

# 염색체 핵형 분류를 위한 계층적 인공 신경회로망 분류기 구현

전계록·최옥환\*·남기곤\*\*·엄상희\*\*·이권순\*\*\*·장용훈\*\*\*

= Abstract =

## The implementation of Hierarchical Artificial Neural Network Classifier for Chromosome Karyotype Classification

K.R. Jun, W.W. Chai\*, K.G. Nam\*\*, S.H. Eom\*\*, K.S. Lee\*\*\*, Y.H. Chang\*\*\*

The research on chromosomes is very significant in cytogenetics since genes of the chromosomes' control revelation of the inheritance plasma. The human chromosome analysis is widely used to study leukemia, malignancy, radiation hazard, and mutagen dosimetry as well as various congenital anomalies such as Down's, Klinefelter's, Edward's, and Patau's syndrome.

The framing and analysis of the chromosome karyogram, which requires specific cytogenetic knowledge is most important in this field. Many researches on automated chromosome karyotype analysis methods have been carried out, some of which produced commercial systems. However, there still remains much room to improve the accuracy of chromosome classification and to reduce the processing time in real clinic environments.

In this paper, we proposed a hierarchical artificial neural network(HANN) to classify the chromosome karyotype. We extracted three or four chromosome morphological feature parameters such as centromeric index, relative length ratio, relative area ratio, and chromosome length by preprocessing from ten human chromosome images.

The feature parameters of five human chromosome images were used to learn HANN and the rest of them were used to classify the chromosome images. The experiment results show that the chromosome classification error is reduced much more than that of the other researchers using less feature parameters.

**Key words** : Chromosome, Hierarchical artificial neural network, Morphological feature, Karyotype

### 서론

세상에 존재하는 모든 생물체는 각각의 형질을 결정하는 유전자를 염색체 내에 포함하고 있다. 즉, 생물체 종(species) 특유의 특성들이 유지 계승될 수 있도록 유전자가 생식세포를 통하여 다음 세대로 유전된다. 생물체의

종들은 각 종별로 개체성이 존재하므로 종이 변화하지 않으면 종이 가지고 있는 염색체의 수와 형태는 항상 일정하게 된다.

세포유전학(cytogenetics) 분야에 있어서는 유전 형질의 발현을 조절하는 유전자가 세포내의 염색체에 존재하므로 염색체에 관한 연구는 상당히 중요하다. 사람의 염

부산대학교 병원 의공학과

Dept. of Biomedical Eng., Pusan National Univ. Hospital

\*부산대학교 병원 산부인과

\*Dept. of Obstetrics & Gynecology, Pusan National Univ. Hospital

\*\*부산대학교 전자공학과

\*\*Dept. of Electronic Eng., Pusan National University

\*\*\*동아대학교 전기공학과

\*\*\*Dept. of Electrical Eng., Dong A University

본 논문은 1996년도 부산대학교 병원 지정진료 연구비에 의하여 수행되었음.

통신저자 : 전계록, (602-739) 부산광역시 서구 아미동 1가 10번지, Tel. (051)240-7535, Fax. (051)240-7534

색체 분석은 여러 가지의 선천성 질환 즉, 다운, 클라인펠터, 애드워드, 그리고 파타우 증후군 등뿐만 아니라 백혈병, 악성 종양, 방사선 장애, 변위유발소 선량법(mutagen dosimetry)의 연구, 그리고 생물학적인 연구 등을 위해 매우 중요하다.

사람의 염색체의 수는 1956년 Tijo와 Levan에 의하여 마흔여섯 개로 밝혀졌으며[1], 그후 1960년 국제 염색체 회의에서 인간의 염색체의 명명과 분류법에 관해 ISCN (international systems for human cytogenetic nomenclature)이라는 규약이 설정되었다. 이 규약에 의하면 정상인의 염색체 수는 스물두쌍의 상 염색체(autosomes)와 한 쌍의 성 염색체(sex chromosome:X,Y)로 구분되며, 여성의 경우는 XX, 남성의 경우는 XY인 성 염색체로 구성된다. 그리고 염색체의 각 쌍은 동원체(centromere)의 위치, 부수체(satellite)의 유무, 상대길이비 등에 의해 분류된다. 이러한 염색체의 수 및 형태학적인 특징(morphological feature)을 핵형(karyotype)이라 하며, 염색체의 군과 염색체의 분류를 번호순으로 나열하여 테이블화한 도형을 핵형분류도(karyogram)라고 한다[2].

염색체의 핵형분류도의 작성과 해석에는 고도로 숙달된 세포유전학 분야의 전문가가 필요하다. 즉, 세포 배양, 염색, 슬라이드 제작, 관찰에 적합한 염색체 군 선택 및 사진촬영, 필름 현상, 염색체 분리, 분리된 염색체의 형태구조학적인 특징을 고려하여 핵형분류도를 작성하고, 염색체를 해석하는 등의 처리과정이 매우 복잡하므로 많은 시간과 노력이 소요된다. 따라서 컴퓨터를 이용하여 염색체의 핵형을 분류하는 방법이 연구되었다.

컴퓨터를 이용한 염색체의 핵형 자동추출 및 해석에 관한 연구는 1964년 Ledly[3]에 의해 처음 시도되어 임상 세포유전학 연구에 유용하게 사용할 수 있음을 제시하였고, Neurath[4]는 현미경의 화상 데이터로부터 염색체의 윤곽선을 추출하고 동원체의 위치를 계산하는 알고리즘을 도입하여 염색체를 해석하는 시스템을 구성하였다. 그리고 1989년 Piper와 Granum[5] 등은 염색체의 면적, 상대농도, 길이, 그리고 농도분포 함수를 사용하여 동원체의 위치를 검출하는 알고리즘을 제안하여 염색체의 핵형을 분류하였고, 1990년 Young[6] 등은 염색체 길이, 동원체 지수, 그리고 염색체 마디들에 관한 특징벡터를 사용하여 염색체의 핵형을 분류하고 해석하는 시스템을 개발하였다. 이러한 연구 결과들에 의해 근래에는 상용화된 시스템이 발표되어 임상에 적용되고 있으나 아직까지는 분류 정확도가 낮으며, 염색체 자동분류 과정에 운용자의 개입이 많아 개선되어야 될 사항들이 여전히 많이 있다.

본 연구에서는 최근 패턴인식 분야에서 많이 연구되고 있는 인공신경망을 이용하여 계층적 인공신경회로망 분류기(hierarchical artificial neural network classifier :

HANNC)를 구성하여 염색체의 핵형 분류기로 사용하였다.

HANNC의 구성은 다량의 입력패턴에 의한 학습 소요 시간을 최소화하며, 염색체를 단계적으로 분류하기 위해 두 단계의 신경회로망으로 분리하였다. 첫 번째 단계의 신경회로망(HANNC1)은 염색체들을 일곱 개의 군(group) 즉, A군(염색체 1-3), B군(염색체 4-5), C군(염색체 6-12, X), D군(염색체 13-15), E군(염색체 16-18), F군(염색체 19-20), 그리고 G군(염색체 21-22, Y) 등으로 분류하기 위해 구성하였다. 그리고 두 번째 단계의 신경회로망(HANNC2)은 일곱 개 군내에 포함된 각각의 염색체를 분류하기 위해 구성하였다.

본 연구를 위하여 임상적인 관점에서 정상으로 판명된 정상인 열명의 염색체를 수집한 후 열명 중 다섯명의 염색체 사진을 영상 전처리(image preprocessing)하였다. 영상 전처리는 비선형적인 구조로 존재하는 염색체 영상에서 특징 파라메타의 추출을 용이하게 하기 위하여, 염색체 중앙축의 화소를 기준으로 32 방향 연장선상에 있는 영상을 추출하여 재구성(reconstruction)하는 방법을 사용하였다[7]. 그리하여 염색체 각각의 동원체지수(centromeric index:C.I.), 상대길이비(relative length:R.L.), 상대면적비(relative area:R.A.)와 같은 특징 파라메타를 구하였고, 이들 파라메타들을 제안한 HANNC의 입력패턴으로 사용하여 HANNC를 학습시켰다. 학습이 끝난 HANNC를 패턴분류기로 사용하여 다른 다섯명에 대한 염색체의 핵형을 분류하였다.

## 염색체의 영상 전처리 및 특징 파라메타 추출

### 1. 염색체의 영상 획득

염색체의 형태학적 구조가 임상적인 관점에서 정상으로 판명된 정상인 열명의 염색체 사진을 스캐너(scanner)를 사용하여 스캔한 후 이를 PC(pentium)에 염색체의 영상화일(image file)로 저장하였다. 염색체의 영상화일을 획득하기 위하여 사용된 하드웨어 환경은 표 1과 같다.

표 1. 염색체의 영상화일 획득에 사용된 하드웨어 환경  
Table 1. Hardware environment for the chromosome image file acquisition

Item	Specification
C.P.U	Pentium 90 MHz
Main Memory	16MB
Hard Disk	1.27GB
Graphic Card	SVGA : 1024×768, 65536 Color, 2MB
Scanner	HP Scanjet 4C
Scanner Adapter	Transparency Adapter

표 2. 특징 파라메타 계산방법

Table 2. The feature parameter calculation equation

특징 파라메타	계산 방법
동원체 지수 (centromeric index : C.I.)	$C.I. = \frac{\text{염색체 단완의 길이}}{\text{염색체의 길이}} = \frac{l_s}{l_c} \leq 0.5$
상대 길이비 (relative length ratio : R.L.)	$l_c = \text{상염색체의 총길이} = \sum_{c=1}^{44} l_c$ $R.L. = \frac{\text{염색체의 길이}}{\text{상염색체의 총길이}} = \frac{l_c}{l_c}$
상대 면적비 (relative area ratio : R.A.)	$S_c = \text{상염색체의 총면적} = \sum_{c=1}^{44} S_c$ $R.A. = \frac{\text{각염색체의 면적}}{\text{상염색체의 총면적}} = \frac{S_c}{S_c}$

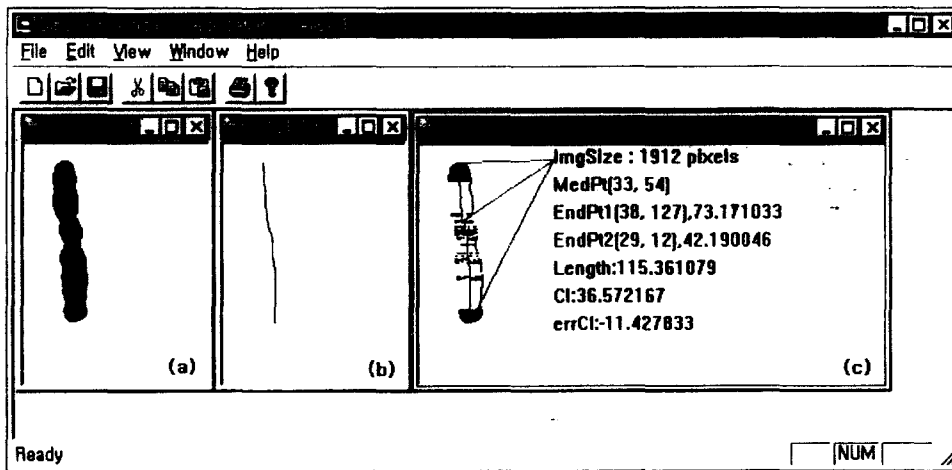


그림 1. 염색체 영상의 전처리 과정과 결과. (a) 이진화된 영상, (b) 세선화 기법을 수행한 결과, (c) 추출된 특징 파라메타 및 영상  
Fig. 1. Chromosome image preprocessing and result. (a) binary image, (b) result of thinning method, (c) extracted characteristic parameter and image

컴퓨터에 비트맵(bitmap : bmp) 파일로 저장된 염색체의 영상화일로부터 개개의 염색체를 분리하기 위하여 일반적으로 영상처리 분야에 많이 사용되는 포토샵(photoshop)을 이용하였으며, 처리과정은 다음과 같다. 염색체 영상의 전체 농도분포를 나타내는 히스토그램(histogram)을 구한 후 문턱치 처리를 수행하여 이진화된 염색체 영상과 배경을 구분한다. 그리고 염색체 영상의 배경을 제거하고 각 염색체의 크기와 형태에 적합한 창들을 설정한 후 창 내의 데이터들을 각 염색체 영상의 부화일(subfile)로 구성하고 이를 저장한다.

## 2. 염색체의 특징 파라메타 추출

패턴 인식에서 특징 파라메타의 추출 및 선택은 패턴 표현 공간을 줄이고 분류의 정확도를 향상시키는데 많은 영향을 준다. 따라서, 적절하고 정확한 특징 파라메타의

선택만이 기본적인 인식율의 향상을 얻어낼 수 있으며, 많은 수의 특징 파라메타들을 사용함으로써 발생할 수 있는 복잡도(complexity)의 감소를 위하여 불필요한 특징들의 제거 작업이 필요하다.

본 연구에서는 ISCN[2]과 HC(human cytogenetics) [8]에서 제시한 3가지 특징 파라메타들을 추출하였고, 이를 이용하여 패턴 인식 및 분류를 수행하였다. 표 2는 본 연구에서 사용한 특징파라메타들의 계산방법을 나타내었다[7].

하나의 염색체 군을 형성하는 마흔여섯 개의 각 염색체들을 창으로 구분한 후, 그림 1과 같은 방법을 적용하면 각 염색체의 특징 파라메타들을 구할 수 있다. 이러한 방법에 의하여 추출된 각 염색체의 특징 파라메타들을 제안된 HANNC의 입력패턴으로 사용한다.

그림 1에 의해 수행된 460개의 염색체 영상에서 추출

표 3. 추출된 염색체의 동원체지수, 상대길이비, 그리고 상대면적비

Table 3. Extracted centromeric index, relative length and relative area ratio of the chromosome

염색체 번호	CI			RL(%)			RA(%)		
	ISCN	추출	오차	ISCN	추출	오차	ISCN	추출	오차
1	48.0	48.40	±0.55	9.08	9.23	±0.28	8.44	8.94	±0.16
2	38.1	39.89	±0.38	8.45	8.23	±0.29	8.02	7.79	±0.34
3	45.9	47.51	±0.93	7.06	7.02	±0.22	6.83	7.03	±0.28
4	27.6	27.99	±0.67	6.55	6.44	±0.23	6.30	6.81	±0.32
5	27.4	27.59	±0.74	6.13	5.99	±0.20	6.08	6.08	±0.38
6	37.7	35.57	±1.19	5.84	5.97	±0.19	5.90	5.87	±0.29
7	37.3	37.86	±0.29	5.28	5.24	±0.18	5.36	5.38	±0.34
8	35.9	32.51	±1.34	4.96	4.76	±0.19	4.93	5.01	±0.29
9	33.3	32.31	±0.84	4.83	4.41	±0.15	4.80	4.77	±0.31
10	31.2	33.77	±0.81	4.68	4.67	±0.17	4.59	4.85	±0.25
11	35.6	37.68	±1.37	4.63	4.63	±0.16	4.61	4.83	±0.37
12	30.9	29.43	±0.29	4.46	4.59	±0.15	4.66	4.46	±0.45
13	14.8	14.77	±0.45	3.64	3.76	±0.14	3.74	3.85	±0.37
14	15.5	13.24	±0.67	3.55	3.64	±0.13	3.56	3.57	±0.29
15	14.9	14.82	±0.48	3.36	3.52	±0.12	3.46	3.38	±0.33
16	40.6	38.11	±0.54	3.23	2.85	±0.11	3.36	2.86	±0.19
17	31.4	30.06	±1.11	3.15	2.99	±0.09	3.25	2.65	±0.23
18	26.1	27.30	±0.97	2.76	2.84	±0.08	2.93	2.87	±0.32
19	42.9	43.50	±0.38	2.52	2.37	±0.09	2.67	2.32	±0.18
20	44.6	43.61	±0.75	2.33	2.31	±0.07	2.56	2.52	±0.22
21	25.7	31.99	±0.47	1.83	2.06	±0.08	1.90	1.82	±0.17
22	25.0	33.22	±0.76	1.68	2.14	±0.07	2.04	2.25	±0.24
X	36.9	38.24	±0.59	5.80	5.31	±0.26	5.12	5.63	±0.32
Y	16.3	16.07	±0.67	1.96	2.22	±0.13	2.15	1.90	±0.31

한 3가지의 형태학적 특징 파라메타를 정리한 결과는 표 3과 같다. 추출된 특징 파라메타들의 정확도를 측정하기 위하여 전체 평균값을 구하였으며, 동원체지수와 상대길이비는 1986년 ISCN에서 발표한 자료와 비교하였고, 상대면적비는 HC의 데이터와 비교하였다.

### 염색체의 핵형 분류를 위한 신경회로망

#### 1. 염색체의 핵형 분류를 위한 HANNC 구현

신경회로망을 사용하여 패턴분류를 수행하는 경우 단점으로 지적되는 것이 신경회로망의 학습에 상당히 많은 시간이 소요된다는 점이다. 또한 더욱 정확한 분류를 위하여 많은 입력 파라메타들을 고려하여야 하는데 이는 학습 소요시간의 증가를 가져오게 되며, 불필요한 파라메타에 대한 선별이 이루어지지 않아 비선형성이 증가되어 오분류할 가능성이 높아진다.

Lerner[9]등은 5개의 다른 형태의 염색체를 사용하여,

piper[5]등이 사용한 MAT 방법을 이용하여 농도파일(density profile : d.p.)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.+길이의 3가지로 입력패턴을 구성하여 다층 신경회로망을 사용하여 패턴분류기를 구성하는 연구를 수행하였으며, 추출한 특징파라메타의 약 70%로서 최종적인 분류수행이 가능하다고 발표하였고, 분류오차에 대한 언급은 없었다. 본 연구에서는 다층 신경회로망과 학습시간을 비교하기 위하여 입력층 뉴런 3개, 은닉층 뉴런 30개, 출력층 뉴런 24개로 다층신경회로망을 구성하여 동일한 특징 파라메타로 학습한 시간은 약 10시간 정도가 소요되었고, 평균 인식율은 약 42%(평균오차 58%)에 그쳤다. 따라서 본 연구에서는 적은 수의 입력 파라메타를 사용하면서 염색체의 핵형 분류에 적합한 신경회로망을 구현하기 위하여 계층적인 구조(hierarchical structure)를 지니는 HANNC를 제안한다.

표 4. 염색체의 군과 각 염색체의 번호

Table 4. Chromosome group and each chromosome number

염색체 군	염색체 번호	염색체 군	염색체 번호
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		

본 연구에서 제안한 HANNC는 학습에 소요되는 시간을 최소화하면서 염색체의 형태구조학적인 특징에 의하여 개개의 염색체를 분류할 수 있도록 계층적인 구조를 지니는 두 단계의 신경회로망으로 구성하였다. 첫 번째 신경회로망(HANNC1)은 염색체 군을 염색체의 형태구조학적인 특징에 의해 일곱 개의 군으로 분류하기 위하여 구성하였다. 참고로 ISCN에서 제안하는 염색체의 군들과 각 군에 포함되는 개개의 염색체들은 표 4와 같다. HANNC1의 주요 기능은 다음과 같다. 영상 전처리 과정과 세션화 기법에 의해 추출된 염색체 특징 파라미터들을 HANNC1의 입력벡터로 사용하여 해당되는 염색체의 군은 1.0을 출력하고, 해당되지 않는 염색체의 군은 0.0을 출력하도록 교사신호를 구성하여 염색체의 특징 파라미터에 관련되는 염색체의 군을 선택하도록 하였다. 두 번째 신경회로망(HANNC2)은 HANNC1에 의해 구분된 일곱 개의 군 내에 존재하는 각각의 염색체를 분류하기 위하여 구성하였다. 즉, HANNC1의 출력에 의해 선택된 염색체의 군에 염색체의 특징 파라미터를 입력하여 해당되는 염색체는 1.0을 출력하고, 해당되지 않는 염색체들은 0.0을 출력하도록 교사신호를 구성하였다.

2. 계층적 신경회로망(hierarchical neural network)

본 연구에서 제안한 HANNC는 계층적인 구조를 지니는 두 단계의 신경회로망 즉, HANNC1과 HANNC2로 구성되었으며, 개략적인 구성도는 그림 2와 같다.

그림 2의 첫 번째 신경회로망에 해당하는 HANNC1의 구성은 그림 3과 같으며 간략한 설명은 다음과 같다. HANNC1의 입력층(input layer)에 입력되는 파라미터는 영상 전처리 과정과 세션화 기법에 의해 추출된 C.I., R.L., 그리고 R.A. 등의 염색체 특징 파라미터이므로 이러

한 특징 파라미터들을 HANNC1의 입력 벡터로 사용할 수 있도록 세 개의 뉴런으로 구성하였다. HANNC1의 중간층(middle layer)의 뉴런의 수는 사용자에게 의해 임의로 조정 가능하도록 N개로 구성할 수 있다. 본 연구에서는 적절한 중간층 뉴런의 개수를 설정하기 위하여, 학습 상수를 고정시킨 뒤에 중간층 뉴런의 수를 가변하여 보았을 경우 최적이라고 판단되는 서른 개의 뉴런으로 구성하였다. 그리고 HANNC1의 출력층은 염색체의 군을 ISCN에서 제안한 형태구조학적인 특징으로 구분 즉, 일곱 개의 군으로 분류하기 위하여 일곱 개의 뉴런으로 구성하였다.

HANNC1의 출력에 의해 구분된 일곱 개의 군 각각에 염색체의 특징 파라미터를 입력하여 각 염색체 군에 속하는 개개의 염색체를 분류하는 HANNC2의 구성은 그림 4와 같다. HANNC2의 입력층은 HANNC1의 출력에 연결될 수 있도록 일곱 개의 뉴런층으로 구성하였다. 그리고 각 뉴런층에 입력되는 파라미터는 네 개의 파라미터 즉, 세 개의 특징 파라미터(C.I., R.L., R.A.)와 염색체 각각의 전체 길이(L)를 나타내는 파라미터이며, 이러한 특징 파라미터들을 입력벡터로 사용하기 위하여 네 개의 뉴런으로 구성하였다. 왜냐하면 염색체를 보다 정확하게 분류할 수 있는 계층적인 구조를 지니는 분류기를 구성하기 위해서는 HANNC1과 HANNC2에서 사용하는 염색체의 특징 파라미터들이 최소한 한 개 이상이 다르든지 또는 전부가 서로 달라야 되기 때문이다.

HANNC2의 중간층은 사용자에게 의해 임의로 조정 가능하도록 N층으로 구성되나, 본 연구에서는 염색체의 형태구조학적인 특징을 감안하여 최소 열 개의 뉴런에서부터 최대 서른 개의 뉴런까지 지정할 수 있도록 구성하였다. 그리고 HANNC2의 출력층은 일곱 개의 군에 포함되어 있는 스물네 개의 염색체를 분류하기 위하여 일곱 개의 뉴런군으로 구성하였으며, 각 뉴런군은 표 2에 정의된 바와 같이 염색체의 각 군별로 분류되는 염색체의 수 만큼 뉴런층으로 구성하였다.

이와 같은 방법으로 구성된 HANNC를 학습시키기 위하여 여러 가지 파라미터들을 사용하여야 한다. 왜냐하면 뉴런과 뉴런을 연결하는 가중치들의 초기값(initial weighted value), 활성화함수의 기울기(steeptness of activation

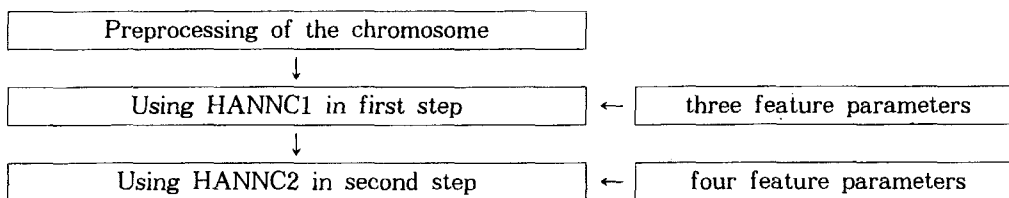


그림 2. 염색체 분류를 위하여 제안한 계층적 인공 신경회로망  
Fig. 2. Proposed HANNC for the chromosome classification

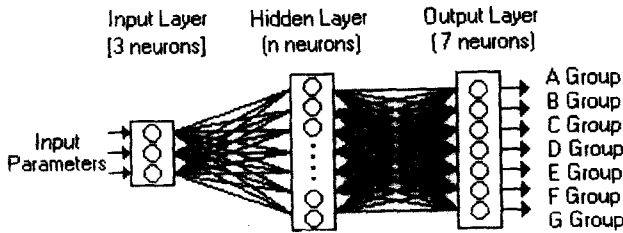


그림 3. 염색체 군을 분류하기 위하여 제안된 HANN1  
 Fig. 3. Proposed HANN1 for chromosome group classification

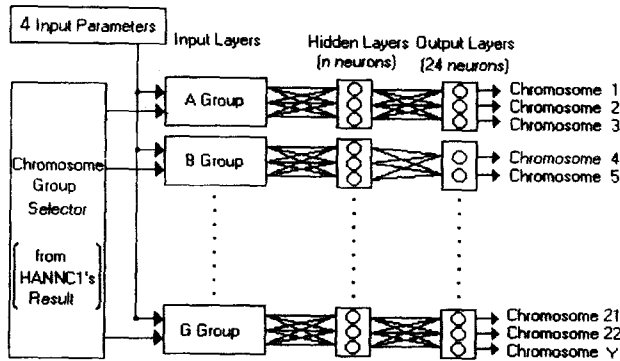


그림 4. 염색체 핵형을 분류하기 위하여 제안된 HANN2.  
 Fig. 4. Proposed HANN2 for chromosome karyotype classification

function), 학습상수(learning constant), 그리고 중간층을 구성하는 뉴런의 개수 등이 HANN1의 학습에 큰 영향을 미치기 때문이다[10]. 따라서 염색체의 특징 파라미터들을 입력벡터로 사용하여 염색체의 핵형을 분류하는 최적의 HANN1을 구현하기 위하여 이러한 파라미터들의 값을 적절히 조정하여야 한다.

본 연구에서는 오차역전파(error backpropagation) 학습 알고리즘을 사용하여 HANN1을 학습하였다. 그리고 HANN1의 학습에 소요되는 시간이 단축시키며, HANN1이 국부적인 극소점(local minimum)에 빠지지 않도록 하기 위하여 모멘트법과 적응학습법을 사용하였다[11]. 이러한 방법들을 사용하여 HANN1의 학습에 필요한 여

러 파라미터들의 적절한 값과 중간층을 구성하는 뉴런의 적절한 개수를 구하기 위하여 여러 가지 실험을 수행한 결과 표 5와 같은 파라미터들을 얻을 수 있었다. 표 5에서 정의한 파라미터들을 사용하여 학습을 수행한 HANN1에 전처리 과정과 세션화 기법에 의해 추출된 염색체의 특징 파라미터를 입력하면 염색체의 형태구조에 적합한 핵형이 분류되도록 하였다.

### 실험결과 및 고찰

본 연구에서 염색체의 형태구조학적인 특징을 고려하여 핵형을 분류하기 위해 구성한 HANN1은 C 언어를 사용하여 구현되었으며, 이를 Pentium 90 컴퓨터에서 시뮬레이션 하였다. 구현된 HANN1은 각 층의 뉴런수, 활성화함수의 기울기, 학습 상수, 가속 상수, 입력 패턴 수, 학습 회수 및 최대 허용 오차 등의 파라미터들을 사용자가 키보드를 통하여 입력할 수 있도록 하여 학습에 유연하게 대처할 수 있게 하였다. 또한 학습이 수행되는 과정을 가시화하기 위하여 학습시 발생하는 오차의 변화를 그래프로 디스플레이되도록 하였으며, 추출된 결과들은 표 형태로 구성된 파일에 저장하였다.

그림 5는 제안된 HANN1에 의해 염색체를 일곱 개의 군으로 분류하기 위하여 HANN1에 염색체의 특징 파라미터들을 입력한 후 학습을 수행시키는 과정에서 발생하는 자승오차의 합(sum-squared error : SSE)을 그래프로 나타내었다. SSE의 변화는 선형적인 형태를 나타내고 있으며, 학습의 수행 회수가 증가하면 할수록 SSE가 계속 감소하는 경향을 보이고 있다.

그림 6은 염색체의 일곱 개 군들 중 A군의 염색체를 분류하기 위하여 제안된 HANN2에 A군의 염색체 특징 파라미터들을 입력한 후 학습을 수행시키는 과정에서 발생하는 SSE의 변화를 그래프로 나타내었다. SSE의 변화가 학습의 초기에는 상당히 완만한 감소 곡선으로 나타나다가 약 80회 정도의 학습이 수행된 후와 약 180회 정도의 학습이 수행된 이후부터 갑자기 급격하게 감소되었다. 이와 같은 현상은 제안된 HANN2의 결합계수들의

표 5. 제안된 HANN1의 학습에 사용된 파라미터의 값과 뉴런의 수  
 Table 5. Parameter values and number of neuron for learning in proposed HANN1

Neural Network	구분 군별	HANN2							
		HANN1 전체	Group A	Group B	Group C	Group D	Group E	Group F	Group G
steepness of activation function		1.0	1.2	1.1	1.0	0.9	0.8	1.0	1.0
learning constant		0.05	0.05	0.05	0.05	0.1	0.5	0.05	0.1
Num. of neuron in hidden layer		30	10	10	30	15	15	10	10

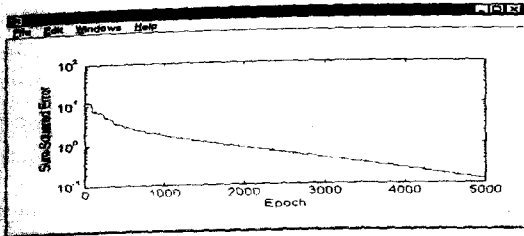


그림 5. 염색체 군에 대한 특징 파라메타들의 학습 중 SSE의 변  
화 과정  
Fig. 5. Variation SSE on feature parameters learning for  
chromosome groups

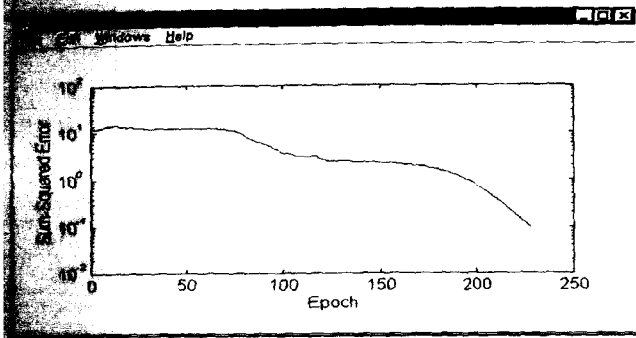


그림 6. A군의 염색체 특징 파라메타들의 학습 중 SSE의 변  
화 과정  
Fig. 6. Variation SSE on feature parameters learning for  
group chromosome

이 국부적인 최소점에 머무르다 최적점(global minimum)을 찾아가는 과정임을 알 수 있다. 그림 5와 그림 6에서 SSE가 적으면 적을 수록 학습이 잘 수행되었음을 알 수 있으며, 학습의 수행 회수(epoch)가 적으면 적을 수록 학습에 소요된 시간이 적음을 알 수 있다.

그림 5와 같이 학습된 HANNC를 패턴 분류기로 사용하여 염색체의 핵형 분류를 수행하였다. 핵형분류에 사용된 학습시에 사용되지 않은 새로운 염색체들의 특징 파라메타이며, 임상적으로 정상으로 판명된 5명분을

학습된 HANNC1을 사용하여 정상인 5명의 염색체 군을 분류한 결과

표 6. The result of chromosome group classification for 5 sample images by means of learned HANNC1

Classification table of the chromosome group									
Group	A	B	C	D	E	F	G	Total	Error Rate(%)
A	30	0	0	0	0	0	0	30	0
B	0	19	1	0	0	0	0	20	5
C	0	0	75	0	0	0	0	75	0
D	0	0	2	28	0	0	0	30	6.67
E	0	0	0	0	30	0	0	30	0
F	0	0	0	0	1	19	0	20	5
G	0	0	0	2	0	0	23	25	8
합계								230	2.61

사용하였다.

표 6은 HANNC1을 사용하여 각 군별 분류를 수행한 결과이다. 여기서 A-G는 개별 염색체 그룹을 나타내며, 세로축의 Total은 분류하여야 할 염색체 수를 나타내며, 가로축의 채색된 부분이 오분류한 염색체의 수를 나타낸다. 오차율은 오분류한 정도를 퍼센트 오차로 나타내었다. 예를 들어 그룹 B의 경우는 분류하여야 할 염색체의 총 수가 20개인데, 이들 중 19개는 바른 분류를 하였으나 1개는 그룹 C로 오분류하였으며 이에 대한 오차율은 5%이다. 그룹 분류시에 나타난 총 오차율은 2.61%이며, 학습에 소요된 시간은 2시간이다.

표 7은 HANNC2를 사용하여 개개의 염색체들의 분류를 수행한 결과를 나타낸 것이다. 여기서, 1-22, X, Y는 개별 염색체들의 번호를 나타내며, 나머지는 표 6과 같다. 예를 들어 4번 염색체의 경우는 분류하여야 할 염색체의 총 수가 10개인데, 이들 중 9개는 바른 분류를 하였으나 1개는 5번 염색체로 오분류하였으며 이에 대한 오차율은 10%이다. 개개의 염색체 분류시에 나타난 총 오차율은 7.39%로 나타났으며, 2시간 35분 동안 학습하였다.

## 결 론

본 연구는 패턴인식 분야에서 많이 연구되고 있는 인공 신경망을 이용하여 염색체의 영상을 구분하고, 염색체의 핵형 분류에 적합한 패턴 분류기의 구성에 연구의 중점을 두었다. 염색체는 핵형분류도에 의하여 염색체 군으로 나누고 염색체 군 내에 개별 염색체들로 구분되어 각종 유전 질환 및 진단에 사용되고 있다. 이러한 핵형분류도에 착안하여 본 연구에서는 계층적 인공신경회로망 분류기(HANNC)를 제안하여 염색체의 핵형 분류기로 사용하였다. 영상 전처리기법을 사용하여 염색체의 형태학적인 특징 파라메타를 연속적으로 추출할 수 있도록 하여 10명에 대한 특징 파라메타를 추출하였고 이를 HANNC의 학

표 7. 학습에 이용되지 않은 정상인 5명의 염색체를 학습된 HANNC2에 적용하여 염색체 핵형을 분류한 결과  
 Table 7. The result of each chromosome karyotype classification for 5 unknown sample images by means of learned HANNC2

Type	The Number of chromosome (the case of unknown chromosome classification)																						Total Error	% Error				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22			X	Y		
1	10																								0	0		
2		10																								0	0	
3			10																							0	0	
4				9	1																					1	10	
5					10																					0	0	
6					1	9																				1	0	
7							10																			0	0	
8							1	7	1		1															3	30	
9									8	2																2	20	
10										10																0	0	
11											10															0	0	
12										1		9														1	10	
13								1			1		8													2	20	
14														10												0	0	
15															9	1										1	10	
16																10										0	0	
17																	9		1							1	10	
18																		10								0	0	
19																			1	8	1					2	20	
20																					10					0	0	
21																						9	1			1	10	
22																							10			0	0	
X											1	1												3		2	20	
Y																									5		0	0
Total Error	0	0	0	0	2	0	1	1	1	3	2	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	17	7.39

습 및 분류에 이용하였다.

본 연구에서 구성한 신경회로망을 학습시키기 위하여 5명분의 특징 파라메타를 입력패턴으로 사용하였다. 학습된 신경회로망을 패턴 분류기로 사용하여 다른 5명분의 염색체를 분류하였다. 염색체 군의 분류 오차는 2.61%이며, 전체 염색체의 분류 오차는 7.39%를 나타내었다. 이 결과는 현재까지 발표된 염색체 패턴 분류기의 분류 오차보다 약간 높은 수치이다. 그러나 적은 수의 입력 파라메타들을 가지고 분류할 수 있는 가능성을 발견할 수 있었으며, 또한 학습에 소요된 시간도 상당히 단축되었다. 분류오차가 타 연구와 비슷한 이유는 염색체 영상을 전처리 하는 과정에서 특징 파라메타의 추출이 용이하지 않은 몇몇 염색체에 의한 것으로 생각되며, 또 신경회로망의 학

습에 사용된 염색체의 특징 파라메타들의 수가 너무 작기 때문인 것으로 판단된다. 따라서, 분류 오차를 감소시키기 위하여 다양하면서 적절한 특징 파라메타의 추출이 추가될 필요가 있으며, 염색체 원 영상의 추출과 전처리 과정에서 특징 파라메타 추출의 정확도를 높여야 할 것이다.

참 고 문 헌

1. J. H. Tijo, and A. levan, "The chromosome number of man", Hereditas, vol.42, pp. 1-16, 1956.
2. An International System for Human Cytogenetic Nomenclature(ISCN), KARGER, 1985.



3. Robert S. Ledly, "High-speed automatic analysis of biomedical picture", Science, vol. 146, pp. 216-223, 1964.
4. Peter W. Neurath, Barkev L. Bablouzian, Tom H. Warms, and Russel C. Serbagi, "Human chromosome analysis by computer-an optical pattern recognition problem", Annals New York Academy of Sciences, vol. 128, pp. 1013-1028, 1965.
5. J. Piper, and E. Granum, "On fully automatic features measurement for banded chromosome classification", Cytometry, Vol. 10, pp. 242-255, 1989.
6. Brian H. Mayall, James D. Tucker, Mari L. Christensen Lucas J. van Vliet, and Ian T Young, "Experience with the athens semi-automated karyotyping system", Cytometry, vol. 11, pp. 59-72, 1990.
7. 장용훈, 이권순, 정형환, 엄상희, 최옥환, 전계록, "염색체 영상의 재구성에 의한 형태학적 특징 파라메타 추출", 대한의공학회지, Vol. 17, No. 4, pp. 545-552, 1996.
8. D. E. Rooney, B. H. Czepulkowski, "Human cytogenetics essential data", John Wiley & Sons, 1994.
9. 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.
10. Jacek M. Zurada, Introduction to artificial neural systems, Info Access Distribution, 1992.
11. B. Widrow, and R. Winter, "Neural nets for adaptive filtering and adaptive pattern recognition", IEEE Computer, pp. 25-39, march, 1988.
12. Rafael C. Gonzalez and Paul Wintz, Digital image processing, Addison-Wesley, pp. 398-403, 1987.
13. Abhihit S, Pandya, and Robert B. Macy, Pattern recognition with neural networks in C++, CRC Press, 1995.
14. K. R. Jun, S.H. Eom, Y.H.Chang, "Optimal Neural Network Classifier for Chromosome Karyotype Classification", AI, Simulation and Planning in High Autonomy Systems, vol. 1, pp. 315-322, March, 1996