

병리진단을 위한 웹기반 혈액영상 분석시스템의 설계 및 구현

김경수*, 이영신**, 김용국***, 이운배***, 김판구***

*메디 정보

**서강 정보 대학

***조선대학교 전자계산학과

Design and Implementation of Web-Based Blood-Cell Analysis System for Pathology Diagnosis

Kyung-Su Kim*, Young-Sin Lee**,

Yong-Kuk Kim***, Yoon-Bae Lee***, Pan-Koo Kim***

*Medi Information

**Seogang Information College

***Dept of Computer Science, Chosun University

요 약

의학분야에서 컴퓨터 활용은 단순히 처리할 데이터의 자동화뿐만 아니라 각종 의학영상들을 자동으로 처리함으로써 의사의 진단을 도와주는 형태로 발전되어 가고 있다. 본 논문에서는 병원의 임상병리과에서 빈번히 수행하는 혈액검사를 자동화하기 위한 것으로 혈액을 자동 분석하는 웹 기반 분석시스템을 구축하였다. 이를 위해 본 논문에서는 혈액 영상으로부터 특징을 추출하기 위한 단계를 서술하고 세포분류를 위한 다층 신경망을 이용해 구현한 내용을 보인다. 또한 본 연구의 결과로 신경망의 학습 효율을 높이기 위한 전처리로서 학습 데이터에 대해 러프 집합 이론을 적용하여 학습 데이터의 차원을 효과적으로 줄일 수 있었다.

1. 서론

최근 초고속정보통신망의 구축과 함께 화상처리기술의 발전, 컴퓨터 그래픽기술의 발전으로 인해 개인병원이나 종합병원 등에서 생성되는 화상들을 디지털화하여 처리하려는 노력들이 활발히 진행되고 있다. 그 예로서 80년대 초부터 필름없는 병원을 목표로 PACS(Picture Archiving and Communication Systems), 원격 의료 진단시스템등이 개발되어 오고 있고, 다양한 의료장비들을 상호 연결하는데 필요한 공통의 통신프로토콜을 위해 1988년에 제안되기 시작하여 1993년에 상당부분 완성된 DICOM(Digital Imaging and Communications

in Medicine) 표준 Version 3.0이 있다[1][6].

병원에서 생성되는 디지털 영상정보의 제조, 전달, 저장, 검진 등에 현재의 컴퓨터 및 통신기술을 적용하여 자동화함으로써 얻을 수 있는 잇점은 다양하다.

이와 마찬가지로 늘어가는 환자들과 건강 검진자들에게 혈액진단의 효용성을 높이고 보다 정확하고 신속한 검진결과를 제공하기 위해서는 지금과 같이 수동적으로 처리하는 것보다는 전문가의 고도의 경험이 필요치 않은 부분에 대해서는 진단자동화를 도입할 필요성이 있다. 또한 국내외적으로 아직 개발된 사례가 없는 혈액 병리진단·검색시스템의 개발은 높은 부가가치를 창출

할 수 있다고 생각한다. 이를 위해 본 논문에서는 현미경으로부터 획득된 영상을 초고속정보통신망을 이용하여 원거리의 혈구분석 서버로 전송하면 혈구세포들의 병적 상태에 대한 자동분석 결과를 다시 클라이언트에 게 리포트 해주는 원격 혈액영상 분석 시스템을 구현하였다.

2. 혈구세포에 대한 소개

말초혈액은 액체성분인 혈장(plasma)과 적혈구, 백혈구, 혈소판 같은 세포성분으로 구성되어 있다. 이 모든 세포성분은 미분화된 간세포인 혈구조작아구에서 유래된다. 이 초기 원시세포들은 조혈 조직내에서 혈구아구(haemocytoblast)로 분화되고 분지계(clone)로 나누어져 적혈구계, 골수계 그리고 거핵구계열의 세포로 성숙되는데 이들 세포 계열들이 위치하는 조직에 따라서 전적아구(rubriblast), 골수아구(myeloblast), 림프아구(lymphoblast), 거핵아구(megakaryoblast), 단아구(monoblast), 형질아구(plasmoblast)와 같은 특이한 혈액 세포들이 생성된다. 이 중 형질아구는 증식과정을 거쳐 최종적으로 혈장이 되고, 전적아구는 적혈구가 되고, 거핵아구는 혈소판을 이루게 된다. 그리고, 단아구는 단핵구가 되고, 림프아구는 림프구, 골수아구는 전골수구를 거쳐 호염기성, 호중성, 호산성으로 분화되고 최종적으로 호염기성 분핵구, 호중성 분엽핵구, 호산성 분엽핵구가 되어 백혈구가 된다[10][11]. 본 논문에서는 적혈구를 분리하기 위한 시스템을 소개하고 정상적혈구, 툽니모양, 유구 적혈구, 표적 적혈구, 구상적혈구, 귀리모양, 겸상적혈구 등 적혈구를 24부류로 분류하였다[10].

3. 혈구영상 분석

본 논문에서는 주로 성숙된 적혈구 영상을 대상으로 적혈구의 병적 상태를 분류하고자 하였다.

3.1 전처리 단계

혈액영상을 자동 분석·진단을 하기 위해서는 먼저 현미경으로부터 혹은 혈액 영상을 스캔한 영상을 입력받는다. 이 때 사용자는 인터넷을 통해 사용자가 혈액 영상을 웹 서버로 업로드하게 된다. 이때 사용된 영상은 현미경을 통해 400배 확대된 주로 적혈구로 구성되어 있는 영상을 사용하기로 한다. 입력된 영상은 컬러 혹은 흑백 영상이 될 수 있는데 컬러 영상인 경우 흑백 영상으로 변환시켜 처리한다. 본 논문에서 입력된 영상에 대해 가장 먼저 하게되는 처리는 혈구를 분리하기 위한 전처리로 적당한 임계치를 결정하여 이진화 영상으로 변환시킨다. 본 논문에서는 혈액영상 이진화를 위해 적용한 원래의 그레이 영상과 이진화된 영상과의 차

이를 나타내는 퍼지 척도를 이용한 방법을 이용하였다.[2]



그림 1. 입력 영상

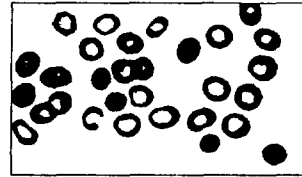


그림2. 이진화된 영상

다음 단계에서는 세그먼트된 혈구세포들에 대해 레이블링을 하게 된다. 일반적으로 레이블링 알고리즘은 크게 Local Neighbourhood 알고리즘과 Divide and Conquer 알고리즘의 두 부류로 나눌 수 있는데 본 논문에서는 Local Neighborhod 알고리즘에 속하는 대표적인 알고리즘은 Iterative Local Operation을 수행하는 알고리즘이다. 그 알고리즘은 이진 영상에서 객체 경계의 첫 화소가 발견될 때까지 영상을 열우선으로 스캔한다. 현재 화소로부터 8방향 이웃화소에 속하는 모든 화소로 전파되는 그 화소에 하나의 소위 'fire'가 세팅된다. 이러한 연산은 영상내의 객체의 모든 화소들이 타고나서(burnt) 불(fire)이 꺼질 때까지 재귀적으로 반복된다. 아래 그림 3은 레이블링된 결과를 나타내고 있다 [3].

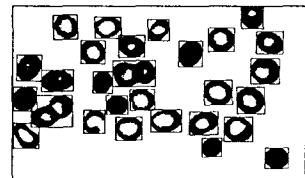


그림 3 레이블링된 영상

3.2 혈구분리단계

레이블링 단계를 거친 후 각각 분리된 혈구세포들 중에 일부는 분리되어 있지 않고 두 개 이상의 세포들이 뭉쳐있는 경우가 생기게 되는데 이는 혈액 도말 슬라이드를 만드는 중에 생기는 것이다. 그런 경우 실제로 임상병리사가 뭉쳐있는 각각의 혈구세포들의 이상유무를

경험적으로 판단하여 분석을 하게 된다. 그러나, 자동진단을 위해서는 각각의 혈구세포들을 분석하기 위해 놓쳐있는 혈구세포들을 분리하는 작업이 필수적이다. 본 논문에서는 지금까지 제안된 형태분할 알고리즘 중에서 비교적 성능이 좋은 Vincent에 의해 제안된 Watershed 알고리즘을 이용하여 형태분할을 수행하였다[5][12].

3.4 인식단계

인식단계에서는 각각의 분리된 적혈구 세포들로부터 이상유무를 구별하고 이상(abnormal) 세포일 경우 24부류의 이상 세포 중에 어떤 부류에 속하는지를 구별하는 단계이다.

3.4.1 특징 추출

세그먼트된 각각의 혈구영상으로부터 신경망 분류기를 학습시키기 위한 학습데이터 추출작업이 필요하다. 일반적으로 화상인식을 위한 학습데이터는 객체로부터 어떠한 공통적인 특징을 추출하여 신경망을 통해 학습시키는 경우와 특징을 추출하지 않고 화상 그 자체를 직접 입력으로 넣어 학습시키는 두 가지 경우로 나눌 수 있다. 본 연구에서는 세그먼트된 혈구세포들의 최소 경계선 박스를 찾아 8×6의 영상으로 재스케일링하여 0.0과 1.0 사이의 값으로 구해 48개의 값을 신경망의 입력으로 사용하였다.

3.4.2 신경망 구조

본 연구에서 적혈구를 분류하기 위해 사용한 신경망 구조는 다음과 같다[4]. 먼저 각각의 혈구세포들에 대해 정규화과정을 통해 얻어진 48개의 특징 데이터를 입력 노드에 입력하고 96개의 은닉층을 거쳐 24부류의 정상 세포를 포함한 이상 적혈구를 분류한다.

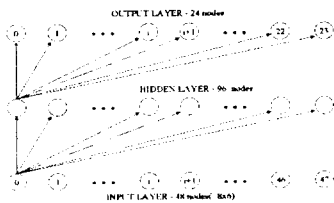


그림 4 적혈구 분류를 위한 다단계 신경망 구조

4. 러프집합을 이용한 전처리

신경망의 학습에 사용하게 될 학습데이터의 크기를

인식율에 영향을 주지 않은 채로 줄이게 되면 네트워크의 크기를 줄이게 되어 학습시간을 단축시킬 수 있다. 따라서, 러프집합의 개념을 적용하여 학습데이터 집합을 지능적으로 줄임으로서 중요한 정보를 추출할 수 있을 뿐만 아니라 학습의 속도를 느리게 하고 네트워크의 성능을 저하시키는 속성들간의 상관관계를 줄일 수 있다[7].

4.1 러프집합의 개념

Pawlak에 의해 제안된 러프 집합은 집합의 대략 포함으로 설명될 수 있고, 모호성(vagueness)과 불확실성(uncertainty)에 대한 새로운 수학적 접근 방법이다[8]. 러프 집합은 자동 분류, 패턴 인식, 학습 알고리즘 등 다양한 분야에 적용될 수 있다[9]. 본 논문에서는 혈구세포의 자동분류를 위한 학습과정의 전처리로서 러프집합 이론을 적용하고자 한다. 그림 1에서 근사 공간 A는 순서쌍 A=(U, R)로 표현되며, 여기서 U는 전체 집합이고 R은 동치 관계이다. (x1, x2)∈R이면 x1과 x2는 근사 공간 A에 식별 불가능이며, 관계 R의 동치류들을 원소 개념(elementary concept)이라 한다. A에서 임의의 원소 개념들의 유한 합집합은 A에서 정의 가능 집합(definable set)이라 한다.

러프 집합 이론에서, 주어진 지식으로 명확히 분류할 수 없는 대상들의 집합 X는 상한 근사(upper approximation : R*(X)), 하한 근사(lower approximation : R-(X)), 상한과 하한 근사는 X의 긍정영역(positive region : POSR(X)), 부정 영역(negative region : NEGR(X)), 경계 영역(borderline region : BNR(X)) 이라 불리는 러프 집합으로 다음과 같이 표현된다.

$$R^*(X) = U \{Y \in U / R : Y \cap X \neq \emptyset\}$$

$$R_-(X) = U \{Y \in U / R : Y \subseteq X\}$$

$$POS_R(X) = R_+ X$$

$$NEG_R(X) = U - R^* X$$

$$BNR(X) = R^* X - R_+ X$$

4.2 러프집합에 의한 전처리

학습에 사용된 데이터는 48개의 속성을 가지고 있었는데 10개의 속성이 줄어들어서 신경망에 학습을 위한 입력노드는 처음에 설계한 48에서 38개의 노드의 구조로 변경되었다.

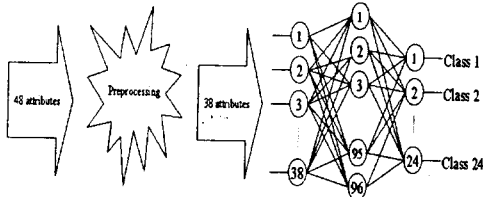


그림 5. 리프집합을 적용한 후의 신경망 구조

5. 실험 및 시스템 구성

5.1 실험

실험은 조선대학교 부속병원 임상병리과에서 얻은 환자의 혈액 샘플을 기반으로 신경망 학습패턴을 획득하였다. 훈련 테스트 셋은 교차 검증(cross-validation)을 위해 24부류의 각각에 대해 50개의 샘플 데이터를 만들었고 이들로 구성된 2400개의 학습패턴 2셋트를 만들어 사용하였다.

실험은 먼저 하나의 학습 패턴으로 신경망을 학습시키고 난 후 다른 하나의 학습 패턴으로 인식율을 계산하였다. 그리고, 또한 학습데이터와 인식 테스트용 학습패턴을 바꾸어 수행하였다. 그 실험결과는 다음과 같다. 표1과 표2에서 사용된 번호는 2절에서 서술한 적혈구의 종류를 말하고 인식된 개수는 학습된 신경망을 가지고 다른 테스트 셋을 인식했을 때의 100개씩을 가지고 있는 각각의 혈구세포들에 대해 올바르게 인식된 혈구세포의 개수이다.

표1과 표2를 통해 테스트 데이터 집합1로 학습시킨 후에 테스트 데이터 집합2를 인식한 결과가 인식율에서 약간 낮음을 알 수 있다. 그리고, 24부류중에서 형태학적인 특성이 정확히 정해져 있는 세포인 경우(정상 적혈구, 톱니 모양, 유규적혈구, 표적 적혈구, 구상 적혈구, 난형 적혈구 등등)에는 100개의 데이터 중에 80% 이상의 인식율을 보이고 있다. 하지만 기존의 자동화가 어려운 이유중의 하나인 형태학적인 특성이 정확하지 않은 혈구세포(귀리, 검상 적혈구, 결정체 혈구, 유규 적혈구 등등)들은 테스트 데이터에 따라 인식율의 차이가 크게 날 수 있음을 보여주고 있다.

5.2 시스템 구성

본 시스템의 초기 화면은 프레임으로 나뉘어진 웹 페이지 중 메뉴와 홈페이지로 나누어진다.



그림 6. 혈액 영상 분석 시스템의 초기화면

다음 그림은 일반적으로 한 환자의 혈액영상을 분석하는데 현미경을 통해 10시야의 영상을 분석하기 때문에 본 시스템에서도 10시야로부터 획득된 10개의 혈액영상을 순차적으로 서버로 보내기 위한 분석메뉴의 초기화면을 보여주고 있다.

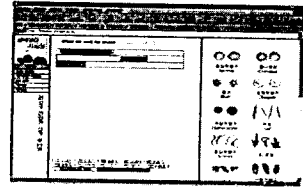


그림 7. 혈액영상 전송 메뉴

다음 그림은 10개의 혈액 영상을 업로드하고 10개의 영상이 분석되어 나온 결과를 보여주고 있다.

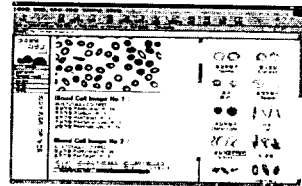


그림 8. 혈액영상 분석 결과 화면

6. 결론 및 향후 연구과제

본 논문에서는 혈액영상을 자동으로 분석하여 혈구세포들의 수를 카운트하고 이상유무를 판별하여 리포트함으로써 병리 진단에 도움을 줄 수 있는 웹 기반 시스템을 구축하였다. 본 논문에서 수행한 적혈구 인식의 결과는 조선대학병원의 임상병리실에서 충분히 참조 가능한 시스템임을 확인하였다. 그러나, 혈구세포의 변이를 인식하여 병리진단을 위한 작업은 다른 검사와의 종합적인 분석을 통해 이루어지고 있다. 따라서, 향후 다른 검사 장비와의 인터페이스 구축을 통해 얻은 결과를 종합적으로 분석해주는 진단 전문가 시스템 구축이 필요하고, 적혈구 외에 진단에 큰 영향을 주는 백혈구의 변

이 인식을 함으로서 보다 완벽한 진단 시스템의 구축이 필요하다.

참 고 문 헌

[1] Brent K. Stewart and Steve G. Langer, "Medical Image Databases and Informatics", IEEE Internation Conf. on image processing, Oct 4-7, Chicago, Illinois, pp.29-33, 1998.

[2] Huang L-K and M-J. J. Wang, "Image thresholding by minimizing the measures of fuzziness", Pattern Recognition, Vol.28, 1:41-51, 1995.

[3] Ioannis Pitas, Digital Image processing Algorithms, Prentice Hall, 1993.

[4] James A. Freeman and David M. Skapura, *Neural Networks : Algorithms, Applications and Programming Techniques*, Addison-Wesley Publishing, 1991.

[5] L. Vincent and P. J. Soille, "Watershed in Digital Spaces : An Efficient Algorithm based on Immersion Simulation", IEEE Trans. PAMI, Vol.13, No.6, pp.583-597, 1991.

[6] Michael W. Vannier and John W. Haller, "Biomedical Image Segmentation", IEEE International Conf. on image processing, Oct 4-7, Chicago, Illinois, pp.20-24, 1998.

[7] Olve Maudal, Preprocessing data for Neural Network Based Classifiers : Rough Sets vs Principal Component Analysis, Ms Thesis, Department of Artificial Intelligence, Univ. of Edinburgh, 1996.

[8] Z. Pawlak, "Rough Sets", International Journal of Computer and Information Sciences, pp.341-356.

[9] Z. Pawlak, "Rough Sets - Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[10] 고려의학 출판부편, 혈액학 도감, 고려의학, 1995

[11] 조한익 편역, 실용 혈액학, 고려의학, 1995.

[12] 최인아외 2인, "Watershed 알고리즘을 이용한 수학적 형태학에 의한 형태 분할", '93 정보과학회 가을 학술발표 논문집, vol.20, No.2, pp.345-348, 1993.

혈구분류번호	인식된 갯수
0	95
1	90
2	88
3	85
4	90
5	75
6	75
7	70
8	78
9	89
10	88
11	70
12	68
13	72
14	65
15	66
16	75
17	66
18	60
19	74
20	60
21	65
22	70
23	86

표1. 테스트 데이터 집합 1에 대한 인식 결과

혈구분류번호	인식된 갯수
0	96
1	88
2	85
3	86
4	86
5	74
6	79
7	68
8	75
9	90
10	90
11	68
12	65
13	75
14	60
15	64
16	71
17	62
18	58
19	74
20	62
21	66
22	68
23	87

표2. 테스트 데이터 집합 2에 대한 인식 결과