Material Removal Rate Prediction in CMP using Tree-based Boosting and Linear Models

CS591 Data Mining Project

Ziye Liu
Dept. of Mechanical Engineering
The Univeristy of Alabma
Tuscaloosa, AL, U.S.
zliu66@crimson.ua.edu

***Abstract*—Chemical-mechanical planarization (CMP) is widely used in semiconductor industry to polish silicon wafer. Material removal rate (MRR) is an important process parameter. A high accuracy predictive model for MRR is critical for the estimation of polishing time. This project aims to develop a model to predict MRR in CMP process. Tree-based boosting (XGBoost) and linear models (elastic net) were experimented. A classifier + two regression models approach was proposed. XGBoost models has better predictive power than linear models. By using the classifier + 2 regression model approach, the error can be further reduced.**

Keywords—chemical-mechanical planarization, XGBoost, elastic net

I. INTRODUCTION

Chemical-mechanical planarization (CMP) is a process of polishing surfaces with the combination of chemical etching and mechanical abrasive polishing (Fig. 1). During the CMP process, the wafer is pressed against a polishing pad. The wafer and polishing pad are rotated in a controlled speed. A slurry composed of abrasive materials and chemicals are dispensed onto the pad. The polishing pad needs to be conditioned after a certain amount of polishing time by a dresser to roughen the pad's surface.

The CMP process has been widely used in the semiconductor industry for silicon wafer polishing. Material removal rate is a key process parameter in CMP process, which is defined as the material of the wafer polished away in a unit time. MRR is important for process monitoring and process control of CMP. The ability to predict MRR is critical for the estimation polishing time and therefore, important for the production of low cost, high quality IC chips.

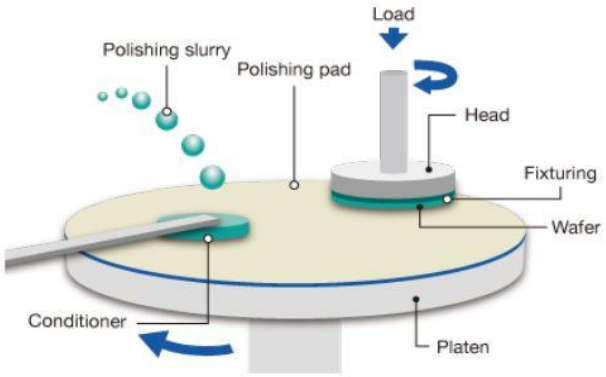

Fig. 1 Illustration of CMP process [1]

Several models have been developed based on the understanding of process mechanics. Preston proposed the first model on MRR. In his model, MRR is proportional to the pressure applied on the wafer and the relative rotational speed between the two rotating bodies [2]. The influence of other process parameters is characterized by a constant K_p (Eq. 1). A similar model was developed by Tseng and Wang. α and β were given a value of 5/6 and 1/2, respectively, based on process mechanics [3]. Luo and Dornfeld have proposed another model, in which MRR was considered as a function of the pressure applied to abrasives, penetration depth, volume removed by a single abrasive, and the number of active abrasives (Eq. 2) [4].

$$MRR = K_p P^{\alpha} V^{\beta} \quad (1)$$

$$MRR = \rho_w N V o l_{removed} \quad (2)$$

Byrane et al. discussed the effect of pad wear on MRR in CMP process [5]. Bastawros et al. developed a physical model that correlates the forces on each individual abrasive particle to the applied nominal pressure. The predicted correlations between MRR and slurry characteristics, i.e., particle size and concentration were verified experimentally [6]. Jeng and Huang developed a micro-contact model which considered the effects of the abrasive particles located between the polishing interfaces, thereby the down force applied on the wafer is carried

both by the deformation of the polishing pad asperities and by the penetration of the abrasive particles. In addition to operational parameters as the applied down force, the model also considered consumable parameters including wafer surface hardness, slurry particle size, and slurry concentration [7].

However, CMP is a very complex process. The combination effect of chemical etching and mechanical polishing is influenced by a lot of process conditions and it is difficult to develop an explicit analytical model to include all the process parameters. Also, some parameters in the developed model is difficult to measure experimentally, which also limits the application of those models.

MRR can be predicted by machine learning approach. For example, PHM (Prognostic Health Management, an academic organization) hosted a data competition to predict MRR in CMP process in 2016. Kong et al integrated nonlinear Bayesian analysis and statistical modeling to estimate and predict process state variables, and therewith to predict the performance measures, such as MRR (MRR), surface finish, surface defects, etc [8]. Lih et al developed an adaptive neuro-fuzzy inference system (ANFIS) based on subtractive clustering (SC) of the input parameter space. Linear statistical models were used to assess the relative significance of process input parameters and their interactions [9]. Wang and Yu developed a new neural network-based run-to-run process control system. It was verified experimentally that the developed system can precisely trace the desired target of MRR [10]. Wang et al developed a Deep Belief Network (DBN) model to reveal the relationship between MRR and polishing operation parameters such as pressure and rotational speeds of the wafer and pad. It was found that DBN outperformed other models such as neural network and support vector regression [11].

Some machining learning models, such as deep learning model, function like a black box. The interpretability of the models is very limited. Another drawback of machine learning model is that the predict capability of the developed model is only valid within the range of given training process window.

In this study, tree-based boosting model (XGBoost) and linear regression model were developed to predict MRR. Since MRR can be classified into two categories, a two-stage approach was proposed. A classifier was trained to determine the MRR group. Then for each class, a regression model was trained separately. It was found

II. Problem Definition

A. Data description

The dataset is provided by PHM (Prognostic Health Management) 2016 data competition. The dataset was divided into three parts by the competition organizer: training dataset, validation dataset, and test dataset. In this project, validation dataset and test dataset will be combined as test dataset. Cross-validation will be used as tune the hyperparameters. Training set has 1972 records and test set has 848 records. The dataset can be obtained from reference [12]. Table 1 summarizes the process attributes provided in the dataset.

Table 1 Process attributes and description

CHAMBER	Chamber in machine for wafer processing
USAGE_OF_BACKING_FILM	A usage measure of polish-pad backing film
USAGE_OF_DRESSER	A usage measure of dresser
USAGE_OF_POLISHING_TABLE	A usage measure of polishing table
USAGE_OF_DRESSER_TABLE	A usage measure of dresser table
PRESSURIZED_CHAMBER_PRESSURE	Chamber pressure
MAIN_OUTER_AIR_BAG_PRESSURE	Pressure related to wafer replacement
CENTER_AIR_BAG_PRESSURE	Pressure related to wafer replacement
RETAINER_RING_PRESSURE	Pressure related to wafer replacement
RIPPLE_AIR_BAG_PRESSURE	Pressure related to wafer replacement
USAGE_OF_MEMBRANE	A usage measure of membrane
USAGE_OF_PRESSURIZED_SHEET	A usage measure of wafer carrier flexible sheet
SLURRY_FLOW_LINE_A	Flow rate of slurry type A
SLURRY_FLOW_LINE_B	Flow rate of slurry type B
SLURRY_FLOW_LINE_C	Flow rate of slurry type C
WAFER_ROTATION	Rotation rate of wafer
STAGE_ROTATION	Rotation rate of stage
HEAD_ROTATION	Rotation rate of head
DRESSING_WATER_STATUS	Status of dressing table
EDGE_AIR_BAG_PRESSURE	Pressure of bag on edge of wafer

B. Research objectives

The major objective of this challenge is to develop a model to predict MRR in CMP process. A reliable MRR prediction can improve the estimates of polishing time, therefor reduce the polishing cost.

C. Evaluation metric

Root mean square error (RMSE) was used to evaluate the performance of developed models.

III. Proposed Solution

A. Tree-based boosting regression model

A tree-based boosting method XGBoost (extreme gradient boosting) was used to predict the relationship between process conditions and MRR. XGBoost is an open-source software library which provides the gradient boosting framework for with interface for C++, Java, Python, R, and Julia. It provides a

"Scalable, Portable and Distributed Gradient Boosting Library" [13]. It has become one of the most popular algorithms in machine learning community (e.g., kaggle). It is the choice of many winning teams for several recent machine learning competitions. In this study, XGBoost python package [14] was used.

Five-fold cross validation was used to tune max tree depth (2, 3, 4, 5, and 6) and the corresponding number of boosting rounds (the number of trees). The best model was then selected to predict the MRR for test dataset. The other parameters used in XGBoost is shown Table 2.

Table 2 Parameters used in XGBoost

eta	0.05
subsample	0.7
colsample_bytree	0.7
min_child_weight	3
gamma	0.3

B. *Classifier + 2 tree-based boosting regression model*

It can be observed from Fig. 2 that the MRR was well separated into two groups. It is possible that there are different mechanisms behind the two groups. The best parameters for the first group may not be suitable for the second group. By splitting the training data into two groups and training the regression models separately, the error might be further reduced. However, before training the regression models, a classifier is needed to determine the given data records belong to which group.

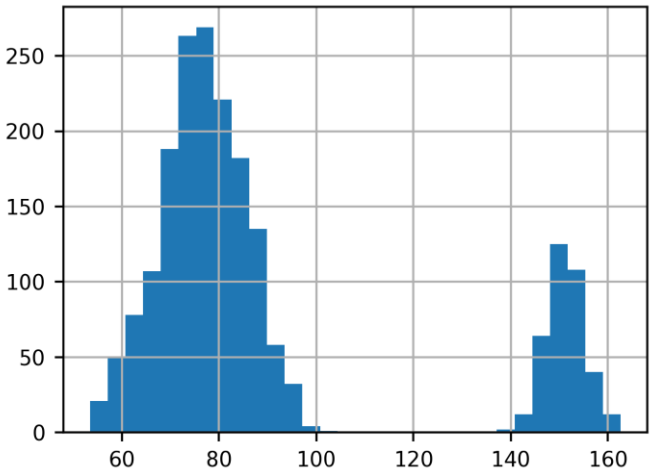

Fig. 2 MRR distribution in the training dataset

Two classes were defined based on MRR, i.e., class 1 for MRR ≥ 120 and class 2 for MRR < 120. A XGBoost classifier was trained. Since the two groups were well separated, the classifier can reach 100% accuracy for both training and test datasets.

Two regression models were trained separately for class 1 and 2. For each model, the max tree depth and number of boosting rounds were tuned with 5-fold cross validation. The two-stage learning approach is illustrated in Fig. 3.

C. *Linear regression model*

Linear regression was also used as a benchmark to compare with XGBoost. The linear regression model used in this project was elastic net. Elastic net was a linear regression models with L1 and L2 regulizations [15]. Thus, elastic net is combination of lasso (L1 penalty) and ridge regression (L2 penalty). The ratio of L1 and L2 penalty is controlled by the parameter α. The total penalty is controlled by the parameter λ. In this study, the python package *sklearn.linear_model.ElasticNet* [16] was used. It should be noted that *sklearn.linear_model.ElasticNet* has different notations with reference [15] and other elastic net packages e.g., R/glmnet. The notations used in reference [15] was used in this report to avoid confusion.

Five-fold cross validation was used to tune α (0.1-1.0) and corresponding λ. When α = 0, it will be completely ridge regression and when α = 1.0, it will be completely lasso regression.

D. *Classifier + 2 linear regression model*

Similarly, a classifier + 2 regression models approach was also applied to linear models. For each model, cross-validation was used to tune α and λ.

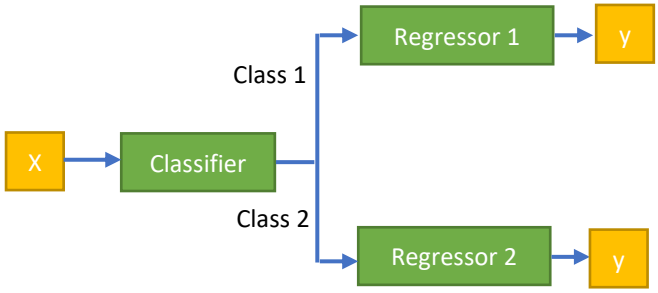

Fig. 3 The classifier + 2 regression model

IV. Evaluations

A. *Single tree-based boosting regression model*

Fig. 4 shows the relationship between the number of boosting rounds of RMSE in cross validation to tune XGBoost parameters. When validation error starts to increase, which means the model is becoming too complex for the model, i.e., over-fitting, the cross-validation will stop and the best number of rounds for the given max tree depth can be found.

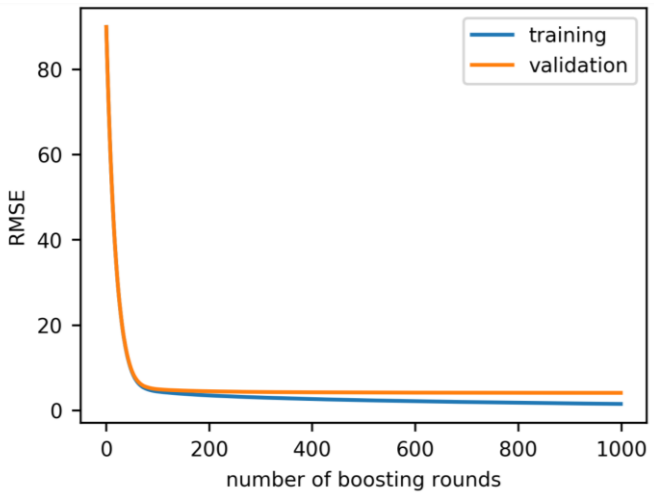

Fig. 4 RMSE vs. number of boosting rounds for training and validation in XGBoost cross validation

The optimal number of rounds and cross-validation error for each max tested tree depth is summarized in Table 3. It can be seen that max tree depth 3 with number of boosting rounds 1025 produces the smallest cross validation error.

Table 3 Optimal number of boosting rounds and CV RMSE

Max tree depth	Optimal number of boosting rounds	CV RMSE
2	862	4.298415
3	1025	4.106870
4	480	4.168649
5	463	4.132332
6	359	4.125890

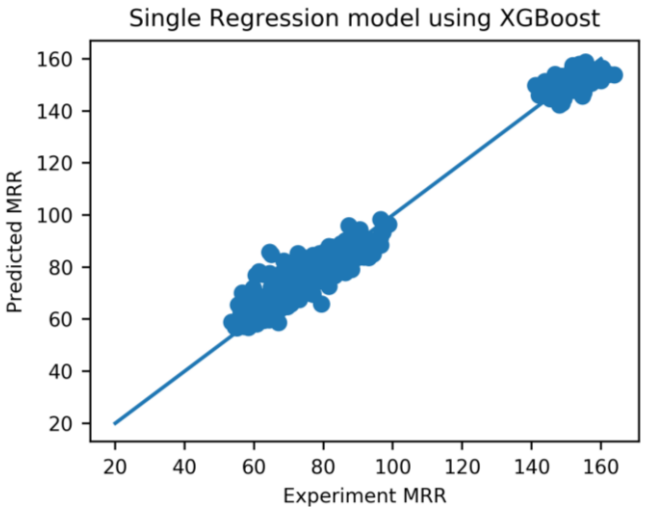

Fig. 5 Predicted MRR for test dataset – single regression model using XGBoost.

The predictive power of the boosting model with the optimized parameter is applied to test dataset. The result is shown in Fig. 5. The RMSE for test dataset is 3.84026.

The importance of the features is shown in Fig. 6. The feature importance obtained by XGBoost is evaluated by F score which is the number of splits on this feature. This information can be used for feature section in a further study. It is also useful for the production since the importance of each process parameter is provided. In the literature, the rotation speed and pressure are the most important process parameters for MRR. However, in this study, it is shown the effect of those process parameters are relative small. This can be explained by the fact that those parameters in practical production is tightly controlled in a small interval and the variation in the parameters is very small. As a result, the tree models have less split on those features and those features have smaller F scores.

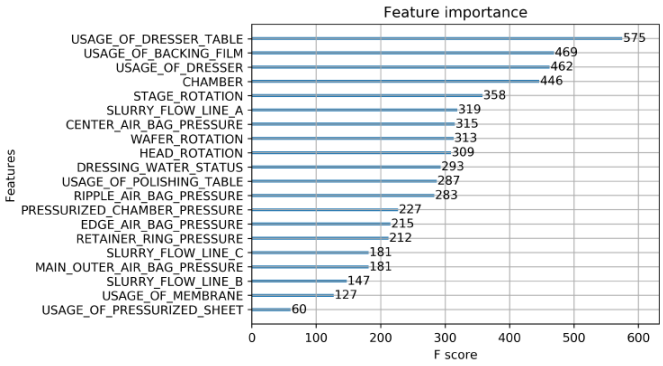

Fig. 6 Feature importance

B. *Classifier + 2 tree-based boosting regression model*

Since class 1 and class 2 are well separated, the developed classifier achieved a classification accuracy of 100%.

Table 4 and 5 summarizes the cross-validation results for class1 and class 2 using different max tree depth and the optimal number of boosting rounds. It can be seen that optimal parameters for two classes are different. Therefore, training the regression model for the two classes separately may achieve a smaller error than training the two models together. However, it should be noted that by training the two models separately, each class has less data records. Therefore, there is higher chance for overfitting.

Table 4 Optimal number of boosting rounds and CV RMSE for different max tree depth (class 1)

Max tree depth	Optimal number of boosting rounds	CV RMSE
2	278	2.621662
3	220	2.590970
4	170	2.613220
5	132	2.600272
6	133	2.635568

Table 5 Optimal number of boosting rounds and CV RMSE for different max tree depth (class 2)

Max tree depth	Optimal number of boosting rounds	CV RMSE
2	910	4.136123
3	643	4.046558
4	605	4.018216
5	415	4.008437
6	458	4.051279

The classifier + 2 boosting regression model was applied to the test dataset. The result is shown in Fig. 7 and the RMSE is 3.57002. Comparing to the single regression model, there is 7% decrease in error.

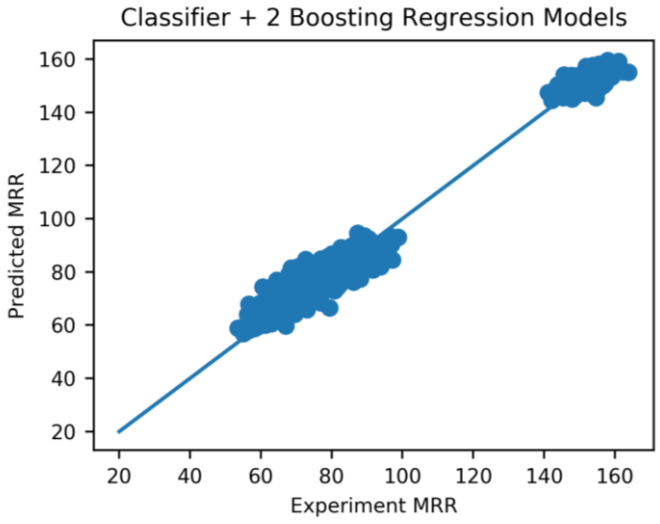

Fig. 7 Predicted MRR for test dataset – two regression models using XGBoost.

C. Elastic net

The optimal parameters for elastic net single regression and two regression models is given in Table 6. The α for class 1 and class 2 are completely different. For class 1, α is 0.1 which the minimum value tried in this study. For class 2, α is 1.0 which makes the model completely lasso regression. The higher α indicates that for class 2, there are some features are not related to MRR. A feature selection maybe helpful to further reduce the error.

Table 6 Optimal parameter for elastic net

Models	α	λ
Single regression	0.02888	1.0
Two regression class1	0.31214	0.1
Two regression class 2	0.00464	1.0

The single and two linear regression models were applied to test dataset. The results are shown in Fig. 8 and 9, resepectively. The RSME for single regression and two regression models are 6.21433 and 5.76161, repspectively. The reduction in RMSE by training the two classes seperately is approximately 7%. It should be noted that there are several "outliers" which predictions are far away from the experiment values. However, there is no such "outliers" in tree-based models, which indicates there might be certain non-linear relationship that linear models cannot capture. Another possible explaination is that for tree-based models, the predicted values will always fall in the range of traning data, while in linear model if one or more features is far away from the training set range, the predicted result might be far away from the training set range. Another observation is that the single regression model shows very weak peditive power for class 2. Two regression model is slightly better than single regression model, but still much worse than boosting models.

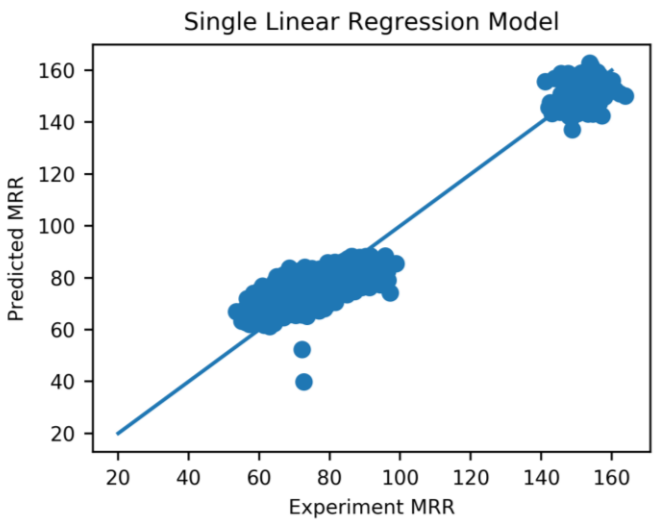

Fig. 8 Predicted MRR for test dataset – single linear regression model using elastic net.

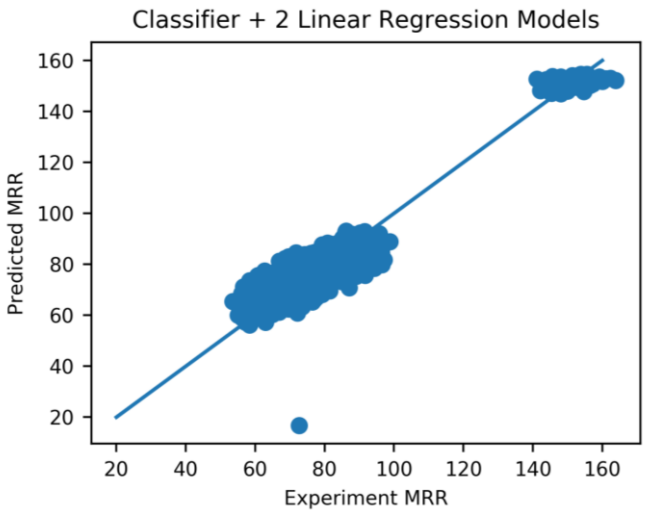

Fig. 9 Predicted MRR for test dataset – two linear regression models using elastic net.

V. Discussion

Table 7 summarizes RMSE for XGBoost and elastic net models. Comparing to linear model, tree-based boosting models show better predictive accuracy. The CMP is complex process. The material is removed by a combination of chemical etching and mechanical polishing. Thus, the relationship between material removal rate and process condition might be non-linear, which explains the relative poor predictive accuracy of linear models.

For both XGBoost models and elastic net, the two-regression model approach shows better accuracy than a single regression model. The regression parameters suitable for class 1 is not optimal for class 2. By training the two classes separately, each class can have their optimal parameters. For each class, there are less training data, a less complex model is developed to avoid over-fitting. For XGBoost models, the single regression model has 1025 trees, while the two regression models have 220 and 415 trees for class 1 and class 2, respectively. The two-regression model approach has less trees in total comparing to the single regression model.

Table 7 RMSE for XGBoost and Elastic net models

	XGBoost	Elastic net
Single regression model	3.84026	6.21433
Two Regression models	3.57002	5.76161

Conclusion

This project experimented the predictive capability of tree-based boosting (XGBoost) and linear models in the estimation of material removal rate in mechanical-chemical planarization. Particularly, a classifier + two regression approach was proposed. The conclusions are given in the followings,

- Tree-based boosting models (RMSE: 3.57002) outperform linear models (RMSE: 5.76161)
- Classifier + 2 regression model approach improves boosting models (~7% decrease in RMSE)
- Classifier + 2 regressor model approach improves linear models (~ 7% decrease in RMSE)

References

[1] AZO Material, Surface Metrology for In-Situ Pad Monitoring in Chemical Mechanical Planarization (CMP), accessed from https://www.azom.com/article.aspx?ArticleID=12527

[2] Preston FW. The theory and design of plate glass finishing machines. J. Glas Tech.. 1927;11:214.

[3] Tseng WT, Wang YL. Re‐examination of Pressure and Speed Dependences of Removal Rate during Chemical‐Mechanical Polishing Processes. Journal of the Electrochemical Society. 1997 Feb 1;144(2):L15-7.

[4] Luo J, Dornfeld DA. Material removal mechanism in chemical mechanical polishing: theory and modeling. IEEE transactions on Semiconductor Manufacturing. 2001 May;14(2):112-33.

[5] Byrne G, Mullany B, Young P. The effect of pad wear on the chemical mechanical polishing of silicon wafers. CIRP Annals-Manufacturing Technology. 1999 Jan 1;48(1):143-6.

[6] Bastawros A, Chandra A, Guo Y, Yan B. Pad effects on material-removal rate in chemical-mechanical planarization. Journal of electronic materials. 2002 Oct 1;31(10):1022-31.

[7] Jeng YR, Huang PY. A material removal rate model considering interfacial micro-contact wear behavior for chemical mechanical polishing. Transactions of the ASME-F-Journal of Tribology. 2005 Jan 1;127(1):190-7.

[8] Kong Z, Oztekin A, Beyca OF, Phatak U, Bukkapatnam ST, Komanduri R. Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear Bayesian analysis and statistical modeling. IEEE Transactions on Semiconductor Manufacturing. 2010 May;23(2):316-27.

[9] Lih WC, Bukkapatnam ST, Rao P, Chandrasekharan N, Komanduri R. Adaptive neuro-fuzzy inference system modeling of MRR and WIWNU in CMP process with sparse experimental data. IEEE Transactions on Automation Science and Engineering. 2008 Jan;5(1):71-83.

[10] Wang GJ, Yu CH. Developing a neural network-based run-to-run process controller for chemical-mechanical planarization. The International Journal of Advanced Manufacturing Technology. 2006 Jul 1;28(9):899-908.

[11] Wang P, Gao RX, Yan R. A deep learning-based approach to material removal rate prediction in polishing. CIRP Annals-Manufacturing Technology. 2017 Apr 21.

[12] PHM society, PHM 2016 Data Challenge, accessed from https://www.phmsociety.org/events/conference/phm/16/data-challenge

[13] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794). ACM.

[14] XGBoost python pakage, Python Package Introduction, acessed from http://xgboost.readthedocs.io/en/latest/python/python_intro.html

[15] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005 Apr 1;67(2):301-20.

[16] Python Sklearn ElasticNet package accessed from http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html