KSMER Journal of The Korean Society of Mineral and Energy Resources Engineers

Vol. 61, No. 1 (2024) pp. 23-32, https://doi.org/10.32390/ksmer.2024.61.1.023

ISSN 2288-0291(print)
ISSN 2288-2790(online)

연구논문

드론 촬영 이미지와 딥러닝 기법을 이용한 태양광 패널의 먼지오염 탐지

한호상[1] · 서장원[2]*

[1]강원대학교 에너지자원융합공학과, [2]강원대학교 에너지자원화학공학과

Detection of Photovoltaic Panel Dust Using Drone Shooting Imagery and Deep Learning Techniques

Hosang Han[1] and Jangwon Suh[2]*

[1]Department of Energy and Mineral Resources Engineering, Samcheok, Korea
[2]Department of Energy Resources and Chemical Engineering, Samcheok, Korea

***Corresponding Author.** Jangwon Suh, jangwonsuh@kangwon.ac.kr, Department of Energy Resources and Chemical Engineering, Kangwon National University, Samcheok, Korea

Received
4 January 2024

Final version Received
6 February 2024

Accepted
27 February 2024

Abstract

Regular detection of dust on photovoltaic (PV) modules is crucial for their maintenance because dust can hinder light harvesting and decrease power generation. The current study utilized drone images of panels and deep learning techniques to detect dust and evaluate the performance of a prediction model. After acquiring and preprocessing the drone images of the panels in Samcheok city, Gangwon state, we applied five deep learning techniques to identify and categorize the panels as dusty or clean. We evaluated the performance of these techniques using a confusion matrix. The convolutional neural network technique demonstrated the highest predictive performance with an area under the curve of 80.6%, classification accuracy of 66.0%, and recall of 94.0%. However, the model's high predictive performance for dusty panel cases was accompanied by a high rate of misclassification for the clean panel cases, indicating the need for improvement. The proposed analytical model can be used to quickly screen and evaluate the dusty panels in large-scale power plants.

Key words : photovoltaic panel, dust, drone, deep learning, samcheok

요 약

태양광 모듈 표면의 먼지오염은 수광장애 요인으로 발전량 저하를 유발하기 때문에 이를 주기적으로 탐지하는 것은 유지관리 측면에서 매우 중요하다. 따라서 본 연구에서는 드론으로 촬영한 이미지와 딥러닝 기법을 이용하여 태양광 패널 규모에서의 먼지오염 여부에 대한 분류 예측 모델의 성능을 평가하였다. 강원특별자치도 삼척시에 위치한 육상 및 옥상 태양광 패널에 대한 드론 이미지를 취득 및 전처리 후 5종의 딥러닝 기법을 적용하여 패널위의 먼지오염 여부를 분류하였으며 혼동행렬을 이용하여 예측결과의 분류 성능을 평가하였다. 그 결과 convolutional neural network 기법이 가장 높은 예측력을 보였으며 Area under the curve는 80.6%, 정확도는 66.0%, 재현율은 94.0%로 계산되었다. 실제로 패널위에 먼지오염이 있는 경우에 높은 예측력을 보인 반면 먼지오염이 없는 경우를 오분류한 비율도 높았기에 이에 대한 개선이 필요할 것으로 판단된다. 본 연구에서 제안한 분석모델은 대규모 태양광 발전소에서 패널위의 먼지오염을 신속하게 평가하고 스크리닝하는 방법론의 하나로 활용될 수 있을 것으로 기대한다.

주요어 : 태양광 패널, 먼지, 드론, 딥러닝, 삼척

Copyright ⓒ 2024, The Korean Society of Mineral and Energy Resources Engineers

서 론

태양광(photovoltaic, PV) 모듈 표면의 그림자(shading), 먼지(dust), 쏘일링(soiling), 낙엽(leaves), 눈(snow), 새의 분비물(bird dropping) 등은 대표적인 수광장애 요인이다(Saini *et al*., 2023). 이와 같은 수광장애 요인들은 태양광 발전량의 저하를 유발하며, 태양광 모듈의 태양 전지(cell)가 직렬로 연결되어 있기 때문에 수광장애 면적 비율에 비해 실제 발전 효율의 저하는 훨씬 크다. 전술한 수광장애 요인 중에는 일반적으로 그림자가 가장 큰 문제로 여겨지고 있지만 이는 태양광 모듈의 설계 및 설치 과정에서 음영지역을 회피하는 방법으로 해결이 가능하다. 반면에 Fig. 1에 나타난 미세먼지나 황사 등과 같은 먼지(오염)는 태양광 발전소가 회피하기 어려운 수광장애 요인에 해당한다. IEA-PVPS(International Energy Agency's Photovoltaic Power Systems Program) 보고서(Schill *et al*., 2022)에 따르면, 태양광 모듈 표면에 쌓인 입자 등 오염물로 인해 발생한 에너지 손실은 전세계 태양광 발전량의 최대 4%로 추정(2018년 기준)되며, 이는 약 30~50억 유로(euro)의 수익 손실에 해당하는 것으로 발표되었다. 국내의 경우 미세먼지나 황사가 태양광 발전에 미치는 영향 수준(정도)에 대해서는 다양한 의견이 존재하지만 악영향을 끼친다는 것은 다양한 문헌의 연구결과(Kaldellis *et al*., 2011; Lee *et al*., 2017; Son and Jeong, 2019; Sung and Cho, 2019)로부터 입증되었다. 따라서 이와 같은 에너지 손실량을 줄이기 위해서는 태양광 모듈 표면의 수광장애 여부를 정기적으로 탐지 및 점검하고, 적절한 세척 등의 유지관리 작업이 이루어져야 할 필요가 있다.

태양광 모듈과 먼지오염에 대한 연구는 국내·외의 다양한 연구자들에 의해 이루어져왔다. 첫 번째 주제는 먼지오염이 태양광 발전량(효율)에 미치는 영향에 대한 것이다. 국내에서 Han *et al*.(2018)은 공기 중 미세먼지에 의한 태양전지의 오염 정도에 대해 분석하였고, Park *et al*.(2019)은 기상기후 자료와 발전량 계측 자료를 이용하여 미세먼지를 포함한 기상특성과 태양광 발전량 간의 회귀모델을 도출하였다. 미세먼지 자료와 인공지능(머신러닝 및 딥러닝 기법)을 이용하여 태양광 발전량 또는 효율 저감량을 예측한 문헌도 발표되었다(You and Bae, 2020; Kim *et al*., 2021). 해외에서도 인도(Rao *et al*., 2014), 이집트(Menoufi *et al*., 2017), 포르투갈(Conceição, *et al*., 2018), 중국(Chen *et al*., 2020) 등 다양한 국가에서 유사한 연구가 다수 보고되었다. 두 번째 주제는 태양광 모듈의 먼지오염 탐지에 관한 것이다. 국내에서는 Kim(2015)가 레이저 광선을 이용한 광학영상 센서로부터 태양광 모듈의 표면 오염을 검출한 바 있다. 해외의 경우 딥러닝 기법을 이용한 태양광 모듈의 먼지 및 핫스팟 분류 연구(Cipriani *et al*., 2020), 태양광 모듈의 먼지 탐지를 위한 시맨틱 영상 분할(image semantic segmentation) 기법 비교(Cruz-Rojas *et al*., 2023), 실험연구 스케일에서 태양광 모듈의 먼지오염을 탐지하기 위한 무인항공기와 딥러닝 기법 적용(Gao and Li, 2023) 연구 등이 수행된 바 있다. 이러한 연구들은 드론 영상의 전처리를 통해 태양광 모듈의 경계를 추출하고, 개별 모듈 표면의 먼지오염 영역을 탐지하였다. 반면에 다수의 태양광 모듈로 구성된 태양광 패널 규모 단위의 먼지오염 탐지를 목적으로 드론 촬영 영상과 딥러닝 기법으로부터 먼지오염의 존재 여부를 분석한 연구는 다소 미흡한 실정이다.

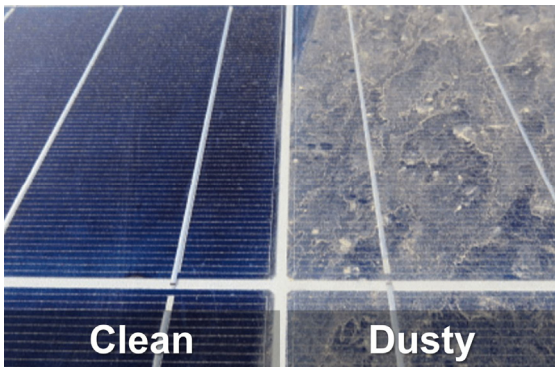

Fig. 1. Exemplary images of clean and dusty PV modules.

본 연구의 목적은 무인항공기(이하 드론)로 촬영 이미지와 딥러닝 기법을 이용하여 태양광 패널 규모에서의 먼지오염(쏘일링 포함) 여부를 분류 예측하는 것이다. 이 연구는 특정 태양광 모듈에 대한 먼지오염 영역을 탐지하는 것이 아니라 세척작업을 위해 태양광 발전소의 패널 규모 수준에서의 먼지오염 여부를 분류하고자 하는 것이다.

연구방법

강원특별자치도 삼척시에 설치되어 있는 육상 및 옥상 태양광 발전소(발전시스템) 일부를 연구지역으로 선정하였다. 2023년 10월 기준으로 강원특별자치도 삼척시에는 401개소의 태양광 발전소가 설치되어 있으며, 태양광 발전 설비 전체 용량은 113 MW(강원특별자치도 전체 용량 대비 6.8%)에 이른다. 그러나 태양광 발전 설비 이용률 관점에서 강원특별자치도의 18개 시군구 중 삼척시의 기상조건 순위가 최하위(18위)에 해당하기 때문에(Renewable Energy Cloud Platform, 2023) 본 연구에서는 삼척시를 연구지역으로 선정하였다.

본 연구에서 태양광 패널의 먼지오염 탐지를 위한 연구방법과 절차를 Fig. 2에 순서도 형태로 제시하였다. 이는 드

론을 이용한 영상자료 취득, 취득된 자료의 딥러닝 적용을 위한 훈련 자료 및 테스트 자료로의 분류, 영상자료의 딥러닝 분석용 정보 추출을 위한 이미지 임베딩, 딥러닝 기법의 하이퍼파라미터 조절 및 적용, 혼동행렬을 이용한 딥러닝 기법의 분류 성능 지표 비교 분석 과정으로 구성된다.

본 연구에서는 파이썬(Python) 기반의 오픈소스 머신러닝 및 데이터 가시화 툴킷 중 하나인 Orange Data Mining 소프트웨어를 활용하여 딥러닝 기법의 태양광 패널 먼지오염 분류 예측 성능을 평가하였다. Fig. 3은 상기 소프트웨어에서 위젯(widget)과 네트워크를 이용하여 설계한 연구 절차(이미지 데이터 입력, 임베딩, 딥러닝 모델 학습 및 분석, 딥러닝 모델 성능 평가)를 보여준다. 주황색 위젯은 데이터를 스프레드시트 형식으로 보여주고, 노란색 위젯은 이미지 데이터의 입력, 임베딩, 가시화 등을 나타내며, 분홍색 위젯은 딥러닝 모델을, 하늘색 위젯은 딥러닝 모델의 평가 및 검증 기능을 수행한다.

자료 취득, 분류 및 전처리

본 연구에서는 Da Jiang Innovation(DJI) 社의 Mavic 2 Enterprise Advanced(이하 Mavic 2) 드론을 이용하여 강

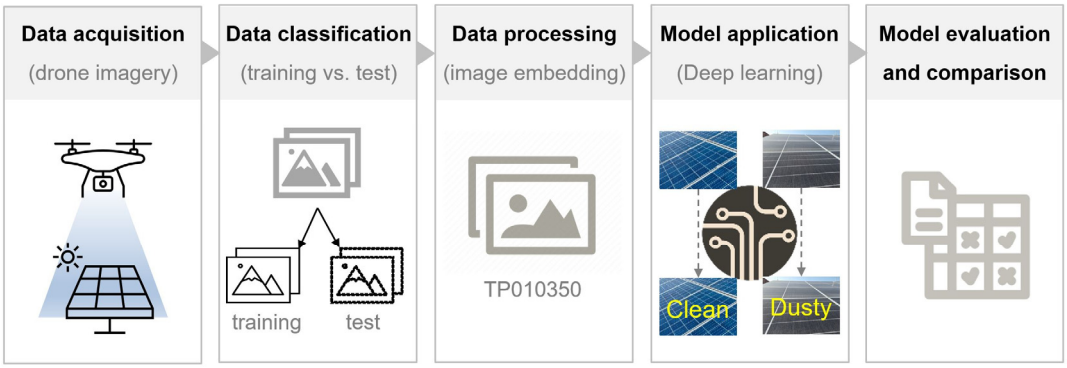

Fig. 2. Flowchart of PV panel dust detection using drone imagery and deep learning techniques.

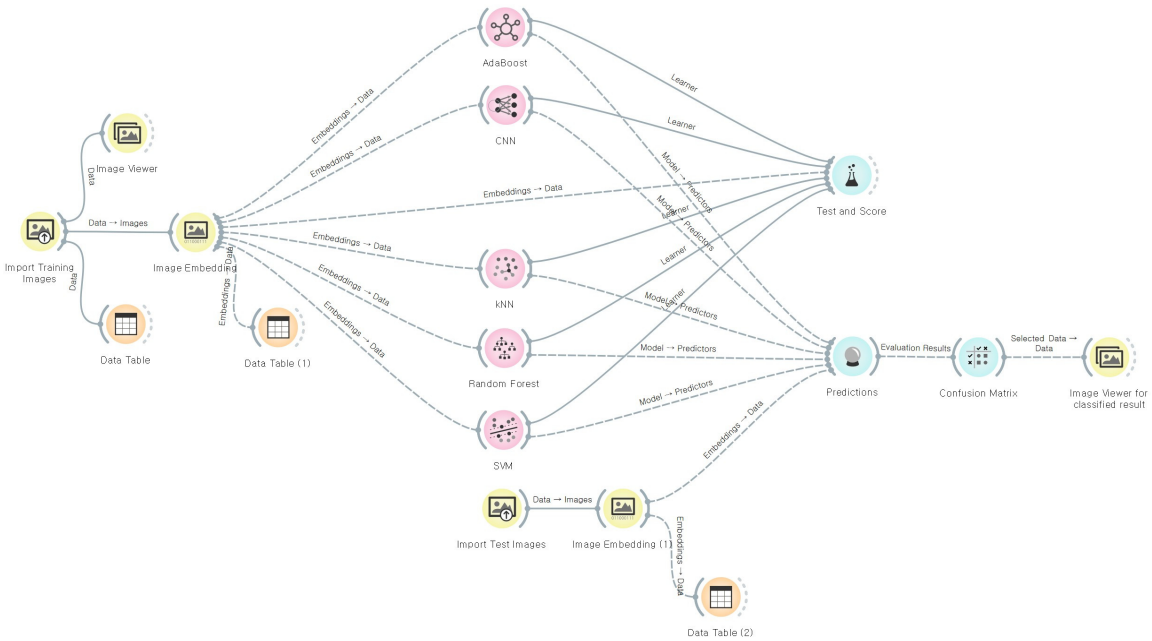

Fig. 3. Design of the widget-based analytical process in Orange Data Mining software for PV panel dust detection.

원특별자치도 삼척시의 태양광 패널에 대한 광학 및 열적외선 영상을 취득하였다. Mavic 2 드론은 통합 고해상도 열화상 센서를 사용하여 현장의 물체를 빠르게 인식할 수 있고, 안정화된 3축 짐벌에 가시광선 카메라와 열화상 카메라를 장착하여 가시광선 및 적외선 영상을 동시에 제공한다는 장점이 있다(Table 1). 태양광 패널의 드론 영상을 취득하기 위해 국제전기표준회(International Electronical Commission, IEC)에서 제시하는 촬영 가이드라인을 참조하였다. 모든 태양광 패널에 대해 충분한 영상자료를 확보하기 위해 IEC 표준에 따라 공간해상도를 3 cm/pixel로 설정하였다. 드론 촬영 날짜를 선정할 때 기온, 습도, 바람, 강수, 일조량 등의 기상 조건을 모두 고려했으며, 일출 후 2시간과 일몰 전 2시간(그림자 문제로 정확성 떨어짐), 정오(강한 빛으로 인해 반사되는 빛이 측정) 시간대를 피하여 오전 10시와 오후 3시경에 촬영을 실시하였다.

강원특별자치도 삼척시에 대한 드론 촬영을 통해 본 연구에서는 먼지가 없는 태양광 패널 영상 500장, 먼지가 있는 태양광 패널 영상 500장 등 총 1,000장의 영상자료를 수집하였다. 일반적으로 딥러닝 분석에서 훈련 자료와 테스트 자료의 비율은 분류 예측 성능에 영향을 미치기 때문에(테스트 자료 대비 훈련 자료의 비율이 높을수록 예측 정확도가 향상됨) 적절한 수준의 비율을 설정하는 것은 매우 중요하다. 그러한 이유로 본 연구에서는 전체 영상자료에 계층적 샘플링을 적용하여 훈련 자료와 테스트 자료가 동일한 비율을 갖도록 각각 800장(80%)과 200장(20%)으로 분류하였다. 또한, 훈련 자료 800장은 먼지오염이 없는 태양광 패널 400장과 먼지오염이 있는 태양광 패널 400장의 영상으로 구성하였다. 테스트 자료 100장 역시 먼지오염이 없는 태양광 패널 50장과 먼지오염이 있는 태양광 패널 50장의 영상으로 구성하였다. 딥러닝 분석을 위해 본 연구에서 취득 및 분류한 영상자료의 개수는 Table 2에 제시되어 있다.

연구 자료 구축 과정에서 드론 영상에 나타난 태양광 패널의 먼지오염 여부, 이른바 실제값(ground truth)에 대한 결정은 두 명의 연구자를 통해 이루어졌다. 사람이 이미지에 대한 데이터 라벨링을 하는 것은 객관성이나 신뢰성 측면에서 문제가 제기될 수 있다. 그래서 먼지오염 여부에 대한 두 명의 판단이 동일한 경우에만 데이터 라벨링을 진행하고, 판단이 상이한 경우에는 연구 자료에 고려하지 않았다. 보수적인 접근을 위해 태양광 패널 영역 전체가 아닌 일부에라도 명확한 먼지오염의 흔적이 있거나 태양광 패널 표면에 색상이 달라 보이는 수준의 먼지오염이 있는 경우에는 먼지오염이 있는 것으로 간주하였다.

드론으로 취득한 영상자료는 비정형 자료이기 때문에 원자료의 형태로 딥러닝 분석을 하기에는 무리가 있다. 영상자료에 딥러닝 기법을 적용하기 위해서는 사람이 이해 가능한 영상을 컴퓨터가 처리 가능한 형태로 변경하는 전처리 작업이 필요하다. 이 과정을 이미지 임베딩(image embedding)이라고 한다. 일반적으로 이미지는 용량, 가로 및 세로 격자셀 개수와 같은 기본 정보만 포함하고 있다. 여기서는 딥러닝 모델이 영상을 구성하는 픽셀(pixel)의 색상 정보로부터 각 이미지의 특징 벡터(vector)를 계산하며, 각 이미지의 데이터 테이블에 이미지를 설명하는 수백~수천 개의 특징 벡터값을 생성한다. 그리고 이 값들은 이미지 분류

Table 1. Specification of DJI Mavic 2 Enterprise Advanced drone

Type	Features	Performance
Aircraft	Dimensions (L × W × H)	Folded: 214 × 91 × 84 mm3
	Takeoff weight (without accessories)	Unfolded: 322 × 242 × 84 mm3
	Max. speed	72 kph (S-mode, without wind)
	Max. flight time	31 min (measured while flying at 25 kph in windless conditions)
M2EA Thermal Camera	Sensor resolution	640 × 512 @30Hz
	Digital zoom	16 ×
	Accuracy of thermal temperature	Measurement: ± 2℃ or ± 2%, whichever is greater.

Table 2. Image data classification for deep learning-based detection of PV panel dust

Data type \ Data size	Clean PV	Dusty PV	Sum (ratio)
Training data	400	400	800 (80%)
Test data	100	100	200 (20%)
Sum (Ratio)	500 (50%)	500 (50%)	1,000 (100%)

등의 딥러닝 분석 기법에 활용된다. 이미지 임베딩의 자세한 적용 방식과 사용 방법에 대한 예시는 Godec *et al*.(2019)의 논문을 참고할 수 있다. 대표적인 임베더(embedder)로는 Inception v3(Szegedy *et al*., 2016), SqueezeNet(Iandola *et al*., 2016), VGG(Simonyan and Zisserman, 2014), Painters, DeepLoc(Almagro Armenteros *et al*., 2017) 등이 있다. 본 연구에서는 ImageNet 데이터 세트에서 정확도가 78.1% 이상인 것으로 입증된 영상 인식 모델이자 장기간 수많은 연구에서 나온 다양한 아이디어가 축적된 것으로 알려진 Inception v3을 임베더로 채택하였다.

자료 훈련 및 검증을 위한 교차검증

딥러닝 기법은 전술한 훈련 자료(training data)의 학습을 통해 태양광 패널 표면의 먼지오염 여부에 대한 알고리즘이나 규칙을 생성해내고, 별도의 검증 자료(validation data)를 이용하여 학습 능력을 평가한 후 학습된 모델을 이용하여 테스트 자료의 먼지오염 여부를 판단한다. 다만 딥러닝 기법의 자료 훈련(학습) 과정에서 일부 또는 특정 자료만 활용할 경우에는 샘플링 과정에서 편향(bias)이 생기거나 전체 훈련 자료의 특성을 반영하지 못하는 단점이 있다. 예를 들어, 만약 800개의 훈련 자료 중에 N개는 딥러닝 모델의 학습에 사용하고 나머지 800-N개는 검증에 사용한다고 가정해보자. 여기서 검증에 사용한 데이터의 분포나 특성이 훈련에 사용한 데이터의 그것과 상이한 경우 이는 편향된 학습 결과를 초래할 수 있다.

이러한 문제를 극복하기 위하여 본 연구에서는 전체 훈련 자료를 모두 검증용 자료로도 사용할 수 있는 계층적 샘플링의 일종인 k-폴드 교차검증(k-fold cross validation) 방법을 사용하였다. 데이터 훈련 과정에서 사용되는 k-폴드 교차검증 방법의 원리를 Fig. 4에 나타내었다. 이 방법은 훈련 자료 집합으로부터 k개의 상호 배타적인 같은 크기의 부분 집합(fold)을 만든 후에 k-1개의 폴드를 훈련용으로, 나머지 1개의 폴드를 검증용으로 활용하는 방식을 반복적으로 수행함으로써 편향을 최소화하고 훈련 과정에 대한 신뢰도를 확보할 수 있는 장점이 있다(Jung and Choi, 2022). 즉, 훈련 자료 전체가 딥러닝 모델의 학습과 검증에 활용되기 때문에 자료를 효율적으로 사용할 수 있는 동시에 자료 선택에 따른 편향성은 최소화 될 수 있다. 다만 K 값이 크면 반복 횟수가 많기 때문에 연산 시간이 증가하는 단점이 있다. 따라서 본 연구에서는 k값을 5로 설정한 5겹 교차검증을 수행하였다.

딥러닝 기법 선정

본 연구에서는 파이썬(Python) 기반의 오픈소스 머신러닝 및 데이터 가시화 툴킷 중 하나인 Orange Data Mining 소프트웨어를 활용하여 딥러닝 기법의 태양광 패널 먼지오염 분류 예측 성능을 평가하였다. 태양광 패널의 먼지 오염 발생 여부를 예측하기 위하여 동일 주제를 다룬 기존 문헌에서 많이 활용된 기법을 검토하였고, 본 연구에서는 AdaBoost(adaptive boost)와 CNN(convolutional neural network), kNN(k-nearest neighbor), RF(random forest), SVM(support vector machine) 등 총 5종류의 기법을 선정

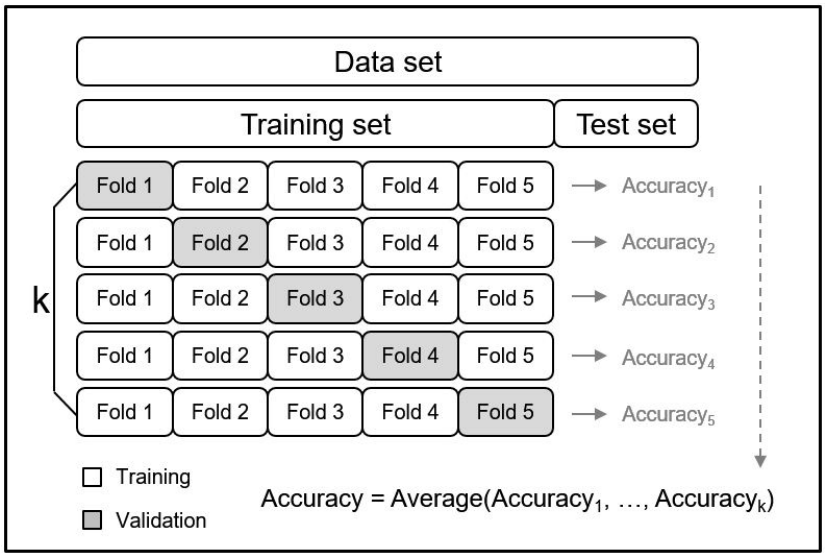

Fig. 4. Schematic diagram for k-fold cross-validation at the data training stage.

하였다. 해당 기법 중에는 머신러닝과 딥러닝 기법이 혼재되어 있으나 본 연구에서는 딥러닝 기법으로 명명하였다.

딥러닝 기법의 예측 성능평가

딥러닝 기법의 예측 성능평가 방법은 최종 예측값의 형태가 연속형(회귀)인지 이산형(분류)인지에 따라 달라진다. 본 연구에서는 태양광 패널의 먼지오염 여부를 예측하는 분류 문제에 해당하기 때문에 이에 적합한 5개의 성능평가 지표를 적용하였다.

Area under the curve(AUC) 지표는 receiver operating characteristic(ROC) 곡선이 이진 존재-부재 변수로 변환될 때 모든 임곗값에 대한 전체 기법의 성능을 요약한다(Lobo *et al*., 2008). ROC 곡선은 이진 분류 문제에서 임곗값이 변화할 때 민감도(Recall)와 1-특이도(Specificity)를 각각 축으로 놓고 해당 조합이 어떠한 형식으로 변하는가를 보여주는 그래프이다. 또한 여러 기법의 ROC 곡선은 비교할 때는 곡선 아랫부분의 면적(AUC)을 이용하며, 해당 면적이 클수록(최대값=1) 성능이 우수한 기법이라고 판단한다.

태양광 패널의 먼지오염 여부와 같은 분류 측기법의 성능을 평가하기 위해서는 주로 혼동행렬(confusion matrix)을 이용한다. 혼동행렬의 평가 요소로는 각각 TP(true positive)와 TN(true negative), FP(false positive), FN(false negative)이 존재하며, 해당 네 가지 요소들을 이용하여 적용된 기법을 성능을 평가한다(Fig. 5). TP와 TN은 실제로 발생된 사건(여기서는 먼지오염의 존재)을 올바르게 예측한 것이며, FP와 FN의 경우 실제 발생여부를 다르게 예측한 것을 의미한다. 전술한 혼동행렬의 4가지 요소들로부터 계산될 수 있는 네 가지 개념이 있는데 이는 각각 Classification accuracy(CA; 정확도)와 Precision(정밀도), Recall(재현율), F1 Score이다. 네 개 지표 모두 1에 가까울수록 높은 예측력을, 0에 가까울수록 낮은 예측력을 보이는 것으로 해석할 수 있다. CA(Fig. 5(a))는 가장 일반적인 딥러닝 기법의 성능 평가 지표로써 전체 데이터 중 실제로 긍정(positive)인 것을 positive로 예측하고 부정(negative)인 것을 negative으로 정확하게 예측하였는지를 보여주는 값이다. Precision(Fig. 5(b))의 경우 기법의 예측값이 얼마나 정확한가를 보여주는 지표로, 기법이 Positive라 예측 분류한 값 중 실측값이 Positive인 비율을 의미한다. Recall(Fig. 5(c))은 실제 실측값 중 기법이 예측한 실측값을 나타내는 지표이며, 실측값이 Positive인 값 중 기법이 Positive인 비율을 표시한다. F1 Score(Fig. 5(d))는 Precision과 Recall 모두를 고려한 수치이며, 이는 두 수치를 조화평균 내어 계산하는 지표이다.

연구결과 및 해석

딥러닝 기법의 훈련 자료 학습 성능평가 결과

딥러닝 기법이 훈련 데이터(강원특별자치도 삼척시에 위치한 태양광 패널 이미지 800장)로부터 먼지오염 여부의 분류 작업을 얼마나 정확하게 학습하였는지(학습력)를 분석하였다. 물론 본 연구의 최종 목적은 테스트 자료에서 높은 분류 예측 정확도를 갖는 것이며, 이는 테스트 데이터를 대상으로 적용한 것이 아니기 때문에 딥러닝 기법의 분류 예측 성능결과(예측력)라고 보기는 어렵다. 본 연구에서는

(a)

		Predicted	
		Positive	*Negative*
Actual	*Positive*	TP	FN
	Negative	FP	TN

$$CA = \frac{TP + TN}{TP + FN + FP + TN}$$

(b)

		Predicted	
		Positive	*Negative*
Actual	*Positive*	TP	FN
	Negative	FP	TN

$$Precision = \frac{TP}{TP + FP}$$

(c)

		Predicted	
		Positive	*Negative*
Actual	*Positive*	TP	FN
	Negative	FP	TN

$$Recall = \frac{TP}{TP + FN}$$

(d)

		Predicted	
		Positive	*Negative*
Actual	*Positive*	TP	FN
	Negative	FP	TN

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Fig. 5. Formulas used to evaluate the categorization prediction performance of deep learning algorithms.

Table 3에 제시된 바와 같이 모든 성능지표(5종)에서 훈련 데이터를 가장 잘 학습한 기법은 CNN 모델로 나타났다. kNN, SVM, RF, AdaBoost 기법이 그 뒤를 이었다. CNN 모델의 CA가 0.866이라는 것은 CNN 모델이 전체 800장의 이미지 중 86.6%에 해당하는 692장에 한 먼지오염 여부를 올바르게 학습했다는 의미로 해석할 수 있다. 5개의 성능 지표값으로 미루어볼 때 과적합에 해당하지는 않는 것으로 판단되었다. 다만 훈련 데이터 학습 방법으로 5-fold cross validation 기법을 사용했기 때문에 아래 결과수치는 매 시행 때마다 일부 변경될 수 있다. 다만 본 연구에서는 반복 시행 과정에서 5개 평가 지표에 대한 5개 딥러닝 알고리즘 순위의 변동은 거의 없었으며, 아래 표에 제시된 값은 10회 반복 시행한 결과의 평균치를 나타낸다.

딥러닝 기법의 테스트 자료 학습 성능평가 결과

본 연구에서 딥러닝 분석의 최종 목적은 훈련 자료(태양광 패널 이미지)의 학습을 통해 도출된 모델(learner)을 테스트 자료에 적용하여 태양광 패널 표면의 먼지오염 여부를 정확하게 예측(predict)하는 것이다. 따라서 진정한 의미에서의 딥러닝 기법의 분류 예측 성능 지표는 훈련 자료가 아닌 테스트 자료의 검증을 통해서만 계산 가능하다. 연구방법 부분에서 전술한 바와 같이 총 200장의 테스트 자료(먼지오염 100장, 비오염 100장)를 대상으로 딥러닝 기법이 예측한 결과와 실제 먼지오염 여부를 비교분석하였다.

Table 4는 태양광 패널의 먼지오염 분류 결과에 대한 딥러닝 모델별 혼동행렬 결과를 제시한 것이다. 본 연구의 경우 TP는 태양광 패널 표면의 먼지오염이 있는 상황을 맞게 예측한 경우(CNN 모델의 경우 94개), TN은 먼지오염이 없는 상황을 맞게 예측한 경우(38개), FP는 먼지오염이 없는 상황을 틀리게 예측한 경우(62개), FN은 먼지오염이 있는 상황을 틀리게 예측한 경우(6개)에 해당한다. 딥러닝 모델과 무관하게 실제로 존재하는 먼지오염은 맞게 예측한 경우가 많았던 반면, 실제로 먼지오염이 없는 경우를 오분류한 경우가 많은 것을 확인할 수 있다.

전술한 혼동행렬의 결과를 토대로 각 딥러닝 알고리즘에 대한 태양광 패널 표면 먼지오염 분류 예측 성능지표를 Table 5에 요약하였다. 훈련 자료의 학습력 결과와 마찬가지로 모든 측면에서 CNN 모델이 가장 높은 분류 예측 성능을 보였다. 그 뒤로는 RF, kNN, SVM, AdaBoost의 순위를 보였다. 이와 같은 결과는 각 딥러닝 모델이 개발되는 과정

Table 3. Learning performance index of each deep learning model for categorizing clean and dusty PV panels

Algorithm	AUC	CA	Precision	Recall	F1 score	Rank
AdaBoost	0.710	0.710	0.706	0.720	0.713	5
CNN	0.931	0.866	0.865	0.868	0.866	1
kNN	0.873	0.812	0.826	0.792	0.809	2
RF	0.855	0.786	0.804	0.757	0.780	4
SVM	0.875	0.781	0.761	0.820	0.789	3

Table 4. Results of the confusion matrix of deep learning algorithms for categorizing clean and dusty PV panels

AdaBoost		Predicted	
		Positive	Negative
Actual	Positive	65	35
	Negative	70	30

CNN		Predicted	
		Positive	Negative
Actual	Positive	94	6
	Negative	62	38

kNN		Predicted	
		Positive	Negative
Actual	Positive	82	18
	Negative	71	29

RF		Predicted	
		Positive	Negative
Actual	Positive	83	17
	Negative	63	37

SVM		Predicted	
		Positive	Negative
Actual	Positive	71	29
	Negative	76	24

에서 어떠한 가정 하에서 만들어지는지와 연관될 수 있다. 나머지 4개 모델과는 다르게 CNN 모델은 동일한 특징이 이미지 내의 여러 부분에 존재(stationarity of statistics)할 수 있고, 작은 영역 안에 픽셀의 종속성(locality of pixel dependencies)이 있는 이미지의 특성을 반영할 수 있는 모델이기 때문인 것으로 판단된다(Krizhevsky *et al*., 2012).

Table 4와 5에 제시된 5개 기법의 순위가 상이한 것으로 볼 때 반드시 높은 학습력을 보인 딥러닝 모델의 예측 정확도가 높은 것은 아님을 확인할 수 있었다. 단순한 가정으로 딥러닝 기법의 사용 목적이 예측에 초점을 두고 있다면 kNN 모델(학습 2위 및 예측 3위)보다는 RF 모델(학습 4위 및 예측 2위)을 선정하는 것이 적절하다고 볼 수 있다.

현재까지 알려진 바로 딥러닝 기법의 예측력은 학습력보다 낮게 나타나는 것이 일반적이다. 다만 아래 표에서 눈여겨볼 점인 Recall은 실제로 먼지오염이 존재하는 전체 경우의 수 대비 정확하게 예측한 비율을 의미한다. 5개 딥러닝 모델 모두 타 지표에 비해 Recall 수치가 높은 것을 확인할 수 있는데 이는 실제로 먼지오염이 존재하는 상황을 정확하게 예측한 비율이 높다는 의미이다. 반면에 Precision의 경우 실제로는 먼지오염이 없는 상황을 반대로 판단한 경우(FP 값)가 많았고, 이 영향으로 0.481~0.603 범위의 값을 나타냈다. 이로부터 추론해볼 때 본 연구에서 훈련 데이터로부터 딥러닝 기법이 학습한 모델은 먼지오염이 존재하는 것으로 판단하는 비율이 높은 것으로 사료된다.

Fig. 6은 태양광 패널의 먼지오염 여부를 오분류한 실제 드론 사진 예시를 보여준다. 전술한 혼동행렬에서 FP와 FN과 같이 오분류된 결과의 원인을 분석하였다. 실제로 깨끗한 모듈을 반대로 분류한 FP의 경우 대부분 이미지 내에 태양광 패널 외에 옥상이나 지붕 등이 회갈색 배경처럼 포함되었거나 태양광 패널의 설치 각도에 따라 빛이 반사되는 사례가 다수였다. 반면에 실제로 먼지오염이 존재하는 모듈을 반대로 깨끗한 모듈로 분류한 FN의 경우 먼지오염이 모듈 전반에 아주 옅게 분포하고 있거나 일부 공간에 작은 덩어리로 분포되어 있는 경우가 많았다. 물론 그림자 유무가 두 사례의 오분류에 영향을 주기도 하였다. 뿐만 아니라 드론이 태양광 패널을 촬영하는 각도, 드론과 태양광 패널 사이의 거리 등과 같은 드론 촬영 조건이나 기상 환경에 따라서도 먼지오염 분류의 정확도가 영향을 받을 수 있을 것이다.

토의 및 결론

본 연구에서는 드론으로 촬영한 영상자료와 5개의 딥러닝 기법을 이용하여 강원특별자치도 삼척시에 설치된 태양광 패널 규모에서의 먼지오염 여부를 예측하고 적용된 기법들의 성능을 검증하였다. 먼지오염 여부에 대한 예측결과와 실측결과의 비교를 통해 도출한 5개 성능지표를 종합적으로 고려했을 때 CNN 모델이 가장 높은 분류 예측 성능

Table 5. Predictive performance index of each deep learning model for categorizing clean and dusty PV panels

Algorithm	AUC	CA	Precision	Recall	F1 score	Rank
AdaBoost	0.475	0.475	0.481	0.650	0.553	5
CNN	0.806	0.660	0.603	0.940	0.734	1
kNN	0.639	0.555	0.536	0.820	0.648	3
RF	0.627	0.600	0.568	0.830	0.675	2
SVM	0.562	0.475	0.483	0.710	0.575	4

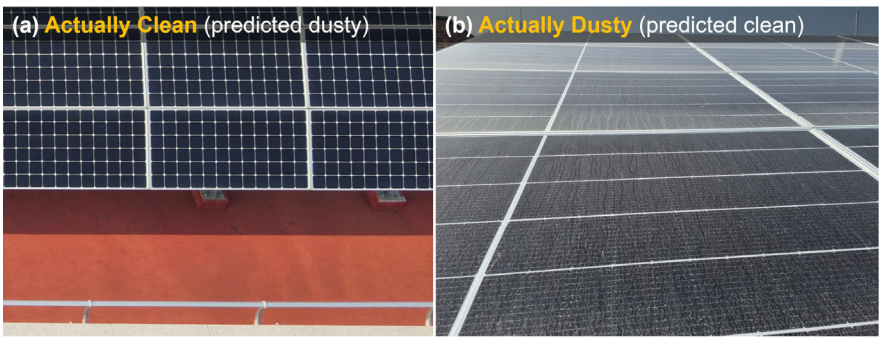

Fig. 6. Exemplary photos of the results of misclassification of dirt on a PV panel. (a) FP, (b) FN.

을 보이는 것으로 분석되었다. 훈련 자료와 테스트 자료 비율을 8:2로 구성했을 때 태양광 패널 표면 먼지 오염에 대한 CNN 기법의 분류 예측 성능은 area under the curve는 80.6%, classification accuracy는 66.0%, precision 60.3%, recall 94.0%, F1 score 73.4%로 평가되었다.

태양광 패널 표면의 먼지오염을 탐지하기 위한 접근법을 제안하였으나 5개 딥러닝 모델 모두 CA 값이 70% 미만에 그쳤다. 혼동행렬의 FP와 FN과 같이 오분류된 것은 이미지 내에 태양광 패널 외에 지붕이나 옥상 등의 다른 배경이 포함되었거나 모듈 내의 그림자 등이 주요한 원인인 것으로 사료되었다. 따라서 향후에는 드론 이미지 내에 태양광 패널 외의 객체(지붕, 옥상, 배경 등)가 포함되었을 때 컴퓨터 비전(computer vision) 기술 등을 적용하여 기타 객체를 자동으로 인식 및 처리함으로써 순수하게 태양광 패널 표면의 먼지오염 여부만을 훈련 단계에서 면밀하게 분석할 수 있는 기술 개발이 필요하다고 판단된다. 뿐만 아니라 드론 촬영 조건이나 기상 환경에 따라서도 먼지오염 분류 정확도가 영향을 받을 수 있기 때문에 추후에는 드론 촬영 방법에 대한 가이드 라인에 대한 연구도 필요할 것으로 사료된다. 또한, 최근 들어 Vision Transformer, Contrastive learning, EfficientNET 등과 같은 새로운 이미지 임베딩 기술이 발표되고 있으므로 향후에는 보다 개선된 이미지 임베더를 적용할 필요가 있을 것으로 사료된다.

태양광 패널의 먼지오염 탐지의 궁극적 목적은 세척(청소)을 통해 수광장애 요인을 제거함으로써 태양광 발전량의 저하를 방지하는 것이다. 따라서 본 연구에서 먼지오염이 존재하는 것으로 맞게 분류된 태양광 패널과 오분류되었지만 실제로 먼지오염이 존재하는 태양광 패널에 대한 세척 등의 유지관리 작업이 이어질 필요가 있다.

사 사

본 연구는 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원(No. NRF-2019R1I1A3A01062541) 및 삼표시멘트 사회공헌 기금으로 운영되는 삼척문화의 이해 플랫폼 사업 지원을 받아 수행된 연구임.

References

Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O., 2017. DeepLoc: prediction of protein subcellular localization using deep learning, *Bioinformatics*, 33(21), p.3387.

Chen, J., Pan, G., Ouyang, J., Ma, J., Fu, L., and Zhang, L., 2020. Study on impacts of dust accumulation and rainfall on PV power reduction in East China, *Energy*, 194, p.116915.

Cipriani, G., D'Amico, A., Guarino, S., Manno, D., Traverso, M., and Di Dio, V., 2020. Convolutional Neural Network for Dust and Hotspot Classification in PV Modules, *Energies*, 13(23), p.6357.

Conceição, R., Silva, H.G., Mirão, J., Gostein, M., Fialho, L., Narvarte, L., and Collares-Pereira, M., 2018. Saharan dust transport to Europe and its impact on photovoltaic performance: A case study of soiling in Portugal, *Solar Energy*, 160, p.94-102.

Cruz-Rojas, T., Franco, J.A., Hernandez-Escpbedo, Q., Ruiz-Robles, D., and Juarez-Lopez, J.M., 2023. A novel comparison of image semantic segmentation techniques for detecting dust in photovoltaic panels using machine learning and deep learning, *Renewable Energy*, 217, p.119126.

Gao, Y. and Li, S., 2023. A deep learning-based method detects dust from solar PV panels through Unmaned Aerial Vehicles, *Journal of Physics: Conference Series*, 2584, p.012019.

Godec, P., Pančur, M., Ilenič, N., Čopar, A., Stražar, M., Erjavec, A., Pretnar, A., Demšar, J., Starič, A., Toplak, M., Žagar, L., Hartman, J., Wang, H., Bellazzi, R., Petrovič, U., Garagna, S., Zuccotti, M., Park, D., Shaulsky, G., and Zupan, B., 2019. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning, *Nature communications*, 10 (1), p.4551.

Han, J., Choi, S., Kim, S., and Jung, Y., 2018. A Study on the Contamination of Photovoltaic Cells by Fine Dust in the Air, *Transactions of Korean Hydrogen and New Energy Society*, 29(3), p.292-298.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., and Keutzer, K., 2016. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size*, arXiv preprint arXiv:1602.07360.

Jung, D. and Choi, Y., 2022. Development of Machine Learning Models for Predicting Air Overpressure in an Open-pit Mine, *Journal of The Korean Society of Mineral and Energy Resources Engineers*, 59(1), p.59-68.

Kaldellis, J.K., Fragos, P., and Kapsali, M., 2011. Systematic experimental study of the pollution deposition impact on the energy yield of photovoltaic installations, *Renewable energy*, 36(10), p.2717-2724.

Kim, S.K., 2015. *Detection of Surface Contamination of Photovoltaic Module Using Image Sensor*, Proceeding of 46th the Korean Institute of Electrical Engineers, Muju, p.1086-1087.

Kim, D., Park, S., Moon, J., and Hwang, E., 2021. *Machine Learning-Based Solar Power Efficiency Prediction Scheme*

Considering PM10 Fine-Dust Data, Proceedings of 23rd the Korean Institute of Information Scientists and Engineers, Online, p.272-274.

Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 25p.

Lee, G., Lee, G., and Kang, S., 2017. A Case Study for Analyzing the Optimal Location for A Solar Power Plant via AHP Analysis with Fine Dust and Weather, *Journal of the Korea Safety Management & Science*, 19(4), p.157-167.

Lobo, J.M., Jimenez-Valverde, A., and Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models, *Global ecology and Biogeography*, 17(2), p.145-151.

Menoufi, K., Fargal, H.F.M., Farghali, A.A., and Khedr, M.H., 2017. Dust accumulation on photovoltaic panels: a case study at the East Bank of the Nile (Beni-Suef, Egypt), *Energy Procedia*, 128, p.24-31.

Park, S., Bang, J., Ruy, I., and Kim, T., 2019. The Prediction of Photovoltaic Power Using Regression Models Based on Weather Big-data and Sensing Data, *The Transactions of the Korean Institute of Electrical Engineers*, 68(12), p.1662-1668.

Rao, A., Pillai, R., Mani, M., and Ramamurthy, R., 2014. Influence of Dust Deposition on Photovoltaic Panel Performance, *Energy Procedia*, 54, p.690-700.

Renewable Energy Cloud Platform, 2023.12.31., https://recloud.energy.or.kr/main/main.do

Saini, R.K., Saini, D.K., Gupta, R., Verma, P., Dwivedi, R., Kumar, A., Chauhan, D., and Kumar, S., 2023. Effects of dust on the performance of solar panels – a review update from 2015-2020, *Energy & Environment*, 34(6), p.2110-2162.

Schill, C., Anderson, A., Baldus-Jeursen, C., Burnham, L., Micheli, L., Parlevliet, D., Pilat, E., Stridh, B., and Urrejola, E., 2022. *IEA-PVPS task 13: Performance, operation and reliability of photovoltaic systems–Soiling Losses – Impact on the Performance of Photovoltaic Power Plants*, Report. International Energy Agency.

Simonyan, K. and Zisserman, A., 2014. *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556.

Son, J. and Jeong, S., 2019. *A Research of the Effects of Particulate Matter on Solar Photovoltaic*, Proceedings of Korea Environmental Policy, Seoul, p.52-54.

Sung, S. and Cho, Y., 2019. Prediction of Photovoltaic Power Generation Based on Machine Learning Considering the Influence of Particulate Matter, *Environmental and Resource Economics Review*, 28(4), p.467-495.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z., 2016. *Rethinking the inception architecture for computer vision*, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p.2818-2826.

You, H. and Bae, S., 2020. *Deep neural network photovoltaic power generation forecasting with fine dust data*, Proceeding of 51st the Korean Institute of Electrical Engineers, Busan, p.547-548.

한 호 상

2022년 강원대학교 공학대학 에너지자원융합공학전공 공학사

현재 강원대학교 에너지자원융합공학과 석사과정
(E-mail; qkfehr@gmail.com)

서 장 원

2013년 서울대학교 공과대학 에너지시스템공학부 공학박사

현재 강원대학교 공학대학 에너지자원화학공학과 부교수
(E-mail; jangwonsuh@kangwon.ac.kr)