Crack depth prediction of Abaqus model by Machine Learning tools

Report submitted to

***Indian Institute of Technology Kharagpur***

***For the fulfillment of the degree of***

**B.TECH PROJECT-II**

By

**Chinta Krishna Mourya**

Roll No. 17NA10011

***Under the guidance of***

**Prof. Mohammed Rabius Sunny**

Department of [Aerospace Engineering](http://www.ae.iitkgp.ernet.in/)

Indian Institute of Technology Kharagpur

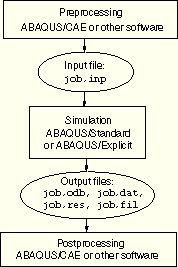
|  |  |  |
| --- | --- | --- |
| Serial No. | Contents | Page No. |
| 1 | Introduction | 3 |
| 2 | Data Acquisition | 4 |
| 3 | Flowchart | 5 |
| 4 | Random Forest Regression | 6 |
| 5 | Results | 8 |
| 6 | 1)Discussion  2) Conclusion | 9  10 |
| 7 | References | 10 |

# List of Contents

# Introduction :

ABAQUS is a suite of powerful engineering simulation programs, based on the finite element method, that can solve problems ranging from relatively simple linear analyses to the most challenging nonlinear simulations and helps in computer aided engineering.

A complete ABAQUS analysis usually consists of three distinct stages: preprocessing, simulation, and postprocessing. These three stages are linked together by files as shown below:



Preprocessing (ABAQUS/CAE)

In this stage you must define the model of the physical problem and create an ABAQUS input file. The model is usually created graphically using ABAQUS/CAE or another preprocessor, although the ABAQUS input file for a simple analysis can be created directly using a text editor.

Simulation (ABAQUS/Standard or ABAQUS/Explicit)

The simulation, which normally is run as a background process, is the stage in which ABAQUS/Standard or ABAQUS/Explicit solves the numerical problem defined in the model. Examples of output from a stress analysis include displacements and stresses that are stored in binary files ready for postprocessing. Depending on the complexity of the problem being analyzed and the power of the computer being used, it may take anywhere from seconds to days to complete an analysis run.

Postprocessing (ABAQUS/CAE)

You can evaluate the results once the simulation has been completed and the displacements, stresses, or other fundamental variables have been calculated. The evaluation is generally done interactively

using the Visualization module of ABAQUS/CAE or another postprocessor. The Visualization module, which reads the neutral binary output database file, has a variety of options for displaying the results, including color contour plots, animations, deformed shape plots, and X–Y plots.

What is Machine Learning model ?

A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data.

Once you have trained the model, you can use it to reason over data that it hasn't seen before, and make predictions about those data.

# Process :

The model I got has some crack. I've changed the crack positions and created one or multiple cracks at a time of different depths, analysed them to get strain values and then in the output, a data file containing printed output of the model and history definition generated by the analysis input file processor and, in Abaqus/Standard, printed output of results written during the analysis run;an output database file containing results for postprocessing with the Visualization module of Abaqus/CAE (Abaqus/Viewer) and, in Abaqus/Standard, diagnostic information;

The dataset collected from abaqus analysis gives the strain of the model as electric potential values and simulation time for a particular depth as well as for multiple depths that we have changed in the crack of the model. This output from abaqus is input for training the machine learning model to predict the crack depth.

With this output we train the machine learning model by **using Random Forest Regression algorithm & Linear Regression algorithm.**

**Flow Chart**

# Random Forest Regression:

**Regression analysis** consists of a set of machine learning methods that allow us to predict a continuous outcome variable (y) based on the value of one or multiple predictor variables (x).

# Random Forest?

Random forest is a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of overcoming over-fitting problem of individual decision tree.  
  
In other words, random forests are an ensemble learning method for classification and regression that operate by constructing a lot of decision trees at training time and outputting the class that is the mode of the classes output by individual trees.

***Random Forest Regression***

***The Random Forest is one of the most effective machine learning models for predictive analytics, making it an industrial workhorse for machine learning.***

***Background***

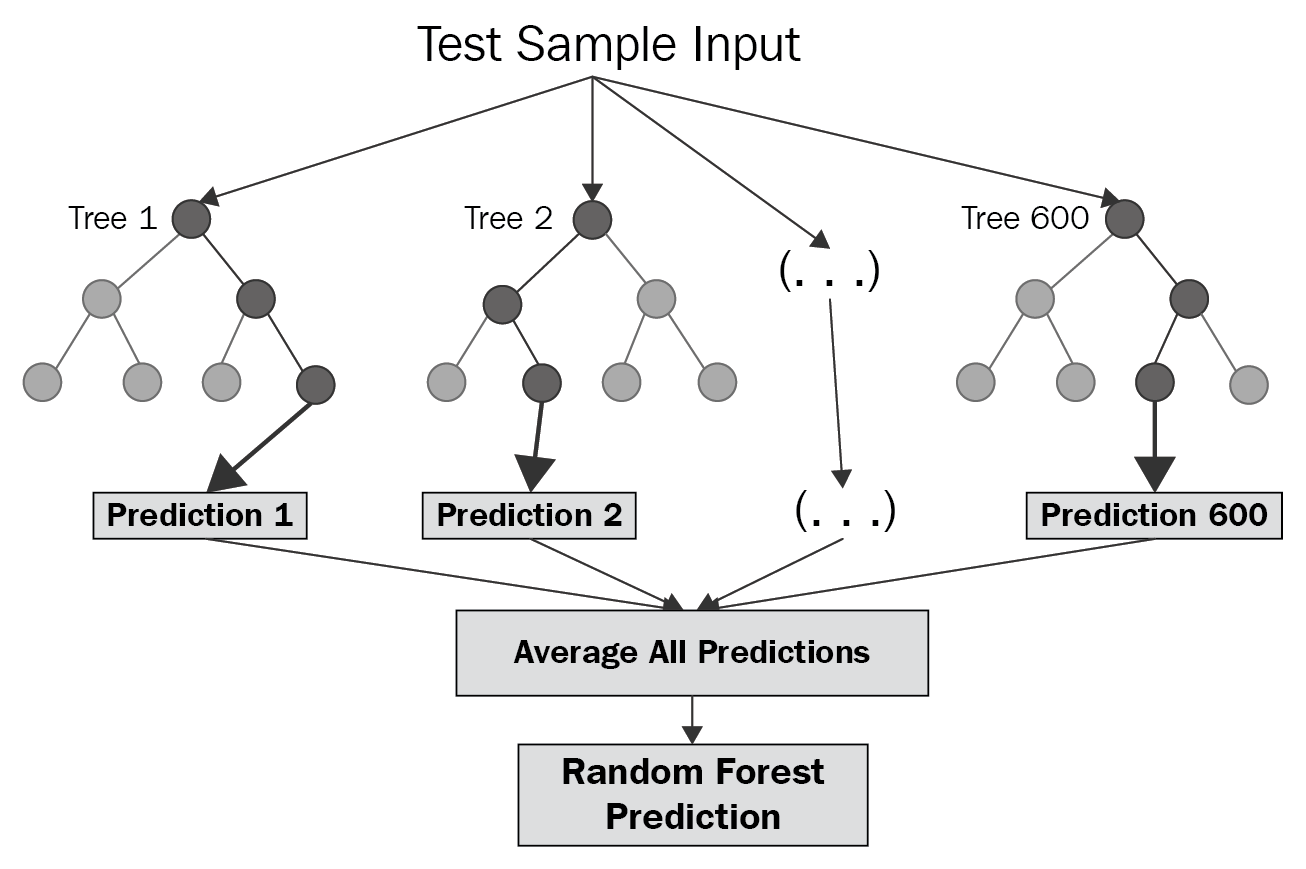
***The random forest model is a type of additive model that makes predictions by combining decisions from a sequence of base models. More formally we can write this class of models as:g(x)=f0(x)+f1(x)+f2(x)+...***

***where the final model g is the sum of simple base models fi***

# Random in 'Random Forest'?

'Random' refers to mainly two process - 1. random observations to grow each tree and 2. random variables selected for splitting at each node. See the detailed explanation in the previous section.  
  
**Important Point :**

*Random Forest does not require split sampling method to assess accuracy of the model. It performs internal validation as 2-3rd of available training data is used to grow each tree and the remaining one-third portion of training data always used to calculate out-of bag error to assess model performance.*



# Model Prediction Error:

.The [most ***popular metric***s for comparing regression models](http://www.sthda.com/english/articles/38-regression-model-validation/158-regression-model-accuracy-metrics-r-square-aic-bic-cp-and-more/) :

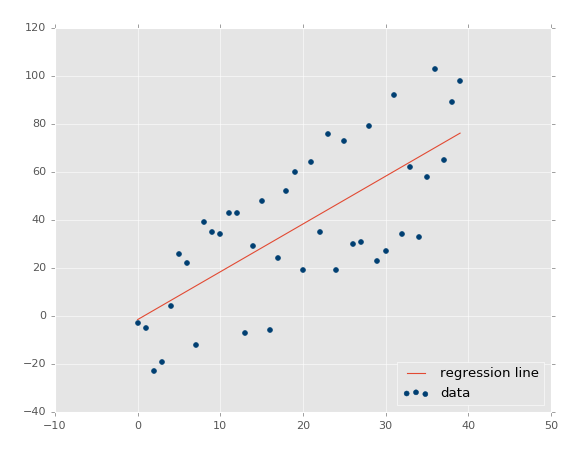
* **Root Mean Squared Error**, which measures the model prediction error. It corresponds to the average difference between the observed known values of the outcome and the predicted value by the model. RMSE is an absolute measure of fit. As the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained variance, and has the useful property of being in the same units as the response variable. ***Lower values of RMSE indicate better fit***. RMSE is a good measure of how accurately the model predicts the response.

### ****Linear Regression:****

The term “linearity” in algebra refers to a linear relationship between two or more variables. If we draw this relationship in a two-dimensional space (between two variables), we get a straight line.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression. If we plot the independent variable (x) on the x-axis and dependent variable (y) on the y-axis, linear regression gives us a straight line that best fits the data points, as shown in the figure below.

We know that the equation of a straight line is basically:



# Results :

The equation of the above line is :

**Y= mx + b**

Where b is the intercept and m is the slope of the line. So basically, the linear regression algorithm gives us the most optimal value for the intercept and the slope (in two dimensions). The y and x variables remain the same, since they are the data features and cannot be changed. The values that we can control are the intercept(b) and slope(m). There can be multiple straight lines depending upon the values of intercept and slope. Basically what the linear regression algorithm does is it fits multiple lines on the data points and returns the line that results in the least error.

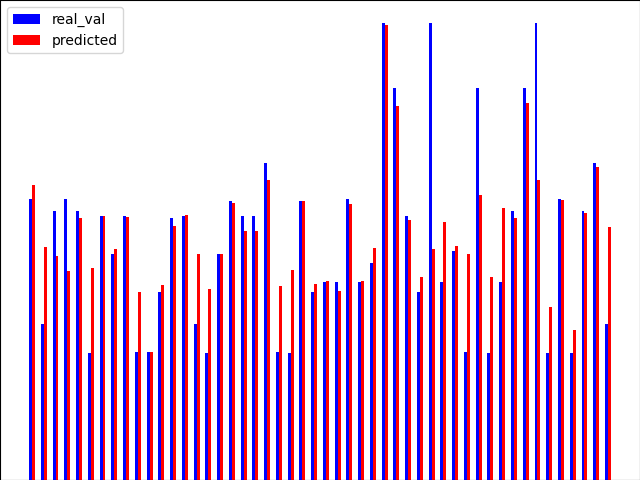
This same concept can be extended to cases where there are more than two variables. This is called multiple linear regression. For instance, consider a scenario where you have to predict the price of the house based upon its area, number of bedrooms, the average income of the people in the area, the age of the house, and so on. In this case, the dependent variable(target variable) is dependent upon several independent variables. A regression model involving multiple variables can be represented as

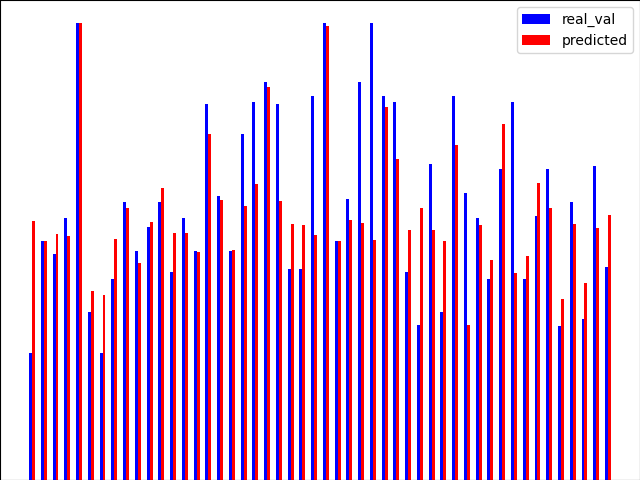
**y = b*0* + m*1*b*1* + m*2*b*2* + m*3*b*3* + ... m*n*b*n***

The Graphs shows comparison between real depths (**blue**) and

predicted depths(**red**) of testing set (various datasets) in training the

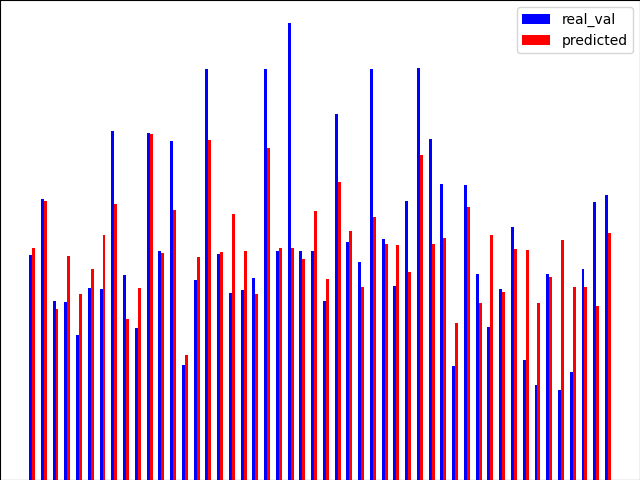
model by the RANDOM FOREST and RMS errors as follows.

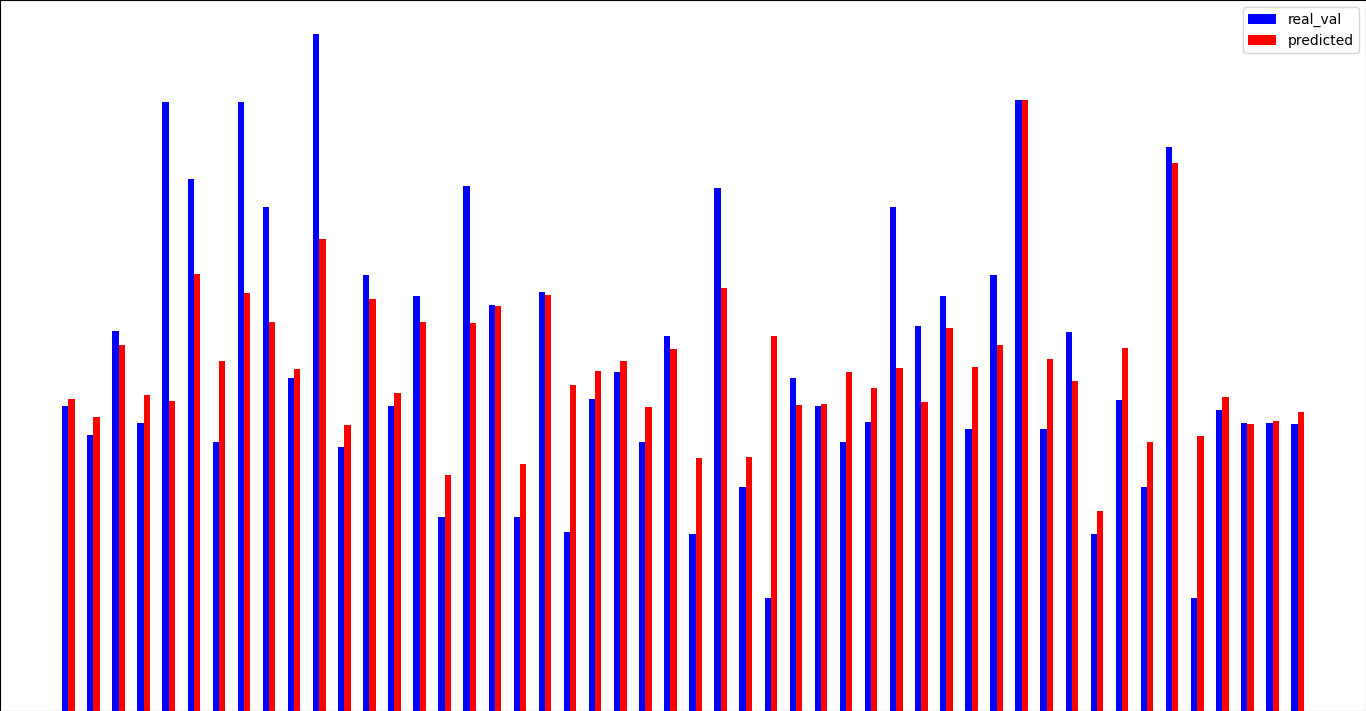




No. of Datasets: **49** & RMS error: **0.0007195**

No. of Datasets: **25** & RMS error: **0.0007631**





No. of Datasets: **97** & RMS error: **0.0007535**

No. of Datasets: **73** & RMS error: **0.0007543**

**The plots are drawn using Linear Regression algorithm :**

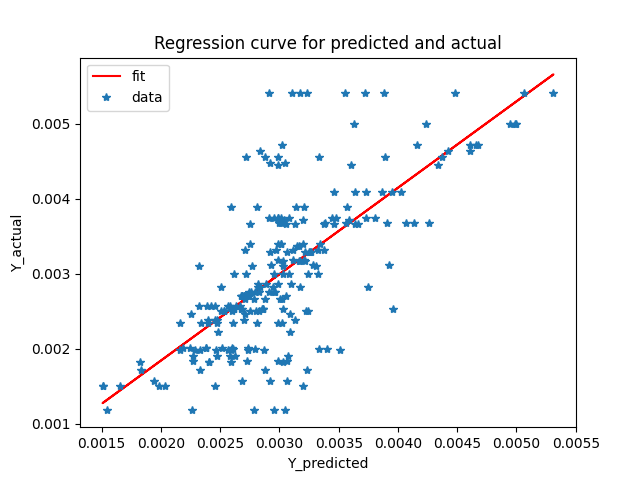
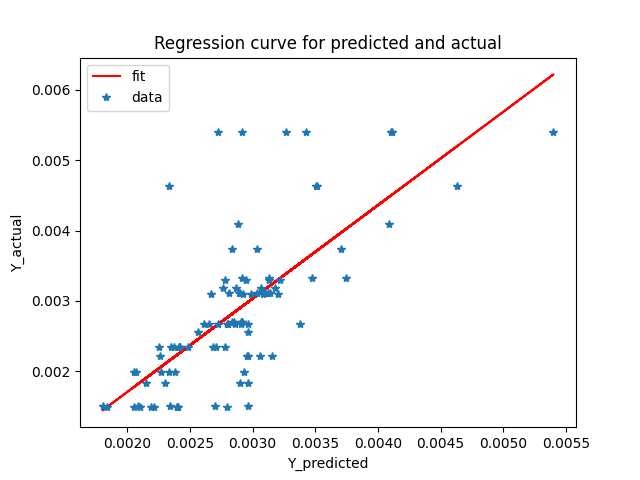


Fig 1: datasets 25 & Regression line: Y=1.32593X ± 0.000943 Fig 2: datasets 49 &Regression line : Y=1.15053X ± 0.000457

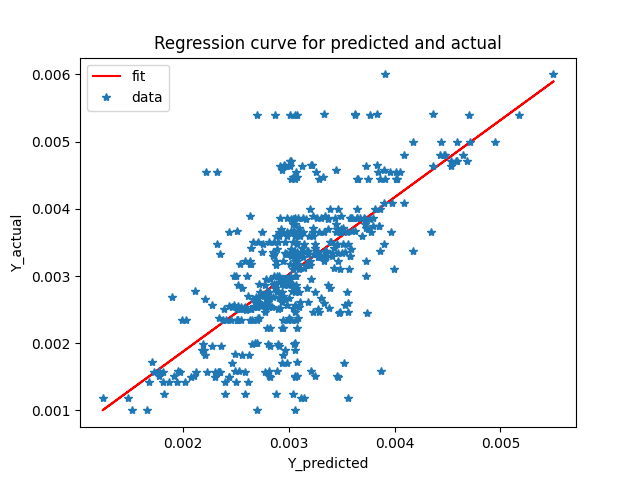
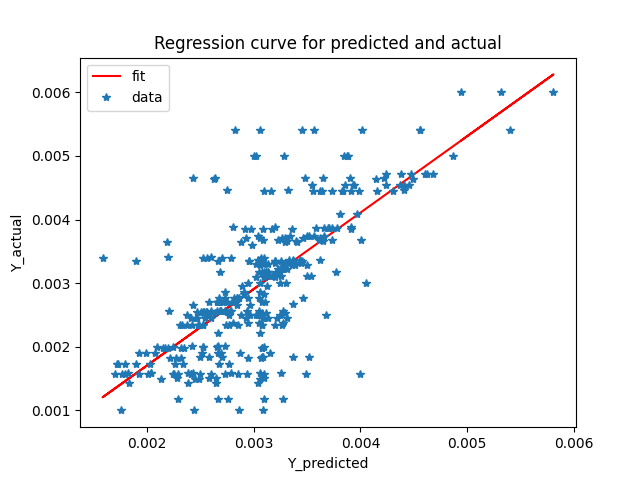


Fig 3: datasets 73 & Regression line : Y=1.20226X ± 0.0007 Fig 4: datasets 97 & Regression line : Y=1.15014X± 0.000430

# Discussion :

In this study, we applied machine learning techniques to solve an inverse

engineering problem of identification of the crack depth, based on the permanent

residual deformation as well as the permanent strain distributions.

We can see in the plots clearly that increase in the datasets, increase in the accuracy.To obtain more accuracy, more data should be collected. The higher the density of the training data, the higher the predicted accuracy.

AI in software testing aims to make testing smarter and more efficient. AI and machine learning apply reasoning and problem solving to automate and improve testing. AI in software testing helps reduce time-consuming manual testing, so teams can focus on more complex tasks, like creating innovative new features.

# Conclusion:

The accuracy of the model is good enough as we have RMS errors very close to zero. If we make more new datasets with more more data, we get very much accuracy. The depth prediction of the crack which we obtained from Machine learning tools works better in the range of 0 to 0.005 units because the maximum width of the model where crack exits is 0.005 units.

The prediction on the crack depth of the crack is not necessarily from the training data, after studying the training data, the machine learning algorithm can automatically use interpolation to find the applied load location that is not in the train data. This is to say that we only need to train algorithm with a limited number of loading sets.

# References:

* A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.
* D. Amaratunga, J. Cabrera, and Y.S. Lee. Enriched random forests. Bioinformatics, 24:2010–2014, 2008.
* Y. Amit. 2D Object Detection and Recognition: Models, Algorithms, and Networks. The MIT Press, Cambridge, 2002.
* L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees. Chapman & Hall, New York, 1984.
* [*Smith, M*](https://www.research.manchester.ac.uk/portal/mike.c.smith.html)*2009,*[ABAQUS/Standard User's Manual, Version 6.9](https://www.research.manchester.ac.uk/portal/en/publications/abaqusstandard-users-manual-version-69(0b112d0e-5eba-4b7f-9768-cfe1d818872e).html)*. Dassault Systèmes Simulia Corp, Providence, RI.*