Prediciton of Mosfet Parmeters using Machine Learning

\*Cancer Diagnosis Detection

Rahul Gupta  
Student, B. Tech., dept of Electronic And Communication Department  
National Institute Of Technology Silchar *Silchar (Assam), India*  
rahulguptaktl123@gmail.com

*Abstract*—For the first time ,a machine learning algorithm is presented to study the effect of source of variability on the perfomance of semiconductor nanodevices. This paper reports the possibility of alternative solution of devices simulationin order to optimize the source variation. As we are approaching to nano scale ,the mosfet application, devices dimensions are getting their scaling limit and it is affecting the gate leakage current ,drain induced barrier lowering etc. to a rise. It is becoming worsening the required characteristic and presentation of the devices. To overcome this some important changes in devices structures and materials will be required for continued transistor reduction and equivalent performance improvement. It is a paper of comparision of unstrained mosfet with performance of N and P channel mosfet with the introduction of machine learning for different conductance and channel length.

Keywords—Introduction, TCAD Simulation for devices Characterstic, ML,Surveys, techniques and methods, Results and Discussion, Conclusion.

1. Introduction

In the era of newly created technologies in the MOSFET industry,it becomes so significant to perform the analysis of the devices structures and their perfomance using the software or the simulation in highly computational ,complex and time consuming tool and structure like TCAD,LTSPICE before going to the fabrication process. An ML model is trained using these TCAD data and we found 1000-10000 TACD data can train an accurate machine .we show that without physical quantities,extraction,performing PCA is essential for the TCAD trained ML model to be applicableto analyze experimental devices characterstics .The physical devices characteristic estimated by ML model which show good agreement with experimental calculation and analysis. Our TCAD -ML framework shows great pledge to accelerate the development of new devices technologies with a important more efficient process of materials and device experimentation.

Machine learning ( ML) has recently gained increased attention for application in semiconductor industry manufacturing .On the other hand ,wafer-level devices variation analysis is crictical for the development of any nascent semiconductor technologies.The variation may be material-related and preocess-related. Today’s analysis practice mostly rely on the extensive devices and material characterization.Many of these characterization are destructive , prohibitively costly, and time-consuming to implement for every devices wafer. An ML assisted variation based on device electrical characteristic is highly desired to allow for more well organised material and devices experimentation.

2. TCAD Simulation for devices Characterstics

1. Problem overview

As projected by Moore's law, the number of transistors in a dense integrated circuit (IC) doubles about every two years [1]. Mathematically this meant the shrinking of transistor dimensions and other improvements which are enabled by advanced integrated circuits and improved semiconductor device fabrication technologies. Since the development of integrated circuits in the 1950s, scaling of metal-oxide semiconductor field-effect transistors (MOSFET) has continued as new technologies to extend complementary metal-oxide semiconductor (CMOS) down to ever smaller technology nodes emerge. However, the CMOS scaling has deviated from Moore's predictions and the scaling standards set by Dennard et al. due to fundadmental physical and technical limitations. As we get closer to atomic dimension, limitations like heat dissipation, leakage current, and channel length modulation have emerged as very significant problems that will inevitably slow down CMOS scaling. [2].

# 2. Limitations of planer Mosfet

This two-dimensional potential results in the degradation of the threshold behavior and introduces the Short Channel Effects:

1. Channel Length Modulation:

Channel length modulation in a MOSFET is brought on by the widening of the depletion layer at the drain as the drain voltage increases. As a result, the drain current is increased and the channel length is shortened.

2. Threshold Voltage variation with channel length Modulation:

In long channel MOSFETs, the channel regulates the charge and carries the majority of it. The threshold voltage starts to decline as we switch to short channels because the source and drain are now also adding to the charge. The threshold voltage (VTH) decreases as a result of the gate in this region having to support less charge. This phenomenon is known as the charge-sharing effect.

A few of the emerging transistors include:

1. FinFET

2. Nanowire FET

3. Nanosheet FET

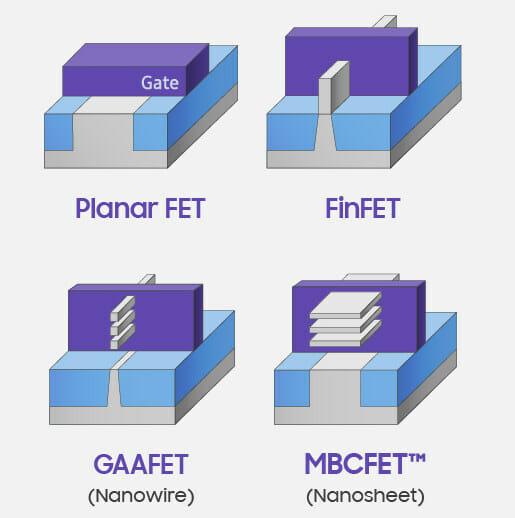


Figure 2: Structure of Nanowire (GAAFET) [2]

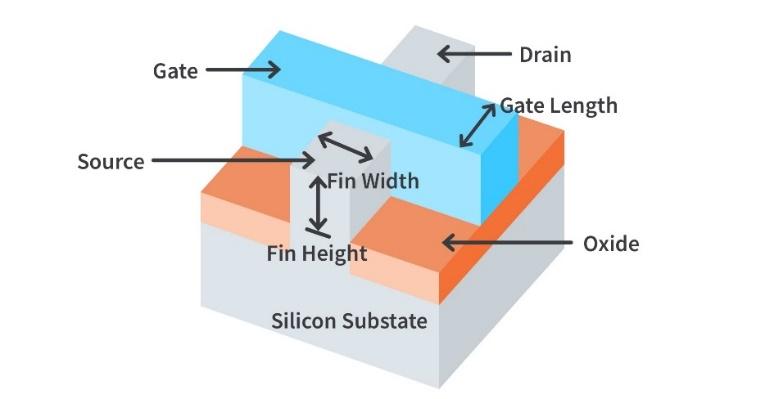


Figure 1: Structure of FinFET [2]

FinFETs are preferred over conventional MOSFETs for a variety of reasons. To increase computational power, computational density must also increase. To accomplish this, more transistors are needed, which results in larger chips. However, it is essential to maintain the area's consistency for practical reasons. The size of the transistor can be decreased as one method of increasing processing power. The ability of the gate electrode to regulate the flow of current in the channel region, however, is diminished when the transistor's size decreases due to the close proximity of the drain and source. Planar MOSFETs exhibit undesirable short-channel effects as a result [3]

Below 28 nm, the leakage current becomes high and the transistor becomes ineffective as the gate length (Lg) decreases below 90 nm. Suppressing the off-state leakage is crucial as the gate length is reduced.

In summary, FinFET devices outperform conventional MOSFET technology in terms of short-channel behavior, switching times, and current density.

3. Why Nanofet wire?

New semiconductor technology must be developed immediately to address problems with cost, speed, density, reliability, and power dissipation. It has taken a lot of work, but among all the contenders, nanowire FET that replaces current silicon-based technology has emerged as one of the most promising answers for continue CMOS scaling.

The nanowires have a larger surface-to-volume ratio and a smaller channel than planar devices made of bulk materials. Additionally, the nanowire FET's gate-surrounding or gate-all-around (GAA) shape allows for good electrostatic gate control over the nanowire channel. In addition, the GAA nanowire transistors allow the optimum short-channel control

for CMOS device scaling while taking into account quantum confinement effects and atomic-scale scattering. The FinFET, which dominates the current at 10 nm or even 7 nm technology nodes, displays strong similarities to the GAA horizontal nanowire FET architecture. To improve scalability beyond the restrictions imposed by the FinFET technology with significantly less complexity than alternative scaling approaches, GAA FETs are thus extremely interesting alternatives in the sub-7 nm nodes [4].

3. ML surveys, techniques and methods

ML algorithms are mainly investigated to decrease computational costs by forecasting the outcome based on the experimental data used to train the ML model. ML has been used in many different research fields, such as biomedicine, signal processing, and the semiconductor industry, due to its viability. Also, complicated processing procedures allow ML to work more carefully in the semiconductor manufacturing industry. In addition, ML uses a variety of statistical models like k-NNs, decision trees, random forests, and linear regression. However, because they maximize the distance between various points, k-NN and linear regression are frequently used for linear dependencies

a. Linear Regression:

Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. It is a fundamental technique in the field of statistics and machine learning, particularly in the context of supervised learning.

The basic idea behind linear regression is to find the best-fitting linear equation that describes the relationship between the independent variable(s) and the dependent variable.

b. K Nearest Neighbour Model:

K-Nearest neighbor (KNN) is a supervised learning technique that determines the similarity of a new data point with available cases and puts the new case into the category which is most similar to the available categories.

By choosing K, the number of nearby observations to use in the algorithm is fixed. The algorithm aims to locate all of these K closest neighbours around an unknown new data point and assigns it to the class with the highest closest neighbours. The Euclidean distance between points is calculated to determine the closeness of points.

c. Decision Tree Model:

Decision tree model is a supervised learning technique that uses a decision tree classifier with features of a dataset as the internal nodes, decision rules represented by the branches and the outcome as the leaf nodes. It gives the graphical representation for finding all possible solutions to a decision based on given conditions.

For predicting the class of the given dataset, the algorithm starts from the root node of the tree which contains the entire dataset. This algorithm compares the values of root attribute with the real dataset attribute and, based on the comparison, finds the best attributes in the dataset using the Attribute Selection Measure (ASM) and makes internal nodes. Every internal node that jumps to the subsequent node after the branch and follows it has a decision made for it. The algorithm continues by performing another comparison between the attribute value for the subsequent node and those of the other sub-nodes. This continues until it reaches the leaf node of the tree.

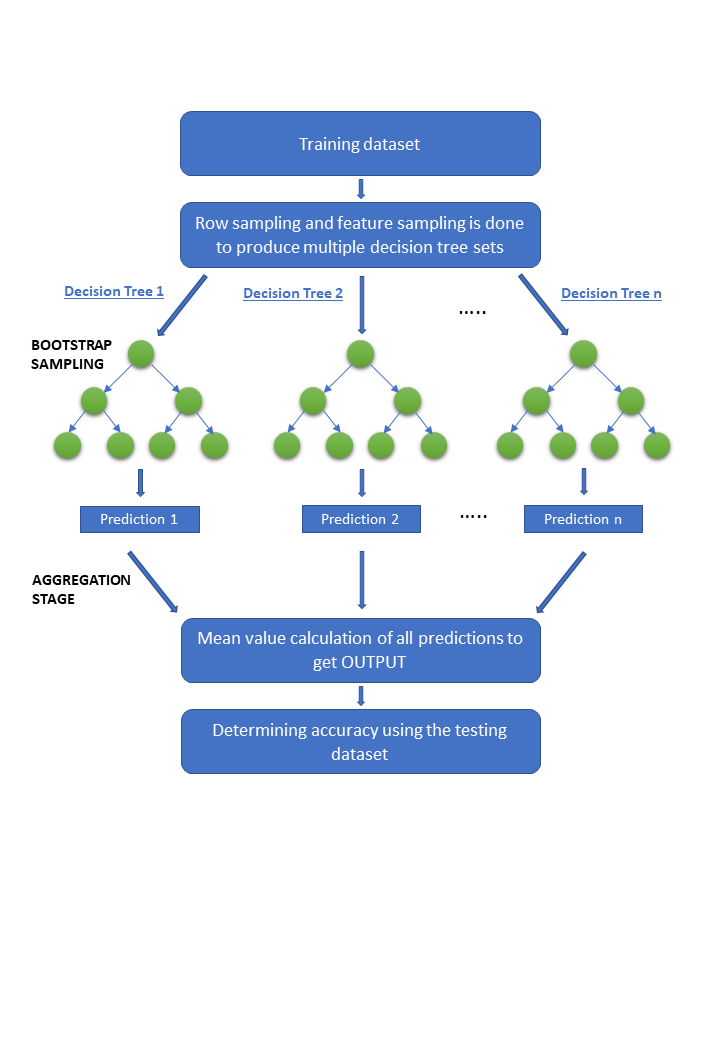


Figure 3: Pictorial representation of Random Forest Regressor (RFR) Model

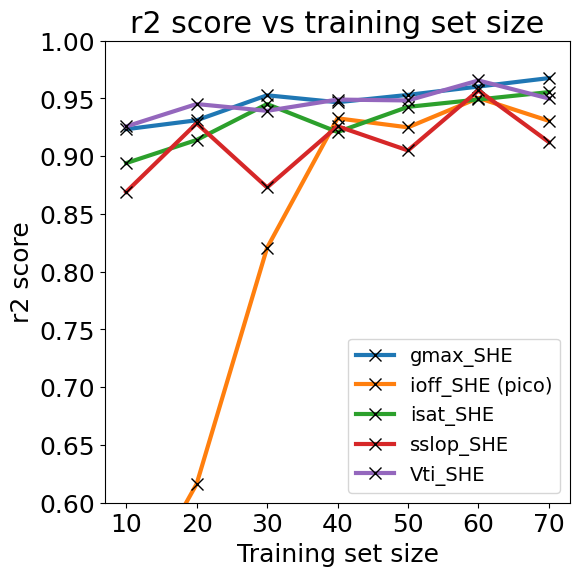
d. Random Forest Regressor Model:

The Random Forest Regressor (RFR) model is a supervised ML algorithm that takes the average of predicted values of multiple decision trees to provide accurate predictions for new data.The input dataset is of m \* n dimensions where m is the total number of data samples and n is size of input features. The RFR model is constructed by creating multiple decision trees in which a subset m’ of the training data (m’ < m) is randomly sampled with replacement (bootstrapped) to create a new dataset for each tree. For each node of the tree, randomly n’ variables (n’ < n) are chosen which forms the basis for the decision at that node. The best split is calculated based on these n’ variables in the training set. Additionally, at each split in the tree, only a random subset of features is considered. This process is repeated until a specified number of trees is grown, resulting in an ensemble of decision trees that are used for predictions. An unbiased average of the resultant predictions from the decision trees is considered to predict output for new input variables. Fig 12 shows the flowchart for the described methodology.

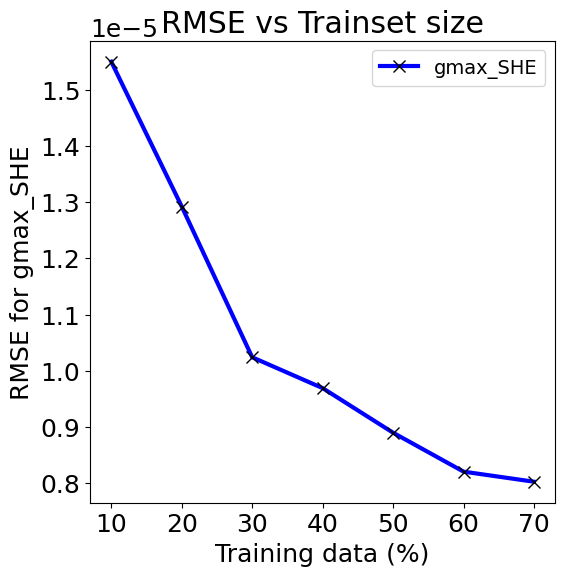
Root Mean Square Error:

The root mean square error (RMSE) function is calculated using the evaluated output by the trained model and the simulated test dataset. For achieving the convergence of error function or to achieve higher accuracy RFR model exhibits ‘low bias’ by populating the trees to its maximum depth and the ‘high variance’ data is converted into ‘low variance’ output by growing each tree on a bootstrap subset of the training dataset and aggregating individual tree predictions to predict the output. The decision trees can be optimised by optimising the hyperparameters such as ‘n\_estimator’ which sets the number of decision trees the model contains, ‘max\_depth’ which sets the maximum depth of a decision tree, ‘min\_samples\_split’ which gives the minimum sample number to split a node and adjusting a few other hyperparameters.

The decision trees are optimized by optimizing the hyperparameters such as ‘no\_of\_estimators’ which sets the number of decision trees the model contains, ‘maxi\_depth which sets the maximum depth of a decision tree,‘mini\_samples\_split’ which gives the minimum sample number to split a node and adjusting a few other hyperparameters

Figure show that variation of score of training data size for

various output parameters for random forest regression model



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| parametersz | **Gmax\_**  **SHE** | Ioff\_  **SHE** | Isat\_  **SHE** | **SSlop\_SHE** | Vti\_  **SHE** |
| no\_of\_  estimators | 50 | 55 | 100 | 50 | 100 |
| Maxi\_  depth | 90 | 90 | 100 | 85 | 100 |
| Mini\_  samples\_  split | 2 | 2 | 6 | 2 | 2 |
| Min\_  samples\_  per\_leaf | 1 | 2 | 2 | 1 | 2 |

figure show that how Rmse is decreasing with increasing

dataset for gmax\_SHE for random forest regression model

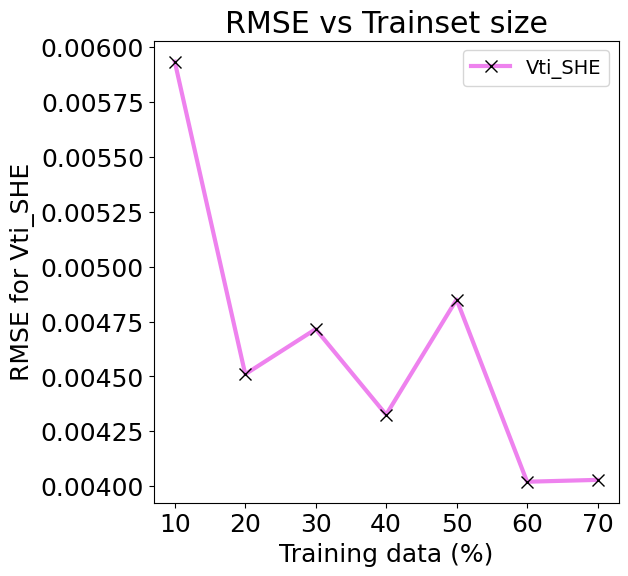
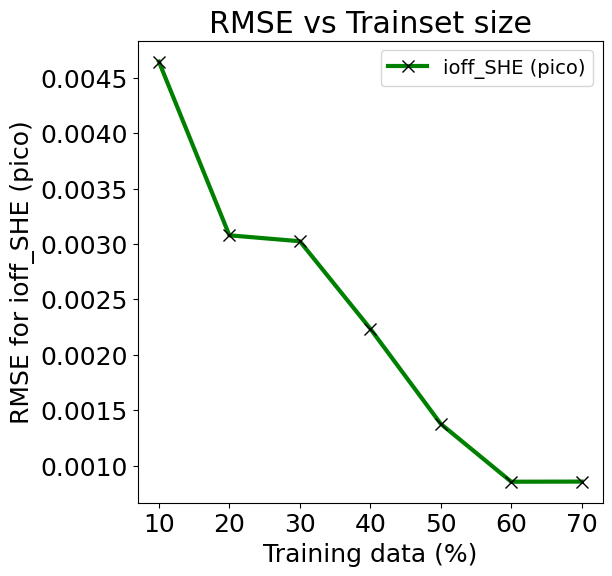
Figure shows the how Rmse error degrades with the increaseing training datasets of Vti\_SHE for random forest regression model

figure show that degradation of Rmse error with increasing in training data size of Ioff\_

SHE for random forest regression model

Figure 1: Pictorial representation of Random Forest Regressor (RFR) Model

Some of the advantages of Random forest regressor model are:

1) It is able to extract significant features even from a small dataset

2) The convergence of RFR is faster than the other ML algorithms

3) It is capable of capturing the non linear dependencies as well.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training Algorithms | Gmax\_  SHE  (10^-5) | Ioff\_  SHE | Isat\_  SHE | sslop\_SHE | Vti\_  SHE |
| Linear Regression | 1.55 | 41 | 0.61 | 84 | 644 |
| K Nearest Neighbour Regression | .948 | 23 | 0.64 | 39 | 395 |
| Decision Tree Regression | 1.54 | 20 | 0.78 | 41 | 529 |
| Random Forest Regressor | 1.14 | 75 | 0.55 | 36 | 442 |

The models proposed in the paper are implemented using the Sklearn library of Python. The dataset obtained is used to train these models and predict future values of diagosis that person has a Malignant and Benign.where malignant means 1 and person has a cancer and benign means 0 and person has not any cancer.

The decision trees are optimized by optimizing the hyperparameters such as ‘no\_of\_estimators’ which sets the number of decision trees the model contains, ‘maxi\_depth’ which sets the maximum depth of a decision tree, ‘mini\_samples\_split’ which gives the minimum sample number to split a node and adjusting a few other hyperparameters.

While training the Machine Learning model, the Root Mean Square Error is calculated. The mathematical formula for calculating the RMSE value for the training or the testing data of the model is given below:

(1)

where N represents the number of samples under consideration and ytrue and ypred represent the actual and predicted values for each of the data point.

Another evaluation parameter considered is the R2 score. The coefficient of determination, or R2, is a measure that provides information about the goodness of fit of a model. It is calculated as given ,

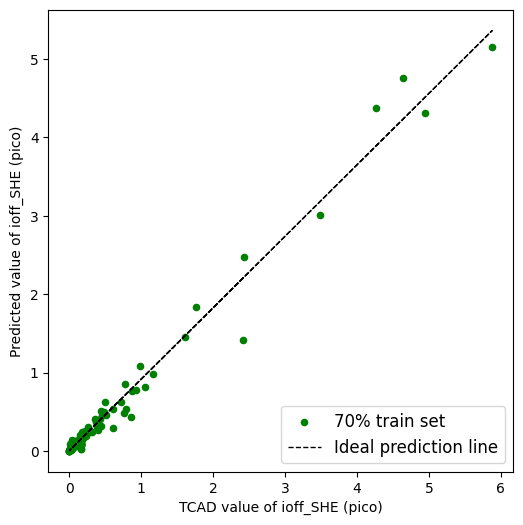
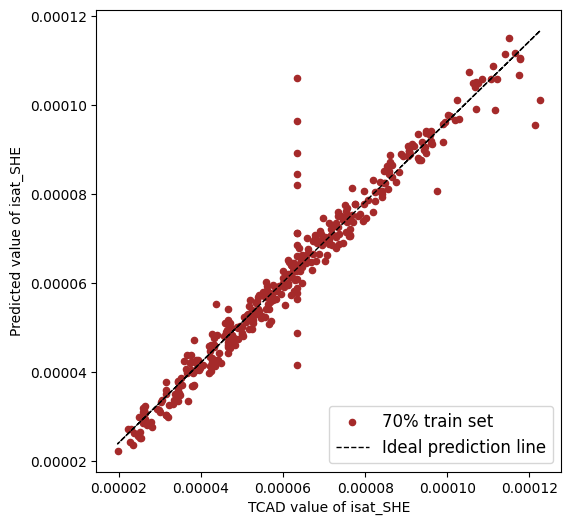
(2)

(3)

The sum of the squares of the data's variance from the mean is the total sum of squares, and the sum squared regression is the sum of the squared residuals. Since it is a percentage, it can only accept values in the range of 0 and 1.

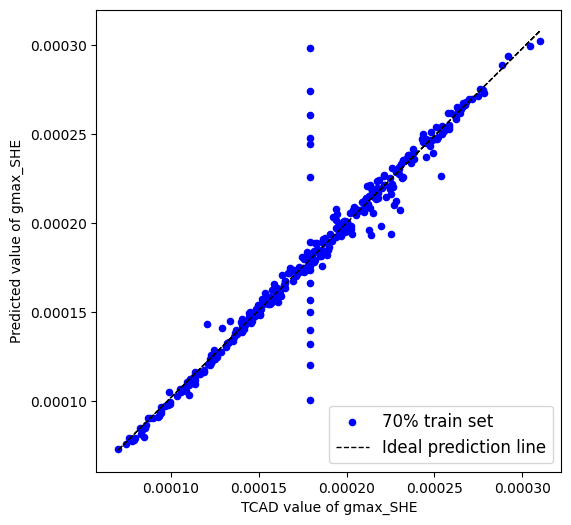
4. Results And Discussions

In the above section we have discussed about the cancer diagnosis and how we know the person has cancer or not from the variou input ideal range. The values of the hyper parameters for the Random Forest Regression algorithm under testing are shown in table 1. Table 2 below shows the best RMSE values obtained for various models for each of the output parameters. The RMSE error values of the RF model are lesser than the errors for other models.

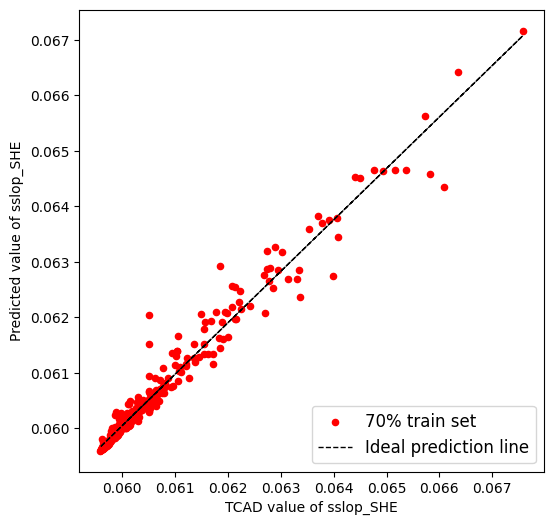
Fig2.Optimal RMSE values of output values for mosfet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Training Algorithms | Gmax\_  SHE | Ioff\_  SHE | Isat\_  SHE | sslop\_SHE | Vti\_  SHE |
| Linear Regression | 0.912 | 0.241 | 0.924 | .602 | .899 |
| K Nearest Neighbour Regression | .945 | 0.929 | 0.934 | .913 | .966 |
| Decision Tree Regression | .907 | .867 | 0.866 | .880 | .931 |
| Random Forest Regressor | .968 | .963 | 0..939 | .923 | .951 |

Fig2.Optimal Score of output values for mosfet .

Fig4. shows the confusion matrix which is plot between the true values and predicted values this matrix showthat how much our model’s predicted values are accurate according to the true values and matrix also tell the more the darker colour implies low prediction values lies there and more the lighter

colour is showing the higher the acurate prediction.For first box tell the there is a only one value for which truth label is true and predicted label is false and second box say the there are 46 values for which truth label and predicted label is true, and third box say that there are 59 values for which there is truth and predicted labels are false mean when the patient has no cancer,our model also predicted that same result as we have.



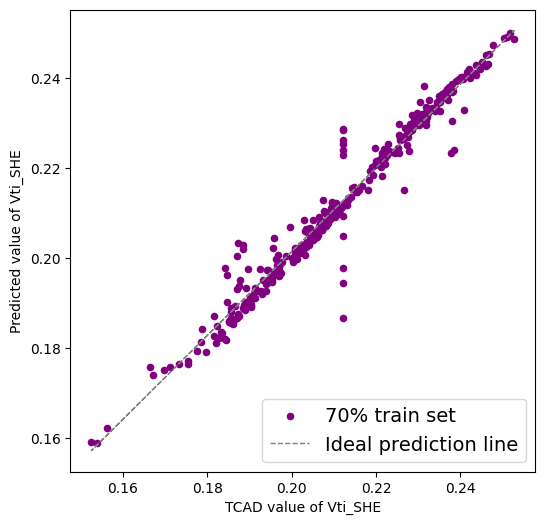


Fig3.scatter plot for predicted value and true value of output values

The above given figures are scatter plot for the differnet model of regression and classification techniques between the true values and predicted values of diagnosis.First one is scatter plot of linear regression and we easily see that the output has only two values 0’s and 1’s.Second one is the scatter plot between true values and predicted values logistic regression is a classification technique so we get this type of scatter plot in which we can see only 0 and 1 values at four points.Third one is Decision Tree Regression which is also a classification technique and we get the same type of scatter plot as same as logistic regression.Fourth is Random Forest Regression in which we get the scatter plot as same of linear regression.Last one is knn which is also a classfication technique

The above figure show the clustering of two ranges of output values 0 and 1.These are the values on graph that the people has cancer or not in which blue show that patient has not any cancer and orange show that the patient has cancer.The two black points between the clusters show the center of both the cluster .

5. Conclusion

The Model that we have discussed is about cancer detection that person has a cancer which is represent by malignment and person hasn’t any cancer which is represent by benign. From this model.The combinantio of both diagnosis and machine learning result used in triaining and prediction.The five regression techniques like Linear Regression,Logistic Regression,Decision Tree forest,Random Forest Regression and K Nearest Neifgbour Regression are explored to observe their ability of prediction with the given datasets.The perfomance of these algorithms are evualated by using the root mean squared error(RMSE) and r2 score as perfomance parameter.The Logistic Regression algorithm show the better prediction accuracy with 94% score.

This work can be extended by generating a few more data and exploring the new deep learning techniques to perform the prediction of output.

END

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**