

현미경 영상 기반 암세포 생존력 관련 표현형 추출

강미선*

순천향대학교 컴퓨터소프트웨어공학과

Microscopic Image-based Cancer Cell Viability-related Phenotype Extraction

Misun Kang*

Department of Computer Software Engineering, Soonchunhyang University
(Manuscript received 2 May 2023 ; revised 17 May 2023 ; accepted 18 May 2023)

Abstract: During cancer treatment, the patient's response to drugs appears differently at the cellular level. In this paper, an image-based cell phenotypic feature quantification and key feature selection method are presented to predict the response of patient-derived cancer cells to a specific drug. In order to analyze the viability characteristics of cancer cells, high-definition microscope images in which cell nuclei are fluorescently stained are used, and individual-level cell analysis is performed. To this end, first, image stitching is performed for analysis of the same environment in units of the well plates, and uneven brightness due to the effects of illumination is adjusted based on the histogram. In order to automatically segment only the cell nucleus region, which is the region of interest, from the improved image, a superpixel-based segmentation technique is applied using the fluorescence expression level and morphological information. After extracting 242 types of features from the image through the segmented cell region information, only the features related to cell viability are selected through the ReliefF algorithm. The proposed method can be applied to cell image-based phenotypic screening to determine a patient's response to a drug.

Key words: Microscopic image, Cancer cell, Image analysis, Feature extraction, Feature selection

1. 서 론

한국인의 주요 사망원인인 암은 일반적으로 동일한 암 종류와 기수를 진단받을 경우, 대부분 비슷한 약물을 이용하여 치료과정을 시작한다. 동일한 약물을 사용했을 때 환자에 따라 부작용이 발생하거나 암치료에 효과적인 반응이 나타나는 등 다른 반응이 나타나는 특징이 있다. 그 원인 중 대표적인 이유는 개인별 유전적 고유 변이 요인에 의한 것이라고 한다

[1,2]. 이 때문에 개인별로 맞춤 치료의 필요성이 대두된다.

개인별 맞춤 치료를 위해서는 실제적으로 여러 종류의 약물을 환자에게 직접 투여해볼 수는 없기 때문에 세포 수준에서의 약물 효능 검증이 필요하다. 암세포의 약물 반응성 측정을 위해 기존에 많이 사용되고 있는 HTS(High Throughput Screening) 방법은 많은 데이터를 빠르게 처리하는 것에만 초점을 맞춘 방법으로 단순히 세포 개수 정보만을 획득하는 간소화된 방법이다. 이를 보완하여 보다 많은 정보를 획득하기 위한 HCS(High-Content Screening) 방법이 사용되고 있다. 기존 HTS에서 추출되는 세포 개수뿐만 아니라 개별 세포별로 유전자와 환경의 영향에 의해 형성된 생물의 형질이 겉으로 드러나는 다양하고 복잡한 세포 표현형 정보 획득을 가능하게 한다. HCS는 높은 공간 해상도의 대용량 현미경 영상 데이터를 생성하여 분석하는 기술로서 수동 분석 시 주관적일 수 있고, 정량적 분석에 한계가 있으며, 시간이 오래 소요되기 때문에 이러한 문제를 보완하기 위해 컴퓨팅 기술을

*Corresponding Author : Misun Kang
Department of Computer Software Engineering, Soonchunhyang University, Asan-si, Chungcheongnam-do 31538, Republic of Korea
Tel: +82-41-530-1382
E-mail: ms.kang@sch.ac.kr

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구입니다(No. RS-2022-00165681). 본 연구에서 사용된 암세포 현미경 영상 데이터는 삼성서울병원 연구센터로부터 제공받아서 수행되었으며 이에 감사드립니다.

기반으로 한 영상 데이터 처리 및 분석 연구가 활발하게 이루어지고 있다[3-5].

본 논문에서는 기존 HTS 방법 대비 많은 정보 획득이 가능한 HCS 방법에 기반한 암세포의 약물 반응성 측정을 위해서 컴퓨팅 기술을 바탕으로 뇌종양 현미경 영상 분석을 통해 2백여 종의 다양한 세포 표현형 특징값들을 추출하고, 추출된 특징들 중에서 약물의 반응성에 따라 변화하는 세포 생존력과 관련이 있는 특징만을 자동으로 선별하는 방법에 대해 제안한다. 연구 수행을 위한 본 논문의 구성은 다음과 같다. 2장에서는 먼저 시야각의 한계 및 고배율 해상도 촬영으로 인해 동일 대상을 부분적으로 나누어 획득한 영상을 하나로 합치는 영상 스티칭 방법을 서술한다. 그리고, 개별 세포 단위로 분할하는 방법을 제안하며, 분할된 영역으로부터 특징을 추출하는 내용에 대해 소개한다. 3장에서는 제안한 방법을 이용하여 실험한 결과를 보여주고, 4장에서는 결론 및 향후 연구 방향을 기술한다.

II. 연구 방법

본 논문에서 제안하는 방법의 순서는 다음과 같다. 고배율 영상을 사용하기 때문에 제한된 FOV(Field Of View)의 한계를 극복하기 위하여 최소 겹침 방법 기반의 영상 스티칭을 적용하고, 조명의 위치에 따라 밝기가 균일하지 못하기 때문에 이를 보정해주는 과정을 수행한다. 그 다음 픽셀의 밝기 값을 기반으로 세포핵의 위치를 탐지하고 개별 세포 단위로 분할한다. 분할된 영역으로부터 다양한 특징들을 추출하여 그 중 암세포의 표현형에 중요한 특징만을 선별한다.

1. 영상 스티칭

본 연구에서는 20X 배율로 촬영한 뇌종양 세포의 형광 현미경 영상을 사용한다. 세포 배양 시 동일 환경인 1개의 웰 플레이트는 FOV의 한계로 인하여 15장의 필드 영상으로 촬영된다. 이를 1장의 영상으로 결합하기 위해서 영상 스티칭 기술이 필요하다[6].

영상 간 중첩되는 영역이 좁을 경우에도 정확하게 스티칭 하기 위해서 영상 밝기값 기반의 삼각 임계값과 NCC(Normalized Cross-Correlation)을 이용한다. 영상 내 관심 영역인 형광 염색된 세포핵의 경우 배경과 대비되어 밝게 나타나는 특성이 있다. 하지만 필드 영상 경계에 위치하는 세포핵 중 작은 일부분만 포함할 경우 어둡게 나타나는 경우가 있어서 배경과 관심 영역을 구분할 때 민감하게 동작하는 삼각 임계값을 이용한다. 삼각 임계값은 영상의 히스토그램 높이와 범위의 대각선에서 최대 거리에 있는 점을 선택하고 끝에 고정 오프셋을 추가하는 방법이다[7]. 이렇게 탐지한 세포핵 위치 정보를 이용하여 수평 또는 수직으로 정렬된 두 영상 사이의 중첩되는 영역을 찾기 위해서 x, y 방향으로 점진적으로 이동시켜 중첩 영역을 확대하면서 NCC를 계산하여 유사도가 높은 영역을 식별하여 나누어 촬영된 필드 영상들을 1장의 영상으로 합치는 과정을 수행한다. NCC 방법은 기존에 영상 밝기값 기반의 스티칭의 경우 영상 전체적인 밝기가 균일하지 않을 경우 적용이 어려운 문제를 보완하기 위해 정규화를 통해 영상 전반적 밝기 차이에 관계없이 원영상의 밝기 차이와 정규화된 영상의 밝기값 패턴을 동시에 비교할 수 있는 방법이다[8]. Fig. 1(a)는 15장의 필드 영상을 영상 밝기값 정보와 NCC 알고리즘을 통해 1장의 영상으로 스티칭한 결과이다.

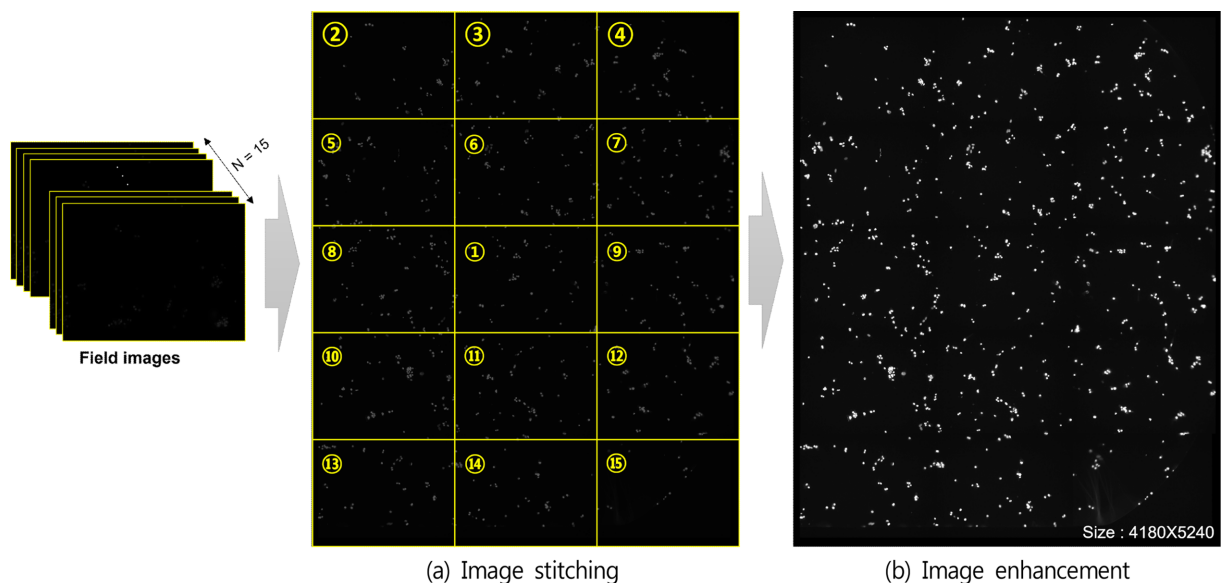


그림 1. 영상 스티칭과 개선 결과

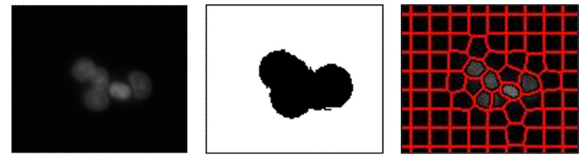
Fig. 1. Image stitching and enhancement result

2. 영상 밝기값 개선

현미경 영상의 경우 조명의 위치에 따라 조명에 가까울수록 형광 발현 강도가 높게 나타나고, 조명에서 멀어질수록 발현 강도가 낮게 나타난다. 이 문제를 보완하여 정확한 분석을 하기 위해 히스토그램 정규화 방법을 통해 영상의 대비도를 향상시킨다. Fig. 1(b)는 각각 촬영된 필드 영상을 1장의 영상으로 스티칭한 후 밝기 값을 개선하여 영상 대비도를 향상시킨 결과이다. 원영상에서 나타나는 조명의 거리에 따른 불균일한 밝기 문제가 해결되고, 영상 전반적으로 밝기값의 대비도가 향상되어 배경과 세포핵 영역 구분이 용이하도록 영상이 개선됨을 확인할 수 있다.

3. 슈퍼픽셀 기반 세포핵 분할

암세포의 경우 특정 환경에서 서로 응집하는 현상이 주요 특징으로 발현될 수 있다. 이 결과를 영상으로 촬영했을 때 일반적인 밝기값만을 이용한 문턱치 알고리즘들을 적용하면 응집된 세포핵을 개별적으로 각각 분할하기엔 한계가 있다. 이러한 한계를 보완하기 위하여 세포핵의 형광 발현 정도값 뿐만 아니라 세포핵의 둥근 형태도 고려한 슈퍼픽셀 기반의 SLIC(Simple Linear Iterative Clustering) 알고리즘을 적용한다. SLIC 알고리즘은 주요 매개변수로 초기 영역 크기 및 세포핵의 형태학적 특징을 반영한 밀집성을 매개변수로 조정한 후 영상을 개별 세포핵 단위로 세분화하는 과정을 수행한다[9,10]. Table 1은 SLIC 알고리즘이 동작하는 과정을 나타내는 표이다[10]. 설정된 매개변수에 의해서 초기 세분화 영역의 크기가 설정되고, 유클리드 거리를 이용하여 k-평균 군집화 방법을 기반으로 주변 값과 비교하여 비슷한 특징에 따라 영역 세분화가 조정되는 방식이다. 세분화되는 과정에서 매끄럽고 규칙적인 경계선을 추출하기 위해서 조정 가능



(a) Raw image (b) Thresholding result (c) SLIC result

그림 2. 영상 분할 결과

Fig. 2. Image segmentation results

한 밀집성 매개변수를 일반적으로 많이 사용되는 10~50 사이의 값을 실험적으로 조정하여 해당 영상에서 가장 효과적인 결과를 보인 30으로 적용하였다.

Fig. 2(a)는 암세포의 주요 특징 중 하나인 응집성이 나타난 일부 영역의 원영상이다. Fig. 2(b)는 원영상에 일반적인 문턱치 알고리즘을 적용한 결과로써 개별 세포 단위로 분할이 되지 않고 여러 개의 세포가 1개로 뭉쳐져서 분할된 결과를 확인할 수 있다. Fig. 2(c)는 응집한 세포 분할을 위해서 형태 정보도 고려한 SLIC 알고리즘 적용 결과이다. 영상 밝기값만 고려했을 때는 서로 붙어있는 세포핵을 분할하는데 한계가 있으나 SLIC 적용 후 각각 둥근 세포핵 특징이 추가적으로 반영되어 개별 세포핵 단위로 분할됨을 확인할 수 있다.

4. 특징 추출

동일 환경인 웰 단위 영상 내 분할된 각 세포핵 영역으로부터 Table 2와 같이 총 242종류의 특징을 추출한다. 앞서 서론에서 소개한 일반적으로 쉽게 사용되는 세포핵의 개수와 관련된 기본 특징을 추출하고, 형태 특징과 관련하여 원형률, 이심률, 볼록성, 오일러 개수, 장축 길이, 둘레, 반지름, 방향성, Zernike[11] 등의 특징들을 다음과 같이 추출한다.

178

표 1. 슈퍼픽셀 기반 영상 세분화 방법

Table 1. Superpixel-based image segmentation method

Input: number of grid compactness Output: Segmented Result
Initialize cluster centers $Center_k = [l_k, a_k, b_k, x_k, y_k]^T$ by sampling pixels at regular grid steps S . Move cluster centers to the lowest gradient position in a 3×3 neighborhood. Set label $l(i) = -1$ for each pixel i . Set distance $d(i) = \infty$ for each pixel i . repeat for each cluster center C_k do for each pixel i in a $2S \times 2S$ region around C_k do Compute the distance D between C_k and i . if $D < d(i)$ then set $d(i) = D$ set $l(i) = k$ end end end Compute new cluster centers. Compute residual error E . until $E \leq \text{threshold}$

표 2. 추출한 특징 정보

Table 2. Extracted feature list information

Characteristic	Number of features
Number of cell nuclei in the image	1
Cell nucleus area shape	141
Cell nucleus intensity	45
Cell nucleus location	24
Cell nucleus intensity threshold value by thresholding Algorithm	4
Neighboring cell nuclei based on the Voronoi algorithm	6
Neighboring cell nucleus	21

- 영역 : 세포핵 영역의 실제 픽셀 개수
- 원형률 : $(4 \times \pi \times \text{세포핵영역}) / \text{경계선}^2$ (1이면 원형)
- 이심률 : 세포핵 영역을 둘러싼 타원의 두 초점 간 거리의 장축에 대한 비(0과 1 사이에 값. 0에 가까울수록 원형에 가까움)
- 볼록성 : 세포핵 영역의 볼록 꺾질 내 차지하는 비율
- 오일러 개수 : 8이웃 연산을 이용하여 찾은 세포핵 영역 내 구멍의 수
- 장축 길이 : 세포핵 영역과 동일한 정규화된 두 번째 중앙 모멘트를 갖는 타원의 장 축 길이
- 단축 길이 : 세포핵 영역과 동일한 정규화된 두 번째 중앙 모멘트를 갖는 타원의 단 축 길이
- 방향성 : 세포핵 영역과 동일한 정규화된 두 번째 중앙 모멘트를 갖는 타원의 장축과 x축 사이의 각도(범위: -90도 ~ 90도)
- Zernike : 세포핵이 분할된 이진 영상에서 1개의 세포핵이 중심에 위치하는 부분 영 상에서 Zernike 다항식을 통해서 추출하는 10개의 특징

그리고, 세포핵에 첨가한 형광 물질의 발현 정도와 관련하여 영상 내 밝기값을 이용하여 세포핵의 경계 영역, 통합 영역으로부터 밝기 정보를 추출하며, 임계값 기반 알고리즘 적용하여 배경과 세포핵 영역 구분 시 적절한 밝기 임계값도 추출한다. 또한, 각 개별 세포 정보뿐만 아니라 영상 내 세포의 전반적인 패턴 정보 추출을 위하여 이웃한 세포와의 관계와 관련된 특징들을 추출한다. 추출 방법은 대표적으로 응집성 확인을 위한 보로노이 테셀레이션(voronoi tessellation)을 이용하여 이웃 세포핵 정보를 추출하고[12,13], 이웃 세포와의 거리 관련 특징 및 각도 정보 등을 추출한다.

앞서 소개한 특징에 대해서 평균값, 중앙값, 최대값, 최소값 등의 통계학적 값들을 추가로 계산하여 특징값으로 사용한다. 또한, 본 논문에서 분석 대상이 되는 세포핵의 경우 픽셀의 개수 즉, 세포핵 영역이 크지 않기 때문에 해당 영역을 슈퍼 픽셀 기반으로 분할하였을 때 분할 경계선을 포함한 결과와

포함하지 않은 결과에 따라 영향을 받을 수 있는 면적, 형태와 관련된 특징들은 경계선을 포함한 영역과 포함하지 않은 영역 각각에 대해 추출한다.

III. 실험 및 결과

본 연구에서는 암세포 영상 분석을 위하여 20X 배율에서 교모세포종의 세포핵을 DAPI 형광 염색하고, AMG 232 약물 첨가 후 PerkinElmer사의 Operetta 형광 현미경으로 촬영된 영상을 사용한다. 20X 배율 영상의 경우 웰플레이트 내 1개의 웰은 15장의 필드 영상으로 구성된다.

각각 다른 웰 환경에 대해 추출한 242 종류의 특징 간의 상관관계를 살펴보기 위하여 각 값들을 색상 히트맵[14]으로 군집화 및 가시화한 결과는 Fig. 3과 같다. 이 결과는 값에 따라 비슷한 값들은 비슷한 색상으로 표시해주는 히트맵과 동시에 데이터를 정규화한 결과를 통해 계층적 군집화를 트리구조로 나타낸 덴드로그램을 표시한 것이다. 이러한 색상과 군집화 트리 구조 결과를 통해서 추출된 특징 242종 중에 암세포의 환경인 약물 첨가 정도를 잘 구분할 수 있는 특징들과 그렇지 않은 특징들이 혼재함을 확인할 수 있다.

이렇게 추출된 다양한 특징들 중에서 실제로 암세포의 약물 처리 등의 환경 변화에 따른 암세포 생존력을 분류하는데 중요하게 작용할 수 있는 특징들만을 선별하는 과정이 필요함을 알 수 있다. 본 연구에서는 주요 특징을 자동으로 선별하기 위하여 ReliefF 알고리즘을 적용한다[15,16]. ReliefF 알고리즘은 특징들을 선별하는 방법 중 효과적인 방법으로써 기존 Relief에서 발전된 방법으로 뒤에 F가 붙는다. 기존 방법에서의 개선점을 살펴보면 복잡한 문제에서 가중치 추정 신뢰도를 높이기 위해서 각 특징 값들의 점수(score)를 계산할 때 단일 값이 아닌 k개의 이웃 수를 사용한다. k는 경험적 테스트를 기반으로 기본값 10을 사용한다.

환경에 따른 암세포의 생존력 분류 기준이 되는 데이터는 병원에서 제공해준 암세포 영상으로부터 분석된 세포 생존력

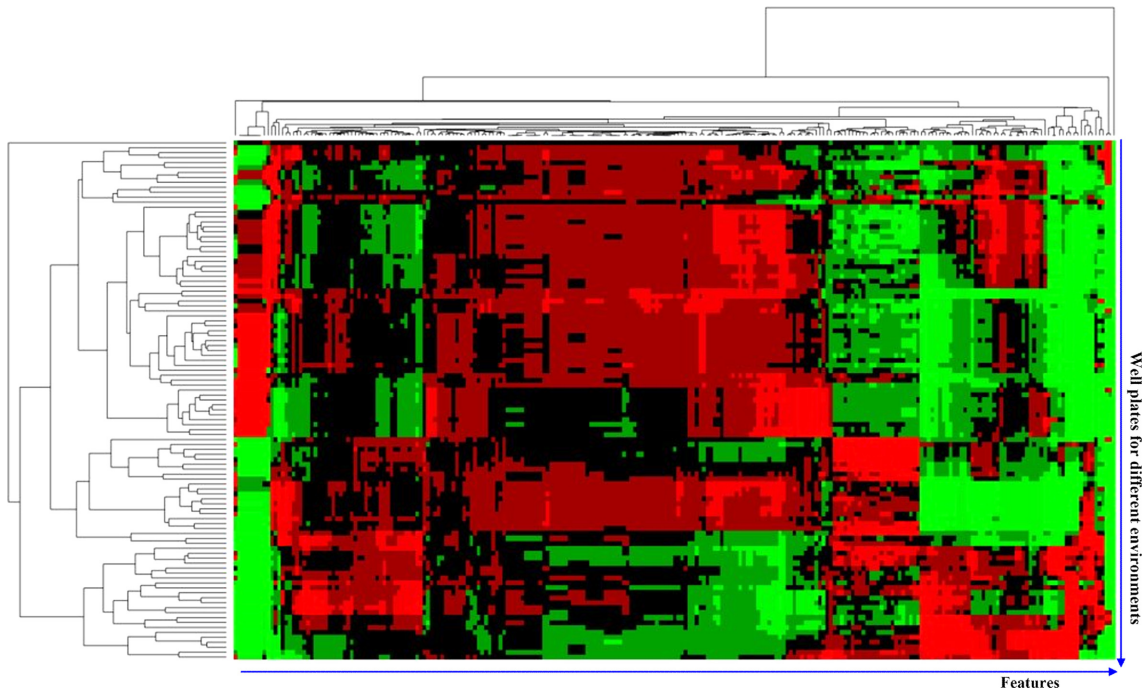


그림 3. 특징 군집화 결과
Fig. 3. Feature clustering results

표 3. 선별된 특징
Table 3. Selected feature types

#	Features	Weight value
1	Median_Neighbors_SecondClosestObjectNumber	0.56255
2	Median_Neighbors_FirstClosestObjectNumber	0.56179
3	StDev_Neighbors_FirstClosestObjectNumber	0.56165
4	Mean_Neighbors_SecondClosestObjectNumber	0.56159
5	Count	0.56151
6	Mean_Neighbors_FirstClosestObjectNumber	0.56141
7	StDev_Neighbors_SecondClosestObjectNumber	0.56131
8	Mean_Neighbors_AngleBetweenNeighbors	0.46002
9	StDev_Neighbors_NumberOfNeighbors	0.44946
10	Median_Neighbors_AngleBetweenNeighbors	0.41658

정보를 이용한다. 제공된 세포 생존력 정보는 암세포에 약물을 첨가했을 때와 약물을 첨가하지 않았을 때 나타나는 변화를 세포의 크기와 형광 발현 정보 및 세포 사멸 마커를 이용하여 계산한 것이다. 본 연구에서는 영상으로부터 추출한 다량의 특징 데이터 242종류 중 세포 생존력 값과 관련된 특징만을 자동으로 선별한다. 선별된 상위 10개 특징은 Table 3과 같다. 약물 반응이 명확하게 나타난 한 환자의 암세포로부터 얻어진 영상을 이용하였으며, 약물을 첨가했을 때와 안 했을 때 달라지는 세포 생존력 차이를 분류하는데 좋은 특징들을 선별한 것이다. 특징들을 살펴보면 기본 특징인 세포의 개수

변화 특징이 5번째로 주요하게 나타났으며, 다른 특징들은 모두 이웃한 세포의 개수와 관련된 특징으로 선별되었음이 확인된다. 이는 영상에 나타나는 세포의 패턴과 밀접한 특징으로 세포의 응집성과도 관련이 있어 약물 첨가에 따른 세포의 개수 변화 및 이동 패턴이 세포 생존력과 밀접한 관련이 있음을 알 수 있다.

IV. 결 론

일반적으로 동일한 암 종류와 기수를 진단받을 경우, 대부

분 비슷한 약물을 이용하여 치료한다. 동일한 약물을 사용했을 때 환자에 따라 반응성은 다르게 나타나는 특징이 있다. 이 때문에 개인별로 맞춤 치료의 필요성이 대두된다. 이를 위해서는 여러 종류의 약물을 환자에게 직접 투여해볼 수는 없기 때문에 세포 수준에서의 약물 효능 검증이 필요하다. 암세포의 약물 반응성 측정을 위해 기존에는 일반적으로 세포 개수의 변화 측정 방법이 많이 이용되었으며, 최근 세포의 유전자와 환경의 영향에 따라 생성된 표현형 정보를 추출하는 연구가 활발하게 진행되고 있다. 본 논문에서는 세포 표현형 정보 추출을 위해 컴퓨팅 기반 영상 분석 기술을 통해서 정량적 정보를 다양하게 추출하고 그 중에서 세포 생존력과 밀접한 특징만을 자동으로 선별하는 방법을 제안하였다. 형광 현미경 영상에 영상 스티칭 알고리즘을 통해 동일 환경인 웰 플레이트 단위 영상을 생성하고, 히스토그램 기반 영상 불균일 밝기를 개선하였다. 개선된 영상에서 세포핵의 형태 및 밝기 정보 등을 고려하여 슈퍼픽셀 기반 알고리즘을 통해 세포핵 영역을 분할하고, 분할된 영역으로부터 200종이 넘는 다양한 특징들을 추출하였다. 또한 추출한 특징들 중 세포 생존력과 밀접한 연관이 있는 특징들만을 추출하는 방법을 제시하였다.

향후 연구과제로는 여러 환자 및 약물에 대한 데이터를 이용하여 분석하는 연구를 수행할 것이다. 또한, 세포핵 뿐만 아니라 분석이 어려운 세포막 부분을 염색한 영상에서의 표현형 정보도 추출하여 세포 생존력과 관련된 특징들을 추출 및 선별하고, 딥러닝을 통해서 보다 다양하고 복잡한 데이터 분석을 수행할 것이다.

References

- [1] Connor S. Glaxo chief: Our drugs do not work on most patients. *The Independent*. 2003;8:1-2.
- [2] Fontanella A. Equivalent/generic drugs use. To whom? When? How?. *Italian Journal of Medicine*. 2014;8(2):77-79.
- [3] Fetz V, Prochnow H, Brönstrup M, Sasse F. Target identification by image analysis. *Natural product reports*. 2016; 33(5):655-667.
- [4] Singh S, Carpenter AE, Genovesio A. Increasing the content of high-content screening an overview. *Journal of biomolecular screening*. 2014;19(5): 640-650.
- [5] Liberali P, Snijder B, Pelkmans L. Single-cell and multivariate approaches in genetic perturbation screens. *Nature Reviews Genetics*. 2015;16(1):18-32.
- [6] Adel E, Elmogy M, Elbakry H. Image stitching based on feature extraction techniques: a survey. *International Journal of Computer Applications*. 2014;99(6): 1-8.
- [7] Ghosh D, Kaabouch N. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*. 2016;34:1-11.
- [8] Seo JH, Yang S, Kang MS, Her NG, Nam DH, Choi JH, Kim MH. Automated stitching of microscope images of fluorescence in cells with minimal overlap. *Micron*. 2019;126:102718.
- [9] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. SLIC superpixels. EPFL-REPORT-149300, 2010.
- [10] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süssstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*. 2012;34(11):2274-2282.
- [11] Teague MR. "Image analysis via the general theory of moments." *JOSA* 70.8 1980;920-930.
- [12] Aurenhammer F. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*. 1991;23(3):345-405.
- [13] Wang J, Ju L, Wang X. An edge-weighted centroidal Voronoi tessellation model for image segmentation. *IEEE Transactions on Image Processing*. 2009;18(8):1844-1858.
- [14] Wilkinson L, Friendly M. The history of the cluster heat map. *The American Statistician*, 2009;63(2):179-184.
- [15] Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore J H. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*. 2018;85:189-203.
- [16] Kononenko I. Estimating attributes: Analysis and extensions of RELIEF. In *ECML*. 1994;94:171-182.