

# 초분광 이미지와 Logistic Regression 모델 기반 마늘 표면 손상 검출 모델 개발

왕평안<sup>1</sup>, 한웅철<sup>1,2\*</sup>

<sup>1</sup>강원대학교 스마트농업융합학과, <sup>2</sup>강원대학교 바이오시스템기계공학과

## Development of Garlic Superficial Injury Detection Model based on Hyperspectral Images and Logistic Regression Model

Pingan Wang<sup>1</sup>, Xiongze Han<sup>1,2\*</sup>

<sup>1</sup>Interdisciplinary Program in Smart Agriculture, Kangwon National University

<sup>2</sup>Department of Biosystems Engineering, Kangwon National University

**요약** 마늘은 강력한 항바이러스성과 면역력을 높이는 특성으로 인해 '천연 항생제'라는 별명을 얻었다. 그러나 썩은 마늘은 메스꺼움이나 식중독을 유발하고, 심지어 사람들의 생명을 위협할 수도 있다. 본 논문은 초분광 카메라 기술을 기반으로 마늘 표면 손상 모델과 기계 학습 분류 모델을 개발하였다. 표면손상 유무 검출을 위해 전처리한 마늘식물은 총 104종으로 이 중 80종의 마늘은 모델 개발 용도, 24종의 마늘은 시험 검증용도로 사용됐다. 우선 ENVI에서 상이한 손상 조건에서의 결과 분석을 위해 식생 지수를 사용하였다. 그런 다음 579개 픽셀에 logistic regression 모델 교육을 진행하였다. 마지막으로, 모델의 실제 검증을 위해 24개의 마늘에 위 모델을 적용하여 검증하였다. 그 결과 NDVI 기술은 마늘 표면의 곰팡이 부분을 명확하게 식별할 수 있었으나, 작은 면적의 손상에 대해서는 파악하지 못하였다. Logistic regression 방법으로 설정된 분류 모형의 정확도는 90%로 나타났으며, 마늘 표면의 경미 손상 구역도 식별할 수 있었다. 이 마늘 손상 식별 모델의 개발은 마늘 손상 부위 검출에 초분광 기술을 확장시켜 적용하였고 이로 인하여 손상된 마늘을 선별하는 비용이 절감되었다.

**Abstract** The potent antiviral and immune-boosting properties of garlic have earned it the moniker "natural antibiotic". On the other hand, rotten garlic can result in nausea, food poisoning, and possibly endanger people's lives. This paper aims to develop a garlic superficial injury model based on hyperspectral camera technology and a machine-learning classification model. One hundred and four garlic plants were pretreated for surface damage, of which 80 and 24 were used for model development and test validation, respectively. First, the Normalized Difference Vegetation Index (NDVI) method was used in ENVI to analyze the results under different damage conditions. Subsequently, 579 pixels were selected for Logistic Regression Model training. Finally, 24 garlic bulbs were used for actual verification of the model. The NDVI technology could identify the mildew part of the garlic surface but could not identify the small area. The accuracy of the classification model established by the logistic regression method can reach 90%. The initial small lesions on the surface of garlic can also be identified. The development of this garlic damage identification model extends the application of the hyperspectral technique in garlic detection. Similarly, the cost of garlic damage detection was reduced.

**Keywords** : Garlic Superficial Injury, Hyperspectral Camera, Logistic Regression, Machine-learning, NDVI

본 연구는 한국연구재단 기본연구지원사업 (NRF-2021R1F1A1055992)의 지원에 의해 이루어진 것임.

\*Corresponding Author : Xiongze Han(Kangwon National University)

email: hanxiongze@kangwon.ac.kr

Received May 8, 2023

Revised May 31, 2023

Accepted July 7, 2023

Published July 31, 2023

## 1. 서론

마늘은 우리 생활에서 흔히 볼 수 있는 조미료로서 식용가치와 약용가치가 높아 많은 선호를 받고 있다. 그러나 상한 마늘을 식용하게 되면 구토, 설사 등의 부작용을 일으켜 생명을 위협할 수 있다. 마늘의 변질, 파손 여부를 사람이 직접 관찰하여 판단하는 경우 오차가 생길 수 있다. 그러므로 마늘의 품질을 결정하기 위해 새로운 마늘 판별 방법의 개발이 필요하다.

초분광 이미지는 RGB 및 다분광 이미지보다 더 많은 파장 정보가 포함되어 있기 때문에 다양한 산업에서 널리 사용하고 있으며, 농업 측면에서는 식물 특성 및 안전성 등의 연구에도 활용되고 있다. 초분광 이미지는 과일의 기계적 손상을 감지할 뿐만 아니라 작물 질병 및 해충의 조기 분석에 사용할 수 있으며 심지어 계란 품질을 분류할 수 있는 것으로 보고 되었다[1-3].

머신러닝을 이용한 손상된 마늘 선별연구는 주로 RGB 카메라로 수행되고 있으나[4], 초분광 카메라 기술은 주로 마늘의 알리신 및 가용성 고형분 (SSC: Soluble Solid Content, 이하 SSC)과 마늘의 공기 건조 후 불균 일성을 감지하는 데 사용되고 있다[5,6]. RGB 카메라와 초분광 카메라를 동시에 사용하여 마늘의 알리신과 손상을 감지하면 비용이 증가하게 된다. 또한, 초분광 카메라도 식물 손상 감지 능력이 있으므로 본 연구에서는 초분광 이미지 기반으로 마늘 손상을 식별하여 비용을 절감할 수 있을 뿐만 아니라 단일 초분광 이미지에서 더 많은 효과적인 정보를 얻을 수 있다.

본 연구에서는 104개의 마늘을 연구 대상으로 선정하였으며, 그 중 80개는 모델 개발을 위해 훈련 데이터에 사용하고, 24개는 개발한 모델의 정확성을 검증하기 위해 사용하였다. 우선 야외에서 모든 마늘에 대한 RGB 이미지와 초분광 이미지를 수집한 후 80개의 마늘 이미지를 영상처리 프로그램 (ENVI: The Environment for Visualizing Images, 이하 ENVI) 에 도입하여 관심 영역 (ROI: Region of Interest, 이하 ROI) 영역을 선택하고 손상 및 정상 영역 마늘의 스펙트럼을 분석 후 식생지수 (NDVI: Normalized difference vegetation index, 이하 NDVI) 결과를 획득하였다. 그리고 정상 및 손상 영역에 대해 총 1000픽셀을 데이터 세트로 구성하고 logistic regression 방법을 사용하여 마늘 분류 모델을 개발하였다. 마지막으로 24개의 마늘에 대한 초분광 이미지를 사용하여 모델을 식별하고 RGB 이미지와 비교하여 개발된 모델을 검증하였다.

## 2. 재료 및 방법

### 2.1 실험 하드웨어 및 데이터 사전 처리

실험에 사용된 마늘은 현지 슈퍼마켓 (춘천·한국)에서 구입하였으며 본 연구를 위해 총 104개 마늘을 준비하였다. 마늘의 다양한 손상 정도를 취득하기 위해 칼을 이용하여 마늘 표면에서 다양한 크기인 손상 면적으로 자르고, 약 5일 동안 25°실내 환경에 두어 다양한 정도로 손상된 마늘 Fig. 1에서 보는 바와 같이 샘플을 얻었다. 104개 마늘을 두 그룹으로 나누어 Fig. 1a에서 보는 바와 같이 80개 마늘은 분류 모델 개발에 사용하였고 나머지 Fig. 1b에서 보는 바와 같이 24개 마늘은 모델의 정확성 검증에 사용하였다.

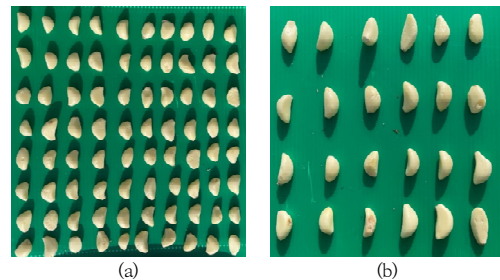


Fig. 1. The garlic samples used in experiment: (a) experimental group and (b) verification group.

마늘 이미지 데이터를 수집하기 위해 초분광 카메라 Specim FX10 (Specim, Spectral Imaging Ltd, Oulu, Finland)를 사용하였으며, 본 카메라는 가시광선 및 근적외선 (VNIR: Visible and Near-Infrared 이하 VNIR) 대역 (400~1000 nm, band 1~56) Table 1과 같이 스펙트럼 데이터를 수집할 수 있다. 초분광 카메라에서 취득한 이미지의 반사율 보정에 사용되는 반사판은 Lastolite LLL LR2050 20-Inch Ezybalance Card 18% (Manfrotto, Ontario, Italy)이다. 초분광 데이터 수집은 데스크톱 소프트웨어 Lumo scanner (Specim, Spectral Imaging Ltd, Oulu, Finland)를 사용하여 스캐너 스테이지 및 샘플 위치를 설정하고 제어하여 연구를 수행하였다.

이미지 데이터 수집을 위해 샘플과 초분광 카메라간의 거리는 약 1.2 m로 설정하였으며, 마늘은 Fig. 2와 같은 플레이트 위에 놓고 초분광 카메라는 삼각 지지대에 부착하였다. 삼각 지지대에 부착되어 있는 전동식 회전 모터는 초분광 카메라를 180° 회전시켜 이미지 스캔할 수

Table 1. Specifications of the SPECIM FX10 hyperspectral camera.

Item	Ventilating fan
Spectral Range	400~1,000 nm
Spectral Bands	224
Spectral FWHM (Full width at half maximum)	5.5 nm
Spatial Sampling	1,024 px
Frame Rate	330 FPS full frame 9,900 FPS with 1 band selected
FOV (Field of view)	38°
F-number	F/1.7
Camera SNR (Signal-to-noise ratio)	600 : 1
Camera Interface	GigE Vision or Camera Link
Dimensions	150 mm x 85 mm x 71 mm
Weight	1.26kg

있다. 초분광 카메라의 데이터 인터페이스는 노트북에 연결되며 Lumo scanner를 사용하여 맑은 날 야외에서 자체 노이즈를 줄이기 위한 Dark current image를 측정 한 후 binning 값을 기본 2배로 설치하여 데이터를 기록했다. 원시 초분광 데이터를 얻은 후 후속 계산 및 모델 훈련을 용이하게 하기 위해 반사판을 사용하여 원시 데이터를 반사율 보정하고 모든 픽셀 값을 0과 1 사이에 분포하도록 정규화 작업을 실시하였다.

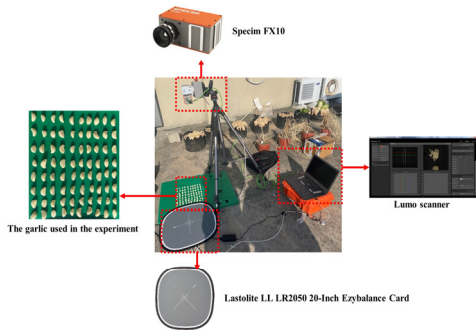


Fig. 2. The composition of the outdoor hyperspectral image collection system.

### 2.2 식생지수

정규화 식생지수 NDVI는 작물의 적색 대역과 근적외선 대역의 반사도 차이와 다중 분광 영상의 적색 대역과 근적외선 대역을 활용하여 식생의 활력도를 나타내는 지수이다[7]. 관측된 다중 분광 위성영상을 활용하면 NDVI를 통해 실제 작물의 상태를 판단할 수 있어서 다양한 분야의 연구에 활용되고 있다[8].

NDVI 계산식은 Eq. (1)과 같다. 여기서 NIR은 근적

외선 반사율, Red는 적외선 반사율이다. 파장대역 39 (810 nm)의 데이터를 근적외선 데이터로 사용하고 파장대역 24 (650 nm)의 데이터를 적색광 데이터로 사용했다. NDVI 이미지는 ENVI에서 계산하고 결과 이미지를 도출하였다.

$$NDVI = (NIR - Red) / (NIR + Red) \quad (1)$$

Where, NIR denotes reflectance of the near-infrared area, Red denotes reflectance of the infrared area.

### 2.3 분류 모델의 개발 및 검증

모델 개발 전에 ENVI에서 마늘의 손상 영역과 정상 영역은 ROI 기능을 사용하여 구분하였으며, Fig. 3과 같이 손상 영역과 정상 영역의 각 파장 대역에서의 반사 평균값을 구했다. 400 nm에서 1000 nm까지의 손상 영역과 정상 영역의 전체 대역에서 반사 값 사이에 차이가 있는 것을 기준으로 전체 56개 파장 대역의 데이터를 후속 모델 훈련의 기본 데이터로 설정하였다.

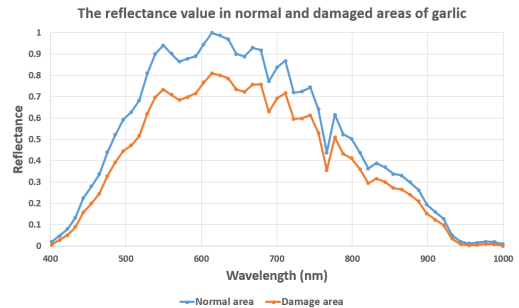


Fig. 3. The reflectance value in normal and damaged areas of garlic.

Logistic regression 모델은 계산 량이 매우 적은 뿐만 아니라 속도가 빠르고 다양한 2차 분류 문제에 널리 사용되고 있다. 본 연구에서 분류 모델 개발을 위해 logistic regression 모델을 선택하여 모델을 취득한 후 검증하였다. 훈련 데이터에서 80개 마늘에 손상과 정상 영역의 비율은 큰 차이가 있었기에 격차를 줄이기 위해 손상 ROI 영역 (300 px)과 정상 ROI 영역 (279 px)의 총 579개 픽셀 위치에서 56개 파장대역의 반사 값을 선택하였고 결과에서 0으로 정상 영역을 나타내고 1로 손상 영역을 표시하여 도출되었다. 데이터의 80%는 모델의 훈련 세트로 사용되며 20%는 모델의 정확성 검증 세트 사용되었다.

모델의 분류 능력을 검증하기 위해 검증조의 24개 마늘의 전체 파장 대역의 반사 값을 사용하여 개발된 logistic

regression 모델에서 검출하였으며, 분류 후의 손상과 정상 픽셀 결과는 ENVI에서 ROI기능을 사용하여 원 사진에서 표시 후 RGB 카메라에서 취득한 사진의 결과와 비교하였다.

### 3. 결과 및 고찰

Fig. 4와 같이 NDVI 사진 결과에서 마늘의 정상 영역과 손상 영역에 대해서 부분적으로 육안으로 관찰이 가능하다. 흑색 점선 영역은 마늘이 심하게 손상되어 변질된 영역이며 정상 영역과 선명한 차이가 있다는 것을 알 수 있지만, 흰색 점선 영역은 경미한 손상으로 인해 정상 영역과 미세한 차이만 있음을 알 수 있다. Fig. 1a의 RGB 사진과 비교하면 NDVI 결과가 마늘 손상 후 변질될 때 더 나은 결과를 얻을 수 있지만 경미한 손상의 결과는 정확하게 식별하기 어려웠다.

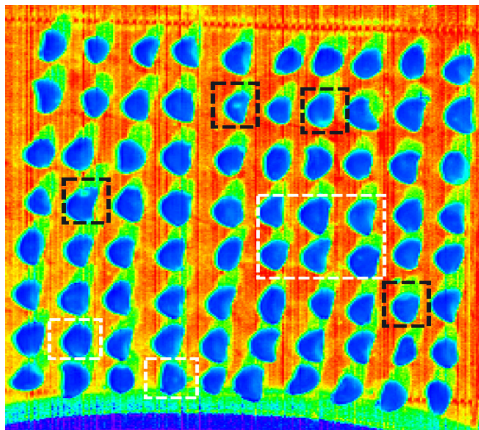


Fig. 4. Results of the experimental group obtained using the NDVI method (the black dotted line frame represents the damaged garlic samples successfully identified, and the white dotted line frame represents the damaged garlic samples not successfully identified).

모델의 정확도 결과는 Table 2와 같다. 정상 영역을 0, 손상 영역을 1로 표현하였으며 모델의 전체 인식률이 90%에 달한다는 것을 알 수 있다. 정상 영역의 예측 성공률과 리콜은 97%, 76%로 나타났고 파손 영역의 예측 성공률과 리콜은 86%, 98%에 달했으며, F1 점수도 각각 85%와 92%로 높게 나타났다. 일반적으로 모델의 예측 정확도 효과는 Fig. 5와 같이 수신자 조작 특성 (ROC:

Receiver Operating Characteristic, 이하 ROC) 곡선으로 판단할 수 있으며, ROC 곡선이 왼쪽 상단에 가까울수록 모델의 정확도가 좋다. ROC 곡선 아래의 구역은 곡선 아래 면적 (AUC: Area Under the Curve, 이하 AUC) 가 되며, 마찬가지로 AUC 영역의 면적이 클수록 모델의 정확도가 높다. 본 연구에서 개발한 모델의 ROC 곡선에서 AUC 결과가 92%에 달한다는 것을 알 수 있어, 본 연구에서 개발한 모델이 마늘의 정상 영역과 손상된 영역을 정확하게 식별할 수 있음을 보여준다.

Table 2. Precision accuracy, recall, and F1-score results of regression model.

Item	Precision	Recall	F1-score
0	0.97	0.76	0.85
1	0.86	0.98	0.92
Accuracy	-	-	0.90
Macro avg	0.92	0.86	0.89
Weighted avg	0.91	0.91	0.90

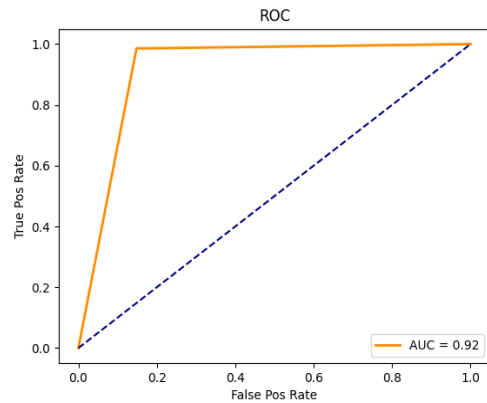


Fig. 5. ROC results of the regression model.

Fig. 6a에서 빨간색 영역은 개발한 모델 기반으로 식별한 파손 결과이고 Fig. 6b는 마늘의 RGB 데이터이다. NDVI 방법의 인식 결과와 비교하여 logistic regression 방법은 마늘 손상 후 변질된 영역을 정확하게 식별할 수 있을 뿐만 아니라 마늘의 경미 손상도 정확하게 식별할 수 있다는 것을 알 수 있다. 그러나 본 연구에서 개발한 모델의 단점은 Fig. 6b에 빨간 점선 영역과 같이 마늘의 뿌리도 손상 영역으로 인식하므로, 앞으로 마늘 뿌리와 마늘 껍질을 식별할 수 있는 정확한 분류 모델을 개발하기 위해 더 다양하고 많은 데이터를 수집이 필요하다.

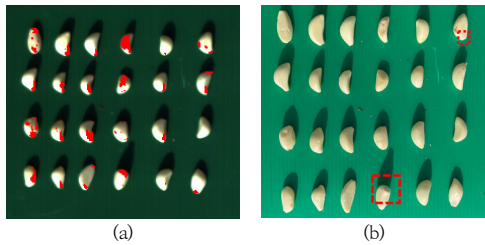


Fig. 6. Identification results of 24 garlic in experimental group: (a) model classification result and (b) RGB camera result.

위에서 언급한 NDVI 방법과 logistic regression 방법의 실험 결과 기반으로 아래와 같이 두 방법에 대한 장단점을 정리할 수 있다. 전통적인 NDVI 방법은 NIR과 Red 두 개의 파장 영역 데이터만을 사용하여 계산되지만, 본 논문에서 제시한 logistic regression 기반의 마늘 표면 손상 검출 방법은 전체 파장 데이터를 분석하므로 데이터 처리량 측면에서 부담이 크다는 것이 단점으로 볼 수 있다. 그러나 전통적인 NDVI 방법은 임계값에 따라 마늘 표면 손상 정도를 결정되며, 여기서 임계값의 판단은 샘플 수와 통계 분석 방법에 따라 주관성이 있을 수 있다. 하지만 본 연구에서 제안한 logistic regression 방법은 0 또는 1 값을 출력하여 객관적인 분류 결과를 도출할 수 있어 더욱 신뢰성이 있는 결과로 볼 수 있다.

#### 4. 결론

본 연구에서는 초분광 이미지 처리 기술과 logistic regression 모델을 기반으로 마늘 표면 손상 검출 모델을 개발하였다. 실외에서 마늘에 대한 초분광 사진을 수집하였고 ENVI에서 원시 사진에 대해서 전처리하고 반사 보정 및 정규화 작업을 통해 NDVI 이미지 결과를 취득하였다. 손상 및 정상 영역의 평균 반사 값을 분석하고 총 579개의 손상 및 정상 영역의 픽셀 데이터를 logistic regression 모델로 도입하여 모델 훈련 후 검증군의 마늘을 사용하여 모델의 정확성에 대해서 검증하였다. 실험 결과는 NDVI 방법과 비교하여 logistic regression 모델의 전체적 정확도가 90%에 도달할 수 있었으며 본 모델은 마늘에 경미한 손상 영역의 식별에 좋은 효과가 나타났다.

향후 연구 방향에서 현장의 자동화가 가능한 분류 기술을 개발하기 위해 다중 조명 조건에서 실시간적인 응용하는 것이 필요하다. 본 연구에서 개발된 마늘 표면 손

상 검출 모델은 실외의 충분한 조명 조건에서 데이터를 수집하여 방사 교정을 통해 이루어졌으나 현장에서 실시간으로 적용 시 고정밀 분류 가능성을 보장하기 위해 다중 조명 조건에서의 연구 데이터를 수집하여 모델 개발 및 분석이 필요하다.

#### References

- [1] N. V. Rivera, J. Gomez-Sanchis, J. Chanona-Perez, J. J. Carrasco, M. Millan-Giraldo et al, "Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning", *Biosystems Engineering*, Vol.122, pp.91-98, Jun. 2014.  
DOI: <https://doi.org/10.1016/j.biosystemseng.2014.03.009>
- [2] S. Srivastava, H. N. Mishra, "Detection of insect damaged rice grains using visible and near infrared hyperspectral imaging technique", *Chemometrics and Intelligent Laboratory Systems*, Vol.221, No.15, pp.104-489, Feb. 2022.  
DOI: <https://doi.org/10.1016/j.chemolab.2021.104489>
- [3] K. S. Yao, J. Sun, C. Chen, M. Xu, X. Zhou, "Non-destructive detection of egg qualities based on hyperspectral imaging", *Journal of Food Engineering*, Vol.325, pp.111-024, Jul. 2022.  
DOI: <https://doi.org/10.1016/j.jfoodeng.2022.111024>
- [4] A. Wang, J. Zhao, "Moldy garlic recognition based on machine vision.", *3rd International Academic Exchange Conference on Science and Technology Innovation*, IEEE, GZ, CHN, pp.772-776, December. 2021.  
DOI: <https://doi.org/10.1109/IAECST54258.2021.9695700>
- [5] A. Rahman, M. A. Faqeerzada, B. K. Cho, "Hyperspectral imaging for predicting the allicin and soluble solid content of garlic with variable selection algorithms and chemometric models", *Journal of the Science of Food and Agriculture*, Vol.98, No.12, pp.4715-4725, Sep. 2018.  
DOI: <https://doi.org/10.1002/jsfa.9006>
- [6] H. H. Cho, S. M. Lee, S. W. Park, R. K. Cho, "Observation for drying non-uniformity of allium vegetables using NIR spectroscopy", *The Korean Society of Food Preservation*, Vol.24, No.1, pp.145-152, Feb. 2017.  
DOI: <https://doi.org/10.11002/kjfp.2014.21.6.838>
- [7] C. J. Tucker, C. O. Justice, S. D. Prince, "Monitoring the grasslands of the Sahel 1984-1985", *International Journal of Remote Sensing*, Vol.7, No.11, pp.1571-1581, 1986.  
DOI: <https://doi.org/10.1080/01431168608948954>
- [8] J. B. Lim, S. E. Cha, M. S. Won, J. H. Kim, Y. R. Ryu et al, "Design of Calibration and Validation Area for Forestry Vegetation Index from CAS500-4", *Korean Journal of Remote Sensing*, Vol.38, No.3, pp.311-326, Jun. 2022.  
DOI: <https://doi.org/10.7780/kirs.2022.38.3.7>

왕 평 안(Pingan Wang)

[정회원]



- 2022년 7월 : 강원대학교 스마트 농업융합학과 (공학석사)
- 2006년 3월 ~ 현재 : 강원대학교 스마트농업융합학과 (박사과정)

<관심분야>

농업로봇, 농업자동화

---

한 응 철(Xiongze Han)

[정회원]



- 2013년 8월 : 서울대학교 바이오 시스템공학과 (공학석사)
- 2017년 2월 : 서울대학교 바이오 시스템공학과 (공학박사)
- 2020년 3월 ~ 현재 : 강원대학교 바이오시스템기계공학과 조교수

<관심분야>

정밀농업, 농업로봇, 원격탐사