



YOLOv5를 이용한 정상/비정상 적혈구 구별 알고리즘 개발

김진구^a, 강지훈^b, 최경철^a, 장웅기^{a,b}, 하호진^{a,b}, 임광석^{a,b}, 김병희^{a,b}, 박용재^{a,b*}

Development of a Normal/Abnormal RBC Detection Algorithm using YOLOv5

Jin-Gu Kim^a, Ji-Hun Kang^b, Kyoung-Cheol Choi^a, Woong-Ki Jang^{a,b}, Ho-Jin Ha^{a,b}, Kwang-Suk Lim^{a,b},
Byeong-Hee Kim^{a,b}, Yong-Jai Park^{a,b*}^a Biohealth-Machinery Convergence Engineering, Kangwon National University^b Department of Smart Health Science and Technology, Kangwon National University

ARTICLE INFO

Article history:

Received	10	March	2022
Revised	28	March	2022
Accepted	29	March	2022

Keywords:

Diagnosis AI algorithm
YOLOv5
Red blood cell
Image processing

ABSTRACT

Blood with cells such as WBC, RBC, and platelets, is often used in health diagnoses using morphological analyses. Among them, RBC is used as an indicator for diagnosing diseases such as diabetes, anemia, and heart failure; however, their accuracy is poor due to subjective analyst interpretations since they change reversibly based on various environmental factors such as viscosity, osmotic pressure, and calcium in the blood. In this study, an algorithm that objectively analyzes the shape of RBC and measures the ratio using YOLOv5, an AI technology, was developed and used to distinguish them as normal or abnormal. Data was augmented using image processing technology due to low data, and the YOLOv5 model was applied after labeling, where it was classified into four classes. The developed algorithm was verified through precision and recall, and subsequently applied to the test data to distinguish normal/abnormal red blood cells to obtain a ratio.

1. 서론

현대의 진단에 있어서 혈액 세포의 형태 분석은 여전히 중요한 진단 방법이다^[1]. 그러나, 세포학적 변수를 정의하기 위한 객관적인 기준이 없어 검토자의 주관적인 해석이 정확도에 영향을 미친다. 따라서 현미경 촬영한 이미지를 전산화된 이미지 분석을 통해 혈액 세포의 형태를 정확하고 객관적으로 평가하고 자동화하는 기술이 진단에 큰 도움을 줄 것으로 판단된다.

혈구 중에서 적혈구는 일반적으로 원반 모양의 형태를 띠고 있다. 그러나 혈액 내의 점도, 삼투압, 칼슘, 산화 질소, 온도 등과 같은 혈액 내 환경에 따라서 가역적으로 형태를 변경할 수 있는

독특한 변형성을 가지고 있기 때문에, 미세혈관을 통과하여 가스를 운반할 수 있다^[2]. 적혈구의 변형성으로 인해 심혈관 질환과 노화에 의해 혈액 내 환경이 변화하면 적혈구의 기형성이 증가한다. 따라서 적혈구의 기형성 평가를 통해서 다양한 질환을 진단할 수 있다. 대표적으로 급성 관상 동맥 증후군(ACS), 허혈성 뇌혈관 질환, 말초 동맥 질환(PAD), 심부전(HF), 심박 세동(AF)과 같은 급성 심혈관 질환뿐만 아니라 고혈압과 당뇨, 빈혈 같은 만성 질환 등을 진단할 수 있다.

혈구의 형태학적 분석은 세포의 크기, 염색양상 및 형태학적 특성을 관찰하여 기술하며 일반적으로 이상이 있는 혈구는 세포 각각의 형태학적 이상의 정도나 이상혈구의 비율 혹은 두 가지를 모

* Corresponding author. Tel.: +82-33-250-6371

E-mail address: yjpark@kangwon.ac.kr (Yong-Jai Park).

두 고려한다^[3-5]. 앞에서 언급한 바와 같이 일관성 있고 각 형태의 등급을 보고하기 위해서는 검사자의 숙련도가 높아 일관되며 검사자에 따라 주관적인 형태학적 특성 조건과 정보를 종합하여 분석하게 된다^[5]. 적혈구의 변형 비율을 보다 정확하게 측정하기 위해서는 수기법으로 변형적혈구를 감별하고 계산하는 과정이 필요하다. 수기법은 형태학적 판정에서 객관적인 정보가 제공되나 전문화된 인력이 필요하고 검사에 소요되는 시간이 많아 인력소모 또한 많다^[6]. 또한 관찰한 세포의 수가 부족한 경우에는 검사결과를 신뢰할 수 없다는 단점이 있다^[7]. 즉 수기법으로 변형 적혈구를 판정하는 것은 현실적으로 많은 어려움을 동반하고 있다. 또한 수기법을 한다고 하더라도 일관된 기준을 유지하기 위한 많은 노력이 필요하다. 형태학적 이상을 글로 표현하거나 한정된 사진 정보로만 이상 정도를 규정하고 검사자의 판단기준과 일치시키는 것은 어려운 측면이 있다. 변형 적혈구를 구별하기 위한 근본적인 해결 방법은 구체적인 수치를 기반으로한 변형적혈구의 판단기준이 필요한 것이다. 다만 아직 포괄적인 기준이 없는 실정이다^[5].

적혈구의 기형성 진단을 위해 적합한 예시로 혈액의 환경에 문제가 있는 대사 증후군과 이와 연관된 질환인 당뇨, 간질환, 혈관 질환 등이 있다. Gyawali, P. et al.^[8]와 Turchetti, V^[9]에 따르면 정상 및 비정상 적혈구의 비율은 89.60:10.4이다. 또한 적혈구의 형태와 염색성에 따른 종류와 심각성에 대하여 Bowls, Discocytes, Kinizocytes, Echinocytes 등으로 나뉘어진다^[10]. 이것에서 marked에 구분될수록 심각한 상태이다. 이와 관련된 보고를 기반으로 정상 및 비정상의 적혈구의 기형성 평가를 활용하기에 용이할 것이다.

본 논문에서는 YOLOv5모델에 혈액 이미지 데이터를 적용하여 정상 및 비정상 적혈구의 비율을 찾아내는 알고리즘을 개발하였다. 정상 및 비정상 적혈구의 비율을 이용하여 다양한 혈액 질환 진단에 유용하게 활용할 수 있을 것으로 기대된다.

2. 방법

2.1 데이터 전처리

이미지 데이터에서 다양한 형태의 비정상 적혈구의 비율을 높이기 위해 자체적으로 합성하였다. 합성 적혈구 영상은 BCCD (blood cell count dataset) 데이터를 기반으로 하여 적혈구 영상에 비정상 적혈구를 추가함으로써 제작되었다. 비정상 적혈구는 질병을 가진 포유류(e.g. 토끼, 개)의 적혈구 영상 중 비정상 적혈구를 추출하여 크기를 조절하여 삽입했다^[11-13]. 그리고 이를 바탕으로 임의의 기존 적혈구 형태를 왜곡시켰다. 비정상 적혈구를 합성하기에 앞서 혈구 및 혈소판을 제외한 혈액 도말 영상의 배경을 편의성을 위해 흰색으로 치환했다. 그리고 원래 적혈구 영상에서

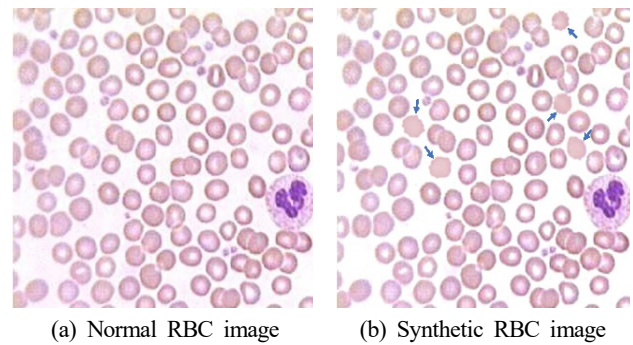


Fig. 1 RBC image (a) normal RBC image, (b) synthetic RBC image. Blue arrows indicate synthetic RBCs imported from and referenced from RBCs of diseased animals

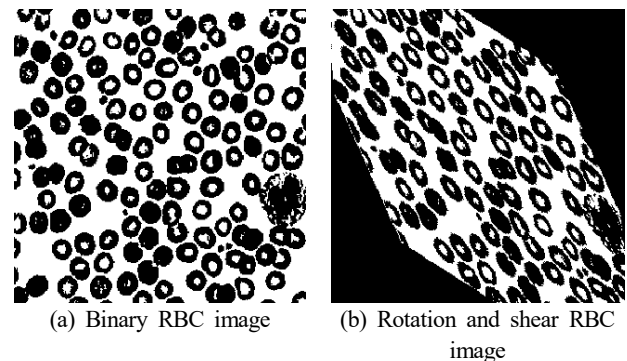


Fig. 2 Augmentation RBC image (a) binary RBC image, (b) rotation and shear RBC image

혈구 및 혈소판을 지우거나 빈 공간에 비정상 적혈구를 합성했다. 원래의 영상과 합성 영상 간의 채도를 조절하여 유사하게 보이도록 하였고, 결과적으로 Fig. 1에 나타난 합성 적혈구 영상을 얻었다. 모든 과정은 Adobe Photoshop 프로그램을 사용하여 진행했다.

본 연구에서는 부족한 데이터 수를 보완하기 위해 기존의 혈액 도말 영상을 전처리하여 사용하였다. 영상은 MATLAB에서 처리되었고, 이진화, 회전, 전단 변형, 밝기 변화 순으로 진행했다. 먼저 RGB 영상은 각각의 색에 다른 가중치를 곱하여 회색조의 영상으로 전환하였다. 그리고 threshold 값을 구하는 방법 중 threshold 값이 이진화의 기준으로 사용되는 Otsu 방법으로 얻은 전역 임계값을 이용해서 이진화 된 혈액 도말 영상을 얻었다(Fig. 2(a)). 이를 $\pm 30^\circ$ 범위 내에서 임의의 값으로 회전시켰고, 영상의 크기는 그대로 유지토록 하였다. 그리고 전단은 수평방향 $\pm 9^\circ$, 수직 방향 $\pm 31^\circ$ 이내의 범위에서 각각 임의의 값을 선택하여 적용하여 변형시켰고 Fig. 2(b)와 같이 영상을 얻었다. 마지막으로 현재 영상의 $\pm 10\%$ 범위에서 임의의 값을 선택되어 밝기가 조절되었다.

2.2 데이터 레이블링

Normal, abnormal, series, cut 클래스로 분류하고 추후 결과

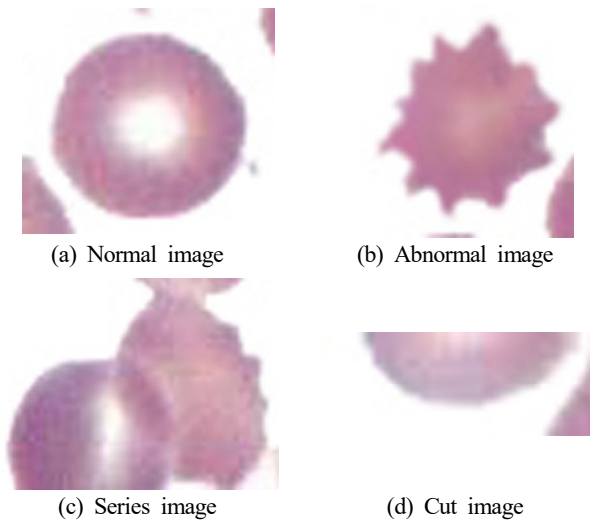


Fig. 3 Labeling RBC image (a) normal image (b) abnormal image (c) series image (d) Cut image

비교를 위해 normal과 abnormal 클래스의 결과만을 가지고 와서 결과를 비교하였다. 정상 적혈구는 그림에 보듯이 적혈구의 원반 모양이 선명하고 내부가 비어있는 이미지를 정의하여 레이블링 하였다. 비정상 적혈구는 그림에 보듯이 톱니 모양의 적혈구, 타원형 적혈구, 저색소성 적혈구, 구상적혈구 내부가 비어있는 적혈구 등 정상에서 벗어난 모든 적혈구의 형상을 abnormal 클래스로 분류하여 레이블링 하였다. 연결된 적혈구는 정상 및 비정상 구별이 모호하여 추후 정상 비정상의 비율 등의 진단 결과에 영향을 끼칠 수 있으므로 series라는 하나의 클래스로 분류하여 레이블링한 뒤 결과에서 제외할 수 있도록 하였다. 사진상 가장자리에 있는 적혈구는 전체 모양이 아닌 일부분이 잘려서 나와 있으므로 연결된 적혈구와 마찬가지로 정상 및 비정상의 구별이 모호하므로 cut이라는 하나의 클래스로 분류한 뒤 결과에 제외할 수 있도록 하였다.

2.3 YOLOv5 모델 설정

YOLOv5는 물체 인식을 목표로 개발된 모델로서 실시간으로 물체 인식이 가능할 만큼 빠른 응답성과 기존의 물체 인식 모델과 비교하여도 낮은 정확성을 특징으로 가지고 있는 모델이다. YOLOv5는 알고리즘 훈련을 위해 기본 이미지와 레이블링 정보를 입력받아 초기의 이미지를 NxN grid cell로 나누고 각 Cell마다 객체를 예측한다. 객체의 예측 정보를 토대로 boundary boxes를 예측하여 각 box에는 위치 좌표 및 confidence score, 클래스에 대한 확률의 정보가 담긴다. 이러한 정보를 바탕으로 객체 인식 모델을 개발하고 validation set을 이용하여 모델을 검증한다. 본 논문에서는 416×416 크기의 73장의 이미지 데이터를 이미지 처리 방식을 적용하여 증강한 후 train data 70%, validation data 20%, test data 10%로 나누어 YOLOv5 모델에 적용하였다.

Table 1 Specification of hardware & software

Configuration	Specifications
CPU	Intel i7-8700
GPU	Nvidia GTX 1050ti
Language	Python 3.7.1
Framework	Pytorch 1.5.0

batch size 16, epoch 1000 적용하였다. 모델 구조는 yolov5s, yolov5m, yolov5l, yolov5x가 있으며, 각각의 모델은 시간에 비례해서 정확도가 올라가는 모습을 보여준다. 본 논문에서는 yolov5m 이후의 모델 구조를 적용 시에 정확도가 비슷하였으며, 시간과 정확도를 고려하여 yolov5m을 적용하였다. YOLOv5 모델에 사용된 하드웨어와 소프트웨어는 Table 1과 같다.

3. 결과

3.1 알고리즘 검증

YOLOv5 모델을 검증하기 위해서 혼동행렬(confusion matrix) 개념을 도입한다. 혼동행렬은 알고리즘의 예측 성능을 통계학적인 표로 정리하여 시각적으로 알 수 있는 지표로 사용되고 있다. 혼동행렬은 레이블링을 통해 정의한 실제 참값을 알고리즘을 통해 예측된 값과 비교하여 다음과 같이 나타낼 수 있다.

- TP : 실제 True 값을 True로 예측
- TN : 실제 False 값을 False로 예측
- FP : 실제 False 값을 True로 예측
- FN : 실제 True 값을 False로 예측

혼동행렬의 구성 요소를 이용하여 정확도(accuracy), 정밀도(precision), 재현율(recall)을 나타내어 알고리즘을 수치적으로 검증한다.

정밀도(precision)는 알고리즘이 True라고 예측한 값 중에서 실제 값 True의 비율로 Eq. (1)과 같이 나타낸다. Fig. 4는 confidence 값에 따른 정밀도의 값을 보여준다. 각각의 클래스의 정밀도의 값은 series, abnormal, cut, normal 순으로 confidence 값 약 0.8에서 높은 precision 값에 도달하는 것을 확인하여 각각의 클래스에 대한 알고리즘의 예측 성능이 얼마나 정밀한지 나타내었다.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

재현율(recall)은 실제 값의 True 중에서 알고리즘이 True라고 예측한 값의 비율로 Eq. (2)와 같이 나타낸다. Fig. 5는 confidence

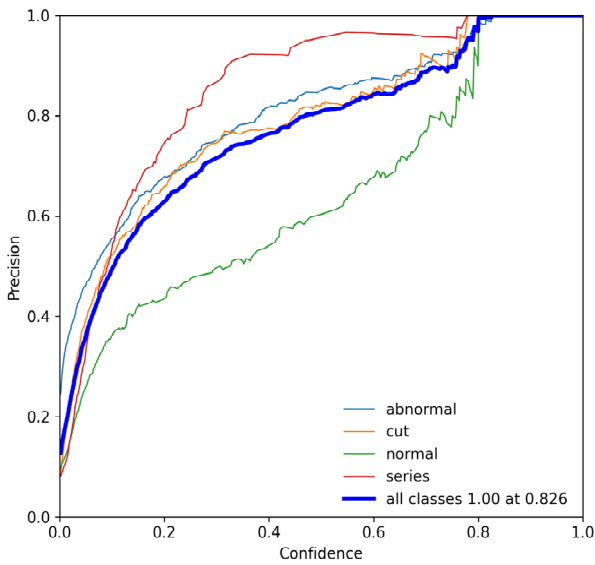


Fig. 4 Precision curve of algorithm

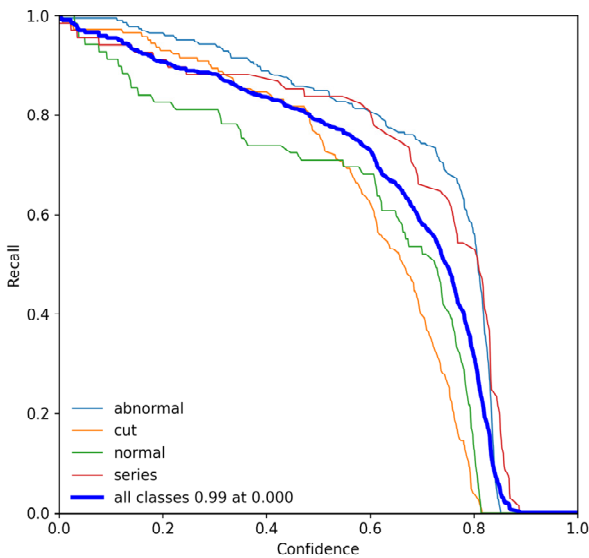


Fig. 5 Recall curve of algorithm

값에 따른 재현율의 값을 보여준다. 각각의 클래스의 재현율의 값은 confidence 값 약 0.8에서 0으로 수렴하는 것을 확인하여 알고리즘의 완전성을 나타내었다.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

정밀도와 재현율 지표를 통해 mean Average Precision (mAP)를 구하여 각 클래스 별로 정확도를 구하였다. Fig. 6는 정밀도와 재현율을 비교한 그래프이며 여기서 각 클래스 별의 그래프 선 아래쪽의 면적을 계산하여 mAP를 구하였다. 그 결과 abnormal은

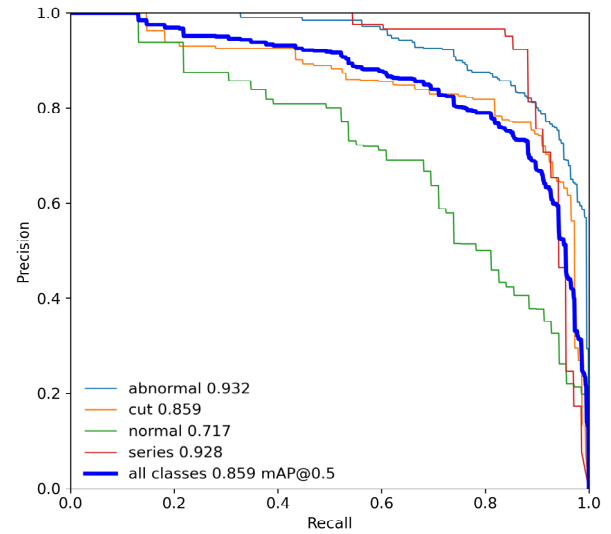


Fig. 6 Precision-recall curve of algorithm

93.2%, cut은 85.9%, normal은 71.7%, series는 92.8%의 정확도를 확인하였다. 마지막으로 앞서 구한 정밀도와 재현율을 통하여 F1 score를 Eq. (3)에 따라 구하였다.

$$(F1 \ score) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

그 결과 abnormal은 96.5%, cut은 92.4%, normal은 83.5%, series는 96.3%의 F1 score를 확인하였다. 기존의 연구에서는 약 96%의 F1 score를 보여 abnormal과 series는 96%로 비슷한 값을 보였지만 cut는 약 4%, normal의 경우는 13% 낮은 F1 score

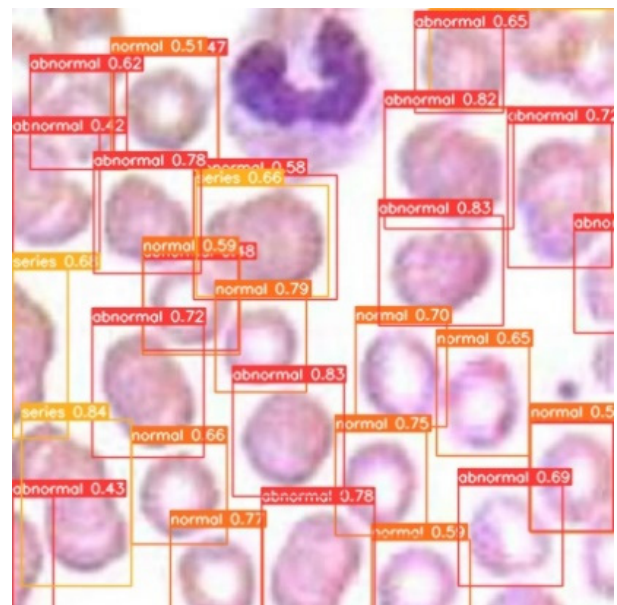


Fig. 7 Applying an algorithm to test data 1

Table 2 The predicted number of normal/abnormal RBC by test data

	Normal	Abnormal	Normal/Abnormal
data 1	17	10	0.59
data 2	19	6	0.32
data 3	23	9	0.39
data 4	19	10	0.53
data 5	21	6	0.29
data 6	32	2	0.06
data 7	21	30	1.43

Table 3 The number of synthetic labeled as abnormal RBC and synthetic predicted as abnormal RBC in test data

	Synthetic	Predicted synthetic
data 1	0	0
data 2	2	2
data 3	0	2
data 4	2	3
data 5	4	4
data 6	5	5
data 7	4	4

값을 보였다.

3.2 알고리즘 적용

제작한 알고리즘을 test data에 적용하여 각각의 클래스를 예측하고 정상/비정상 비율을 구하였다. Fig. 7의 알고리즘이 예측한 이미지를 보면 모든 적혈구에 대해 bounding box가 표시되어 적혈구에 대하여 정확하게 판단하는 것을 볼 수 있다. 또한, 정상 적혈구의 비해 비정상 적혈구의 비율이 높은 것을 볼 수 있으나 전체적으로 확률이 비정상 적혈구가 더 낮은 것을 확인하였다. 비율을 구할 때에는 seires와 cut 클래스를 제외한 후 정상/비정상의 비율을 구하였으며 각각의 데이터에 대해서는 Table 2에 정리하였다. Test data의 혈액 이미지에서 합성적혈구는 평균적으로 2~3개가 추가되었다. 총 7개의 test data 중 data 3, 4의 클래스 분류 과정에서 경계에 위치한 적혈구들은 cut 클래스로 분류되었고, 합성적혈구도 예외로 취급되지 않았다. 그래서 abnormal 클래스를 가진 합성적혈구의 수는 data 3에서는 0개, data 4에서는 2개로 나타났다. 하지만, 예측 결과에서는 abnormal 클래스로 분류된 합성적혈구는 각각 2개, 3개로 예측되었다. 이는 cut 클래스로 분류되었던 합성적혈구가 예측 data에서는 abnormal 클래스로 분류되었기 때문이다. 나머지 test data에서는 합성적혈구를 예측이 정확한 것으로 확인되었다. test data 라벨링 과정에서 합성적혈구 개수와 예측된 결과에서 합성적혈구의 수는 Table 3에

기재했다.

4. 결론

본 논문에서는 혈액 세포의 형태를 객관적인 평가를 위해 YOLOv5를 이용하여 진단이 가능하도록 정상/비정상 적혈구를 구별하는 알고리즘을 개발하고 적용하였다. 데이터의 개수를 확보하기 위하여 이미지 전처리를 통해 데이터를 증가하였다. 데이터는 정상과 비정상의 비율을 계산하기 위해 정상, 비정상 이외의 형태는 제외되어야 하므로 normal, abnormal 클래스와 제외되는 클래스인 series, cut으로 총 4개의 클래스로 나누어 레이블링하였다. 알고리즘의 성능 평가를 위해 정밀도와 재현율을 지표로 적용하여 각각의 클래스에 대한 mAP를 구하여 클래스별 정확도는 abnormal은 93.2%, cut은 85.9%, normal은 71.7%, series는 92.8%를 확인하였다. 알고리즘을 실제로 적용하여 모든 적혈구가 bounding box로 표시되는 것을 확인하였으며, 적혈구의 정상과 비정상 비율을 구하였다.

본 논문에서는 적은 데이터를 가지고도 적혈구의 정상과 비정상을 확인하고 비율을 구하는 알고리즘을 개발하였다. 향후 데이터를 추가로 확보하고 정상 및 비정상에 대한 정확한 레이블링을 통해 전체적인 정확도를 높여 당뇨 빈혈 등의 질환에 대한 진단이 가능할 것으로 보인다.







후 기

이 연구는 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 한국연구재단에서 부여한 과제번호: 2020R1I1A3073575/NRF-2020R1F1A1072926).

References

- [1] Merino, A., Puigví, L., Boldú, L., Alférez, S., Rodellar, J., 2018, Optimizing Morphology through Blood Cell Image Analysis, *Int. J. Lab. Hematol.*, 40:1 54-61, <https://doi.org/10.1111/ijlh.12832>.
- [2] Kim, J. H., Lee, H. Y., Shin, S. H., 2015, Advances in the Measurement of Red Blood Cell Deformability: A Brief Review, *J. Cell. Biotechnol.*, 1:1 63-79, <https://doi.org/10.3233/JCB-15007>.
- [3] Choi, J. M., Lee, W. S., 2013, Development of a Poikilocyte Measuring Method Using Image Analysis Software, *Lab. Med. Online*, 3:6 6-14, <https://doi.org/10.3343/lmo.2013.3.1.6>.
- [4] Hyun, B. H., Gulati, G. L., Ashton, J. K., 1991, Differential

- Leukocyte Count: Manual or Automated, What Should It Be?, *Yonsei Med. J.*, 32:4 283-291, <https://doi.org/10.3349/ymj.1991.32.4.283>.
- [5] Taylor, J. J., 2010, Blood Cell Morphology: Grading Guide, *Arch. Pathol. Lab. Med.*, 134:8 1229-1230, <https://doi.org/10.5858/2009-0646-BK.1>.
- [6] Briggs, C., Longair, I., Slavik, M., Thwaite, K., Mills, R., Thavaraja, V., Foster, A., Romanin, D., Machin, S. J., 2009, Can Automated Blood Film Analysis Replace the Manual Differential? An Evaluation of the CellaVision DM96 Automated Image Analysis System, *Int. J. Lab. Hematol.*, 31:1 48-60, <https://doi.org/10.1111/j.1751-553X.2007.01002.x>.
- [7] Rümke, C. L., 1985, The Imprecision of the Ratio of Two Percentages Observed in Differential White Blood Cell Counts: A Warning, *Blood Cells*, 11:1 137-140.
- [8] Gyawali, P., Richards, R. S., Bwititi, P. T., Nwose, E. U., 2015, Association of Abnormal Erythrocyte Morphology with Oxidative Stress and Inflammation in Metabolic Syndrome, *Blood Cells Mol. Dis.*, 54:4 360-363, <https://doi.org/10.1016/j.bcmd.2015.01.005>.
- [9] Turchetti, V., De Matteis, C., Leoncini, F., Trabalzini, L., Guerrini, M., Forconi, S., 1997, Variations of Erythrocyte Morphology in Different Pathologies, *Clin. Hemorheol. Microcirc.*, 17:3 209-215.
- [10] Constantino, B. T., 2015, Reporting and Grading of Abnormal Red Blood Cell Morphology, *Int. J. Lab. Hematol.*, 37:1 1-7, <https://doi.org/10.1111/ijlh.12215>.
- [11] Papasouliotis, K., Murphy, K., 2021, Pictorial Guide to Canine and Feline Blood Smears, Part 2: Abnormal Findings – Erythrocytes, *In Practice*, 43 363-372, <https://doi.org/10.1002/inpr.99>.
- [12] Christopher, M. M., Hawkins, M.G., Burton, A.G., 2014, Poikilocytosis in Rabbits: Prevalence, Type, and Association with Disease, *PloS One*, 9:11 e112455, <https://doi.org/10.1371/journal.pone.0112455>.
- [13] Lenske, E., Padula, A. M., Leister, E., Boyd, S., 2018, Severe Haemolysis and Spherocytosis in a Dog Envenomed by a Red-bellied Black Snake (*pseudechis porphyriacus*) and Successful Treatment with a Bivalent Whole Equine IgG Antivenom and Blood Transfusion, *Toxicon*, 151 79-83, <https://doi.org/10.1016/j.toxicon.2018.07.005>.

	<p>Jin-Gu Kim M.Sc. candidate in the Department of Interdisciplinary Program in Biohealth-machinery convergence engineering, Kangwon National University. His research interest is Machine. E-mail: jink4763@gmail.com</p>
	<p>Ji-Hun Kang M.Sc. candidate in the Department of Interdisciplinary Program in Biohealth-machinery convergence engineering, Kangwon National University. His research interest is Machine. E-mail: jihunkang0426@kangwon.ac.kr</p>
	<p>Kyoung-Cheol Choi M.Sc. candidate in the Department of Interdisciplinary Program in Biohealth-machinery convergence engineering, Kangwon National University. His research interest is biological. E-mail: crude925@gmail.com</p>
	<p>Woong-Ki Jang Post-doctoral associate in the Department of Smart Health Science and Technology, Kangwon National University. His research interest is Micro/nano scale surface texturing technologies and digital twin. E-mail: wkddndrl@kangwon.ac.kr</p>
	<p>Ho-Jin Ha Professor in the Department of Smart Health Science and Technology, Kangwon National University. His research interest is Machine. E-mail: hojinha@kangwon.ac.kr</p>
	<p>Kwang-Suk Lim Professor in the Department of Smart Health Science and Technology, Kangwon National University. His research interest is Antibody-drug conjugate, immuno-cancer cell therapy / gene and drug delivery system E-mail: kslim@kangwon.ac.kr</p>



Byeong-Hee Kim

Professor in the Department of Smart Health Science and Technology, Kangwon National University. His research interest is micro and nano system design and precision control of machine tools.

E-mail: kbh@kangwon.ac.kr



Yong-Jai Park

Professor in the Department of Smart Health Science and Technology, Kangwon National University. His research interest is Soft Robotics, Novel Mechanism Design, Bio-inspired Robot and Robot Application

E-mail: yjpark@kangwon.ac.kr