

임상적 의사결정지원시스템에서 순차신경망 분류기를 이용한 급성백혈병 분류기법

임선자[†], 이반빈센트^{**}, 권기룡^{***}, 윤성대^{****}

Acute Leukemia Classification Using Sequential Neural Network Classifier in Clinical Decision Support System

Seon-Ja Lim[†], Ivan Vincent^{**}, Ki-Ryong Kwon^{***}, Sung-Dae Yun^{****}

ABSTRACT

Leukemia induced death has been listed in the top ten most dangerous mortality basis for human being. Some of the reason is due to slow decision-making process which caused suitable medical treatment cannot be applied on time. Therefore, good clinical decision support for acute leukemia type classification has become a necessity. In this paper, the author proposed a novel approach to perform acute leukemia type classification using sequential neural network classifier. Our experimental result only cover the first classification process which shows an excellent performance in differentiating normal and abnormal cells. Further development is needed to prove the effectiveness of second neural network classifier.

Key words: Acute Leukemia Classification, Sequential Neural Network, Clinical Decision Support System, K-Means Clustering

1. 서 론

의료 서비스는 공공복지를 달성하기 위한 가장 중요한 측면 중 하나이다. 대부분의 병원에서 이전의 의료 서비스 시스템은 일반적으로 충분한 표준으로 간주되어 왔지만 진단 프로세스의 시간 지연 감소 및 효율성 개선에서 의료서비스를 향상시킬 수 있는 몇 가지 측면이 있다. 따라서 신뢰성 있고 시간 효율

적이며 저렴한 진단 시스템이 필요하다.

백혈구 암 또는 백혈병은 백혈구를 공격하는 가장 치명적인 질병이다. 체내에서 백혈구는 면역체계와 질병 예방을 담당한다. 혈액세포는 골수 내부에 형성된다. 세포는 초기 단계를 폭발 또는 미성숙세포라고 한다. 일반적으로 혈액세포는 신체가 요구하는 대로 조절된 방식으로 생산된다. 그러나 백혈병은 비정상적인 백혈구의 생성이 제어가 안되며 정상적인 백혈

※ Corresponding Author: Ki-Ryong Kwon, Address: (48513) 45, Yongso-ro, Nam-Gu, Busan, Korea, TEL: +82-51-629-6257, FAX: +82-51-629-6230, E-mail: krkwon@pknu.ac.kr

Receipt date: Jan. 13, 2020, Approval date: Jan. 20, 2020

[†] Dept. of Computer Engineering, Pukyong National University (E-mail: lsja76@gmail.com)

^{**} Dept. of IT Convergence and Applications Eng., Pukyong National University (E-mail: Ivan.vincent@gmail.com)

^{***} Dept. of IT Convergence and Applications Eng., Pukyong National University

^{****} Dept. of Computer Engineering, Pukyong National University (E-mail: sdyoun@pknu.ac.kr)

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2016R1D1A3B03931003, No. 2017R1A2B2012456) and the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2020-2016-0-00318) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and Ministry of Trade, Industry and Energy for its financial support of the project titled "the establishment of advanced marine industry open laboratory and development of realistic convergence content"

구와 적혈구 및 혈소판과 같은 기타 혈액 성분의 활동을 억제한다

백혈병 분류의 경우, 컴퓨터 비전 기반 방법은 진단 시스템 성능을 향상 시키는 대체방법 중 하나로 간주되었다. 본 논문에서는 백혈병 질병의 우수한 분류 결과를 얻기 위해 일부 이미지처리 방법론과 함께 컴퓨터 비전 기반 알고리즘이 적용된다.

백혈병은 그 심각도 수준에 따라 두 가지 유형으로 구분된다. 첫째 유형은 급성백혈병으로, 환자 상태의 급속한 악화 상태로 파악 가능하며 통계적으로 대부분의 환자는 치료가 제대로 되지 않으면 약 6개월 동안 증상이 지속된다. 둘째 유형은 만성 백혈병으로, 진단 후 약 2년이 지나고 치료를 받지 않으면서 점차 악화되는 경우이다.

각 백혈병 유형은 영향을 미치는 백혈구 유형에 따라 2가지 등급으로 분류 될 수 있다. 그들은 림프구가 있고 골수성 백혈병이다. 골수성 백혈병에서 적혈구, 백혈구 및 혈소판으로 성숙하는 골수세포에서 비정상적인 세포 성장이 발생한다. 림프구 백혈병에 있는 동안 림프구가 되는 골수세포에서 비정상적인 세포 성장이 발생한다. 이 매개변수를 기반으로, 프랑스-미국-영국 (FAB) 분류에 의해 결정되는 4가지 광범위한 백혈병 분류는 급성골수성 백혈성(AML), 급성림프성 백혈병(ALL),만성골수성 백혈병(CML) 및 언급되지 않은 각 아형과 함께 만성림프성 백혈병(CLL)이 있다[1-3].

유동 세포 계측법, 면역 표현형 및 분자 프로빙과 같은 몇 가지 고급 의료 검사 또는 검사를 포함하여 백혈병 분류를 시도하는 많은 연구가 있다 [4,5]. 이 검사는 세포의 세포학 특성뿐만 아니라 여러 가지 화학적 특성을 사용하기 때문에 백혈병 세포를 분류하는 데 매우 높은 성능을 발휘한다. 면역-표현형 방법은 유세포 분석법이 구별 할 수 있을 정도로 명백한 특정 항원으로 세포를 표지함으로써 유세포 분석법의 성능을 향상시킬 수 있다 [6].

그럼에도 불구하고 모든 시험은 비용이 많이 들고 시간이 많이 걸리며 작업자의 능력과 지구력에 달려 있다. 따라서, 말초 혈액 도말 이미지의 컴퓨터 비전 기반 현미경 검사에서의 이미지처리 사용은 빠르고 경제적이며 신뢰할 수 있는 백혈병 진단 방법을 생성하기 위한 표준 백혈병 결정지원 기술로서 바람직하다. 많은 연구자들이 컴퓨터 비전 기반 백혈병 분류

에 관한 훌륭한 연구를 수행했다고 Soltanzadeh 및 Scotti 등은 모든 이미지에 대한 데이터베이스를 생성할 뿐만 아니라 말초혈액도말 이미지에서 각각의 검출된 단일 세포에 기초하여 4가지 백혈병 유형 사이에 우수한 분류를 수행한다. 다른 연구는 Mohapatra 등이 컬러 클러스터링 기술에 기초하여 백혈병 유형을 검출하였다. 컴퓨터 비전 기반 알고리즘은 진단 처리시간을 단축할 뿐만 아니라 혈액암 전문의 진단 결과의 정확성을 향상시킨다. 본 실험에서 이미지 데이터베이스 소스는 충분하지 않지만 분류기는 백혈병 유형을 올바르게 분류할 수 있을 것으로 예상된다.

본 논문에서는 순차 신경망 분류기를 이용하여 급성백혈병 유형을 분류하는 새로운 접근법을 제안한다. 제안한 방법은 PCA 방법을 이용한 특징추출분석, 관심 영역을 추출하기 위한 몇 가지 종래의 이미지 전처리, 이미지 클러스터링 및 이미지 분할방법을 이용한다. 본 논문에서 제안한 분류기는 암 데이터 중 정상 이미지와 암세포 이미지를 97.75 % 정확도로 구분하였다.

2. 관련연구

2.1 백혈병 개요

2.1.1 급성골수성 백혈병(AML)

급성골수성 백혈병은 골수에서 초기 형태의 골수성 세포에서 발생한다. AML사례는 일반적으로 노인에서 발생하며 45세 이전에는 잘 발생하지 않는다. AML는 백혈병 4가지 유형 중에서 가장 많은 사망자가 발생한다. 핵 면적은 전체 세포 면적의 약 50-60% 이고 약 40-50%는 세포질이다. 때로는 “Auer rods”라고 불리는 짧은 자주색 줄무늬가 세포질영역에서 발견되어 AML 식별이 더 쉽다.

2.1.2 급성 림프 구성 백혈병 (ALL)

급성림프구성 백혈병(ALL)은 초기 단계의 골수 림프구 세포에서 발생하는 두 종류 급성백혈병 중 하나이다. ALL은 아동에게 발생하는 4가지 유형 중에서는 가장 흔하지만, 성인에게는 가장 덜 흔한 유형이다. 아동은 통상 성인보다 더 공격적인 치료를 더 견디기 때문에, 사망은 성인들 중 발생한다. 백혈병에 감염된 대부분의 백혈구는 적혈구 크기의 약 2배이며, 일반적인 백혈구는 평균 적혈구와 비슷한

크기이다. 대부분의 감염된 세포 중, 핵은 전체 세포의 약 80-90%를 차지하고 세포질은 약 20-30% 차지한다. 이 이는 AML과 ALL을 구별하는 가장 명확한 특징인 반면, ALL과 만성 유형을 구별하는 또 다른 주요 특징은 세포 내부에 너무 많은 과립이 있어 모든 세포의 표면이 매끄럽다는 점이다.

2.1.3 만성골수성 백혈병 (CML)

만성골수성 백혈병은 골수세포를 공격하고 특히 발달상태에서 CLL과 유사한 행동을 한다. CML 진단 시 평균 연령은 약 64세 정도이다. CML 세포는 세포질 면적 비 측면에서 AML과 유사하다. CML은 또한 세포 중 세포질의 많은 부분을 보여주며 대부분의 CML 세포는 성숙 백혈구로 4개 유형 중 가장 분화 가능한 세포이다. CML 핵은 대부분의 경우 모양이 둥글고 개발 중인 AML과 달리 차별화 가능한 모양과 질감으로 발전한다. CML의 이러한 특징은 다른 유형의 백혈병세포와 CML을 구별 가능하게 한다.

2.1.4 만성림프구성 백혈병(CLL)

만성림프구성 백혈병은 초기 단계의 림프구를 공격하는 ALL과는 달리 림프구를 바로 공격한다. CLL 세포는 소량의 세포질을 갖는 ALL과 유사한 특성을 가지며, 세포영역의 80-90%가 핵이 차지한다. ALL과 CLL의 차이점은 핵 영역 내부에 핵의 존재 유무이다. CLL은 기본적으로 발달이 완료된 성숙한 세포이기 때문에, CLL 세포내에 뉴클레오티드가 전혀 검출되지 않는다. 뉴클레올로스의 존재는 CLL 세포에 대해 커버렛 변환을 수행하면 검출 가능하다. 이 기능은 ALL과 CLL을 구별하는 주요기능이다. Fig. 1은 백혈병의 4 가지 분류유형이다.

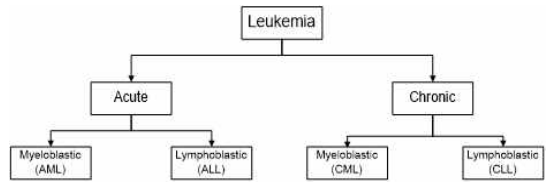


Fig. 1. Leukemia Class Classification.

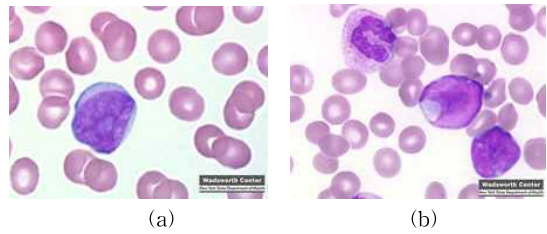


Fig. 2. (a) Acute Myeloid Leukemia (AML), (b) Acute Lymphocytic Leukemia (ALL).

Fig. 2는 AML과 ALL의 세포이고, Table 1은 AML과 ALL의 세포 특성을 비교한 것이다.

Table 2에서 만성골수성 백혈병(CML)은 다른 유형들 중에서도 생존율이 가장 높음을 알 수 있다. 평균 연령이 64세 이상인 미국 시민들 중 잘 나타나는 매우 드문 유형의 백혈병이다. 두 번째로 높은 생존율은 급성림프성 백혈병으로 CML과는 약간 다른 새로운 사례 수를 보였지만 사망자 수는 크게 다르다. 생존율이 71%인 만성림프성 백혈병, 생존율 45%인 급성골수성 백혈병이 그 뒤를 잇는다.

3. 제안한 분류 방법론

3.1 프레임워크와 제약

말초혈액도말이미지에 존재하는 백혈구 수에 따라 정상 및 비정상 백혈구 이미지를 대략 분류할 수

Table 1. Cytology properties comparison of AML and ALL

Parameter	AML	ALL
Blast size	Large and uniform	Small to medium; variable
Chromatin	Finely dispersed	Rather coarse
Nucleoli	1 to 4, often prominent	Absent or 1 or 2; indistinct
Cytoplasm	Moderately abundant	Scant to moderate
Granules	Granules often present	Granules lacking in nearly all cases
Auer rods	60 - 70% of cases	Absent
Myelodysplasia	Often present	Absent

Table 2. Statistical data from American Cancer Society after observing all Leukemia cases (including both children and adult) in United States for 2014

Leukemia Type	Newly Diagnosed Cases	Number of Deaths	Survivability	Average Age Diagnosed
ALL	6,020	1,440	76%	< 5 years old
AML	18,860	10,460	45%	66 years old
CLL	15,720	4,600	71%	72 years old
CML	5,980	810	86%	64 years old

있기때문에, 우리는 두 가지 방법을 제안한다. 첫째는 주로 이미지 전처리에 관한 사전분류이며, 둘째는 추출과 신경으로 구성된 두 개의 신경망 분류절차로 구성된 순차적 분류절차이다. 네트워크분류기와 전체 프레임워크는 Fig. 3에 설명되어 있다.

이 실험에서는 전체 평면 혈액도말 이미지가 실험 대상으로 사용된다. 우리는 훈련된 네트워크를 얻기 위해 이 알고리즘을 통해 처리된 90개의 훈련이미지와 함께 총 100개의 이미지를 사용했으며 10개의 테스트 이미지가 교차검증 이미지로 보존되었다. 단일 세포 이미지 대신, 완전한 평면세포 이미지가 사용되기 때문에, 이 시스템은 자동 백혈병 결정지원 시스템상에서 빌딩 모듈이 될 가능성이 높다. 제안된 신경망 분류기를 추가적으로 계속 사용함으로써 각 분류기의 계산비용을 절감하면서 SVM 알고리즘에 비해 우수한 결과를 확보할 수 있다.

이 방법은 새로운 백혈병 분류방법이며 가장 우수한 결과치를 보장하기 위한 제한조건도 제시해준다. 예를 들면 데이터베이스의 모든 이미지를 이미지처리 변경 없이 배율과 크기를 일정하게 처리해야 하는 제약조건도 있다. 이는 왜냐하면 이미지를 하나의 이미지 안에 여러 개의 핵이 있는 전체 이미지로 관찰

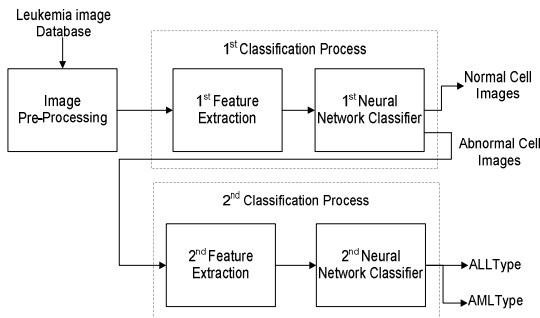


Fig. 3. The main framework of the proposed methodology.

하기 위함이다. 원하는 분류결과에 적절한 컬러 클러스터링 프로세스를 유지하기 위해서는 각 이미지에 표준 컬러링 시스템이 적용되어야 한다

3.2 이미지 전처리, 클러스터링 및 분할

3.2.1 RGB에서 CIE L*a*b 색 공간변환

이미지 전처리는 샘플 이미지를 읽는 데서 시작하여 RGB에서 CIE L*a*b로 색상공간을 변환한다. 이 계층은 광도계층 L, 색도계층 a* 및 계층 b*로 구성되어 3차원에서 2차원으로 색 치수를 줄여 색상기반 클러스터링 프로세스를 단순화한다. Fig. 4는 각 계층의 색 공간 변환 및 시각화 결과를 보여준다.

3.2.2 k-평균군집화

결과 이미지는 색상정보를 기반으로 이미지를 3

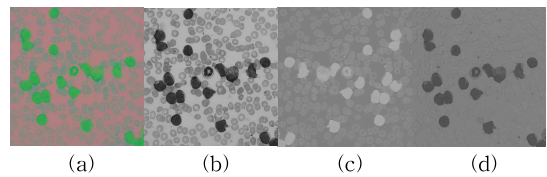


Fig. 4. (a) Result of color space transformation RGB to CIE L*a*b, (b-d) display the luminosity layer L, chromaticity *a layer, and *b layer respectively.

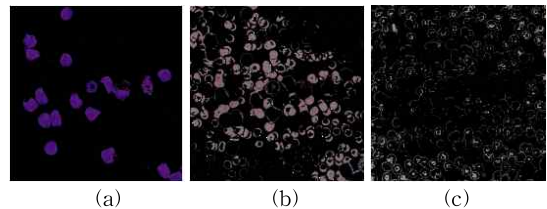


Fig. 5. Results of k-Means Clustering: (a) WBC's Nucleus Cluster, (b) WBC's Cytoplasm and Red Blood Cells cluster, and (c) Background Cluster.

개의 클래스로 분리하여 Fig. 5에서와 같이 k-평균 군집화에 사용된다[4]. 이렇게 분리하는 이유는 백혈구, 적혈구, 혈소판으로 구성된 혈액도말 이미지 자체의 특성 때문이다. 백혈구 세포는 다른 혈액 성분에 비해 색상특성이 매우 뚜렷하기 때문에 k-평균 알고리즘은 백혈구 세포를 제1, 제2 및 제3 클러스터로서 핵, 세포질 및 배경 등으로 분류한다.

3.2.3 OTSU 임계값, 형태적 필터, 영역개구부 및 마스크작성

핵 세분화 프로세스의 경우, 먼저 관심영역으로 핵 클러스터를 선택하고 OTSU 자동 임계값에서 시작하여 일련의 마스크 작성기술을 적용하여 첫 번째 클러스터에서 이진(binary)이미지를 만들고 결과 이미지가 관심영역에서 영역 열기 및 닫기와 같은 형태적 필터를 적용하게 되면, 원하지 않는 세그먼트 영역을 제거할 수 있다. Fig. 6 에서와 같이 이 실험에서는 1 픽셀 직경의 디스크 커널 모양의 영역 개구부 형태 필터와 5 픽셀 직경의 디스크 커널 모양의 영역 형태 필터를 사용한다. 직경 수는 제공된 이미지 데이터베이스를 사용한 실험에 기반한다. 여기서 핵 분할방법의 실제 과정은 형태학적 필터 커널의 직경 값을 결정하는데 있다. 이는 실험에서 사용하는 이미지에 실제로 의존하기 때문이다. 이전 단계의 결과는 두 가지로 핵 분할 이미지 및 추가단계에서 사용될 핵 분할 마스크이다.

3.3 그룹화된 세포분리 절차

실험을 통해 관찰된 바와 같이, 각각의 말초혈액도말 이미지 프레임에서 그룹화 된 백혈구가 이미지 프레임에 나타날 가능성이 높음을 알 수 있다. 문제

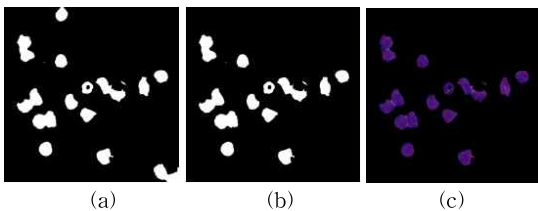


Fig. 6. (a) Nucleus Segmented Mask before border cleaning, (b) Nucleus segmented image after border cleaning, (c) Nucleus segmented image as a result of mask convolution with original image.

는 이로 인해 다음 프로세스에 문제가 발생할 수 있다는 점이다. 특히 각 셀의 모양 피쳐 계산이 이런 이유로, 각각의 개별 백혈구에 대해 형상특징계산이 실행되도록 그룹화 된 세포분리기술이 수행될 필요가 있다.

유역알고리즘은 원형 혈액세포를 분할 하는데 매우 적합하다. 유역알고리즘은 모든 픽셀에서 0이 아닌 가장 가까운 픽셀까지의 거리를 계산하는 거리변환에 따라 각 백혈구를 서로 다른 “캐치먼트 분지(catchment basins)”로 분할한다. 이 알고리즘은 거의 원형 셀에서는 분할성능이 매우 우수하지만 여전히 불규칙하거나 임의의 모양 개체를 분할하는 데는 다소 결함이 있다. 이 실험에서는 모든 샘플 이미지에 단일유역변환만 사용하여 핵 분할 이미지와 치수가 동일한 “단일 핵 마스크 이미지”를 생성하였다.

3.4 기능선택 및 추출

3.4.1 정상 및 비정상 세포 이미지의 특징선택 및 추출[7,8]

첫 번째 분류과정에서는 Fig. 7 에서와 같이 에너지, 엔트로피, 대비, 상관관계 및 백혈구의 Hausdorff 치수 값과 같은 정상세포와 비정상 세포를 구별하기

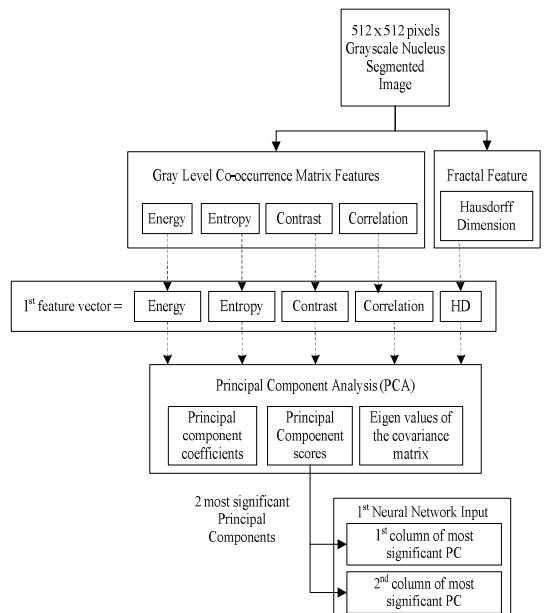


Fig. 7. Feature selection procedure for the first feature vector.

만하면 된다. 이는 매우 우수한 차별화 능력을 인식하고 보여주는 간단한 기능인 텍스트로 분류된다 [9,10].

정규화된 대칭 GLCM의 ij^{th} 요소에 대해 P_{ij} 를 설정하고 이미지의 그레이 레벨수에 대해 N 을 설정한다. 각 기능의 정의는 다음 섹션에서 설명한다.

1) 에너지(Energy)는 계산된 픽셀의 각 GLCM 값을 기반으로 이미지의 동질성 측정

$$Energy = \sum_{i,j=0}^{N-1} (P_{i,j})^2 \quad (1)$$

2) 엔트로피(Entropy)는 입력 이미지의 질감을 특성화하는데 사용할 수 있는 무작위성에 대한 통계적 측정값이다.

$$Entropy = - \sum_{i,j=0}^{N-1} \ln(P_{i,j}) P_{i,j} \quad (2)$$

3) 대비(Contrast)는 국소(Regional) 공유행렬의 다른 순간을 나타낸다. 이는 이미지 식(3)에 나타난 대조도 또는 다수의 국소적 변동을 측정한다.

$$Contrast = \sum_{i,j=0}^{N-1} P_{i,j} (i-j)^2 \quad (3)$$

4) 상관관계(Corelation)는 이미지 식(4)에 나타난 선형의존성의 국소적 패턴을 측정한다.

$$Correlation = \sum_{i,j=0}^{N-1} P_{i,j} \frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (4)$$

여기서 μ 는 GLCM 평균 (GLCM에 기여한 관계에서 모든 픽셀의 강도 추정값임)이며 σ^2 는 GLCM에 기여한 관계에서 모든 참조 픽셀의 분산값이다. μ 및 σ^2 는 식(5) 및 식(6)에 따라 계산된다.

$$\mu = \sum_{i,j=0}^{N-1} i(P_{i,j}) \quad (5)$$

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{i,j} (i-\mu)^2 \quad (6)$$

5) Hausdorff 치수(dimension)는 다양한 양적 측정을 위한 의학 및 과학연구에서 널리 사용되는 프랙탈 차원의 하나이다[6]. 이 실험에서 Hausdorff 차수는 핵 주변의 거칠기를 측정하는데 사용되며 정상 및 비정상 세포 이미지를 분류할 수 있는 필수기능으로 간주된다. 박스카운팅(Box Counting) 방법에 기초한 Hausdorff 치수 절차는 먼저 핵 분할 이미지의

그레이스케일 이미지로부터 이진(binary) 이미지를 획득하고 에지(edge)검출기술을 적용하여 핵 경계를 추적함으로써 계산된다. 다음 단계는 경계모서리를 사각형격자로 겹쳐서 식(7)을 사용하여 Hausdorff 치수를 계산한다.

$$HD = \frac{\log(N)}{\log(N(s))} \quad (7)$$

여기서 (N) 은 중첩된 그리드의 체공수이고 $(N(s))$ 은 점유된 체공 또는 상자수이다. HD 값이 높을수록 핵 주변에서 거칠기가 더 높음을 의미한다.

이 실험에서 모든 샘플 이미지는 많은 핵으로 구성되어 있기 때문에 HD 값은 정상 세포와 비정상 세포 사이에 큰 차이를 보여준다. 따라서 Hausdorff 치수는 첫 번째 분류절차에서 매우 중요한 기능이다.

3.4.2 골수성 및 림프구성 백혈병 세포 이미지에 대한 특징선택 및 추출

급성림프구성과 골수성 백혈병 유형을 분류할 두 번째 분류기의 경우, 분류과정에는 각 세포 모양에 대한 여러 가지 특징과 같은 더 복잡한 특징이 필요하다. 이러한 기능을 사용하면 ALL과 AML의 차이를 잘 구분할 수 있다.

(1) 세포영역: 각 백혈병 세포는 서로 다른 특성을 가지고 있으며, 가장 두드러진 특징 중 하나는 백혈구 영역이다. 분절된 세포 면적을 측정함으로써 실제 백혈구가 차지하는 픽셀 수에 대한 정보를 얻을 수 있다. 이 실험에서 카운트된 셀 면적은 하나의 프레임 내에서 검출된 모든 셀의 총계가 식(8)을 따른다.

$$Cell Area = \sum_i \sum_j p(i,j) \text{ where } p(i,j) = \begin{cases} 1, & \in \text{side cell RoI} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

(2) 핵 영역: 백혈구는 주로 핵으로 구성되며 세포 질로 둘러싸여 있다. 백혈병 유형은 이 기능에 따라 차별화될 수 있으므로 각 백혈구에서 핵 영역 정보를 얻는 것이 중요하다. 핵 영역 계산법은 다양하다. 예를 들면 핵 관심영역(RoI) 내부의 점유 픽셀을 계산하는 것이다

$$Nuc Area = \sum_i \sum_j p(i,j) \text{ where } p(i,j) = \begin{cases} 1, & \in \text{side vcleus RoI} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

(3) 세포질영역: 백혈구를 인정하는 또 다른 중요

한 특징은 세포질영역이다. 이는 세포질과 핵 영역 사이의 비율계산에 사용된다. 세포질영역을 성공적으로 계산함으로써 AML과 ALL의 차이를 알 수 있다. 세포영역과 핵 영역은 이미 잘 알려져 왔기 때문에 세포영역은 세포영역에서 핵 영역을 빼는 식으로 간단히 계산할 수 있다.

$$Cyt\ Area = Cell\ Area - Nc\ Area \quad (10)$$

(4) 핵 대세포질 면적비율(N:C 비율) : 이 특징은 핵과 세포질 사이의 면적비율을 나타낸다. 여기서 모든 세포에서 핵과 세포질 사이의 영역을 비교한다. 이 정보는 AML이 ALL보다 더 큰 세포질 면적을 가지므로 ALL이 AML보다 더 큰 NC비를 갖기 때문에 유형 분류에 사용된다. N:C 비율의 증가는 일반적으로 암세포 활동의 증가를 의미한다.

$$Nc: Cyt\ Ratio = \frac{Nc\ Area}{Cyt\ Area} \quad (11)$$

(5) 핵 대 세포면적비율 (N: 셀 비율) : 이 기능은 핵과 전체 세포면적 사이의 면적비율을 나타낸다. 이 비율정보는 AML 및 ALL 미분 알고리즘의 주요 특징 중 하나로 매우 중요하다. 왜냐하면 ALL이 AML보다 핵 대 세포 면적비율이 더 크다고 기존 문헌에서 연구되고 있기 때문이다.

$$Nc: Cell\ Ratio = \frac{Nc\ Area}{Cell\ Area} \quad (12)$$

3.5 순차신경망 분류기

3.5.1 신경망분류기 연결

신경망방법은 충분한 훈련 후 분류성능에서 우수한 정확도로 잘 알려져 있기때문에, 본 논문에서 좋은 분류결과를 얻었으며, 또 시험절차에서 처리시간을 단축하기 위해 두 신경망의 연결을 실험하였다. 순차 신경망 분류기 아키텍처를 구축하기 위해서는 두 신경망 분류기가 하나의 순차 시스템에 연결되어야 한다. 따라서 이 논문에서 연결과정은 두 번째 분류기에서 볼 수 있는데, 여기에서 AML과 ALL 백혈병 유형을 분류하기 위해 두 번째 분류기가 모든 비정상 세포를 처리한다. 이 실험에 사용된 신경망 구조는 서로 간에 최적의 처리시간을 보장하기 위해 가능한 단순하게 만들어졌다.

3.5.2 제1신경망 분류기

제1신경망 분류기의 목적은 정상 세포와 비정상 세포를 분류하는 것이다[11-13]. 이 분류기의 입력은 가장 중요한 주성분의 처음 두 열이므로 분류기의 각 입력노드(node)를 차지한다. 각 각의 숨겨진 노드는 입력-숨김 가중치 (Xi_k) 및 숨겨진 노드바이어스 (Bi_j)와 함께 입력로드 ($Wi_{k,j}$)로 부터 일정한 로드가 부여되었다.

$$ni_j = \sum_{k=1}^2 \sum_{j=1}^3 Xi_k \cdot wi_{k,j} + Bi_j \quad (13)$$

결과계산은 전달함수에 ($Fn(ni_j)$) 의해 처리되고 출력 계층에 wo_j 에 다른 가중치를 생성하고 출력노드 (Bo)의 바이어스와 합산하여 식 (14)에 따라 출력노드값(no)을 생성한다.

$$no = \sum_{j=1}^3 Fn(ni_j)wo_j + Bo \quad (14)$$

$$output = \sum_{j=1}^3 \left(\sum_{k=1}^2 Xi_k \cdot wi_{k,j} + Bi_j \right) wo_j + Bo \quad (15)$$

따라서 최종출력은 입력계층에서 숨겨진 계층으로의 선택된 전달함수는 접선 시그모이드(sigmoid)이며, 식(16)과 Fig. 8(a)에 표시된 접선 sigmoid 함수 형태와 같다.

$$\tan (sig (x)) = \frac{2}{(1 + \exp (-2x))^{-1}} \quad (16)$$

은닉계층에서 출력 계층으로의 전송함수는 방정식(17)을 따르는 일반적인 순수라인함수 또는 선형 전달함수로서 Fig. 8(b)와 같은 형태이다.

$$f(x) = x \quad (17)$$

전체 신경망 분류기 아키텍처는 Fig. 9에서 같이 볼 수 있다.

3.5.3 제 2신경망 분류기

제 2 신경망분류기에서, 정상 및 비정상 세포이미

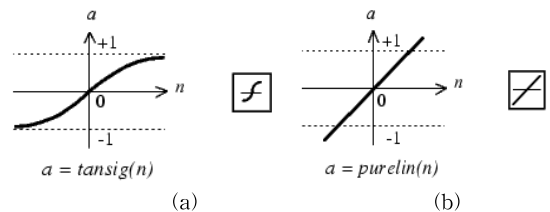


Fig. 8. (a) Tangential sigmoid transfer function and (b) linear transfer function.

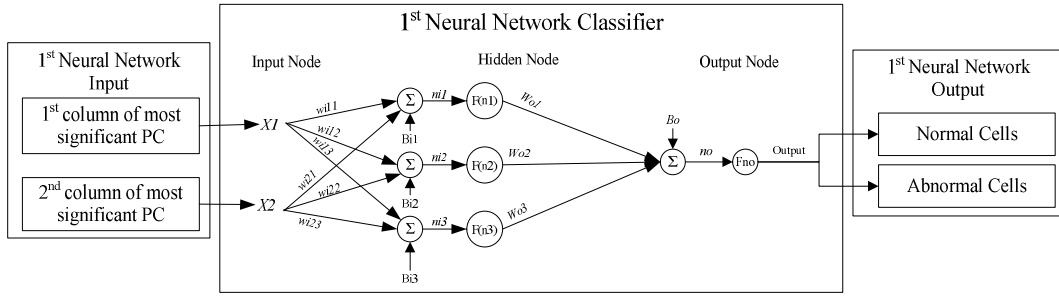


Fig. 9. First Neural Network Classifier Architecture which consists of 2 input nodes, 3 hidden nodes, and 1 output node to classify normal and abnormal cell images.

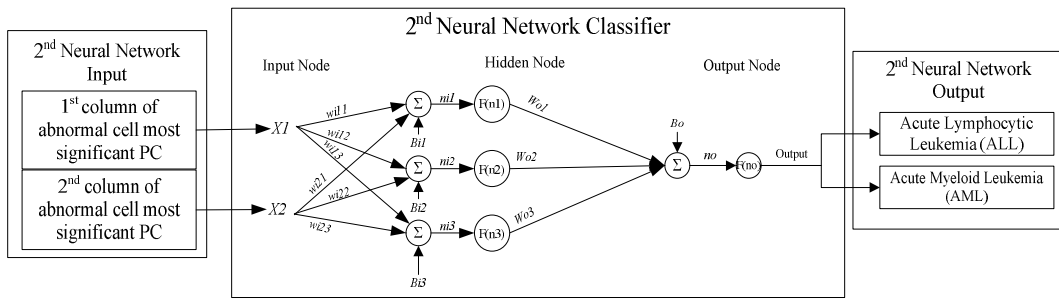


Fig. 10. Second Neural Network Classifier Architecture which consists of 2 input nodes, 3 hidden nodes, and 1 output node to classify ALL and AML cell images.

지를 분류하는 대신, 급성림프구성 백혈병(ALL)과 급성 골수성 백혈병(AML)을 분류한다. 그러나 이 분류기는 Fig. 10과 같이 이전 분류기와 동일한 네트워크 아키텍처를 사용한다.

두 가지 신경망 모두 LM (Levenberg-Marquardt) 알고리즘을 사용하여 학습하며, 지상진리 (ground truth) 데이터베이스 정보에서 수집된 사전 결정된 목표에 도달한다. LM 알고리즘은 비선형 실수값 함수의 제곱의 합으로 표현되는 다변량 함수의 최소값을 찾는 반복기법이다. 이 알고리즘은 신경망 훈련체계를 포함하여 광범위한 분야에서 널리 채택된 비선형 최소제곱 문제에 대한 표준기술이 되었다.

LM 알고리즘은 실제로 두 가지 최소화 방법의 조합으로, 기울기 하강 (gradient descent) 방법과 Gauss-Newton 방법이며 다음과 같이 간단히 설명한다. 기울기 하강 방법은 목적함수의 기울기와 반대방향으로 매개변수 값을 업데이트하는 일반적인 최소화 방법이다. 간단한 목적함수의 최소값을 찾기 위한 수렴 알고리즘으로 인식된다. 이 방법은 식(18)에 제시된 파라미터에 대한 카이제곱 목적함수의 기울기로 모

델링 된다.

$$\frac{\partial}{\partial p} X^2 = (y - \hat{y}(p))^T W \frac{\partial}{\partial p} (y - \hat{y}(p)) = -(y - \hat{y}(p))^T W \left[\frac{\partial \hat{y}(p)}{\partial p} \right] \quad (18)$$

여기서 $[\partial \hat{y} / \partial p]$ 는 매개변수 p 의 변화에 대한 함수 \hat{y} 의 국소 감도를 나타낸다. 따라서 가파른 하강방향으로 파라미터를 이동시키는 미분 (h)은 식(19)과 같다.

$$h_{g,d} = \alpha J^T W (y - \hat{y}) \quad (19)$$

여기에서 α 는 가장 가파른 방향으로 계단의 길이를 나타낸다.

LM 알고리즘을 형성하는 또 다른 방법은 Gauss-Newton 방법이며 제곱 목적함수의 합을 최소화하는 방법이기도 한다. 가우스-뉴턴방법 공식도출은 식 (20)에 기술된 것과 유사한 공식을 만든다.

$$[J^T W J] h_{gn} = J^T W (y - \hat{y}) \quad (20)$$

이상 두 가지 방법 모두에서 LM 알고리즘은 식 (21)에 따라 기울기 하강업데이트와 가우스-뉴턴 업데이트 사이의 파라미터 업데이트를 변경한다.

$$[J^T W J + \lambda I] h_{lm} = J^T W (y - \hat{y}) \quad (21)$$

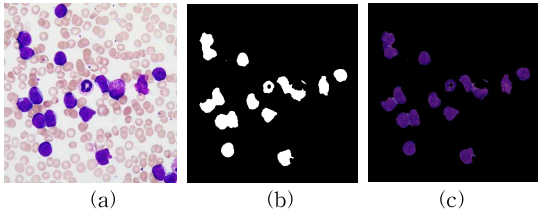


Fig. 11. (a) Original blood smear image, (b) binary mask of segmented region of interest, (c) nucleus segmented image.

여기서 λ 는 알고리즘이 따라야 할 업데이트를 나타내는 LM 알고리즘의 주요 매개변수이다. λ 값이 작으면 LM 알고리즘은 Gauss-Newton 업데이트를 따르고, 반면에 λ 값이 크면 LM 알고리즘은 기울기 하강업데이트를 따른다. 알고리즘이 해당 솔루션에 접근한 경우 일반적으로 로컬 최소값으로 빠르게 수렴하면 λ 값이 크게 초기화되고 점차 작아지는 가우스-뉴턴 방법이 점점 작아지고 있다. LM 알고리즘의 최종 업데이트 방정식은 식(22)로 표현된다.

$$[J^T W J + \lambda \text{diag}(J^T W J)] h_{lm} = J^T W (y - \hat{y}) \quad (22)$$

식(22)는 이 신경망 학습알고리즘에서 기본표현으로 간주된다

4. 실험결과

4.1 이미지 전처리, 클러스터링 및 분할결과

본 실험에 사용된 혈액 영상에서 Fig. 11(a)는 훈련 절차에 사용되는 90개 이미지의 예를 나타낸 것이

다. 섹션 3.2에 언급된 바와 같이 이미지 전처리, 클러스터링 및 분할 프로세스를 거친 후, 이진 이미지 마스크 및 핵 분할 이미지는 각각 Fig. 11(b) 및 11(c)에 나타난 바와 같이 생성된다.

4.2 특징추출결과

이 시각화를 통해 추출된 기능을 가장 중요하거나 덜 중요한 기능으로 평가할 수 있다. 이 경우 첫 번째 PCA 결과는 추출된 기능이 매우 중요하다는 것을 일관되게 나타낸다. 이전실험을 통해 대조 및 Hausdorff의 치수특성이 정상 및 비정상 세포를 구별할 수 있는 가장 중요한 특성으로 알려져 있다.

Fig. 12(a)는 PCA 변환 결과를 보여준다. 그림에서 알 수 있듯이, 빨간색 원 표시는 비정상 세포를 나타내고 파란색 x:표시는 정상 세포를 나타낸다. 그 사이에는 녹색선이 있으며 비정상 세포 이미지와 정상 세포 이미지 사이의 차이 경계선을 추정한다. Fig. 12(b)는 비정상 세포 이미지에 대한 두 번째 주요 성분분석을 보여준다. 빨간색 원 표시는 모든 셀 이미지를 나타내고 파란색 "x" 표시는 AML 셀 이미지를 나타낸다. 이 그림은 모든 마크가 그래프의 왼쪽을 차지하고 AML 마크가 여전히 그래프의 오른쪽에 흩어져 있기 때문에 선택된 피치가 높은 의미를 가지고 있음을 분명히 보여준다. 적은 수의 파란색 표시는 AML 이미지 데이터베이스의 부족으로 인해 발생하므로 추가연구가 필요하다

4.3 신경망 분류 결과

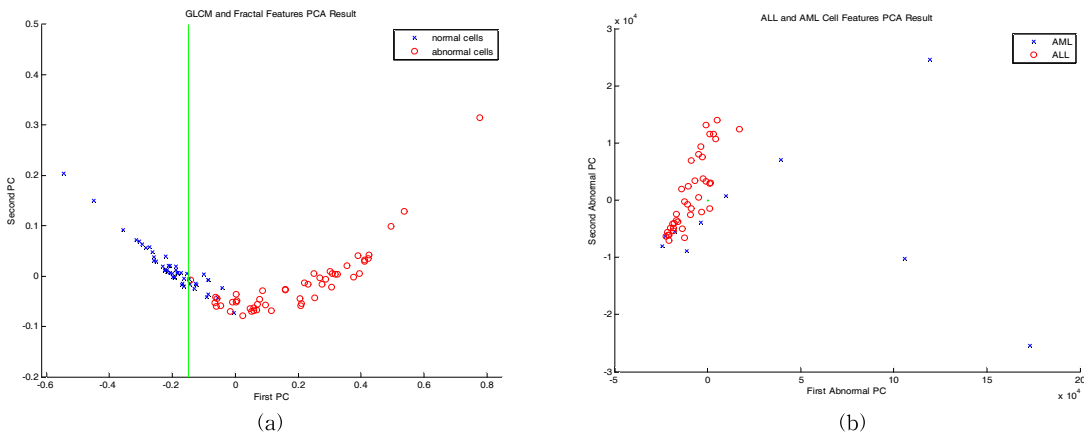


Fig. 12. (a) First PCA result, (b) second PCA result.

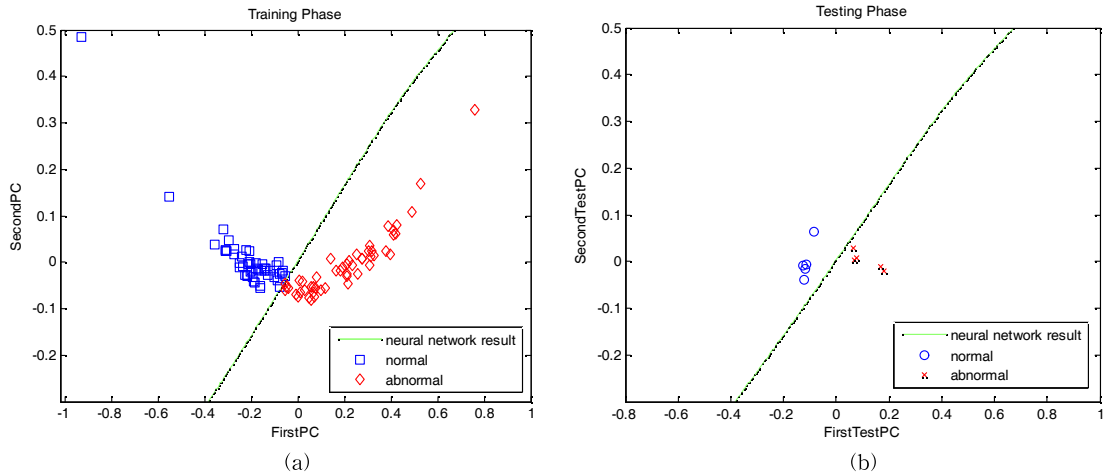


Fig. 13. (a) Verification result, (b) Validation result.

제 1 신경망 검증단계 분류결과는 Fig. 13(a)에 나와 있으며, 검증단계 분류결과는 Fig. 13(b)에 나타나 있다. 검증절차는 훈련된 네트워크에 동일한 입력훈련 이미지 데이터베이스를 적용하여 수행된다. 검증 성능은 Madukhar 등을 능가하는 97.7%의 정확도를 보인다. Madhukar 방법[4]의 정확도 93.5% 나타낸다. Table 3은 신경망 분류기를 사용한 결과의 데이터이다.

5. 결 론

본 논문에서 제안된 방법은 100개의 완전한 혈액도말 이미지를 활용하였다. 90개의 이미지는 훈련 이미지에 사용되는 반면 10개의 이미지는 유효성검사 이미지로 보존하였다. 이 실험에서, 정상 및 비정상 세포 이미지를 구별하기 위해 GLCM 및 프랙탈 특징

이미지가 데이터베이스로부터 추출되었으며, AML 및 ALL 백혈병유형을 구별하기 위해 비정상 세포 이미지에서만 5개의 기하학적 특징이 추출되었다. 순차신경망 분류성능은 97.7%의 정확도를 나타냈으며 93.5%의 정확도를 보이는 기존 참문헌들을 능가하였다. 이러한 실험결과는 자동 백혈병 진단 결정지원시스템을 위한 구성모듈 중 하나로 간주될 것으로 예상되지만, AML과 ALL 백혈병 유형 사이의 적절한 분류결과를 얻기 위해서는 향후에 추가연구가 필요하다.

REFERENCES

[1] E. Berner, *Clinical Decision Support Systems: State of the Art*, Agency for Healthcare Research and Quality Publication, Birmingham, Alaba., 2009.

[2] The American Society of Hematology Website, <http://www.hematology.org>(accessed october 8, 2014)

[3] V.F.R.D. Labati, "ALL-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing," *Proceeding of 18th IEEE International Conference on Image Processing*, pp. 2045-2048, 2011.

[4] S. Kumar, S. Mishra, P. Asthana, and Pragya, "Automated Detection of Acute Leukemia Using K-mean Clustering Algorithm,"

Table 3. Neural network classifier result

Parameter	Verification performance	Validation performance
	Value	Value
True Positive (TP)	40	5
False Positive (FP)	1	0
True Negative (TN)	48	5
False Negative (FN)	1	0
Specificity (TPR)	97.6%	100%
Sensitivity (SPC)	98%	100%
Accuracy	97.7%	100%

- Advances in Computer and Computational Sciences*, Vol. 1, No. 3, pp. 655-670, 2018.
- [5] C. Vununu, K.W. Kang, S.H. Lee, and K.R. Kwon, "Pyramidal Deep Neural Networks for the Accurate Segmentation and Counting of Cells in Microscopy Data," *Journal of Korea Multimedia Society*, Vol. 22, No. 3, pp. 335-348, 2018.
- [6] S. Shafique and S. Tehsin, "Computer-aided Diagnosis of Acute Lymphoblastic Leukaemia", *Computer Math Methods Med*, Vol. 2018, No. 1, pp. 1-13, 2018.
- [7] D.S.K.A.Z. Chitade, "Color Based Image Segmentation Using K-means Clustering," *International Journal of Engineering Science and Technology*, Vol. 2, No. 10, pp 5319-5325 2010.
- [8] H.A.R. Soltanzadeh, "Extraction of Nucleolus Candidate Zone in White Blood Cells of Peripheral Blood Smear Images Using Curvelet Transform," *Hindawi Publishing Corporation, Computational and Mathematical Methods in Medicine*, Vol. 2012, No. 1, pp. 1-12, 2012.
- [9] L.C.D. Ruberto, "White Blood Cells Identification and Counting from Microscopic Blood Image," *World Academy of Science, Engineering and Technology*, Vol. 73, No. 1, pp. 363-370, 2013.
- [10] S. Mohapatra and D. Patra, "Automated Leukemia Detection Using Hausdorff Dimension in Blood Microscopic Images," *Proceeding of International Conference of Emerging Trends in Robotics and Communication Technologies*, pp. 64-68, 2010.
- [11] N.H. Harun, "Automated Classification of Blast in Acute Leukemia Blood Samples Using HMLP Network," *Proceedings of the 3rd International Conference on Computing and Informatics*, pp. 55-60, 2011.
- [12] S.A.C.M. Madhukar, "New Decision Support Tool for Acute Lymphoblastic Leukemia Classification," *Proceeding of Society of Photographic Instrumentation Engineers- imaging Science and Technology Electronic Imaging*, pp. 1-12, 2012.
- [13] I. Vincent, "Feature Selection Using Principal Component Analysis for Leukemia Classification," *Proceeding of the 10th International Conference on Multimedia Information Technology and Applications 2014*, pp. 206-207, 2014.



Seon-Ja Lim

She received the M.S. in computer education from Kyung-sung University in 2001, and also Ph.D. candidate in computer engineering from Pukyong National University. She worked at Busan University of Foreign

Studies as adjunct Professor from 2006-2016. And She is currently a adjunct professor at Tongmyong University. Her research interests are in the area of big data analysis, image processing, artificial intelligence, neural network.



Ivan Vincent

Graduated as Bachelor of Engineering majoring in Electrical Engineering from School of Electrical and Informatics of Institut Teknologi Bandung in 2012. Graduated as Master of Engineering majoring in Digital

Signal and Image Processing from IT Convergence and Application Engineering Department of Pukyong National Universtiy in 2015. Currently working as a professional web-developer that constantly learning to utilize cutting edge technology such as Artificial Inteligence and Machine Learning in professional environment.



Ki-Ryong Kwon

He received the B.S., M.S., and Ph.D. degrees in electronics engineering from Kyungpook National University in 1986, 1990, and 1994 respectively. He worked at Hyundai Motor Company from 1986-1988 and at Pusan

University of Foreign Language from 1996-2006. He is currently a professor in Department of IT Convergence and Application Engineering at the Pukyong National University. He has researched University of Minnesota in USA in 2000-2002 with Post-Doc. and Colorado State University on 2011-2012 with visiting professor. He was the President of Korea Multimedia Society in 2015-2016. His research interests are in the area of digital image processing, multimedia security and watermarking, bioinformatics, weather radar information processing, and machine learning.



Sung-Dae Youn

He received the B.S., M.S. degrees in electronics engineering from Kyungpook National University, and received the Ph.D. degree in computer science from Pusan National University in 1997. He is currently a professor

in Department of Computer Engineering at the Pukyong National University. His research interests are in the area of parallel processing, multicasting communications, big data.