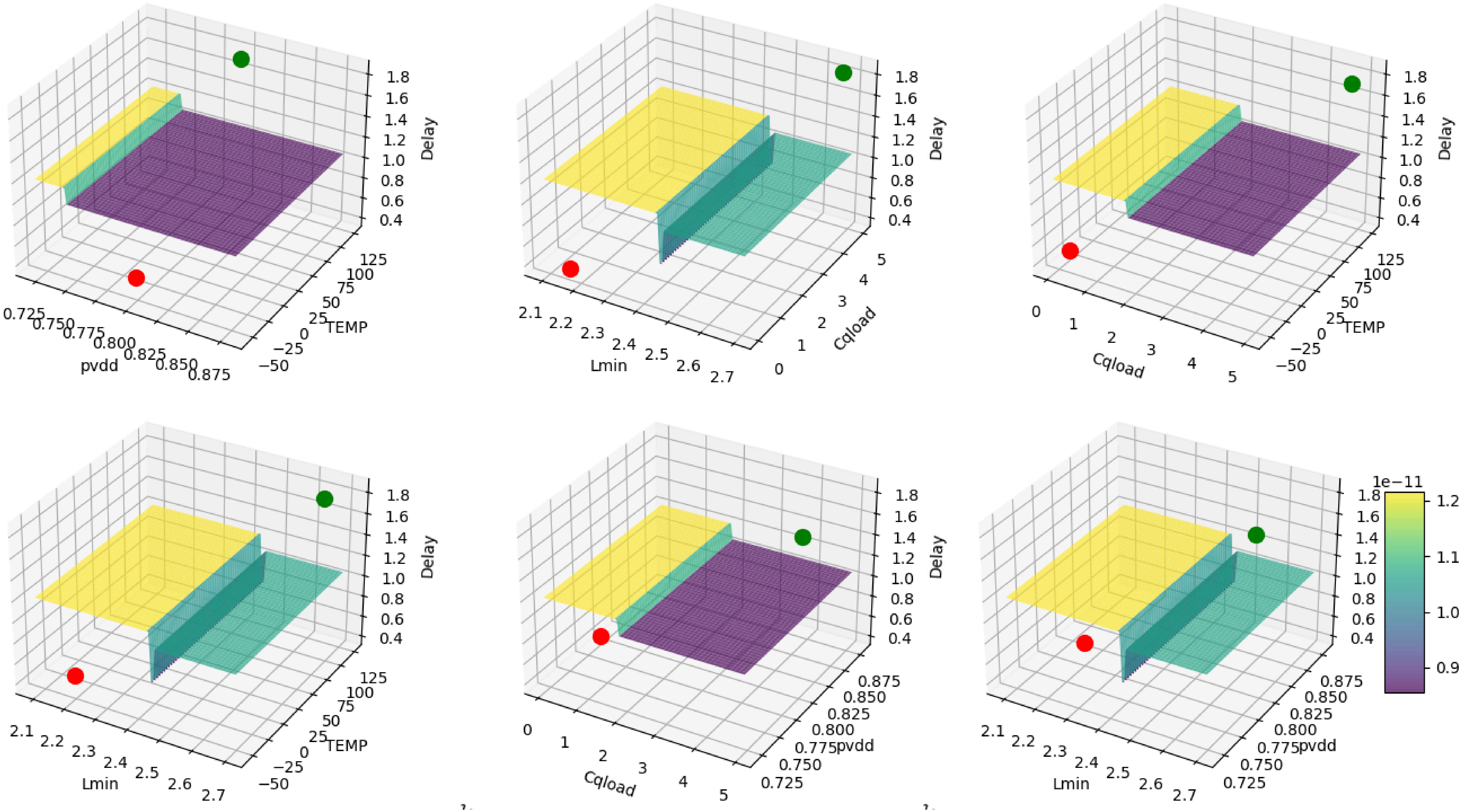
3D Plots

3D Surface Plots for INVERTER (delay) gate



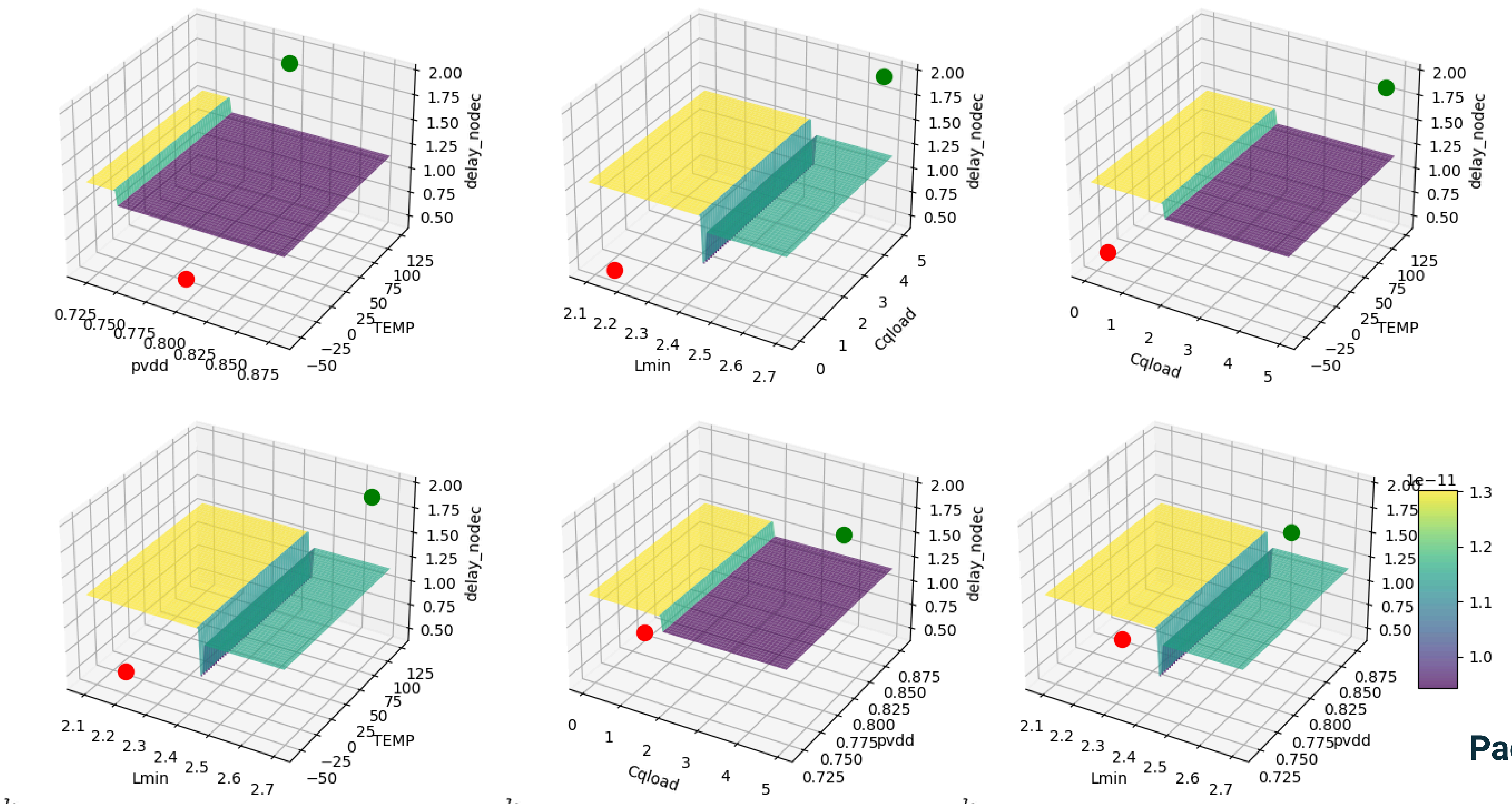
3D Plots

3D Surface Plots for AND3 (delay) gate



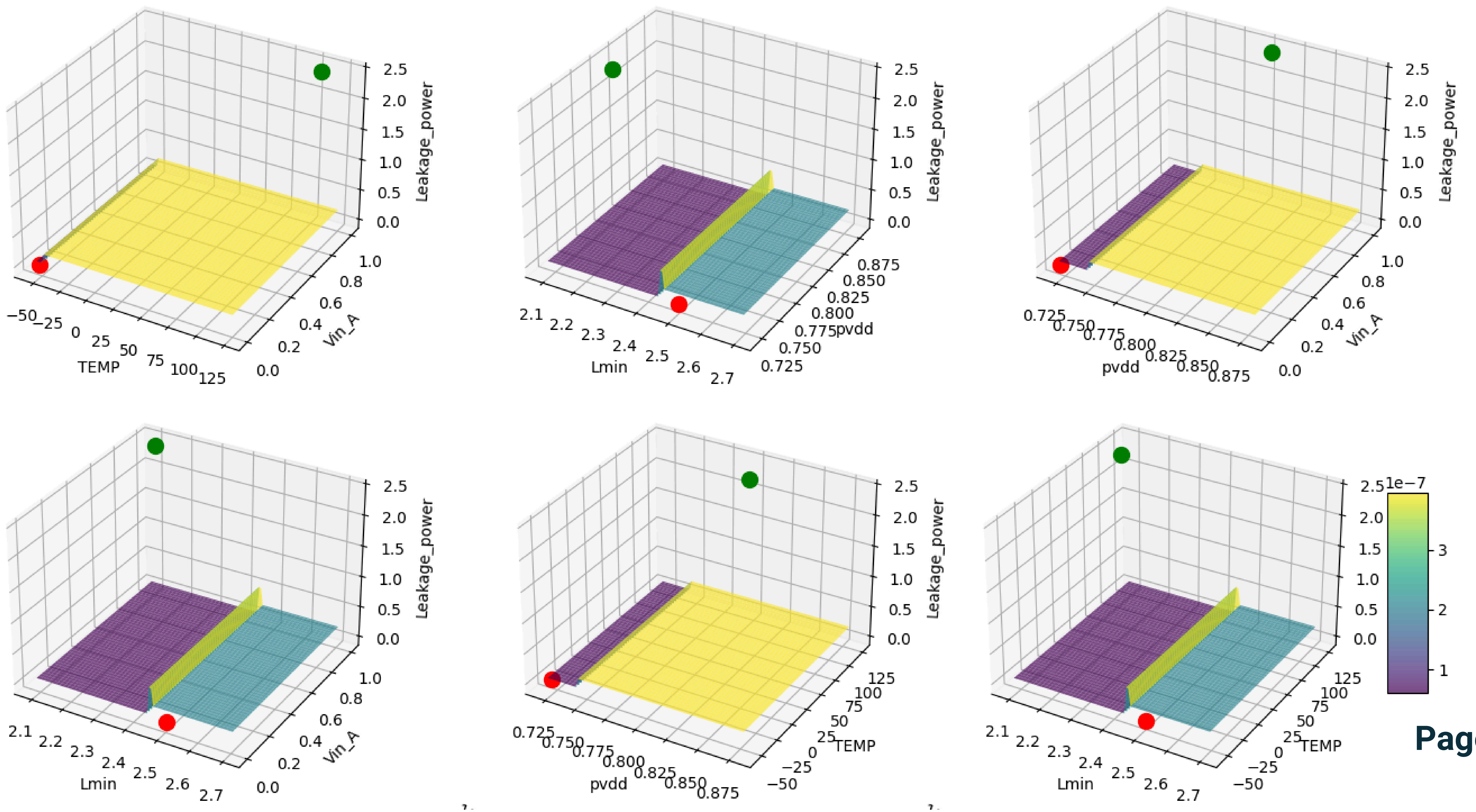
3D Plots

3D Surface Plots for OR2 (delay) gate



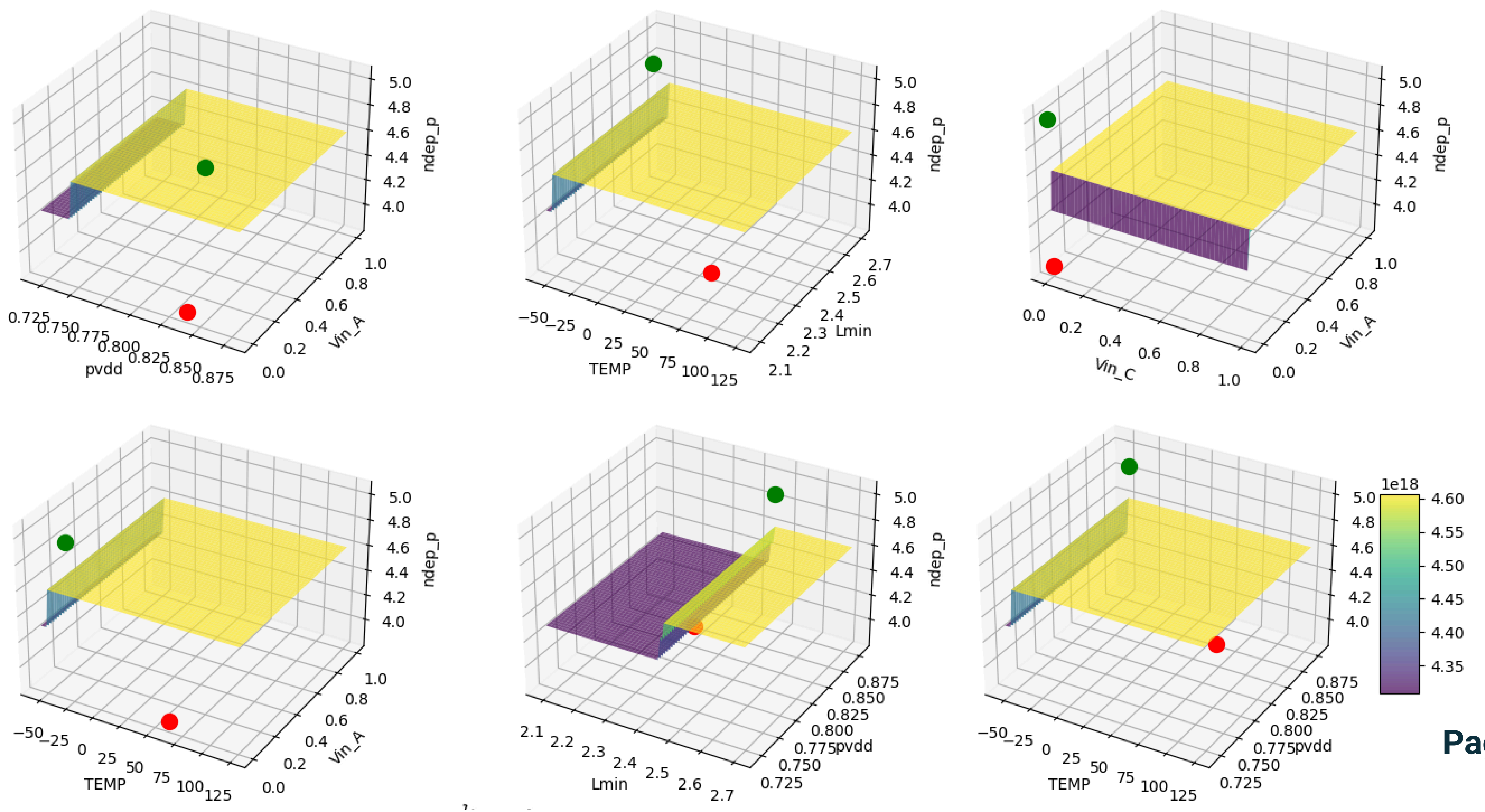
3D Plots

3D Surface Plots for INVERTER (leakage) gate



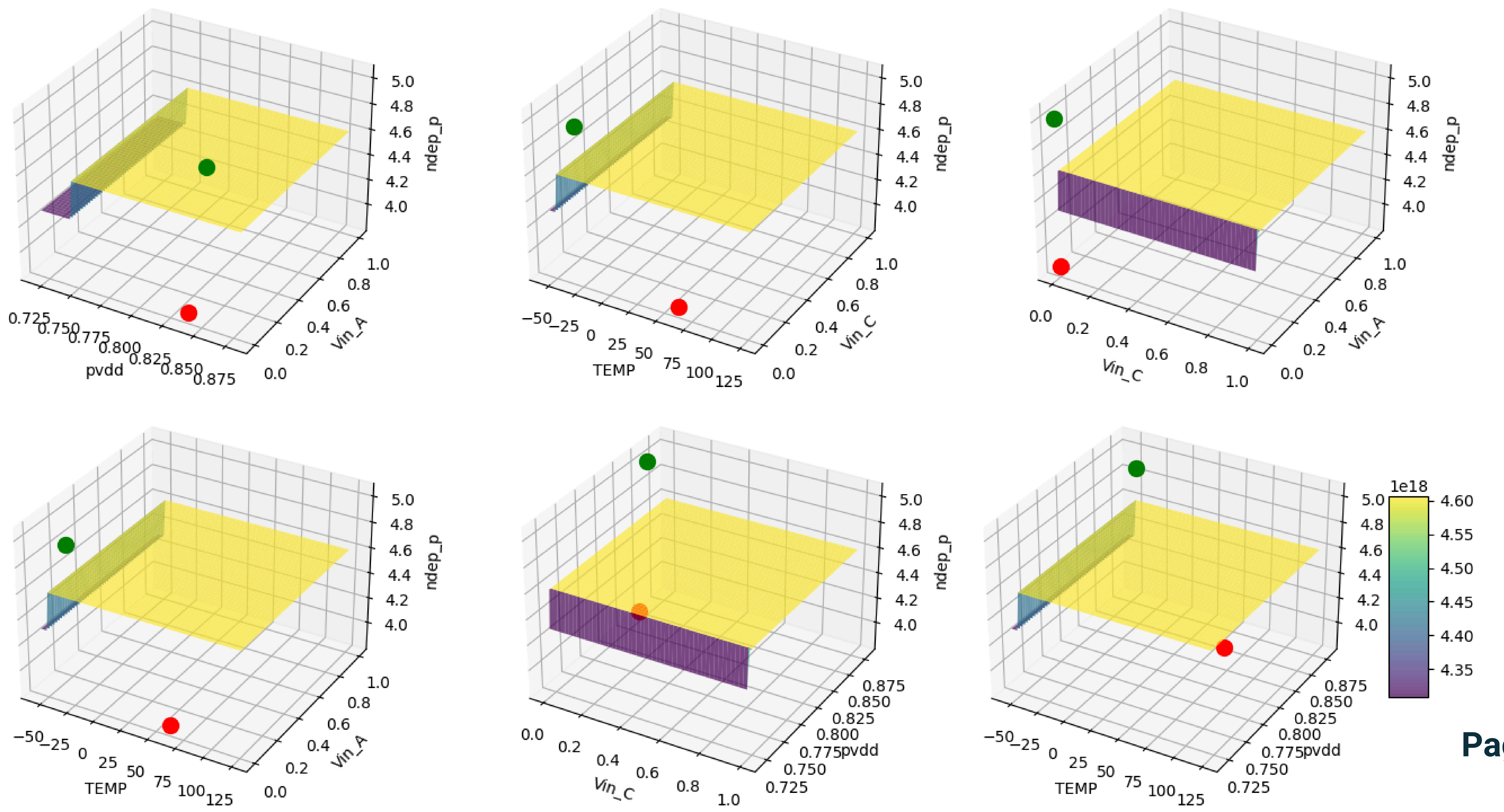
3D Plots

3D Surface Plots for AND3 (leakage) gate



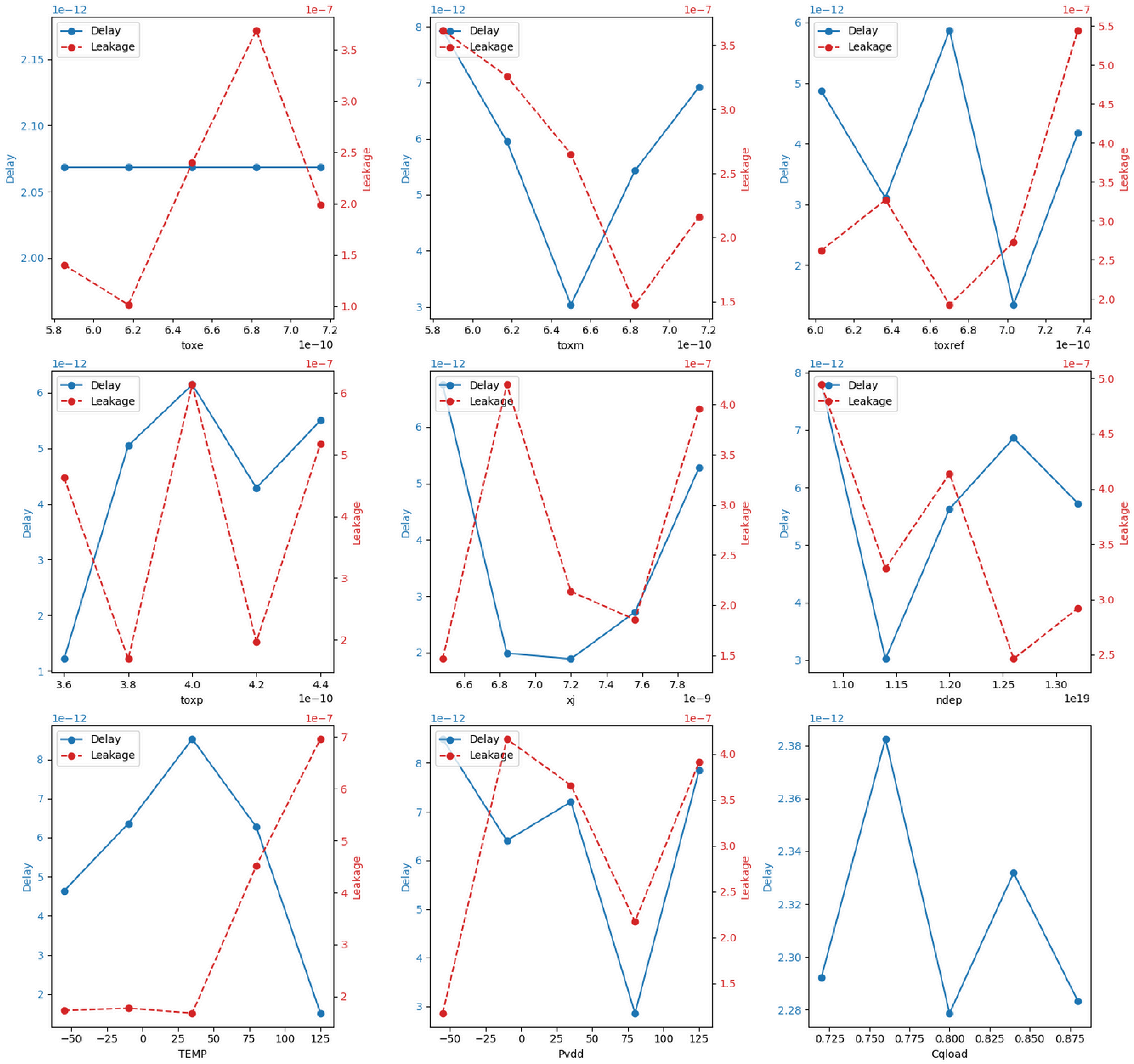
3D Plots

3D Surface Plots for OR2 (leakage) gate



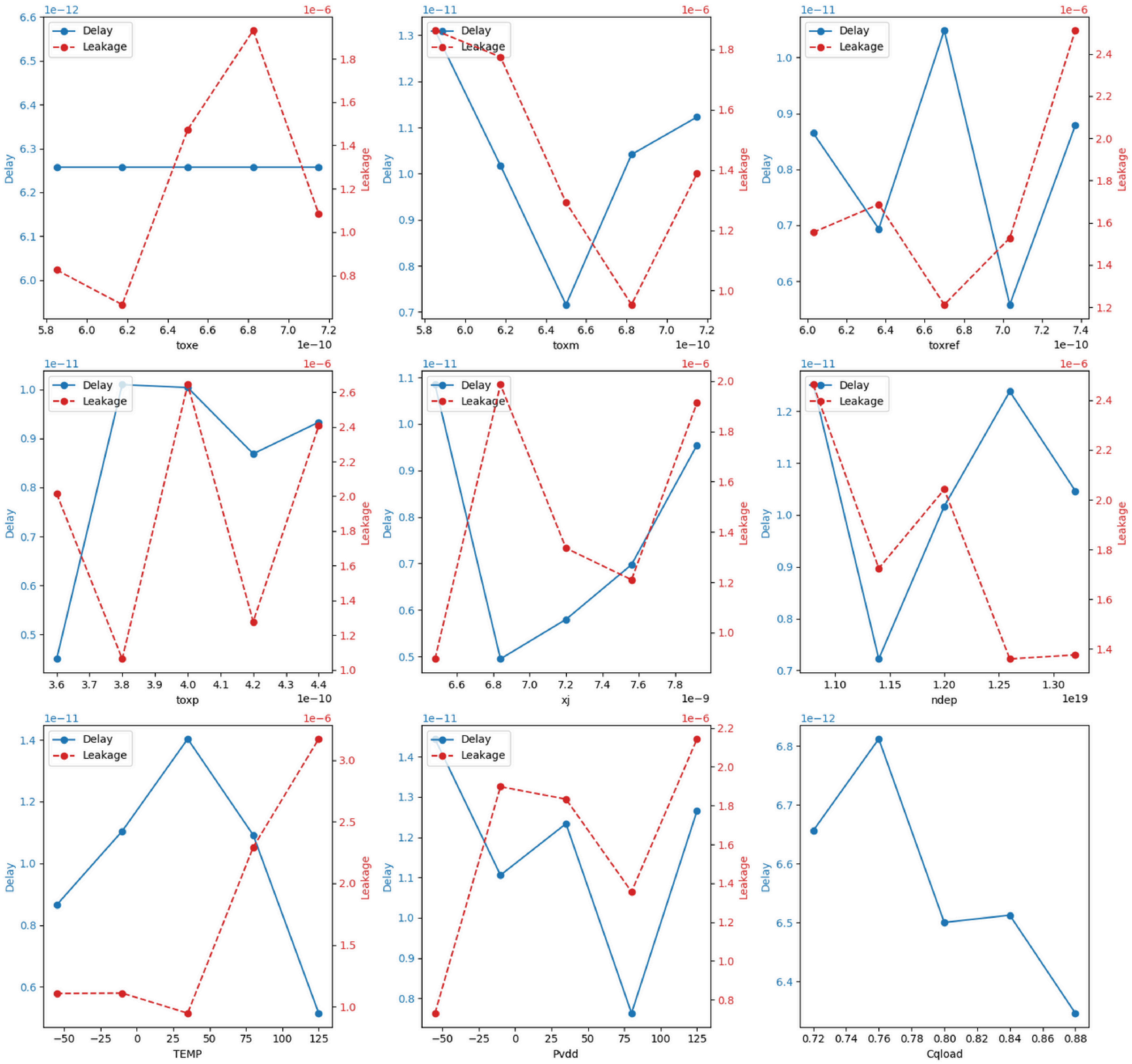
Dual Axis Plots

Variation of Delay and Leakage for Inverter for various technology parameters



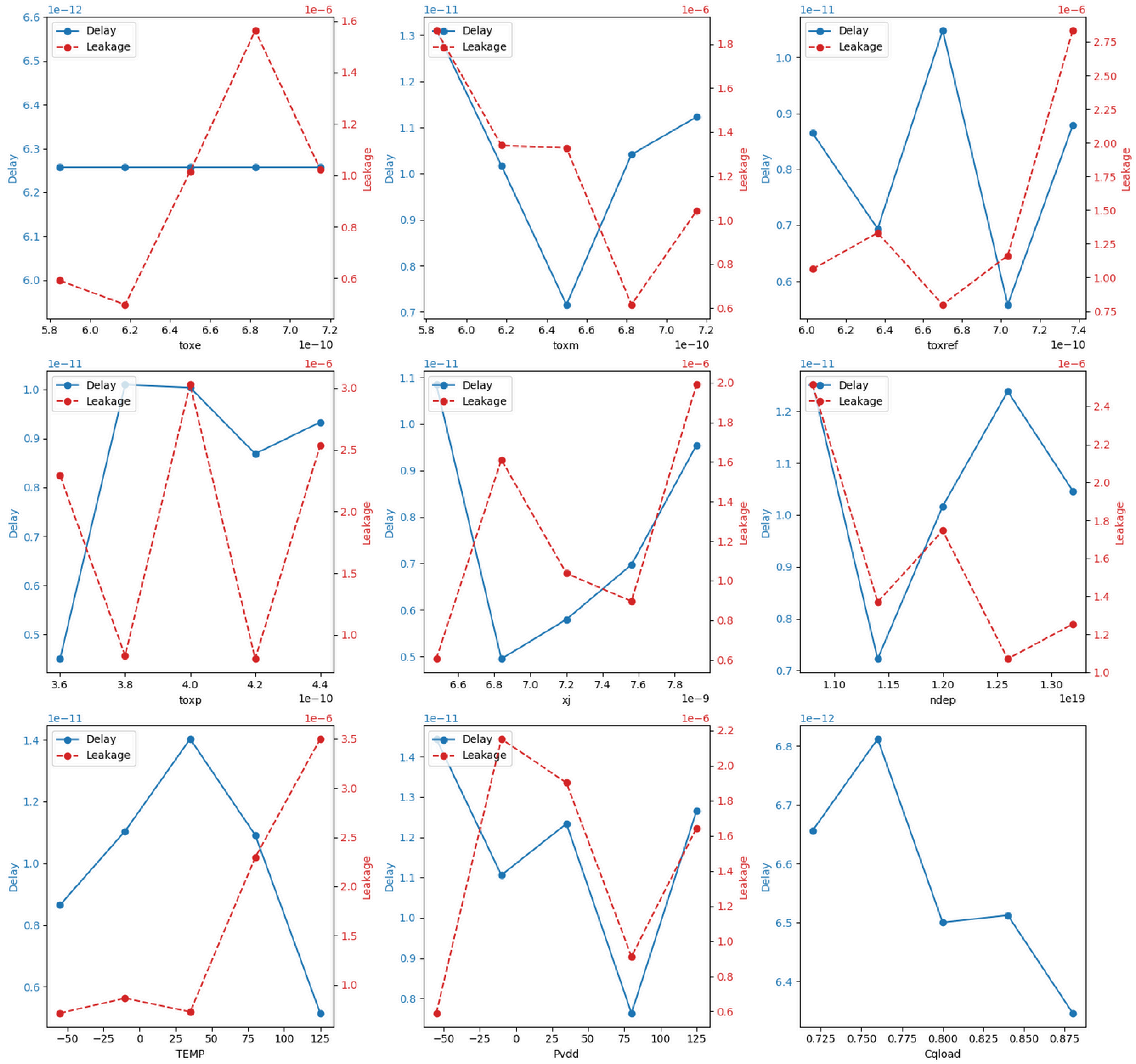
Dual Axis Plots

Variation of Delay and Leakage for AND3 for various technology parameters

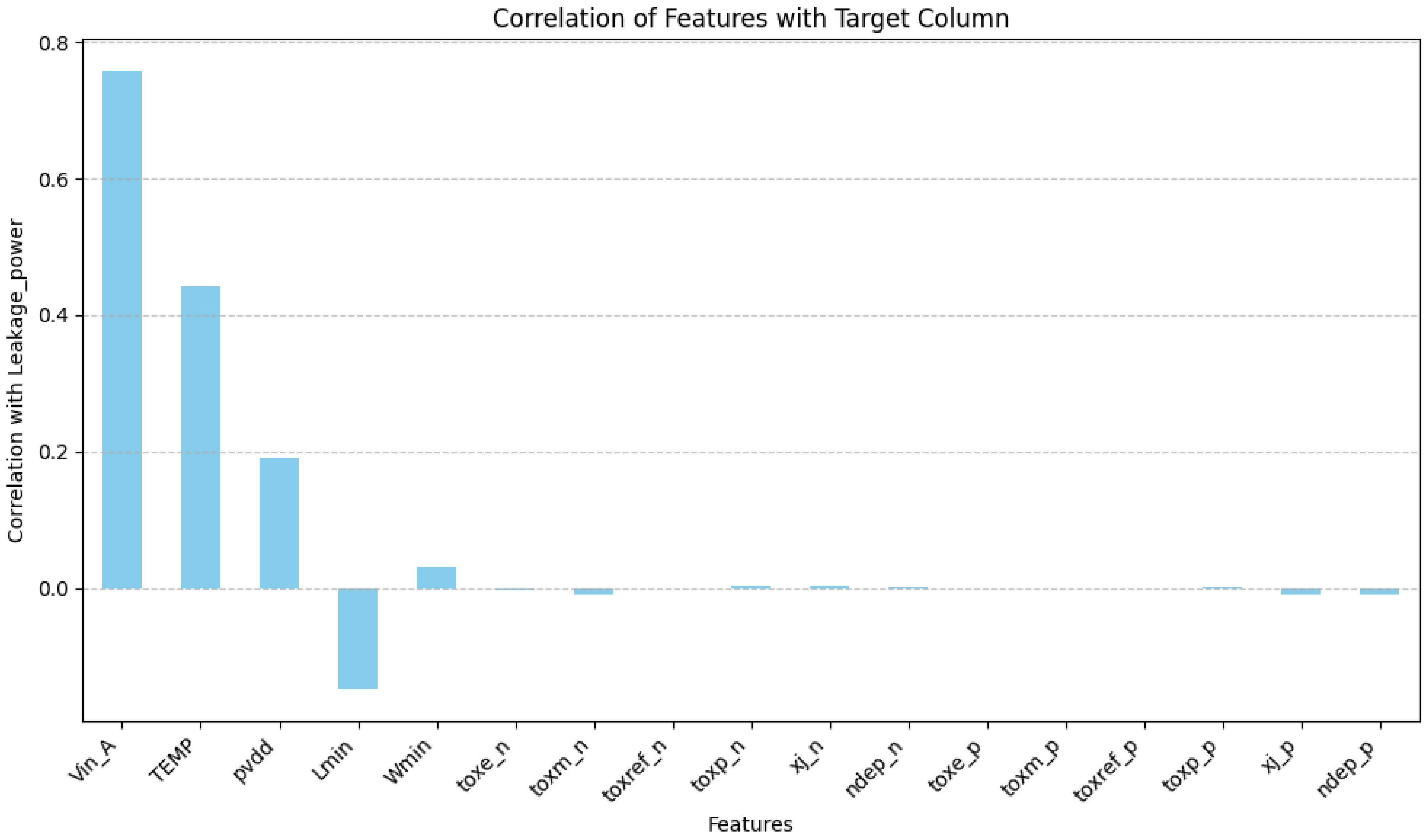


Dual Axis Plots

Variation of Delay and Leakage for OR2 for various technology parameters



Correlation of Leakage Power with Input Parameters

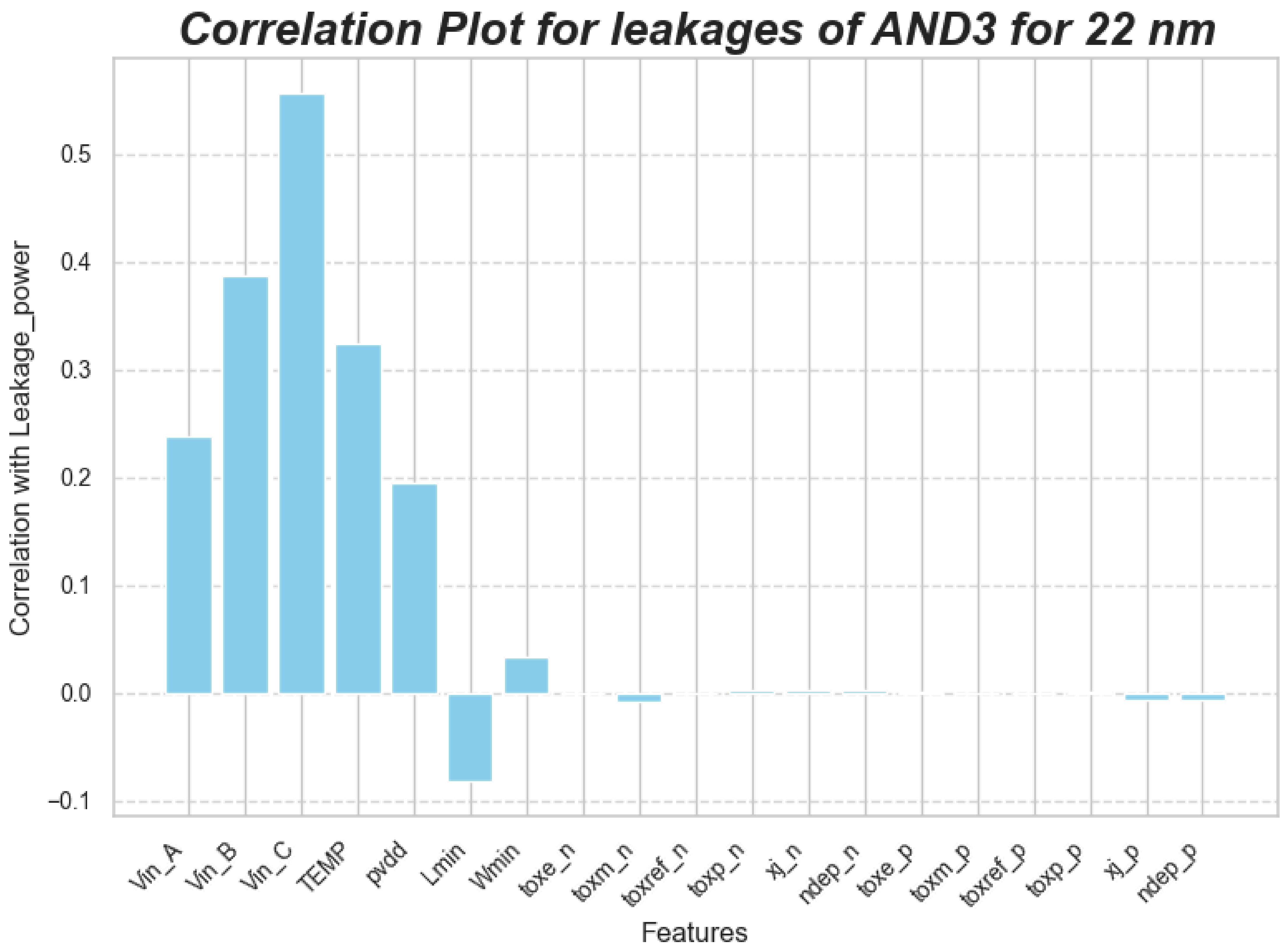


Inverter leakage for 22nm_MGK

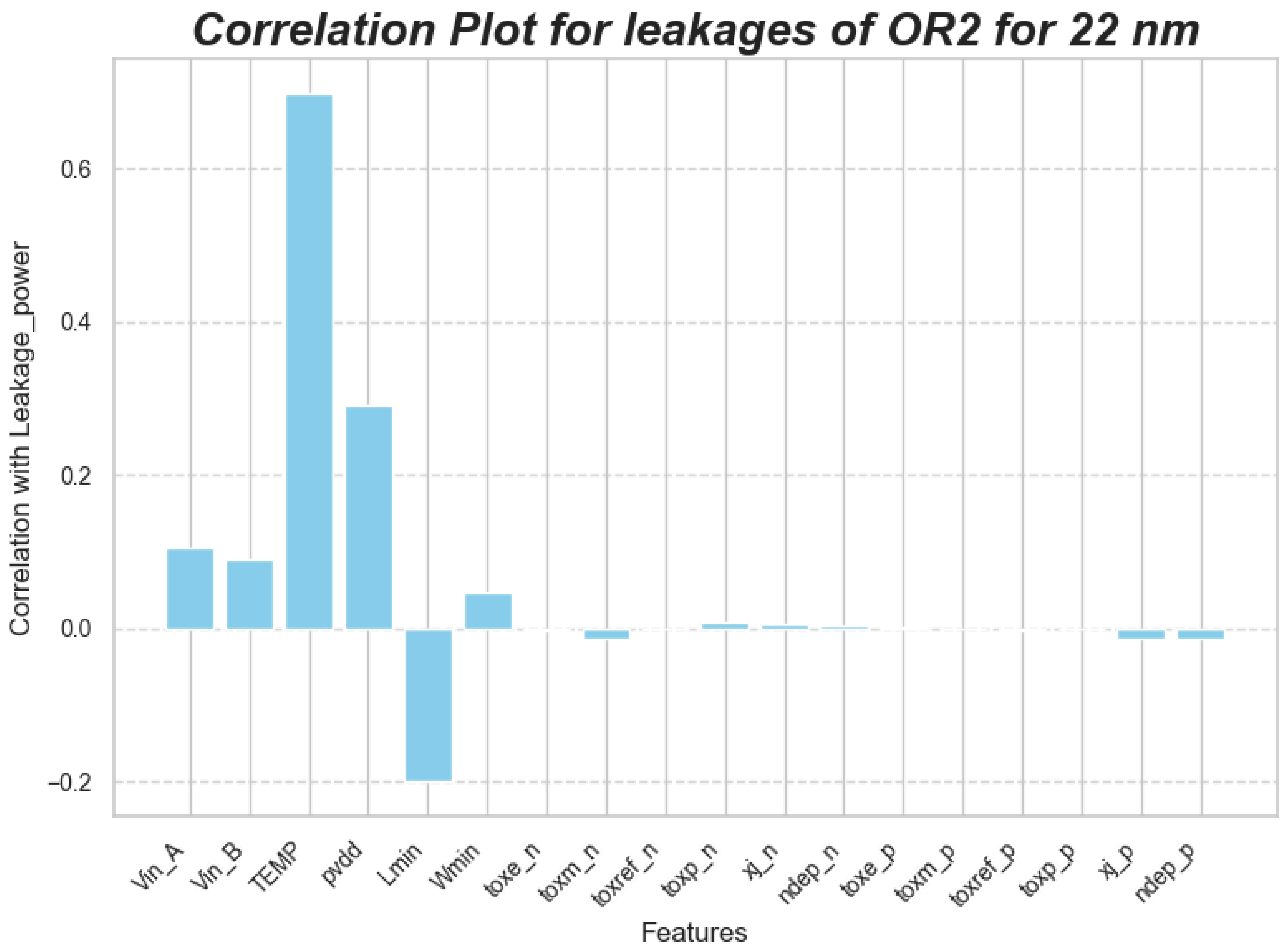
The more the correlation according to the graph, the more sensitive are those parameters to the leakage power.

Also, we can observe that Leakage power increases with decrease in L_min .

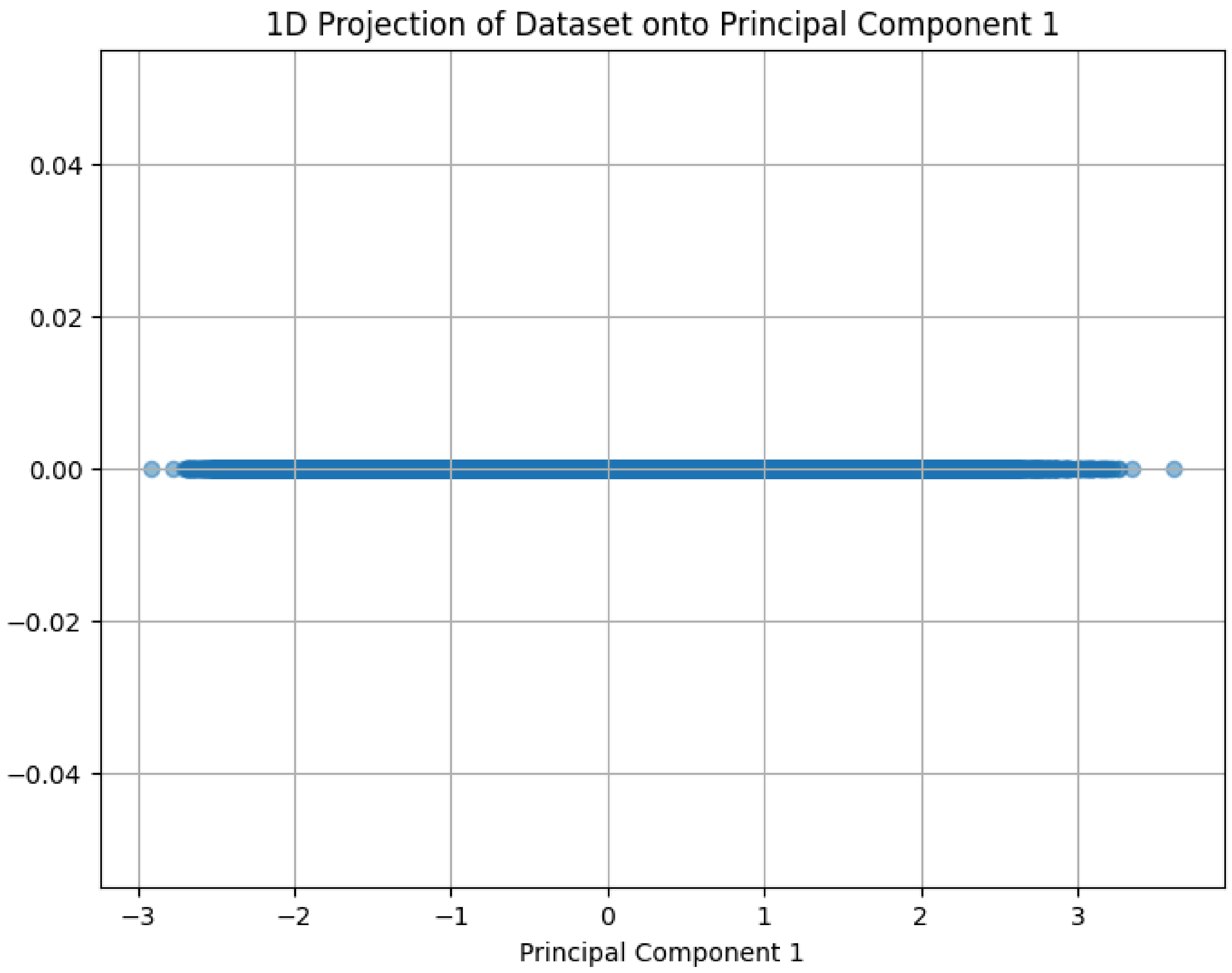
Correlation of Leakage Power with Input Parameters



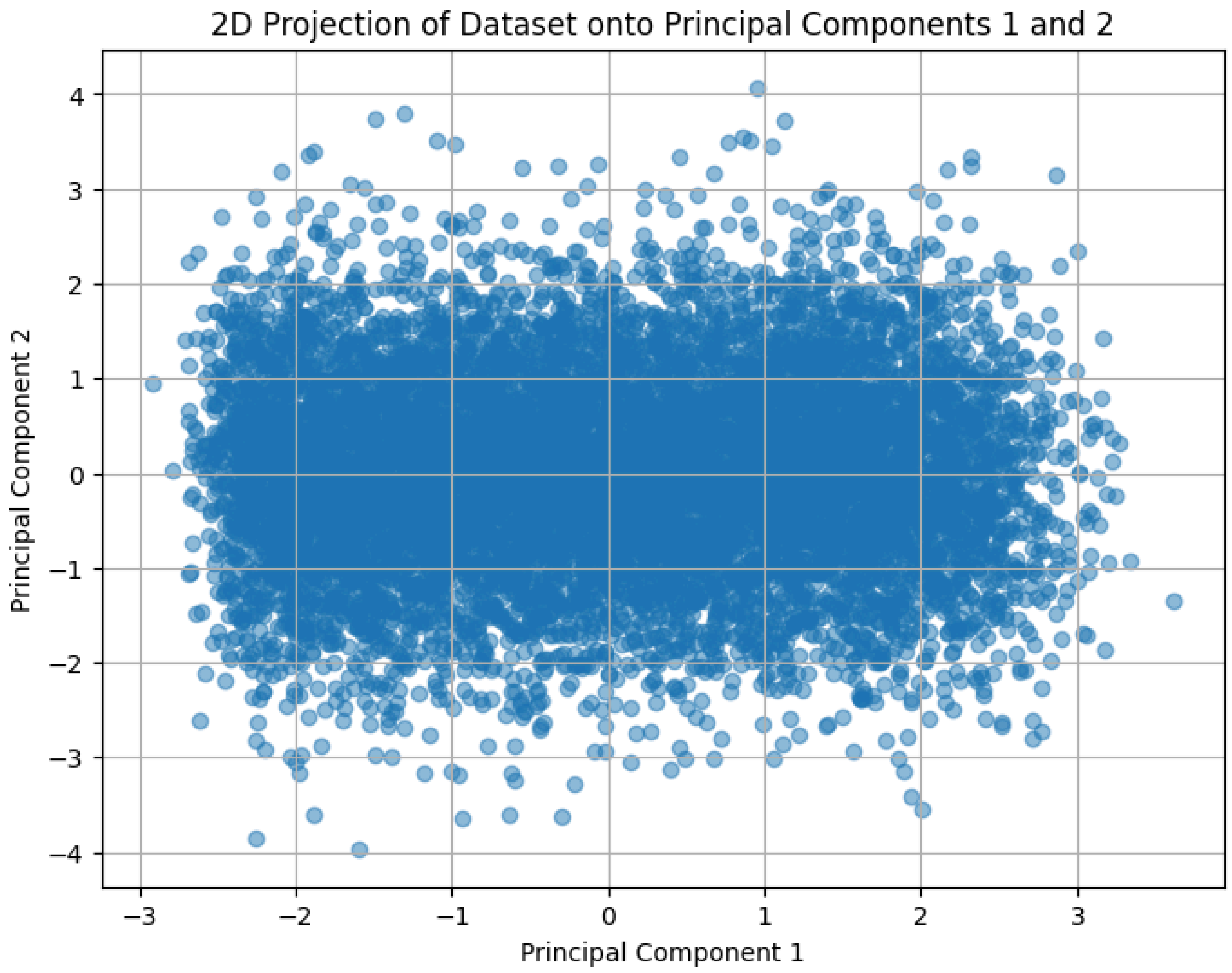
Correlation of Leakage Power with Input Parameters



Prinicipal Component Analysis (PCA)

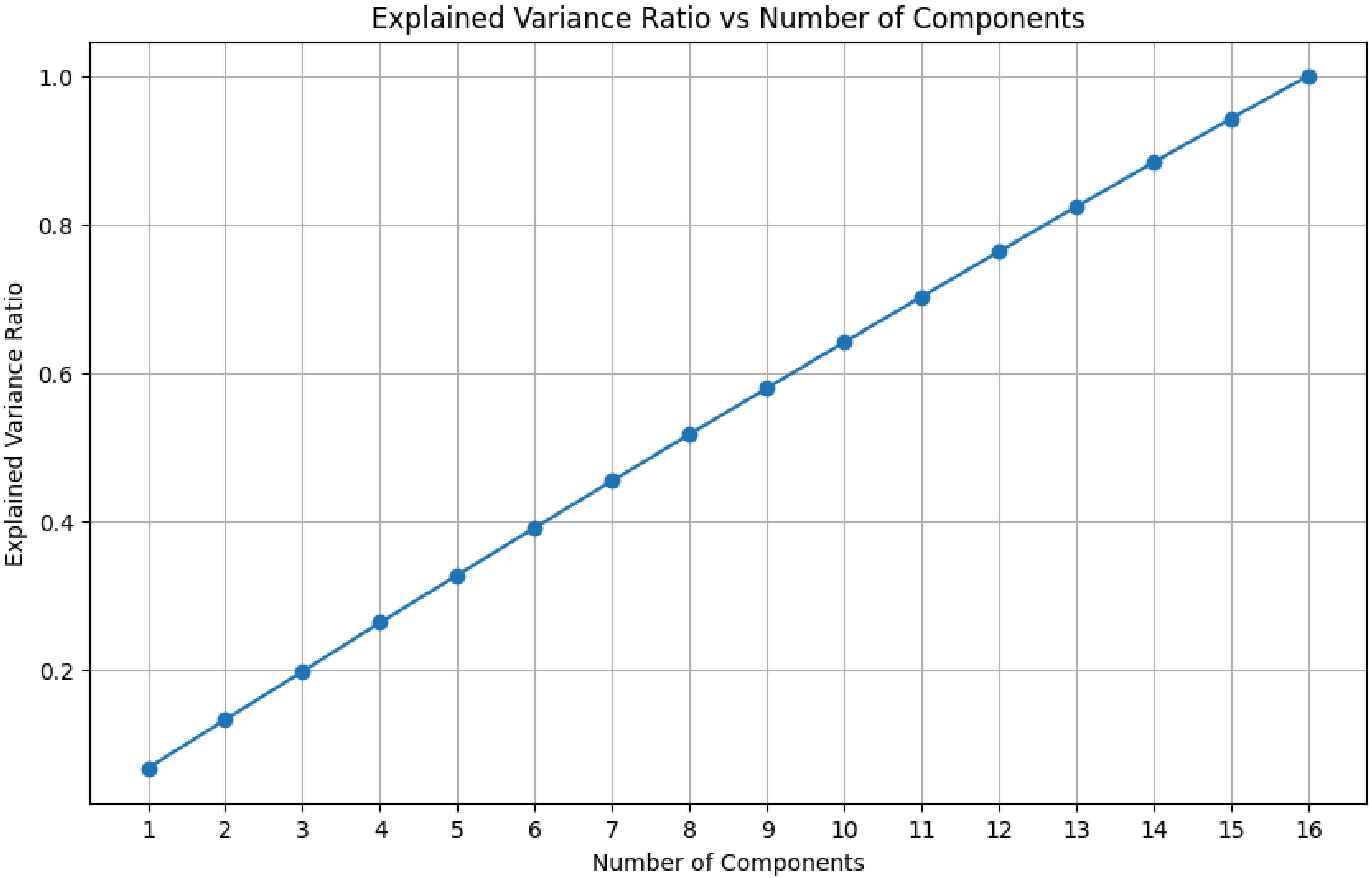


Visualization of PCA with 1 component for Inverter delay of 22nm_MGK



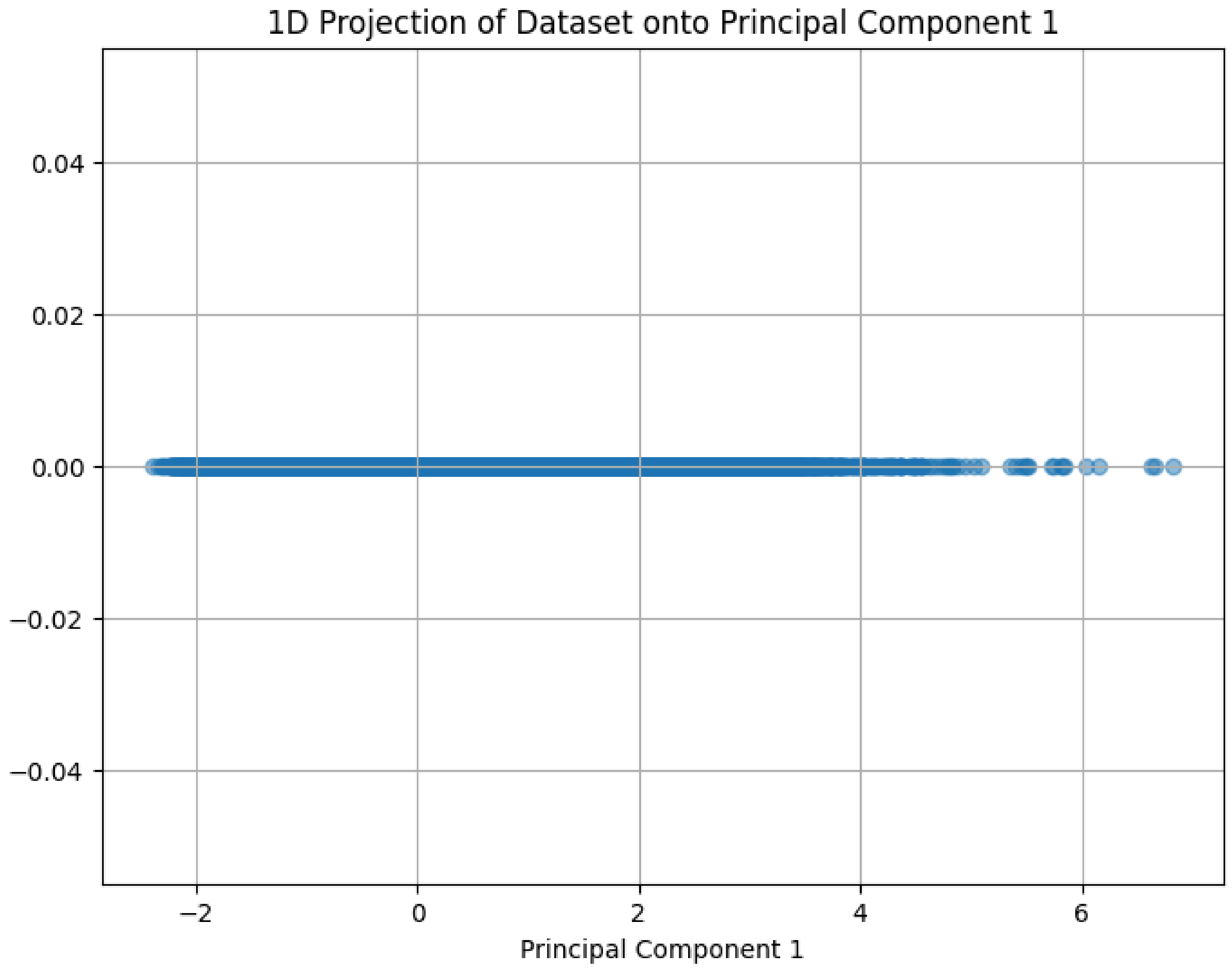
Visualization of PCA with 2 component for Inverter delay of 22nm_MGK

Prinicipal Component Analysis (PCA)

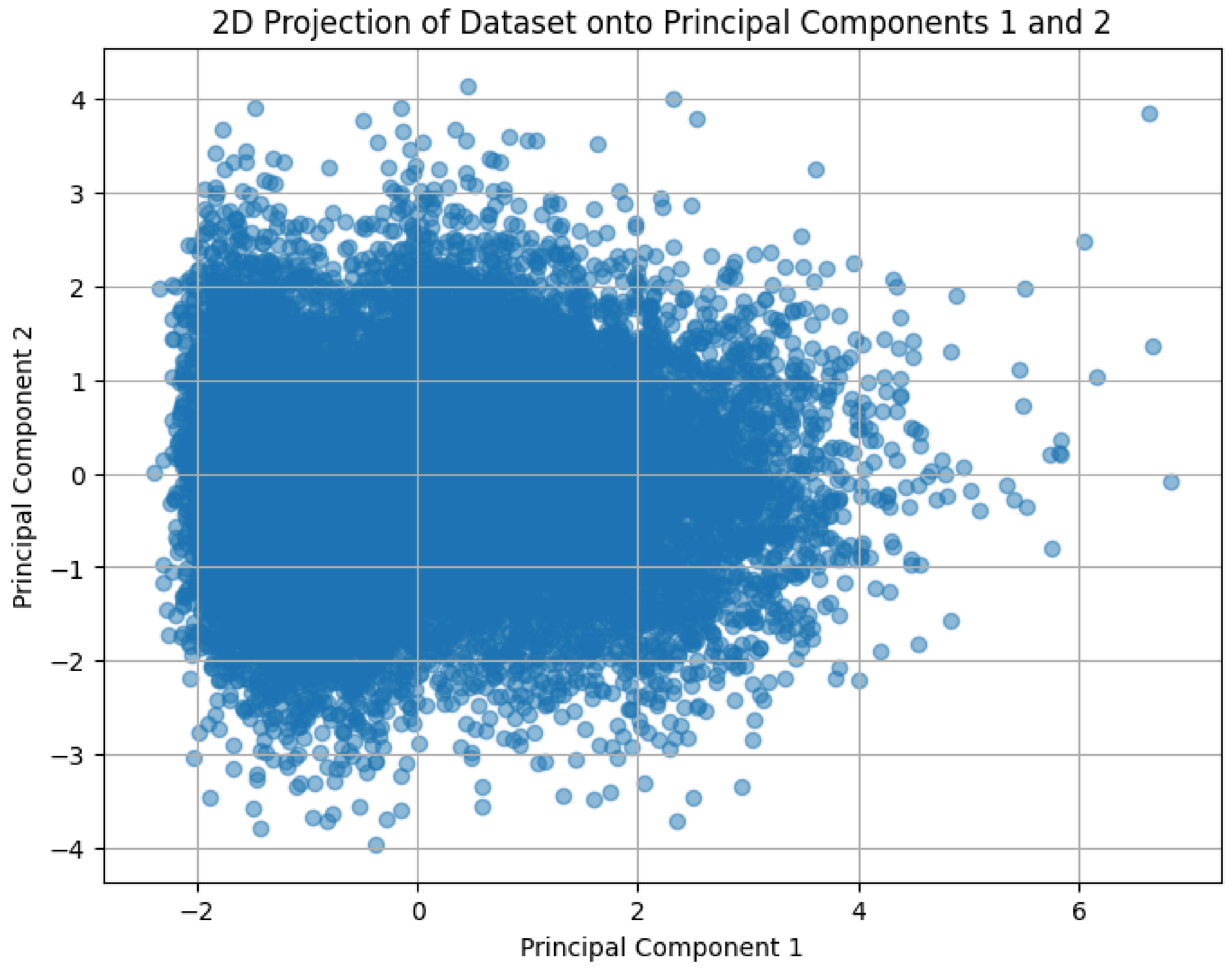


This plot depicts that explained variance ratio increases with increase in no. of input parameters and hence, all parameters have an effect on the INVERTER_delay for 22nm_MGK

Prinicipal Component Analysis (PCA)

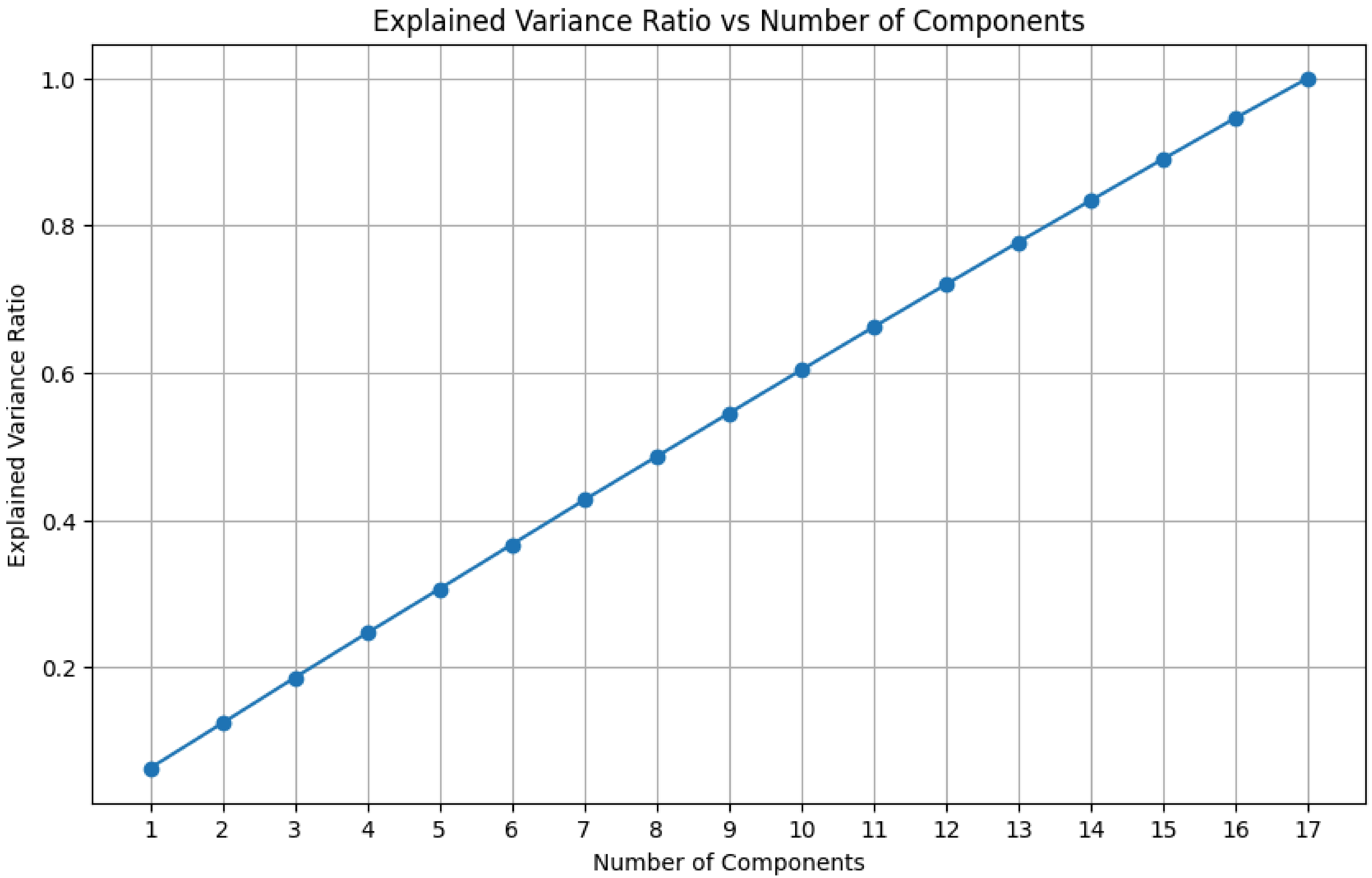


Visualization of PCA with 1 component for leakage of 22nm_MGK



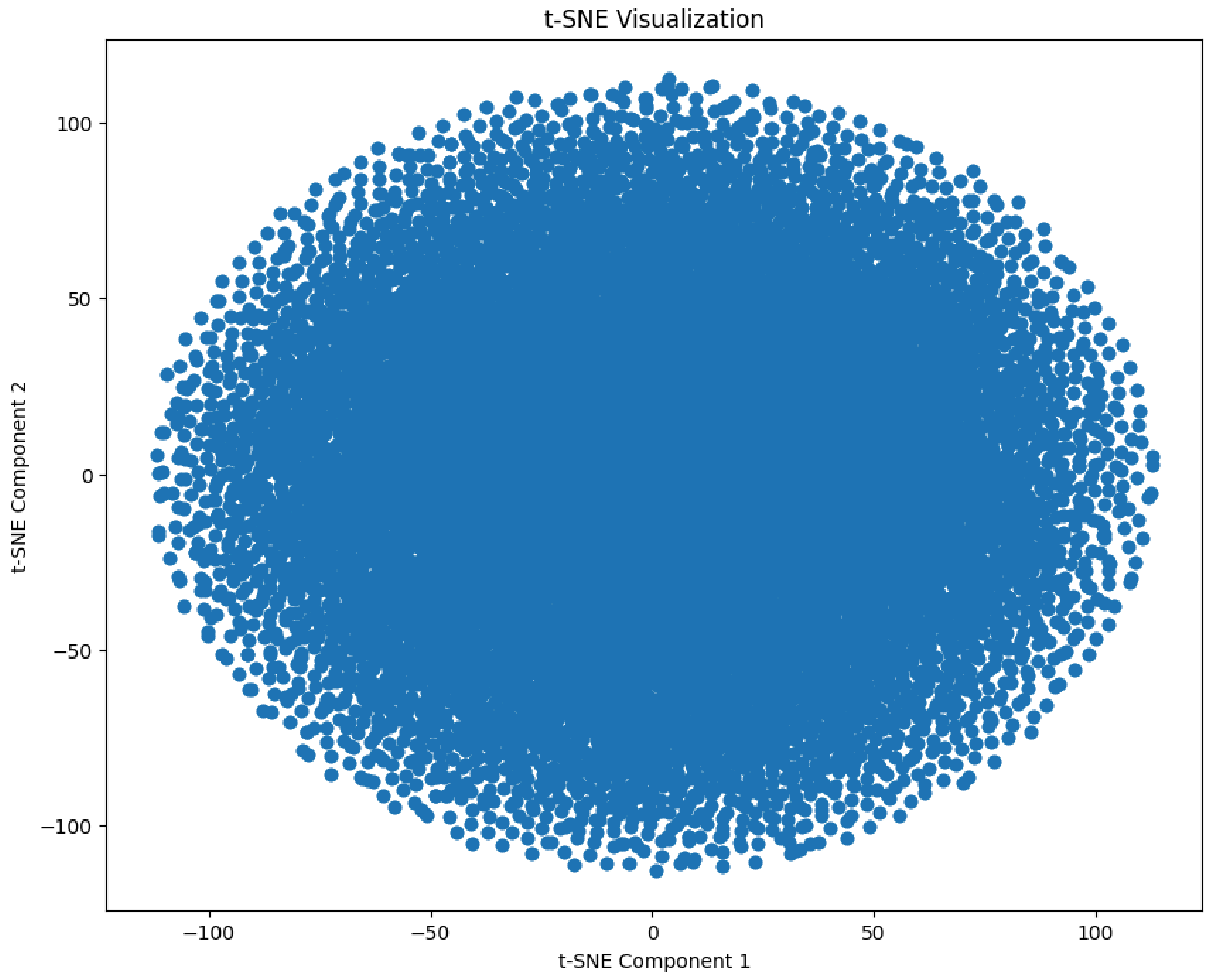
Visualization of PCA with 2 components for Inverter leakage of 22nm_MGK

Prinicipal Component Analysis (PCA)



This plot depicts that explained variance ratio increases with increase in no. of input parameters and hence, all parameters have an effect on the INVERTER_leakage for 22nm_MGK

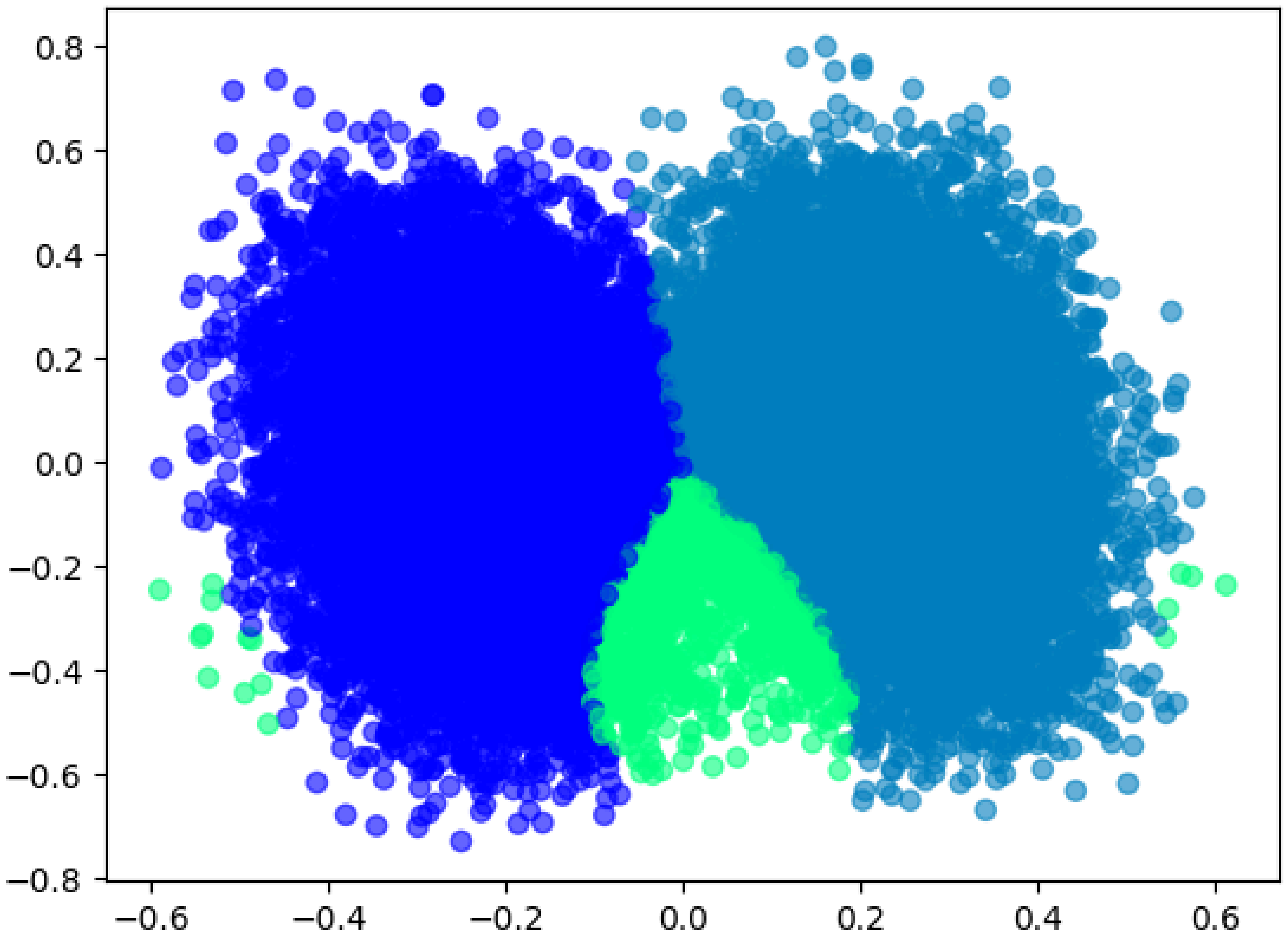
t-SNE Visualization



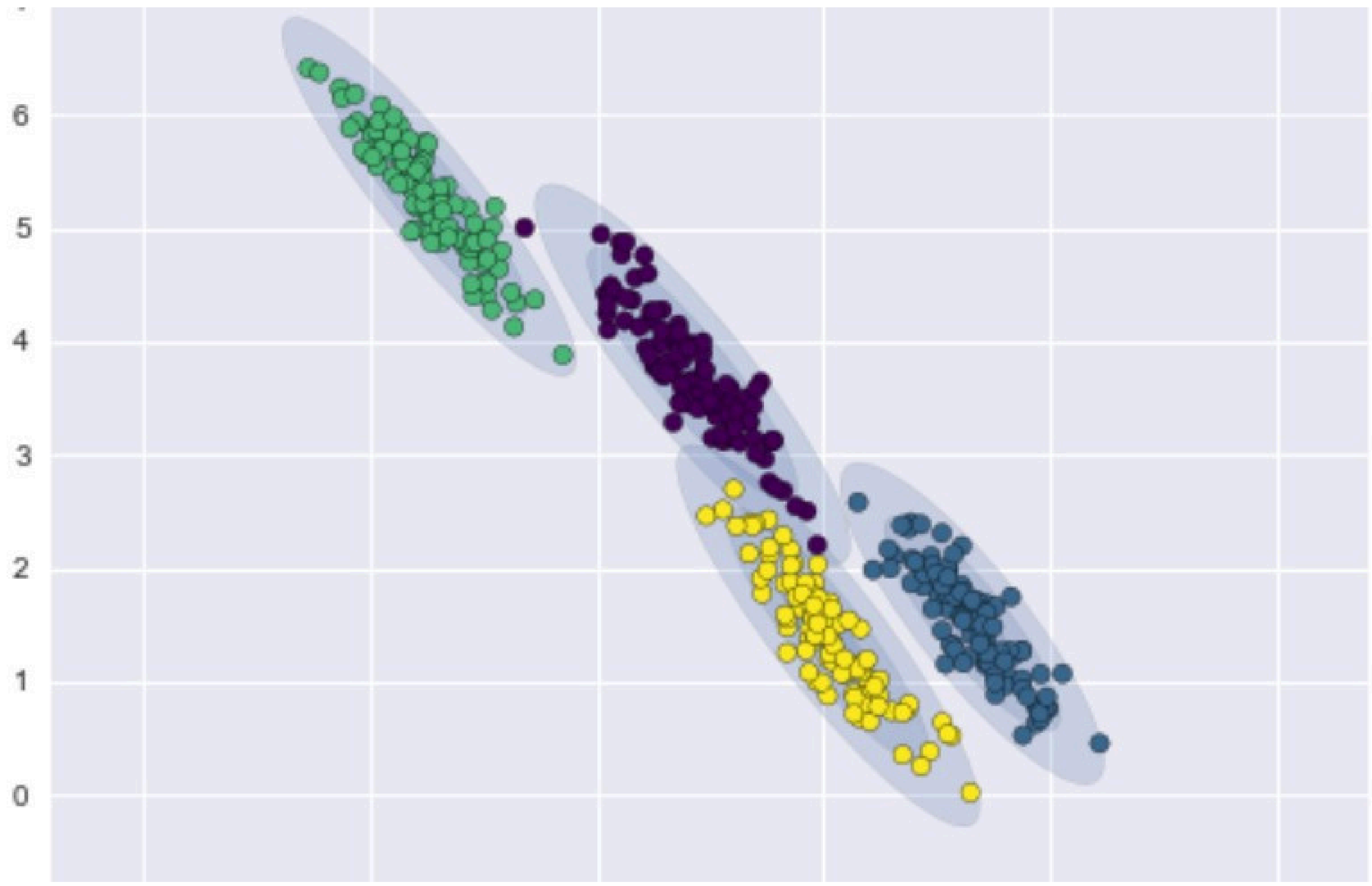
t-SNE is concerned with preserving small pairwise distances whereas, PCA focuses on maintaining large pairwise distances to maximize variance.

t-SNE preserves the relationships between data points in a lower-dimensional space, making it quite a good algorithm for visualizing complex high-dimensional data.

Gaussian Mixture Models

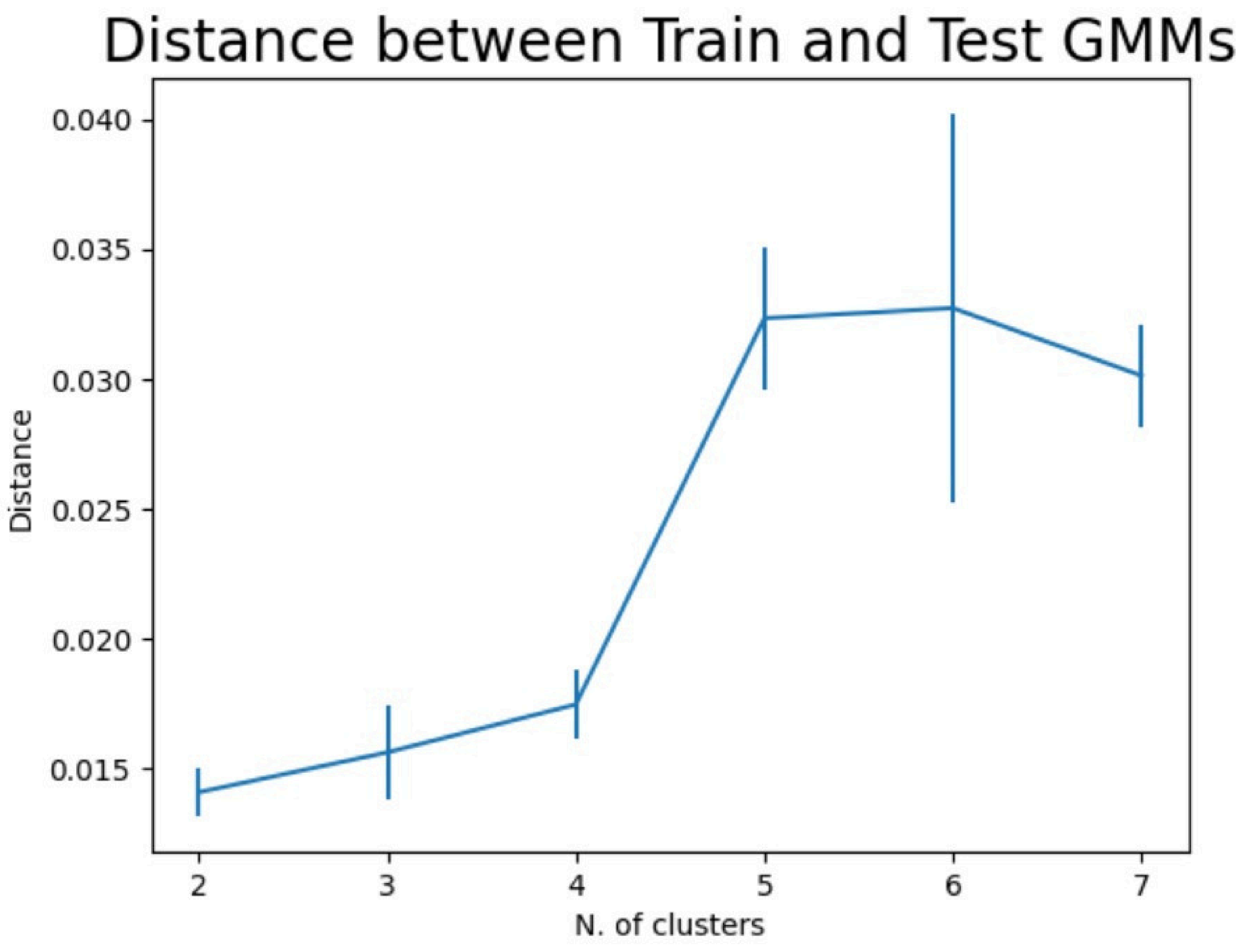


Visualization of GMM with 3 components for Inverter leakage of 22nm_MGK



Depicts that GMM can handle oblong clusters i.e. clusters of any shape and hence, are advantageous over k-means clustering which can only handle circular clusters effectively

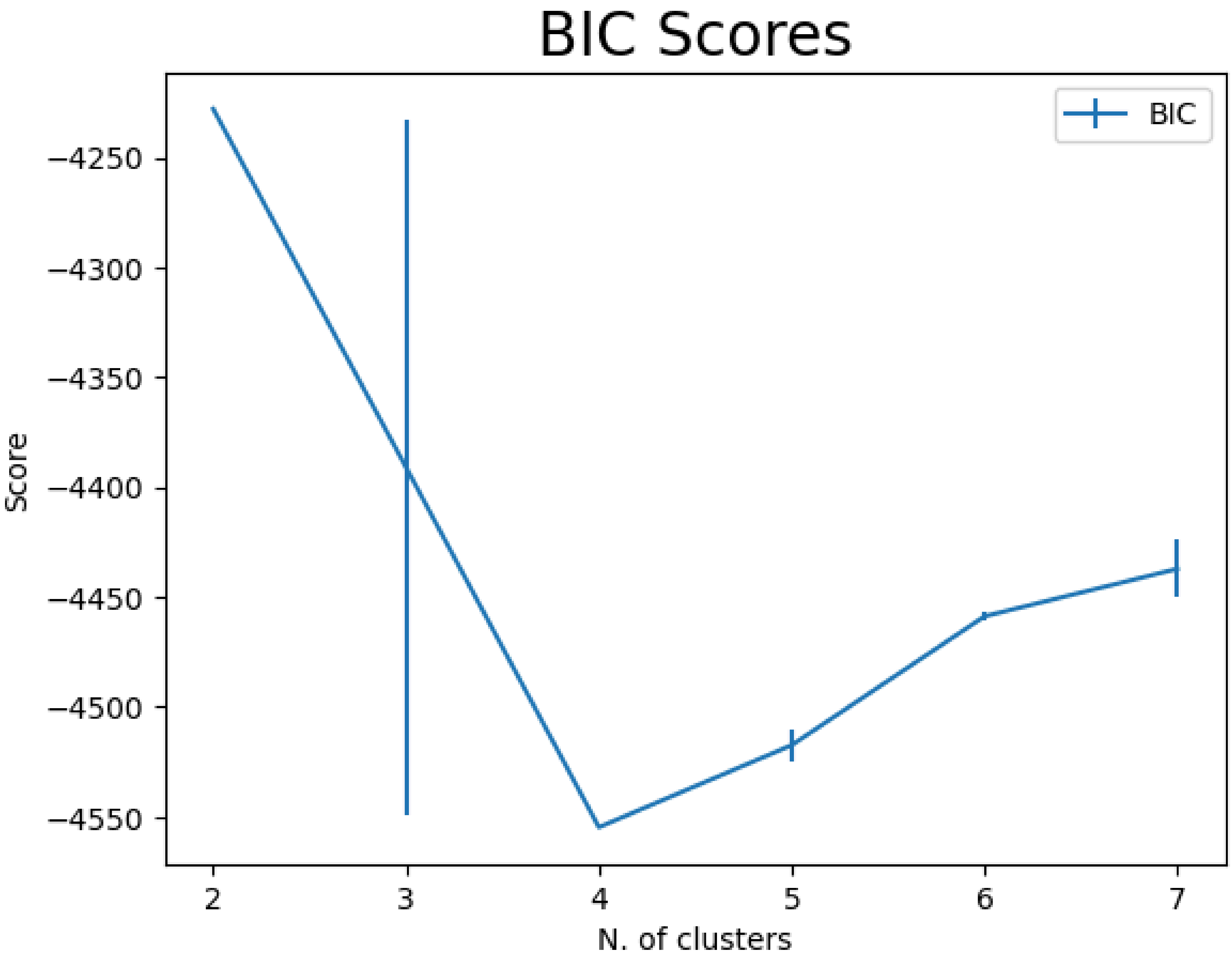
Gaussian Mixture Models



Silhouette score checks how much the clusters are compact and well separated. The more the score is near to one, the better the clustering

The lesser the JS-distance between the two GMMs, the more the GMMs agree on how to fit the data

BIC Score Analysis



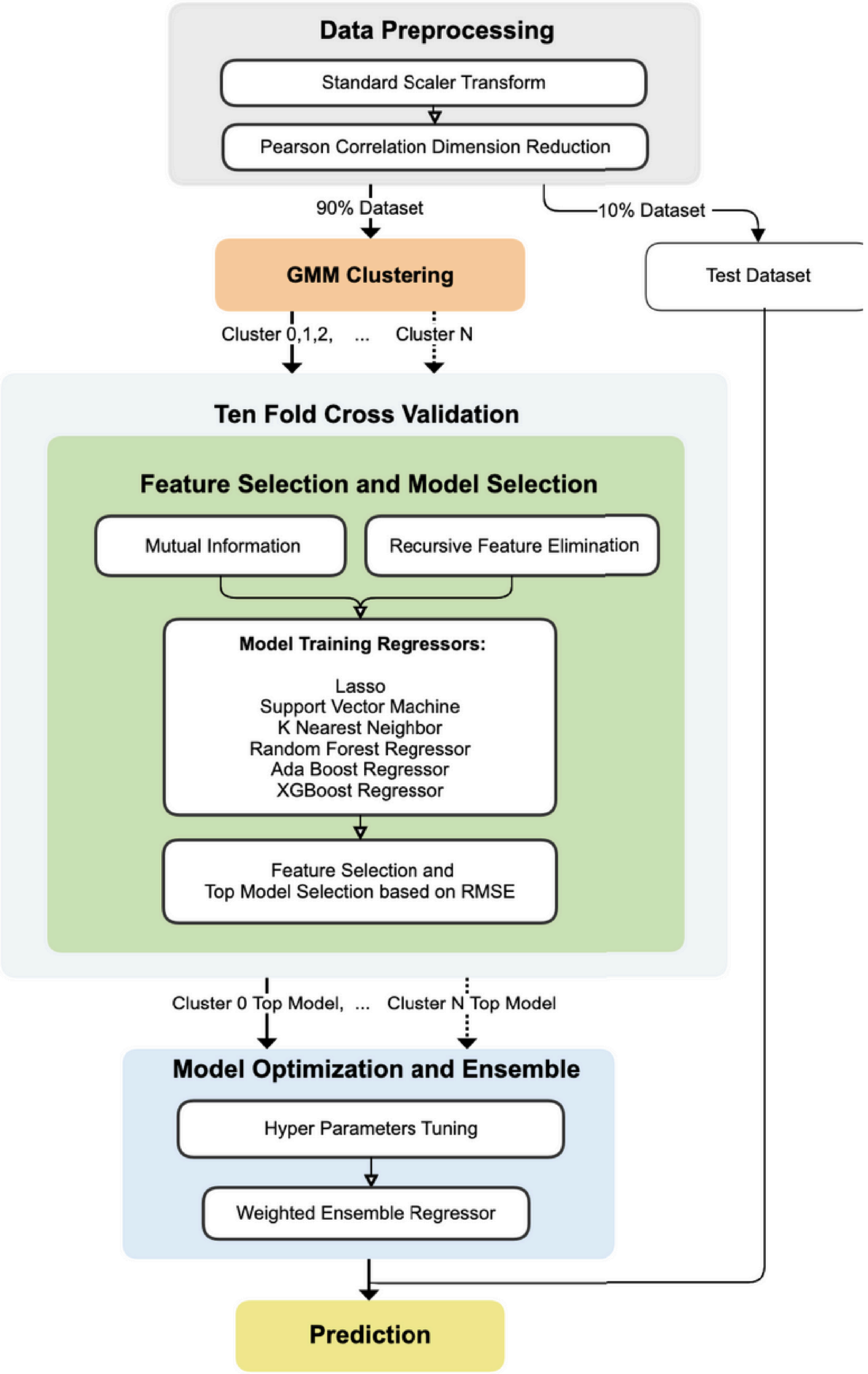
Bayesian **I**nformation **C**riterion gives us an estimation on how much is good the GMM in terms of predicting the data we actually have.

The **lower is the BIC, the better** is the model to actually predict the data we have

ML Framework

- We utilise an weighted ensemble method based approach to utilise the benifits of various models.
- It is a widely used technique.
- We have used the following base models.

ML Framework flowchart



Lasso Regression

- The Lasso regression is one type of linear regression model to solve high dimensional problems.
- It uses L1 regularization for feature selection which reduces model complexity and improves the prediction performance.

Support Vector Machine (SVR):

- Support vector machine (SVM) uses hyperplanes to maximize the separation between classes.
- Kernel tricks can be used to solve for both linear and non-linear problems.

K Nearest Neighbor Regressor(KNN):

- K Nearest Neighbor (KNN) is a simple and powerful algorithm.
- It classifies a new datapoint based on a similarity measure which is the distance function.
- The ease of interpretation and implementation of KNN makes the algorithm widely usable in pattern recognition and many other areas.

Random Forest Regressor (RFR):

- A Random Forest Regressor consists of a collection, or ensemble, of decision trees.
- Each decision tree is trained independently on a random subset of the training data and a random subset of the features.

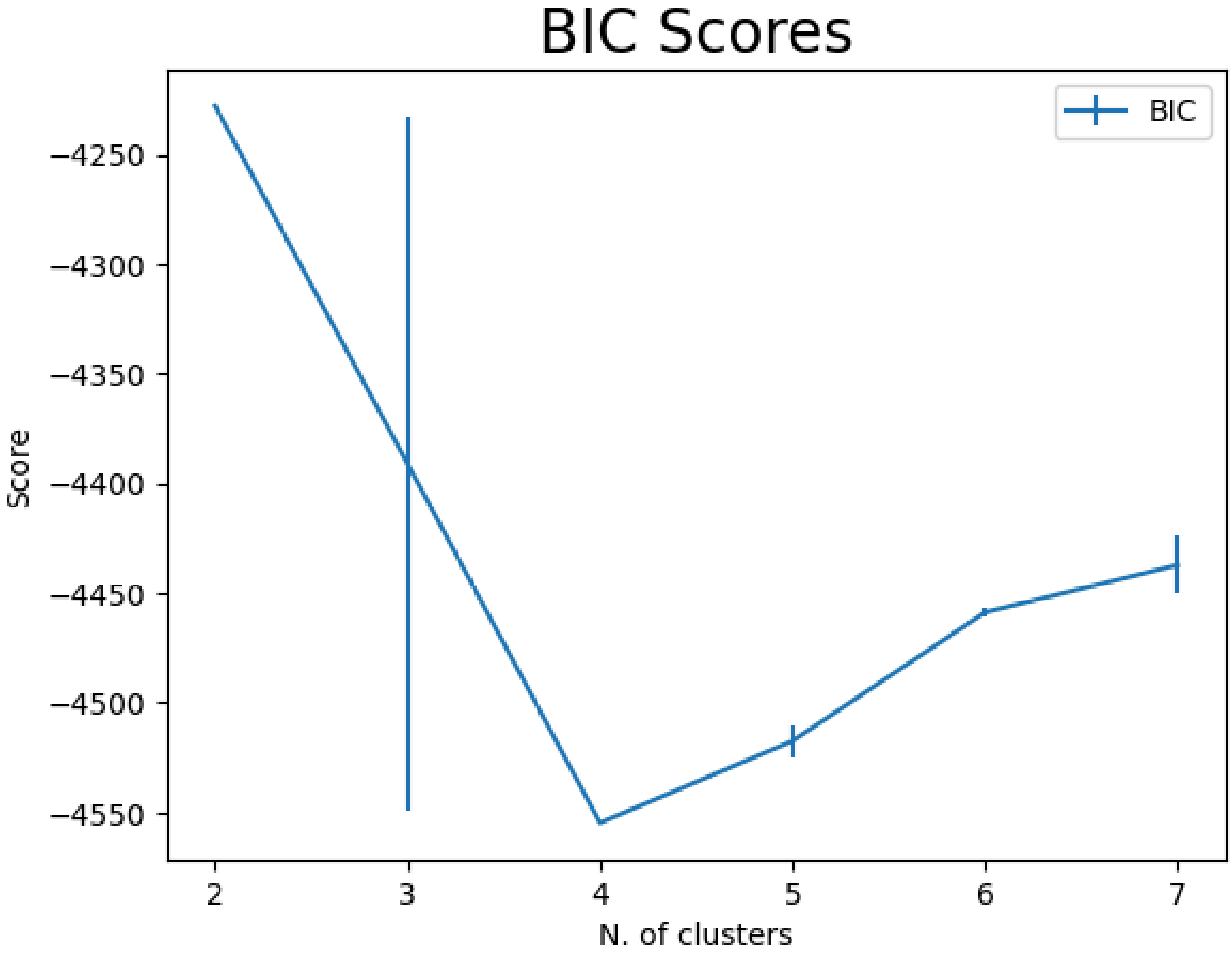
Adaptive Boosting (ADA):

- AdaBoost operates iteratively, where each weak learner is trained on a weighted version of the training data. Initially, all data points are given equal weights.
- During each iteration, the weights of misclassified data points are increased, while the weights of correctly classified data points are decreased.

XGBoost (XGB):

- Extreme Gradient Boosting
- Sequentially trains an ensemble of weak learners (typically decision trees) to improve predictive accuracy.
- It builds models in a stage-wise fashion, where each new model corrects the errors made by the previous ones.

BIC Scores

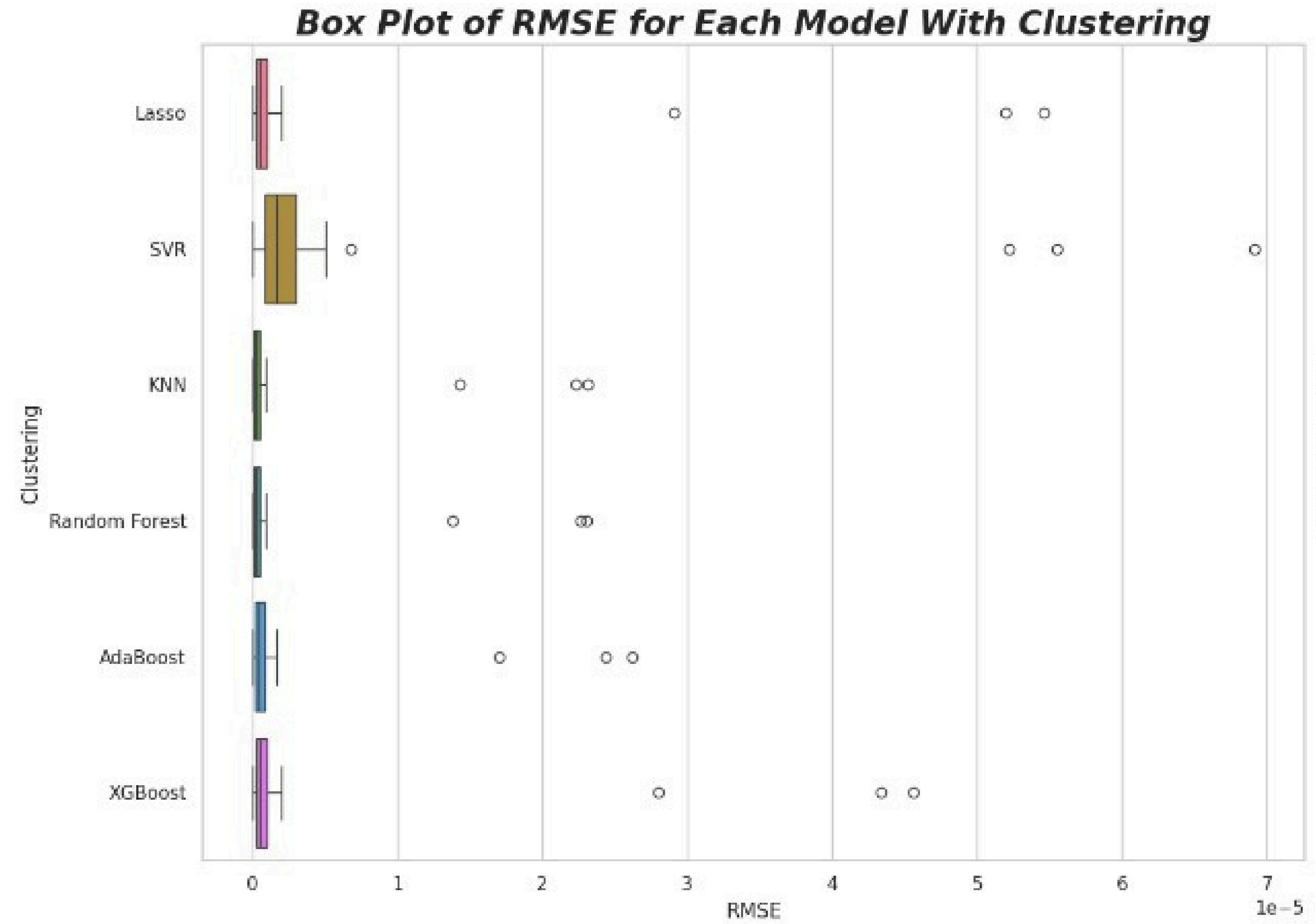


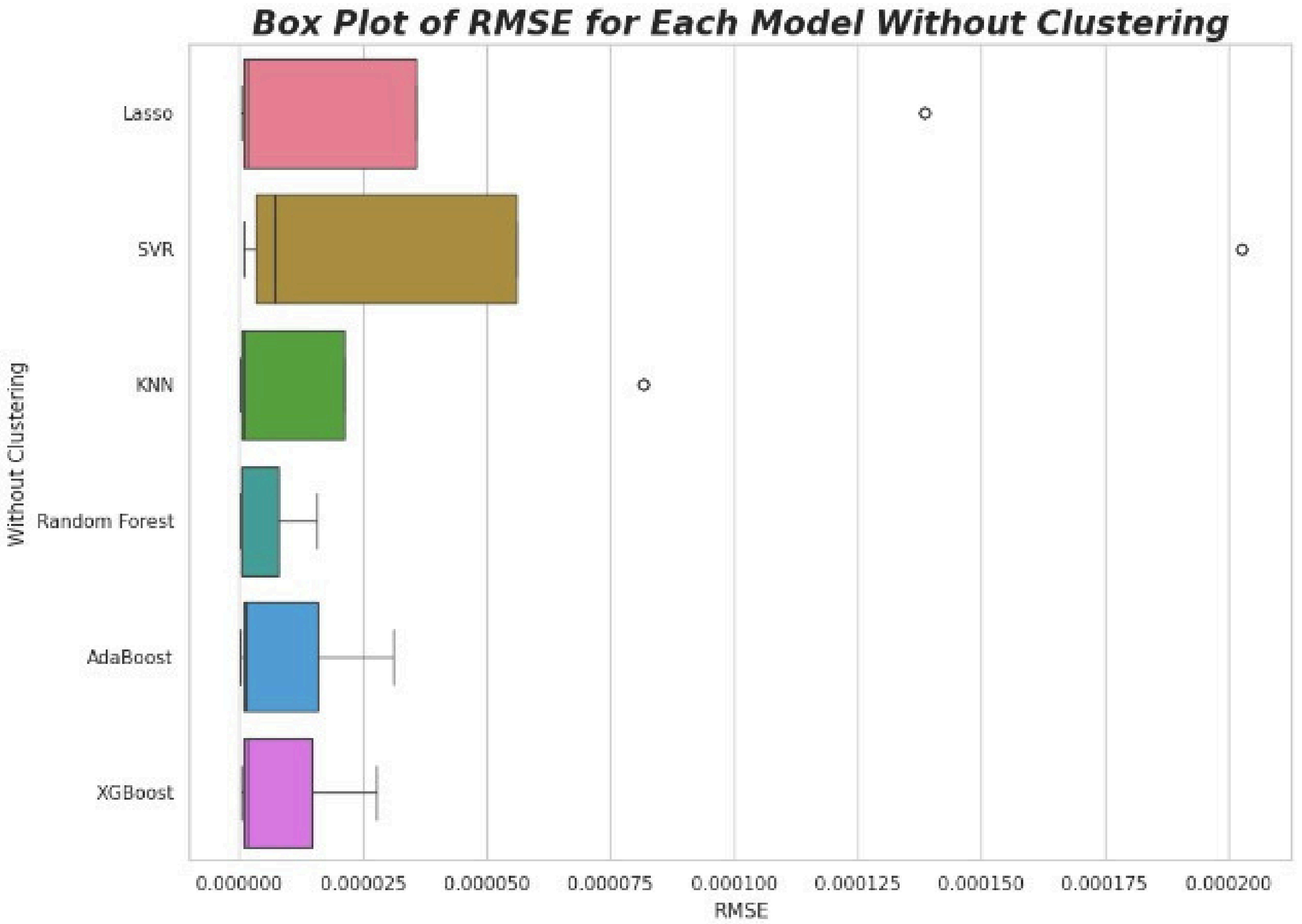
BIC Scores
BIC
Score
−4250
−4300
−4350
−4400
−4450
−4500
−4550
2
3
4
5
6
7
N. of clusters

Mean and variance of RMSE for the models used

	Model	Mean with Clustering	Std with Clustering	Mean without Clustering	Std without Clustering
0	AdaBoost	2.763856746384353e-06	6.975854876060281e-06	8.70025029625515e-06	1.3528421495915124e-05
1	KNN	2.342359914193569e-06	6.221539485457054e-06	2.084044246244653e-05	3.605294890967665e-05
2	Lasso	5.2419875087489685e-06	1.4332486885735182e-05	3.535839474333753e-05	6.112859476434277e-05
3	Random Forest	2.3291685689679736e-06	6.220144202867451e-06	4.309792553558716e-06	6.945643575362161e-06
4	SVR	7.71545977237079e-06	1.7936722346758145e-05	5.383504104314506e-05	8.822923871772546e-05
5	XGBoost	4.597797948448914e-06	1.2176004928889916e-05	7.96270467374659e-06	1.210871747382054e-05

Box Plots





Lower RMSE with lesser variance gives optimal results !!

ML Model Results

The R2 score obtained by our model is:

R2 score:0.9315904798240491

From the analysis above,we can conclude that the optimal No. of **GMM clusters** for each gate on basis of **BIC score**:

- **INVERTER:4**
- **XOR : 5**
- **AND_2 : 5**
- **AND_3: 10**
- **OR_2 : 5**

Conclusion

- Real process data sets were used to test and verify the robustness and validity of our procedure and the results indeed clearly prove that the **GMM clustering** with **weighted ensemble approach** provides for a much more improved prediction model for future root cause analysis and effective process optimization.
- The limitation for GMM is that it is **sensitive** to the **initial guesses** of the parameter values and can get stuck at **local minima**.
- The model selection and optimization method can be further explored to reduce the **training time** and computational resource.
- Besides, there is still scope for low yield root cause analysis flow enhancement.

References

[1] Abbas, Z., & Olivieri, M. (Year). Impact of technology scaling on leakage power in nano-scale bulk CMOS digital standard cells. Microelectronics Journal (2014)

[2] Z. Abbas, V. Genua and M. Olivieri, "A novel logic level calculation model for leakage currents in digital nano-CMOS circuits," 2011 7th Conference on Ph.D. Research in Microelectronics and Electronics, Madonna di Campiglio, Italy, 2011.

[3] Z. Abbas, A. Mastrandrea and M. Olivieri, "A Voltage-Based Leakage Current Calculation Scheme and its Application to Nanoscale MOSFET and FinFET Standard-Cell Designs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 22, no. 12, pp. 2549-2560, Dec. 2014

[4] Deepthi Amuru, Andleeb Zahra, Harsha V Vudumula, Pavan K Cherupally, Sushanth R Gurram, Amir Ahmad, Zia Abbas, “AI/ML algorithms and applications in VLSI design and technology”

International Institute
Of Information
Technology Hyderabad

THANKS

FOR YOUR ATTENTION

Mentor: Dr. Zia Abbas

TA: Mohammad Akhtar, Dheekshith Akula, Ritwik Goswami,
Chetan Mittal, Mehul Saraswat, Nouduru Venkata Raghavendra

Vedant Nipane, 2021102040
Pronoy Patra, 2021112019
Ashish Chokhani, 2021102016
Mayank Shivhare, 2021102026

07 May, 2024