

Diagnosis of Spondylopathy Using Mahalanobis Taguchi System

Jung Eui Hong[†]

Department of Industrial and Management Engineering, Korea National University of Transportation

Mahalanobis Taguchi System을 이용한 척추질환 환자의 진단에 관한 연구

홍 정 의[†]

한국교통대학교 공과대학 산업경영공학과

The Mahalanobis-Taguchi System is a diagnosis and predictive method for analyzing patterns in multivariate cases. The goal of this study is diagnosis of the spondylolisthesis from biomedical data that is derived from the shape and orientation of the pelvis and lumbar spine. The data set has six attributes including pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and grade of spondylolisthesis and two class including normal and abnormal. From University of California at Irvine machine learning repository, 100 normal and 150 spondylolisthesis patient's data were used for this study. Mahalanobis Taguchi System (MTS) application process and the diagnosis results were described in this paper.

Keywords : MTS(Mahalanobis Taguchi System), Spondylolisthesis, Diagnosis

1. 서 론

인간의 질병을 보다 효과적이고 경제적으로 진단하기 위해서는 질병의 진단에 꼭 필요한 검사항목을 설정하고 이를 정확하게 분석하는 일이 무엇보다 중요하다. 이를 위해서는 의료 측정 데이터가 가지는 의미를 보다 정확히 분석하고 이해함으로써 보다 정확한 진단에 접근 할 수 있다. 본 연구는 그동안의 자동차의 브레이크 설계, 유방암의 진단 그리고 음성분석통한 파킨슨병의 진단 등 다수의 변수를 가지고 변수간의 상호작용을 배제 할 수 없는

비선형 다변량 시스템의 해석을 통해 얻은 연구 결과를 바탕으로 척추 질환 환자의 골반과 요추의 형태와 구조에 대한 데이터를 이용하여 척추전방 전위 증의 진단에 적용하여 그 진단의 정확성 여부를 평가 하고자 한다.

대부분의 다 변량 시스템은 변수들의 변화에 따른 측정값의 변화를 얼마나 정확히 예측하는 것이 중요하다. 그러나 일반적으로 이러한 해석을 위해서는 변수의 수와 변수 값의 형태에 따라 많은 양의 데이터가 필요로 할 수 있다. 의료분야에서 질병의 진단을 위한 데이터의 획득은 검사비용과 시간 그리고 검사를 위한 환자의 고통을 고려하지 않을 수 없다. 따라서 검사 항목 중에서 질병의 진단에 반드시 필요한 검사항목만을 선정한다면 진단을 위한 시간과 경제적인 비용을 낮출 수 있을 것이다.

시상면 상에 정렬된 인간의 척추에는 일련의 교대적 만곡(Curvature)이 나타난다. 경부 만곡, 흉부만곡, 요부 만곡 그리고 천미만곡이 그것으로 모두 생리적인 만곡이다. 이러한 만곡은 사람이 바로 설 때 몸의 중력평형을 유지하기 위하여 이루어진 것이다[1]. 특히 시상 면에서

Received 24 May 2012; Accepted 28 October 2012

[†] Corresponding Author : jehong@ut.ac.kr

© 2012 Society of Korea Industrial and Systems Engineering

This is Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited(<http://creativecommons.org/licenses/by-nc/3.0>).

의 척추 정렬은 가장 중요한 요소이면 정상적인 만곡은 체중을 고르게 분포시키며 각 관절에 대한 부담을 극소화 시켜 무리가 가지 않도록 조절하고 연부조직에 대한 긴장과 변형을 방지하여 좋은 자세를 유지 하게 된다. 본 연구는 골반지표(Pelvic parameter) 측정치를 이용하여 이들 측정치가 척추전방 전위 증에 미치는 영향을 MTS를 이용하여 분석하여 진단 척추 질병의 진단에 활용하고자 한다.

Rocha Neto[2]는 신경망 이론과 Support Vector Machine을 이용하여 척추전방전위 증의 진단에 대한 연구를 진행 하였다. 알고리즘의 특징상 전체 데이터의 40%와 80%를 학습에 활용하였다.

2. Mahalanobis Taguchi System

마할라노비스 거리(Mahalanobis Distance)는 변수들의 표준편차와 상관관계를 고려하여 두 지점의 거리를 나타내는 방법이다. 따라서 모든 변수들이 표준화(generalization)되어 있고 각각 독립적이라면 마할라노비스 거리는 유클리디안 거리와 일치한다. 강건 설계 방법을 고안해 낸 다구찌는 임의로 선택된 정상 집단의 마할라노비스 거리를 기초로 마할라노비스 평면(Mahalanobis Space)을 설정하고 이평면을 구성하는 변수들의 상관관계를 이용하여 새로운 측정값의 마할라노비스 거리를 계산하여 정상 집단과 비정상집단을 판단할 뿐만 아니라 변수들이 시스템에 미치는 영향까지 분석하는 마할라노비스 다구찌 시스템(Mahalanobis Taguchi System)[3]방법을 고안해 냈다.

MTS 방법을 이용한 정확한 진단을 위해서는 변수들의 상관관계 매트릭스를 구성할 기준 MS 공간을 설정하는 것이 중요하다. 여러 개의 변수를 가진 다차원 공간에서의 관측된 패턴은 변수들 간의 상호 상관관계에 영향을 받으며 종종 이러한 상관관계를 무시한다면 잘못된 해석을 내릴 수 있다.

MTS에서 MS는 선정된 정상그룹과 비정상 그룹의 데이터를 표준화하여 두 그룹을 구별 할 수 있는 기준 공간으로 이를 이용하여 정상 그룹과 비정상 그룹의 MD 거리 값이 분명하게 구분되어 진다면 설정된 MS는 유효한 것으로 판단할 수 있다. MS가 구해지면 변수가 시스템에 영향을 미치는 정도를 판단하기 위하여 S/N비(Signal to Noise Ratio)와 OA(Orthogonal Array)를 이용한다.

MTS의 일반적인 적용절차는

첫째, 표준이 되는 정상 집단으로부터 판단에 적용될 데이터를 선정한다. 모든 데이터는 식 (1)을 이용하여 일

반화 과정을 통해 변환된다. 다음으로 선정된 데이터와 변수들의 상관관계를 이용하여 MS를 구성한다. 이때 사용된 정상 데이터에 의해 계산된 마할라노비스 거리의 평균값은 1에 근접한다.

$$Z_i = \frac{X_i - m}{\sigma} \quad (1)$$

여기서 m 은 변수의 평균 값이고, σ 는 표준편차 그리고 X_i 는 임의 측정값이다.

다차원 공간에서의 마할라노비스 거리는 변수들 간의 상관관계를 계산함으로써 구할 수 있다. 이러한 MD의 통계학적인 의미는 임의의 측정값이 선택 집단의 중간값으로부터 얼마나 근접해 있는가를 의미한다. 아래의 공식은 마할라노비스 거리를 계산하는 공식이다.

$$MD_j = D_j^2 = \frac{1}{k} Z_{ij}^T C^{-1} Z_{ij} \quad (2)$$

여기서 C^{-1} 은 변수들 간의 상관계수를 포함한 상관행렬의 역행렬이고 Z_{ij} 는 일반화된 데이터 값이다. T는 표준 벡터의 transpose 벡터이다.

둘째, 이렇게 구해진 MS 공간의 유효성을 판단한다. 이를 위해서 MS 공간 밖의 측정값 즉 비정상 또는 건강하지 않은 데이터를 이용한다. 이러한 집단의 MD를 구하기 위해서 정상 또는 건강한 집단의 기준 MS를 구성하는 평균, 표준편차 그리고 상관행렬을 이용한다. 기준 MS 공간이 유효하다면 비정상 또는 건강하지 않은 집단의 MD 값은 정상 또는 건강한 집단의 MD 값보다 훨씬 커서 구별이 뚜렷할 것이다.

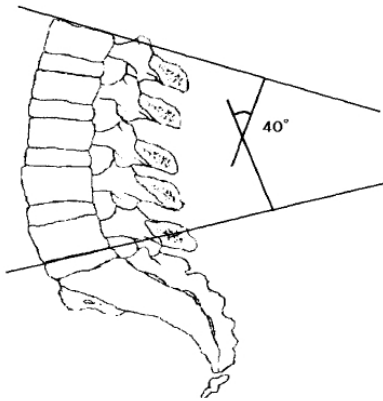
셋째, 변수 중에서 측정값에 영향을 미치지 않거나 적게 미치는 변수를 찾아내서 제거하여 시스템의 해석을 쉽게 하는 일이다. 이러한 목적을 위해서는 직교 배열표와 SN비가 유용하게 활용되어 질 수 있다. 직교배열표의 열은 실험변수들을 배열 하였고 행은 실험의 조합을 나타낸다. 즉 직교 배열표의 Level 1은 변수를 사용하는 경우를 의미 하고 Level 2는 변수를 사용하지 않는 경우를 의미 한다. 따라서 직교배열표의 조건에 따라 변수들은 사용되어지거나 무시될 수 있으며 이를 바탕으로 SN비를 계산 할 수 있다. 망대특성의 SN비를 구하는 식은

$$SN \text{ ratio} = \tau = -10 \log \left(\frac{1}{t} \sum_{t=1}^t \frac{1}{D_i^2} \right) \quad (3)$$

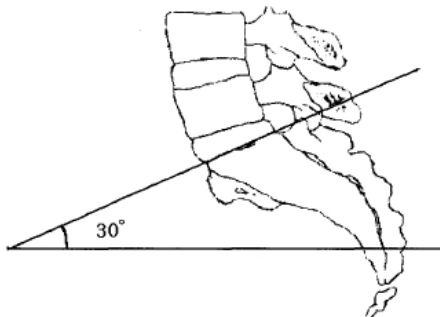
여기서 τ 는 S/N ratio이고 t는 비정상 그룹의 개수이며, D_i^2 는 i번째의 마할라노비스 거리(MD)이다.

3. MTS를 이용한 척추전방 전위증(Spondylolisthesis)의 진단

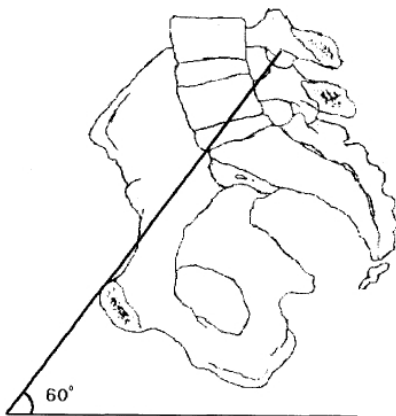
인체의 방사선 사진에서 측정 할 수 있는 골반 지표 (Pelvic Parameter) 들에는 변하지 않는 해부학적 지표와 위치에 따라 변하는 지표로 나눌 수 있는데 골반입사각 (Pelvic Incidence)는 전자에 속하고 천추 경사(Sacral Slope)와 골반 경사(Pelvic Tilt)는 후자에 속한다[5].



<Figure 1> Lumbar Lordosis Angle



<Figure 2> Sacral Slope



<Figure 3> Pelvic Tilt

3.1 측정 변수

본 연구를 위한 척추에 관한 생의학 데이터는 UCI machine learning depository의 Vertebral Column 데이터를 이용하였으며, 사용된 데이터는 100명의 정상인과 150명의 척추전방 전위증 환자의 척추관련 데이터를 <Table 1> 같은 변수로 구성되어 있다.

<Table 1> Variables of Data[2]

	Variable
A1	Pelvic Incidence
A2	Pelvic Tilt
A3	Lumbar Lordosis Angle
A4	Sacral Slope
A5	Pelvic Radius
A6	Level of Slipping

골반 입사각(Pelvic Incidence)은 시상면 균형에 중요한 척추-골반 지표이며 시상면상 만곡의 모양을 결정하고 균형을 조절하는 가장 중요한 지표로 인식되고 있다.

<Table 2> Normal Group's Data

A1	A2	A3	A4	A5	A6
36.42	13.88	20.24	22.54	126.08	0.18
51.08	14.21	35.95	36.87	115.8	6.91
34.76	2.63	29.5	32.12	127.14	-0.46
48.9	5.59	55.5	43.32	137.11	19.85
46.24	10.06	37	36.17	128.06	-5.1
46.43	6.62	48.1	39.81	130.35	2.45
39.66	16.21	36.67	23.45	131.92	-4.97
45.58	18.76	33.77	26.82	116.8	3.13
66.51	20.9	31.73	45.61	128.9	1.52
82.91	29.89	58.25	53.01	110.71	6.08
50.68	6.46	35	44.22	116.59	-0.21
89.01	26.08	69.02	62.94	111.48	6.06
54.6	21.49	29.36	33.11	118.34	-1.47
34.38	2.06	32.39	32.32	128.3	-3.37
45.08	12.31	44.58	32.77	147.89	-8.94
47.9	13.62	36	34.29	117.45	-4.25
53.94	20.72	29.22	33.22	114.37	-0.42
61.45	22.69	46.17	38.75	125.67	-2.71
45.25	8.69	41.58	36.56	118.55	0.21
33.84	5.07	36.64	28.77	123.95	-0.2

골반 경사각(Pelvic Tilt)은 제 1천추 상연의 중심에서 고관절 축을 잇는 중심을 지나는 수직선과의 각으로 측정하며 요추 전만각(Lumbar Lordosis Angle)은 제 1요추에의 상연을 그은 선과 제 5요추체의 상연이 만나는 각으로 10도에서 60도까지의 분포를 보인다. 천추 경사각(Sacral Slope)은 제 1천추의 상연과 수평선과의 각으로 결정된다. Pelvic Radius는 골반반지름을 말하고 Level of Slipping은 제 5요추와 천골의 판이 미끄러진 정도를 나타낸다. 다음은 진단을 위한 Mahalanobis Space 설정을 위해 정상인 그룹에서 임의로 20명 환자를 선택하였고 척추전방 전위증 환자의 그룹에서 10명의 데이터를 선택하여 비정상 그룹(Abnormal)으로 선택하였다.

<Table 3> Abnormal group's Data

A1	A2	A3	A4	A5	A6
63.36	20.02	67.5	43.34	131	37.56
67.51	33.28	96.28	34.24	145.6	88.3
76.31	41.93	93.28	34.38	132.27	101.22
73.64	9.71	63	63.92	98.73	26.98
56.54	14.38	44.99	42.16	101.72	25.77
80.11	33.94	85.1	46.17	125.59	100.29
95.48	46.55	59	48.93	96.68	77.28
74.09	18.82	76.03	55.27	128.41	73.39
87.68	20.37	93.82	67.31	120.94	76.73
48.26	16.42	36.33	31.84	94.88	28.34

3.2 MTS의 적용

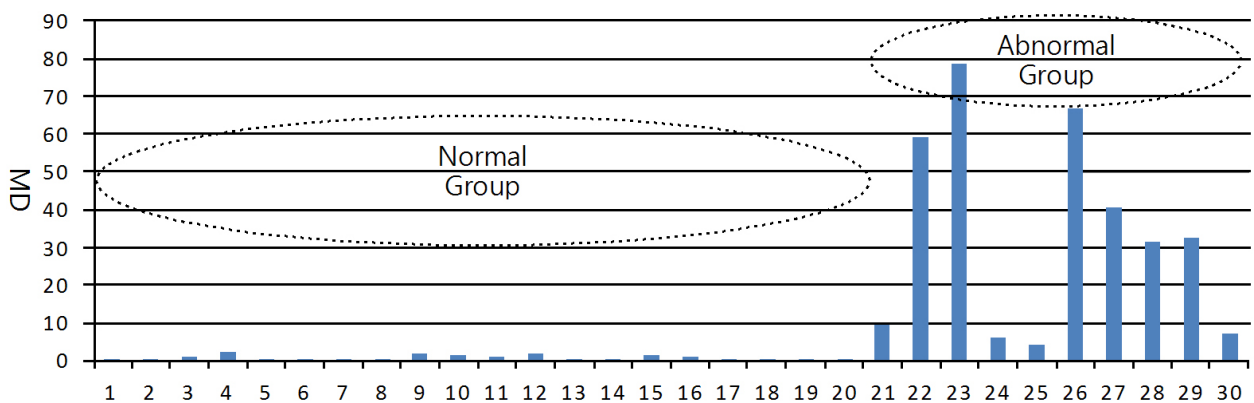
선정된 정상그룹으로부터 Generalization과정을 걸쳐서 각각의 변수들의 상관계수를 계산하고 이를 이용해서 정상 그룹의 MD 값을 계산하였다. <Table 2> 정상그룹을 <Eq. 1>을 이용하여 계산한 상관행렬은 <Table 4>와 같다.

<Table 4> Correlation Matrix

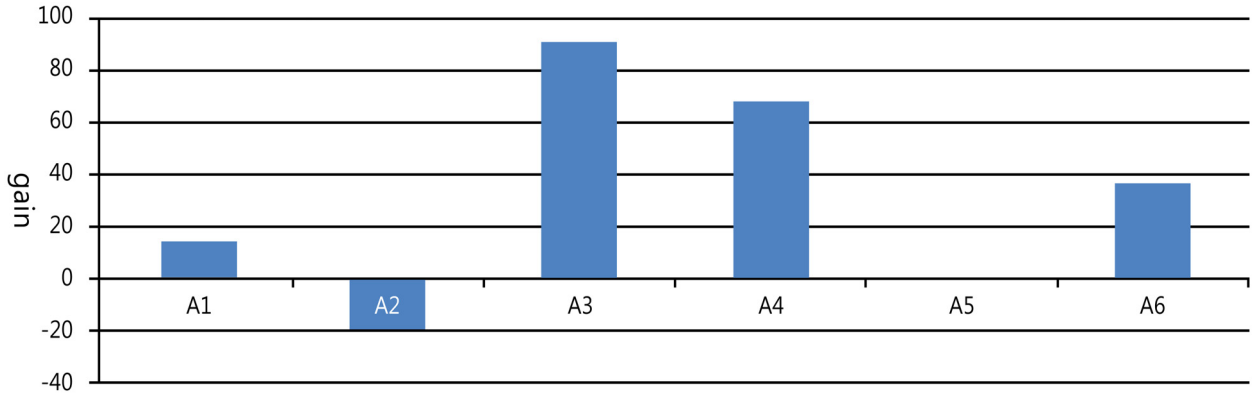
	A1	A2	A3	A4	A5	A6
A1	1.00	0.80	0.67	0.87	-0.47	0.33
A2	0.80	1.00	0.30	0.39	-0.46	0.05
A3	0.67	0.30	1.00	0.78	-0.06	0.46
A4	0.87	0.39	0.78	1.00	-0.34	0.46
A5	-0.47	-0.46	-0.06	-0.34	1.00	-0.19
A6	0.33	0.05	0.46	0.46	-0.19	1.00

일반적인 MTS 절차에 따라 정상집단으로부터 상관행렬과 역행렬 함수를 얻을 수 있고 이를 바탕으로 각각의 데이터에 대한 MD의 값을 계산 할 수 있다 20명 정상인 데이터의 MD의 평균값을 계산한 결과 평균 MD = 0.96 이었고 이는 정상 집단의 ϵ MD 값은 1에 근접함으로 적절한 기준집단 선정이 이루어 졌다고 판단 할 수 있다. 또한 정상그룹을 Mahalanobis Space(MS)를 기준으로 계산한 척추전방 전위증 환자의 MD 값을 계산한 결과 그 평균값은 37.2로 정상 환자의 MD 값으로부터 구별됨을 알 수 있다. 따라서 선택된 정상그룹은 환자의 진단에 유효할 것으로 판단된다. <Figure 4>는 정상인 그룹의 데이터에 의해 설정된 MS를 기준으로 계산된 비정상 그룹의 MD 값을 비교하여 도식화 한 그래프이다. 이러한 모든 계산과정은 Matlab 소프트웨어를 이용하여 계산 하였다. 정상그룹의 MD의 최대값은 2.9를 나타냈고 비정상 그룹 즉 척추전방 전위증 환자의 MD의 최소값은 4.85였다. 따라서 선택된 정상그룹을 이용하여 구성된 MS는 정상그룹과 비정상 그룹을 구분 가능하므로 효과적으로 척추전방 전위증 환자의 진단에 활용될 수 있을 것으로 판단 할 수 있다.

다음은 정상그룹을 기준으로 설정된 MS를 이용하여 250건의 전 데이터에 대하여 진단을 실시하였다. 즉 선택된 정상그룹의 데이터로부터 생성된 상관행렬을 이용



<Figure 4> Comparison of MD



<Figure 5> The Gain of Variables

하여 전 측정 데이터의 MD 값을 계산하고 최고의 진단 성공률을 나타내는 threshold value를 설정하여 진단을 실시하였다. 정상그룹과 척추전방 전위증을 구별하는 최적의 threshold value는 Simulation 결과 3으로 나타났다. 진단된 결과는 정상인을 척추전방 전위증 환자로 진단한 Type 1 error와 척추전방 전위 전환자를 정상으로 진단하는 Type 2 error의 수 그리고 정확히 진단한 진단율은 다음과 같다.

$$\text{Gain} = \frac{\text{SN비}_{\text{해당변수를 사용한 경우}}}{\text{SN비}_{\text{해당변수를 사용하지 않은 경우}}} \quad (4)$$

gain은 해당 변수를 사용할 경우 MD 값의 SN비와 사용하지 않을 경우의 SN비의 차이로 정의 할 수 있다. 식 (4)에서 gain 값이 클수록 해당 변수는 마할라노비스 거리 (MD)에 긍정적인 효과를 미치는 것으로 판단 할 수 있다. 따라서 큰 gain 값을 가지는 변수는 척추전방 전위 증에 매우 긍정적인 영향을 미치는 것으로 볼 수 있으며 음의 gain 값을 가지는 변수는 척추전방 전위증병의 진단에 큰 영향을 미치지 못하는 변수로 판단 할 수 있다. 즉 해당 변수가 척추전방 전위 증의 진단에 큰 영향을 미치지 않는다면 추후 진단을 위한 데이터를 확보를 위한 시간적 경제적 비용을 줄이기 위해 해당변수는 진단에서 제외 할 수도 있을 것이다. <Figure 5>는 변수에 따른 gain 값을 나타내고 있다. <Figure 5>에서 음의 gain값을 나타낸 변수 A2(Pelvic Tilt)를 변수 항목에서 제외하고 5개의 변수만을 가지고 정상그룹과 비정상 그룹의 MD 값을 비교하였다. <Figure 6>은 A2 변수를 제거한 후 나머지 변수만을 가지고도 정상그룹과 비정상 그룹이 분명히 구분됨을 알 수 있다. 마지막으로 5개의 변수만을 사용하여 250명 환자에 대한 척추전방 전위 증의 진단을 실시하여 모든 변수를 사용하여 진단한 결과와 비교하였다. <Table 6>에서와 같이 하나의 변수를 진단에서 제외한 결과가 오히려 모든 변수를 사용하여 진단한 결과 값보다 높게 나타남을 알 수 있다.

<Table 5> Test Result

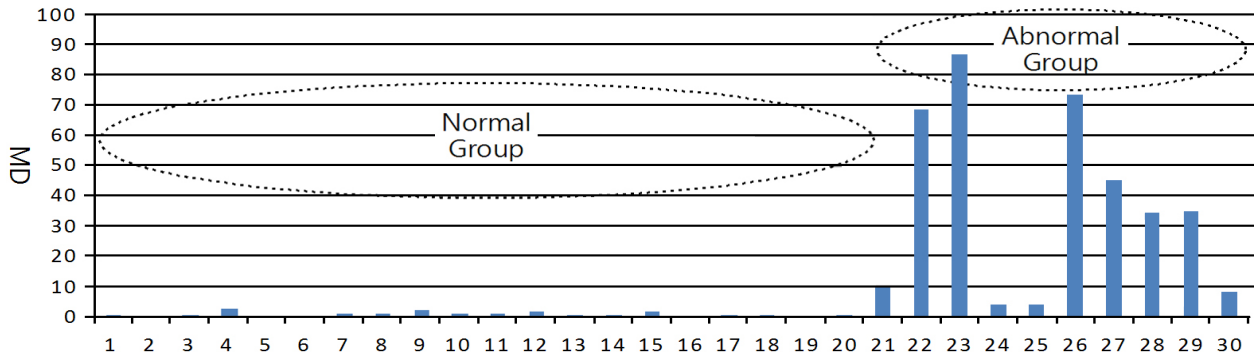
	Error
Type 1 error	6(2.4%)
Type 2 error	5(2%)
Correct %	95.60%

3.3 유효 변수의 선정

다음으로 직교 L8 직교 배열표와 망대특성의 SN비를 이용하여 각각의 변수가 진단에 미치는 중요도를 분석하였다[5]. 6개의 변수를 적절하게 배치하기 위해서는 L8 직교 배열표(OA)를 사용하였으며 망대특성의 SN비를 사용하였다(<Eq.3> 참조). 직교 배열 표에서 Level 1은 해당 변수를 사용한 경우로 Level 2는 해당 변수를 사용하지 않은 경우로 하여 각각의 변수를 사용했을 경우와 사용하지 않았을 경우의 MD값의 변화를 이용하여 SN비를 계산 하였다. 각각의 변수들이 MD값에 미치는 영향을 평가하기 위해 정상 집단 보다 MD의 변화량이 큰 비정상 집단을 이용하여 직교 배열표의 조합에 따라 MD 값을 계산하였으며 이를 근거로 각각의 변수에 대한 SN비를 계산하였다.

<Table 6> Test Results after Optimization

	Using 6 Variables	Using 5 Variables
Type 1 error	6(2.5%)	6(2.5%)
Type 2 error	5(2%)	4(1.5%)
Correct %	95.60%	96%



<Figure 6> Comparison of MD after Using 5 Variables

4. 결론

MTS 기법은 새로운 데이터 해석방법(Data Analytic Method) 으로 암과 같은 질병의 진단, 문자 인식, 음성인식, 최적화 등의 분야에서 활발한 연구가 이루어지고 있고 성공적인 적용사례도 여러 논문을 통해 발표되고 있다[6, 7, 8]. 특히 정량적인 변수 값과 정성적인 Class값을 가지는 의학 분야의 진단을 위한 MTS 기법은 적절한 정상집단 및 비정상 집단의 선정을 통해 정상인과 환자를 구별할 수 있는 Mahalanobis Space만 만들 수 있다면 진단에 필요한 의학적인 전문 지식 없이도 효과적으로 질병의 진단에 활용 할 수 있다.

본 연구는 UCI Machine Learning Repository에 공개되어 있는 데이터를 활용하였다. 척추전방 전위 증에 영향을 미치는 6개의 척추의 구조관련 정량적인 데이터와 환자유무를 나타내는 정성적 데이터로 구분되어 있고 100개의 환자군 그리고 150개의 정상인군으로 나누어져 있다.

본 연구는 20명의 정상그룹과 10명의 환자그룹을 임의로 선별하여 진단을 위한 MS를 구성하였고 이를 통해 250명의 환자를 진단한 결과 약 96%의 진단 정확도를 얻을 수 있었다. 또한 각각의 변수들이 진단에 미치는 영향을 분석하여 진단에 영향을 미치지 않는 것으로 판단된 A2(Pelvic Tilt)를 변수에서 제외한 경우 진단의 성공률이 오히려 높아진 것으로 나타났다.

Acknowledgement

본 연구는 한국교통대학교 학술연구비의 지원을 받아 수행된 연구임.

References

[1] Jeong, D.-U., Yeo K.-C., Yoon, I.-A., Kang, H.-S., and Moon, S.-I., A clinical Study on Correlation between Cervical, Lumbar Lordosis and Low back Pain. *The*

Journal of Korean Acupunctual and Moxibustion Society, 2009, Vol. 26, No. 2, p 15-23.

[2] Rocha Neto, A.R., Barreto, G.A., and Cardoso, J.S., Diagnostic of Pathology on the Vertebral Column with Embedded Reject Option. Proceedings of the 5th Conference on Pattern Recognition and Image Analysis, Gran Canaria, Spain. *Lecture Note on Computer Science*, Vol. 6669, p 588-595.

[3] Taguchi, G. and Jugulum, R., *The Mahalanobis-Taguchi Strategy : A Pattern Technology System*, John Wiley and Sons, Inc., 2002.

[4] Taguchi, G. and Jugulum, R., New Trends in Multivariate Diagnosis. *Indian Journal of Statistics*, 2000, Vol. 62, Series B, No. 2, p 233-248.

[5] Cudney, E.A., Hong, J.-E., Rajesh Jugulum, Kioumars Paryani, Kenneth M. Ragsdell, and Genishi Taguchi, An Evaluation of Mahalanobis Taguchi System and Neural Network for Multivariate Pattern Recognition. *Journal of Industrial and System Engineering*, 2007, Vol. 1, No. 2, p 139-150.

[6] Kwon, H.J., Kim, M.J., and Choi, Y.D., The Influence of Lumbosacral Angle, Pelvic level and Symptoms by Standing Lumbar Traction on HIVD Patients. *The Journal of Korean Academy of Orthopaedic manual Physical Therphy*, 1999, Vol. 5, No. 1, p 5-16.

[7] Hong, J.E., Diagnosis of Parkinson's Disease by Voice Disorder Using Mahalanobis Taguchi System. *J. of Society of Korean Industrial and System Engr.*, 2009, Vol. 32, No. 4, p 215-222.

[8] Hong, J.E. and Cudney, E.A., Taguchi, G., Jugulum, R., Paryani, K., and Ragsdell K., *A Comparison study of Mahalanobis Taguchi System and nueral network for multivariate pattern recognition*. 2005 ASME IMECE Proceedings, Orlando FL.

[9] UCI Machine Learning Repository www.ics.uci.edu/~mllearn/.