

멀티모달 학습 기법을 활용한 고추 건조도 예측 연구

황호석¹, 손승우², 문애경³

¹ 한국전자통신연구원

² 매사추세츠 로웰 대학

³ 한국전자통신연구원

ghtjr991123@naver.com, seungwoo_son@uml.edu, akmoon@etri.re.kr

A Study on the Prediction of Red Pepper Dryness Using Multimodal Learning Techniques

Hoseok Hwang¹, Seung Woo Son², Aekyung Moon³

¹ Electronics and Telecommunications Research Institute

² University of Massachusetts Lowell

³ Electronics and Telecommunications Research Institute

요 약

본 논문은 고추의 건조도를 정확하게 예측하기 위해 이미지 데이터와 시간데이터를 결합한 멀티모달 방식을 제안한다. 기존 이미지 기반 예측은 건조 후반부에서 정확도가 저하되는 한계가 있었으며, 이를 보완하기 위해 이미지 특징과 시간 데이터를 융합하고, 시간 차원의 확장과 가중치 조정을 통해 건조 진행의 동적 특성을 반영한다. 실험 결과, 멀티모달 모델은 단일 이미지 기반 모델보다 평균 3.07% 더 낮은 SMAPE 값이 나오며 우수한 성능을 보였다. 이를 통해 멀티모달 모델이 건조도 예측에 효과적으로 활용될 수 있음을 확인하였다.

1. 서론

고추는 김치, 양념, 고추장 등 다양한 가공식품의 원료로 사용되며 널리 사용되나 수분 함량이 높아 수확 후 품질 저하 또는 부패에 취약하다[1]. 이러한 문제를 해결하기 위해 건조를 하여 고추의 유통기한을 늘리고 더 오랫동안 품질을 유지할 수 있다[2].

햇빛에 말려 건조를 진행하는 전통적인 방법으로는 시간이 많이 걸리고 환경에 따라 최종 수분 함량을 제어하기 어려워 제품 품질이 낮아질 수 있다[3]. 이러한 문제를 해결하기 위해 농산물 건조기가 널리 사용되고 있으며, 이를 통해 기후 조건과 관계없이 대량의 농산물을 빠르게 건조할 수 있다. 그러나 건조가 적절하게 이루어졌는지 확인하기 위해 여전히 사람이 직접 농산물을 꺼내어 상태를 평가해야 하는 비효율적인 과정이 요구된다. 따라서, 건조 과정을 자동화하고 건조도를 정량적으로 평가할 수 있는 방법이 필요하다.

기존 연구에서는 이미지 기반 특징을 추출한 뒤 회귀 분석 기법을 적용하여 건조 시간을 예측하는 방법

이 활용되었다[4]. 이러한 방식은 효과적이거나 건조 후반부에 이미지 변화가 거의 없을 경우 정확도가 감소한다는 한계점이 존재한다[5].

이를 극복하기 위해 최근에는 서로 다른 유형의 데이터를 동시에 활용하는 멀티모달 학습 기법이 주목받고 있다[6]. 고추 건조도 예측의 경우 이미지 데이터는 표면 색상과 형태 변화를 반영하고, 시간 데이터는 건조 진행 상황의 연속성을 나타낸다. 따라서, 두 정보를 결합하면 단일 데이터만 사용하는 것보다 더 정확하고 강건한 예측이 가능하다.

본 연구에서는 이러한 관점을 바탕으로, 시간 차원의 확장과 가중치 조정을 활용한 멀티모달 모델을 설계하였다. 제안된 모델은 건조 과정의 동적 특성을 효과적으로 반영하여 높은 예측 정확도를 확보하고자 한다.

2. 고추 건조도 예측 방법

본 연구에서는 고추 건조도 탐지를 위해 이미지와 시간 데이터를 결합한 모델을 사용하였다. 전체적인

구조는 그림 1과 같으며 들어온 이미지에 대해 객체 분할(object segmentation)모델을 사용하여 고추 영역을 분할한다. 그런 다음 분할된 객체 이미지와 건조 진행 시간을 멀티모달 모델을 사용하여 건조도를 예측한다.



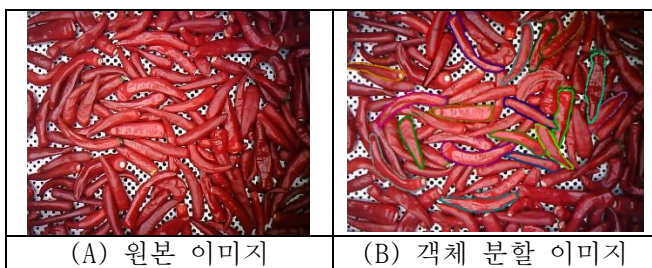
(그림 1) 고추 건조도 예측 흐름도

2-1. 데이터셋

고추 건조도의 실제 값(Ground Truth)을 시간별로 확보하기 위해 주기적으로 건조기에서 고추를 꺼내 사진을 촬영하고 무게를 측정한다. 촬영된 이미지는 시각적 변화를 기록하는 역할을 하고 측정된 무게 값은 건조도의 실제 값으로 사용된다. 고추는 가정용 고추 건조기와 농업용 고추 건조기 2개를 사용하여 촬영되었고 40 시간 동안 1시간 주기로 건조기에 꺼내 촬영 및 각 무게를 측정했다. 그 결과, 총 14,398개의 이미지와 무게 데이터를 수집하고 모델 학습 및 검증에 사용한다.

2-2. 딥러닝 모델을 이용한 객체 분할

건조도 예측은 이미지 전체가 아닌 개별 고추 단위로 수행된다. 그러나, 수집된 원본 이미지에는 여러 개의 고추가 있으며 따라서 고추를 정확히 분할하는 것은 필수적이다. 이를 위해 본 연구에서는 Yolo-seg 모델을 학습 및 사용하여 원본 이미지인 그림 2-A에서 고추 객체를 배경과 분리하여 그림 2-B와 같이 신뢰도 0.75 이상의 여러 개의 고추들이 검출되며 이중 가장 높은 신뢰도를 지닌 4개의 고추 이미지를 분할하여 저장한다[7].



(그림 2) Yolo-seg 모델의 객체 분할

2-3. 데이터 전처리 및 특징 추출

데이터 전처리는 모델 학습 효율 및 일관성 있는 결과를 위해 필수적이다. 따라서, 멀티모달 입력으로 활용되는 시간 데이터와 이미지 데이터에 대해 각각 전처리 과정을 거친 후 CNN을 이용하여 이미지 특징을 추출한다[8].

시간 데이터는 건조 시작 시점부터 종료 시점까지 일정 간격으로 기록되었으며, 값의 범위는 0에서 400 사이로 주어진다. 모델 학습 시 각 변수의 스케일 차이가 최적화 과정에 영향을 주는 것을 방지하기 위해, 본 연구에서는 Min-Max 정규화를 적용하여 모든 값을 0부터 1사이의 구간으로 변환한다[9]. 이렇게 변환된 시간 데이터에 대해 차원 확장 및 가중치를 조절함으로써 가장 적합한 값을 찾는다.

객체 분할이 되어 나온 고추 이미지 데이터는 다양한 크기와 비율로 수집되었기 때문에 모델에 일관되게 적용하기 위해 크기 조정 과정을 수행하였다. 또한, 사전 학습된 모델인 EfficientNet에서 사용한 ImageNet 데이터셋의 평균과 표준 편차를 사용하여 정규화 하여 특징 추출에 균일한 밝기와 색상 분포를 보장한다[10,11]. 이렇게 변환된 이미지 데이터의 특징을 학습하기 위해 CNN 모델인 EfficientNet을 사용하여 분류 계층을 제거하고, 은닉 계층에서 도출된 임베딩 벡터를 이미지 특징으로 추출한다.

2-4. 멀티 모달 기반 고추 건조도 예측

앞선 단계에서 추출된 CNN 기반 이미지 특징은 고추의 색상, 형태, 질감 변화를 반영한다. 그리고 Min-Max 정규화를 거친 시간 데이터는 건조 진행 정도를 나타내는 보조 특성으로 사용한다. 본 연구에서 멀티모달 처리는 이미지 특징과 시간 데이터를 각각 변환한 뒤 결합하는 방식으로 수행된다. 먼저, 원본 이미지는 전처리 과정을 거쳐 CNN을 통해 고차원 벡터로 변환된다. 동시에 건조 시간 데이터는 별도의 완전연결층(Fully Connected Layer)을 통과하여 저차원 벡터로 변환된다. 이후 두 벡터는 결합 연산을 통해 융합되며, 멀티모달 벡터는 다층 퍼셉트론(MLP) 구조의 회귀 모듈에 입력되어 최종적으로 각 고추 단위의 건조도를 0부터 1사이의 범위 내 연속 값으로 출력한다.

3. 실험

3-1. 데이터 준비

객체 분할을 통해 만들어진 데이터는 랜덤하게 train, val, test로 사용된다. 그러나, 일반적인 분할 방식은 각 시점에서의 고추 데이터가 균일하게 분할되지 않는 한계점이 존재한다. 이를 해결하기 위해 각 시점의 고추 데이터들을 그룹화 한 뒤 그룹화된 고추마다 ID를 부여한다. 그리고 이러한 객체 ID를 기반으로 데이터 분할이 수행한다. 이러한 방식을 통해 특정 시점의 데이터가 불균형 하게 분할되는 문제점을 방지하여 일관성 있는 학습 데이터를 얻을 수

있다. 훈련, 검증 및 테스트 세트의 비율은 6:2:2로 설정되며 훈련 세트는 모델 훈련에 사용되며, 검증 세트는 하이퍼파라미터를 조정하고 과적합을 방지하기 위해 조기 정지를 구현하는 데 사용되고 테스트 세트는 모델의 성능을 평가하는 데 사용된다.

3-2. 평가지표

본 연구에서는 멀티모달 기반 고추 건조도 예측 모델의 성능을 정량적으로 평가하기 위하여, 회귀 분석에서 널리 사용되는 MSE(Mean Squared Error), MAE(Mean Absolute Error), RMSE(Root Mean Squared Error), R^2 (R-squared)를 사용한다. 해당 지표들은 예측 값과 정답 값이 차이를 나타내는 지표들로 해당 값이 낮을수록 모델이 정확함을 뜻한다. 그러나 고추 건조도의 특성상, 건조도가 진행될수록 측정 값은 0으로 수렴한다. 이 경우 일반적인 회귀 지표들은 값이 작을수록 상대적인 오차를 크게 반영하지 못하는 문제가 발생한다. 특히, 건조도 값이 낮은 구간에서는 작은 수치 차이가 실제 건조 상태에 큰 의미를 가질 수 있으므로, 보다 민감하게 반응할 수 있는 지표가 필요하다. 이를 해결하기 위해 본 연구에서는 SMAPE(Symmetric Mean Absolute Percentage Error) 추가적으로 사용한다[12]. SMAPE는 수식 1과 같이 예측 값과 실제 값의 차이를 양쪽 값의 평균으로 나누어 상대적인 오차를 백분율로 환산하는 방식으로 정의된다.

$$SMAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|) / 2}$$

(수식 1) SMAPE 수식

SMAPE는 전통적인 MAPE(Mean Absolute Percentage Error)에 비해 실제 값이 0에 가까운 경우 발생할 수 있는 분모 문제를 완화하며, 예측 값과 실제 값이 모두 작은 영역에서도 안정적으로 상대 오차를 측정할 수 있다. 따라서 건조도 예측과 같이 값이 0으로 수렴하는 데이터 특성을 갖는 문제에서 특히 유용하다.

3-3. 실험결과

표 1은 다양한 차원 확장 및 시간 가중치 설정에 따른 모델 성능을 비교한 결과이다. 여기서 모델 표기법 A/B에서 A는 시간 데이터의 차원 확장 정도를, B는 시간 가중치 값을 의미한다. 예를 들어, 1/1은 차원 확장을 적용하지 않고 가중치 1.0을 적용한 경우를 의미하며, 64/2는 시간 데이터를 64차원으로 확장한 뒤 가중치 2.0을 적용한 경우를 의미한다.

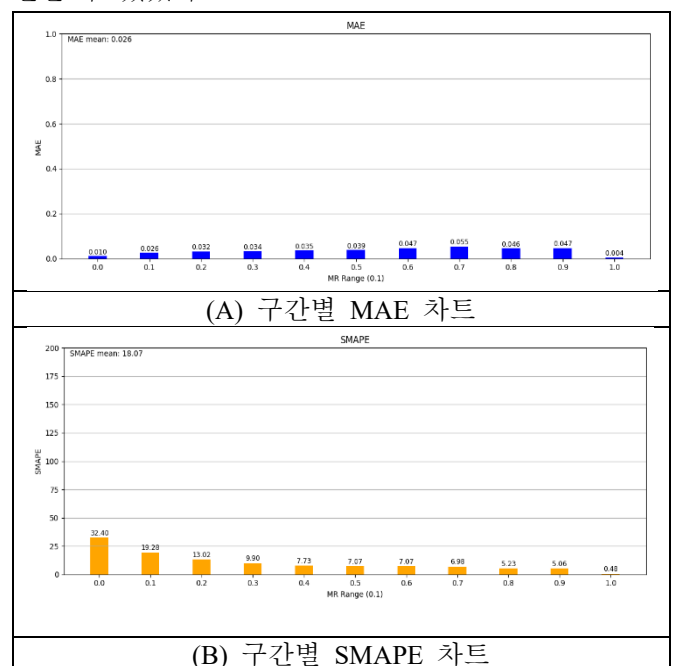
실험 결과 단일 이미지 기반의 CNN 모델보다 멀티모달 모델이 우수한 것을 확인하였고 멀티모달 사이

에서 기존의 성능 지표들은 크게 차이가 나지 않았으나 SMAPE 값을 비교하였을 때 64차원에 가중치 2.0 설정이 가장 낮은 SMAPE 값을 보이며 가장 좋은 성능을 보이는 것을 확인하였고 128차원에 가중치 2.0은 SMAPE가 약간 높게 나오나 MAE와 RMSE가 더 낮게 나옴을 확인하였다.

<표 1> 모델별 실험 결과

모델	MSE	MAE	RMSE	R^2	SMAPE
CNN	0.00	0.03	0.04	0.98	23.36
Multi-Modal (1/1)	0.00	0.03	0.05	0.98	23.36
Multi-Modal (1/2)	0.00	0.03	0.04	0.98	23.01
Multi-Modal (64/1)	0.00	0.02	0.03	0.99	18.79
Multi-Modal (64/2)	0.00	0.03	0.04	0.99	18.07
Multi-Modal (128/1)	0.00	0.03	0.04	0.99	22.05
Multi-Modal (128/2)	0.00	0.02	0.03	0.99	18.16
Multi-Modal (256/1)	0.00	0.02	0.03	0.99	21.83
Multi-Modal (256/2)	0.00	0.02	0.03	0.99	29.39

또한, 건조도가 낮아짐에 따라 정확도의 변화를 분석하기 위해 건조도 구간별 성능 분석을 추가적으로 진행하였다. 이러한 구간별 분석은 모델이 건조 초기, 중기, 후기 등 서로 다른 단계에서 어떠한 특성을 보이는지 확인할 수 있도록 한다. 결과는 그림 3과 같으며 MAE는 전반적으로 유사한 결과가 나왔으나 SMAPE는 값이 낮아질수록 오차가 높아지는 것을 확인할 수 있었다.



(그림 3) 구간별 평가지표 차트

4. 결론

본 논문에서는 이미지 특징과 건조 시간을 합친 멀티모달 모델을 사용하여 고추의 건조도를 예측하였다. 제안한 모델은 이미지만 사용하였을 때 보다 시간 차원 1 가중치 1 과 시간 차원 256 가중치 2 를 제외한 나머지 멀티모달이 평균 3.07% 더 낮은 SMAPE 값이 나와 더 좋은 결과가 나오는 것을 확인하였다. 또한, 추가적으로 시간 차원 확장 및 가중치를 조절하며 확인한 결과 64 차원 가중치를 2 배로 주었을 때 18.07 으로 가장 낮은 SMAPE 가 나오며 가장 정확하게 건조도를 예측하였다. 다만, 구간별로 분석을 진행한 결과 MAE 는 정확하게 나왔으나 SMAPE 는 건조도가 낮아질수록 오차가 커지는 경향이 보여 추가적인 학습 방법이 필요할 것으로 판단된다. 향후 연구에서는 MSE 만을 사용하는 것이 아닌 MSE 와 SMAPE 를 복합적으로 사용해 학습하여 건조도가 낮은 구간에서도 낮은 SMAPE 가 나오도록 하여 모든 구간에서 정확하게 건조도를 예측하도록 할 계획이다.

Acknowledgement

본 논문은 안동 중소기업 ICT 융합기술경쟁력강화사업[25AD1100]의 연구결과로 수행되었음. 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [25ZD1110, 대경권 지역산업 기반 ICT 융합 기술 고도화 지원사업]

참고문헌

- [1] Krzykowski A, Rudy S, Polak R, Biernacka B, Krajewska A, Janiszewska-Turak E, Kowalska I, Żuchowski J, Skalski B, Dziki D. "Drying of Red Chili Pepper (*Capsicum annuum* L.): Process Kinetics, Color Changes, Carotenoid Content and Phenolic Profile". *Molecules*. vol. 29, no. 21, pp. 5164, 2024.
- [2] Miao Liu, Liu Hu, Na Deng, Yongjian Cai, Hui Li, Bo Zhang, and Jianhui Wang. "Effects of different hot-air drying methods on the dynamic changes in color, nutrient and aroma quality of three chili pepper (*Capsicum annuum* L.) varieties." *Food chemistry: X*, vol. 22, 101262. 29, 2024.
- [3] Anjaneyulu and Sharangi, "Study the Drying Behavior of Red Pepper (*Capsicum annuum* L.)." *International Journal of Bio-resource and Stress Management*, vol. 13, no. 8, pp.822-829, 2022
- [4] G. Romano, D. Argyropoulos, M. Nagle, M.T. Khan, J. Müller, "Combination of digital images and laser light to predict moisture content and color of bell pepper simultaneously during drying." *Journal of Food Engineering*, vol. 109, Issue 3, pp.438-448, 2012
- [5] Ganesh C Bora, Rohit Pathak, Mojtaba Ahmadi, Purbasha Mistry, "Image processing analysis to track colour changes on apple and correlate to moisture content in drying stages." *Food Quality and Safety*, vol. 2, Issue 2, pp.105-110, 2018
- [6] Shichen Li, Chenhui Shao, "Multi-modal data fusion for moisture content prediction in apple drying." *Manufacturing Letters*, vol. 44, pp. 1316-1325, 2025
- [7] Niu, Kunlong, Chongyang Wang, Jianhui Xu, Chuanxun Yang, Xia Zhou, and Xiankun Yang. "An Improved YOLOv5s-Seg Detection and Segmentation Model for the Accurate Identification of Forest Fires Based on UAV Infrared Image." *Remote Sensing*, vol. 15, no. 19, pp. 4694, 2023
- [8] Shichen Li, Chenhui Shao, "Multi-Modal Fusion of In-Situ Video Data and Process Parameters for Online Forecasting of Cookie Drying Readiness", arXiv:2504.15599, 2025.
- [9] Patro, S. Gopal Krishna, Kishore Kumar Sahu. "Normalization: A Preprocessing Stage." arXiv:1503.06462, 2015.
- [10] Russakovsky, O., Deng, J., Su, H. et al. "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015.
- [11] Mingxing Tan, Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv:1905.11946, 2019.
- [12] Chen C, Twycross J, Garibaldi JM. "A new accuracy measure based on bounded relative error for time series forecasting." *PLoS One*, vol. 12, no. 3, 2017.