

IoT 및 딥 러닝 기반 스마트 팜 환경 최적화 및 수확량 예측 플랫폼

최호길* 안희학** 정이나*** 이병관****

A Smart Farm Environment Optimization and Yield Prediction Platform based on IoT and Deep Learning

Hokil Choi* Heuihak Ahn** Yina Jeong*** Byungkwan Lee****

요약 본 논문은 농장의 바이오 센서 데이터를 수집해서 농장에서 재배중인 농작물의 질병을 진단하고, 그 해 수확량을 예측하는 IoT 및 딥 러닝 기반 스마트 팜 환경 최적화 및 수확량 예측 플랫폼을 제안한다. 이 플랫폼은 현재 날씨, 토양 미생물 등 수집 가능한 모든 정보를 수집하여 작물이 잘 성장할 수 있도록 농장 환경을 최적화하고, 농장에서 재배 중인 작물의 잎을 이용하여 작물의 질병을 진단하고, 그리고, 농장의 모든 정보를 사용하여 올해 수확량을 예측한다. 실험 결과 AEOM(Agricultural Environment Optimization Module)의 평균 정확도는 RF(Random Forest)보다 약 15%, GBD(Gradient Boosting Tree)보다 약 8% 높고, 데이터가 증가해도 RF나 GBD에 비해 정확도가 덜 감소한다. 선형 회귀에 따르면 정확도의 기울기는 ReLU의 경우 $-3.641E-4$, Sigmoid의 경우 $-4.0710E-4$, 계단함수의 경우 $-7.4534E-4$ 이다. 따라서 ReLU 사용시 정확도 기울기가 가장 낮으므로 테스트 데이터의 양이 증가함에 따라 ReLU는 다른 두 가지 활성화 기능보다 더 정확하다. 본 논문에서 제안한 EOYPP는 농장 전체를 관리하는 플랫폼으로 실제 농장에 도입된다면 국내 스마트 팜의 발전에 크게 이바지할 것이다.

Abstract This paper proposes “A Smart Farm Environment Optimization and Yield Prediction Platform based on IoT and Deep Learning” which gathers bio-sensor data from farms, diagnoses the diseases of growing crops, and predicts the year's harvest. The platform collects all the information currently available such as weather and soil microbes, optimizes the farm environment so that the crops can grow well, diagnoses the crop's diseases by using the leaves of the crops being grown on the farm, and predicts this year's harvest by using all the information on the farm. The result shows that the average accuracy of the AEOM is about 15% higher than that of the RF and about 8% higher than the GBD. Although data increases, the accuracy is reduced less than that of the RF or GBD. The linear regression shows that the slope of accuracy is $-3.641E-4$ for the ReLU, $-4.0710E-4$ for the Sigmoid, and $-7.4534E-4$ for the step function. Therefore, as the amount of test data increases, the ReLU is more accurate than the other two activation functions. This paper is a platform for managing the entire farm and, if introduced to actual farms, will greatly contribute to the development of smart farms in Korea.

Key Words : Agricultural, Analysis System, Artificial Intelligence System, CNN, Smart Farm

1. 서론

현재 농업에 대한 많은 사람들의 인식은 힘든 일이

지만 소득이 적다는 것이다. 농촌 인구의 감소와 농촌의 고령화 문제가 심각한 수준이다. 이러한 농촌의 문제를 해결할 수 있는 스마트 팜이 많은 관심을 받고 있

This work was supported by the Ministry of SMEs and Startups grant funded by the Korea government (No. G21S269 429901)

*Department of Computer Engineering, Catholic Kwandong University

**Corresponding Author : Department of Software, Catholic Kwandong University (hhahn@cku.ac.kr)

Received November 16, 2019

Revised December 17, 2019

Accepted December 23, 2019

다. 스마트 팜(Smart Farm)은 관찰, 처방, 작업, 결과 분석 4개의 단계로 ICT 기술을 사용한다. 관찰 단계는 여러 가지 센서를 이용하여 농경지의 환경과 작물 생육 상태를 조사하여 기초 데이터를 만들어내는 단계이다. 처방단계는 관찰 단계에서 만들어낸 데이터를 이용하여 문제점을 해결하나 개선하는 방법을 추출하는 단계이다. 작업 단계는 처방 단계에서 결정한 방법들을 수행하도록 기계장비인 로봇, 드론 등을 이용해 작업을 수행하는 단계이다. 결과분석 단계는 빅데이터, AI 등을 이용해 산출된 데이터를 분석하여 수정하고 보완하는 작업을 통해 농업을 환경을 점점 개선하는 단계이다. 이러한 스마트 팜의 활용성 때문에 전 세계적으로 다양한 방법으로 사용하고 있다. 스마트 팜 산업이 이 전부터 농업에 사용되었지만 활발하게 사용된 기간은 얼마 되지 않는다. 따라서 몇 가지의 문제점이 제시된다. 먼저, 농작물을 관리하고 수확하는 기술은 많이 사용되고 있으나 병충해를 예방하기 위한 기술은 부족하다. 농작물의 생육 환경을 빠르게 개선하기 위해서는 실시간적인 작물, 환경이 분석되고 개선 방법을 제시해야 한다. 또한, 작물의 관리뿐만 아니라 작물의 수확량을 예측하여 효과적인 스마트 팜의 개선이 필요하다. 그러므로 본 논문에서는 농업환경 구성 및 수확량을 예측하기 위하여 IoT 및 딥 러닝 기반 스마트 팜을 위한 환경 최적화 및 수확량 예측 플랫폼(EOYPP)을 제안한다.

2. 관련 연구

2.1 스마트 팜

최휘민은 국내에 보급되는 스마트팜 기기는 업체마다 제품 규격이 다르고 소프트웨어와 하드웨어가 별도로 개발되고 있는 등 호환성과 품질이 떨어져, 저가의 스마트 팜 기기를 사용하는 농가의 피해가 자주 발생하여 기기의 이상을 파악할 수 있는 시스템이 필요하다. 그러므로 국내·외 스마트팜 정책 동향과 국내의 스마트 농업 동향을 살펴보고 스마트팜의 오작동과 고장에 의한 이상 상황을 분석하여, 미리 예방하고 문제 상황이 발생할 때 사용자에게 알려주는 시스템을 설계했다[1].

황성일은 수경 인삼 식물공장의 사례를 통하여, 스

마트 팜 팩토리를 ICT 기반에서 설계하였다. 스마트 팜 팩토리는 환경정보를 수집하여 시설물을 제어하는 기존의 식물공장 개념을 확장하여 식물의 생육을 위한 정보 수집과 시설 운영 그리고 관리의 자동화가 가능하고, 수집된 정보를 이용하여 최적의 작물 생육을 위한 정보의 분석과 적용을 가능하도록 한 전사적인 식물공장 관리 시스템이다[2].

이시현은 두 개의 범용 임베디드 컴퓨팅 플랫폼을 이용하여 사용자의 요구사항을 쉽게 적용하고, 시스템의 유연성과 시장의 적기성을 증대시키기 위한 목적으로 스마트 팜에 요구되는 기본적인 기능을 설계하였다. 임베디드 컴퓨팅 플랫폼은 스마트 팜의 모터 제어와 센서 등의 기능을 제어하고 임베디드 컴퓨팅 플랫폼과 통신 기능을 수행한다[3].

안원영은 축산 사료의 측정 장치를 IoT 기반으로 개발하고, 자동 주문과 배송 시스템을 구현하여 관련 영역의 효율을 극대화하는 방안을 제안했다. 구체적으로는 첫째, 온·습도 측정 센서와 초음파 센서를 이용하여 온·습도 정보와 사료 저장소인 피드빈 내 사료 잔량을 모으는 센서를 개발했다. 둘째, 축산농가에 설치된 게이트웨이를 거쳐서 정보를 SCM 서버에 전송하는 방법을 구현했다[4].

박용주는 번식 축사 환경에 적용하도록 THI 지수를 계산하여, 농장 환경 자동 제어 시스템과 무선 센서 네트워크를 구현했다. 먼저 축산물 온도와 심박수 및 운동량과 같은 환경 정보를 수집하고, 온·습도와 조도 등 농장 환경 데이터를 수집하여 THI 지수를 계산한다. 둘째, 자동으로 최상의 상태를 유지하도록 수집된 데이터를 이용하여 지능형 농장의 지붕 개방과 진동 팬을 제어한다[5].

2.2 수확량 예측

김영환은 우리나라의 여섯 개 주요 수종에 대해 임분 수확 예측모델을 이용하여 간벌 시업에 따른 임분의 생장 변화와 수확량을 예측했다. 간벌 시나리오의 설정을 위해 간벌 시기를 15년에서 40년까지 5년 단위로 구분하되 간벌의 주기는 10년 이상이 되도록 하고, 간벌로 제거되는 비율을 전체 60%를 넘지 않도록 제한했다. 이런 제한조건을 적용하여 열 가지 간벌 시

나리오를 설정하고 수종별 수확량과 지수별 수확량을 비교 분석했다. [7]

성제훈은 정밀 농업의 비료나 농약과 같은 농작물 처리의 변화된 적용으로 현장 내에서 질작 효율의 향상을 위해 수율과 토양 특성 사이의 관계에 대한 완벽한 이해를 위해 측정된 토양 특성과 쌀 수확량 사이의 기능적 관계를 조사했다[8].

김주희는 경제적 방제수준을 설정하고자 고추 흰가루병이 수량에 미치는 영향을 조사하고 시험했다. 고추 흰가루병의 병 발병 정도에 따른 수량의 변화에 대한 분석을 10단계로 구분하여 시험했다. 흰가루병 발생수준, 과장과 과경 그리고 과중은 부의 상관관계를 형성하고, 발생의 증가에 따라 수량이 감소하여 유의한 부의 상관관계가 형성된다[9].

오정원은 최상의 품질 상태에서 과일을 수확하고 좋은 가격으로 출하하도록, 수확 시기에 기계 학습을 적용하는 방법을 제안했다. 온실 내에서 재배되는 과일의 내부 온습도 값과 색상 값 그리고 무게 값을 온습도 센서와 색상 센서 그리고 무게 센서를 이용하여 수확 시기 예측에 필요한 인자를 실시간으로 수집하고 확보했다[9].

손영모 외 4명은 삼나무의 생장 및 수확량 평가와 탄소저장 및 흡수량을 알아보기 위하여 수행했다. 직경 분포 추정을 위하여 흉고직경, 흉고단면적, 수고 등 생장인자별 생장 추정식을 도출하고, 이들에 대한 적합성을 검증했다. 그리고 삼나무에 대한 지역별 임지생산력을 파악할 수 있는 지위지수를 Schumacher 모델로서 개발하였다. [10]

3. IoT 및 딥 러닝 기반 스마트 팜을 위한 환경 최적화 및 수확량 예측 플랫폼 설계

3.1 AEOM 설계

AEOM(Agricultural Environment Optimization Module)은 현재 날씨, 토양 미생물 함량 등 농장에서 수집 가능한 모든 정보를 수집하여 작물이 잘 성장할 수 있도록 농장 환경을 최적화한다. AEOM은 그림 1과 같이 신경망 모델을 활용하여 농장의 온도, 토양의 영양제량, 스프링클러의 작동(수분 보충량), CO2 농도

조절 등을 수행한다. AEOM은 과거 환경 데이터와 농장 관리 데이터를 트레이닝 데이터 세트로 사용하여 신경망 모델을 학습시킨다. 여기서 과거 환경 데이터는 과거 특정 날짜의 날씨, 토양 상태 등을 의미하고, 농장 관리 데이터는 그 환경 데이터에 따라 농장에서 이루어진 대처(영양제 공급, 수분 보충 등)를 의미한다.

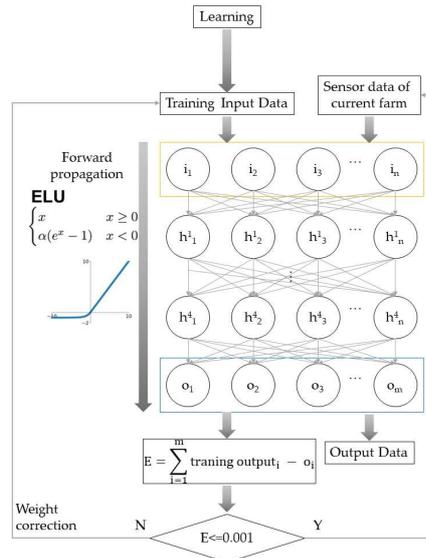


그림 1. AEOM의 신경망 모델
Fig 1. Neural network model of AEOM

AEOM은 다음과 같이 동작한다. 첫째, AEOM은 우선 신경망 모델의 입력, 출력, 연결 강도, 학습률, 편향(bias)을 초기화한다. 수식 1은 AEOM의 입력을, 수식 2는 출력을, 수식 3은 연결 강도를 나타낸다.

$$I = \{i_1, i_2, i_3, \dots, i_n\} \quad (1)$$

$$O = \{o_1, o_2, o_3, \dots, o_m\} \quad (2)$$

$$W = \{w_{11}^1, w_{12}^1, w_{13}^1, \dots, w_{nm}^5\} \quad (3)$$

여기서 w_{11}^1 중 위 첨자는 연결된 히든 레이어가 몇 번째 히든 레이어인지를 의미하고 아래 첨자의 왼쪽은 직전 레이어의 노드 번호, 오른쪽은 다음 레이어의 노드 번호를 의미한다. 예를 들어 w_{11}^1 은 입력레이어의 1번 노드와 첫 번째 히든레이어의 1번 노드 사이의 연결 강도를 의미한다. AEOM이 가지는 히든레이어는 수식 4에 표현된다. 본 논문에서는 4개의 히든레이어를 사용하였으며, 히든 레이어의 노드 수는 입력 레이어

어의 노드 수와 같다.

$$H = \{h_1^1, h_2^1, h_3^1, \dots, h_n^4\} \quad (4)$$

둘째, AEOM은 입력과 출력 벡터를 0으로 설정하고, Glorot Initialization 방식을 사용하여 연결 강도를 초기화하고, 학습률은 0.001로, 편향은 0으로 초기화한다. 수식 5는 Glorot Initialization를 통하여 AEOM이 연결 강도를 초기화하는 것을 나타낸다. 여기서, n_{in} 는 직전 레이어의 노드 수를 의미하고, n_{out} 은 음 레이어의 노드 수를 의미한다.

$$W = 2/n_{in} + n_{out} \quad (5)$$

셋째, AEOM은 트레이닝 데이터를 사용하여 신경망 모델을 학습시킨다. AEOM은 초기화된 연결 강도, 트레이닝 입력 데이터 과 활성화 함수인 ELU함수를 사용하여 트레이닝 입력 데이터의 결과를 계산한다. 수식 6은 ELU함수를 나타내고, 수식 7은 각 노드의 결과를 계산하는 방법을 나타낸다.

$$\begin{aligned} \text{if } (x > 0) f(x) &= x \\ \text{else } f(x) &= \alpha(e^x - 1) \end{aligned} \quad (6)$$

$$h_a^j = ELU\left(\sum_{k=1}^n h_k^{j-1} w_{ka}^j\right) \quad (7)$$

$$* h_a^0 = i_a, h_a^5 = o_a, \quad a = 1 \text{ to } 5$$

넷째, AEOM이 수식 7을 통하여 트레이닝 입력 데이터에 대한 결과 O를 계산했다면, 트레이닝 출력 데이터와 손실함수를 이용하여 연결 강도를 수정한다. 수식 8, 9, 10은 손실함수를 이용한 오차 신호 δ 를 계산하고, 수식 11, 12, 13은 AEOM이 연결 강도를 수정하는 과정은 나타낸다.

$$\delta_{o_i} = o_i - t_{o_i} \quad (8)$$

$$\delta_{h_i^4} = \delta_{o_i} w_i^4 \quad (9)$$

$$\delta_{h_i^c} = \sum_{k=1}^n \delta_k^{c+1} w_{ik}^{c+1} \quad (10)$$

$$w_{ij}^4 = w_{ij}^4 + \alpha h_i^5 \delta_{o_j} \quad (11)$$

$$w_{ij}^k = w_{ij}^k + \alpha h_i^k \delta_{h_j^k} \quad (12)$$

$$w_{ij}^1 = w_{ij}^1 + \alpha o_i \delta_{h_j^1} \quad (13)$$

다섯째, AEOM의 학습이 끝나면 AEOM은 현재 농장의 상태를 입력 I로 사용하고 수식 7을 통하여 농장 환경을 최적화하는 출력 O를 계산한다.

3.2 CIAM 설계

본 논문에서는 CNN을 생성하여 농작물의 질병을 진단할 수 있는 CIAM(Crop Image Analysis Module)을 제안한다. CIAM은 작물 별로 서로 다른 CNN모델을 생성한다. 즉, 벼에 대한 CNN, 토마토에 대한 CNN등 작물 마다 하나의 CNN모델이 생성된다. 이 CNN들은 모두 같은 방법을 이용하여 생성된다. CIAM의 입력은 128*128이미지를 3개의 채널(R, G, B)로 나눈 128*128*3의 이미지이다. CIAM의 출력 노드의 수는 각 작물이 가질 수 있는 질병의 수이다. 예를 들어 CIAM이 토마토의 CNN을 생성한다면, 그 CNN의 출력 노드의 수는 토마토가 가질 수 있는 질병 11개와 정상상을 의미하는 노드 1개, 총 12개가 된다. CIAM의 출력은 0~1사이의 값이며, 출력 노드 중 가장 확률 값이 높은 노드를 해당 작물의 질병으로 결정한다. CIAM은 세번의 컨블루션(Convolution)과정과 서브 샘플링(sub-sampling)과정, 그리고 2개의 레이어를 사용하는 완전연결(Fully Connected) 신경망 연산을 사용하여 CNN을 생성한다. CIAM의 컨블루션 과정과 서브 샘플링과정은 식물의 앞에서 질병의 특징을 추출하고, 완전연결 신경망은 CIAM의 최종 출력인 질병명과 질병의 가능성을 계산한다.

컨블루션과정은 이미지를 스캔하여 이미지의 특징을 추출하고 그 결과인 컨블루션 피쳐 맵(Convolutional Feature map)을 결과로 출력하는 과정이다. 컨블루션과정에는 n*n의 정방 행렬을 이용하여 이미지를 스캔하는 필터와 필터의 이동 단위인 Stride가 사용된다. CIAM은 4*4의 필터를 사용한다. 필터의 개수는 첫 번째 연산에서는 20개, 두 번째 연산에서는 40개, 3번째 연산에서는 60개를 사용한다. 세 번의 컨블루션 과정에서 Stride는 모두 1을 사용한다. 컨블루션은 이미지와 필터 사이에 ReLU함수를 사용하여 컨블루션 피쳐 맵을 계산한다.

컨블루션과정을 통하여 계산된 컨블루션 피쳐 맵은 서브 샘플링을 통하여 간소화된다. CIAM은 서브 샘플

링 기법 사이에 하나인 맥스 풀링(max pooling)을 사용한다. 맥스 풀링이란 일정 크기의 이미지 픽셀 사이에 가장 특징이 강한 영역을 추출하여 하나의 픽셀로 만드는 것이다. CIAM은 2*2 맥스 풀링 기법을 사용하여 컨블루션 피쳐 맵을 간소화한다. 그림 2는 세 번의 컨블루션과정과 서브 샘플링과정을 거친 이미지의 컨블루션 피쳐 맵을 나타낸다.

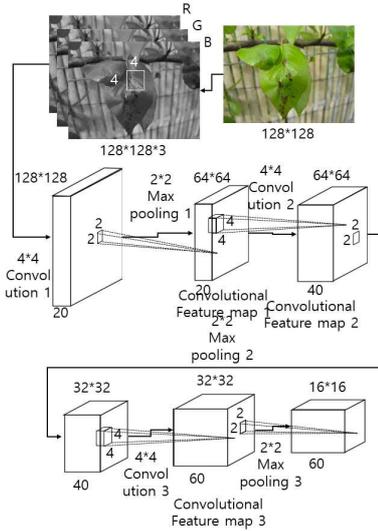


그림 2. 컨블루션과정과 서브 샘플링과정
Fig 2. Convolution and sub-sampling courses

그림 2는 세 번의 컨블루션과정과 서브 샘플링과정을 거친 이미지의 컨블루션 피쳐 맵을 나타낸다. 그림 2에서 CIAM의 가지는 최종 컨블루션 피쳐 맵의 크기는 16*16이며, 채널은 60개가 된다. CIAM은 세 번의 컨블루션과정과 서브 샘플링과정이 종료된 16*16*60의 컨블루션 피쳐 맵을 배열로 바꾸고, 완전연결 신경망을 사용해 작물의 질병의 수와 같은 크기의 배열을 생성한다. CIAM은 신경망의 활성화 함수로 Softmax를 사용하고, Softmax의 손실 함수를 이용하여 신경망을 학습시킨다. Softmax의 결과는 0~1사이의 값이며, CIAM은 작물의 질병 목록 중 Softmax값이 가장 높은 노드의 결과를 사용자에게 전달한다. 수식 14과 수식 15는 Softmax함수와 Softmax의 손실 함수를 나타낸다.

$$softmax f(x_i) = \frac{e^{x_i}}{\sum_{k=1}^n e^{x_k}} \quad (14)$$

$$loss function dE = - \sum_{k=1}^n t_k \ln(y_k) \quad (15)$$

수식 15에서 t_k 는 학습 데이터이며, y_k 는 신경망의 결과다. CIAM이 E를 사용하여 완전연결 신경망을 학습하면, CIAM은 이미지를 인식하여 해당 작물이 어떤 질병에 걸렸는지 진단하고, 그 결과를 사용자에게 알린다.

3.3 YCM 설계

본 논문에서 제안하는 YCM(Yield Calculation Module)은 현재 기상데이터와 작물의 상태를 계산하여 작물 수확량을 예측한다. YCM은 작물 수확량을 정확하게 예측하기 위해 ANN을 기반으로 설계되었다. YCM은 입력으로 CIAM의 출력(질병 명), 작물 이름, 수확까지 남은 시간(일 단위), 해당 지역의 현재 기온, 습도, 강수량(강설량), 일조량, 지면 온도, 현지 기압, 지면의 수분 증발량을 사용한다. 즉, YCM의 입력 레이어 노드의 수는 10개가 된다. 수식 16은 YCM의 입력을 나타낸다.

$$X = \{x_1, x_2, x_3, x_4, \dots, x_{10}\} \quad (16)$$

YCM은 출력으로 입력된 작물의 예상 수확량을 계산한다. 즉, YCM의 출력 노드의 수는 1개이며, y 로 표시한다. YCM은 히든 레이어를 5개 사용하고, 각 히든 레이어의 노드 수는 20개로 설계한다. YCM에서 사용하는 히든 레이어는 수식 4에서 표현된다.

$$H = \{h_1^1, h_2^1, h_3^1, h_4^1, \dots, h_{20}^5\} \quad (17)$$

수식 17에서 위 첨자는 히든 레이어의 번호를 의미하며, 아래 첨자는 노드 번호를 의미한다. YCM의 각 노드는 서로 연결되어 있으며, 노드 사이의 연결 강도를 가중치로 표현한다. 가중치는 수식 18에서 표현된다.

$$W = \{w_{1,1}^1, w_{1,2}^1, w_{1,3}^1, w_{1,4}^1, \dots, w_{20}^6\} \quad (18)$$

수식 18에서 $w_{1,1}^1$ 는 입력 레이어의 1번 노드와 첫 번째 히든 레이어의 1번 노드 사이의 가중치를 의미한다. 다음으로 YCM의 학습률, 초기 가중치가 설정되어야 하고 학습 방법이 설명되어야 한다. YCM의 학습률은 0.003으로, 초기 가중치는 모두 0.05로 설정된다. YCM은 학습방식으로 Back-Propagation을 사용한다. YCM은 학습을 위하여 손실 함수와 활성화 함수를 사용한다. 수식 19는 YCM의 손실 함수를 나타내며, 수식 20은 YCM의 활성화 함수를 나타낸다.

$$E = \frac{1}{n} \sum_k (y_k - d_k)^2 = (y - d)^2 \quad (19)$$

$$f(x) = \max(0, x) \quad (20)$$

수식 19는 YCM의 손실 함수인 평균 제곱 오차 함수이며, 수식 20은 YCM의 활성화 함수인 ReLU 함수이다. 수식 19에서 n 은 출력 노드의 수다. YCM의 출력 노드는 1개이기 때문에, 손실 함수 E 는 간단하게 표현될 수 있다. 평균 제곱 오차 함수와 ReLU함수는 노드의 결과가 음수일 때 낮은 정확도를 갖지만, 계산 속도가 빠르다는 장점이 있다. YCM의 입력력 데이터는 모두 양수이기 때문에, 수식19와 수식20은 YCM의 성능 저하 없이 빠른 속도를 보장한다. YCM이 활성화 함수를 이용해 출력 노드까지 노드 값을 계산했다면, YCM은 back-propagation 학습을 위하여 출력 노드의 값을 사용하여 출력 노드의 오차 신호 δ_y 를 계산한다. 수식 21은 출력 노드의 오차 신호를 나타낸다.

$$\delta_y = (d - y) \quad (21)$$

여기서 d 는 학습 데이터이고, y 는 YCM의 출력이다. δ_y 가 계산되었다면, YCM은 δ_y 를 이용하여 마지막 히든 레이어의 오차 신호 $\delta_{h_i^5}$ 를 계산한다. 식 22는 마지막 히든 레이어의 오차 신호를 계산하는 것을 나타내고, 수식 23은 1~4 번째 히든 레이어의 오차 신호를 계산하는 것을 나타낸다.

$$\delta_{h_i^5} = \delta_y w_i^6 \quad (22)$$

$$\delta_{h_i^c} = \sum_{k=1}^{28} \delta_{h_k^{c+1}} w_{i,k}^{c+1} \quad (23)$$

모든 레이어의 오차 신호가 계산되었다면, YCM은 그것을 사용하여 노드 사이의 가중치를 수정한다. YCM이 오차 신호를 사용하여 ANN의 연결 강도를 수정하면, 학습 데이터를 사용하여 학습 데이터를 사용한 ANN의 E(수식 19)가 일정 수준 이하로 떨어지면 학습을 종료한다.

4. 성능분석

4.1 AEOM의 성능분석

AEOM은 최적의 농장 환경을 계산하기 때문에, 실시간성보다 정확성이 더 중요하다. AEOM의 정확성을 검증하기 위하여, 본 논문은 학습되지 않은 입력 데이터 세트를 10개에서 50개까지 증가시키면서 AEOM, RF(Random Forest)와 GBD(Gradient Boosting Tree)의 정확도를 측정하였다. 실험은 일반 PC에서 진행되었으며, 실험 환경은 표 1에 나타낸다.

(표 1) 실험 환경
(Table 1) Experiment environment

분류	실험에 사용된 사양
CPU	intel i5-8500
RAM	8GB
GPU	nvidia GTX 1050 2GB
OS	Window 10 Edu 64bit
Framework	Torch

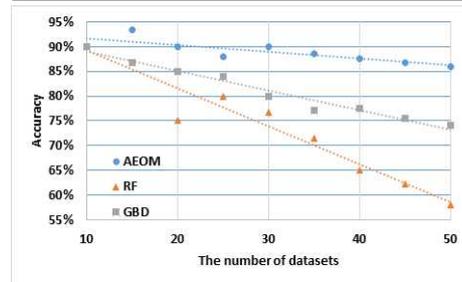


그림 3. AEOM, RF와 GBD의 정확도
Fig 3. Accuracy of AEOM, RF and GBD

그림 3은 AEOM, RF, GBD의 정확도를 나타낸다. 그림 3의 점은 데이터 세트의 개수마다 AEOM, RF, GBD의 정확도를 나타내고, 점선은 데이터 세트가 10개부터 50개까지 증가할 때 AEOM, RF, GBD의 정확도 감소 추이를 나타내는 추세선이다. AEOM의 정확도는 데이터 세트의 개수가 15개일 때 93.33%, 25개

일 때 88%, 35개일 때 88.57%, 50개일 때 86%로 측정되었다. AEOM의 정확도는 데이터 세트의 개수가 15개일 때 93.33%, 25개일 때 88%, 35개일 때 88.57%, 50개일 때 86%로 측정되었다. AEOM의 정확도는 데이터 세트의 개수가 15개일 때 93.33%, 25개일 때 88%, 35개일 때 88.57%, 50개일 때 86%로 측정되었다. AEOM의 평균 정확도는 88.90%이고 데이터 세트가 증가함에 따라 정확도 추세선의 기울기는 -0.0013이다. RF의 평균 정확도는 73.89%이고 데이터 세트가 증가함에 따라 정확도 추세선의 기울기는 -0.0077이다. GBD의 평균 정확도는 81.1%이고 데이터 세트가 증가함에 따라 정확도 추세선의 기울기는 -0.004이다. 즉, AEOM의 평균 정확도는 RF보다 약 15%, GBD보다 약 8% 높고, 데이터가 증가해도 RF나 GBD에 비해 정확도가 덜 감소한다. 따라서 AEOM은 스마트 팜의 농장 환경을 가장 잘 최적화할 수 있다.

4.2 CAIM의 성능분석

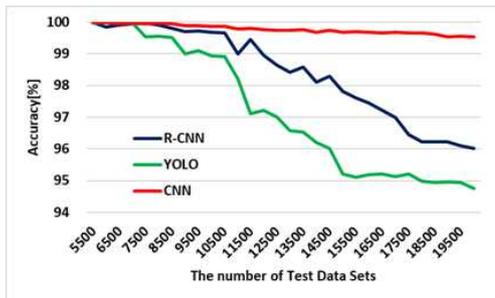


그림 4. CNN, R-CNN 및 YOLO의 정확도
Fig 4. Accuracy of convolutional neural network (CNN) compared with R-CNN and YOLO models.

이 논문에서 CAIM에 사용된 CNN은 작동 시간 및 정확도 면에서 R-CNN 및 YOLO와 비교된다. 잎 이미지와 질병 이름으로 구성된 5,000-20,000개의 테스트 데이터 세트를 사용하여 실험을 진행하였다. 실험 환경은 AEOM을 실험한 환경과 같다. 그림 4는 테스트 데이터 세트 수에 따른 CNN, R-CNN 및 YOLO의 정확도를 나타낸다. 실험 결과, CNN의 평균 정확도는 99.6 %, R-CNN의 평균 정확도는 96.1 %, YOLO의 평균 정확도는 94.2 %로 측정되었다. CNN의 평균 정확도는 R-CNN보다 약 3.5 % 더 높고 YOLO보다 약

5.4 % 더 높다.

4.3 YCM의 실험

본 논문에서는 YCM의 성능을 분석하기 위해 세 가지 실험을 수행한다. 먼저, YCM의 트레이닝 데이터 세트를 사용하여 five-fold cross-validation을 수행했다. YCM은 이미지의 90%를 트레이닝 데이터 세트로 사용하고 나머지 10%를 테스트 데이터 세트로 사용한다. 또한 트레이닝 데이터 세트를 나누는 기준인 K를 1에서 50으로 늘려 cross-validation의 정확도를 측정한다. 둘째, ReLU는 YCM에서 활성화 함수로 사용되며 Sigmoid, 계단 함수와 비교된다.

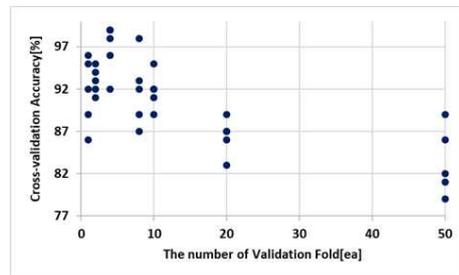


그림 5. YCM의 교차검증 정확도
Fig 5. Cross-validation accuracy of the YCM

그림 5에서 five-fold cross-validation에 사용되는 유효성 검사 횟수는 1 ~ 50이다. cross-validation의 평균 정확도는 fold가 1개일 경우 91.6%, fold가 2개일 경우 93%, fold가 4개일 경우 96.8%, fold가 8개일 경우 91.8%, fold가 20개일 경우 86.4%, fold가 50개일 경우 83.4%로 측정되었다. YCM은 4개의 validation folds를 사용하여 학습 속도를 최적화해야 한다.

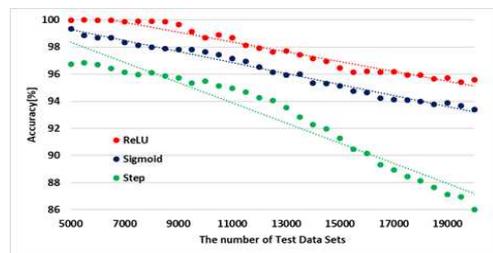


그림 6. 활성화 함수에 따른 YCM의 정확도
Fig 6. Accuracy of the YCM by activation function.

그림 6은 각 활성화 함수에 따른 YCM의 정확도를 나타낸다. YCM이 활성화 함수로 ReLU를 사용했을 때 평균 정확도는 97.8332%이고, Sigmoid를 사용했을 때 정확도는 96.2454%, 계단 함수를 사용했을 때는 92.759%로 측정되었다. 선형 회귀에 따르면 정확도의 기울기는 ReLU의 경우 $-3.641E-4$, Sigmoid의 경우 $-4.0710E-4$, 계단 함수의 경우 $-7.4534E-4$ 로 측정되었다. 따라서 ReLU는 다른 두 함수보다 높은 정확도를 가지면서 데이터가 늘어나도 두 함수들 보다 정확도가 감소되지 않는다.

5. 결론

본 논문에서는 현재 날씨, 토양 미생물 함량 등 농장에서 수집 가능한 모든 정보를 수집하여 작물이 잘 성장할 수 있도록 농장 환경을 최적화 하는 AEOM(Agricultural Environment Optimization Module)와 CNN을 생성하여 농작물의 질병을 진단할 수 있는 CIAM(Crop Image Analysis Module) 그리고 현재 기상데이터와 작물의 상태를 계산하여 작물 수확량을 예측하는 YCM(Yield Calculation Module)으로 구성되는 "IoT 및 딥 러닝 기반 스마트 팜을 위한 환경 최적화 및 수확량 예측 플랫폼"을 설계한다. 제안하는 플랫폼의 기대효과는 다음과 같다. 첫째, 제안하는 CIAM은 R-CNN보다 약 3.5%, YOLO보다 약 5.4프로 높은 정확도를 갖는다. 둘째, 제안하는 CYM은 ReLU를 활성화 함수로 사용하여 약 97%의 정확도를 갖는다. 셋째, 실시간적인 작물, 환경이 분석이 가능하다. 넷째, 작물의 관리뿐만 아니라 작물의 수확량을 예측하여 효과적인 스마트 팜 환경을 개선할 수 있다.

REFERENCES

- [1] Choi, Hwi-Min, Kim, Joo-Man, "Anomaly Detection System of Smart Farm ICT Device", The Institute of Internet, Broadcasting and Communication, Vol. 19, No. 2, pp.169-174, 2019.
- [2] Sung-Il Hwang, Jong-Moon Joo, Seong-Yong Joo "ICT-Based Smart Farm Factory Systems through the Case of Hydroponic Ginseng Plant Factory", JThe Journal of Korean Institute of Communications and Information Sciences, Vol 40, No.4, pp.780-790, 2015.
- [3] Sihyun Lee, "Fundamental Functionality Design of a Smart Farm Using an Embedded Computing Platform", Journal of the Institute of Electronics Engineers of Korea, Vol 55, No.4, pp.151-157, 2018.
- [4] Wonyoung An, YunHi Chang, "A Study on the Livestock Feed Measuring Sensor and Supply Management System Implementation based on the IoT.", kiiect, Vol. 10, No. 5, pp.442-454, 2017.
- [5] Yongju Park, Jun Moon, "Smart Dairy Management System Development Using Biometric/Environmental Sensors and Farm Control Gateway.", IEMEK Journal of Embedded Systems and Applications, Vol. 11, No. 1, pp.15-20, 2016.
- [6] Kim Young-hwan, Kim Tae-wook, Won Hyun-kyu, Lee Kyung-hak, Shin Man-yong, "Estimation of Timber Production by Thinning Scenarios Using a Forest Stand Yield Model.", Korean Society Of Forest Science, Vol. 101, No. 4, pp.592-598, 2012.
- [7] Jae Hun Sung, Dong Hun Lee, "Rice Yield Prediction Based on the Soil Chemical Properties Using Neural Network Model.", Korean Society for Agricultural Machin, Vol. 30, No. 6, pp.360-365, 2005.
- [8] Ju-Hee Kim, Seong-Soo Cheong, Ju-Rak Yim, Ki-Kwon Lee, "Yield Loss Assessment and Determination of Control Thresholds for Powdery Mildew of Chili pepper (Capsicum annum L).", The Korean society of pesticide science, Vol. 19, No. 2, pp.113-118, 2015.
- [9] Jung-won Oh, Haeng-gon Kim, Il-tae Kim, "Design and implementation of fruit harvest time predicting system based on machine learning.", Smart media journal, Vol. 8, No. 1, pp.74-71, 2019.
- [10] Yeong Mo Son, Jin Taek Kang, Jeong Sun Hwang, Park Hyun, Kang Su Lee, "Assessment and Prediction of Stand Yield in Cryptomeria japonica Stands.", Korean

Society Of Forest Science, Vol. 104, No. 3, pp.421-426, 2015.

저자약력

최 호 길(Ho-kil Choi) [정회원]



- 1977년 2월 : 한국항공대학교 항공관리학과(이학사)
- 1988년 1월 : 연세대학교 산업대학원 컴퓨터공학과 전자계산 전공(공학석사)
- 2019년 3월 ~ 현재 : 가톨릭관동대학교 일반대학원 컴퓨터공학과 박사과정
- 1980년 6월 : KCC정보통신(주) 근무
- 1999년 12월 : (주)강원랜드 정보전략실장
- 2018년 3월 ~ 현재 : 세인테크(주) 대표이사

<관심분야> IoT기반 스마트 팜, 스마트 공장, 스마트 시티, S/W 공학, S/W 개발

안 희 학(Heui-hak Ahn) [정회원]



- 1981년 2월 : 송실대학교 전자계산학과(공학사)
- 1983년 2월 : 송실대학교 전자계산학과(공학석사)
- 1994년 8월 : 송실대학교 전자계산학과(공학박사)
- 1984년 4월 ~ 현재 : 가톨릭관동대학교 소프트웨어학과 교수

<관심분야> 시스템소프트웨어, 프로그래밍 언어, 멀티미디어, 컴퓨터보안

정 이 나(Yi-na Jeong) [정회원]



- 2011년 2월 : 가톨릭관동대학교 컴퓨터학과(공학사)
- 2013년 8월 : 가톨릭관동대학교 컴퓨터학과(공학석사)
- 2018년 8월 : 가톨릭관동대학교 컴퓨터학과(공학박사)
- 2017년 3월 ~ 현재 : 가톨릭관동대학교 컴퓨터학과 초빙교수

<관심분야> 빅데이터, 네트워크, 프로그래밍 언어

이 병 관(Byung-Kwan Lee) [정회원]



- 1979년 2월 : 부산대학교 기계설계학과(공학사)
- 1986년 2월 : 중앙대학교 전자계산학과(이학석사)
- 1990년 2월 : 중앙대학교 전자계산학과(공학박사)
- 1988년 3월 ~ 현재 : 가톨릭관동대학교 소프트웨어학과 교수

<관심분야> 네트워크보안, 빅데이터, 데이터마이닝, 사물인터넷