

SSVM(Stepwise-Support Vector Machine)을 이용한

반도체 수율 예측

안대웅¹ · 고효현² · 김지현³ · 백준걸^{2*} · 김성식²

¹(주)하이닉스반도체 / ²고려대학교 정보경영공학부 / ³광운대학교 경영대학

A Yields Prediction in the Semiconductor Manufacturing Process Using Stepwise Support Vector Machine

Daewoong An¹ · Hyo-Heon Ko² · Jihyun Kim³ · Jun-Geol Baek² · Sung-Shick Kim²

¹Hynix Semiconductor

²Division of Information Management Engineering, Korea University

³School of Business Administration, Kwangwoon University

It is crucial to prevent low yields in the semiconductor industry. Since many factors affect variation in yield and they are deeply related, preventing low yield is difficult. There have been substantial researches in the field of yield prediction. Many researchers had used the statistical methods. Many studies have shown that artificial neural network (ANN) achieved better performance than traditional statistical methods. However, despite ANN's superior performance some problems such as over-fitting and poor explanatory power arise. In order to overcome these limitations, a relatively new machine learning technique, support vector machine (SVM), is introduced to classify the yield. SVM is simple enough to be analyzed mathematically, and it leads to high performances in practical applications. This study presents a new efficient classification methodology, Stepwise-SVM (SSVM), for detecting high and low yields. SSVM is step-by-step adjustment of parameters to be precisely the classification for actual high and low yield lot. The objective of this paper is to examine the feasibility of SVM and SSVM in the yield classification. The experimental results show that SVM and SSVM provides a promising alternative to yield classification for the field data.

Keyword: support vector machine, semiconductor yield classification, semiconductor manufacturing process

1. 서론

반도체 산업은 21세기 정보화 사회를 선도하는 첨단 핵심부품 산업이다. 또한 높은 부가가치를 지니고 있어 경쟁이 가장 높은 산업 중 하나이다. 최근 고도정보화 사회 진입과 첨단산업

발전에 따라 전 산업의 핵심요소로 반도체 제품은 종류도 다양해지고, 급속히 수요가 증가하고 있다. 최근 세계 시장의 단일화 추세에 영향 받아 반도체 시장은 치열한 글로벌 경쟁 체제에 놓여있다. 치열한 경쟁 속에서 반도체 제조업체들은 생산성 향상과 고부가가치화를 통해 경쟁력 확보를 위한 기업의

본 과제는 한국소프트웨어진흥원의 SW공학 요소기술 개발과 전문인력 양성사업의 결과물임을 밝힙니다.

*연락처 : 백준걸 교수, 136-701 서울시 성북구 안암동 고려대학교 정보경영공학부, Fax : +82-2-929-5888, E-mail : jungeol@korea.ac.kr
투고일(2009년 06월 26일), 심사일(1차 : 2009년 07월 29일, 2차 : 2009년 08월 06일), 게재확정일(2009년 08월 11일).

전사적 노력을 기울이고 있다(Pieter, 2000).

반도체 산업의 경쟁력을 유지하기 위해 생산주기, 재작업율, 공정변동, 공정재고, 수율 등 공정 및 품질 성능지표들의 관리를 수행하고 있다. 특히 반도체 산업의 성능 지표들 중 투입 수에 대한 양품 비율을 의미하는 수율이 기업 경쟁력의 핵심 평가 지표로 사용되고 있다. 신제품의 경우 신속한 양산 목표수율의 확보는 시장 선점과 원가 절감을 통한 가격 경쟁력 확보에 필수적인 성공요건이다. 따라서 반도체 산업에서 경쟁력 창출을 넘어 기업의 생존을 위해서라도 정확한 수율관리는 필수적이다(Kim *et al.*, 1998). 그러나 반도체 제조의 수율관리는 수백 개의 제조공정을 수개월에 걸쳐 진행하기 때문에 복잡하고 어렵다. 또한 반도체 제조공정이 초미세 나노 공정으로 구성되고, 제품이 초소형화 됨에 따라 관리에 많은 어려움을 겪고 있다. 특히 각 공정에서 발생된 작은 영향이 누적되어 최종 제품에 불량으로 나타나기 때문에 관리가 더욱 어려운 상황이다.

반도체 제조 기업들은 안정된 수율 관리를 위해 초정정 환경과 정기/비정기적으로 장비 유지보수 및 작업자 교육을 실시한다. 또한 설비·공정·작업 표준 제정 및 준수, 통계적 공정관리, 수율 분석용 시스템 운영 등을 통해 정확한 수율 관리 노력을 수행하고 있다. 하지만 대부분의 수율 관리 방법은 단변량 관리도를 통해 해당 공정의 공정변수 산포만을 분석하는 간접 관리로 진행되고 있다. 따라서 개별공정에서 관리하는 공정변수의 이상 유무만을 판단하면, 여러 단계의 공정을 거치면서 관련 있는 누적된 오차들은 파악할 수 없게 된다.

목표 수율 확보를 위해 반도체 제조 기업은 최종 수율관리를 수행한다. 또한 최종수율 관리를 위해 Lot 히스토리 검사 등에 많은 노력을 기울이고 있다. 특히 올바른 최종 수율 관리는 높은 고수율 확보 및 저수율 발생 예방과 무검사 활동, 수율개선, 검사비용 감소 등의 효과를 얻을 수 있다. 따라서 최종 수율관리의 효과적이고 올바른 문제 해결을 위해 본 연구는 반도체 최종검사 공정에서 수율 예측을 통한 고수율(High-yield)과 저수율(Low-yield)을 분류하는 효율적인 방법을 제시한다.

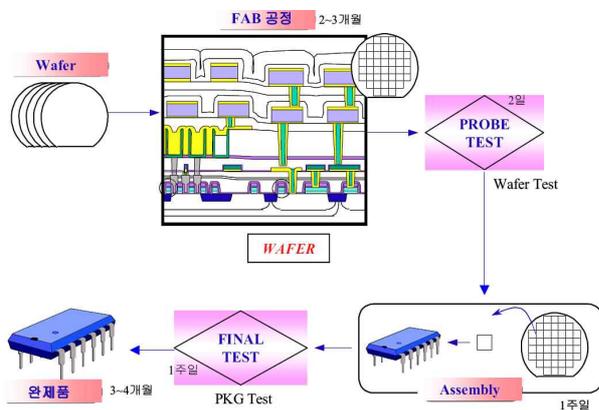


Figure 1. Flow Chart of Semiconductor Manufacturing

반도체 제조공정은 300~400여 개의 단위공정으로 구성되

며 완제품의 칩이 되기까지 일반적으로 3~4개월이 걸린다. 반도체 공정은 <Figure 1>과 같이 FAB(Wafer Fabrication) 공정, Probe 검사(Probe Test) 공정, 조립(Assembly) 공정과 최종검사(Final Test) 공정 단계로 구분할 수 있다(Uzsoy *et al.*, 1992).

FAB 공정은 웨이퍼(wafer) 표면에 여러 종류의 막을 형성하여 하나의 웨이퍼에 수백 개의 칩(chip)을 구성하는 공정이다. Probe 검사 공정은 웨이퍼내의 칩이 정상적인 전기적 특성을 가지는지 검사하여 양/불량 칩을 판별하는 공정이다. 조립공정(Assembly)은 만들어진 웨이퍼 내의 칩을 분리하여 칩의 전기적, 물리적 특성을 향상시키고 외부의 기계적, 물리적 충격으로부터 칩을 보호하기 위하여 형성화시켜 준다. 최종 검사공정(Final Test)은 조립공정을 거쳐 완성된 칩이 제대로 동작하는지 칩의 전기적 특성, 기능 및 신뢰성을 검사하여 양/불량을 판별하는 공정이다. 이상의 공정에서 Probe 검사는 FAB 공정의 문제를 파악하고 불량 원인을 찾아내어 후속공정 수율저하를 예방하는 조치를 수행한다. 특히 최종검사 수율의 정확한 예측을 통해 수율 저하를 유발시키는 Lot을 분류하고 효과적인 사전조치를 수행한다.

반도체 수율은 제조단계에 따라 FAB 수율, Probe 수율, 조립 수율(Assembly 수율), 최종검사 수율(Final Test 수율, 이하 최종 수율)로 구분한다. <Figure 2>는 반도체 제조공정에서 사용하는 각각의 수율 계산 방법을 보여준다. FAB 수율은 wafer in(투입된 웨이퍼장수) 대비 wafer out(중간에 깨지는 등의 불량이 발생하지 않고 제대로 나온 웨이퍼 장수)의 비율이다. Probe 검사

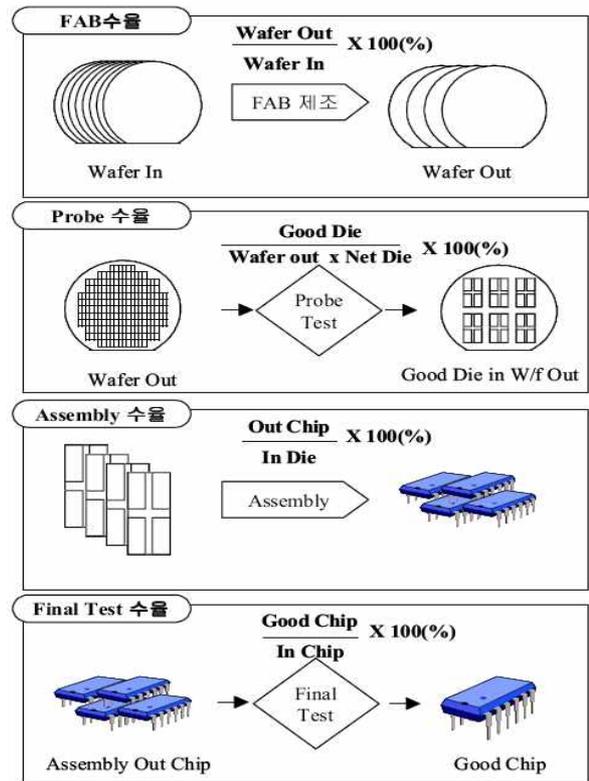


Figure 2. Semiconductor yield calculation(Baek and Nam, 2002)

수율은 wafer out된 총 칩수(wafer out × net die) 대비 Probe 검사 결과 양품인 칩의 비율이다. 넷다이(net die)는 한 웨이퍼에 만들어지는 칩의 총수를 의미하며, 제품에 따라 다르지만 보통 수백~수천 개의 칩 수량이다. 조립 수율과 최종수율은 각각 조립 작업과 최종 검사 후 투입된 칩 중 양품 칩의 비율이다.

네 종류의 수율 중에서 FAB 수율과 조립수율은 통상 거의 100%에 가깝다. FAB 수율은 우수한 설비 도입과 개선 등을 통해 불량률의 원인을 사전에 제거하고 산출 되어 높은 수율을 지닌다. 또한 FAB 수율은 칩이 아닌 웨이퍼가 불량인지를 검사하기 때문에 웨이퍼 내의 칩 불량은 고려되지 않는다. 조립수율은 칩을 조립하는 과정상태의 불량 여부만을 판단한다. 또한 조립수율은 양품만을 대상으로 가공을 진행한다. 최근 설비능력의 향상으로 가공 상의 불량은 거의 발생하지 않아 조립수율은 FAB수율과 동일하게 높은 수율을 가진다. Probe 수율은 실제 칩 하나하나를 개별 검사하므로 Lot 불량을 판단하는 FAB수율에 비해 낮은 수율을 가진다. 또한 Probe 검사는 제품에 영향을 줄 수 있는 원인들에 대해 검사를 수행한다. Probe 검사는 미세한 이상에 대해서도 정확한 불량검출을 수행하여 낮은 수율을 가지게 된다. 최종검사는 제품에 영향을 줄 수 있는 모든 검사를 수행한다. Probe검사를 통해 양품으로 진행되더라도 최종검사를 거치면 불량으로 판정되는 칩이 발생한다. 불량 칩은 공정을 진행하면서 생성되는 진행성 결함에 의해 발생한다. 또한 Probe 검사가 비용 및 시간적인 제약으로 모든 검사를 수행할 수 없어서 발생되기도 한다. 최종 수율이 낮을 경우 조립공정과 최종검사까지 진행되는 동안 소요되는 시간과 비용의 손실이 증가하게 되어 생산비용 및 원가의 상승을 초래한다. 또한 최근 메모리칩이 대용량화됨에 따라 최종검사 시간이 증가하고 있다. 특히, 특수기능 제품의 고정밀도 개별 검사 요구로 인해 최종검사비용 및 시간은 기하급수적으로 상승한다. 따라서 불량으로 판정될 칩이 불필요한 최종검사를 진행하여 시간 낭비와 비용 상승을 초래하지 않게 효율적인 최종수율 관리가 필요하다(Kim and Bac, 1995).

최종수율의 향상을 위해 Probe 검사 단계는 Lot별 최종수율 예측을 수행하여 필요한 조치를 수행한다. 하지만 현재 예측 방법은 단변량 회귀분석에 의한 단순 예측을 수행하여 고/저 수율의 오분류를 빈번하게 발생시키고 있다. 또한 오분류에 따른 불필요한 조치로 오히려 최종수율에 악영향을 주고 있다. 따라서 본 연구는 최종수율 향상을 위해 효과적인 예측과 분류를 통해 검사비용 절감, 생산성 향상, 고객 납기 준수 및 서비스를 향상시키는 효율적인 수율 분류 방법을 제안한다.

수율 예측과 관련하여 대부분의 기존 연구들은 FAB공정의 데이터를 이용한 Probe 검사 수율을 예측하였다. Ciciani and Jazeolla(1991)는 웨이퍼 내의 결점(defect)에 통계적인 분포를 적용하여 웨이퍼 단위의 칩 수율을 수리적으로 예측하는 모델을 개발한 단일변량분석을 수행하였다. 하지만 단일변량분석연구는 기본적으로 한 변수에 의해 수율예측을 하는 경우 다른 변수의 영향을 고려하지 못한다는 한계가 있다.

Croiser(1988)는 단일변량분석의 한계를 극복하기 위해 다수의 변수를 하나의 모형에 통합하는 다변량분석 연구를 진행하였다. 그러나 다변량분석의 통계적 기법은 수많은 공정 변수와 엄청난 양의 데이터로 인해 모든 변수를 분석하기에는 너무 많은 시간이 소요된다. 따라서 실제 공정에서는 일부 변수만을 선별하여 분석하고 있다.

다변량분석을 개선한 방법으로 다변량판별분석이 있다. 판별 분석은 변수들의 분포가 다변량정규분포이고, 각 집단의 분산 및 공분산 구조가 동일해야 사용가능하다. 하지만 실제 현장의 데이터는 집단의 구조가 서로 상이하어 판별 분석의 기본 가정을 충족시키지 못해 널리 사용되지 못하고 있다.

Odom and Sharda(1990)는 인공지능망을 예측분야에 적용하였다. Odom and Sharda(1990)는 연구에서 판별 분석과 인공지능망모형을 적용하여 예측률을 비교하였으며, 인공지능망 모형이 판별 분석에 비해 우수한 결과를 보였다. Tam and Kiang(1992)도 인공지능망을 이용한 예측을 수행하였다. 또한 예측 결과를 판별분석, Logit, k-최근접 이웃방법(k-nearest neighbor), 귀납적 추론(ID3)의 결과와 비교하였다. 연구 결과에서 인공지능망에 의한 모형이 예측성과 적응력 등에서 다른 방법에 비해 우수한 결과를 보였다. Kang *et al.*(1998)은 귀납적 의사결정트리, 후방전파(back propagation)를 사용한 신경망, SOM(Self-Organized Map)을 예측에 이용하였다.

이 외에도 많은 연구들이 인공지능망을 이용하여 우수한 예측정확성을 보였다. 그러나 인공지능망의 결과는 설명력이 부족하여 예측결과의 원인을 설명하기 어렵다는 한계가 있다. 또한 기존의 연구가 주로 소량의 제한적 자료를 대상으로 하여 실제적인 문제에 일반화 시키는데 한계점을 가지고 있다. 특히 인공 신경망 모형은 모델 생성에 있어 과도적합(Over fit)의 문제와 인공지능망의 구조를 설계하는데 많은 시간과 노력이 필요하다는 단점도 있다.

본 연구는 이상의 단점을 개선하기 위한 방법으로 최근 우수한 분류 알고리즘으로 알려진 SVM을 수율예측에 사용한다. SVM은 1995년 통계학자인 Vapnik에 의해 개발된 학습기법으로, 입력공간과 관련된 비선형문제를 고차원 특징공간의 선형 문제로 대응시켜 나타내기 때문에 수학적으로 분석하는 것이 수월하다(Hearst *et al.*, 1998). 또한 SVM은 조정해야 할 파라미터의 수가 많지 않아 비교적 간단하게 학습에 영향을 미치는 요소들을 규명할 수 있다. 최근 몇 년간 SVM을 사용한 다양한 연구가 진행되었다. 특히 SVM은 문서분류, 영상인식, 문자인식 등에서 뛰어난 일반화 성능을 보여주었다(Joachims, 1998; Osuna *et al.*, 1997). Meyer *et al.*(2003)은 16개의 분류 알고리즘과 SVM의 성능을 비교한 결과 SVM이 소수의 경우를 제외하고 다른 알고리즘에 비해 동등하거나 우수하다는 결론을 제시하였다. 그러나 기존의 연구들은 대부분 Probe 수율 예측에 한정되어 있고 최종수율 예측에 대한 연구는 찾아볼 수 없었다. 따라서 본 연구는 최종 수율 관리의 효과적이고 올바른 문제 해결을 위해 최종검사 공정의 정확한 수율 예측에 SVM을 이용한 효과적인

방법을 제시한다. 제안된 방법은 정확한 Lot단위의 수율 관리를 위해 Lot별 Process Control Monitor(이하 PCM) 데이터와 Probe 검사 이력에 근거하여 정확한 최종수율을 예측하고 고수율과 저수율 Lot을 효율적으로 구분한다.

2. 본론

2.1 문제 개요

본 연구에서 다루는 공정은 FAB OUT 이후 Probe 검사를 실시할 때 최종수율을 고려하지 않고 모든 Lot에 대해 동일한 조건으로 검사를 실시한다. 최종수율이 좋지 않은 경우 1~2개월의 시간을 소요하여 검사된 제품들을 재작업 하여 고수율 여부를 확인한다.

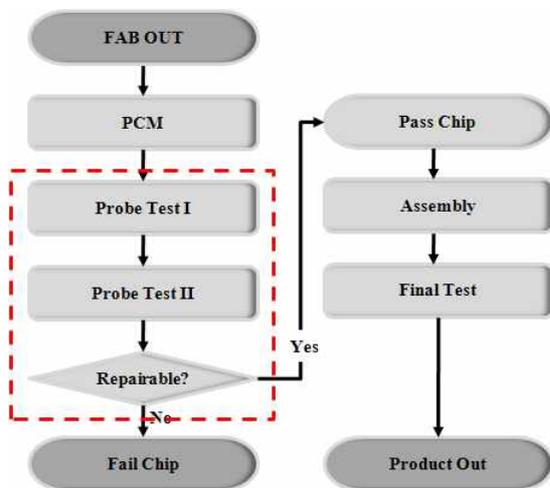


Figure 3. Flow Chart of Current Semiconductor Test Process

<Figure 3>은 현재 FAB OUT 이후 최종검사까지의 필요한 공정 절차를 나타내고 있다. 현재 반도체 검사공정은 FAB OUT 이후 가장 먼저 단위소자의 특성을 측정하는 PCM 공정을 진행한다. PCM 공정은 단위소자의 전압, 전류 및 기타 전기적 특성을 측정하여 값을 얻는다. 또한 PCM 공정은 하나의 Lot중에서 통상 세장의 웨이퍼를 샘플링 측정된 평균을 Lot을 대표하는 측정값으로 한다. PCM 공정을 거친 Lot은 두 단계로 구성된 Probe 검사 공정을 진행한다. Probe 검사는 1단계 사전 정보를 통한 기본 검사 후에, 2단계 심화 검사를 수행한다. 두 단계의 Probe 검사 공정은 FAB 공정에서 발생한 이상을 검출하고 전기적 검사를 통해 구제(Repair)할 수 있는 칩을 선별한다. 특히, 칩 선별을 통해 반도체가 패키지(Package) 형태로 가공되기 전에 재가공을 통한 수율 향상을 지원하게 된다. 또한 각 웨이퍼의 수율을 측정하고, 불량 발생 칩을 분류하여 DC 불량, Function불량에 대한 불량률 데이터를 웨이퍼별로 수집한다. 모든 Probe 검사가 완료되면 각 Lot의 웨이퍼 상 칩들은 개별 칩으로

분리되어 조립 공정을 통해 완성된 제품형태를 갖는다. 완성된 칩들은 최종검사 공정에서 개별 패키지의 전기적 특성 및 외관특성을 검사하여 최종 양품으로 선별된다.

이상의 공정 절차에서 모든 Lot은 최종수율을 고려하지 않고 동일한 정보를 가지고 두 단계의 Probe 검사를 진행하고 있다. 또한 Lot의 제품들이 최종검사에서 불량으로 판명될 가능성이 있는 경우도 아무런 대응 조치 없이 Probe 검사를 진행하고 있다. 무분별한 Probe 검사가 진행되면 최종수율 손실이 발생하여 다양한 비용 손실이 연쇄적으로 발생된다. 또한 고수율과 저수율 Lot이 섞일 경우 목표량 산출을 위해 많은 검사를 수행하게 된다. 최종 검사량이 많을수록 장비 사용으로 인한 부하 증가와 시간 손실이 발생한다. 현재의 방식으로는 반도체 칩에 불량이 발생하더라도 그 칩이 모든 공정을 마친 뒤 불량을 확인할 수 있고, 문제를 개선하기 전까지는 불량이 계속 발생된다. 개선 전까지 대응 조치가 이루어 지지 않는다면 수율 손실이 발생하거나 고객이 원하는 수준의 제품을 필요한 양만큼 생산하지 못하게 된다. 그러나 고수율로 예상되는 Lot은 우선적으로 Probe 검사를 실시하면 수율 손실이 적게 발생하여 목표량 달성을 위해 최종검사에서 시간과 비용을 절약하게 된다. 반면 저수율로 예상되는 Lot은 현재의 검사조건으로 Probe 검사를 실시하게 되면 최종검사의 수율 손실이 많이 발생하게 된다. 따라서 Probe 검사 단계에서 강화된 조건의 검사를 수행하여 최종수율을 개선해야 한다. 이상의 방법으로 저수율로 예상되는 Lot과 고수율로 예상되는 Lot을 구분하여 Probe 검사 조건을 이원화한다면 최종수율을 향상할 수 있다. 이원화된 Probe 검사를 통해 불량을 확인하고 다시 고수율을 확보할 수 있는 1~2개월의 시간을 줄이게 된다면, 불필요한 시간 낭비와 불량을 예방하여 수율 개선 및 제조비용의 절감을 가져오게 된다. 또한 고수율 Lot을 정확히 선별하여 최종검사를 진행하게 되면 고객의 요구를 맞출 수 있는 생산량을 적기에 납품할 수 있으며, 고객에게 필요한 양만큼의 최종검사를 진행함으로써 전체적인 검사비용을 줄일 수 있다. 반면 저수율로 예상되는 Lot은 Probe 검사단계에서 취할 수 있는 효과적인 대응 조치를 통해 수율을 개선할 수 있다.

- 예측 정확도(Classification Accuracy) :

$$CA[\%] = \frac{cc}{cc+uc} \times 100 \tag{1}$$

cc : correctly classified

uc : uncorrectly classified

본 연구의 성능을 비교하는 척도로는 예측정확도와 오분류율을 사용한다. 예측정확도는 식 (1)과 같으며, 전체 중에서 정확히 분류한 비율을 의미한다. 오분류율은 식 (2)와 같으며, 고수율로 예측한 것 중에서 실제저수율이 발생한 경우와 저수율로 예측한 것 중에서 실제 고수율이 발생한 경우가 전체 중에서 차지하는 비율이다. 그러나 본 연구는 고수율로 예측된 것

이 실제 저수율로 발생한 경우패널티를 주어 예측 성능뿐만 아니라 실제 현장에서의 위험 요인까지 반영해야 한다.

- 오분류율(Misclassification Rate) :

$$MR[\%] = \left(\frac{2}{3} \frac{rl}{ph} \times \frac{1}{3} \frac{rh}{pl} \right) \times 100 \quad (2)$$

rl : real low yield(실제저수율)
 rh : real high yield(실제고수율)
 pl : predict low yield(저수율예측)
 ph : predict high yield(고수율예측)

따라서 본 연구는 반도체 검사 공정의 분류 정확도 향상과 오분류의 감소를 위해 각 경우에 대해 서로 다른 가중치를 주고 오분류는 최소화하고 정확도는 최대화하는 분류 알고리즘을 제안한다. 각 경우에 대한 가중치는 <Table 1>과 같다.

Table 1. Weight of Each Case

경우	예측	고수율	고수율	저수율	저수율
	실제	고수율	저수율	고수율	저수율
가중치		1(w ₁)	3(w ₂)	0.5(w ₃)	2(w ₄)

정확도에 따른 가중치는 실제 저수율을 정확히 예측하는 것이 중요하므로 2로 부여하고, 실제 저수율을 고수율로 예측한 경우 패널티를 3으로 크게 하였다.

이상의 가중치를 적용한 본 연구의 목적함수는 식 (3)과 같다. 식 (3)의 분자는 정확하게 분류하는 경우이며, 분모는 오분류한 경우이다. 각 경우의 결과에서 오분류가 줄고 정확한 분류가 진행 된다면 목적 함수는 최대가 된다.

- 목적함수 :

$$Maximize \frac{w_1 \times (rh/ph) + w_4 \times (rl/pl)}{w_2 \times (rl/ph) + w_3 \times (rh/pl)} \quad (3)$$

2.2 분류기를 이용한 반도체 검사 공정 제안

본 연구는 반도체 검사공정에서 최종수율을 예측하는 분류기를 통해 예측된 Lot들을 이원화된 Probe 검사를 실시하여 최종수율을 향상시키는 방법을 제안한다. 현재 FAB OUT 이후의 검사 공정은 PCM과 Probe 검사를 모든 Lot에 동일하게 적용하고 있다. 또한 최종검사에서 저수율이 발생할 가능성이 있는 경우에도 아무런 대응 조치 없이 Probe 검사를 진행한다. 따라서 최종수율 손실이 발생하고 불량 칩으로 인한 시간과 비용에서 많은 손실을 야기하고 있다. 그러나 본 연구는 저수율 발생 가능 Lot을 PCM과 Probe 검사 1단계 종료 후 생성된 데이터를 기반으로 고/저수율 분류기를 사용하여 분류한다. 고/저수율 분류기는 생성된 데이터를 기반으로 모델을 만들고 모델 기반의 고/저수율 예측을 통해 이원화된 Probe 검사 2단계를 수

행한다. 본 연구에서 제안하는 검사공정은 <Figure 4>와 같다.

<Figure 4>에서 FAB Out된 Lot은 1단계 Probe 검사 후 제안된 분류기를 통해 고/저수율로 분류된다. Lot이 고수율로 분류되면 일반적인 2단계 Probe 검사를 진행한다. 그러나 Lot이 저수율로 분류되면 적절한 조치를 취해 강화된 조건의 2단계 Probe 검사를 수행한다. 이원화된 2단계 Probe 검사를 통해 최종검사공정의 잠재 불량을 제거함으로써 최종수율은 향상된다. 그러나 잘못된 분류로 실제 고수율 Lot이 강화된 Probe 검사를 수행한다면, 잘못된 조치로 인해 오히려 불량칩을(Over-kill) 만들게 되고 전체적인 수율 손실은 더 크게 발생하게 된다. 따라서 분류기는 올바른 예측을 통해 정확한 고/저수율 Lot을 구분해야 한다. 따라서 본 연구에서 분류기는 일반 SVM 모델을 개선하여 순차적으로 미세한 분류까지 가능하도록 Stepwise SVM(SSVM) 모형으로 구성한다. SSVM 모형은 SVM을 적용할 때 단계적으로 파라미터를 조정하여 수율 예측의 오류를 최소화하기 위한 방법이다. 따라서 제안된 SSVM 모형은 일반 SVM 모형보다 분류가 어려운 경계 구간의 일정 범위를 단계적으로 확대하여 세밀하고 정확하게 분류한다.

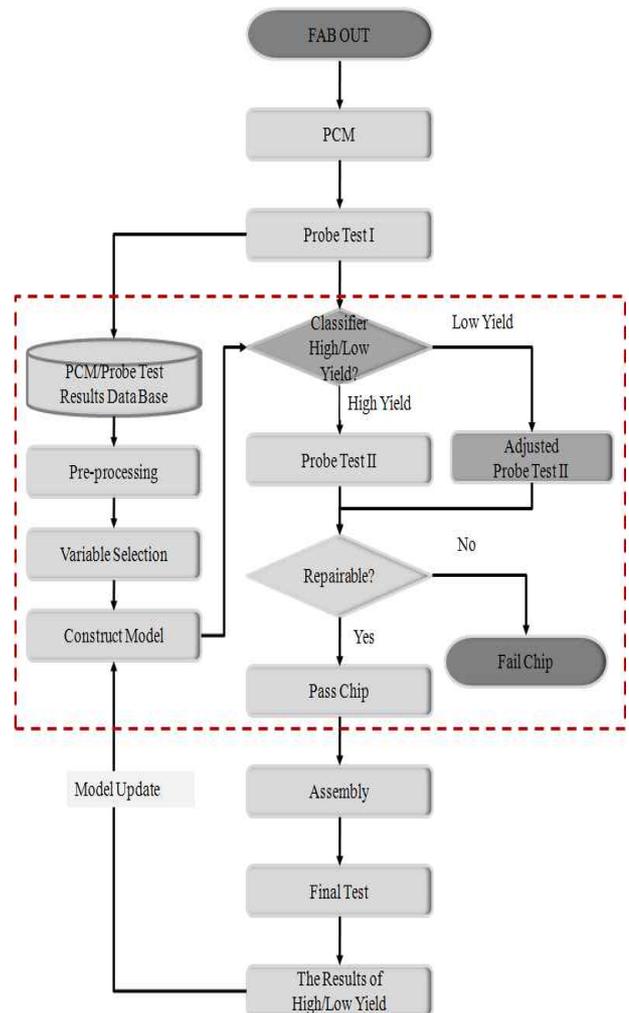


Figure 4. Flow Chart of Semiconductor Test Process using Classifier

일반적으로 SVM은 입력벡터를 고차원의 특징공간으로 사상시킨 후 두 클래스(class) 사이의 마진(margin)을 최대화시키는 분리경계면(hyperplane)을 찾는 것이 목적이다. 선형분리가 가능한 경우에는 분리경계면을 선형식으로 나타낼 수 있으나 비선형의 경우에는 커널함수를 이용하여 선형모형으로 근사화할 수 있다. 사용할 수 있는 커널함수는 여러 가지가 있으며, 일반적으로 많이 사용되는 커널함수로는 가우시안 RBF와 다항식커널을 들 수 있다(Vapnik, 1995).

가우시안 RBF :

$$K(x, y) = \exp\left(-\frac{1}{\sigma^2} \|x - y\|^2\right)$$

다항식 커널 :

$$K(x, y) = (xy + 1)^d$$

여기서 d 는 다항식 차수이고, σ^2 은 가우시안 RBF 커널의 대역폭이다. 분리 가능한 문제에 있어서 계수의 하한은 0이다. 하지만 분리가 불가능한 문제에서 SVM은 계수의 하한 이외에 상한 C 를 추가함으로써 일반화된 결과를 얻을 수 있다.

SSVM 모형은 <Figure 5>와 같이 구성된다. SSVM 모형에서 각 SVM 모형은 단계에 따라 정확한 분류가 수행될 수 있도록 독립적으로 파라미터를 선정하여 모형을 생성한다. SVM 모형의 구축에 있어서 커널함수의 선택은 가장 중요한 문제 중의 하나이다. 본 연구에서는 SVM의 커널함수로서 가장 널리 사용되며 오류에 강건(robust)한 가우시안 RBF(Radial Basis Function)를 사용한다. SVM의 성능에 있어서 커널함수의 상한 C 와 커널 파라미터 σ^2 가 중요한 역할을 한다. 따라서 실제 저수율을 고수율로 예측하는 오분류를 최소화하는 모형을 생성하기 위해 적절한 상한 C 와 커널 파라미터 σ^2 를 단계적으로 선정한다. 단계적으로 파라미터를 선정하는 방법은 우선 고수율을 가장 잘

분류하는 첫 번째 상한 C_1 와 커널 파라미터 σ_1^2 를 선정한다. 첫 번째 선정된 파라미터에 의해 고수율로 분류된 데이터는 제거하고 남아있는 데이터 중에서 고수율을 가장 잘 분류하는 두 번째 상한 C_2 와 커널 파라미터 σ_2^2 를 선정한다. 동일한 방법으로 n 번째 파라미터는 $n-1$ 단계에서 선정된 파라미터에 의해 고수율로 분류된 데이터는 제거하고 남아있는 데이터 중에서 고수율을 가장 잘 분류하는 n 번째 상한 C_n 와 커널 파라미터 σ_n^2 를 선정한다. 단계적으로 파라미터를 선정하므로 SSVM은 분류가 어려운 구간을 확대해가면서 세밀하고 정확한 분류를 가능하도록 한다. 고수율 검출을 위한 SSVM모형은 각 영역별로 규정된 고수율을 설정하여 학습시켜 모형을 생성한다. 또한 단계별 학습은 해당 구간에서 모형을 생성하면 모형에 의해 검출된 학습 데이터는 제외한다. 이상의 단계를 통해 생성된 모형은 실제 공정에서 신규 Lot이 투입되면 단계별 분류기에서 PCM, Probe 검사 결과를 이용하여 최종수율을 분류에 사용된다. SSVM 모형 구축과 각 모형에서 사용되는 수식은 다음과 같다.

$$\{x_i, y_i\}, y_i \in \{-1, 1\}, i = 1, \dots, N \quad (5)$$

$$\phi(x; \theta) \in \{-1, 1\} \quad (6)$$

$$\theta_n = \{\text{upperbound } C_n, \text{kernel parameter } \sigma_n^2\}, \quad (7)$$

$$n = 1, \dots, k \quad (8)$$

$$F_n(x) = \phi(x; \theta) \quad (8)$$

$$\theta_n = \arg \min_{\theta} \sum_{i=1}^N \exp(-y_i F_n(x_i)) \quad (9)$$

$$F_{n+1}(x) = \phi\left(x - \sum_{i=1}^N x_i; \theta\right), \forall i \in ph \quad (10)$$

$$\theta_{n+1} = \arg \min_{\theta} \sum_{i=1}^N \exp(-y_i F_{n+1}(x_i)), \forall i \notin ph \quad (11)$$

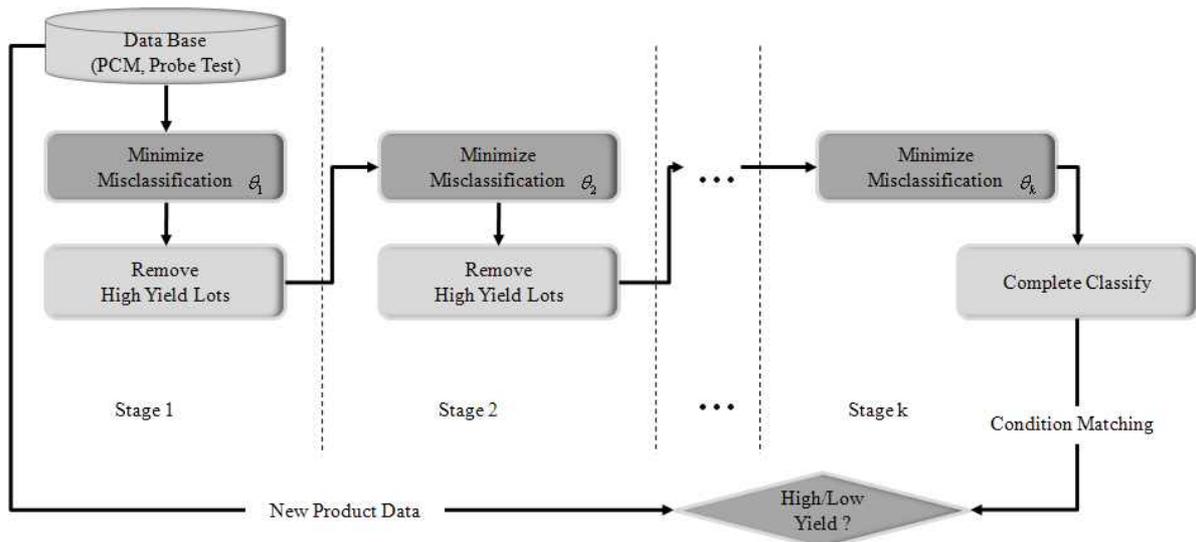


Figure 5. Classifier Design based on SSVM

- i : 데이터 인덱스
- N : 데이터의 총 개수
- n : 분류기의 단계별 인덱스
- k : 단계의 총 수
- x_i : Lot별 PCM 및 Probe 검사공정 결과 데이터(입력 변수)
- y_i : 최종 검사공정 결과
- $\phi(x; \theta)$: SVM 분류 결과
- θ : 모형파라미터 조합
- $F_n(x)$: SVM 분류 결과

본 연구는 식 (5)에서 Lot별 PCM 및 Probe 검사공정 결과 데이터를 x_i 로 정의하고, 최종 검사공정 결과를 y_i 로 정의한다. 최종 검사공정 결과는 고수율과 저수율로 존재하며 결과 값은 항상 1과 -1이다. 식 (6)에서 SVM 분류기에 의해 분류된 결과는 $\phi(x; \theta)$ 로 정의하며, x 는 입력되는 변수를 나타낸다. 또한 θ 는 SVM 모형을 구축하는데 필요한 파라미터 조합을 나타낸다. $\phi(x; \theta)$ 을 통해 각 Lot은 1과 -1의 분류를 통해 고수율 Lot을 분류한다. 식 (7)은 SSVM모형의 각 분류기에서 필요한 모형파라미터를 정의한다. 식 (8)은 θ_n 의 모형파라미터를 가지고 분류기에 의해 검출된 결과를 $F_n(x)$ 에 저장한다. 식 (9)는 단계별로 정확한 검출과 오분류 최소를 위한 θ_n 을 찾는다. 식 (10)은 이전단계에서 사용한 데이터를 제외하고 분류된 결과를 나타낸다. 식 (11)은 다음 단계의 모형파라미터를 계산한다. 또한 다음 단계의 SVM 모형들은 식 (9)와 식 (10)을 반복하여 계획한 단계만큼 구성하게 된다.

이상의 수식을 통해 본 연구에서 제안한 SSVM 모형 구축과정은 다음과 같다.

- Step 1 : 실험에 필요한 PCM 및 Probe 검사공정 결과 데이터를 수집하고 분류 단계 n 을 설정한다.
- Step 2 : 수집된 데이터 중에 학습에 필요한 데이터와 모형 검증에 필요한 데이터를 분할한다.
- Step 3 : n 을 1로 설정하고 고수율 분류 수준에 따른 오분류를 최소화하는 θ 를 학습데이터를 통해 산출한다.
- Step 4 : 모형 검증용 데이터 집합을 통해 분류 기준에 해당하는 분류가 수행되면 현 단계의 학습을 종료한다.
- Step 5 : 학습된 n 단계의 모형을 통해 학습 데이터 중에 고수율로 분류된 데이터는 학습 데이터에서 제외한다.

- Step 6 : n 단계까지 진행이 완료되면 Step 5로, Otherwise n 을 1만큼 증가시키고 Step 2로 이동
- Step 7 : SSVM 모형에 대해 전체 검증용 데이터를 사용하여 최종 분류를 수행하고 고수율 기준을 통과하면 모형생성을 완료, Otherwise Step 2로 이동

이상의 반복적 학습과 검증을 통해 생성된 SSVM은 공정에 적용되어 신규 생산 Lot이 발생하면 정확하고 체계적으로 분류를 실행한다. 분류 단계는 검증 단계와 같이 첫 번째 모형부터 Lot의 정보를 입력하여 분류한다. 만일 입력된 정보가 규정된 단계 내에서 고수율로 판명될 경우 일반 Probe 검사를 수행한다. 그러나 입력된 정보가 모든 분류기를 거처도 최종적으로 저수율 Lot으로 판명될 경우, 해당 Lot은 강화된 2단계 Probe 검사가 수행되도록 분류된다.

3. 실험계획 및 결과분석

본 연구의 성능을 평가하기 위해 현재 재직 중인 반도체 회사의 본사와 해외 FAB의 데이터를 이용하여 현장 적용 가능성과 성능에 대해 평가하였다. 성능 평가를 위해 실무 담당자들과 협의하여 최종수율 평균이 80%인 제품을 대상으로 하였다. 또한 최종검사 공정의 고/저수율 구분 기준은 70%로 사용하였다. 평균 보다 낮은 70%로 기준을 선정한 것은 현재 생산되는 Lot들의 수율이 가장 혼재하는 구간이고, 70%를 기준으로 비용손실 정도의 차이가 크게 나타난다.

대상 데이터는 2008년 1월부터 2008년 12월까지 수집하였다. 기업의 데이터 보안 관리상 성능 평가에서는 과거에 생산된 제품 Lot중 고수율 Lot 516개와 저수율 Lot 517개, 총 1,033개의 Lot 데이터를 사용하였다. 수집된 Lot의 데이터는 현장에서 사용 중인 PCM의 21개와 Probe 검사의 100개 데이터이다. PCM 데이터는 동일 Lot에 포함된 3장의 웨이퍼를 샘플링 하여 측정하여 평균한 값이다. Probe 검사 데이터는 Lot에 포함된 전수 웨이퍼를 측정한 평균 값이다. PCM 데이터는 각 단위소자의 특성에 대한 변수로서 트랜지스터(Transistor), 저항(Resistor), 콘덴서(Capacitor)등의 전기적 특성을 측정한 것이다. Probe 검사 데이터는 1단계 Probe 검사결과로 1단계 Probe 검사의 수율, DC 불량 종류에 따른 불량률, Function불량 종류에 따른 불량률, 반도체 동작의 기본 특성값, 리던던시(Redundancy) 사용률, 기타 Probe

Table 2. Example of PCM, Probe Test Result Variable

Lot	Yield	E1	E2	E3	E4	E20	E21	P1	P2	P10	P11	P12	P13	P14	P15	P99	P100
3233	88.12915	14.8533	749.786	2438.47	-3.282	0.294402	-0.34922	76.6	81.1	10.3	0.2	11.3	10.4	2.7		6.9	4.5
3280	97.42188	17.7652	937.8	3314.22	-3.231	0.296806	-0.27158	77.3	80	10.5	0.1	11.3	11.5	1.1		6.1	2.8
2358	73.31256	15.0602	802.578	2525.11	-3.193	0.286647	-0.28315	76.1	81.4	11.4	0.1	13	8.4	2.4		4.6	5.3
2810	75.89391	16.1435	527.363	2072.41	-3.230	0.276475	-0.36022	57.6	63.2	15	0.3	20.1	19.5	2.5		4.3	5.7
3422	94.99499	16.3684	760.278	2485.3	-3.278	0.310144	-0.26005	71.7	75.2	12.7	0.1	13.5	14.1	1.3		8.9	3.5
3458	77.03863	17.6724	887.686	3270.05	-3.231	0.303492	-0.27403	55.8	59.2	31.4	0.2	30.9	11.2	1.1		7.2	3.4
3587	65.33421	17.4317	1138.06	2983.05	-3.30	0.305705	-0.24754	74.6	80.9	12.5	0.2	13.3	8.6	0.9		7	6.3
3678	95.11981	16.453	903.451	2785.8	-3	0.30286	-0.26958	80	82.6	11.9	0.2	12.6	7.6	1.1		3.7	2.6
2968	99.75625	14.6333	733.034	2332.52	-3.34	0.293436	-0.26999	42.5	49.8	19.6	0.6	21.3	31.8	3.6		11.5	7.2
2981	99.32288	15.5326	671.426	2271.76	-3.16	0.283952	-0.25322	70.5	78.5	12.9	0.1	16.6	7.7	5.7		7.2	8

검사에서 측정된 값들이다. <Table 2>는 PCM과 Probe 검사 결과를 가지고 SSVM에서 사용될 변수들을 나열하였다.

3.1 비교대안 알고리즘

본 연구는 제안하는 방법의 성능 평가를 위해 현재 현장에서 사용 중인 방법과 기존 예측 방법을 비교대안으로 객관적인 성능 비교를 통해 제안된 방법의 우수성을 입증한다. 비교 대안은 전통적 통계기법이며 현재 공정에서 사용 중인 판별분석(선형 판별분석(Linear Discriminant Analysis, LDA)과 비선형 판별분석(Quadratic Discriminant Analysis, QDA) 및 본 연구에서 참조하여 사용한 일반 SVM이다. 선형 판별분석은 기존 현장에서 사용하는 방법으로 가장 보편적인 실 공정 예측방법이며, 이차판별분석은 선형 판별분석을 개선한 방법이다. 또한 일반 SVM은 기존 예측 방법 중에 가장 좋은 성능을 보여주는 방법이다.

판별분석(Discriminant Analysis)은 관측대상이 속하는 집단을 예측하기 위한 수법이다. 집단의 예측이란, 예를 들면 양품/불량품, 구입자/비구입자 집단 어느 쪽에 속하는가를 예측하는 것이다. 이와 같은 예측을 판별이라 하고, 집단의 수가 둘인 경우를 2군 판별, 집단의 수가 셋 이상의 경우를 다군 판별이라고 한다. 판별분석 방법은 m 개의 집단이 존재하고 수집된 데이터의 소속 집단을 이미 알고 있는 경우에 사용한다. 만일 소속 집단이 불분명한 대상이 발생하면 판별 분석은 수집되어 있는 데이터에 의거해 소속 집단을 예측한다. 판별함수는 관측대상을 집단 간 차이 최대화, 오분류 비율 및 오분류에 따른 비용 최소화 등의 목적에 맞게 수립된다. 또한 판별 분석은 집단 간 공분산이 같으면 선형판별분석(LDA)을 사용하고, 집단 간 공분산이 다르면 비선형 방법인 이차판별분석(QDA)을 사용한다.

3.2 성능평가 절차

본 연구의 성능 평가는 <Figure 6>과 같은 절차로 진행된다.

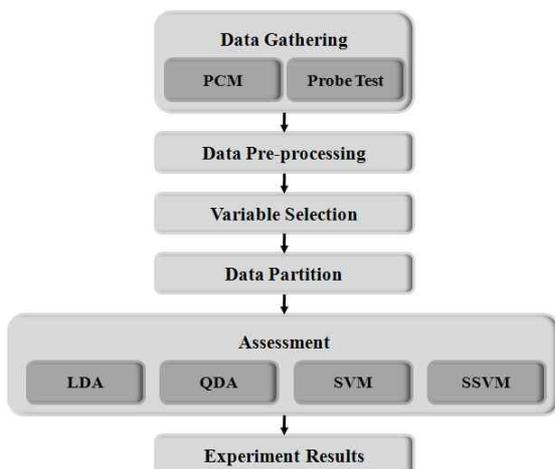


Figure 6. Performance Assessment Procedure

<Figure 6>에서 데이터 수집(Data Gathering)은 예측에 필요한 총 121개의 데이터를 PCM과 Probe 검사 결과를 수집한다. 전처리과정(Data Pre-processing)은 유효한 데이터 생성을 위해 수집된 데이터를 가공한다. 전처리과정은 먼저 이상치 제거를 위하여 각 변수 분포의 양측 1%의 데이터를 제거하고, 결측치를 각 비율의 평균값으로 보정하여 유효한 데이터 셋을 완성한다. 변수선정(Variable Selection)은 단일변량분석의 과정인 t-test를 통하여 고수율 또는 저수율 분류에 유의한 변수($P < 0.01$)들을 가려낸다. 또한 단일변량분석을 통해 선택된 변수들에 대해 다변량분석인 로지스틱 회귀분석 방법의 단계별 변수선정방법(stepwise method with forward selection)을 사용하여 최종적으로 변수($P < 0.05$)를 선정한다. 데이터 분할(Data Partition)은 모델 수립과 평가를 위해 데이터를 분리한다. 전체 데이터 중에서 모형구축용으로 총 데이터의 80%를 사용하고, 나머지 20%는 모형검증용 표본으로 사용한다. 모형 구축용 데이터는 고수율과 저수율을 50:50의 비율로 시간적으로 무관하게 랜덤으로 선정한다. 모형 구축은 5-fold cross validation을 통해 최소 Mean Squared Error(MSE)를 가지는 모형을 대표 모형으로 생성한다. 5-fold cross validation은 전체 5개의 폴더 중에 1개는 모델 구성, 다른 한 개는 모델 적합도 평가, 나머지 3개는 모델 테스트를 수행하고, 각 폴더를 교차 평가하여 총 ${}_5C_3$ 의 조합을 통해 최소 MSE를 가지는 모형을 수립한다. 또한 수립된 모형의 실질적 평가는 모형검증용 표본을 사용한다. 모형검증용 표본은 전체 데이터를 5구간으로 나누어 차례로 사용한다. 따라서 객관적 성능 평가는 5번을 수행한다. 5-fold cross validation은 모델 수립에 따른 일반성 확보와 샘플링에 따른 오류를 줄일 수 있게 한다(Weiss *et al.*, 1991). 평가(Assessment)는 동일한 데이터를 기준으로 LDA, QDA, 일반 SVM, 및 제안된 SSVM에 적용하여 수행된다. 또한 모형검증용 표본을 통해 대안들 간의 비교를 위한 최종 성능평가 결과(Experiment Results)를 얻는다.

3.3 성능 평가 결과 분석

성능평가는 본사 및 해외 FAB 데이터를 사용하고 절차에 따라 실험하여 결과를 비교하였다. 우선 본사 FAB 데이터를 사용한 예측정확도 결과는 <Figure 7>과 같다.

예측정확도는 SSVM이 가장 높았고 일반 SVM, LDA, QDA의 순서로 평가되었고 LDA나 QDA에 비하여 많은 차이를 보이고 있다. 평가 데이터의 Set 4는 모든 대안에서의 결과가 보는 바와 같이 비슷하게 나타난다. 이는 Set 4의 평가 데이터들이 고수율과 저수율의 경계가 명확하여 선형 판별로도 쉽게 분리되는 특징을 가지고 있어 큰 차이를 보이지 않았다. 따라서 평균을 구할 때는 Set4의 결과는 혼재한 데이터들의 일반적인 검출 성능 평가를 위해 제외하였다. 평가 데이터 Set 4를 제외하면 데이터가 혼재한 일반 Set의 경우 SSVM이 예측정확도에서 평균적으로 우수하였다. 반면 기존 LDA, QDA보다 우수하지

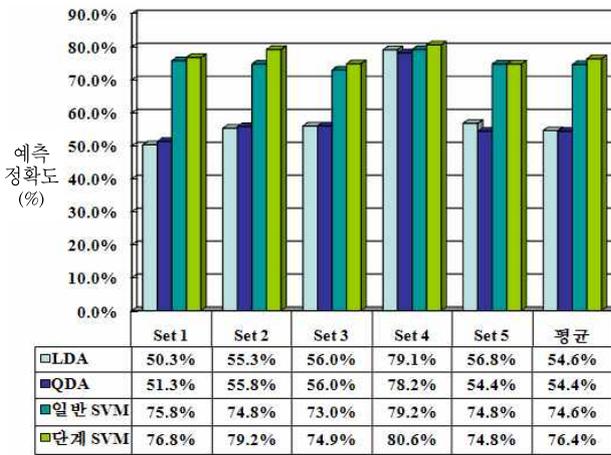


Figure 7. The Results of Each Set(Domestic FAB)

만 일반 SVM은 모형파라미터 설정 한번으로 혼재한 데이터 구간의 분류를 수행하긴 했지만 제안된 SSVM 보다는 낮았다. 그러나 제안된 SSVM은 분류수준을 준수하며 오분류를 최소화하는 모형파라미터 설정을 단계적으로 지시함으로써, 혼재한 데이터 구간을 단계별로 확대하여 세밀하고 정확하게 검출을 수행할 수 있어 성능의 우수함을 보였다.

Table 3. The Results of Domestic FAB

예측	실제	LDA	QDA	일반SVM	SSVM
고수율	고수율	64.3%	64.5%	67.6%	69.0%
고수율	저수율	45.3%	46.3%	18.2%	15.3%
저수율	고수율	35.7%	35.5%	32.4%	31.0%
저수율	저수율	54.7%	53.7%	81.8%	84.7%
예측정확도		54.6%	54.4%	74.6%	76.4%
오분류율		40.7%	41.1%	23.6%	21.0%
목적함수 결과		1.129	1.098	3.268	3.885

<Table 3>은 비교대안별 예측정확도와 오분류율 및 목적함수 결과를 정리한 것이다. 또한 예측 수율과 실제 수율의 정확도 및 오분류율도 나열하였다. <Table 3>에서 일반 SVM은 선형 판별분석보다 예측정확도와 오분류율에서 각각 20.0%p, 17.1%p의 차이를 보인다. 특히 고수율로 예측한 Lot이 저수율로 판명되는 경우가 27.1%p의 차이를 보인다. 그러나 본 연구에서 제안한 SSVM은 일반 SVM보다 1.8%p 예측정확도와 2.6%p 오분류율을 개선하였다. 또한 고수율로 예측한 Lot이 저수율로 판명되는 경우가 2.9%p 개선하였다. 실제 공정은 저수율 Lot의 정확히 검출하는 것이 가장 중요하다. 실제 공정의 목적과 부합하여 생각하면 84.7%의 저수율 예측정확도를 보여주는 SSVM이 비교대안 중 가장 좋은 성능을 가지고 있다.

본 연구에서 정확한 분류 및 오분류에 대하여 가중치를 적용한 목적함수 결과는 제안된 SSVM이 가장 큰 값을 보인다. 판별분석은 저수율 Lot을 고수율로 예측하는 오분류가 가장 많

다. 하지만 SSVM이 높은 목적함수 값을 갖는 것은 정확한 분류 및 적은 오분류를 수행하고, 특히 가중치가 높은 저수율 Lot을 고수율로 오분류하지 않기 때문이다. 실제 현장에서는 판별분석 및 단순 선형회귀분석 방법에 따라 고수율과 저수율을 분류하고 있다. 이러한 단순한 방법에 비하여 단계별 고수율 검출을 통해 정확한 저수율을 예측하는 SSVM은 실제 현장에서 우수한 성능을 가질 수 있다. 또한 정확한 저수율 예측이 수율 향상과 제조비용 절감의 중요한 부분임을 고려할 때 본 연구에서 제안한 SSVM은 실제 공정에서 많은 기여를 할 것으로 생각된다. 그러나 고수율 Lot의 예측 결과는 대안들에 비해 크게 개선되지 않았다. 고수율 Lot의 예측이 대안들과 유사한 것은 데이터를 수집할 때 고수율과 저수율의 비율을 50 : 50으로 강제적으로 조정하려해서 고수율 예측을 상대적으로 어렵게 만들었기 때문이다. 또한 본 성능 평가를 실제 현장에서 수행하려고 단계를 세단계로 제약하여 최적의 단계 구성을 수행하지 못했기 때문이다. 만일 최적의 단계 설정이 된다면, 고수율의 분류도 향상될 것으로 기대한다. 따라서 정확한 분류 수행을 위해 최적의 단계 설정에 대한 추후 연구가 필요하다.

해외 FAB에 대한 비교평가는 국내 FAB의 평가와 동일한 절차에 따라 실행하였다. 해외 FAB의 데이터는 본사와 동일한 제품으로 총 1,015개 Lot중 고수율 465개 Lot과 저수율 550개 Lot을 수집하였다. 해외 FAB 데이터를 이용한 성능평가 결과는 <Figure 8>과 같다.

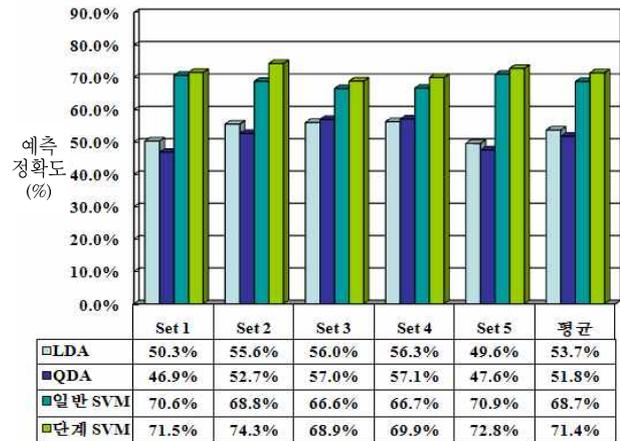


Figure 8. The Results of Each Set(Overseas FAB)

예측정확도는 본사의 결과와 유사하게 나타났다. 평균적으로 본사 대비 예측정확도는 낮지만, 공정 환경이 다른 곳이라도 제안된 SSVM이 성능평가에서 대안들보다 좋은 결과를 보였다. 또한 SSVM을 적용했을 때의 예측정확도, 오분류율 및 목적함수 결과가 <Table 4>에서 보듯 대안들에 비하여 우수함을 알 수 있다. 제안된 SSVM은 선형판별분석보다 저수율 Lot을 고수율로 예측한 경우가 22.2%p 개선되었으며, 저수율 Lot을 저수율로 예측한 경우가 22.2%p 개선되었다. 특히 고수율 Lot을 고수율로 예측한 경우의 개선 정도는 본사의 경우 4.7%p,

해외팩의 경우는 11.1%p로 본사 대비 5.4%p 상승하였다. 고수율 Lot의 개선이 본사보다 높은 것은, 해외 FAB이 아직 안정화되지 않아 고수율 Lot이 적어 단계별 모형파라미터 조정을 통해 쉽게 고수율 Lot을 검출하였기 때문이다. 그러나 본사 대비 많은 저수율 Lot으로 인해 예측정확도에선 5%p 낮으며, 저수율 Lot을 저수율로 예측하는 경우 8.6%p 낮게 나타난다. 이런 현상은 본 연구에서 제안한 SSVM이 고수율 Lot검출을 통해 저수율 Lot을 구별하기 때문이다. 그러나 해외 FAB의 경우 본사와 다른 환경과 제품의 성숙도를 고려할 때 현재의 결과는 다소 긍정적으로 평가되고 있다.

Table 4. The Results of Overseas FAB

예측	실제	LDA	QDA	일반SVM	SSVM
고수율	고수율	53.5%	52.3%	61.9%	64.6%
고수율	저수율	46.1%	48.7%	24.4%	23.9%
저수율	고수율	46.5%	47.7%	38.1%	35.4%
저수율	저수율	53.9%	51.3%	75.6%	76.1%
예측정확도		53.7%	51.8%	68.7%	71.4%
오분류율		46.3%	48.2%	30.0%	28.6%
목적함수 결과		0.999	0.930	2.312	2.430

이상의 성능평가 결과는 본사와 해외 FAB 모두에서 본 연구의 제안한 방법이 가장 우수한 결과를 보여주고 있다. 이는 본 연구에서 제안한 SSVM이 일반 SVM 모형에 비해 단순한 파라미터 선정으로 분류하지 못한 미세한 분류를 단계적으로 가능하게 만들었기 때문이다. 또한 실무 현장에서 고수율 예측 Lot이 저수율 Lot로 발생하는 것을 예방하는 것은 중요하다. 특히 저수율 Lot의 사전 검출 및 예방은 고수율 Lot 검출보다 매우 중요하다. 이상의 실무적 고려사항을 적용했을 때 SSVM은 기존 판별분석을 대체하여 많은 수익창출과 정확한 공정관리를 가능하게 할 것이다.

4. 결론 및 추후연구

본 연구는 반도체 제조공정에서 최종수율 예측을 통한 수율 향상과 불량에 대한 사전예방을 할 수 있는 분류 방법을 제안하였다. 제안된 방법은 1단계 Probe 검사 완료 후 분류기를 이용하여 고수율과 저수율을 구분하고 2단계 Probe 검사를 이원화시켜 최종수율을 향상하는 방법이다.

제안된 방법에 사용된 분류기는 예측 성능이 우수하기로 알려진 SVM을 보다 정밀한 예측이 가능하도록 개선한 SSVM을 제안하였다. 또한 본 연구는 제안된 분류기를 수율 예측으로 널리 사용 중인 기존 LDA, QDA 및 본 연구의 표준 참조 모델로 사용한 SVM과 성능을 비교하였다. 성능 비교는 현장의 실제 데이터를 이용하였고, 대안들과 객관적이며 오분류를 최소화

할 수 있는 방법을 사용하였다. 또한 성능평가 결과 본 연구에서 제안된 SSVM의 우수성을 확인하였고, 저수율 발생 Lot의 정확한 예측을 통해 실무 현장에서의 적용 가능성도 확인하였다. 또한 향후 수율 예측 분야의 분류문제에 있어서 여러 공장 및 공정에서 유용하게 사용될 것으로 생각된다.

그러나 본 연구에서 나타난 고수율 Lot의 예측력 문제를 향상시킬 수 있는 방법에 대한 추후 연구가 필요하다. 또한 SVM을 사용할 때 객관적으로 파라미터를 선정할 수 있는 방법 및 SSVM의 단계를 언제 종료할 것인지에 대한 추후 연구도 필요하다. 특히 모델 수립에 필요한 유효 변수를 효과적으로 추출할 수 있다면 SSVM의 성능이 더욱 향상될 것으로 본다. 따라서 많은 변수 중에 유의한 변수를 추출하는 변수 선정에 대한 연구도 필요하다.

참고문헌

- Baek, D. H., Nam, J. G. (2002), Improved Semiconductor Yield System using Datamining, *Spring Semannual Conference of Korean Operations Research And Management Society*, 298-305.
- Ciciani, B. and Jazeolla, G. (1991), A Markov Chain-Based Yield Formula for VLSI Fault-Tolerant Chips, *IEEE Transactions on Computer-Aided Design*, **10**(2), 252-259.
- Crosier, R. B. (1988), Multivariate Generalizations of Cumulative Sum Quality-control Schemes, *Technometrics*, **30**, 539-549.
- Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., and Scholkopf, B. (1998), Support vector machines, *IEEE Intelligent System*, **13**(4), 18-28.
- Joachims, T. (1998), Text categorization with support vector machines, *Proceedings of the European Conference on Machine Learning* 10th European Conference on Machine Learning, 137-142.
- Kang, B. S., Lee, J. H., Shin, C. K., Yu, S. J., and Park, S. C. (1998), Hybrid machine learning system for integrated yield management in semiconductor manufacturing, *Expert Systems with Applications*, **15**, 123-132.
- Kim, T. S., Bae, G. J. (1995), Research of TEST Trend for High density memory product, *The Institute of Electronics Engineers of Korea*.
- Kinam Kim et al. (1998), DRAM Technology Perspective for Gigabit Era, *IEEE Trans. Electron Devices*, **45**(3), 598-608.
- Meyer, D., Leisch, F., and Hornik, K. (2003), The support vector machine under test, *Neurocomputing*, **55**, 169-186.
- Odom, M. and Sharda, R. (1990), A neural network model for bankruptcy prediction, *Proceedings of the International Joint Conference on Neural networks*, II-163-II-168.
- Osuna, E., Freund, R., and Girosi, F. (1997), Training support vector machines : an application to face detection, *Proceedings of Computer Vision and Pattern Recognition*, 130-136.
- Pieter Pete B. (2000), 2000 begins with a revised industry roadmap, *Solid State Technology*, 31-44.
- Tam, K. and Kiang, M. (1992), Managerial applications of neural networks, *Management Science*, **38**(7), 926-947.
- Uzsoy, R., Lee, C., 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. (1995), The Nature of Statical Learning Theory, *Springer*, New York.
- Weiss, S. and Kulikowski, C. (1991), Computer Systems That Learn, *Morgan Kaufmann Publishers, Inc*.

**안대응**

부경대학교 전자공학과 학사
 현재 : 하이닉스반도체 DRAM 개발사업부
 고려대학교 정보경영공학부석사과정
 관심분야 : Data mining, Machine Learning

**고효현**

명지대학교 산업공학과 학사
 고려대학교 산업공학과 석사
 현재 : 고려대학교 정보경영공학부 박사과정
 관심분야 : 생산관리, e-Business, APC

**김지현**

고려대학교 산업공학과 학사
 고려대학교 산업공학과 석사
 University of Michigan 산업공학과 박사
 현재 : 광운대학교 경영학부 교수
 관심분야 : 생산 및 품질 관리 시스템, APC,
 Data Mining

**백준걸**

고려대학교 산업공학과 학사
 고려대학교 산업공학과 석사
 고려대학교 산업공학과 박사
 현재 : 고려대학교 정보경영공학부 부교수
 관심분야 : 지능형 이상 진단, 첨단공정제어
 (APC), 데이터 마이닝(Data Mining)
 응용

**김성식**

고려대학교 기계공학과 학사
 고려대학교 산업공학과 석사
 Southern Methodist University 산업공학과 석사
 Southern Methodist University 산업공학과 박사
 현재 : 고려대학교 정보경영공학부 교수
 관심분야 : 생산 및 재고관리 시스템, CIM/
 ERP/SCM, APC