

## S-MTS를 이용한 강판의 표면 결함 진단

김준영

고등기술연구원 플랜트 SE팀  
(joonykim@outlook.com)

차재민

고등기술연구원 플랜트 SE팀  
(dr.jaemin.cha@gmail.com)

신중욱

고등기술연구원 플랜트 SE팀  
(paulshin@iae.re.kr)

염충섭

고등기술연구원 플랜트 SE팀  
(csyeum@iae.re.kr)

.....

강판 표면 결함은 강판의 품질과 가격을 결정하는 중요한 요인 중 하나로, 많은 철강 업체는 그동안 검사자의 육안으로 강판 표면 결함을 확인해왔다. 그러나 시각에 의존한 검사는 통상 30% 이상의 판단 오류가 발생함에 따라 검사 신뢰도가 낮은 문제점을 갖고 있다. 따라서 본 연구는 Simultaneous MTS (S-MTS) 알고리즘을 적용하여 보다 지능적이고 높은 정확도를 갖는 새로운 강판 표면 결함 진단 시스템을 제안하였다. S-MTS 알고리즘은 단일 클래스 분류에는 효과적이지만 다중 클래스 분류에서 정확도가 떨어지는 기존 마할라노비스 다구찌 시스템 알고리즘(Mahalanobis Taguchi System; MTS)의 문제점을 해결한 새로운 알고리즘이다. 강판 표면 결함 진단은 대표적인 다중 클래스 분류 문제에 해당하므로, 강판 표면 결함 진단 시스템 구축을 위해 본 연구에서는 S-MTS 알고리즘을 채택하였다. 강판 표면 결함 진단 시스템 개발은 S-MTS 알고리즘에 따라 다음과 같이 진행하였다. 첫째, 각 강판 표면 결함 별로 개별적인 참조 그룹 마할라노비스 공간(Mahalanobis Space; MS)을 구축하였다. 둘째, 구축된 참조 그룹 MS를 기반으로 비교 그룹 마할라노비스 거리(Mahalanobis Distance; MD)를 계산한 후 최소 MD를 갖는 강판 표면 결함을 비교 그룹의 강판 표면 결함으로 판단하였다. 셋째, 강판 표면 결함을 분류하는 데 있어 결함 간의 차이점을 명확하게 해주는 예측 능력이 높은 변수를 파악하였다. 넷째, 예측 능력이 높은 변수만을 이용해 강판 표면 결함 분류를 재수행함으로써 최종적인 강판 표면 결함 진단 시스템을 구축한다. 이와 같은 과정을 통해 구축한 S-MTS 기반 강판 표면 결함 진단 시스템의 정확도는 90.79%로, 이는 기존 검사 방법에 비해 매우 높은 정확도를 갖는 유용한 방법임을 보여준다. 추후 연구에서는 본 연구를 통해 개발된 시스템을 현장 적용하여, 실제 효과성을 검증할 필요가 있다.

**주제어** : 빅데이터, 데이터마이닝, 다중 클래스 분류, 마할라노비스 다구찌 시스템, 강판 표면 결함 진단

.....

논문접수일 : 2016년 11월 17일    논문수정일 : 2016년 12월 22일    게재확정일 : 2016년 12월 25일  
원고유형 : 일반논문                      교신저자 : 김준영

### 1. 서론

강판의 표면 결함은 강판의 품질과 가격에 중대한 영향을 미치는 요인 중 하나이다. 그 동안 많은 철강업체는 강판 표면 검사를 위해 검사자의 시각에 의존하는 방법을 사용해 왔다. 구체적

으로 검사자는 약 150~800m/분의 빠른 속도로 지나가는 강판 상·하면을 비추는 거울 앞에서 두 거울을 바라보며 지나가는 결함들을 확인한 후 검사자의 직관적인 감에 의해 등급을 매겼으나, 이와 같은 시각에 의한 검사 방법은 통상 30%의 오·판단이 발생함에 따라 신뢰도에 심각

한 문제점을 가지고 있다(Kim et al., 2006). 따라서 보다 지능적이고 높은 정확도로 결함을 진단하기 위한 방법이 산업현장으로부터 꾸준히 요구되고 있다.

이러한 필요성에 부합하고자 기계학습과 데이터마이닝 알고리즘을 적용해 강판의 표면 결함을 진단하기 위한 연구들이 수행되어 왔다. 구체적으로 Enhanced Classification Table (ECT)과 Probabilistic Neural Network (PNN)를 활용해 냉연 강판 표면 결함을 분류한 연구(Song et al., 2007), K-Nearest Neighbor (KNN) 기법 및 Tree-Structured Neural Network (TSNN)를 적용한 강판 표면 결함 분류기 개발 연구(Moon et al., 2007), Decision Tree와 Multi-Layer Perception Neural Network (MLPNN), Logistic Regression (LR), Support Vector Machine (SVM), Treebagger 알고리즘 등을 이용한 강판 표면 결함 진단 연구(Fakhr and Elsayad., 2012; Tian et al., 2015 ; Simic et al., 2015)가 있다.

전술한 알고리즘들 이외에도 마할라노비스 다구찌 시스템(Mahalanobis Taguchi System; MTS)은 결함을 진단할 수 있는 알고리즘 중 하나로서, 마할라노비스 거리(Mahalanobis Distance; MD) 개념과 다구찌 품질공학 이론에 기반하여 대상 문제의 정상 그룹과 비정상 그룹을 분류하고 최적화하는 다변량 패턴 분류 알고리즘이다(Taguchi et al., 2002). MTS는 척추질환 환자의 진단(Hong, 2012), 회전기기의 이상 진단(Park et al., 2008), 쿨링 팬과 인덕션 모터의 이상 진단(Jin et al., 2013), 베어링 이상 진단(Ahmet et al., 2010) 등 다양한 분야의 진단 문제에 널리 적용되고 있지만, 3개 이상의 클래스를 분류하는 다중 클래스 분류 문제에서는 정확도가 낮은 한계점을 가지고 있다. 따라서 본 연구에서는 다

양한 강판 표면 결함을 분류하기 위해 다중 클래스 분류가 가능한 진보된 MTS 알고리즘인 Simultaneous MTS (S-MTS)를 적용하여(Cha and Kim et al., 2016) 강판 표면 결함 진단 시스템을 개발하고, 선행 연구에서 제시하고 있는 타 알고리즘 기반 결함 진단 시스템과의 성능 비교를 통하여 이에 대한 효과성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 다변량 패턴 분석 알고리즘인 MTS에 대해 설명하고, 3장에서는 본 연구에 적용된 S-MTS 알고리즘 설명과 그 외 MTS 알고리즘들과 비교 분석을 수행한다. 4장에서는 S-MTS 기반 강판 표면 결함 진단 시스템을 구축하고, 선행 연구들과 정확도를 비교함으로써 본 진단 시스템의 성능을 확인한다. 마지막으로 5장에서는 결론과 함께 논문을 마무리한다.

## 2. MTS (Mahalanobis Taguchi System)

본 장에서는 S-MTS의 이론을 설명하기에 앞서, S-MTS의 기본이 되는 MTS 이론에 대해 간략하게 설명한다. MTS는 정상 그룹의 중심점을 기준으로 각 개별 데이터를 단위 공간 안에 표현한 후 이들이 중심점으로부터 얼마나 멀리 떨어져 있는지를 MD로 계산하게 된다. 여기서 MD가 클수록 정상 그룹이 아닐 가능성이 높아지며, 클래스 구분을 위한 문턱값(Threshold Value)을 초과하면 비정상 그룹으로 판단한다. 이와 같은 MTS는 다변량 데이터를 MD라는 하나의 종합 지표로 쉽게 표현할 수 있으며, 거리 편차를 극대화하는 변수를 선택적으로 선정함으로써 클래스 구분을 보다 효과적으로 할 수 있다는 특징을 가지고 있다. MTS의 설계 절차는 다음과 같이

크게 4단계로 구성된다.

첫 번째 단계는 정상 그룹의 MS를 구축하는 단계이다. 이를 위해 우선 적절한 변수로 구성된 정상 상태의 데이터를 수집하고, 정상 데이터를 기반으로 한 정상 그룹을 선정한다. 다음으로, 정상 그룹으로부터 MS를 도출한다. 중심점으로부터 거리 평균이 1.0이 되는 MS를 도출하기 위해서는 정상 그룹 데이터를 정규화 해야 되며, 정규화는 측정값에서 평균값을 뺀 값을 표준편차로 나누어준다. 이때 수식은 <Eq. 1>과 같다. 여기서  $i$ 는 변수의 번호,  $x_i$ 는  $i$ 번째 변수의 임의의 값,  $m_i$ 는  $i$ 번째 변수의 평균,  $s_i$ 는  $i$ 번째 변수의 표준편차이다.

$$Z_i = \frac{x_i - m_i}{s_i} \quad \text{<Eq. 1>}$$

이렇게 도출한 정규화 된 벡터( $Z$ )는 상관계수로 구성된 상관행렬( $C$ )의 역행렬을 이용하여 정상 그룹의 MD를 계산하며, 이에 대한 수식은 <Eq. 2>와 같다. 여기서  $D_i^2$ 는  $i$ 번째의 MD,  $k$ 는 변수의 개수,  $Z$ 는  $i$ 번째 변수의 정규화 된 벡터이고,  $Z_i^T$ 는  $Z$ 의 전치행렬이며,  $C^{-1}$ 는 상관행렬의 역행렬이다.

$$MD_i = D_i^2 = \frac{1}{k} Z_i C^{-1} Z_i^T \quad \text{<Eq. 2>}$$

해당 계산식에서 공분산행렬의 역행렬 대신 상관행렬의 역행렬을 사용하는 이유는 상관행렬은  $\pm 1$  범위 내에 있는 단위 값인 상관계수로 구성되어 있기 때문에 공분산행렬에 비해 계산과 해석이 빠르다는 장점을 활용하기 위함이다.

두 번째 단계에서는 앞서 구축된 MS의 유효

성을 확인한다. 이를 위해, 우선 비정상 데이터로 구성된 비정상 그룹을 선정하고, 정상 그룹의 MS를 기준으로 비정상 그룹의 MD를 계산한다. 이때, 비정상 그룹을 정상 그룹과 같은 공간에서 비교하기 위해, 비정상 그룹의 측정값에서 정상 그룹의 평균값을 뺀 후 이를 정상 그룹의 표준편차로 나눔으로써 비정상 그룹을 표준화 한다. 이에 대한 수식은 <Eq. 3>과 같다. 여기서  $y_i$ 는  $i$ 번째 변수의 임의의 값,  $m_i$ 는 정상 그룹의  $i$ 번째 변수의 평균,  $S_i$ 는 정상 그룹의  $i$ 번째 변수의 표준편차이다.

$$Z_i = \frac{y_i - m_i}{S_i} \quad \text{<Eq. 3>}$$

이어서 비정상 그룹의 MD를 계산하게 되며, 계산된 비정상 그룹의 MD와 앞선 단계에서 계산된 정상 그룹의 MD를 비교하였을 때 두 그룹간의 차이가 명확히 구분된다면 초기 설정한 MS 공간이 유효성 있게 구축되었다고 평가할 수 있다.

세 번째 단계에서는 예측 능력이 높은 유용한 변수를 설정한다. 여기서 예측 능력이 높은 변수는 다른 변수들에 비해 정상 그룹과 비정상 그룹간의 차이를 보다 크게 하여 그룹간의 구분을 명확하게 하는데 도움이 되는 변수를 의미하는데, 본 절차는 이들을 선별함으로써 시스템의 정확도를 높이기 위함이다. 변수들의 예측능력을 평가하기 위해 2수준계 직교배열표와 SN비(Signal-to-Noise Ratio; SNR)를 활용한다. 우선 내측배열과 외측배열에 변수와 신호인자를 배치하여 각 실험 조건에 따라 MD를 계산하고, 이로부터 SN비를 도출한다. SN비의 수식은 변수값의 특성에 따라 달라지는데, 변수 값이 클수록

좋은 특성을 의미할 경우에는 망대특성 (Larger-the-better)의 SN비, 변수 값이 작을수록 좋은 특성을 의미할 경우에는 망소특성 (Smaller-the-better)의 SN비, 변수 값이 특정 목표 값에 가까울수록 좋은 특성을 의미할 경우에는 망목특성(Nominal-the-better)의 SN비, 알려진 신호인자가 있는 경우에는 동특성(Dynamic)의 SN비 계산식을 사용한다. 이 중 동특성의 SN비의 예측능력이 정확도가 가장 높은 것으로 알려져 있으며, 이에 대한 수식은 <Eq. 4>와 같다 (Taguchi et al., 2002). 여기서 n은 비정상 그룹의 시료 수,  $S_m$ 은 평균제곱합(Sum of squares due to mean),  $V_e$ 는 오차 분산(Error variance)이다.

$$SN\ Ratio = 10 \log_{10} \left[ \frac{1}{n} \frac{(S_m - V_e)}{V_e} \right] \quad \text{<Eq. 4>}$$

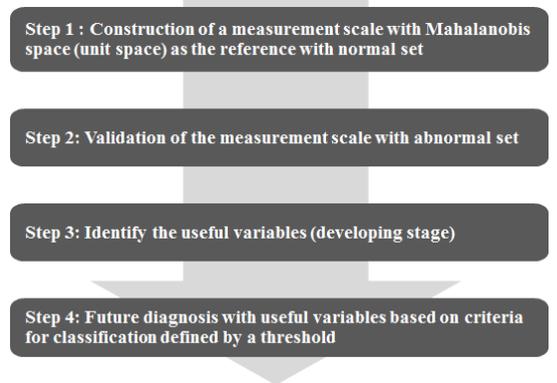
이어서, 도출된 실험 조건 별 SN비를 이용하여 변수 별로 예측 능력이 있는지 여부를 판단하기 위해, SN비 이득(SNR Gain)을 계산한다. 이를 위해서 변수 별로 해당 변수가 사용된 실험 조건들의 SN비 평균에서 해당 변수가 사용되지 않은 실험 조건들의 SN비 평균의 차를 구하면 해당 변수의 SN비 이득이 된다. 이에 대한 수식은 <Eq. 5>와 같다. 여기서 i는 변수의 번호이고,  $\overline{SN_i^+}$ 은 i번째 변수가 사용된 SN비의 평균이고,  $\overline{SN_i^-}$ 은 i번째 변수가 사용되지 않은 SN비의 평균이다.

$$SNR\ Gain_i = \overline{SN_i^+} - \overline{SN_i^-} \quad \text{<Eq. 5>}$$

도출된 변수 별 SN비 이득이 음의 값을 갖는 경우 해당 변수는 예측 능력이 없는 것으로, SN

비 이득이 양의 값을 갖는 경우 예측 능력이 있는 것으로 해석할 수 있으며, 그 값이 클수록 예측 능력이 높은 주요 변수임을 의미한다.

마지막으로 네 번째 단계에서는 예측 능력이 높은 변수로 구성된 진단 시스템을 구축한다. 앞선 단계에서 선정된 예측 능력이 높은 변수만을 이용해 진단 시스템을 개발한 후 이들의 MD를 계산하여 정상 그룹과 비정상 그룹의 판단에 활용한다. <Figure 1>은 전술한 MTS의 설계 절차를 보인다.



<Figure 1> MTS Process (Taguchi et al., 2002)

### 3. Simultaneous MTS (S-MTS)

본 장에서는 본 연구에서 적용한 S-MTS의 이론을 설명한다. 기존 MTS는 기본적으로 기준이 되는 그룹의 기준 MS를 구축한 뒤 비교할 그룹이 기준 MS로부터 얼마나 떨어져 있는지 MD로 계산하여 클래스를 판별하는 방식인데, 이 방식은 분류해야 할 클래스가 다수일 경우에도 기준이 되는 그룹의 MS에 의해서만 클래스를 분류하기 때문에 근본적으로 다중 클래스간의 분류 정

확도가 떨어지는 문제점을 갖고 있다. 이러한 문제점을 극복하기 위해 S-MTS는 다양한 비교 그룹 별로 MS를 구축하여 MD를 계산한 뒤, 이들 중 최소의 MD를 갖는 클래스를 해당 비교 그룹의 클래스로 판별하는 방식을 채택하였다. S-MTS에서 ‘Simultaneous’의 의미는 각 상태 별 MS를 기준으로 비교 그룹의 MD를 동시에 비교함으로써 적절한 클래스를 식별하는 데서 기인한다. S-MTS의 설계 절차는 크게 4단계로 구성된다.

첫 번째 단계는 다양한 상태 별 참조 그룹을 선정하고 각 상태 별 MS를 구축하는 단계이다. 이를 위해 대상 시스템의 다양한 상태를 식별하고, 적절한 변수로 구성된 다양한 상태 별 데이터를 수집해 이를 기반으로 참조 그룹을 선정한다. 다음으로 상태 별 참조 그룹의 MD를 계산해 상태 별 참조 그룹 MS를 구축한다. 이때 중심점으로부터 거리 평균이 1.0이 되는 MS를 도출하기 위해 상태 별 참조 그룹의 데이터를 정규화하며, 정규화는 측정값에서 평균값을 뺀 값을 표준편차로 나누어 계산한다. 이에 대한 수식은 <Eq. 6>과 같다. 여기서  $c$ 는 식별이 가능한 상태,  $x_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 임의의 값,  $m_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 평균,  $s_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 표준편차이다.

$$Z_{c,i} = \frac{x_{c,i} - m_{c,i}}{s_{c,i}} \quad \text{<Eq. 6>}$$

이렇게 도출한 상태 별 참조 그룹의 정규화 된 벡터는( $Z_c$ )는 상관계수로 구성된 상태 별 상관행렬( $C_c$ )의 역행렬을 이용하여 상태 별 참조 그룹의 MD를 계산하며, 이에 대한 수식은 <Eq. 7>과 같다. 여기서  $k_c$ 는  $c$ 상태의 변수 개수,  $Z_{c,i}$ 는  $c$ 상

태의  $i$ 번째 변수의 정규화 된 벡터이고,  $Z_{c,i}^T$ 는  $Z_{c,i}$ 의 전치행렬이며,  $C_c^{-1}$ 는  $c$ 상태의 상관행렬의 역행렬이다.

$$MD_{c,i} = D_{c,j}^2 = \frac{1}{k_c} Z_{c,i} C_c^{-1} Z_{c,i}^T \quad \text{<Eq. 7>}$$

두 번째 단계에서는 앞서 구축된 상태 별 참조 그룹 MS의 유효성을 확인한다. 이를 위해 이미 상태를 알고 있는 다양한 상태 별 비교 데이터를 수집해 상태 별 비교 그룹을 선정하고, 상태 별 참조 그룹 MS를 기준으로 비교 그룹의 MD를 계산한다. 이때, 비교 그룹을 참조 그룹과 같은 공간에서 비교하기 위하여 비교 그룹의 측정값에서 참조 그룹의 평균값을 뺀 후 이를 참조 그룹의 표준편차로 나눔으로써 비교 그룹을 표준화한다. 이에 대한 수식은 <Eq. 8>과 같다. 여기서  $y_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 임의의 값,  $m_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 평균,  $s_{c,i}$ 는  $c$ 상태의  $i$ 번째 변수의 표준편차이다.

$$Z_{c,i} = \frac{y_{c,i} - m_{c,i}}{s_{c,i}} \quad \text{<Eq. 8>}$$

이어서 비교 그룹의 MD를 계산하게 되며, 계산 결과를 바탕으로 비교 그룹의 MD가 참조 그룹의 상태 별 MD와 가장 차이가 적은 상태를 해당 비교 그룹의 상태로 식별한다. 다음으로 비교 그룹의 식별된 상태와 실제 상태가 일치하는지 확인한다. 이를 통해 구축된 상태 별 MS가 다양한 상태를 잘 식별할 수 있는지 여부, 다시 말해 상태 별 MS의 유효성을 확인한다.

세 번째 단계에서는 예측 능력이 높은 유용한 변수를 설정한다. 여기서 예측 능력이 높은 변수

는 다른 변수들에 비해 상태 간 구분을 명확하게 하는데 도움이 되는 변수를 의미하며, 본 절차는 이들을 선별함으로써 시스템의 정확도를 높이기 위함이다. S-MTS에서 변수 예측능력을 평가하기 위한 절차는 기본적으로 MTS의 절차와 비슷하지만, 다양한 상태를 고려하여 약간의 차이가 존재한다. 구체적으로, MTS에서는 변수들의 예측능력을 평가하기 위해 1개의 직교배열표를 활용하여 실험 조건 별로 MD를 계산하고, 이로부터 SN비와 SN비 이득을 도출한다. 하지만, S-MTS에서는 우선 상태 별로 SN비와 SN비 이득을 도출하고, 다음으로 도출된 상태 별 SN비와 SN비 이득을 더하여 모든 상태를 고려한 총괄 SN비 이득(Overall SNR Gain)을 도출한다. 이는 상태 별로 해당 상태와 다른 상태와의 구분에 영향을 주는 변수들이 다를 수 있기 때문에 모든 상태 간 식별을 종합적으로 고려하기 위함이다. 이에 대한 수식은 <Eq. 9>, <Eq. 10>과 같다. 여기서  $\overline{SN_{c,i}^+}$  은 i번째 변수가 사용된 c상태의 SN비 평균이고,  $\overline{SN_{c,i}^-}$  은 i번째 변수가 사용되지 않은 c상태의 SN비 평균이다.

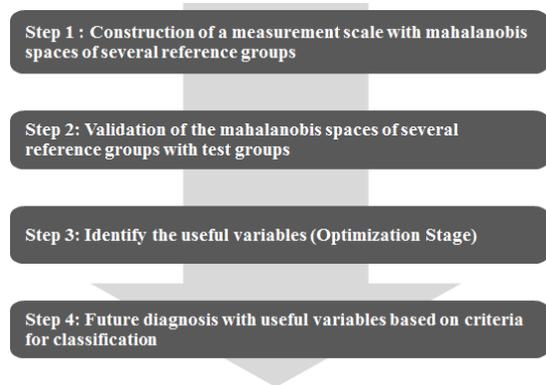
$$SNR\ Gain_{c,i} = \overline{SN_{c,i}^+} - \overline{SN_{c,i}^-} \quad \langle \text{Eq. 9} \rangle$$

$$Overall\ SNR\ Gain_i = \sum_c^n SNR\ Gain_{c,i} \quad \langle \text{Eq. 10} \rangle$$

도출된 총괄 SN비 이득이 음의 값을 갖는 경우 해당 변수는 예측 능력이 없는 것으로, 총괄 SN비 이득이 양의 값을 갖는 경우 예측 능력이 있는 것으로 해석할 수 있으며, 그 값이 클수록 예측 능력이 높은 주요 변수에 해당한다.

마지막으로, 네 번째 단계에서는 예측 능력이 높은 변수로 구성된 진단 시스템을 구축한다. 예측 능력이 높은 변수만으로 진단 시스템을 개발

하여 강관 표면 결함 분류를 수행한다. <Figure 2>는 전술한 S-MTS의 절차를 보인다.



<Figure 2> S-MTS Process  
(Cha and Kim et al., 2016)

S-MTS와 같이 다중 클래스 분류에서 낮은 정확도를 갖는 MTS를 개선하기 위한 선행 연구들이 존재하는데, 대표적으로 정상 그룹 및 비정상 그룹의 MS를 구축하고 관찰 데이터의 MD를 가장 작게 하는 클래스를 해당 데이터의 클래스로 분류하는 Multiclass MTS (MMTS) 연구(Su and Hsiao, 2009), 회전 기기의 상태 모니터링을 위해 정상 상태의 표준 공간 (Standard Space)과 다양한 이상 상태 별 공간(Faulty Space)를 구축해 이들을 서로 구분하는 Improved MTS 연구(Ren et al., 2011)가 있다. 선행 연구들과 S-MTS의 비교 결과는 <Table 1>과 같다. 해당 방법들은 하나의 MS만 구축함에 따라 다중 클래스 분류가 어려운 MTS의 근본적인 문제를 해결하기 위해 다수의 MS를 구축하고 이 중 최소 MD를 갖는 클래스를 해당 데이터의 클래스로 분류한다는 점에서 유사한 접근 방법을 가지고 있으나, MD 계산과정에서 나타나는 다중공선성

〈Table 1〉 Theoretical Comparison between MMTS, Improved MTS and S-MTS

Category	MMTS (2009)	Improved MTS (2011)	S-MTS (2016)
No. of MS	Multiple	Multiple	Multiple
Solution for multicollinearity	Gram Schmidt Orthogonalization	Removal of Multicollinearity Variables	Removal of Multicollinearity Variables
Classification of Classes	Minimum MD	Minimum MD	Minimum MD
Characteristics for SNR	Larger-the-better	Not Available	Dynamic

(Multicollinearity) 문제를 해소하는 방법에서 차이가 나타난다. 다중공선성 문제란, 상관관계가 높은 변수들이 포함된 다변량 데이터의 경우 비정칙행렬(Singular Matrix)이 발생함에 따라 MD 계산이 불가능한 현상을 의미하는데, 이와 같은 문제를 해결하고자 S-MTS와 Improved MTS는 상관관계가 높은 변수 자체를 제거하는 방법을 채택한 반면, MMTS는 데이터를 직교변환하여 변수간 상관관계를 없애는 방법을 채택하고 있다. 이뿐만 아니라 SN비 계산식에서도 차이가 있는데, S-MTS는 MD 값을 신호인자 수준값으로 하는 동특성(Dynamic) SN비 계산식을 사용하는 반면, MMTS는 변수 값이 클수록 좋은 특성을 의미하는 망대특성 SN비 계산식을 사용하는 특징이 있다.

## 4. S-MTS를 이용한 강판 결함 진단

### 4.1 강판 표면 결함 데이터 (Steel Plate Faults Data Set)

본 연구를 위해 University of California at Irvine (UCI) Machining Learning Repository의

Steel Plate Faults Data Set을 활용하였다 (Semeion, 2016). 강판 표면 결함 데이터는 결함 중에서도 발생빈도가 가장 큰 7가지 결함(Pastry, Z\_Scratch, K\_Scratch, Stains, Dirtiness, Bumps, Other Faults)을 구분하기 위해 결함 발생과 관련이 있을 것으로 예상되는 변수를 기록한 데이터이다. 다시 말해, 27개 변수(측정항목)를 기록한 1,941개의 강판 표면 측정 데이터로 <Table 2>와 같이 구성되어 있다. 참고로 S-MTS 강판 표면 결함 진단에는 MATLAB R2013과 EXCEL을 이용하였음을 밝혀둔다.

〈Table 2〉 Steel Plate Faults Data Set

Class	Type of Faults	No. of classes
C1	Pastry	158
C2	Z_Scratch	190
C3	K_Scratch	391
C4	Stains	72
C5	Dirtiness	55
C6	Bumps	402
C7	Other Faults	673
Total		1,941

### 4.2 강판 표면 결함 데이터 전처리

본 단계는 S-MTS를 적용하기 이전에 앞서 강판 표면 결함 데이터의 전처리 작업을 진행하였

다. 대표적으로 해당 데이터는 동일 내용이 기록된 중복 변수가 존재하는데, 강판 종류가 A300 타입일 경우 변수 12의 값을 1로, 변수 13의 값은 0으로 표시하고 있으며, 반대로 강판의 종류

(Table 3) Variables of Steel Plate Faults

Before Data Treatment		After Data Treatment	
No.	Variable Name	No.	Variable Name
1	X_Minimum	1	X_Minimum
2	X_Maximum	2	X_Maximum
3	Y_Minimum	3	Y_Minimum
4	Y_Maximum	4	Y_Maximum
5	Pixels_Areas	5	Pixels_Areas
6	X_Perimeter	6	X_Perimeter
7	Y_Perimeter	7	Y_Perimeter
8	Sum_of_Luminosity	8	Sum_of_Luminosity
9	Minimum_of_Luminosity	9	Minimum_of_Luminosity
10	Maximum_of_Luminosity	10	Maximum_of_Luminosity
11	Length_of_Conveyer	11	Length_of_Conveyer
12	Type Of Steel_A300	12	Type Of Steel
13	Type Of Steel_A400	13	Steel_Plate_Thickness
14	Steel_Plate_Thickness	14	Edges_Index
15	Edges_Index	15	Empty_Index
16	Empty_Index	16	Square_Index
17	Square_Index	17	Outside_X_Index
18	Outside_X_Index	18	Edges_X_Index
19	Edges_X_Index	19	Edges_Y_Index
20	Edges_Y_Index	20	Outside_Global_Index
21	Outside_Global_Index	21	LogOfAreas
22	LogOfAreas	22	Log_X_Index
23	Log_X_Index	23	Log_Y_Index
24	Log_Y_Index	24	Orientation_Index
25	Orientation_Index	25	Luminosity_Index
26	Luminosity_Index	26	SigmoidOfAreas
27	SigmoidOfAreas		

가 A400 타입일 경우와 변수 12는 0으로, 변수 13에는 1로 기록하고 있다. 이러한 경우 변수들 간의 상관관계때문에 S-MTS를 적용하기 어려워 <Table 3>과 같이 변수 12(Type Of Steel\_ A300)와 변수 13(Type Of Steel\_A400)을 하나의 변수(Type Of Steel)로 통합해 A300 타입과 A400 타입을 각각 0과 1로 표시함으로써 중복 변수 문제를 해결하였다.

또한 7가지 강판 표면 결함 중 Other Faults (C7)은 정확히 무슨 결함인지 정의되지 않거나, 여러 결함이 복합적으로 발생한 것을 나타낸다. 이는 나머지 결함과 달리 결함 종류 및 발생 원인이 정의되지 않아 별도의 클래스로 구분하는데 한계가 있다. 따라서 본 연구는 클래스 구분이 불가능한 Other Faults(C7)를 제외한 나머지 6가지 강판 표면 결함 데이터(1,268개)에 S-MTS를 적용하였다.

#### 4.3 참조 그룹 선정 및 MS 구축

본 단계는 S-MTS의 첫번째 단계로써 강판 표면 결함 별로 참조 그룹의 MS를 구축한다. 이를 위해 본 과정에는 참조 그룹별로 총 6개의 MS(MSC1, MSC2, MSC3, MSC4, MSC5, MSC6)가 구축된다. 이 과정에서 사용되는 데이터는 <Table 4>와 같이 70:30 비율에 따라 888개 데이터를 참조 그룹 선정 및 MS 구축에, 나머지 380개의 데이터는 비교 그룹 식별 및 MS 유효성 확인으로 사용하였다.

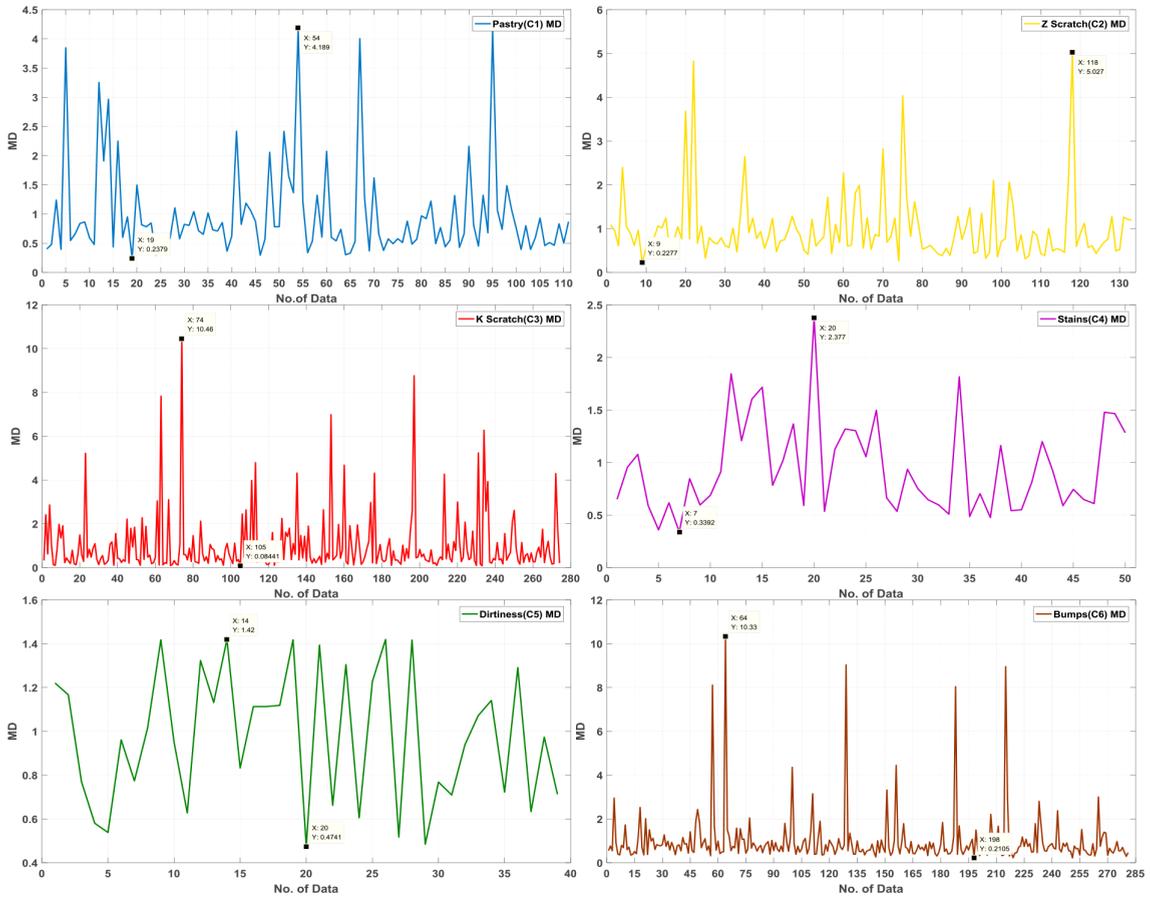
<Table 4> Reference and Test Group

Class	No. of Reference Group	No. of Test Group
C1	111	47
C2	133	57
C3	274	117
C4	50	22
C5	39	16
C6	281	121
Total	888	380

상관행렬의 역행렬( $R^{-1}$ )을 이용해 6가지 강판 표면 결함 별 참조 그룹 MD를 계산한 결과, 참조 그룹의 평균 MD가 0.942-0.996 범위 내에 존재하여 참조 그룹의 MS가 잘 구축되었다고 평가할 수 있다. <Table 5>는 각 참조 그룹 별 최소, 최대, 평균 MD를 제시하고 있으며, 계산된 강판 결함 별 MD는 <Figure 3>과 같다.

<Table 5> MDs of Reference Group

Class	Min. MD	Max. MD	Ave. MD
C1	0.238	4.189	0.991
C2	0.228	5.027	0.992
C3	0.084	10.462	0.996
C4	0.339	2.377	0.942
C5	0.474	1.420	0.973
C6	0.211	10.330	0.996



〈Figure 3〉 MDs of Reference Group

#### 4.4 비교 그룹 식별 및 MS 유효성 확인

본 단계는 S-MTS의 두번째 단계로써 비교 그룹을 식별하고, 첫번째 단계에서 구축한 참조 그룹 MS의 유효성을 확인한다. 이를 위해 본 과정에서는 비교 그룹에서 사용한 데이터를 제외하 나머지 380개의 데이터를 비교 그룹으로 선정하였다.

이후 앞서 구축한 강판 표면 결함 별 참조 그룹  $MS_{Cj}$  ( $MS_{C1}$ ,  $MS_{C2}$ ,  $MS_{C3}$ ,  $MS_{C4}$ ,  $MS_{C5}$ ,  $MS_{C6}$ )을 이용하여 비교 그룹 MD를 계산하는데, 이때

강판 표면 결함 별 참조 그룹( $MS_{Cj}$ )의 평균과 표준편차를 이용하여 비교 그룹( $C_i$ )의 MD를 계산한다. 즉, C1 결함의 비교 그룹 MD( $MD_{C1}$ )는 참조 그룹의  $MS_{Cj}$ 인  $MS_{C1}$ ,  $MS_{C2}$ ,  $MS_{C3}$ ,  $MS_{C4}$ ,  $MS_{C5}$ ,  $MS_{C6}$ 별로 총 6개의 MD( $MD_{C1-MS_{C1}}$ ,  $MD_{C1-MS_{C2}}$ ,  $MD_{C1-MS_{C3}}$ ,  $MD_{C1-MS_{C4}}$ ,  $MD_{C1-MS_{C5}}$ ,  $MD_{C1-MS_{C6}}$ )가 계산되며, 나머지 5가지 결함(C2, C3, C4, C5, C6)도 마찬가지로 각각 6개의 MD가 산출된다. <Figure 4>는 각 비교 그룹 MD 및 최소 MD를 제시하고 있다.

<b>C1(Pastrv) MD</b>							
No.	MD <sub>C1-MS1</sub>	MD <sub>C1-MS2</sub>	MD <sub>C1-MS3</sub>	MD <sub>C1-MS4</sub>	MD <sub>C1-MS5</sub>	MD <sub>C1-MS6</sub>	Min. MD
1	1.18	9.85	86.83	-6116516468298770.00	111.44	3.51	MD <sub>C1-MS1</sub>
2	0.78	23.60	1264.12	-3822829898441600.00	6339.20	24.09	MD <sub>C1-MS1</sub>
3	0.36	3.98	99.19	-1529128226410790.00	6.58	0.74	MD <sub>C1-MS1</sub>
4	0.67	9.17	59.30	3575.14	151.30	0.89	MD <sub>C1-MS1</sub>
5	1.69	11.85	292.01	-6116518708332570.00	6907.54	2.26	MD <sub>C1-MS1</sub>
6	0.62	11.37	34.74	-382282835124775.00	4783.70	1.93	MD <sub>C1-MS1</sub>
...							
47	0.49	6.77	745.12	-21503384062321900.00	4348.27	2.32	MD <sub>C1-MS1</sub>
<b>C2(Z Scratch) MD</b>							
No.	MD <sub>C2-MS1</sub>	MD <sub>C2-MS2</sub>	MD <sub>C2-MS3</sub>	MD <sub>C2-MS4</sub>	MD <sub>C2-MS5</sub>	MD <sub>C2-MS6</sub>	Min. MD
1	7.46	0.88	50.32	-3440541347585100.00	6059.40	1.21	MD <sub>C2-MS2</sub>
2	5.69	1.07	50.77	-3440543701038820.00	5352.50	1.31	MD <sub>C2-MS2</sub>
3	1.49	0.46	57.84	-3440544862226540.00	6308.10	0.69	MD <sub>C2-MS2</sub>
4	0.60	0.86	61.43	-3440541055365940.00	5463.79	1.28	MD <sub>C2-MS1</sub>
5	3.92	0.80	59.00	-3440541484817280.00	6235.54	0.93	MD <sub>C2-MS2</sub>
6	6.69	0.47	55.40	-3440546431699110.00	5973.80	0.78	MD <sub>C2-MS2</sub>
...							
57	0.55	0.58	59.45	-3440541029399770.00	5746.13	0.89	MD <sub>C2-MS1</sub>
<b>C3(K Scratch) MD</b>							
No.	MD <sub>C3-MS1</sub>	MD <sub>C3-MS2</sub>	MD <sub>C3-MS3</sub>	MD <sub>C3-MS4</sub>	MD <sub>C3-MS5</sub>	MD <sub>C3-MS6</sub>	Min. MD
1	1531.51	131.09	0.16	-382272365699669.00	11213.80	170.68	MD <sub>C3-MS3</sub>
2	3.31	5.03	5.20	-382282542369806.00	202.33	3.28	MD <sub>C3-MS6</sub>
3	2085.80	99.92	0.39	-382268024166748.00	35238.75	258.11	MD <sub>C3-MS3</sub>
4	1410.29	100.71	0.17	-382272150622461.00	13915.71	137.52	MD <sub>C3-MS3</sub>
5	34786.69	2457.56	2.88	-382236733286551.00	173515.30	1406.07	MD <sub>C3-MS3</sub>
6	1572.54	128.43	0.22	-382272457695116.00	10178.50	149.29	MD <sub>C3-MS3</sub>
...							
117	4.43	6.60	1.29	-382282519854286.00	203.51	4.55	MD <sub>C3-MS3</sub>
<b>C4(Stains) MD</b>							
No.	MD <sub>C4-MS1</sub>	MD <sub>C4-MS2</sub>	MD <sub>C4-MS3</sub>	MD <sub>C4-MS4</sub>	MD <sub>C4-MS5</sub>	MD <sub>C4-MS6</sub>	Min. MD
1	4.07	4.01	13.07	0.86	83.3	2.26	MD <sub>C4-MS4</sub>
2	6.46	6.11	12.34	0.49	142.1	3.30	MD <sub>C4-MS4</sub>
3	8.73	5.07	15.68	1.00	120.4	3.42	MD <sub>C4-MS4</sub>
4	24.52	5.62	14.93	1038.68	154.2	2.72	MD <sub>C4-MS6</sub>
5	16.36	8.25	13.64	0.61	121.3	5.01	MD <sub>C4-MS4</sub>
6	15.73	4.95	12.98	1.93	201.5	2.85	MD <sub>C4-MS4</sub>
...							
22	4.80	9.16	11.52	0.79	143.9	3.37	MD <sub>C4-MS4</sub>
<b>C5(Dirtiness) MD</b>							
No.	MD <sub>C5-MS1</sub>	MD <sub>C5-MS2</sub>	MD <sub>C5-MS3</sub>	MD <sub>C5-MS4</sub>	MD <sub>C5-MS5</sub>	MD <sub>C5-MS6</sub>	Min. MD
1	0.66	3.26	369.94	-9557061852529520.00	1.02	0.81	MD <sub>C5-MS1</sub>
2	2.14	4.50	395.95	-9557096464955730.00	6.29	31.08	MD <sub>C5-MS1</sub>
3	1.46	3.93	375.20	-9557077628430090.00	2.92	13.75	MD <sub>C5-MS1</sub>
4	2.19	8.61	28.92	-382281160102445.00	2.89	4.44	MD <sub>C5-MS1</sub>
5	1.21	3.80	379.88	-9557076931099870.00	0.62	11.87	MD <sub>C5-MS5</sub>
6	31.13	8.52	27.63	-3673735593472120.00	3.06	1.84	MD <sub>C5-MS6</sub>
...							
16	1.97	14.52	118.64	61512.02	5649.09	25.83	MD <sub>C5-MS1</sub>
<b>C6(Bumps) MD</b>							
No.	MD <sub>C6-MS1</sub>	MD <sub>C6-MS2</sub>	MD <sub>C6-MS3</sub>	MD <sub>C6-MS4</sub>	MD <sub>C6-MS5</sub>	MD <sub>C6-MS6</sub>	Min. MD
1	1.26	3.52	2.96	-382282451148820.00	26.14	0.37	MD <sub>C6-MS6</sub>
2	23.45	1.42	43.58	-1529130748259950.00	6257.89	0.78	MD <sub>C6-MS6</sub>
3	1.68	8.80	22.58	-3440546378108060.00	111.14	0.52	MD <sub>C6-MS6</sub>
4	13.31	2.56	41.35	-1529131159271650.00	5901.23	0.58	MD <sub>C6-MS6</sub>
5	0.85	2.01	1875.61	-59731630496284700.00	20.05	0.90	MD <sub>C6-MS6</sub>
6	3.43	10.04	33.18	-3440542173586580.00	49.18	0.39	MD <sub>C6-MS6</sub>
...							
121	0.92	7.67	59.14	-6116520439809140.00	129.15	0.26	MD <sub>C6-MS6</sub>

〈Figure 4〉 MDs of Test Group

(Table 6) Accuracy of S-MTS for Steel Plate Faults

Class	Correct	Incorrect	Accuracy (%)
C1	40	7	85.11
C2	48	9	84.21
C3	109	8	93.16
C4	18	4	81.82
C5	3	13	18.75
C6	108	13	89.26
Average	326	54	85.79

이때 최소 MD를 갖는 강판 표면 결함을 해당 비교 그룹의 강판 표면 결함으로 식별하였으며, 이것이 실제 결함과 일치하는지 여부를 검증한 결과 <Table 6>과 같이 평균 정확도가 85.79%인 것으로 확인하였다. 강판 표면 결함 별 정확도를 살펴보면, C5(Dirtiness)의 경우 C1(Pastry) 데이터와의 높은 상관관계로 인해 C1(Pastry) 결함으로 오판함에 따라 상대적으로 낮은 정확도를 가지고 있었으나, 이를 제외한 나머지 강판 표면 결함에서는 84% 이상의 정확도를 갖는 것으로 확인하였다.

#### 4.5 유용한 변수 선정

본 단계는 S-MTS의 세번째 단계로써 강판 표면 결함 식별에 도움이 되는 유용한 변수를 선정한다. 여기서 말하는 유용한 변수란 강판 표면 결함들 간의 구분을 명확하게 해주는 변수로써, 보다 효과적인 결함 분류 시스템을 구축하기 위해 필요한 단계이다. 이를 위해 본 연구에서는 직교배열표를 활용하여 실험 조건 별 MD를 계산한 후 각 변수들의 SN비 이득을 도출하여 예측능력이 높은 유용한 변수를 선정하였다. 이때

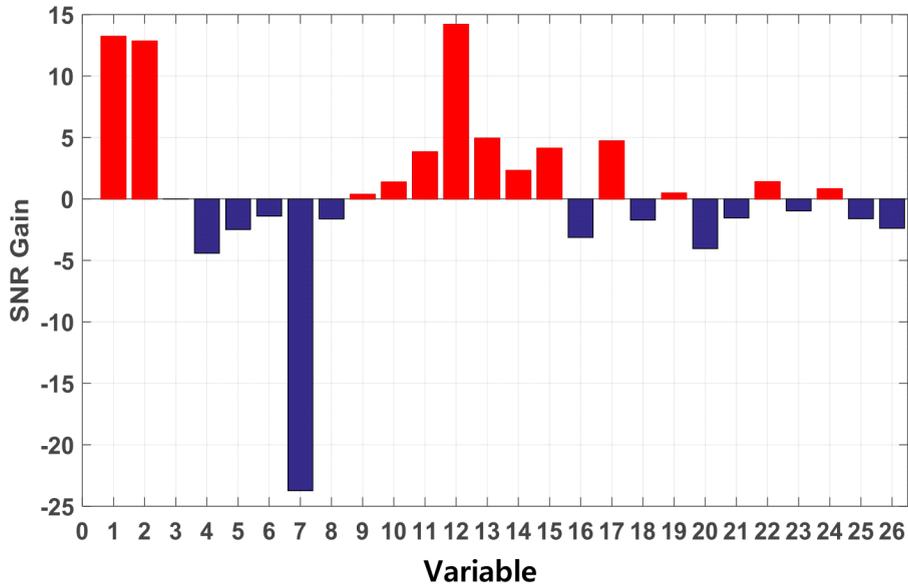
사용한 직교배열표는 26개 변수를 다루면서, 가장 적은 실험 횟수를 갖는 L32 직교배열표를 사용하였으며, SN비 계산식은 망대특성 SN비 보다 정확한 예측이 가능한 것으로 알려진 동특성 SN비를 사용하였다.

이로부터 강판 표면 결함 별 총 6개의 SN비 이득(SNR Gain<sub>C1</sub>, SNR Gain<sub>C2</sub>, SNR Gain<sub>C3</sub>, SNR Gain<sub>C4</sub>, SNR Gain<sub>C5</sub>, SNR Gain<sub>C6</sub>)이 계산되었으며, 이들의 총합인 총괄 SN비 이득은 <Figure 5>에서 제시하고 있다. 본 결과를 해석하자면 총괄 SN비 이득이 양의 값을 갖는 변수는 강판 표면 결함들 사이의 구분을 명확하게 하는 유용한 변수이며, 반대로 음의 값을 갖는 변수는 결함 식별에 부정적 영향을 주는 변수로 해석할 수 있다. 따라서 <Figure 6>과 <Table 7>에서 표시한 바와 같이 총괄 SN비 이득의 크기에 따라 변수 12, 1, 2, 13, 17, 15, 11, 14, 22, 10, 24, 19, 9 순서대로 강판 표면 결함 식별에 크게 기여하는 유용한 변수임을 확인하였다.

S-MTS를 이용한 강판의 표면 결함 진단

Class	Level	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	A25	A26
C1	L1	-29.6	-31.2	-29.6	-30.0	-30.6	-30.7	-31.7	-30.9	-30.9	-30.5	-30.4	-30.6	-30.6	-30.5	-30.7	-31.6	-29.3	-31.4	-29.9	-30.8	-31.3	-30.1	-31.1	-30.5	-30.5	-31.0
	L2	-31.7	-30.1	-31.6	-31.2	-30.6	-30.6	-29.6	-30.4	-30.4	-30.8	-30.9	-30.7	-30.7	-30.8	-30.6	-29.6	-31.9	-29.9	-31.3	-30.5	-29.9	-31.2	-30.2	-30.8	-30.7	-30.2
	SNR Gain <sub>C1</sub>	2.1	-1.1	2.0	1.2	0.0	-0.1	-2.1	-0.5	-0.5	0.2	0.4	0.1	0.2	0.3	0.0	-2.0	2.6	-1.4	1.4	-0.3	-1.4	1.2	-0.9	0.2	0.2	-0.8
C2	L1	-33.4	-31.6	-32.6	-33.5	-33.2	-33.3	-35.6	-33.1	-33.2	-33.5	-33.5	-32.7	-33.4	-32.7	-32.8	-33.5	-32.6	-33.1	-33.6	-33.4	-33.4	-33.0	-33.4	-33.0	-33.2	-33.5
	L2	-32.9	-34.6	-33.7	-32.7	-33.1	-32.9	-30.6	-33.1	-33.0	-32.8	-32.7	-33.5	-32.8	-33.5	-33.4	-32.7	-33.6	-33.1	-32.6	-32.8	-32.8	-33.3	-32.8	-33.2	-33.1	-32.7
	SNR Gain <sub>C2</sub>	-0.5	3.0	1.1	-0.8	-0.1	-0.5	-5.1	0.0	-0.2	-0.7	-0.8	0.8	-0.6	0.8	0.5	-0.8	1.0	0.0	-1.0	-0.6	-0.5	0.3	-0.6	0.2	-0.1	-0.9
C3	L1	-36.8	-36.1	-37.6	-37.6	-36.3	-36.9	-40.2	-36.6	-36.3	-34.9	-35.5	-36.4	-36.6	-36.4	-36.6	-36.6	-36.7	-36.6	-36.7	-36.6	-36.5	-36.6	-36.5	-36.6	-36.6	-36.6
	L2	-36.4	-37.1	-35.6	-35.5	-36.9	-36.3	-33.0	-36.5	-36.8	-38.3	-37.7	-36.7	-36.6	-36.7	-36.6	-36.6	-36.5	-36.5	-36.4	-36.5	-36.6	-36.6	-36.6	-36.5	-36.6	-36.6
	SNR Gain <sub>C3</sub>	-0.4	1.0	-2.0	-2.1	0.6	-0.6	-7.2	-0.1	0.5	3.4	2.2	0.3	0.0	0.2	0.0	0.0	-0.2	-0.1	-0.3	-0.2	0.1	0.0	0.1	-0.1	0.0	0.0
C4	L1	-165.9	-170.3	-172.3	-171.3	-171.9	-170.7	-172.9	-171.3	-171.2	-172.2	-170.3	-166.7	-166.9	-172.4	-171.0	-171.2	-171.8	-171.3	-171.4	-172.0	-171.3	-171.3	-171.2	-171.1	-171.6	-171.8
	L2	-177.2	-172.7	-170.7	-171.6	-171.2	-172.3	-170.1	-171.5	-171.8	-170.9	-172.8	-176.3	-176.1	-170.6	-172.0	-171.8	-171.3	-171.5	-171.6	-171.0	-171.7	-171.3	-171.8	-171.9	-171.4	-171.3
	SNR Gain <sub>C4</sub>	11.3	2.4	-1.5	0.1	-0.7	1.6	-2.8	0.0	0.6	-1.3	2.5	9.6	9.2	-1.8	1.0	0.6	-0.5	0.0	0.2	-1.1	0.4	-0.5	0.6	0.8	-0.1	-0.5
C5	L1	-46.4	-45.3	-46.3	-47.3	-47.8	-47.5	-46.9	-47.2	-47.0	-46.8	-46.8	-46.2	-47.6	-46.5	-46.6	-47.3	-46.2	-47.0	-46.7	-47.5	-46.9	-46.6	-46.8	-47.0	-47.5	-46.9
	L2	-47.4	-48.5	-47.6	-46.6	-46.1	-46.3	-47.0	-46.6	-46.9	-47.0	-47.1	-47.7	-46.3	-47.3	-47.3	-46.6	-47.7	-46.8	-47.1	-46.4	-47.0	-47.2	-47.0	-46.9	-46.3	-47.0
	SNR Gain <sub>C5</sub>	1.0	3.2	1.3	-0.7	-1.7	-1.2	0.1	-0.6	-0.1	0.2	0.3	1.5	-1.3	0.8	0.7	-0.7	1.5	-0.3	0.4	-1.1	0.1	0.6	0.2	-0.1	-1.2	0.1
C6	L1	-42.8	-40.4	-43.1	-43.7	-42.9	-43.0	-46.0	-42.9	-42.6	-42.8	-43.1	-41.7	-43.9	-41.7	-41.7	-42.7	-42.5	-42.6	-42.8	-43.1	-42.8	-42.7	-42.8	-42.8	-42.8	-42.8
	L2	-42.5	-44.9	-42.2	-41.6	-42.4	-42.3	-39.3	-42.4	-42.7	-42.5	-42.2	-43.6	-41.4	-43.6	-43.6	-42.6	-42.8	-42.7	-42.5	-42.2	-42.5	-42.6	-42.5	-42.5	-42.5	-42.5
	SNR Gain <sub>C6</sub>	-0.3	4.4	-0.9	-2.0	-0.5	-0.6	-6.6	-0.4	0.0	-0.4	-0.8	1.9	-2.5	2.0	1.9	-0.2	0.3	0.1	-0.2	-0.9	-0.2	-0.1	-0.4	-0.2	-0.4	-0.3
Overall	SNR Gain	13.2	12.9	0.0	-4.4	-2.5	-1.4	-23.7	-1.6	0.4	1.4	3.8	14.2	4.9	2.3	4.1	-3.1	4.7	-1.7	0.5	-4.0	-1.5	1.4	-1.0	0.8	-1.6	-2.4

〈Figure 5〉 SNR and Overall SNR Gain



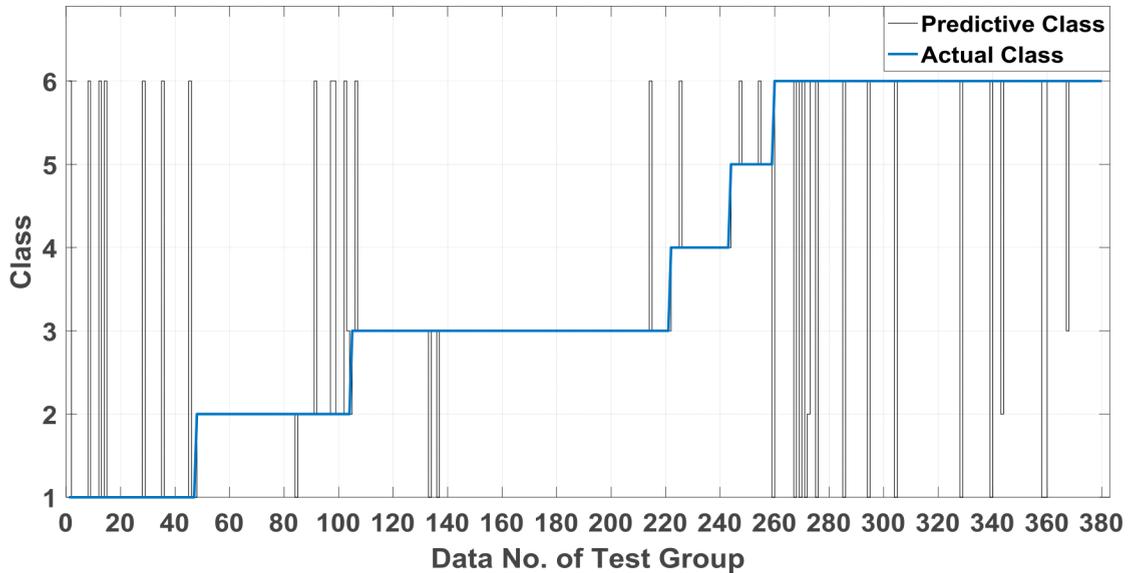
〈Figure 6〉 Overall SNR Gain of Each Variables

〈Table 7〉 List of Useful Variables

Priority	No.	Variables
1	12	Type Of Steel
2	1	X_Minimum
3	2	X_Maximum
4	13	Steel_Plate_Thickness
5	17	Outside_X_Index
6	15	Empty_Index
7	11	Length_of_Conveyer
8	14	Edges_Index
9	22	Log_X_Index
10	10	Maximum_of_Luminosity
11	24	Orientation_Index
12	19	Edges_Y_Index
13	9	Minimum_of_Luminosity

#### 4.6 유용한 변수 기반 진단

본 단계는 S-MTS의 네번째 단계로 세번째 단계에서 선정한 유용한 변수 13개의 예측능력을 검증하고자 해당 변수만을 이용해 강판 표면 결함 진단 실험을 재수행하였다. 이에 대한 결과는 <Figure 7>과 <Table 8>에서 제시한 바와 같이 13개 변수만 사용하였음에도 불구하고 정확도 측면에서 전체 26개 변수를 모두 사용하였을 때보다 5% 높은 90.79%의 정확도를 갖는 것으로 확인하였다. 이는 C1(Pastry)과 C5(Dirtiness) 데이터 간의 높은 상관관계로 인해 낮은 정확도를 보였던 C5 식별에도 적절한 변수 최적화가 이루어졌다는 것으로 해석할 수 있다.



〈Figure 7〉 Result of S-MTS using Useful Variables

〈Table 8〉 Accuracy of S-MTS using Useful Variables

Class	S-MTS using All Variables (%)	S-MTS using Useful Variables (%)
C1	85.11	85.11
C2	84.21	89.47
C3	93.16	96.58
C4	81.82	95.45
C5	18.75	81.25
C6	89.26	88.43
Average	85.79	90.79

#### 4.7 토의

강판 표면 결함 진단은 데이터마이닝에서 많은 관심이 집중되는 분야 중 하나로서 다양한 알고리즘을 도입 및 적용하는 연구가 늘고있다. 따라서 본 절에는 S-MTS 기반 강판 표면 결함 진단

시스템의 정확도와 선행 연구들에서 제시하고 있는 타 알고리즘 기반 강판 표면 결함 진단 시스템 간의 정확도를 비교하고자 한다. 비교 대상 알고리즘은 Decision Tree, Multi Perception Neural Network (MLPNN), Logistic Regression (LR), Support Vector Machine (SVM), Tree Bagger Random Forest, Grid Search (GS), Genetic Algorithm (GA), Particle Swarm Optimization (PSO)이며, <Table 9>는 선행 연구들에서 수행한 실험 결과를 기반으로 각 알고리즘의 정확도를 정리한 것이다. 표에서 보이는 바와 같이 S-MTS 기반 강판 표면 결함 진단 시스템의 정확도는 MLPNN, LR, GS, GA, PSO 대비 작게는 6%, 크게는 27% 정도 우수한 정확도를 갖고 있으며, Decision Tree와 SVM, Tree Bagger Random Forest에 비해 작게는 4%, 크게는 9% 정도 낮은 정확도를 갖는 것으로 확인되었다. 기본적으로 해외 선진 엔지니어링 사의 결함 분류 정확도가

〈Table 9〉 Comparison of Accuracy

Authors (Year)	Classification Methods	Accuracy
Fakhr et al. (2012)	Decision Tree C5.0	98.09%
	Multi Perception Neural Network (MLPNN)	79.14%
	Logistic Regression (LR)	62.99%
Jin et al. (2013)	Decision Tree	94.38%
	Multiplayer perceptron	83.87%
	Logistic Regression (LR)	72.64%
Simic et al. (2014)	Support Vector Machine (SVM)	95.89%
	TreeBagger Random Forest	99.55%
Tian et al. (2015)	Grid Search (GS)	77.7~77.8%
	Genetic Algorithm (GA)	77.2~78.0%
	Particle Swarm Optimization (PSO)	78.0~78.8%
Kim et al. (2016)	Simultaneous MTS (S-MTS)	90.79%

75~80% 수준에 불가하다는 사실을 인지한다면, S-MTS 기반 강판 표면 결함 진단 시스템은 타 알고리즘과 비교해보아도 실제 현장에 적용할 수 있을 만큼 충분한 신뢰성을 갖는 것으로 판단할 수 있다. 또한 S-MTS는 타 알고리즘과 다르게 다양한 측정변수들의 예측능력 평가를 바탕으로 변수 최적화가 가능하다는 이점이 있다. 다시 말해, 클래스 분류에 영향을 적게 미치는 변수를 찾아 이를 제거할 수 있다면 현장에 설치되는 불필요한 측정 센서의 개수를 줄일 수 있고, 이는 설비 유지보수 및 에너지 비용을 절감하는데 탁월한 효과를 기대할 수 있다.

## 5. 결론

본 연구에서는 철강 산업의 중요한 이슈인 강판 표면 결함 진단을 위해 다중 클래스 분류에

진보된 알고리즘인 S-MTS에 기반한 강판 표면 결함 진단 시스템을 제안하였다. 강판 표면 결함은 강판 품질, 가격과 직결되는 중요한 품질 지표로서 이를 관리하기 위해 육안으로 강판 표면 검사를 수행하고 있으나, 상당한 오·판단으로 인해 신뢰도에 심각한 문제를 갖고 있다. 따라서 본 연구는 S-MTS 알고리즘을 적용해 보다 높은 정확도를 가진 새로운 강판 표면 결함 진단 시스템을 고안하고자 하였다. 결함 진단은 S-MTS 절차 따라 6가지 강판 결함 별 참조 그룹 MS를 구축하고, 이를 기반으로 비교 그룹의 MD를 계산한 뒤 최소 MD를 갖는 강판 결함을 해당 비교 그룹의 강판 결함으로 식별하였다. 이로부터 나온 S-MTS 기반 강판 표면 결함 분류 정확도는 90.79%의 높은 정확도를 갖는 것으로 확인되었다. 이러한 본 연구 결과는 실제 현장 적용하는데 있어 충분한 신뢰성을 가지는 것으로 해석할 수 있다. 또한 변수 최적화를 통해 불필요한 측

정 센서 개수를 줄여 설비 유지보수 비용 및 에너지 소비 절감이 가능할 것으로 예상된다.

본 연구는 S-MTS 기반 강판 표면 결함 진단 시스템의 성능 비교를 위해 기존 문헌들의 실험 결과를 기반으로 비교 분석한 한계가 있다. S-MTS 강판 표면 결함 진단 시스템의 분류 성능을 보다 정확하게 검증하기 위해서는 동일 실험 조건에서 타 알고리즘 기반 결함 진단 시스템의 정확도를 비교할 필요가 있다. 따라서 추후 연구에서는 이러한 한계 요인을 극복한 추가 실험을 진행할 예정이며, 이와 동시에 본 연구에서 제안한 S-MTS 기반 강판 표면 결함 진단 시스템의 현장 적용을 통해 실제 효과성까지 검증함으로써 해당 결과를 기반으로 분류 정확도를 개선할 계획이다.

## 참고문헌(References)

- Ahmet, S., S. Jagannathan, C. Saygin, "Mahalanobis Taguchi System (MTS) as a Prognostics Tool for Rolling Element Bearing Failures", *Journal of Manufacturing Science and Engineering*, Vol.132, No.5(2010)
- Cha, J. M., J. Y. Kim, J. U. Shin, and C. S. Yeom, "A Method for Improving Multiclass Classification Performance of Mahalanobis Taguchi System", *Proceedings of the Korea Society of IT Service Conference*, Vol.2016, (2016), 411~414.
- Fakhr, M. and A. M. Elsayad, "Steel plates faults diagnosis with data mining models", *Journal of Computer Science*, Vol.8, No.4(2012), 506~514.
- Hong, J. E., "Diagnosis of Spondylopathy Using Mahalanobis Taguchi System", *Journal of Society of Korea Industrial and Systems Engineering*, Vol.35, No.4(2012), 10~15.
- Jin, X. and T. W. S. Chow, "Anomaly Detection of Cooling Fan and Fault Classification of Induction Motor using Mahalanobis-Taguchi System", *Expert Systems and Applications*, Vol.40, (2013), 5787~5795.
- Kim, C. H., S. H. Choi, W. J. Joo, and G. B. Kim, "Classification of Surface Defect on Steel Strip by KNN Classifier", *Journal of the Korean Society for Precision Engineering*, Vol.23, No.8(2006), 80~88.
- Moon, C. I., S. H. Choi, W. J. Joo, G. B. Kim, and H. K. Kim, "Development of a Neural Network Classifier for the Classification of Surface Defects of Cold Rolled Strips", *Journal of the Korean Society for Precision Engineering*, Vol.24, No.4(2007), 76~83.
- Park, S. G., W. S. Park, Y. Y. Lee, D. S. Kim, and J. E. Oh, "A Fault Diagnosis on the Rotating Machinery Using MTS", *Transactions of the Korean Society for Noise and Vibration Engineering*, Vol.18, No.6(2008), 619~623.
- Ren, J., Y., Cai, X., Xing, and J., Chen, "A method of multi-class faults classification based-on Mahalanobis-Taguchi system using vibration signals", *Proceedings of 9<sup>th</sup> IEEE International Conference on Reliability, Maintainability and Safety (ICRMS)*, Vol.2011, (2011), 1015~1020.
- Semeion, *Steel Plates Faults Diagnosis Dataset*, UCI Repository of machine learning databases, Irvine, CA: University of California, Department of Information and Computer Science, 2016. Available at <https://archive.ics.uci.edu/ml/datasets/Steel>

- Plates Faults (Downloaded 1 September, 2016).
- Simić, D., V. Svirčević, and S. Simić, “An Approach of Steel Plates Fault Diagnosis in Multiple Classes Decision Making”, *Hybrid Artificial Intelligence Systems*, Vol.8480, (2014), 86~97.
- Song, S. J., H. J. Kim, S. H. Choi, and J. H. Lee, “Classification of Surface Defects on Cold Rolled Strips by Probabilistic Neural Networks”, *Journal of the Korean Society for Nondestructive Testing*, Vol.17, No.3(1997), 162~173.
- Su, C.T. and Y.H., Hsiao, “Multiclass MTS for simultaneous feature selection and classification”, *IEEE Transactions on Knowledge and Data Engineering*, Vol.21, No.2(2009), 192~205.
- Taguchi, G., and R. Jugulum, *The Mahalanobis-Taguchi Strategy: A Pattern Technology System*, John Wiley & Sons, New York, 2002.
- Tian, Y., M. Fu, and F. Wu, “Steel plates fault diagnosis on the basis of support vector machines”, *Neurocomputing*, Vol.151, (2015), 296~303.

## Abstract

## Steel Plate Faults Diagnosis with S-MTS\*

Joon-Young Kim\*\* · Jae-Min Cha\*\*\* · Junguk Shin\*\*\* · Choongsub Yeom\*\*\*

Steel plate faults is one of important factors to affect the quality and price of the steel plates. So far many steelmakers generally have used visual inspection method that could be based on an inspector's intuition or experience. Specifically, the inspector checks the steel plate faults by looking the surface of the steel plates. However, the accuracy of this method is critically low that it can cause errors above 30% in judgment. Therefore, accurate steel plate faults diagnosis system has been continuously required in the industry. In order to meet the needs, this study proposed a new steel plate faults diagnosis system using Simultaneous MTS (S-MTS), which is an advanced Mahalanobis Taguchi System (MTS) algorithm, to classify various surface defects of the steel plates. MTS has generally been used to solve binary classification problems in various fields, but MTS was not used for multiclass classification due to its low accuracy. The reason is that only one mahalanobis space is established in the MTS. In contrast, S-MTS is suitable for multi-class classification. That is, S-MTS establishes individual mahalanobis space for each class. 'Simultaneous' implies comparing mahalanobis distances at the same time. The proposed steel plate faults diagnosis system was developed in four main stages. In the first stage, after various reference groups and related variables are defined, data of the steel plate faults is collected and used to establish the individual mahalanobis space per the reference groups and construct the full measurement scale. In the second stage, the mahalanobis distances of test groups is calculated based on the established mahalanobis spaces of the reference groups. Then, appropriateness of the spaces is verified by examining the separability of the mahalanobis diatances. In the third stage, orthogonal arrays and Signal-to-Noise (SN) ratio of dynamic type are applied for variable optimization. Also, Overall SN ratio gain is derived from the SN

---

\* This work was partly supported by the Energy Technology Development Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, and Republic Korea (No. 2014520400090). And it was also partly supported by the Industrial Core Technology Development Program, (No. 10063187) funded by the Ministry of Trade, Industry & Energy (MI, Korea).

\*\* Corresponding Author: Joon-Young Kim

Plant SE Team, Institute for Advanced Engineering (IAE)

175-28, Goan-ro 51beon-gil, Baegam-myeon, Cheoin-gu, Yongin-si, Gyeonggi-do, 17180, Republic of Korea

Tel: +82-10-9829-2010, Fax: +82-31-330-7850, E-mail: joonykim@outlook.com

\*\*\* Plant SE Team, Institute for Advanced Engineering (IAE)

ratio and SN ratio gain. If the derived overall SN ratio gain is negative, it means that the variable should be removed. However, the variable with the positive gain may be considered as worth keeping. Finally, in the fourth stage, the measurement scale that is composed of selected useful variables is reconstructed. Next, an experimental test should be implemented to verify the ability of multi-class classification and thus the accuracy of the classification is acquired. If the accuracy is acceptable, this diagnosis system can be used for future applications. Also, this study compared the accuracy of the proposed steel plate faults diagnosis system with that of other popular classification algorithms including Decision Tree, Multi Perception Neural Network (MLPNN), Logistic Regression (LR), Support Vector Machine (SVM), Tree Bagger Random Forest, Grid Search (GS), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The steel plates faults dataset used in the study is taken from the University of California at Irvine (UCI) machine learning repository. As a result, the proposed steel plate faults diagnosis system based on S-MTS shows 90.79% of classification accuracy. The accuracy of the proposed diagnosis system is 6-27% higher than MLPNN, LR, GS, GA and PSO. Based on the fact that the accuracy of commercial systems is only about 75-80%, it means that the proposed system has enough classification performance to be applied in the industry. In addition, the proposed system can reduce the number of measurement sensors that are installed in the fields because of variable optimization process. These results show that the proposed system not only can have a good ability on the steel plate faults diagnosis but also reduce operation and maintenance cost. For our future work, it will be applied in the fields to validate actual effectiveness of the proposed system and plan to improve the accuracy based on the results.

**Key Words** : Big Data, Multiclass Classification, Simultaneous MTS (S-MTS), Mahalanobis Taguchi System (MTS), Steel Plates Faults Diagnosis

Received : November 17, 2016   Revised : December 22, 2016   Accepted : December 25, 2016  
Publication Type : Regular Paper   Corresponding Author : Joon-Young Kim

## 저자 소개



**김준영**

포항공과대학교 플랜트시스템엔지니어링학과에서 석사 학위를 마쳤으며, 현재 고등기술연구원 플랜트 시스템엔지니어링(SE) 부서의 연구원으로 재직 중이다. 주요 연구분야는 빅데이터 기반 플랜트 설비 진단 및 예측, 통계분석, 데이터 마이닝 기법과 응용, 시스템엔지니어링 기반 플랜트 FEED Package 표준화 및 개발, 요구 기반/모델 기반 플랜트 시스템 아키텍처 설계 등이다.



**차재민**

포항공과대학교 산업경영공학과에서 석박사 학위를 마쳤으며, 현재 고등기술연구원 플랜트 시스템엔지니어링(SE) 부서의 선임연구원으로 재직 중이다. 주요 연구분야는 플랜트 특성을 고려한 데이터마이닝 기법 개발 및 적용, 데이터마이닝 기반 플랜트 설비 진단/예측, 모델 기반 시스템 엔지니어링 (MBSE) 등이다.



**신중욱**

광운대학교 전파공학과에서 석사 학위를 마쳤으며, 현재 고등기술연구원 플랜트시스템엔지니어링(SE) 부서의 선임연구원으로 재직 중이다. 주요 연구분야는 시스템엔지니어링에 기반 플랜트 FEED 개발, 모델 기반 플랜트 시스템 아키텍처 설계, 플랜트 설비 진단 및 예측 등이다.



**염충섭**

고등기술연구원 플랜트 시스템엔지니어링(SE) 부서를 맡고 있으며, 아주대학교 시스템공학과에서 박사학위를 받았다. 국가 미래성장동력 추진단장(초임계 CO<sub>2</sub> 발전시스템)을 맡고 있으며, 플랜트 생애주기에 걸친 시스템엔지니어링 적용 연구에 관심을 가지고 있다.