Journal of the Korean Institute of Industrial Engineers
Vol. 41, No. 6, pp. 572-578, November 2015.
ISSN 1225-0988 | EISSN 2234-6457
http://dx.doi.org/10.7232/JKIIE.2015.41.6.572

<Application Research>

반도체 제조 가상계측 공정변수를 이용한 웨이퍼 수율 예측

남완식 · 김성범†

고려대학교 산업경영공학과

A Prediction of Wafer Yield Using Product Fabrication Virtual Metrology Process Parameters in Semiconductor Manufacturing

Wan Sik Nam · Seoung Bum Kim

Department of Industrial Management Engineering, Korea University

Yield prediction is one of the most important issues in semiconductor manufacturing. Especially, for a fast-changing environment of the semiconductor industry, accurate and reliable prediction techniques are required. In this study, we propose a prediction model to predict wafer yield based on virtual metrology process parameters in semiconductor manufacturing. The proposed prediction model addresses imbalance problems frequently encountered in semiconductor processes so as to construct reliable prediction model. The effectiveness and applicability of the proposed procedure was demonstrated through a real data from a leading semiconductor industry in South Korea.

Keywords: Classification, Yield prediction, Virtual metrology, Semiconductor industry

1. 서 론

반도체 산업은 개인용 컴퓨터에서 모바일까지 첨단 산업의 발전과 수요처의 다변화 및 고도화에 따라 시장규모가 지속적으로 성장하고 있다. 반도체 제조 공정은 생산 초기 단계부터 제품을 완성하는 최종 단계까지 수많은 공정변수들을 모니터링하는 복잡한 과정으로 구성되어 있다(Shin and Park, 2000). <Figure 1>은 수백 개의 정밀 공정으로 진행되는 반도체 제조 공정을 요약하여 보여주고 있다. 대표적으로 원재료가 투입된 후 제작공정(이하 FAB 공정), 프로브(Probe) 검사, 조립, 그리고 패키지 검사 등 크게 네 단계로 구분 할 수 있다(Uzsoy *et al.*, 1992). FAB 공정은 잉곳(Ingot)을 가공하여 만든 웨이퍼(Wafer) 25매를 1랏(Lot)으로 구성하여 FAB에 투입한 후, 패턴 생성 공정, 층 생성 공정 등의 세부 공정을 통하여 웨이퍼 위에 수천 개의 집적 회로를 형성하는 단계이다. 프로브 검사 공정은 웨이퍼 상태에서 각각의 칩(Chip)을 전기적으로 테스트하여 양품과 불량품으로 판별하는 단계이다. 조립 공정은 프로브 검사 공정 이후에 웨이퍼 단위에서 칩 단위로 분리하여 칩의 전기적, 물리적 특성을 향상시키고 외부의 기계적, 물리적 충격으로부터 칩을 보호하기 위한 단계이다. 마지막으로 패키지 검사 공정은 조립된 칩의 전기적 특성 및 기능, 신뢰성 등을 검사하여 양품과 불량품으로 구분하는 단계이다(Baek and Han, 2003). 특히, 본 연구에서 중점적으로 다룰 FAB 공정은 반도체 제조 공정 중에서 기술적으로 가장 복잡하고, 자본 집약적인 단계 중 하나이다. 또한 FAB 공정은 하루에 수만 매의 웨이퍼를 생산하고 있으며 한 웨이퍼 당 1,000개 이상의 공정변수가 생성되고 있어 데이터 관점에서 볼 때 방대한 양의 데이터가 생성됨을 알 수 있다(Shin and Park, 2000).

<Figure 2>는 반도체 FAB 공정의 대표적인 9개 단위 공정을 보여주고 있다. 각 단위 공정 단계에서 생성되는 공정 데이터, 설비 데이터, 그리고 웨이퍼 히스토리 정보와 같은 수많은 공정변수는 서로 상호 연관되어 웨이퍼 수율에 영향을 준다(Chen *et al.*, 2006). 수율은 반도체 품질의 대표적인 측정 기준으로(Kumar *et al.*, 2006), 제조 공정에서 높은 수율을 유지하는 것은

† 연락저자 : 김성범 교수, 02841 서울시 성북구 안암로 145 고려대학교 산업경영공학과, Tel : 02-3290-3397, Fax : 02-929-5888, E-mail : sbkim1@korea.ac.kr
2015년 2월 9일 접수; 2015년 6월 16일 수정본 접수; 2015년 6월 16일 게재 확정.

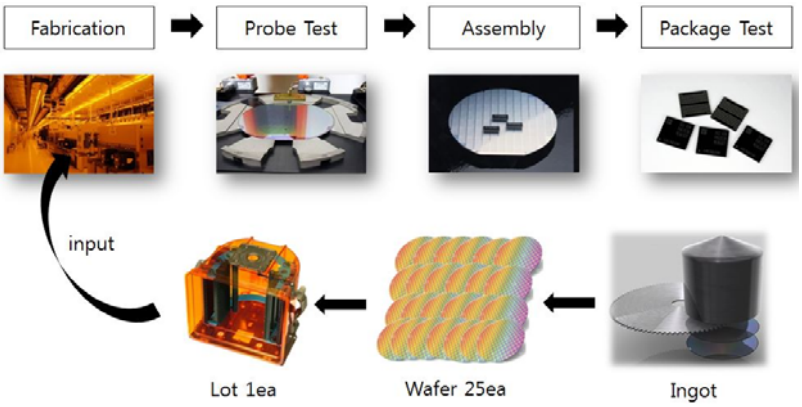

Figure 1. Semiconductor manufacturing process

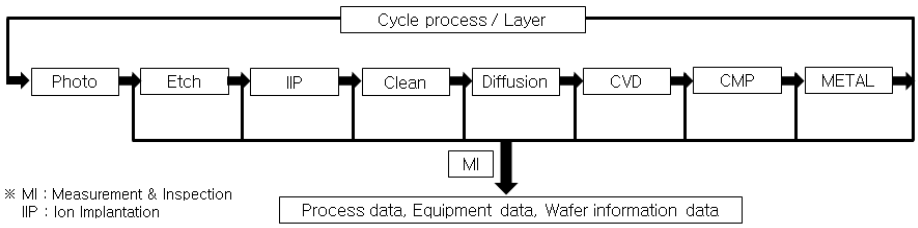

Figure 2. Semiconductor fabrication process

반도체 산업에 있어서 매우 중요하다고 하겠다(Shin and Park, 2000). 일반적으로 반도체 수율은 투입량 대비 완성된 양품의 비율을 의미하며, <Figure 1>에서 설명한 반도체 제조 공정 단계에 따라서 FAB 수율, 프로브 수율, 조립 수율, 그리고 최종 수율로 구분된다(Baek and Han, 2003). FAB 수율은 투입된 웨이퍼 매수 대비 프로브 검사에 도달하기 전 폐기되지 않은 웨이퍼 매수의 비율이다. 프로브 수율은 프로브 검사 전 웨이퍼 내 총 다이 수 대비 프로브 검사 결과 양품인 다이 수의 비율이다. 조립 수율과 최종 수율은 각각 투입 양 대비 완성된 양품의 비율로 나타낸다. 반도체 제조 공정에서는 일반적으로 FAB 수율과 프로브 수율이 가장 중요한 품질 측정 기준이다(Park *et al*., 1997). 특히, 반도체 제품의 보다 빠른 수율 예측으로 저수율 제품의 원인을 찾아 이를 개선하여 조기에 고수율 제품으로 전환 생산하는 것은 급변하는 반도체 시장 대응을 위한 필수 조건이다.

수율 예측과 관련된 대표적인 연구들은 다음과 같다. Murphy (1964)은 FAB 공정에서 웨이퍼에 유발되어 품질 불량을 야기하는 결점의 밀도 분포 함수를 이용하여 수율을 예측하는 모델을 구축하였다. Chunningham *et al*.(1995)은 서로 다른 반도체 제조 공장의 환경을 대표하는 변수를 이용하여 선형 회귀 모델을 사용하여 FAB 전체 수율을 예측하는 모델을 구축하였다. Shin and Park(2000)은 FAB 공정에서 발생된 공정변수를 이용하여 변수선택기법, 인공신경망기법, 메모리 기반 추론(Memory Based Reasoning)기법을 사용하여 랏 단위 프로브 수율을 예측하는 모델을 구축하였다. Li *et al*.(2006)은 저수율의 원인이 되는 FAB 공정의 공정변수를 이용하여 유전자 프로그래밍(Genetic Programming) 기법을 사용하여 랏 단위로 프로브 수율을 예측하고 분류하는 모델을 구축하였다. Chien *et al*.(2007)은 FAB 공정에서 측정된 데이터를 변수로 이용하여 K-평균 군집분석, Kruskal-Wallis 검정, 의사결정나무기법을 사용하여 랏 단위로 프로브 수율을 예측하는 모델을 구축하였다. 또한 An *et al*.(2009)은 공정 제어 모니터링 데이터와 프로브 검사 공정에서 출력된 데이터를 변수로 이용하여 SSVM(Stepwise Support Vector Machine) 기법을 사용하여 랏 단위의 최종 수율을 고수율과 저수율로 구분하는 분류 예측 모델을 구축하였다.

이와 같이 수율 예측과 관련된 기존 연구들은 FAB 공정 초기단계에서부터 프로브 검사 공정 이후 단계까지 측정 및 검사되는 공정변수 데이터를 이용하여 랏 단위 수율 예측 모델을 구축하였다. 하지만 반도체 제조 공정은 갈수록 높은 복잡도와 향상되는 정밀도로 인하여 보다 정확한 품질 관리가 요구되기 때문에(Kang *et al*., 2012), 지금까지의 랏 단위 수율 예측 모델은 한계를 가지고 있다. 그러므로 웨이퍼 단위로 수율을 관리하기 위해서는 제조 과정에서 각각의 웨이퍼에 대한 수율을 정확하게 예측 할 필요가 있다. 웨이퍼 단위로 수율을 예측하여 보다 정밀하고, 정확한 제조운영이 이루어지면 결국 랏 전체의 수율을 높이게 되어 최종적으로 FAB 공정 전체의 품질이 향상된다. 이를 해결하기 위해 가상계측 방법론이 연구되었다.

가상계측과 관련된 대표적인 연구는 다음과 같다. Chen *et al*.(2005)은 웨이퍼와 웨이퍼간의 고급 공정 관리(Advanced Process Control)를 위해 가상계측을 이용하여 웨이퍼의 공정 상태를 예측하는 방법을 제안하였다. Khan *et al*.(2008)은 가상계측을 활용하여 공장 전체 관리(Factory-Wide Control)를 위한

방법을 제안하였다. Kang *et al.*(2011)은 다양한 데이터마이닝 기법들을 사용하여 가상계측 예측 시스템을 제안하였다. Lynn *et al.*(2012)은 반도체 식각(Etch) 공정을 진행한 웨이퍼의 상태를 모니터링하기 위한 가상계측 모델을 구축하였다.

<Figure 3>은 가상계측과 실제계측의 개념을 보여 주고 있다. 실제계측은 계측 장비의 측정 능력 한계와 측정 시간 단축을 위해 샘플링 방법으로 계측이 이루어지기 때문에 모든 랏과 모든 웨이퍼의 공정변수 정보를 얻을 수 없다. 이와 달리 가상계측은 실제계측을 하지 않은 웨이퍼도 데이터를 통해 구축한 예측 모델을 통해 공정변수 정보를 얻을 수 있는 기법이다(Kang *et al.*, 2009). 가상계측으로 얻은 모든 웨이퍼에 대한 품질 정보를 통하여 보다 정확하게 웨이퍼 단위로 수율을 예측할 수 있고, 반도체 제조 단계에 있는 각각의 공정을 더욱 정확하게 관리 할 수 있으며, 최종 FAB 공정 전체 품질을 향상 시킬 수 있다(Chang *et al.*, 2006).

본 연구에서는 FAB 공정에서 생성된 가상계측 데이터를 공정변수로 사용하여 웨이퍼 단위로 프로브 수율을 고수율 혹은 저수율로 판별하는 예측 모델을 제안하고자 한다. 이를 통하여 FAB 공정 단계에서 웨이퍼 단위로 수율을 예측함으로써 조기에 저수율 제품을 선별 폐기하고 신속하게 고수율 제품을 전환 생산하여 반도체 제조 생산성 향상 및 경비 절감 효과 그리고 고수율 제품 생산을 통한 고객 신뢰성 향상을 기대할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 본 연구에서 제안하는 반도체 제조 공정으로부터 생성된 가상계측 데이터를 이용하여 웨이퍼 수율을 분류 예측하는 방법에 대해 서술하였고, 제 3장에서는 실제 반도체 제조 공정에서 추출한 가상계측 데이터에 제안기법을 적용하여 효과를 입증하였다. 제 4장에서는 본 연구의 결론 및 기대효과와 함께 한계점에 대해 논의하고 향후 연구 방향을 모색하였다.

2. 반도체 웨이퍼 수율 예측 모델링 방법

기존 수율 예측 모델은 수율 예측 가능변수 및 변수 적용 알고리즘에 따라 다양하게 연구되어왔다. 본 연구에서 제안하는 수율 예측 모델은 반도체 제조 공정에서 생성된 가상계측 데이터를 기반으로 웨이퍼를 고수율, 저수율로 구분하는 분류 모델을 구축하는 것이다. 제안하는 수율 예측 모델링 프로세스는 <Figure 4>와 같이 4단계로 이루어지며 자세한 설명은 다음과 같다.

• Step 1 : 데이터 전처리

<Figure 4>에서 보여주듯이 첫 번째 단계는 데이터 전처리 작업이다. 도메인 지식을 바탕으로 가상계측 데이터 중 웨이퍼 수율을 대표하는 특징을 갖는 변수를 선택한다. 도메인 지식 외에도 데이터마이닝 알고리즘을 기반으로 하는 여러 가지 변수선택 기법들이 있으나, 엔지니어의 경험을 통해 변수를 선택하는 방법이 가장 강건하므로, 본 연구에서는 최종적으로 도메인 지식을 바탕으로 웨이퍼 수율에 영향을 주는 변수를 선정 하였다.

가상계측은 FAB 공정 진행 중인 웨이퍼의 공정 설비 데이터를 설명 변수로 하고 해당 웨이퍼를 실제로 계측하여 얻은 품질 지표를 목표 변수로 하여 예측 모델을 구축하기 때문에 예측 모델의 정합성에 따라서 가상계측 결과의 신뢰도가 달라진다(Kang *et al.*, 2012). 그러므로 가상계측 데이터 중에서 신뢰도가 낮은 관측치를 제거하기 위해 비지도 이상치 탐지(Unsupervised Anomaly Detection)기법을 사용하였다(Chandola *et al.*, 2009). 본 연구에서 사용한 이상치 탐지 기법은 모든 가상계측 데이터의 관측치를 단일 클래스로 가정한 후 단일 클래스 분류 알고리즘을 사용하여 정의된 단일 클래스의 정상 범위를 벗어나는 관측치를 이상 데이터로 판별하여 제거하는 방법을 사용하였다(Ferreira *et al.*, 2009).

• Step 2 : 데이터 불균형 해결

현실에서 다루는 많은 분류 문제는 불균형 상태에 놓여있으며(Chen *et al.*, 2004), 이는 예측 알고리즘의 성능을 저하시키는 주요 요인 중 하나이다(Kang and Cho, 2006). 반도체 제조 공

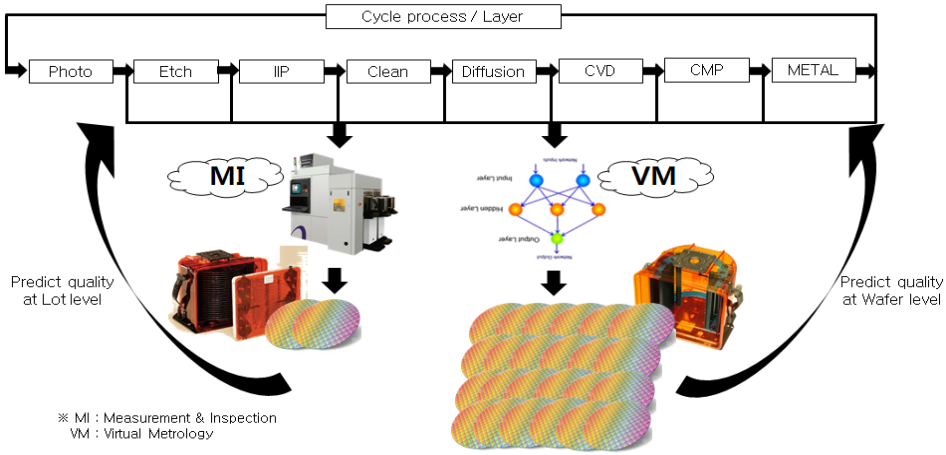

Figure 3. The concepts of actual metrology and virtual metrology

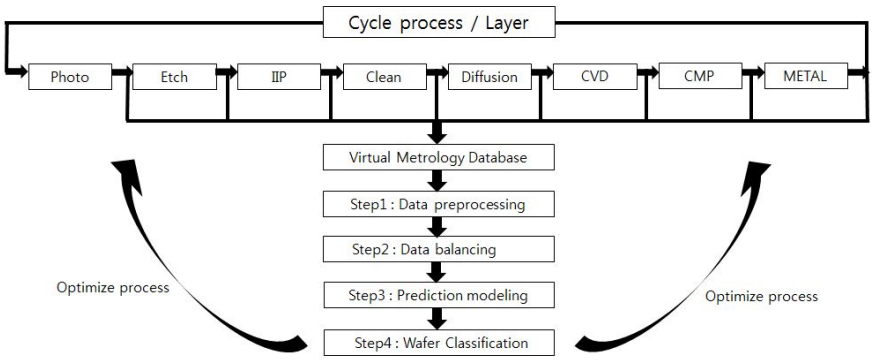

Figure 4. Overview of the proposed prediction modeling procedure

정 또한 제품의 수율이 안정화 되면서 고수율 웨이퍼와 저수율 웨이퍼의 비율이 심각한 불균형 상태가 된다. 그러므로 웨이퍼 수율을 고수율, 저수율로 구분하는 분류 모델 구축을 위해서는 데이터 불균형 문제를 해결해야 한다. 본 연구에서는 여러 가지 방법 중 SMOTE(synthetic minority over-sampling technique) 기법을 사용하여 데이터 불균형 문제를 해결하였다. SMOTE 알고리즘은 소수 계층 과다추출, 다수계층 과소추출의 조합으로 데이터 불균형 문제를 해결하는 방법이다. 샘플링 방식으로 K이웃근접기법을 사용하여 새로운 소수 계층의 데이터를 생성하여 소수 계층의 결정 경계를 넓히는 효과를 가져와 기존의 복원 샘플링 기법이 가지고 있던 데이터 과적합 문제를 해결해주는 장점이 있다(Chawla *et al*., 2002).

- **Step 3 : 수율 예측 모델 구축**

이 단계에서는 전 단계에서 전 처리한 가상계측 데이터를 독립변수로 구성하고, 제품의 수율을 품질 기준에 의하여 고수율, 저수율로 구분한 범주형 변수를 종속변수로 활용하여 분류모델을 구축한다. 본 연구에서는 분류 문제에 널리 사용되고 있는 다양한 분류 알고리즘을 적용하여 비교해 보았다. 예측 모델의 성능을 비교 검증하기 위해 고려된 기법들은 다음과 같다.

결과에 대한 해석이 쉽고 학습 시간이 짧은 의사결정나무(Quinlan, 1986), 최적의 초평면으로 분류 문제를 해결하는 SVM(Vapnik, 2000), 선형 모델로 분류 문제를 해결하는 로지스틱 회귀분석(Hosmer and Lemeshow, 1989), 패턴인식이나 데이터 마이닝 등 다양한 분야에 응용되고 있는 인공신경망(Rosenblatt, 1958) 기법을 사용하였고, 마지막으로 랜덤 포레스트(이하 RF) 기법을 사용하였다. RF 알고리즘은 나무 유도 과정에서 임의의 변수선택을 사용하여 학습 데이터의 부트스트랩(Bootstrap) 샘플로 유도된 잘라내지 않은 분류 또는 회귀 나무의 앙상블이다(Chen *et al*., 2004). RF 알고리즘은 많은 경우 다른 분류 알고리즘과 비교하여 높은 예측력을 보여주고 있다(Breiman, 2001). 무엇보다도 모델 구축 시 사용자가 결정해야하는 매개변수의 수가 적고, 그 값에 덜 민감하기 때문에 사용자 입장에서 모델 구축이 매우 용이하다고 하겠다(Liaw and Wiener, 2002).

본 연구에서는 R의 데이터마이닝 알고리즘을 사용하여 수율 예측 모델을 구축하였고, 각 알고리즘에 사용된 R 패키지와 최적의 매개변수는 <Table 1>과 같다.

Table 1. R packages and optimal parameters of the prediction models

Algorithm	Package	Optimal parameters
Decision Tree	tree	prune size : 12
SVM	e1071	kernel : radial, cost : 1, gamma : 0.022
Logistic Regression	stats	threshold : 0.4
ANN	nnet	hidden layer : 2
Random Forest	randomForest	ntree : 500, mtry : 4

Step 4 : 새로운 제품의 수율 예측

Step 1부터 Step 3까지의 단계적인 과정을 거쳐 구축된 분류 모델은 새로운 제품의 수율을 예측하는데 사용된다. 즉, 새로운 제품의 가상계측 정보를 구축된 모델에 넣으면 해당 제품의 품질이 고수율인지 저수율인지 예측되는 것이다.

3. 실험계획 및 결과

3.1 실험 데이터

본 연구에서 제안하는 웨이퍼 수율 예측 모델의 효용성을 입증하기 위해 국내 반도체기업의 실제 가상계측 데이터를 적용하였다. 실험에 사용한 데이터는 1개월간 생산된 웨이퍼의 가상계측 데이터이며 사용 변수는 90개, 고수율 웨이퍼의 관측치는 548개, 저수율 웨이퍼의 관측치는 67개로 구성되었다. 고수율 웨이퍼와 저수율 웨이퍼의 비율은 9대 1로 범주간 심한 불균형을 보여주고 있다. 데이터 형태는 연속형 독립변수와 범주형 종속변수로 이루어졌다. <Table 2>에서 보듯이, 실험결과의 안정성을 확보하기 위해 10-Fold Cross Validation으로 데이터를 처리하여 진행하였다(Kohavi, 1995).

Table 2. 10-fold cross validation data for an experiment

Data	Class	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Train	High	494	494	491	491	500	492	495	491	491	493
	Low	59	59	62	62	53	62	59	63	63	61
Test	High	54	54	57	57	48	56	53	57	57	55
	Low	8	8	5	5	14	5	8	4	4	6

3.2 실험 성능 평가

일반적으로 분류문제의 성능 척도는 <Table 3>을 통해 정확도(Accuracy)를 계산하여 사용한다. 하지만 정확도는 범주간 불균형이 심한 데이터의 경우 관측치가 많은 범주에 편향되기 때문에 올바른 평가척도로 사용되기 어렵다. 예를 들어, 소수 범주에 속한 데이터가 0.1%이고, 다수 범주에 속한 데이터가 99.9%인 이진 분류 문제를 생각해보자. 이 경우 정확도를 측정 지표로 사용하게 되면, 모든 데이터를 다수 범주로 분류함으로써 99.9%라는 높은 정확도를 얻을 수 있다. 하지만 소수 범주에 속한 0.1%의 데이터를 전혀 판별하지 못하는 문제점을 갖게 된다(Kang *et al*., 2006).

Table 3. Contingency table

		Predict	
		Positive	Negative
Actual	Positive	True Positive(TP)	False Negative(FN)
	Negative	False Positive(FP)	True Negative(TN)

Note) TP : the proportion of low yielde that are correctly predicted, FN : the proportion of low yields that are incorrectly predicted as high yields, TN : the proportion of high yields that are correctly predicted, FP : the proportion o high yields that are incorrectely predicted as low yields.

따라서 본 연구에서는 불균형 데이터 분류 문제에서 평가척도로 사용하고 있는 식 (1)의 민감도(Sensitivity)와 특이도(Specificity)의 기하평균(이하 GM)을 사용하였다(Kubat *et al*., 1998).

$$GM = \sqrt{Sensitivity \times Specificity} \quad (1)$$

민감도는 실제 총 저수율 대비 모델이 올바르게 저수율이라고 예측한 경우의 비율을 의미하며 식 (2)와 같이 구할 수 있으며, 특이도는 실제 총 고수율 대비 모델이 올바르게 고수율이라고 예측한 경우의 비율을 나타내며 식 (3)과 같이 구할 수 있다. GM은 민감도와 특이도의 곱의 제곱근으로 계산되어 불균형 데이터에도 한 범주에 편향되지 않은 성능척도로 사용될 수 있다.

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (2)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (3)$$

3.3 실험 결과

<Table 4>는 가상계측 데이터의 불균형 정도에 따른 RF 모델 성능을 나타낸 것이다. SMOTE(100 : 100)은 소수계층 과다추출과 다수계층 과소추출 비율을 동일하게 구성하여 샘플링 한 것이다. SMOTE(200 : 100) 부터 SMOTE(500 : 100) 조건은 소수계층 과다추출 샘플링 개수를 단계적으로 증가시킨 것이다. SMOTE를 미적용한 원 데이터를 사용하여 구축한 예측 모델은 GM 기준 21%의 예측 정확도를 보였으나, 고수율과 저수율의 비율을 1 : 1로 동일하게 샘플링하여 예측 모델을 구축한 SMOTE (100 : 100) 조건은 GM 기준 76%의 예측 정확도를 나타내었다. 또한 저수율 데이터의 샘플링을 과도하게 할수록 예측 정확도가 낮아지는 것을 확인 할 수 있었다.

Table 4. Performance of RF for different degrees of imbalance

Dataset	Sensitivity	Specificity	GM
SMOTE(100 : 100)	69%	84%	76%
SMOTE(200 : 100)	51%	92%	68%
SMOTE(300 : 100)	46%	94%	66%
SMOTE(400 : 100)	45%	95%	65%
SMOTE(500 : 100)	42%	96%	63%
Without Using SMOTE	4%	99%	21%

<Table 5>는 본 연구에서 제안한 수율 예측 방법을 다양한 데이터마이닝 분류 알고리즘에 동일하게 적용하여 각 모델의 성능을 비교한 것이다. 가장 우수한 예측 모델은 RF 알고리즘을 사용한 예측 모델로써 76%의 정확도를 보이고 있고, 그 뒤로 support vertor machine 이 74%로 비슷한 정확도를 보이고 있다. 특이한 점은 인공신경망 알고리즘을 사용한 예측 모델은 민감도 기준 72%의 예측 정확도를 나타내었다. 즉, 소수 범주의 분류 문제 해결에 있어서는 인공신경망 알고리즘을 사용한 예측 모델이 우수한 성능을 나타낸다고 할 수 있다.

4. 결 론

통상적으로 반도체 산업에서 수율 예측 모델은 FAB 공정 단계부터 조립공정 전 까지 샘플링으로 측정 및 검사된 공정변수를 통해서 랏 단위로 구축 되었다. 하지만 샘플링을 통해 수집된 공정변수 데이터는 모든 웨이퍼의 품질지표를 대변하는데

Table 5. Experiment result(for all of the methods)

Dataset	Model	Sensitivity	Specificity	GM
SMOTE (100 : 100)	RF	69%	84%	76%
	SVM	64%	85%	74%
	ANN	72%	67%	69%
	Logistic	69%	66%	67%
	D-Tree	66%	69%	67%
SMOTE 미적용	RF	4%	99%	21%
	SVM	16%	99%	40%
	ANN	30%	90%	52%
	Logistic	31%	95%	55%
	D-Tree	28%	93%	51%

한계가 있다. 또한 샘플링 된 공정변수를 통해 구축된 수율 예측 모델 또한 정확한 품질을 예측하는데 한계가 있다.

본 연구에서는 FAB 공정에서 생성된 가상계측 데이터를 기반으로 데이터의 전처리, 데이터의 불균형 문제를 해결한 후 데이터마이닝 분류 알고리즘을 통해서 웨이퍼 단위로 수율을 판별 할 수 있는 예측 모델을 제안하였다. 이렇게 구축된 예측 모델은 FAB 공정 단계에서 제품의 품질 상태를 구분함으로써 생산성 향상에 기여할 수 있을 것이다. 또한 FAB 공정 단계에서 고수율, 저수율 제품을 분석하여 최적의 공정조건을 찾으므로 제조 공정 능력 향상이 가능하다. 제안기법의 한계점으로는 수율 예측 모델의 변수로 이용되는 가상계측 데이터의 정합성에 따라 수율 예측 모델의 성능이 결정되는 것이다. 추후 과제로는 가상계측으로부터 얻어진 오차를 반영한 모델을 구축하여 보다 현실적인 모델을 구축하는 것이다.

참고문헌

An, D., Ko, H. H., Gulambar, T., Kim, J., Baek, J. G., and Kim, S. S. (2009), A semiconductor yields prediction using stepwise support vector machine, *In Assembly and Manufacturing, ISAM 2009, IEEE International Symposium*, 130-136.

Baek, D. H. and Han, C. H. (2003), Application of data mining for improving and predicting yield in wafer fabrication system, *KIISS*, **9**(1), 157-177.

Breiman, L. (2001), Random forests, *Machine learning*, **45**(1), 5-32.

Chandola, V., Banerjee, A., and Kumar, V. (2009), Anomaly detection : A survey, *ACM computing surveys (CSUR)*, **41**(3), 15.

Chang, Y. J., Kang, Y., Hsu, C. L., Chang, C. T., and Chan, T. Y. (2006), Virtual metrology technique for semiconductor manufacturing, In Neural Networks, *IJCNN 2006, International Joint Conference* on, IEEE, 5289-5293.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002), SMOTE : Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, **16**, 321-357.

Chen, C., Liaw, A., and Breiman, L. (2004), Using random forest to learn imbalanced data, *University of California, Berkeley*.

Chen, P., Wu, S., Lin, J., Ko, F., Lo, H., Wang, J., and Liang, M. (2005), Virtual metrology : a solution for wafer to wafer advanced process control, *In Semiconductor Manufacturing, ISSM 2005, IEEE International Symposium*, 155-157.

Chen, Y.-T., Yang, H.-C., and Cheng, F.-T. (2006), Multivariate Simulation Assessment for Virtual Metrology, *Proc. IEEE Int. Conf. on Robotics and Automation(ICRA 2006)*, 1048-1053.

Chien, C. F., Wang, W. C., and Cheng, J. C. (2007), Data mining for yield enhancement in semiconductor manufacturing and an empirical study, *Expert Systems with Applications*, **33**(1), 192-198.

Cunningham, S. P., Spanos, C. J., and Voros, K. (1995), Semiconductor yield improvement: results and best practices, *Semiconductor Manufacturing, IEEE Transactions on*, **8**(2), 103-109.

Ferreira, A., Roussy, A., and Condé, L. (2009), Virtual metrology models for predicting physical measurement in semiconductor manufacturing, *In Advanced Semiconductor Manufacturing Conference, ASMC 2009, IEEE/SEMI*, 149-154.

Hosmer, D. and Lemeshow, W. (1989), Applied Logistic Regression, *Ed. John Wolfley and Sons*, 8-20.

Kang, P. and Cho, S. (2006), EUS SVM : Ensemble of under-sampled SVMs for data imbalance problems, *Proceedings in Korean Industrial Engineering Conference*, 291-298.

Kang, P. and Cho, S. (2006), EUS SVMs : Ensemble of under-sampled SVMs for data imbalance problems, *In Neural Information Processing, Springer Berlin Heidelberg*, 837-846.

Kang, P., Kim, D., Lee, H. J., Doh, S., and Cho, S. (2011), Virtual metrology for run-to-run control in semiconductor manufacturing, *Expert Systems with Applications*, **38**(3), 2508-2522.

Kang, P., Kim, D., Lee, S. K., Doh, S., and Cho, S. (2012), Estimating the reliability of virtual metrology predictions in semiconductor manufacturing : A novelty detection-based approach, *Journal of the Korean Institue of Industrial Engineers*, **38**(1), 46-56.

Kang, P., Lee, H. J., Cho, S., Kim, D., Park, J., Park, C. K., and Doh, S. (2009), A virtual metrology system for semiconductor manufacturing, *Expert Systems with Applications*, **36**(10), 12554-12561.

Khan, A. A., Moyne, J. R., and Tilbury, D. M. (2008), Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares, *Journal of Process Control*, **18**(10), 961-974.

Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, *In IJCAI*, **14**(2), 1137-1145.

Kubat, M., Holte, R. C., and Matwin, S. (1998), Machine learning for the detection of oil spills in satellite radar images, *Machine learning*, **30**(2/3), 195-215.

Kumar, N., Kennedy, K., Gildersleeve, K., Abelson, R., Mastrangelo, C. M., and Montgomery, D. C. (2006), A review of yield modelling techniques for semiconductor manufacturing, *International Journal of Production Research*, **44**(23), 5019-5036.

Li, T. S., Huang, C. L., and Wu, Z. Y. (2006), Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system, *Journal of Intelligent Manufacturing*, **17**(3), 355-361.

Liaw, A. and Wiener, M. (2002), Classification and Regression by randomForest, *R news*, **2**(3), 18-22.

Lynn, S., Ringwood, J., and MacGearailt, N. (2012), Global and local virtual metrology models for a plasma etch process, *Semiconductor Manufacturing, IEEE Transactions*, **25**(1), 94-103.

Murphy, B. T. (1964), Cost-size optima of monolithic integrated circuits, *Proceedings of the IEEE*, **52**(12), 1537-1545.

Park, K. S., Jun, C. H., and Kim, S. Y. (1997), The comparison and use of yield models in semiconductor manufacturing, *IE interfaces*, **10**(1), 79-93.

Quinlan, J. R. (1986), Induction of decision trees, *Machine learning*, **1**(1), 81-106.

Rosenblatt, F. (1958), The perceptron : a probabilistic model for information storage and organization in the brain, *Psychological review*, **65**(6), 386.

Shin, C. K. and Park, S. C. (2000), A machine learning approach to yield management in semiconductor manufacturing, *International Journal of Production Research*, **38**(17), 4261-4271.

Uzsoy, R., Lee, C. Y., and Martin-Vega, L. A. (1992), A review of production planning and scheduling models in the semiconductor industry part I : system characteristics, performance evaluation and production planning, *IIE Transactions*, **24**(4), 47-60.

Vapnik, V. (2000), The Nature of Statistical Learning Theory, S*pringer*.