

Predictive Models for Equipment Fault Detection in the Semiconductor Manufacturing Process

Sathyan Munirathinam and Balakrishnan Ramadoss

Abstract—Semiconductor manufacturing is one of the most technologically and highly complicated manufacturing processes. Traditional machine learning algorithms such as uni-variate and multivariate analyses have long been deployed as a tool for creating predictive model to detect faults. In the past decade major collaborative research projects have been undertaken between fab industries and academia in the areas of predictive modeling. In this paper we review some of these research areas and thus propose machine learning techniques to automatically generate an accurate predictive model to predict equipment faults during the wafer fabrication process of the semiconductor industries. This research paper aims at constructing a decision model to help detecting as quickly as possible any equipment faults in order to maintain high process yields in manufacturing. In this research, we use WEKA tool and R languages for implementing our proposed method and other five machine learning discovery techniques.

Index Terms—Predictive model, semiconductor manufacturing process, machine learning, data classification, feature selection, R language, and python language.

I. INTRODUCTION

The semiconductor industry is one of the most technology-evolving and capital-intensive market sectors. Effective prediction of fault detection in equipment is necessary to prevent abrupt equipment breaking down and is also benefit to improve productivity, reduce costs and repairing time. Machine Learning (ML) is the engineering of methods that enable computers to adapt their behavior based on empirical data. ML analyzes data and uses the theory of statistics to build mathematical models to predict events in the future. Machine learning methodologies are nowadays applied in many industrial and scientific environments including technology-intensive manufacturing and in general every data-intensive field that might benefit from reliable predictive capabilities, like the semiconductor industry.

Semiconductor manufacturing [1] is a highly complex manufacturing process composed of hundreds of steps. The number of process steps in wafer fabrication, typically over 500, and the amount of data recorded during the entire production process, this produces a vast amount of monitoring data. The major processes in most semiconductor industries [5], [16] are in the following

sequence: production of silicon wafers from pure silicon material, fabrication of integrated circuits onto the raw bare silicon wafers, assembly by putting the integrated circuit inside a package to form a ready-to-use product, and testing of the finished products.

In recent years, many manufacturing equipments are equipped with sensors to facilitate real-time monitoring of the production process. These production-state and equipment-state sensor data provide an opportunity for efficient control and optimization. Unfortunately, such measurement data are so overwhelming that timely detection of any fault during the production process is difficult. In this paper, we study the problem of accurate detection of equipment fault states in the wafer fabrication process. The dataset is donated by McCann *et al* [20] and publicly available for re-experimentation. The fault detection model can be automatically built from the existing sensor data by means of machine learning techniques. An efficient and effective approach to monitor the health state of equipment and predict impending failure has long been an interest of researcher communities as well as industries. It is the BIG data that makes predictive maintenance a reality. There is no general technique followed for Predictive Model and generally it is tailored to a specific business problem (see Fig. 1).

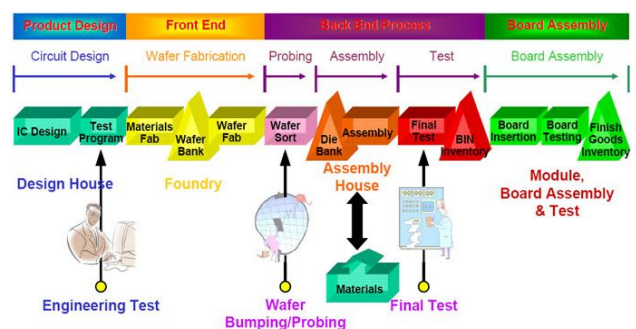


Fig. 1. Overview of semiconductor manufacturing processes.

Machine learning algorithms have traditionally been focused on simple prediction modeling. Given observations that have been generated by an unknown stochastic dependency, the goal is to infer a law that will be able to correctly predict future observations generated by the same dependency. Statistics, in contrast, has traditionally focused on “data modeling”, i.e., on the estimation of a probability law that has generated the data. During recent years, the boundaries between the two disciplines have become blurred and both communities have adopted methods from the other.

Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model based on

Manuscript received December 14, 2014; revised March 1, 2015. This work was carried as part of PhD research at Bharathiar University, India.

Sathyan Munirathinam is with Micron Technology and PhD researcher at Bharathiar University, India (e-mail: Sathyan.Munirathinam@gmail.com).

Balakrishnan Ramadoss is with National Institute of Technology, Trichy, India (e-mail: brama@nitt.edu).

inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions (see Fig. 2).



Fig. 2. Prediction technique in semiconductor Fabs.

Machine learning is about the application of learning algorithm to build model that can best characterize underlying data and accurately predict the class of unlabeled data. In modern industry with numerous automated machines, a large amount of data has been automatically collected. Engineers may potentially use these raw data to identify specific hidden patterns such as the process fault model to assist the timely investigation of the root causes of the defects.

The organization of this paper is as follows. The review of related work to intelligent methods for predicting failure in manufacturing process is presented in Section II. Then the proposed predictive model for fault detection technique is explained and demonstrated in Section III. A series of experimentation and results are presented in Section IV. Finally, Section V concludes the paper with discussion of our future research direction.

II. PROBLEM DEFINITION

Semiconductor industry statistics show that most semiconductor capital equipment suffers at least 8% unscheduled downtime and loses another 7% to scheduled maintenance (see Fig. 3).

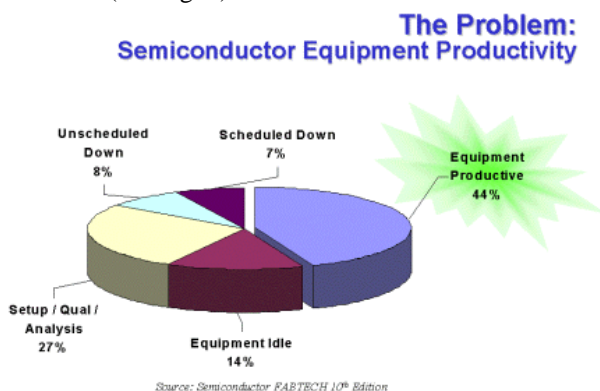


Fig. 3. Source from FABTECH 10th edition.

Maximizing process tool uptime has remained a core challenge for the manufacturing of advanced semiconductors over time. Predictive Maintenance is the process of discovering when equipment needs maintenance in order to avoid a catastrophic failure. The PdM problems usually suffer due to the fact that the lack of a sufficient amount of observations to prepare a reliable statistical model. The research paper provides a method for predictive maintenance in semiconductor equipments, comprising the following steps: collecting and preprocessing the raw data, then performing classification through a statistic

classification model.

III. RELATED WORK

In most manufacturing processes, cost, quality, and delivery time are key factors for enterprises to attain long-term competition. During the manufacturing processes, process engineers have to monitor and identify the specific characteristics of abnormal products as soon as possible [6], [7], [15], [23], [29], [33], [34] Process control is crucially important to the semiconductor industries that operate the multistage manufacturing systems on the product scale of lesser 300 nanometers [18]. Pham and Afify [35] reviewed machine learning techniques in the manufacturing domain. They evaluated the several machine learning techniques and examined application areas in which they have been successfully deployed.

Modern technology in semiconductor manufacturing enables real time process control with the measurement data obtained from the equipment sensors and the final electrical test. With such high volume of data recorded during the entire production process, effective monitoring and optimal process control by investigating and analyzing these data are difficult work for process engineers. Traditional process control methodology like uni-variate and multivariate control charts is no longer an efficient method to control manufacturing systems with hundreds of processing stages. Instead automatic and advanced process control method is required.

Ison and colleagues [13], [14] proposed a decision tree classification model to detect fault of plasma etch equipment. The model was built from the five sensor signal data. Many researchers also studied the fault detection problem during the etch process. Goodlin *et al* [9] proposed to build a specific control chart for detecting specific type of faults. They collected tool-state data directly from the etcher. These data consist of 19 variables. The work of Spitzlsperger and colleagues [21] was also based on the statistical method. They adopted the multivariate control chart method to maintain changes in the mean and standard deviation coefficients by remodeling technique.

Most work on fault detection methods has studied the process control problem with a few features of tool-state and process-state measurement data. McCann and his team [17] proposed a rather different setting in which the measurement data from the wafer fabrication process contain as much as 590 features. With such abundant features or variables, feature selection techniques [10] are obviously necessary in order to improve both the prediction and the computational performances.

In this paper, we also analyze the wafer fabrication data [20] collected from 590 sensors with the last feature is a label stating pass or fail state. The observed data contain 1,463 pass cases with only 104 fail cases. In this work not only a feature selection method for extracting the post discriminative sensors is proposed, but also a boosting technique is devised to deal with highly imbalance between the pass and fail cases.

IV. RESEARCH PROCESSES

The CRISP-DM and SEMMA methodologies are most widely used by the data mining community. CRISP-DM is

easier to use than SEMMA as it provides detailed neutral guidelines that can be used by anyone in the data mining field. Our research process is modeled based on the CRISP-DM model. CRISP-DM model is a machine learning process model that describes commonly used approaches that machine learning experts use to tackle problems. A review and critique of machine learning process models in 2009 called the CRISP-DM the “de facto standard for developing machine learning and knowledge discovery projects” (see Fig. 4).

- Business Understanding
- Data Understanding
- Data Preparation
- Data Modeling
- Model Evaluation
- Model Deployment

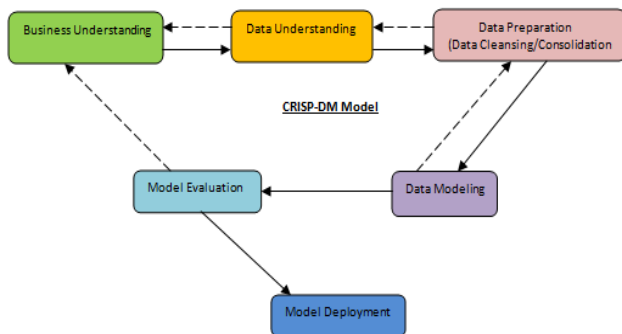


Fig. 4. Overview of CRISP-DM model.

- 1) Business Understanding. The machine learning/machine learning project should start by defining its goals in terms of business requirements. This specification should then be converted into a machine learning problem definition.
- 2) Data Understanding. To effectively operate on the data in the later phases, some knowledge has to be obtained on the characteristics of the data itself. It's very important to understand the sensor data generated from semiconductor equipment.
- 3) Data Preparation. This is the process of producing rich data from raw data. Typical pre-processing tasks are noise-cleaning, feature extraction, feature reduction and feature selection.
- 4) Data Modeling. In this phase, a number of machine learning techniques are proposed and their parameters are adjusted to the specific problem.
- 5) Model Evaluation. This stage involves further evaluation of the techniques of sufficient quality. Particular attention has to be directed to possible problems that have not been previously considered. It is also necessary to be confident that the methods will actually deal with the original goals of the project.
- 6) Model Deployment. This last phase involves the necessary steps to make the user able to exploit the machine learning method developed in the previous steps.

V. SECOM DATASET

This dataset used in this research, namely SECOM (Semiconductor Manufacturing) dataset [40], consists of manufacturing operation data and the semiconductor quality

data. This dataset consists of 590 manufacturing operation variables and 1 quality variable for 1115 instances. The SECOM dataset [20] contains 1567 examples taken from a wafer fabrication production line. Each example is a vector of 590 sensor measurements plus a label of pass/fail test. Among the 1567 examples, there are only 104 fail cases which are labeled as positive (encoded as 1), whereas much larger amount of examples pass the test and are labeled as negative (encoded as -1). This is a 1:14 proportion. SECOM dataset is the data about semiconductor manufacturing process. The goal of SECOM dataset is a good or bad semiconductor from manufacturing process. There are two classes in 1567 instances including 104 fails, and they have some missing values. More information about the SECOM dataset is available under Appendix A.

VI. PROPOSAL FOR PREDICTIVE MODEL FOR FAULT DETECTION

The imbalance of pass and fail examples in addition to the large number of measurement data obtained from hundreds of sensors make this dataset a difficult one to accurately analyze. It is thus our main focus to devise a method based on machine learning techniques to build an accurate model for fault detection. The framework of our study is presented in Figure 1. Feature selection techniques in our study are ranging from simply removing features with a constant value and features containing numerous missing values (more than 55% of values are missing), to statistical based analysis such as chi-square and principal component analysis (PCA) and information theoretical based such as gain ratio.

We also devise a cluster based technique call *Average Diff* to analyze discrimination power of each feature. On the model building phase, we apply four methods to induce the fault- detection model namely decision tree, naïve Bayes, logistic regression, and k-nearest neighbor. The dataset is in a form of matrix; rows represent each observation or instance and columns represent features which are values recorded from each sensor. The steps in our proposed method for creating an accurate model to detect fault case from highly imbalance data with numerous features are as follows:

A. Data Preparation Phase

Data collection and preparation is the first and crucial step in developing a predictive model. Data preparation is an indispensable step in order to convert various data forms and types into proper format that is meaningful to machine learning predictive model. During semiconductor manufacturing process, larger amounts of data are collected during processing on a regular basis. The collected data comprise all the variables including the predictor variables that can be used for establishment of prediction models. Data available are “horizontal”: huge amount of different variable available (to be reduced) and few observations available in the same operating conditions. With variable selection, correlation analysis and PCA we want to reduce the number of regressors while with data clustering we aim to increment observations usable for modeling. There are more than one

hundred statistical variables (means, variances, maximum and minimum values, etc.,) have been collected from the machine (see Fig. 5).

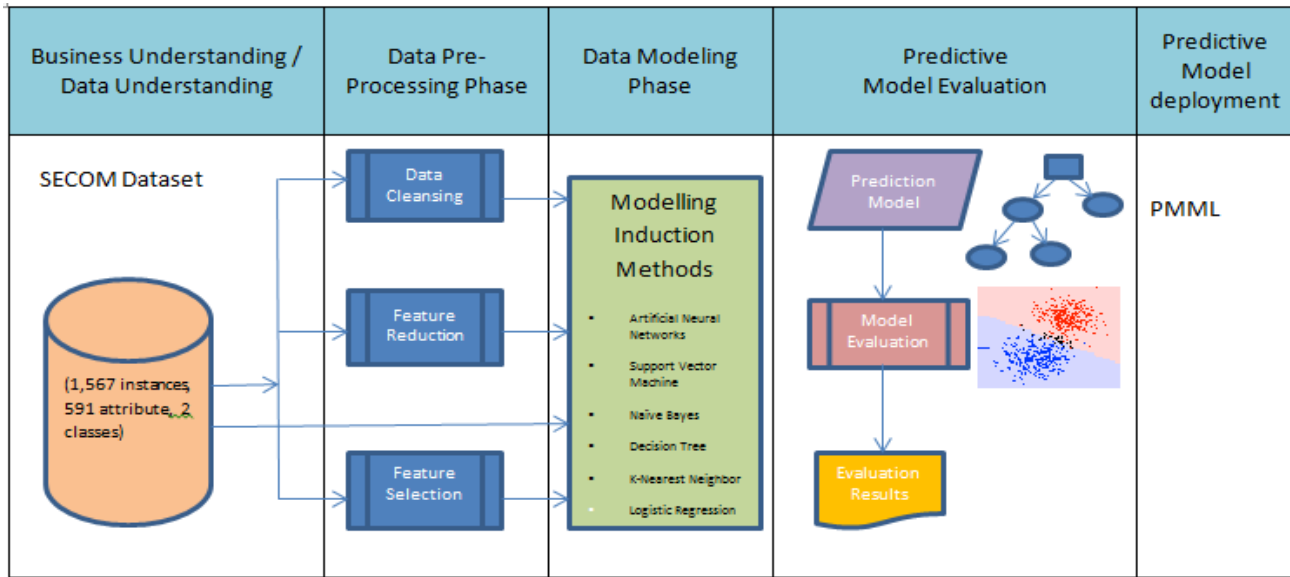


Fig. 5. The proposed method to generate predictive model.

B. Data Cleansing Phase

Data cleansing phase is an important phase in predictive modeling. The extracted SECOM data set needs to be cleansed for anomalies and also the data should be normalized for all the range of values of raw data varies widely. A data cleansing procedure discards 452 instances with null and missing values.

C. Feature Scaling

“Feature Scaling” is the technique that followed to normalize the data set. The goal of the data cleansing is to obtain a complete cleansed data set that can be modeled with outliers removed and solutions for handling of missing data applied. To normalize the input data set, the continuous variables were transformed on a linear scale to a value with a range of 0 to 1 or -1 to 1. Ordinal data were spaced equally over the same range. Missing values were substituted with the class mean. Data with different scales can induce instability in neural networks (Weigend and Gershenfeld, 1994). In order to normalize the raw data of input and output the following normalization equation is used:

$$x_{\text{norm}} = 2 \times \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} - 1$$

where x is the data to be normalized, i.e., and x_{\min} and x_{\max} are minimum and maximum values of the raw data. In such a way, all the inputs and the desired outputs are normalized within the range of ± 1 .

D. Feature Reduction

One of the major challenges in semiconductor manufacturing process is that the number of tool variables is usually very high [2]; in such setting, variable selection techniques often prove to be useful [7]. Investigate data observed from each sensor, i.e. data in each column. If the data appear to be a single value, then remove that feature. Count in each column the „not available” or missing values.

If data are missing more than 55%, then remove that feature (see Fig. 6).

E. Feature Selection

To find the most influential parameters is the important step in the whole process of predictive modeling. It is vital to note that the success of any predictive model largely depends on the predictor variables that are selected to use as the model inputs. The most elementary method for selection for parameters is the use of a trial/error method. But with this method, it is a tedious and complicated processing considering the number of parameter variables. With most modeling methods, it is also possible to use all data as input variables to avoid this “predictor selection” step, but this approach could result in bad model behavior. Doing this will result in following things:

- Over-fitting
- More Computational Power
- Less Prediction Accuracy

So we adopted a strategy of combining three techniques in selecting the parameter. Below are the 3 techniques followed in selecting the right variables for prediction Variable selection in the field of semiconductor manufacturing can be classified into two categories:

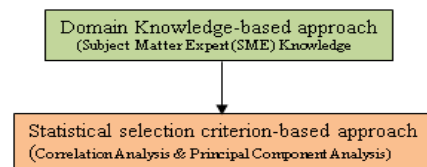


Fig. 6. Feature selection techniques.

F. Subject Matter Expert (SME) Knowledge

More than thousand parameters are used by the semiconductor equipment to control the overall process. In order to reduce the parameter set to reasonable selection we took the benefit of the existing expertise of equipment engineers. We were able to reduce the parameter selection from 1000 variables to 80+ variables. But still 80+ variables

are too much for developing model. It should be reduced even further for a reasonable set of variables.

G. Correlation Analysis (CA)

Next step to reduce the set of parameters was done with correlation analysis. Correlation Analysis is useful for determining the direction and strength of the association (linear relationship) between two variables. This technique is performed to omit parameters bringing little information to the dataset. The strength of the linear association between two variables is quantified by the correlation coefficient. For every couple of measurement parameters (x_1, x_2) we compute the correlation coefficient. The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient or “Pearson’s correlation coefficient”, commonly called simply the “correlation coefficient”. The population correlation coefficient $P_{x,y}$ between two random variables X and Y with expected values μ_x and μ_y and standard deviations σ_x and σ_y is defined as

$$\begin{aligned} p_{x,y} &= \text{corr}(X,Y) \\ &= \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \\ &= \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y} \end{aligned}$$

where μ_x , μ_y and σ_x , σ_y are respectively the mean and the standard deviations of x , y . E is the expected value operator, cov means covariance. Variables/Predictors showing a pairwise correlation of ± 0.8 or higher are removed, Coefficient $P_{x,y}$ between two random variables X and Y with. Below figure shows the correlation matrix. The

reduced parameter set contains $p=28$ predictors (see Fig. 7- Fig. 8).

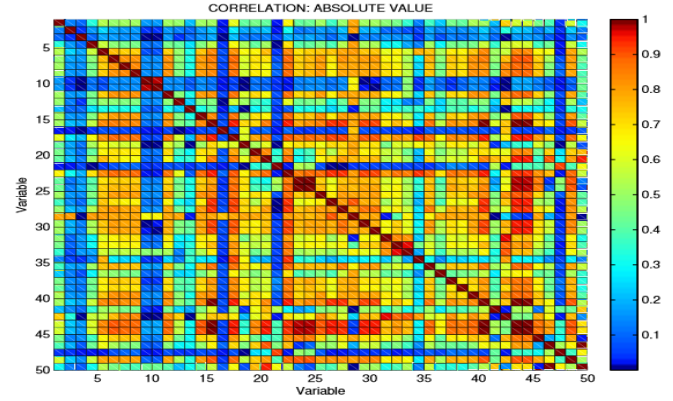


Fig. 7. Correlation matrix (CA).

The algorithm is as follows:

- 1) Calculate the correlation matrix of the predictors
- 2) Determine the two predictors associated with the largest absolute pair-wise correlation (call them predictors A and B)
- 3) Determine the average correlation between A and other variables. Do the same for predictor B.
- 4) If A has a larger average correlation, remove it; otherwise remove predictor B.

Repeat Step 1–Step 4 until no absolute correlations are above the threshold.

The idea is to first remove the predictors that have most correlated relationships.

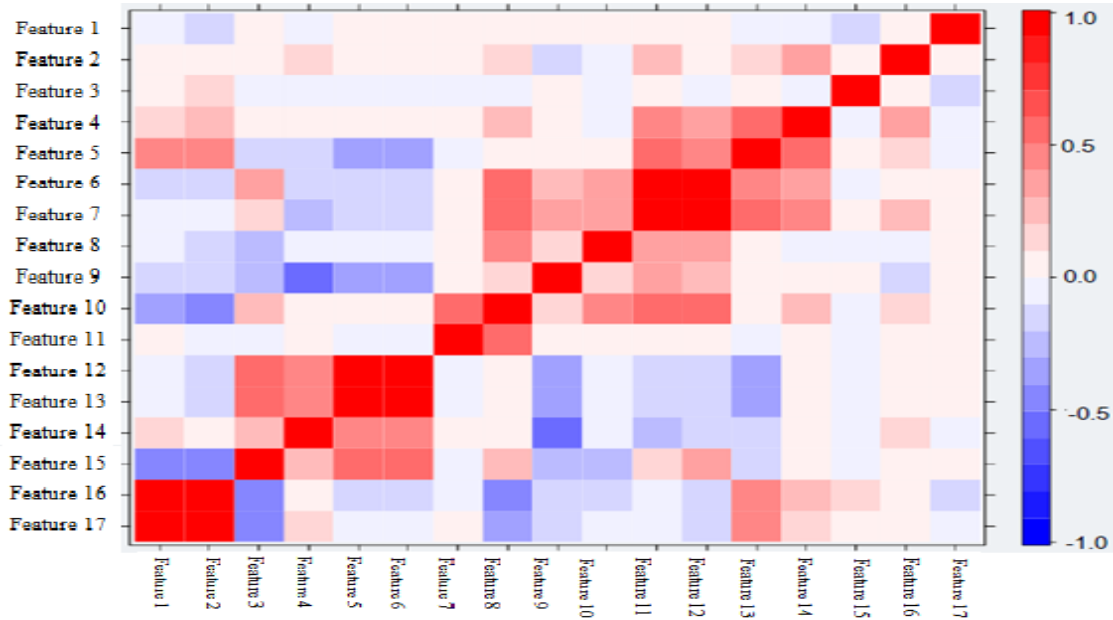


Fig. 8. Correlation analysis matrix (CAM).

Apply two statistical based feature selection techniques: chi-square and principal component analysis (PCA), and save the result as two separate datasets.

H. Variable Component Analysis (VIA)

Next step to pick the top variables for our prediction model was done using Variable Importance Analysis (VIA). In the context of variable importance analysis, we used 3

techniques:

- 1) Chi-Square
- 2) Decision Trees
- 3) Random Forest Trees

Gain ratio based technique is used and also cluster-based feature selection technique, called *AverageDiff*. *AverageDiff* clusters data into two clusters (fail cluster and pass cluster). It compares value differences in every feature of the fail cluster mean and the pass cluster mean. Then

ranking features in descending order according to the magnitude of mean differences computed in Steps 2–Steps 5, and output the ranked features.

I. Model Building Phase

The SECOM data set is divided into two datasets, train data and test data. Each data set maintains the same proportion of pass and fail cases. Over-sampling the fail cases in the train data by duplicating the fail cases to be the same amount as the pass cases. For datasets from Step 3–Step 8, evaluate model accuracy with 10-fold cross validation technique. Dataset from Step 9 is evaluated with the test set. Build a prediction model with Decision Tree (DT), Naïve Bayes (NB), k-Nearest Neighbor (KNN), Logistic regression, Artificial Neural Network (ANN) and Support Vector Machine (SVM) algorithms (see Fig. 9–Fig. 12).

We assess the model performance based on the below eight performance metrics:

- 1) True positive rate (TP rate or recall),
- 2) False positive rate (FP rate or false alarm).
- 3) Precision,
- 4) F- measure,
- 5) MCC
- 6) ROC Area
- 7) PRC
- 8) PRC Area
- 9) Recall

The computation methods of these matrices are given below figure:

$$TP \text{ Rate (Rate of true positives)} = \frac{TP}{TP + FN}$$

$$\text{Sensitivity/Precision} = \frac{TP}{TP + FP}$$

$$F - \text{Measure} = \frac{2TP}{2TP + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$TP \text{ Rate (Rate of false positives)} = \frac{FP}{FP + TN}$$

CONFUSION MATRIX		Predicted Class	
		Class = +1	Class = -1
	Class = +1	TP	FN
	Class = -1	FP	TN

Fig. 9. Confusion matrix.

The symbols in the matrices can be explained as follows:

TP = true positive or number of fail cases that are correctly identified as fail,

FP = false positive or number of pass cases that are incorrectly identified as fail cases,

FN = false negative or number of fail cases that are misclassified as pass cases,

TN = true negative or number of pass cases that are

correctly classified as pass cases.

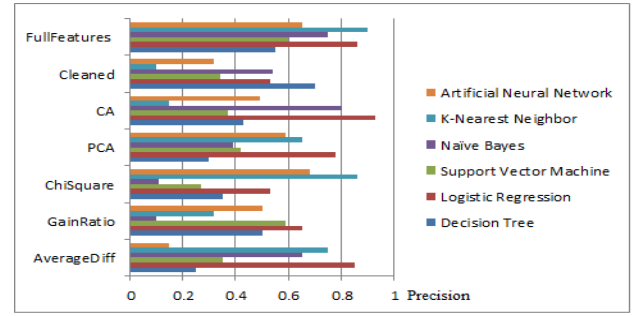


Fig. 10. Precision on different predictive models.

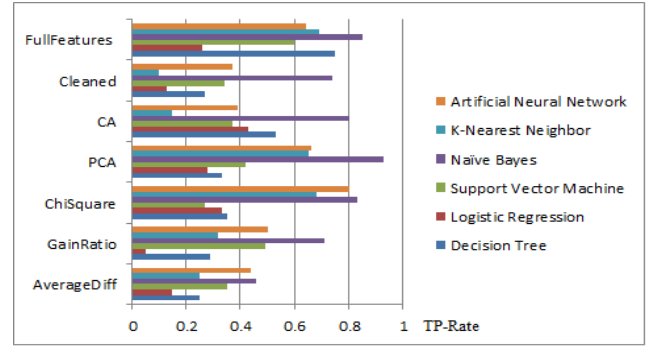


Fig. 11. TP-Rate on different predictive models.

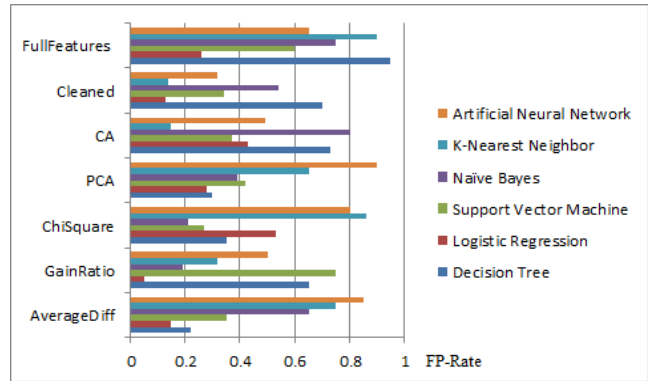


Fig. 12. FP-Rate on different predictive models.

VII. EXPERIMENTATION AND RESULTS

A. Results of Feature Selection Techniques

For the assessment of the created models, usually goodness-of-fit metrics are used to review the prediction capability of the model. We use the WEKA software [32] to perform a series of experiments. The first part of our study aims at selecting principal features that show the most discrimination power of differentiating fail cases from pass cases. In the cleaning step, we remove 137 features that contain a single value and lots of missing values. From the remaining 454 features, we select the best 168 features (to maintain around 95% of variances) by means of principal component analysis (PCA), Chi-square test, gain ratio computation, and our own *Average Diff* method.

The fault detection models are then derived from each feature selected data. We want the model that shows the highest values of TP rate, precision, and F-measure, but the lowest value in FP rate. The experimental results on the four model measurement metrics are shown in Fig. 3– Fig. 6.

For this specific data domain, it can be noticed that feature selection can considerably improve the accuracy of

fault detection models. The proposed *AverageDiff* method contributes the most to decision tree model, whereas the gain ratio method is the best feature selection method for the naive Bayes and logistic regression model building approaches. The k-nearest neighbor method (in which k was set to be one on our experiments because it yields the best result) needs a cleaned dataset without any other feature selection facility.. It is worth mentioning here that for such a large number of features like this application the neural network and support vector machine approaches consume so much memory that they cannot run to completion.

Among the four model building methods, naive Bayes model can detect fault cases at the success rate as high as 90%, but the false alarm (FP rate) is also as high as 80% as well. We compare the TP rate versus the FP rate of each model and provide the result in Fig. 7.

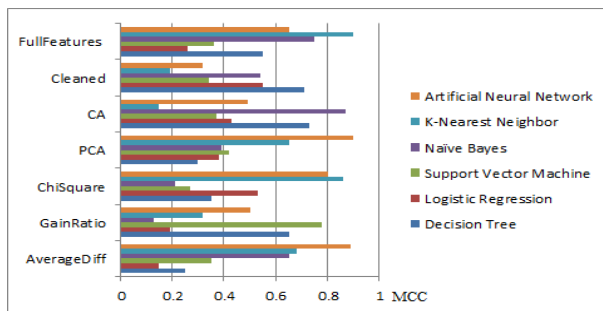


Fig. 13. MCC on different predictive models.

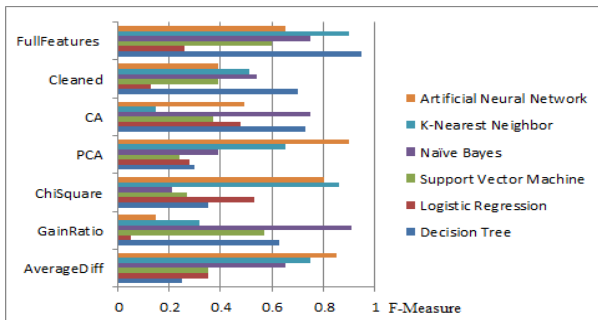


Fig. 14. F-Measure on different predictive models.

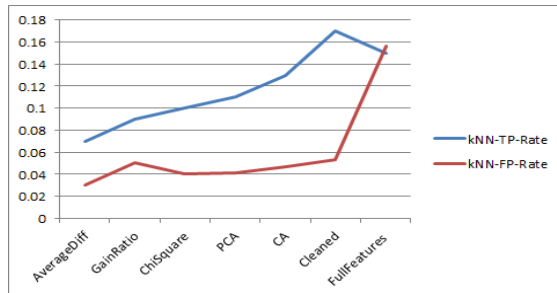


Fig. 15. TP-Rate vs. FP-Rate for k-NN.

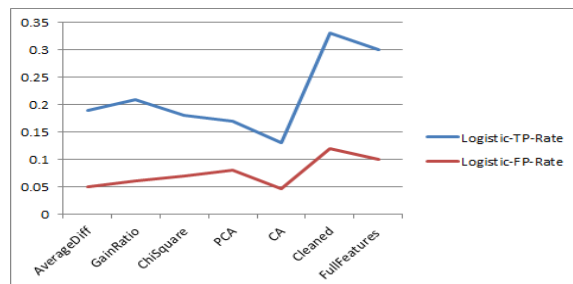


Fig. 16. TP-rate vs. FP-rate for logistic.

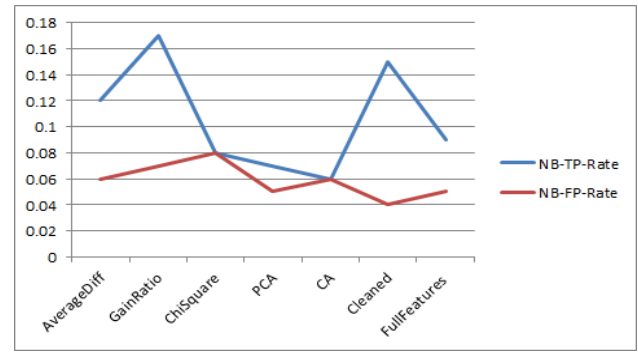


Fig. 17. TP-Rate vs. FP-Rate for naive Bayes.

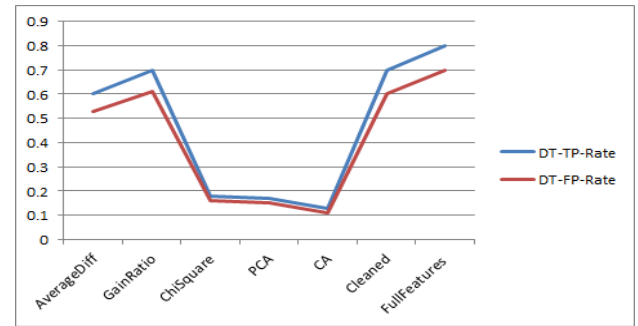


Fig. 18. TP-rate vs. FP-rate for decision tree.

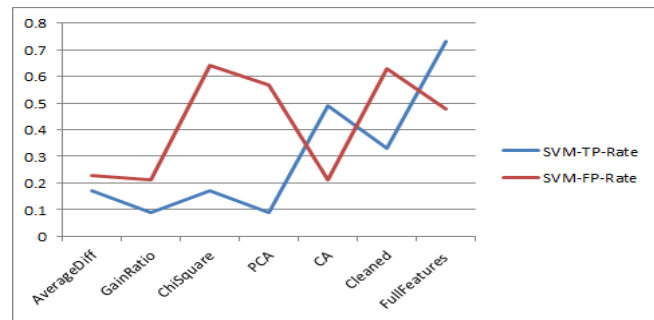


Fig. 19. TP-rate vs. FP-rate for SVM.

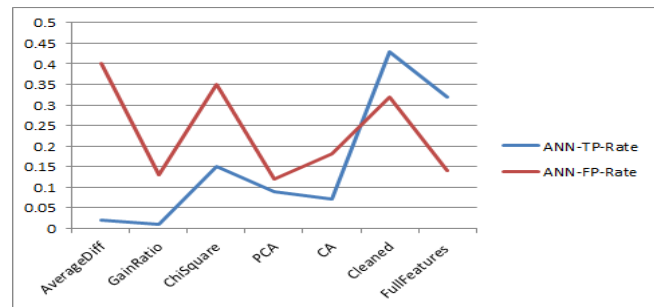


Fig. 20. TP-Rate vs. FP-Rate for ANN.

B. Comparative of Machine Learning Models

In the SECOM dataset, the number of failure test points is very few (just 104 instances) comparative to the number of pass test (1463 instances). It is therefore a difficult task to build automatically the accurate predictive model that can detect such minority cases. We thus propose the idea of over-sampling the minority cases. The process start by separating the SECOM dataset into a train set and a test set. The test set contains 468 instances in which 59 instances are fail test and 409 are pass test. The train set contains 45 instances of fail test and 1054 of pass test (see Fig. 13-Fig. 20).

We then duplicate the number of fail test in the training data to be 1096 instances. The fault detection models are built from this minority case boosting training

dataset. The models are then evaluated their classification performances by the separated test dataset. The classification error matrices of models built from the five different machine learning models are given in Fig. 8 and the performance criteria are summarized in Table 1. In this research, we comparatively study five machine learning algorithms that create these predictive models with minority class datasets. Classifying minority class is difficult because size of the class is too small. Many researchers have tried to solve this problem, Alhamamdy and Ramanohanarao [1] proposed an algorithm called EPDT (Emerging Pattern Decision Tree) to train a decision tree that can classify minority class (see Fig. 21-Fig. 23).

A 3-fold cross-validation approach was used for performance evaluation. The entire data set of SECOM data was divided with random generator into 3 subsets. Two of the 3 subsets were combined and used for training.

Data from the third subset served as the evaluation set during the training phase. The entire process was repeated 2

more times by rotating the subset used as the evaluation set during training.

SECOM Dataset		
Subset 1	Subset 2	Subset 3

ML Algorithms/	WEKA – Performance Matrix					
	KN N	LR	N B	Decision Tree	ANN	SV M
TP	0.98	1.0	0.746	1.0		
FP	0.24	0.335	0.352	0.161		0.2
Pre	0.37	0.301	0.234	0.472		0.6
F-	0.54	0.463	0.356	0.641		0.5
M	0.2	0.27	0.3	0.41		0.
RO	0.5	0.46	0.2	0.43		0.
PR	0.1	0.32	0.2	0.42		0.
Rec	0.3	0.31	0.4	0.51		0.

Fig. 21. WEKA performance matrix.

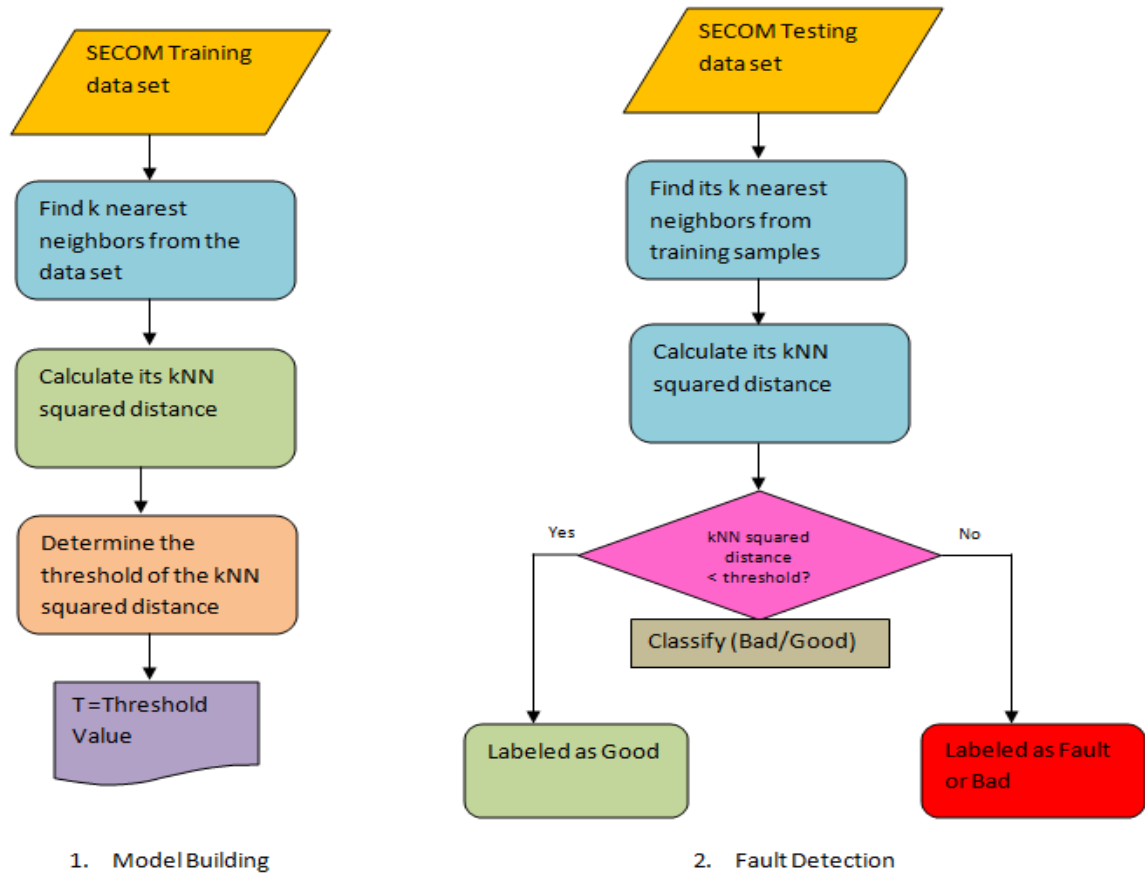


Fig. 22. Overview of K-NN PDM.

C. K-Nearest Neighbor

Due to the unique characteristics of the semiconductor processes, such as nonlinearity in the data, a predictive model using the k-nearest neighbor rule (PD-kNN) is developed in this paper. The kNN rule is an intuitive concept and its basic idea is given as the following: For a given unlabeled sample x , the kNN rule finds the k-nearest labeled samples in the training data set and assigns x to the class that appears most frequently within the k-subset (i.e., k-nearest neighbors).

The proposed predictive model using the kNN rule (PD-

kNN) is based on the fault sample's distance to the nearest neighboring training samples must be greater than a normal samples' distance to the nearest neighboring training samples. The idea is that to determine a threshold (t) with certain confidence level. The proposed method consists of two parts:

- 1) Predictive Model Building
- 2) Fault Detection using Classification

1) Predictive model building

The kNN squared distance is defined as the sum of squared distances of sample i to its k-nearest neighbors.

$$D^2_i = \sum_{j=1}^k d^2_{ij}$$

where d^2_{ij} denotes squared Euclidean distance from sample i to its j th neighbor.

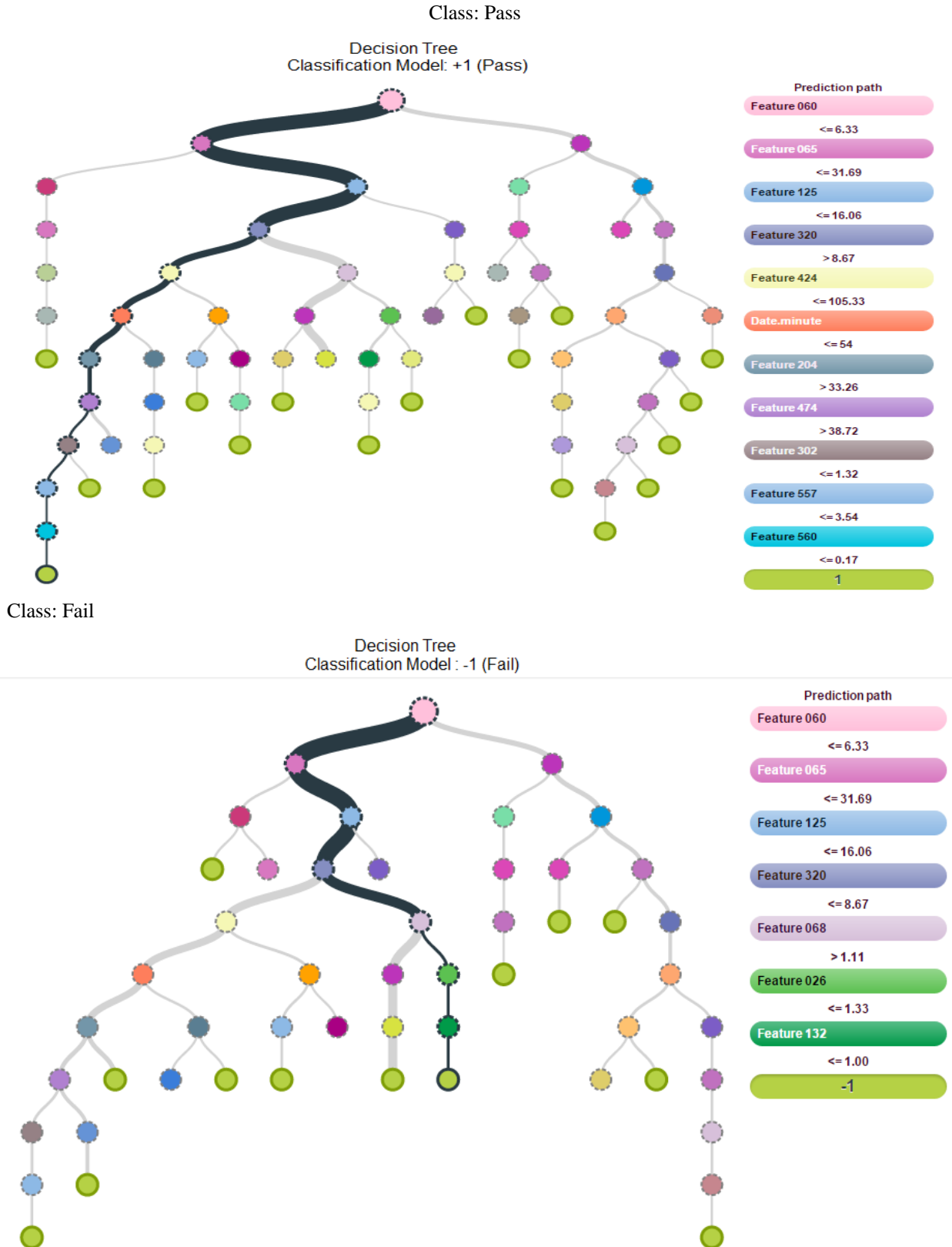


Fig. 23. SECOM decision tree model.

2) Fault detection using classification

For an incoming training dataset x , the fault detection part consists of 3 steps:

- 1) Finding x 's k-nearest neighbors from the training data set.

- 2) Calculation of x 's kNN squared distance D^2_x

- 3) D^2_x Comparison of against the threshold T .

If $D^2_x \leq T$, it is classified as a normal, Otherwise, it is detected as a fault.

Confusion Matrix:

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual Class	Class= 1	58	1
	Class= -1	98	311

D. Decision Tree

Decision tree induction is a powerful technique to discover a tree model for future event prediction. A decision tree is a tree-shaped structure that represents sets of decisions. Each non-terminal node represents a test or decision to be carried out on a single attribute value (i.e., input variable value) of the considered data item, with one branch and sub-tree for each possible outcome of the test.

Confusion Matrix

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual Class	Class= 1	59	0
	Class= -1	66	343

E. Naïve Bayes

A Naïve Bayes classifier is a probabilistic-based classifier that has been initially introduced by Duda and Hart [36] in 1973. SECOM data set is an imbalanced dataset and Naïve Bayes is a frequently proposed to imbalanced dataset problems. Naive Bayes induction algorithm is very popular in classification field. The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited in Semiconductor industry, when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods (see Fig. 24-Fig. 27).

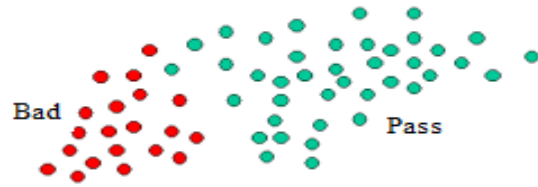


Fig. 24. SECOM naïve bayes model.

Confusion Matrix:

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual Class	Class= 1	44	15
	Class= -1	144	265

In this confusion matrix, the model correctly predicted the pass class for predictive maintenance 44 times and incorrectly predicted it 15 times. The model correctly predicted the negative class for predictive 144 times and incorrectly predicted it 265 times.

The Naïve Bayes algorithm 'is based on conditional probabilities'. Bayes' Theorem is the foundation of this technique. It uses a formula that 'calculates a probability by counting the frequency of values and combinations of values in the historical data'. It is as follows:

$$\text{Prob}(B \text{ given } A) = \text{Prob}(A \text{ and } B) / \text{Prob}(A)$$

We propose a new model on the classification performance of naive bayes for predicting the equipment maintenance in Semiconductor manufacturing processes.

1) Support vector machine

Recently SVM which was developed by Vapnik (1995) is one of the methods that are receiving increasing attention with remarkable results. SVM implements by constructing an optimal separating hyper plane in the hidden feature space using quadratic programming to find a unique solution

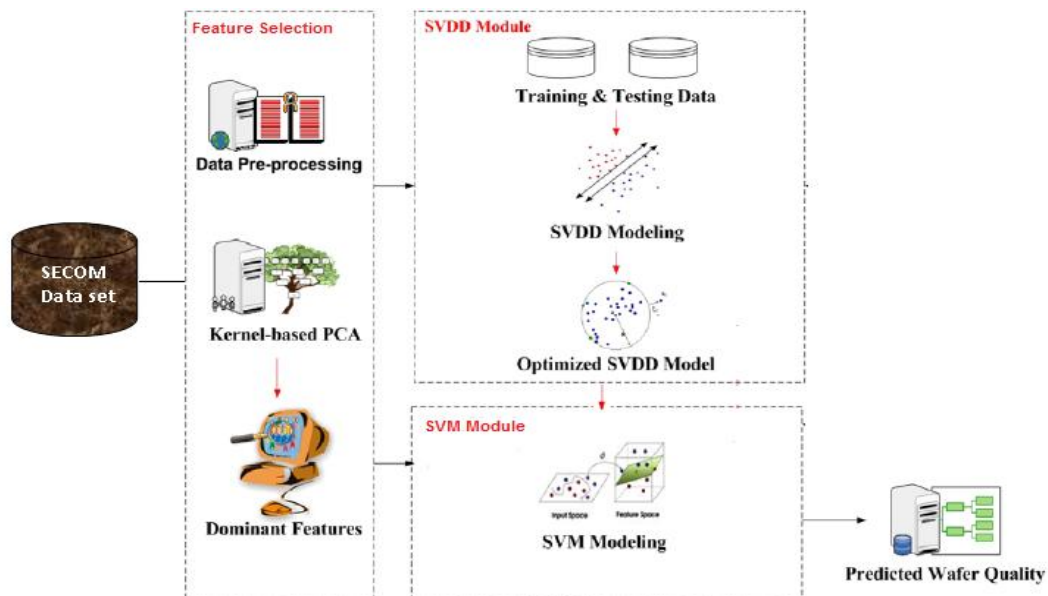


Fig. 25. SECOM SVM model.

Kernel-based PCA is used in order to implement Support Vector Machine for classification. KPCA is a non-linear

extension of PCA. This research chooses support vector data description (SVDD) to construct the predictive model.

The advantage of the SVDD is that an estimate of the error on the target set can be obtained immediately by looking at the fraction of support vectors. For SVM modeling, we use MAPE (Mean Absolute Percentage Error), MaxError (Maximal Error) and RMSE (Root Mean Square Error) indexes to guarantee the performance of SVM model.

the future until it exceeds a threshold value The popular approach for developing artificial faults into specific equipment, acquire data representative of each fault and train the nets to classify them, cross-validating and testing the trained system with data not used for training. The setup of network parameters in an artificial neural network includes learning trials, learning rate, and the momentum correlation coefficient. Back-propagation neural.

A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

Confusion Matrix:

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual Class	Class= 1	38	23
	Class= -1	127	79

1) Logistic regression

Logistic Regression is one of the simpler classification models. For predicting unexpected variations in the equipment, we use a simple approach of historic trend analysis. We use a regression model to estimate the trend.

Confusion Matrix:

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual Class	Class= 1	59	0
	Class= -1	137	272

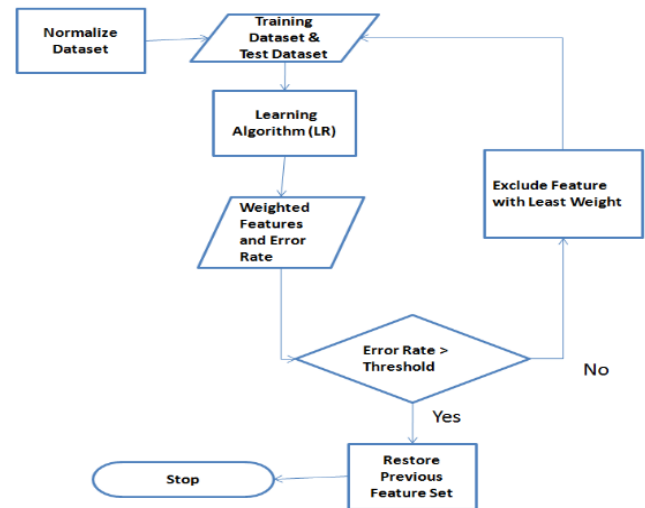
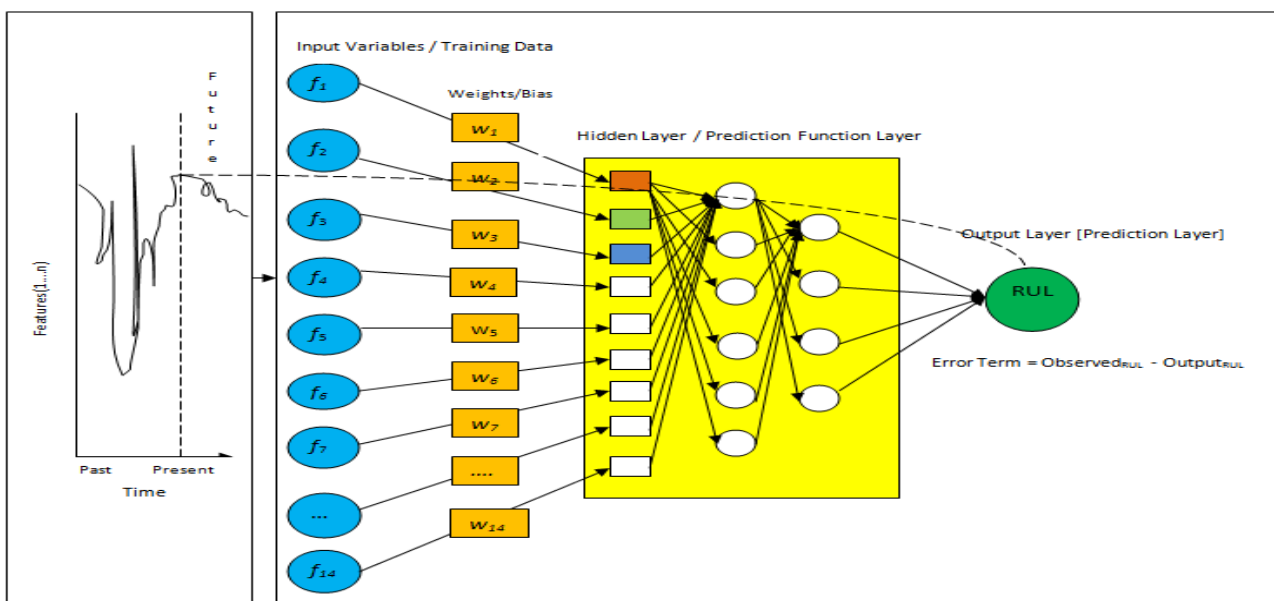


Fig. 26. Logistic regression flowchart.

2) Artificial neural network (ANN)

Artificial Neural Network (ANN) is a fast-growing method which has been used in different industries during recent years. Artificial Neural Network (ANN) based methods have been extensively investigated for equipment health condition prediction. Artificial neural network (ANN) methods have shown great promise in achieving more accurate equipment remaining useful life prediction. Network (BPN) model is a learning model in the neural network and the most representative one. Back-propagation neural network has improvements towards the increases the hidden layer and uses a smooth differentiable transfer function. The proposed method in our research is **14-H-1** ANN Model (14 variables with hidden layer function and 1 output). As it can be seen from the ANN architecture, the input layer corresponds to the 14 measure parameters. The desired outputs from the ANN are unique combinations of 28 primary numbers arranged in a graphical display. The training process of the ANN stopped when it showed the best performance based on a selected number of hidden neurons and weights for the network. The ANN can be named as 14-H-32 according to its structure.



ANN Model: 14-H-1
Fig. 27. Proposed ANN model.

The important element in ANN model is to find the best transfer function for each of the nodes. We have applied sigmoid function (called as logistic function) and its formula looks like this:

$$f(\text{input}) = \frac{1}{1 + e^{-\text{input}}}$$

Here f is the activation function that activates the neuron, and e is widely used mathematical constant that has the approximate value of 2.718. In the researching process, we have faced problem due to incompleteness or missing of the data. Due to the restriction of time and ability, the training of the artificial neural back-propagation network is still not flawless.

VIII. CONCLUSION

The semiconductor industry is one of the most capital-intensive industries with a high of capital investment on equipment's. Optimization of manufacturing equipment's has received significant attention and shown to be a necessary competitive advantage. There are exciting challenges and opportunities for the engineers and researchers to develop a new standard for this vigorously growing industry. A good classification prediction model is beneficial for the prediction in semiconductor manufacturing fabrication process. Most semiconductor manufacturing is highly complex and has produced constantly hundreds of metrology data that are awaiting for process engineers to analyze for the purpose of maintaining efficient operations and getting optimum yield of high quality products. For such a large volume of measurement data, automatic data analysis technique such as data mining is essential.

APPENDIX A: SEMCOM: FACT SHEET.

Repository URL:
<http://www.causality.inf.ethz.ch/data/SECOM.zip>

Title: SEmi COnductor Manufacturing

Authors: Michael McCann, Yuhua Li, Liam Maguire, Adrian Johnston
 Contact name, email and website: Michael McCann, mccann-m15@email.ulster.ac.uk, www.isrc.ulster.ac.uk

Key facts: The data consists of 2 files the dataset file SECOM consisting of 1567 examples each with 591 features a 1567 x 591 matrix and a labels file containing the classifications and date time stamp for each example. The Dataset is presented with features in columns each representing a recorded measurement and product examples in rows. The labels file then represents a simple pass/fail classification corresponding to each row in the dataset where -1 corresponds to fail and +1 corresponds to a pass. A date time stamp for each each pass/fail is also provided in the labels file corresponding to a selected functionality test. The data contains null values varying in intensity depending on the individuals features corresponding to data points with no recorded measurements in the original metrology data.

Keywords: Causal discovery, feature selection, semiconductor manufacturing, industry, business.

REFERENCES

- [1] H. Alhammady and K. Ramamohanarao, "The application of emerging patterns for improving the quality of minority-Class classification," in *Proc. the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2004, pp. 207-211.
- [2] H. Alhammady and K. Ramamohanarao, "Using emerging patterns and decision trees in minority-class classification," in *Proc. the International Conference on Data Mining Series*, 2004, pp. 315-318.
- [3] J. Burez and D. V. D. Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications*, vol. 36, 2009, pp. 4626-4636.
- [4] G. Costa, M. Guarascio, G. Mamco, R. Ortale, and E. Ritacco, "Rule learning with probabilistic smoothing," in *Proc. the International Conference on Big Data Analytics and Knowledge Discovery*, 2009, pp. 428-440.
- [5] E. Epistola, *Directory of Dealers in Used Electronic and Semiconductor Manufacturing*, 2011.
- [6] R. Furferi and L. Governi, "Development of an artificial vision inspection system for real-time defect detection and classification on circular knitting machines," *WSEAS Transactions on Computers*, vol. 5, no. 6, pp. 1186-1193, 2006.
- [7] M. Gallo, G. Guizzi, and V. Zoppoli, "An integrated approach to develop a simulation model in manufacturing processes," *International Journal of Systems Applications, Engineering and Development*, vol. 1, pp. 137-145, no. 4, 2007.
- [8] Z. Ge and Z. Song, "Semiconductor manufacturing process monitoring based on adaptive substatistical PCA," *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 1, pp. 99-108, 2010.
- [9] B. E. Goodlin, D. S. Boning, H. H. Sawin, and B. M. Wise, "Simultaneous fault detection and classification for semiconductor manufacturing tools," *Journal of Electrochemical Society*, vol. 150, no. 12, pp. G778-G784, 2003.
- [10] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.
- [11] S. Han, B. Yuan, and W. Liu, "Minority class mining: progress and prospect," in *Proc. the Chinese Conference on Pattern Recognition*, 2009, pp. 1-5.
- [12] Q. P. He and J. Wang, "Fault detection using the k- nearest neighbour rule for semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, 2007, pp. 345-354.
- [13] A. M. Ison, W. Li, and C. J. Spanos, "Fault diagnosis of plasma etch equipment," in *Proc. the IEEE Int. Symposium on Semiconductor Manufacturing*, 1997, pp. B49-B52.
- [14] A. M. Ison and C. Spanos, "Robust fault detection and fault classification of semiconductor manufacturing equipment," in *Proc. Int. Symposium on Semiconductor Manufacturing*, Tokyo, 1996.
- [15] H. D. Lin, C. Y. Chung, and W. T. Lin, "Principal component analysis based on wavelet characteristics applied to automated surface defect inspection," *WSEAS Transactions on Computer Research*, vol. 3, no. 4, 2008, pp. 193-202.
- [16] G. S. May and C. J. Spanos, *Fundamentals of Semiconductor Manufacturing and Process Control*, John Wiley & Sons, 2006.
- [17] M. McCann, Y. Li, L. Maguire, and A. Johnston, "Causality challenge: benchmarking relevant signal components for effective monitoring and process control," in *Proc. the JMLR Workshop*, 2008, pp. 277-288.
- [18] S. J. Qin, G. Cherry, R. Good, J. Wang, and C. A. Harrison, "Semiconductor manufacturing process control and monitoring: A fab-wide framework," *Journal of Process Control*, vol. 16, 2006, pp. 179-191.
- [19] C. Seiffert, T. M. Khoshgoftar, and J. V. Hulse, "Improving software-quality predictions with data sampling and boosting," *IEEE Transactions on Systems, Man, Cybernetics-Part A: Systems and Humans*, vol. 39, no. 6, 2009, pp. 1283-1294.
- [20] SECOM, SEmiCOnductor Manufacturing, [Online]. Available: <http://www.causality.inf.ethz.ch/repository.php>.
- [21] G. Spitzlsperger, C. Schmidt, G. Ernst, H. Strasser, and M. Speil, "Fault detection for a via etch process using adaptive multivariate methods," *IEEE Transactions on Semiconductor Manufacturing*, vol. 18, no. 4, 2005, pp. 528-533.
- [22] J. Stefanowski and S. Wilk, "Selective pre-processing of imbalanced data for improving classification performance," in *Proc. the International Conference on Big Data Analytics and Knowledge Discovery*, 2008, pp. 283-292.

- [23] S. K. Subramaniam, S. H. Fusin, R. S. S. Singh, and A. H. Hamidon, "Production monitoring system for monitoring the industrial shop floor performance," *International Journal of Systems Applications, Engineering & Development*, vol. 3, no. 1, 2009, pp. 28-35.
- [24] E. Tafazzoli and M. Saif, "Application of combined support vector machines in process fault diagnosis," in *Proc. the American Control Conference*, 2009, pp. 3429- 3433
- [25] J. V. Hulse and T. Khoshgoftaar, "Knowledge discovery from imbalanced and noisy data," *Data and Knowledge Engineering*, vol. 68, 2009, pp. 1513-1542.
- [26] G. Verdier and A. Ferreira, "Adaptive mahalanobis distance and k-nearest neighbour rule for fault detection in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 1, pp. 59-68, 2011.
- [27] Z. Wang, S. Ji, L. Zhang, Y. Huan, and G. Yang, "Research on automatic detecting technology of Razor's assembling quality based on the machine vision system," *WSEAS Transactions on Computers*, vol. 5, no. 6, pp. 1319-1324, 2006.



knowledge discovery. He presented lots of papers on predictive modeling.

Sathyan Munirathinam is a business intelligence engineer at Micron Technology, and a PhD scholar from Bharathiar University, Tamil Nadu, India. He received his bachelor and master degrees in computer science from Madurai Kamaraj University, Madurai in India. He received his second master degree from Illinois Institute of Technology, Chicago, USA. His current research includes predictive analytics, data mining, machine learning, databases and



warehousing and XML database management. He is a life member of ISTE, New Delhi India.

Balakrishnan Ramadoss is a computer science professor from Department of Computer Applications, National Institute of Technology, Tiruchirapalli, India. He received his M.Tech degree in computer science and engineering from Indian Institute of Technology, Delhi, India. He received his PhD in applied mathematics from Indian Institute of Technology, Powai. His research interests are data mining, data