

# CNN 을 활용한 수박 당도 예측

강다영, 김채민, 유근영, 이다형, 김현희  
동덕여자대학교 정보통계학과

die\_nacht25@naver.com, dwu20191042@gmail.com, u102402@naver.com, leeda99@naver.com, heekim@dongduk.ac.kr

## Prediction of the Sugar Content of Watermelon based on Convolutional Neural Network

Da-Young Kang, Chae-Min Kim, Geun-Young Yoo, Da-Hyung Lee, Hyon Hee Kim  
Department of Statistics and Information Science, Dongduk Women's University

### 요 약

수박의 이미지와 수박의 무게 데이터를 활용해 수박의 당도를 예측하고 모델의 정확도를 측정한다. 과피가 얇고, 부피가 작은 과일의 경우 휴대용 비파괴 당도 측정기를 통해 비교적 간편하게 당도 측정이 가능하다. 하지만 수박은 과피도 두껍고, 부피도 크기 때문에 넓은 장소와 비용을 부담해야 하는 선별장에만 당도를 측정할 수 있는 실정이다. 본 논문에서는 줄무늬가 끊어지지 않고, 원형이 아닌 타원형이 맛있는 수박이라는 속설에 부합하는 수박이 실제로 맛있는 수박인지를 확인하고자 수박 이미지를 수집하여 당도에 따라 이미지를 분류한 다음, CNN 을 적용하여 수박 당도 예측을 실시하였다. 실험 결과 타원형 수박은 당도가 높은 것으로 나타났으나 줄무늬가 끊어진 수박과 끊어지지 않은 수박 간의 당도 차이는 없는 것으로 나타났다. 향후 수박의 당도에 영향을 미칠 수 있는 다양한 변수를 활용하여 정확도를 높인다면 현재 사용되고 있는 비파괴 당도 측정기를 보완할 수 있을 것으로 기대된다.

### 1. 서론

최근 인공지능 기술의 발전에 따라 이미지 인식 분야에서 딥러닝 기술이 가장 주목받고 있다[1]. 이에 발맞추어 농업 분야에서도 이를 활용한 다양한 연구가 진행되고 있다. 과일의 경우 당도가 소비자의 구매 결정에 큰 영향을 미치는데, 사과, 배, 귤과 같은 작은 과일의 경우 휴대용 비파괴 당도 측정기를 활용할 수 있지만, 수박과 같은 부피가 큰 과일의 경우 이를 통한 방법이 불가능하다. 따라서 수박의 경우, 농가에서 사용하는 거대한 분류 장치인 비파괴 당도 측정기를 통한 분류에 의존할 수밖에 없다. 하지만 모든 수박 농가에서 비파괴 당도 측정기를 사용하는 것은 아니며, 이를 사용하여 분류했다라도 소비자가 수박을 구매할 때 본인이 구매하는 수박의 당도가 어느 정도인지 파악하고 구매하기는 힘들다. 위의 기술을 수박 이미지 데이터로 당도를 측정하는 학습에 적용하여 수박의 사진만으로 당도를 예측할 수 있다면, 소비자들이 편리하게 이용할 수 있을 것이다. 본 연구에서는 수박 이미지 데이터를 기반으로 CNN

이미지 인식 및 분류 알고리즘을 이용하여 수박의 당도를 예측해 주는 시스템을 제시한다. 수박 이미지와 수박의 무게 데이터를 통해 학습 모델을 구축하고 이미지로만 이진 분류와 삼진 분류 그리고 이미지와 무게 데이터를 혼합한 데이터를 이용하여 이진 분류와 삼진 분류를 각각 시행한 후 이미지 학습을 통해 수박의 당도를 예측하는 모델의 정확도를 비교해 보고자 한다.

향후 이 모델을 지속해서 발전시킨다면, 일차적으로 소비자는 구매하는 수박의 당도를 직접 확인하고 구매할 수 있게 될 것이다. 또한 수박 농가에서도 더는 많은 공간과 비용이 있어야 하는 비파괴 당도 측정 과정을 거치지 않아도 된다. 애플리케이션을 통해 당도를 확인할 수 있게 되면 수박을 수확하지 않은 채로 개별적 수박의 당도를 예측할 수 있어 적절한 수확시기에 대한 대응을 할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 사용된 데이터의 수집 과정과 분류 기준 및 속성에 관해 설명한다. 제 3 장에서는 본 연구에서 제시한 CNN 모델의 구조에 관해 설명한다. 제 4 장에서는 성능

측정과 무게와 당도 간의 상관관계에 대해 보이고, 제 5 장은 결론 및 향후 연구를 제시한다.

## 2. 관련 연구

과일의 당도는 다양한 방법을 통해 분류되어 왔다. 과일 색상 선별 알고리즘, 사용자 Interface 및 선별 상태 표시 알고리즘을 통해 예측하거나[2], 과피가 얇은 과일의 경우 근적외선을 이용하여 당도를 측정하는 반사식 비파괴 당도 측정기를 통해 당도를 측정하고, 이 값을 통해 분류할 수 있다. [3]

앞의 연구처럼, 과일의 분류 방법에 대한 연구는 많은 알고리즘을 통해 발전되어 왔다. 본 연구에서는 상대적으로 작은 과일에만 한정되어 있던 당도 분류 모델을 부피가 큰 수박에도 적용될 수 있다는 것을 보인다. 이를 위하여 수박 이미지 데이터를 CNN 알고리즘[4]을 통해 당도 기준으로 학습하고 분류한다는 점에서 기존 연구들과 차별점이 있다.

## 3. 실험 설계

### 3.1 데이터 수집 과정 및 분류 예측 기준

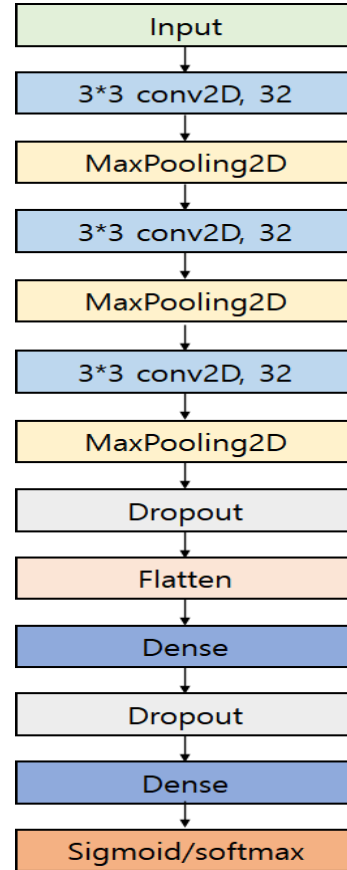
본 연구에서 사용한 데이터는 2021년 6월 충북 음성에 위치한 대소농협 협동조합 수박 선별장의 약 1,000개의 수박으로, 오차를 최대한 줄이기 위해 직접 동일한 위치에서 3초 간격으로 사진 촬영을 진행하였다. 촬영한 수박의 이미지와 비파괴 당도 측정기에서 측정한 각 수박의 당도와 무게로 번호를 매겨 매치한 뒤 ‘번호\_당도\_무게’ 형식으로 이미지와 엑셀을 정리하였다. 그리고 이진 분류 학습을 위해 이미지들을 수박의 당도가 11.5 Brix 보다 낮으면 ‘bad’, 그보다 높으면 ‘good’로 분류하였다. 또한 삼진 분류 학습을 위해서는 11.0 Brix 보다 낮으면 ‘bad’, 11.0~12.0 Brix 사이는 ‘good’, 12.0 Brix 이상은 ‘best’로 분류하였다.

### 3.2 CNN 모델 구조

본 논문에서는 모델의 정확도 측정을 위해 (그림 1)의 구조를 갖는 CNN 모델을 제안하고 학습을 진행하였다.

(그림 1)은 본 연구에서 사용된 CNN 모델의 구조이다. 모델의 입력(Input)은 965개의 이미지 데이터이며, 모델의 출력(Output)은 해당 이미지가 속한 분류 카테고리이다. 각 입력 데이터는 padding= ‘same’, activation= ‘relu’인 3\*3 filter를 이용한 Conv2D를 거친 후 filter 값 중 제일 큰 값을 가져와

resize 하는 MaxPooling2D를 수행한다. 이를 반복한 후 Flatten 하여 모델을 1차원 배열로 변환한 후 256개의 층을 쌓은 후 dropout 정규화를 거친다. 마지막으로 1층을 더 쌓은 후 sigmoid/softmax 활성화 함수 과정을 수행한 뒤 분류 결과를 출력하는 모델을 만들었다.



(그림 1) CNN 모델 구조

## 4. 실험 결과

### 4.1 성능 평가

실험을 위해서 구글의 colab을 사용하였으며, 이미지만을 이용한 이진 분류와 삼진 분류, 그리고 이미지와 수박 무게를 혼합한 이진 분류와 삼진 분류로 나누어 실시하였다. 실험 결과 모델의 성능은 (표 1)로 정리하였다.

	이미지 이진분류	이미지 삼진분류	이미지+kg 이진분류	이미지+kg 삼진분류
accuracy	0.89	0.77	0.82	0.69
loss	0.31	0.63	0.51	0.85

(표 1)

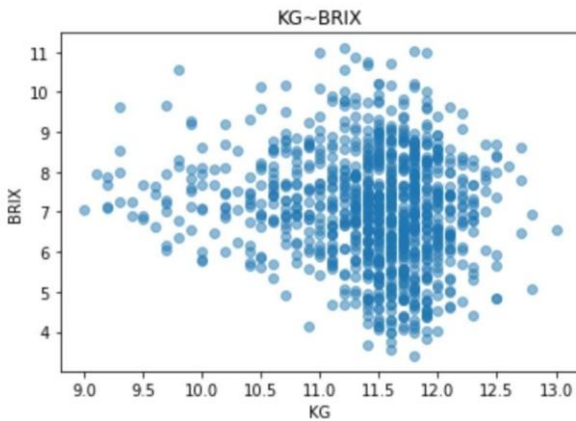
먼저 이미지만을 이용해 실험을 진행했을 때, accuracy 와 loss 가 이진 분류와 삼진 분류에서 각각 (0.89 와 0.31), (0.77 과 0.63)이 나오는 것을 확인할 수 있었다. 또한 이미지 데이터에 ‘무게’ 속성을 추가하여 이진 분류와 삼진 분류를 진행해보았을 때, accuracy 와 loss 가 이진 분류와 삼진 분류에서는 각각 (0.82 와 0.51), (0.69 와 0.85)로 다소 차이가 있었다. (표 1)을 통해 해당 실험에서 이미지만을 사용하여 이진 분류를 실행한 모델의 정확도가 0.89로 가장 높게 나타났다.

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

(수식 1)

(수식 1)은 피어슨 상관계수의 공식이다. [5]

이를 사용하여 무게와 당도 사이의 상관계수를 구해보았다. Brix 와 kg 간의 상관계수는 -0.1297 이며 (그림 2)로 이를 확인할 수 있다. 이것으로 보아 무게와 당도는 서로 영향을 주지 않는다는 것을 알 수 있다.



(그림 2) kg-brix 상관관계 그래프.

#### 4.2 실제 데이터로 예측

흔히 줄무늬가 끊어지지 않고 수박의 모양이 타원형에 가까울수록 맛있는 수박이라는 속설이 존재한다. 이 속설처럼 실제로 둥근 수박과 타원형 수박, 줄무늬가 끊어지지 않은 수박과 끊어진 수박 중 학습 모델이 어느 것을 맛있다고 예측할지 실제 데이터로 실행해 보았다. 비교적 높은 성능을 보여준 이미지 데이터만 이용한 이진 분류 모델과 삼진 분류 모델에 적용해 보았다. 이미지만으로 이진 분류하는

모델에 (그림 3)과 같은 이미지를 적용하여 예측해 보았다. 그 결과 둥근 수박의 당도를 ‘bad’, 타원형 수박의 당도를 ‘good’ 이라고 예측하였고 이는 각 수박의 실제 값과 같았다. 그다음 이미지로만 삼진 분류하는 모델에 (그림 3)과 같은 이미지를 적용하여 예측해 보았다. 그 결과 둥근 수박의 당도를 ‘good’, 타원형 수박의 당도를 ‘best’ 라고 예측하였고 이 또한 각 수박의 실제 값과 같았다.



(그림 3) 둥근 수박 vs 타원형 수박

이미지만으로 이진 분류하는 모델에 (그림 4)과 같은 이미지를 적용하여 예측해 보았다. 그 결과 두 수박의 당도를 모두 ‘bad’ 라고 예측하였다. 줄무늬가 끊어진 수박은 수박의 실제 당도와 같았지만, 줄무늬가 끊어지지 않은 수박은 실제 당도와 차이를 보였다. 그 다음 이미지로만 삼진 분류하는 모델에 (그림 4)와 같은 이미지를 적용하여 예측해 보았다. 그 결과 줄무늬가 끊어진 수박과 줄무늬가 끊어지지 않은 수박 모두 ‘best’ 라고 예측하였다. 하지만 실제로 각 수박의 당도를 체크했을 때 줄무늬가 끊어진 수박의 당도는 ‘bad’ 로 예측과 달랐지만, 줄무늬가 끊어지지 않은 수박의 당도는 ‘best’ 로 일치하였다.



(그림 4) 줄무늬가 끊어진 수박 vs 끊어지지 않은 수박

#### 5. 결론

본 연구에서는 수박의 이미지를 학습시키고, 이를 특정 당도를 기준으로 분류한 뒤 CNN 알고리즘을 통해

당도 예측의 정확도를 측정하고자 하였다. 이진 분류와 삼진 분류를 진행하였을 때 0.89 와 0.77 로 이진 분류의 정확도가 더 높았다.

또한, '무게(kg)'라는 속성을 추가했을 때의 성능을 측정한 결과 이진 분류의 경우 0.82, 삼진 분류의 경우 0.69 로 해당 속성을 추가하지 않았을 때가 더 높은 정확도를 보였다. 그래서 당도와 무게 간의 상관계수를 구해본 결과,  $-0.1297$  로 무게는 수박의 당도에 영향을 끼치지 않는다는 것을 확인할 수 있었다.

해당 시스템에 실제 데이터를 적용해 예측해 본 결과 둥근 수박과 타원형 수박은 이진 분류 모델, 삼진 분류 모델 모두 예측을 잘하였지만, 줄무늬가 끊어진 수박과 끊어지지 않은 수박의 경우 두 모델 모두 정확한 예측을 하지 못하였다. 이는 같은 농장에서 수확한 수박 이미지만을 학습하였기 때문에 수박 당도 결정에 영향을 미칠 수 있는 다른 속성들을 고려하지 못했다는 한계점으로 보였다. 따라서 향후 연구에선 다양한 조건의 수박 데이터를 수집하여 기후, 지형 등의 수박 당도에 영향을 미치는 속성값을 추가하고자 한다. 추가적인 속성을 적용하여 학습을 진행한다면 더 좋은 성능을 기대할 수 있을 것으로 보인다.

### 참고문헌

- [1] Soo-Ho Jeong, Meong-hun Lee, Hyun Yoe, "Fruit Classification System Using Deep Learning", Journal of Knowledge Information Technology and Systems(JKITS), Vol. 13, No. 5, pp. 589-595, October 2018
- [2] Min-Hyeok So, Seung-Hwan Yu, Hack-Yoon Kim, " Fruit Color Sorting System using Image Processing", Korean Institute of Information Technology, pp 20-23, 2013
- [3] SangHoon Lee, JongHa Kim, HyeMin Song, Hyun Kim, "A Study on the Measurement System of Apple Sugar Content Based on Smartphone Camera", The Institute of Electronics and Information Engineers, pp 784-787, 2019
- [4] Nakyong Lee, Jooyeon Kim, Junho Shim, "Empirical Study on Analyzing Training Data for CNN-based Product Classification Deep Learning Model", Society for e-Business Studies 26.1, pp 107-126, 2021
- [5] ChangWoo Hong, Kyeon Hur, "Prognosis of the Remaining Useful Life of a Turbofan Engine Using Deep Neural Network and Pearson Correlation Coefficient", Journal of the KNST, 4(1), pp 78-83, 2021.