

### 계층형 신경회로망을 이용한 염색체 핵형 분류

장 용훈<sup>\*</sup>, 이 영진<sup>\*\*</sup>, 이 권순<sup>\*\*</sup>

<sup>\*</sup>동주대학 전산정보처리과, <sup>\*\*</sup> 동아대학교 전기공학과

## Karyotype Classification of Chromosome Using the Hierarchical Neu

Yong Hoon Chang<sup>\*</sup>, Young Jin Lee<sup>\*\*</sup>, Kwon Soon Lee<sup>\*\*</sup>  
Dongju College<sup>\*</sup>, Dong-A University<sup>\*\*</sup>

**Abstract** - The human chromosome analysis is widely used to diagnose genetic disease and various congenital anomalies. Many researches on automated chromosome karyotype analysis have been carried out, some of which produced commercial systems. However, there still remains much room for improving the accuracy of chromosome classification. In this paper, We proposed an optimal pattern classifier by neural network to improve the accuracy of chromosome classification. The proposed pattern classifier was built up of two-step multi-layer neural network(TMANN). We reconstructed chromosome image to improve the chromosome classification accuracy and extracted four morphological features parameters such as centromeric index(C.I.), relative length ratio(R.L.), relative area ratio(R.A.) and chromosome length(C.L.). These Parameters employed as input in neural network by preprocessing twenty human chromosome images. The experiment results shown that the chromosome classification error was reduced much more than that of the other classification methods.

### 1. 서 론

염색체(chromosome)의 세포유전학적인 해석은 인간의 유전학적 진단 및 동·식물에 관한 유전학적 연구를 위하여 널리 사용되고 있다. 염색체의 분석은 핵형 분석(karyotype analysis)을 통하여 이루어지고 있으며 임상 의학 및 세포 유전학적 연구 분야 등에서 다양한 선천적 염색체성 유전질환 및 백혈병, 악성 종양의 진단 그리고 생물학적 연구 등을 위하여 매우 중요한 실험으로 알려져 있다. 특히 태아의 핵형 분석은 태아의 염색체의 이상을 조기에 판별하는 산전 유전 진단을 위한 필수적인 검사 방법으로 알려져 있다[1].

컴퓨터를 이용한 염색체 자동 분류에 관한 연구가 1964년 Ledly에 의하여 최초로 수행되어 임상에서 사용할 수 있음을 제시하였다. 1989년 Piper 등은 염색체 영상에서 28개의 형태학적 특징 파라메타를 추출하여 패턴 분류를 위하여 통계적 패턴 분류 방법인 maximum likelihood방법을 사용하여 염색체를 자동 분류하는 연구를 수행하였으며, 1990년 Erik 등은 총 38개의 염색체 밴드 패턴정보를 6단계로 code화하여 구분론적 패턴 분류방법인 markov network방법을 사용하여 염색체를 분류하는 연구를 수행하였다[2-7]. 최근에는 인공 신경회로망을 이용하여 염색체를 분류하는 연구방법들이 제안되었다. 1994년 Lerner 등은 염색체 영상에서 농도 파일(density profile)에서 64개, 동원체 지수 및 염색체 길이 등 총 66개의 특징 파라 메타를 추출하여 two-layer feedforward neural network을 사용하여 패턴분류기를 구성하는 연구를 수행하였으며, 추출한 특징파라메타의 약 70%이상을 사용하여

야만 염색체를 분류하는 수행이 가능하다고 발표 하였다[8-9]. 영상의 패턴을 인식하고 분류하는 수행과정에서 원시영상(raw image)의 특징 정보를 소실하지 않고 특징 파라메타를 추출할 수 있는 영상 전처리방법이 아주 중요하다. 그러나 염색체 영상은 형태적으로 굵어진 염색체가 많이 존재하는 관계로 인하여 같은 번호의 염색체일지라도 동일한 특징 파라메타의 추출이 용이하지 않다. 동일한 부류의 패턴 정보들은 패턴 벡터 공간(pattern vector space)상에서 군집화를 형성하여 야만 다른 패턴들과의 분류가 용이 하다. 염색체 영상을 MAT(medial axis transform)방법을 사용하여 염색체를 분류하는 연구를 수행하였으나 동일한 번호의 염색체에서 특징 파라메타의 편차가 많은 관계로 인하여 인식율의 향상에 한계가 있었다[10-11]. 따라서 원시 영상정보의 손실이 적으며, 동일한 부류의 영상 패턴을 군집화 하는 영상 처리방법이 요구 된다. 본 연구 에서는 염색체 영상의 재구성 기법(reconstruction algorithm)을 제시하고 이를 이용하여 추출한 염색체 패턴 정보를 신경 회로망의 입력패턴으로 사용하였다.

최근 패턴 인식 분야에서 사용되는 인공 연구에서는 염색체를 분류하는 경우 다층 신경회로망(multi-layer neural network)의 구조를 사용하여 학습과 분류에 사용하고 있다. 그리고 분류의 정확도를 높이기 위하여 많은 양의 패턴 입력 정보들이 사용되고 있다. 그러나 다층 신경 회로망에서 입력 파라메타의 수가 증가하면 학습시간이 아주 많이 소요되며, 어떤 파라메타들은 학습을 방해하는 요인으로 작용한다.

따라서 본 연구에서는 장용훈[12]등이 제시한 염색체 핵형 분류를 위한 최적 패턴 분류기인 2단 다층 인공 신경회로망(two-step multi-layer neural network : TMANN)을 사용 하였다. 이러한 TMANN은 두단계의 신경회로망으로 구성된다. 첫단계의 신경회로망을 TMANN1이라 하며, 이는 염색체의 균을 먼저 분류한다. 두 번째 단계의 신경 회로망을 TMANN2라 하며, 개별 염색체를 분류하는 신경회로망이다. 염색체의 패턴 특징 정보는 임상적으로 정상으로 판명된 20명의 염색체 영상을 재구성하여 동원체 지수(centromeric index:C.I.), 상대 길이비(relative length:R.L.), 상대 면적비(relative area:R.A.) 및 염색체 길이(chromosome length:C.L.)등 4종류의 정보를 추출했다.

본 연구에서는 추출한 패턴 정보를 사용하여 다음과 같이 핵형 분류를 수행하였다. TMANN1에는 C.I., R.L., 및 R.A. 인 3종류를 사용하였으며, TMANN2에는 입력 패턴을 두 가지로 하여 실험을 수행하였다. 첫 번째는 C.I., R.L., 및 R.A.의 3종류를 사용하였으며, 두 번째는 C.I., R.L., R.A. 및 C.L.인 4종류의 패턴 특징 정보를 사용하였다. 이상과 같은 수행에 있어 추출된 20명의 패턴 정보중 10명분은 학습 입력으로 사용하였으며, 나머지 10명분은 분류 대상 입력으로 사용하여 실험결과를 타 연구자들의 결과와 비교하였다.

### 2. 재구성에 의한 형태학적 패턴 정보 추출

염색체의 관찰은 유사분열(mitosis) 중기(metaphase)의 세포에서 가장 용이하다. 따라서 유사 분열 중기의 세포를 현미경을 통하여 관찰한 후, 이를 촬영한 필름을 스캐너(scanner)를

사용하여 임상적인 관점에서 정상인으로 판명된 20명 즉, 920개의 염색체를 촬영하여 영상파일을 저장하였다. 그림 1은 분류하기 이전의 염색체 영상을 나타낸다. 염색체는 형태적으로 곱어져 있으므로 같은 번호의 염색체에서 추출한 특징파라메타의 편차가 심하다. 따라서 비선형적인 형태의 염색체를 선형적으로 재구성하여 특징파라메타를 추출하면 편차를 줄일 수 있다. 이러한 방법을 사용하여 특징 파라메타를 추출하면, 패턴 벡터 공간상에서 동일한 부류별로 군집화된 패턴 클래스가 구하여 진다. 이렇게 군집화된 패턴 클래스는 염색체의 분류를 향상에 기여한다. 따라서, 적절하고 군집화된 특징 파라메타의 선택이 인식율을 향상시킬 수 있으며, 많은 양의 특징 파라메타들을 사용하여 나타낼 수 있는 복잡도의 감소를 위하여 적절한 특징 정보를 선택하여야만 한다. 염색체 영상의 형태학적인 특징 파라메타의 계산방법은 표 1과 같다.

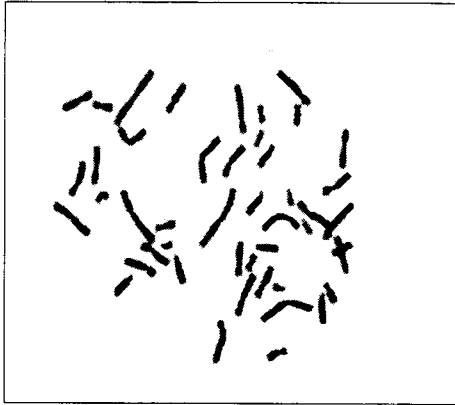


그림 1 유사분열 중기의염색체의 영상.

표 1. 특징파라메타 계산방법.

| 특징 파라메타 | 계산 방법  |
|---------|--|
| 동원체 지수  | $C.I = \frac{\text{염색체 단원의 길이}}{\text{염색체의 길이}} = \frac{l_s}{l_c} \leq 0.5$  |
| 상대 길이비  | $l_i = \text{상염색체의 총길이} = \sum_{c=1}^4 l_c$<br>$R.L. = \frac{\text{염색체의 길이}}{\text{상염색체의 총길이}} = \frac{l_c}{l_i}$  |
| 상대 면적비  | $S_i = \text{상염색체의 총면적} = \sum_{c=1}^4 S_c$<br>$R.A. = \frac{\text{각염색체의 면적}}{\text{상염색체의 총면적}} = \frac{S_c}{S_i}$ |

본 연구에서 사용하기 위한 특징 파라메타는 염색체의 형태가 비선형적인 구조를 지니고 있다. 따라서 염색체 영상의 중앙축의 화소를 기준으로 32방사 방향의 염색체 폭을 구하여 선형적인 형태의 염색체 영상으로 재구성하였다. 재구성 영상에서 특징 파라메타를 추출하였으며, 영상을 재구성하는 알고리즘은 다음과 같으며, 이에 대한 순서도는 그림 2와 같다.

- 세선화 알고리즘을 수행하여 염색체의 중앙 축을 구한다. 이때 염색체 양끝에서 잘린 부분(양끝의 약 10%씩)은 맨 끝점의 좌표와 다음 점의 좌표와의 각도를 이용하여 연장한다.
- 설정된 염색체 중앙 축의 화소( $x_c^k, y_c^k$ )를 원점으로 선택한다.
- 원점을( $x_c^k, y_c^k$ )로 하여  $\Delta\theta=2\pi/M$ 의 등 각도로 M개의 길이를 계산한다. 여기서 k는 중앙 축의 화소에 대한 알고리즘의 반복 수행 횟수이며, M은 32이다.

- 길이 성분 i에 대하여 식(1)을 사용한 문턱치 예지 검출을 행하여  $H_{thr}$ 값이 문턱치  $H_{th}$ 보다 크면 길이 성분 i에 대한 수행을 종료하고 이의 좌표( $x_e^m, y_e^m$ )를 저장한다.  
$$H_{thr}(r^i) = \frac{g(r^i+1)+g(r^i)-g(r^i-1)-g(r^i-2)}{2}$$
 (1)  
여기서  $r^i$ 는 길이 성분 i에 대한 방사 방향의 증분을 나타내며,  $g(r)$ 는 염색체 각 화소의 농도 정보치를 나타낸다.
- M개의 방향에 대해 (d)를 반복 수행한다.
- (e)가 종료되면 식(2)와 같이 원점의 좌표( $x_c^k, y_c^k$ )와 저장된 좌표( $x_e^m, y_e^m$ )의 유클리드 거리를 계산한다.  
$$l_{ik} = \sqrt{(x_e^m - x_c^k)^2 + (y_e^m - y_c^k)^2}$$
 (2)
- (f)의 수행이 완료되면 식(3)과 같이  $l_{ik}$ 와  $\pi$ 각만큼 떨어진  $l_{ik-\pi}$ 의 길이를 합하여 재구성을 위한 길이  $l_i$ 를 구한다.  
$$l_i = l_{ik} + l_{ik-\pi}$$
 (3)
- 길이  $l_i$ 중에서 가장 짧은 길이  $L_i$ 를 식(4)와 같이 구한다.  
$$L_i = \text{Min} [\text{Sort} (l_i)]$$
 (4)
- 중앙축의 총화소 수 k개를 만족할 때까지 (d)에서 (h)까지 반복 수행한다.

이상과 같이 수행하면 염색체를 재구성하기 위한 전체 선분이 결정된다. 결정된 선분상의 화소들을 이용하면 염색체 영상을 재구성 할 수 있다

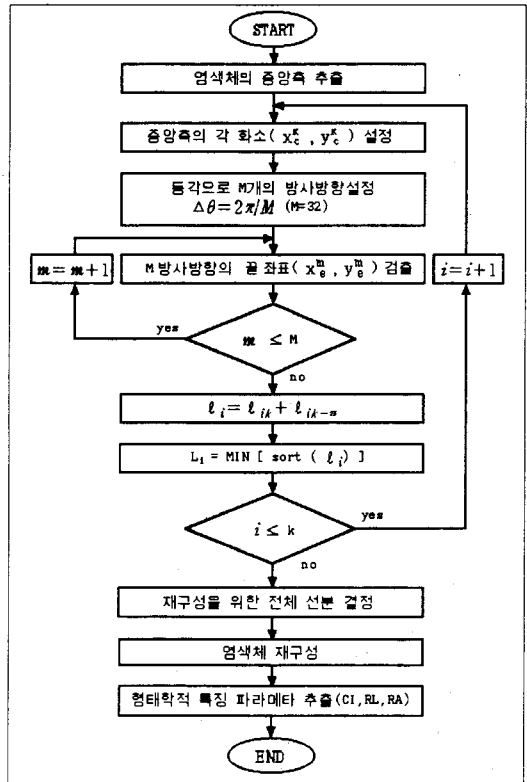


그림 2 염색체 영상의 재구성을 위한 순서도.

그림 3은 재구성 알고리즘을 사용하여 1번 염색체의 원시영상과 재구성된 염색체 영상 및 추출된 특징 파라메타의 결과를 나타내었다. 표 2는 염색체 영상을 재구성하여 추출한 특징 파라메타를 ISCN(an international system for human cytogenetic nomenclature)과 HC(human cytogenetic)의 데이터와 비교하였다. 같은 번호의 염색체 특징파라메타의 차이는 표준편차로 나타내었다.

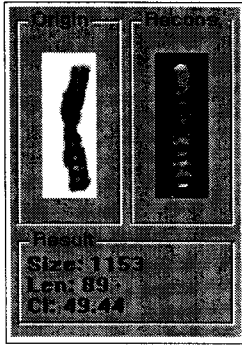


그림 3 염색체 영상의 재구성 및 특징 파라메타.

표 2 염색체 재구성 영상에 의한 형태학적 특징 파라메타.

| NO | Centromeric Index |                | Relative Length |                |      |             | Relative Area  |     |                |     | Chromosome Length |     |
|----|-------------------|----------------|-----------------|----------------|------|-------------|----------------|-----|----------------|-----|-------------------|-----|
|    | ISC N             | Reconstruction | ISC N           | Reconstruction |      | HC          | Reconstruction |     | Reconstruction |     | Mean              | S.D |
|    |                   |                |                 | Mean           | S.D  |             | Mean           | S.D | Mean           | S.D |                   |     |
| 1  | 48.0              | 48.41 ± 0.41   | 9.08            | 7.97 ± 0.61    | 8.44 | 8.45 ± 0.61 | 88.2 ± 2.71    |     |                |     |                   |     |
| 2  | 38.1              | 41.70 ± 2.40   | 8.45            | 7.74 ± 0.55    | 8.02 | 8.50 ± 0.58 | 78.7 ± 2.47    |     |                |     |                   |     |
| 3  | 45.9              | 45.01 ± 0.81   | 7.06            | 6.12 ± 0.29    | 6.83 | 6.32 ± 0.47 | 64.6 ± 2.36    |     |                |     |                   |     |
| 4  | 27.6              | 29.44 ± 2.94   | 6.55            | 6.01 ± 0.28    | 6.30 | 6.48 ± 0.45 | 56.5 ± 2.01    |     |                |     |                   |     |
| 5  | 27.4              | 26.88 ± 0.59   | 6.13            | 5.86 ± 0.33    | 6.08 | 6.05 ± 0.41 | 63.2 ± 1.84    |     |                |     |                   |     |
| 6  | 37.7              | 37.11 ± 1.80   | 5.84            | 5.51 ± 0.31    | 5.90 | 6.02 ± 0.43 | 54.5 ± 1.78    |     |                |     |                   |     |
| 7  | 37.3              | 38.22 ± 0.94   | 5.28            | 5.17 ± 0.41    | 5.36 | 5.07 ± 0.37 | 54.8 ± 1.82    |     |                |     |                   |     |
| 8  | 35.9              | 34.12 ± 0.86   | 4.96            | 4.67 ± 0.33    | 4.93 | 4.94 ± 0.32 | 51.4 ± 1.79    |     |                |     |                   |     |
| 9  | 33.3              | 33.23 ± 0.43   | 4.83            | 4.52 ± 0.31    | 4.80 | 4.62 ± 0.31 | 47.8 ± 1.58    |     |                |     |                   |     |
| 10 | 31.2              | 32.69 ± 0.61   | 4.68            | 4.60 ± 0.32    | 4.59 | 4.48 ± 0.29 | 44.6 ± 1.47    |     |                |     |                   |     |
| 11 | 35.6              | 34.38 ± 0.76   | 4.63            | 4.63 ± 0.34    | 4.61 | 4.56 ± 0.27 | 45.2 ± 1.52    |     |                |     |                   |     |
| 12 | 30.9              | 31.35 ± 0.89   | 4.46            | 4.38 ± 0.32    | 4.66 | 4.41 ± 0.26 | 45.7 ± 1.55    |     |                |     |                   |     |
| 13 | 14.8              | 15.14 ± 0.48   | 3.64            | 3.49 ± 0.28    | 3.74 | 3.40 ± 0.24 | 38.4 ± 1.23    |     |                |     |                   |     |
| 14 | 15.5              | 15.18 ± 0.95   | 3.55            | 3.59 ± 0.19    | 3.56 | 3.49 ± 0.25 | 32.3 ± 1.14    |     |                |     |                   |     |
| 15 | 14.9              | 14.65 ± 0.72   | 3.36            | 3.52 ± 0.18    | 3.46 | 3.35 ± 0.27 | 38.1 ± 1.18    |     |                |     |                   |     |
| 16 | 40.6              | 40.26 ± 0.83   | 3.23            | 3.01 ± 0.09    | 3.36 | 2.92 ± 0.19 | 33.4 ± 1.21    |     |                |     |                   |     |
| 17 | 31.4              | 30.48 ± 0.78   | 3.15            | 2.92 ± 0.12    | 3.25 | 2.81 ± 0.16 | 27.2 ± 0.97    |     |                |     |                   |     |
| 18 | 26.9              | 26.18 ± 0.84   | 2.76            | 2.80 ± 0.17    | 2.93 | 2.73 ± 0.20 | 25.3 ± 0.87    |     |                |     |                   |     |
| 19 | 42.9              | 43.09 ± 1.65   | 2.52            | 2.77 ± 0.21    | 2.67 | 1.95 ± 0.15 | 26.1 ± 0.84    |     |                |     |                   |     |
| 20 | 44.6              | 43.75 ± 1.48   | 2.33            | 2.61 ± 0.19    | 2.56 | 2.20 ± 0.12 | 24.5 ± 0.75    |     |                |     |                   |     |
| 21 | 25.7              | 25.54 ± 1.34   | 1.83            | 1.81 ± 0.13    | 1.90 | 1.35 ± 0.09 | 19.4 ± 0.62    |     |                |     |                   |     |
| 22 | 25.0              | 25.23 ± 0.72   | 1.68            | 1.96 ± 0.14    | 2.04 | 1.63 ± 0.13 | 21.3 ± 0.69    |     |                |     |                   |     |
| X  | 36.9              | 35.21 ± 1.08   | 5.80            | 4.54 ± 0.28    | 5.12 | 4.79 ± 0.29 | 41.6 ± 1.46    |     |                |     |                   |     |
| Y  | 16.3              | 16.60 ± 0.66   | 2.22            | 2.20 ± 0.16    | 2.15 | 1.76 ± 0.12 | 20.7 ± 0.72    |     |                |     |                   |     |

### 3. 신경회로망을 이용한 패턴 분류기 구성

신경회로망은 상호 연결된 뉴런에 의하여 입력과 출력사이에서 비선형 사상(nonlinear mapping)을 하며, 이러한 사상에 의해 변환, 필터, 최적화, 연상기억 및 분류 등의 기능을 수행할 수 있다.

염색체들은 형태학적으로 유사한 패턴을 가지는 부류들을 모아 염색체 군을 형성하고 있으며, 이를 기준으로 핵형 분류도(karyogram)를 작성하여 염색체의 분류 및 진단에 활용하고 있다. 따라서 이러한 특징을 근거로 하여 염색체를 인식하고 분류하기 위하여 2단 다중 신경회로망을 사용한다. 첫 번째 단계의 다중 신경회로망(TMANN1)은 염색체를 7개 군으로 분류하기 위하여 구성하였고, 두 번째 단계의 다중 신경회로망(TMANN2)은 각 군내에 속한 개개의 염색체(1-22, X, Y)를 분류하기 위하여 7개의 다중 신경회로망으로 구성하였다. TMANN의 학습과 분류에 사용된 입력 파라메타들은 2절에서 구한 4종류의 특징 파라메타들을 사용한다. 표 3은 A에서 G까지인 7개의 염색체 군과 각 군에 포함되는 염색체 번호를 나타내었다.

본 연구에서 수행한 TMANN은 그림 4와 같이 구성하였다. 그리고 TMANN1에는 3종류의 특징 파라메타를 사용하였으며, TMANN2에는 두 가지의 입력 패턴을 사용하였다. 즉, 첫

번째 입력패턴을 1st pattern이라 하고 C.I., R.L., 및 R.A.이며, 두 번째 입력패턴을 2nd pattern이라 하고 C.I., R.L., R.A., 및 C.L.을 사용하였다. TMANN1은 입력층 3개, 은닉층 30개, 출력층 뉴런 7개로 구성되며, 출력변수는 7개의 염색체 군(A~G)을 나타낸다. TMANN1에 의해 분류된 염색체 군의 출력은 TMANN2에 구성되어 있는 7개의 다중 신경회로망 중에서 해당되는 신경회로망을 선택한다.

표 3 염색체 군과 염색체 번호.

| Chromosome Group | Chromosome Number | Chromosome Group | Chromosome Number |
|------------------|-------------------|------------------|-------------------|
| Group A          | 1번 ~ 3번           | Group E          | 16번 ~ 18번         |
| Group B          | 4번 ~ 5번           | Group F          | 19번 ~ 20번         |
| Group C          | 6번 ~ 12번, X       | Group G          | 21번 ~ 22번, Y      |
| Group D          | 13번 ~ 15번         |                  |                   |

TMANN2는 개개의 군에 해당되는 각각의 염색체를 분류하는 7개의 다중 신경회로망으로 구성되었고, 입력층 뉴런은 첫 번째의 실험에서는 3개, 두 번째 실험에서는 4개이며, 분류결과를 나타내는 출력층 뉴런의 번호는 염색체의 번호와 일치하도록 구성하였다.

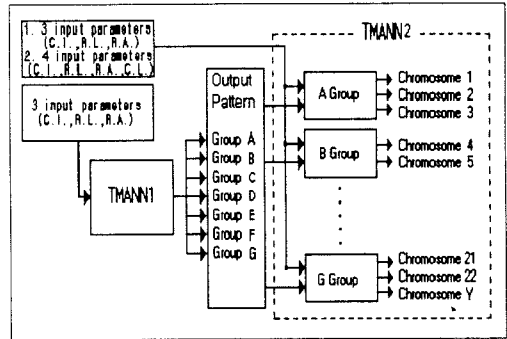


그림 4 TMANN의 구조.

신경회로망에서는 가중치의 초기값, 활성화함수의 기울기, 학습 상수(learning constant), 중간층 뉴런의 개수 등이 학습에 큰 영향을 미치게 된다. 따라서 염색체의 특징벡터를 입력으로 하여 학습되는 신경회로망의 최적 상태를 정확히 정의할 수 없으므로 다양한 형태의 실험이 요구된다.

본 연구에서는 지도학습인 오차 역전파 학습알고리즘을 사용하였으며, 학습요소들의 적절한 값과 중간층 뉴런의 적절한 수를 선정하기 위하여 여러 번의 실험을 수행한 결과 표 4와 같은 파라메타를 얻을 수 있었다. TMANN2에서의 조건은 1st pattern과 2nd pattern일 때와 동일한 조건으로 하여 신경회로망을 학습시켰다. 이때 사용된 교사신호는 분류와 일치하는 뉴런의 교사신호는 1.0이며 나머지는 0.0이다. 또한 학습 시간을 단축하고 신경회로망이 국부적인 극소점(local minimum)에 빠지지 않게 하기 위하여 모멘트법과 적응 학습법을 사용하였다.

표 4 신경회로망의 파라메타.

| 파라메타      | 신경회로망  |  | Group A |      | Group B |      | Group C |      |         |     |
|-----------|--------|--|---------|------|---------|------|---------|------|---------|-----|
|           | TMANN1 |  | 1st     | 2nd  | 1st     | 2nd  | 1st     | 2nd  |         |     |
| 활성함수의 기울기 | 1.0    |  | 1.2     | 1.2  | 1.1     | 1.1  | 1.0     | 1.0  |         |     |
| 학습 상수     | 0.05   |  | 0.05    | 0.05 | 0.01    | 0.01 | 0.05    | 0.05 |         |     |
| 중간층 뉴런의 수 | 30     |  | 10      | 10   | 10      | 10   | 30      | 30   |         |     |
| 파라메타      | 신경회로망  |  | Group D |      | Group E |      | Group F |      | Group G |     |
|           | TMANN1 |  | 1st     | 2nd  | 1st     | 2nd  | 1st     | 2nd  | 1st     | 2nd |
| 활성함수의 기울기 | 0.9    |  | 0.9     | 0.8  | 0.8     | 1.0  | 1.0     | 1.0  | 1.0     |     |
| 학습 상수     | 0.1    |  | 0.1     | 0.5  | 0.5     | 0.05 | 0.05    | 0.1  | 0.1     |     |
| 중간층 뉴런의 수 | 15     |  | 15      | 15   | 15      | 10   | 10      | 10   | 10      |     |

#### 4. 실험결과 및 고찰

패턴 분류기로 구현된 신경회로망은 각 층의 뉴런수, 활성화 함수의 기울기, 학습 상수, 가속 상수, 입력 패턴 수, 학습 회수 및 최대 허용 오차 등의 파라메타들을 키보드를 통하여 입력하므로써 신경회로망의 학습에 유연하게 대처할 수 있도록 설계하였고, 또한 학습되는 과정을 가시화하기 위하여 오차의 변화를 그래프 처리하였으며, 추출된 결과는 표로 구성하여 파일로 저장하였다.

그림 5와 그림 6은 각각 재구성된 염색체의 1st pattern과 2nd pattern을 사용하여 A그룹을 분류하는 TMANN2에서의 학습과정을 SSE(sum-square error)의 변화로 나타낸 것이다.

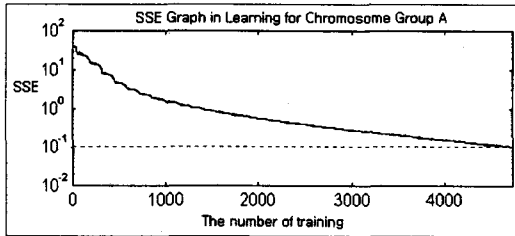


그림 5 1st pattern입력시 TMANN2의 자승 오차의 변화.

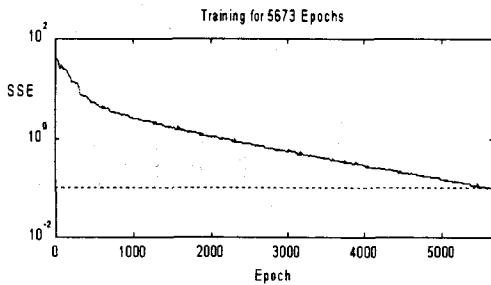


그림 6 2nd pattern입력시 TMANN2의 자승 오차의 변화.

염색체의 분류대상인 10명의 염색체를 패턴분류기로 분류한 결과 TMANN1에 의한 염색체 군의 분류에서는 오분류없이 정확히 100%를 분류하였다. TMANN2에 있어서 1st pattern은 전체 분류 대상 460개 중 오분류 결과는 21개로 4.56%의 분류 오차율을 나타내었으며, 2nd pattern에 의한 오분류 결과는 15개로 3.26%의 분류 오차율을 나타내었다. 각 군의 분류 오차는 표 5와 같다. 표 6은 본 연구에서 수행한 방법과 타 연구 방법에 의한 분류오차의 결과를 비교하여 나타내었다.

표 5. 염색체 분류 오차.

|       | Group A |    | Group B |    | Group C |     | Group D |    | Group E |    | Group F |    | Group G |    | Total |       |
|-------|---------|----|---------|----|---------|-----|---------|----|---------|----|---------|----|---------|----|-------|-------|
|       | 1s      | 2n | 1s      | 2n | 1s      | 2n  | 1s      | 2n | 1s      | 2n | 1s      | 2n | 1s      | 2n | 1st   | 2nd   |
| 오분류개수 | 0       | 0  | 0       | 0  | 1       | 9   | 6       | 4  | 0       | 0  | 3       | 2  | 0       | 0  | 21    | 15    |
| 분류오차율 | 0       | 0  | 0       | 0  | 2.1     | 1.1 | 1.0     | 0  | 0       | 0  | 0.4     | 0  | 0       | 0  | 4.56% | 3.26% |

표 6. 염색체 분류방법 및 결과.

| Researches | Methods                    | Classification error(%) |      | feature parameters |     | Remarks                             |
|------------|----------------------------|-------------------------|------|--------------------|-----|-------------------------------------|
|            |                            | 1st                     | 2nd  | 1st                | 2nd |                                     |
| Proposed   | TMANN                      | 4.56                    | 3.26 | 3                  | 4   | 염색체 영상의 재구성 알고리즘에 의한 입력 패턴 추출       |
| Lerner     | Two-layer N.N.             | 연급 없음                   |      | 66                 |     | 특징파라메타의 약 70% 사용시 분류가능성 제시          |
| Piper      | Maximum likelihood         | 5.8 ~ 21                |      | 28                 |     | 영상의 질에 의한 오분류의 범위                   |
| Lucas      | Non-parametric baye's rule | 4.0 ~ 11.5              |      | 7                  |     | 곱이진 염색체 분류대상에서 제외 영상의 질에 의한 오분류의 범위 |
| Erik       | Markov Network             | 7.9                     |      | 38                 |     |                                     |

Lerner등은 농도파일(density profile)에서 64개, 동원체 지수 및 염색체 길이 등 66개의 특징 파라메타를 추출하였다. 그리고 추출한 특징파라메타를 64d.p., 64d.p. +c.i. 및 64d.p.+c.i.+long.의 3가지로 입력패턴을 구성하여 two-layer feedforward neural network을 사용하여 패턴분류기를 구성하는 연구를 수행하였으며, 추출한 특징파라메타의 약 70%로서 최종적인 분류수행이 가능하다고 발표하였고, 분류오차에 대한 언급은 없었다.

Piper 등에 의해 수행된 연구는 염색체의 면적, 상대농도, 길이, 동원체 지수 및 가장농도곡선 등 28개의 특징 파라메타를 사용하여 maximum likelihood방법으로 데이터 베이스를 구축하여 분류한 결과 영상의 질에 따라 5.8%~21%의 분류 오차가 나타났다. Lucas등은 염색체 길이, 동원체 지수 및 5개의 농도정보 등 총 7개의 특징 파라메타를 사용하여 non-parametric baye's rule을 적용하여 영상의 질에 따라 분류오차가 4.0~11.5%로 우수한 결과를 나타내었지만, 형태학적으로 많이 구부러진 염색체는 실험 대상에서 제외된 단점을 가진다.

#### 5. 결론

염색체 영상은 곱이진 경우가 대부분이며 같은 번호의 염색체에서도 동일한 패턴정보가 구해지지 않는 비선형적인 특징을 지니고 있다. 패턴 벡터 공간상에서 같은 부류의 패턴 정보들이 일정한 범주 내에서 군집화를 구성하여야만 다른 부류의 패턴들과의 분류가 상호 용이하다. 따라서 비선형성의 특징을 많이 지니는 염색체 영상의 패턴 정보를 추출하기 위하여 염색체 영상을 골개 구성하는 영상 재구성 알고리즘을 제시하여 20명에 대하여 4종류의 형태학적인 특징 정보인 C.I., R.L., R.A., 및 C.L.의 패턴 정보를 추출하였다. 패턴인식 분야에 많이 사용되고 있는 인공 신경회로망을 이용하여 핵형 분류에 적합한 패턴 분류기를 구성하기 위한 연구를 수행하였다. 실험 결과 3개의 입력 패턴정보를 사용하여 염색체 군을 분류한 결과오차는 없었으며, 전체 염색체에 대한 평균 분류 오차는 3종류의 패턴 정보 사용시 4.56%를 나타내었으며, 4종류의 패턴 정보를 사용했을 때는 3.26%로 감소하였다.

본 연구는 학습시간을 최소화하여 염색체를 분류하기 위한 패턴 분류기를 구현하였으며, 염색체의 분류오차에서도 만족한 결과를 나타내었다. 또한 분류에 필요한 패턴 정보를 같은 부류들을 군집화 시키므로써 비선형성을 내포하는 염색체의 패턴 정보를 선형화 하여 분류오차에서도 타 연구에 비하여 우수한 결과를 나타내었다. 분류 오차가 나타나는 염색체 군의 신경회로망에 다른 패턴 정보를 부가하여 패턴 분류기를 개선하면 분류오차를 더욱 감소시킬 수 있을 것으로 생각된다.

#### 참고 문헌

- (1) An International System for Human Cytogenetic Nomenclature(ISCN), KARGER, 1985.
- (2) Robert S. Ledly, "High-speed automatic analysis of biomedical picture," Science, Vol. 146, pp. 216-223, 1964.

- [3] J. M. Cho and D. H. Hong, "Computer-assisted karyotyping system of Giemsa-stained chromosomes (II)," *Proc. of 1989 Korea-Japan Joint Conference on MBE*, pp.19-23, Sep. 21-22, 1989.
- [4] Lucas J. van Vliet, Ian T Young and Brian H. Mayall, "The athena semi-automated karyotyping system," *Cytometry*, Vol. 11, pp. 51-58, 1990.
- [5] Brian H. Mayall, James D. Tucker, Mari L. Christensen, Lucas J. van Vliet and Ian T. Young, "Experience with the athena semi-automated karyotyping system," *Cytometry*, Vol. 11, pp. 59-72, 1990.
- [6] Jens Gregor and Erik Granum, "Finding chromosome centromeres using band pattern information," *Comput. Biol. Med.*, Vol. 21, No. 1/2, pp. 55-67, 1991.
- [7] Jim Piper and Erik Granum, "On fully automatic features measurement for banded chromosome classification," *Cytometry*, Vol. 10, pp. 242-255, 1989.
- [8] Lerner B., Levinstein M., Rosenberg B., Guterman H., Dinstein I. and Romem Y., "Feature selection and chromosome classification using a multilayer perceptron neural network," *IEEE International Conference on Neural Networks*, Vol. IV, 6/7, pp. 3540-3545, Jun. 28-Jul. 2, 1994.
- [9] Key Rok Jun, Sang Hee Eom and Young Hoon Chang, "Optimal neural network classifier for chromosome karyotype classification," *AI Simulation, 96, Int. SCS*, pp. 315-318, march, 1996.
- [10] 전계록, 최육환, 남기곤, 엄상희, 이권순, 장용훈, "염색체 핵형 분류를 위한 계층적 인공 신경회로망 분류기 구현," *대한의용생체공학회지*, Vol. 18, No. 3, pp. 233-241, 1997.
- [11] 장용훈, 이권순, 정형환, 엄상희, 전계록, "신경회로망을 이용한 염색체 영상의 최적 패턴 분류기 구현에 관한 연구," *대한전기학회지*, Vol. 46, No. 7, pp.1129-1134.