

# 기상 데이터 기반 감귤 당도 예측 인공지능 모델 개발

## Development of Artificial Intelligence Model for Predicting Citrus Sugar Content based on Meteorological Data

서동민

한국과학기술정보연구원 연구데이터공유센터/과학기술연합대학원대학교 데이터&HPC과학

Dongmin Seo(dmseo@kisti.re.kr)

### 요약

감귤의 품질은 일반적으로 당도와 산도로 결정된다. 특히, 당도는 감귤의 맛을 결정하기 때문에 매우 중요한 요소이다. 현재 농가에서 가장 많이 사용하는 감귤 당도 측정 방법은 휴대용 착즙당도계 및 비파괴당도계로 측정하는 방식이다. 이 방법은 개인이 손쉽게 측정 가능한 방법이지만, 감귤농협 정식기계보다 당도수치의 정확성이 떨어지며 특히 0.5 Brix 이상 오차 차이가 발생해 현장에서 사용하기에는 아직 많이 부족하다. 또한, 현재 시점의 측정이지 예측 측정이 아니다. 그래서 본 논문에서는 기존 수집된 감귤 당도 값과 기상 데이터(평균 온도, 습도, 강우량, 일사량, 평균 풍속)을 기반으로 측정되지 않은 날짜의 감귤 당도를 0.5 Brix 이하 오차 범위 내에서 예측하는 AI 모델을 제안했다. 또한, 성능평가를 통해 제안하는 예측 모델이 제주 성산 지역에 대해서는 절대 평균 오차가 0.1154, 하원 지역에 대해서는 0.1983인 것을 확인했다. 마지막으로 제안한 모델은 0.5 Brix 이하 오차 차이를 지원하며, 예측 측정을 지원하는 기술이기에 그 활용성에 있어 진보성이 매우 높을 것으로 기대된다.

■ 중심어 : | 기상 데이터 | 감귤 | 당도 | 인공지능 | 예측 |

### Abstract

Citrus quality is generally determined by its sugar content and acidity. In particular, sugar content is a very important factor because it determines the taste of citrus. Currently, the most commonly used method of measuring citrus sugar content in farms is a portable juiced sugar meter and a non-destructive sugar meter. This method can be easily measured by individuals, but the accuracy of the sugar content is inferior to that of the citrus NongHyup official machine. In particular, there is an error difference of 0.5 Brix or more, which is still insufficient for use in the field. Therefore, in this paper, we propose an AI model that predicts the citrus sugar content of unmeasured days within the error range of 0.5 Brix or less based on the previously collected citrus sugar content and meteorological data (average temperature, humidity, rainfall, solar radiation, and average wind speed). In addition, it was confirmed that the prediction model proposed through performance evaluation had a mean absolute error of 0.1154 for Seongsan area and 0.1983 for the Hawon area in Jeju Island. Lastly, the proposed model supports an error difference of less than 0.5 Brix and is a technology that supports predictive measurement, so it is expected that its usability will be highly progressive.

■ keyword : | Meteorological Data | Citrus | Sugar Content | Artificial Intelligence | Prediction |

\* 본 연구는 중소벤처기업부 2020년도 산연 Collabo R&D사업의 「IoT·블록체인·AI 기반 클린 농산물 유통 관리 추적 플랫폼 개발(S2895529)」사업과 한국과학기술정보연구원 「연구데이터와 인프라의 공유·활용 체제 구축(K-21-L01-C04)」사업으로부터 지원받아 수행된 연구임.

접수일자 : 2021년 02월 26일

수정일자 : 2021년 03월 12일

심사완료일 : 2021년 03월 15일

교신저자 : 서동민, e-mail : dmseo@kisti.re.kr

## I. 서론

인공지능(Artificial Intelligence, AI)은 인간의 학습능력, 추론능력, 지각능력, 논증능력, 자연언어의 이해능력 등을 인공적으로 구현한 컴퓨터 프로그램 또는 이를 포함한 컴퓨터 시스템이다[1]. 최근 많은 해외 국가들은 AI를 새로운 국가 경쟁력으로 인식하고 AI를 기반으로 한 다양한 산업 활성화에 주력하고 있다. 일례로, 미국 정부는 미국의 AI 미래 준비, AI 국가 R&D 전략계획, AI 자동화와 경제를 기반으로 2019년 전 연방 차원의 AI 전략인 'AI 이니셔티브'를 발표했고, 연구 개발 투자와 AI 거버넌스 표준화 인력 및 기술보호를 강조하고 있다. 일본은 2019년 'AI 전략 2019 사람, 산업, 지역, 정부 모두의 AI'를 발표 및 4대 목표와 7대 전략 목표를 선정 후, 전 분야의 기술혁신과 더불어 경제·사회문제 해결에 주력하고 있다. EU는 2018년 'Communication on AI' 전략을 발표하고 AI를 기술·경제·사회·윤리 등 국가 전반의 혁신 기회로 인식하고 글로벌 AI 선두국가 도약을 위해 2022년까지 총 15억 유로 투자 계획을 수립했다. 중국은 2017년 민관협력 '차세대 AI 발전계획'을 수립 후, AI를 국가 전략산업으로 인식, 정부 주도의 대규모 투자, 인력양성을 추진하는 한편, 선도 기업 지정과 특허플랫폼을 육성하고 있다. 국내 과기정통부에서는 세계적 수준의 AI 기술력 및 R&D 생태계 확보를 위해 2018년 'I-Korea 4.0 실현을 위한 인공지능 R&D 전략'을 수립했다[2].

최근 AI 기술은 IT 기술뿐만 아니라 의료, 금융, 제조, 복지 등 다양한 분야에서 지능형 융합 서비스로 차별적 부가가치를 창출하는데 활용되고 있다. 일례로 뉴욕 맨해튼의 마운트 시나이 병원은 간, 직장, 전립선암 등의 질병발생을 94%의 정확도로 예측하는 딥러닝 기반의 AI 알고리즘을 이용하고 있다. 일본 요코하마은행은 고객의 재무정보, 거래 결제정보와 약 10만개의 회사들을 분석하고 이를 이용하여 고객의 수익성을 예측 후, 예측된 정보를 기반으로 대출 심사 및 금리를 결정하는 인공지능 기반 신용평가 시스템을 도입해 이용하고 있다. 미국 제너럴 모터스는 조립 로봇에 탑재된 카메라를 이용해 로봇 부품의 고장 사전 징후를 탐지하는 AI 기반의 장비 관리 솔루션을 이용하고 있다. 이 솔루

션의 테스트 결과 7천 개가 넘는 로봇 중에서 72건의 부품 고장 사례를 탐지했고 사전 감지를 통해서 작업다운 시간을 최소화했다. 중국 화웨이는 AI와 증강현실을 활용해 청각장애 아동들이 수화로 글을 번역해 읽는 것을 돕는 모바일 서비스를 만들었다.

그 외에도 AI가 농업 분야에 활용되면서 농업이 안고 있는 많은 난제를 해결하려는 다양한 사례들이 나타나고 있다. 미국 Harvest Automation은 AI 기술을 기반으로 인간은 할 수 없는 작물의 상태 분석 및 토양 분석을 통해 농원이나 온실 안을 자율적으로 이동하면서 물을 주거나 농약·제초제·비료 등을 살포할 수 있는 로봇을 개발했다[3][4]. 일본 파나소닉은 토마토의 색깔이나 모양, 장소를 특정하고 적절한 수확 시기를 판단하여 자동으로 수확하는 AI 기반 로봇을 개발했다. 파나소닉은 토마토 수확에 AI를 도입함으로써 기존에는 약 80%였던 수확률을 96%까지 향상시킬 수 있었다. 이처럼 농업 분야에서의 AI 활용은 일손이 부족한 농업 지원, 수확량 및 품질 향상, 생산 비용 절감 등 새로운 가치를 창출하고 있다.

국내에서 감귤은 제주도의 특산물인 감귤류를 통틀어 이르는 말로 현재 제주산 감귤은 주로 가공 식품보다는 과실로 소비되고 있으며, 영양학적으로 비타민, 식이섬유, 유기산 및 유리당의 공급원으로 이용되고 있다. 특히, 한방학적으로 감귤의 효능을 보면 위장장애, 천식, 가래, 식욕부진 및 동맥경화 등에 효과가 있는 것으로 알려져 있다. 최근 연구결과, 감귤 과피가 체내 콜레스테롤 수치를 감소시키고, 지방세포 분화를 억제시키는 효능이 있다고 밝혀져 체중감량에 좋은 과일로 알려지고 있다[5]. 또한, 제주특별자치도감귤출하연합회 보도 자료에 따르면, 제주 도내 감귤산업의 전체 시장 규모가 1조원을 웃도는 것으로 추산돼 지역경제에서 절대 비중을 차지하고 있는 것으로 분석됐다[6]. 이에 좋은 품질의 감귤 수확은 매우 중요하다.

감귤의 품질은 일반적으로 당도와 산도로 결정된다. 특히, '2018 식품소비행태조사 통계보고서'에 따르면 감귤은 한국인이 좋아하는 과일 4위로 '국민 과일'이지만, 신맛이 강하면 꺼리는 경우도 많고 같은 귤이라도 고령층의 경우는 새콤한 맛보다는 단맛이 강한 편을 선호한다고 설명하고 있다[7]. 그래서 감귤 재배 농가에서

는 수확 전까지 수차례 감귤의 당도를 측정한다.

감귤 당도 측정 방법은 크게 2가지 방법이 사용되고 있다. 첫 번째는 감귤농협의 농협기술센터에서 껍을 착즙해 정식기계로 측정하는 방식이다. 이 방법은 가장 정확한 측정 방법이지만, 파괴식 측정 방식이며 농가에서는 농협기술센터로 매번 직접 방문해야 하는 번거로움이 있다. 두 번째는 휴대용 착즙당도계 및 비파괴당도계로 측정하는 방식이다. 이 방법은 개인이 손쉽게 측정 가능한 방법이지만, 감귤농협 정식기계보다 당도 수치의 정확성이 떨어지며 특히 0.5 Brix(1 Brix는 물 100g에 당분이 1g 녹아 있는 것을 의미) 이상 오차 차이가 발생해 현장에서 사용하기에는 아직 많이 부족하다. 또한, 두 방법 모두 현재 시점의 측정이지 예측 측정이 아니다.

본 논문에서는 기존 수집된 감귤 당도 값과 기상 데이터(평균 온도, 습도, 강우량, 일사량, 평균 풍속)을 기반으로 측정되지 않은 날짜의 감귤 당도를 예측하는 AI 모델을 제안했다. 특히, 제안한 모델은 0.5 Brix 이하 오차 차이를 지원하며, 예측 측정을 지원하는 기술이기에 그 활용성에 있어 진보성이 매우 높을 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 기술과 관련된 것들에 대해 기술한다. 3장에서는 제안하는 AI 기반 감귤 당도 예측 모델에 대해 기술한다. 4장에서는 다양한 환경서의 성능평가를 통해 제안하는 예측 모델의 우수성을 입증한다. 마지막으로 5장에서는 결론 및 향후연구에 대해 기술한다.

## II. 관련 연구

### 1. 과수 생육·품질 관리시스템

농촌진흥청 국립원예특작과학원 과수 생육·품질 관리 시스템에서는 [그림 1]과 [그림 2]와 같이 제주 내 지역별 감귤 당도와 기상 데이터를 제공하는 웹페이지를 서비스하고 있다[8]. 본 연구에서는 이 시스템에 대한 웹 크롤링을 통해 감귤 당도 예측 모델의 학습 데이터와 테스트 데이터로 사용할 기존 감귤 당도 데이터와 평균 온도, 습도, 강우량, 일사량, 평균 풍속 데이터를 획득했다.

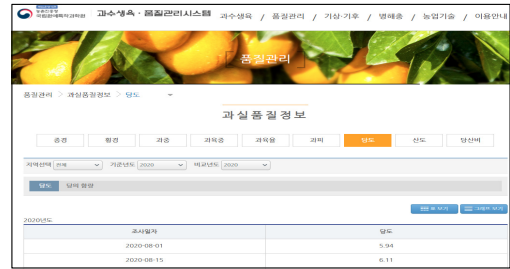


그림 1. 감귤 당도 데이터를 제공하는 과실품질정보 웹페이지

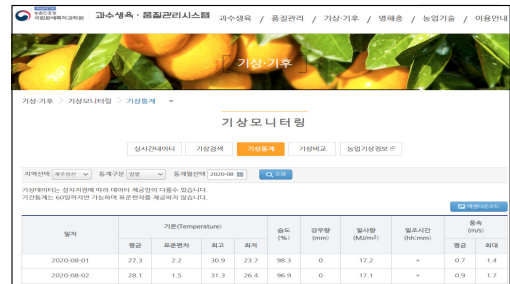


그림 2. 기상 데이터를 제공하는 기상모니터링 웹페이지

### 2. 과일 분야 인공지능 기술

과일의 재배와 수확에 있어서 과실 상태를 미리 판단하는 것은 수확량 예측에 도움이 될 뿐만 아니라, 재배 현장 관리에도 큰 도움이 된다. [9]에서는 복숭아 영상 이미지 분석을 통해 미성숙 복숭아를 판별하는 인공지능 경망 모델을 개발했다. 자연 상태에서는 여러 광조건 때문에 성숙 과실과 미성숙 과실을 구분하기가 상당히 어려운 측면이 있는데, 이 연구에서 개발한 인공신경망은 85%의 정확도를 제공한다. 물론, 다소 낮은 정확도를 보이고 있지만, 이 연구는 최초로 자연광 이미지를 통해 미성숙 복숭아를 선별하는 알고리즘을 제안했다는 점에서 연구적인 의미가 있다. [10]에서는 복숭아를 촬영한 동영상 이미지에 대하여 영역기반 합성곱 신경망(Region Convolutional Neural Network)을 기반으로 한 YOLOv3 알고리즘을 적용하여 복숭아 표피 분석을 통한 불량품 분류 인공지능 모델을 개발했다. 이 연구는 농작물들 중에서도 외상에 가장 취약한 복숭아를 대상으로 분석모델을 제안했다는 점에서 연구적인 의미가 있다. [11]에서는 가시광선-근적외선 스펙트럼을 이용해 배의 당도를 예측하는 인공신경망 모델을 개발했다. 이 연구에서는 대역폭을 조절한 분광자료 입

력 값을 설정하고 배의 당도를 출력 값으로 설정한 후 역전파 알고리즘을 통해 학습시킴으로써 높은 수준의 예측 성능을 보일 수 있음을 검증했다[12].

### 3. 온도 기반 감귤 당도 예측 기술

출원번호 10-2018-0166842 특허에서는 온도 데이터를 이용한 감귤 당도 예측 기술을 제안하고 있다. 이 특허에서 제안하는 기술은 크게 데이터 수집부, 당도 검량선 모델링부, 당도 예측부로 분류된다. 데이터 수집부에서는 감귤의 착과 시점으로부터 소정 시간 간격마다 측정된 감귤의 당도 값과 해당 시점의 온도 데이터를 기상청으로부터 수집해 데이터베이스에 저장한다. 당도 검량선 모델링부에서는 데이터베이스에 저장된 당도 데이터와 온도 데이터를 기반으로 [그림 3]과 같은 당도 검량선을 모델링한다.

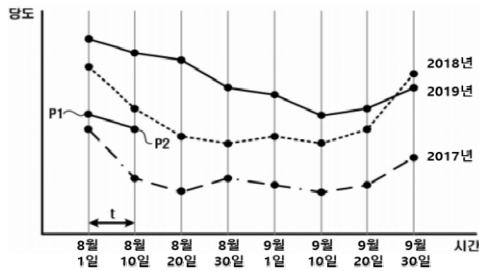


그림 3. 출원번호 10-2018-0166842 특허에서 제안하는 당도 검량선

당도 예측부에서는 모델링된 당도 검량선을 이용해서 앞으로의 감귤 당도를 예측한다. 일례로, 2020년 8월 1일 측정된 당도 값 P1을 획득 후, P1과 가장 가까운 검량선(2017년)을 기준으로 이후 당도 값 P2를 예측한다. 단, 기존 수집 온도 데이터를 가지고 이후 날짜들에 대한 평균 온도를 획득하고 기준 검량선의 온도 차에 대한 당도 값을 보정해 예측 값을 보정한다. 단, 특허에는 온도 차에 대한 당도 값을 어떻게 보정하는지에 대해서는 명확히 설명되어 있지 않다. 결론적으로, 본 특허에서 제안하는 기술은 AI 기반 기술이 아니며, 기준 당도 검량선을 선택하기 위해서 예측을 원하는 날짜의 당도 값이 요구되고 감귤의 당도에 영향을 주는 다양한 기상 데이터가 있음에도 온도만을 사용하고 있

어서 예측 정확도를 높일 수 있는 여지가 있다.

## III. 제안하는 AI 기반 감귤 당도 예측 모델

### 1. 시스템 구성도

본 논문에서는 기존 수집된 감귤 당도 값과 기상 데이터인 평균 온도, 습도, 강우량, 일사량, 평균 풍속을 기반으로 측정되지 않은 날짜의 감귤 당도를 예측하는 AI 예측 모델을 개발했다. 특히, 개발한 예측 모델은 실제 감귤의 당도와 예측된 감귤의 당도 오차가 0.5 Brix 이하인 예측 정확도를 제공한다. [그림 4]는 제안하는 예측 모델에 대한 시스템 구성도를 보여준다. 제안하는 시스템은 웹 크롤러, AI 데이터 생성기, AI 기반 감귤 당도 예측 모델로 분류된다.

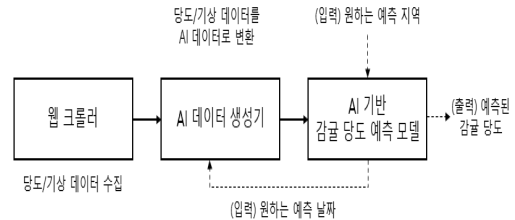


그림 4. 제안하는 예측 모델의 시스템 구성도

### 2. 웹 크롤러

웹 크롤러는 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하고 원하는 데이터를 수집하는 컴퓨터 프로그램이다. 검색 엔진과 같은 여러 사이트에서는 데이터의 최신 상태 유지를 위해 웹 크롤러를 사용한다. 본 논문에서 개발한 웹 크롤러는 Python의 Selenium library[13]를 기반으로 개발된 웹 크롤러이며, 농촌진흥청 국립원예특작과학원 과수 생육·품질 관리시스템에서 제공하는 감귤 당도 데이터를 제주 지역 별, 특정 조사일자 별 그리고 월 상반기·하반기 별로 최신 데이터를 실시간 수집해서 엑셀 파일로 저장하는 기능을 제공한다. 특히, 개발한 웹 크롤러는 기존 URL 기반의 웹 페이지 전환이 아닌, Selenium을 이용해 가상 사용자 액션(Virtual User Action)을 기반으로 한 웹 페이지 전환 기능을 제공함으로써 Javascript로 작성된 웹 페

이지 내의 데이터 수집 기능을 제공한다.

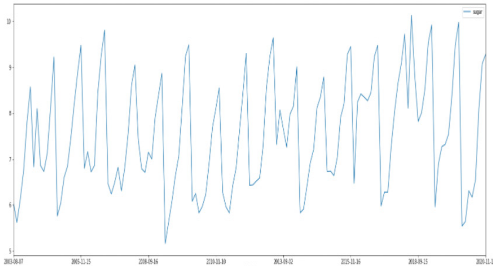


그림 5. 성산 지역, 특정 조사일자 별 당도 데이터 시각화

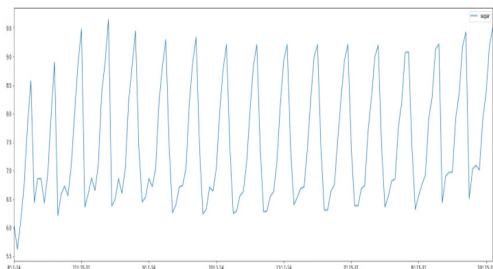


그림 6. 성산 지역, 월 상반기·하반기 별 당도 데이터 시각화

[그림 5]는 웹 크롤러를 통해 수집된 성산 지역에 대한 특정 조사일자 별 측정된 당도 데이터를 시각화한 것이다. [그림 6]은 성산 지역에 대한 월 상반기·하반기 별 측정된 당도 데이터를 시각화한 것이다. [그림 5]와 [그림 6]을 보면 패턴을 가지고 당도 데이터가 변화하는 것을 확인할 수 있다. 즉, 통계적 예측이 가능하다는 것을 확인할 수 있다.

### 3. AI 데이터 생성기

AI 개발에 있어서 이루고 싶은 목적에 맞춰 가능한 많은 정보를 모았다면 이를 AI 모델에서 사용할 수 있게 잘 정제된 데이터 형식으로 바꿔야 한다. 이 작업은 모든 AI 개발에 있어서 첫 단추이자 가장 중요한 작업이다. 본 논문에서는 웹 크롤러를 통해 농촌진흥청 국립원예특작과학원 과수 생육·품질 관리시스템에서 수집된 감귤 관련 당도 데이터와 평균기온, 습도, 강우량, 일사량, 평균풍속 데이터를 AI 기반 감귤 당도 예측 모델에서 사용할 수 있는 데이터 형식으로 바꿔주는 AI 데이터 생성기를 개발했다. 농촌진흥청 국립원예특작과

학원 과수 생육·품질 관리시스템 내 농작업 일정 자료에 따르면 제주도는 감귤의 열매가 만들어지는 시점인 개화기가 일반적으로 5월부터 시작되며, 당도( $y_t$ )는 매년 7월 15일을 시작으로 11월 15일을 마지막으로 제공한다. 이에 본 연구에서는 5월 1일부터 11월 14일까지의 총 198일의 기상 데이터( $j$ )와 측정된 당도 데이터를 가지고, AI 기반 감귤 당도 예측 모델에서 사용할 비용 함수(cost function)를 수식 1과 같이 제안했다. 수식 1에서  $t$ 는 학습용 데이터로 사용되는  $y_t$ 의 총 수를 나타내며,  $x_1$ 은 평균기온,  $x_2$ 는 습도,  $x_3$ 은 강우량,  $x_4$ 는 일사량,  $x_5$ 는 평균풍속을 나타낸다. 그리고 제안하는 비용 함수에서 사용할 학습용 데이터를 [표 1]과 같이 제안했다[14].

$$H(x_1, x_2, x_3, x_4, x_5) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5,$$

$$cost(W) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n (H(x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij}, x_{5ij})) - y_i \quad (1)$$

[표 1]에서 기상 데이터 컬럼은 5월 1일을 시작으로 당도가 측정되기 전일까지의 수집된 기상 데이터와 '0.0'로 구성된다. 지역명은 특정 지역에 대한 당도 예측을 위한 구분자로만 활용된다. 당도 측정일은 각 row 데이터의 당도가 언제 측정된 것인지를 확인하기 위한 구분자로만 활용된다. 즉, AI 기반 감귤 당도 예측 모델은 [표 1]에서 기상 데이터와 당도 데이터만을 사용한다. 또한, 예측을 원하는 특정 일이 있으면, 해당 년도 5월 1일부터 해당 일 전날까지의 기상 데이터를 입력으로 사용하면 된다. 만약, 측정된 기상 데이터가 없다면 기존 수집된 데이터의 평균 데이터를 사용하면 된다.

### 4. AI 기반 감귤 당도 예측 모델

본 논문에서 개발한 AI 기반 감귤 당도 예측 모델은 198개의 기상 데이터와 1개의 당도 데이터를 가지고 예측하는 선형 회귀 모델이다. 즉, 개발한 예측 모델은 수식 1의 cost를 최소화하는 가중치를 찾는 것이다. 본 논문의 예측 모델은 Python 3.7 환경에서 tensorflow 2.4.0의 keras 2.4.3을 가지고 개발했다. 특히, 최적의 cost를 찾기 위한 방법으로 손실(loss) 함수는 평균 제

표 1. 제안하는 AI 학습용 데이터 생산 지역 예시

지역명	당도 측정일	기상 데이터															당도	
		5월 1일					5월 2일					...	11월 14일					
		평균 기온	습도	강우량	일사량	평균 풍속	평균 기온	습도	강우량	일사량	평균 풍속	...	평균 기온	습도	강우량	일사량		평균 풍속
성산	2017-07-15	19.2	99.3	0.0	16.3	0.3	18.3	99.7	0.0	9.9	0.3	...	0.0	0.0	0.0	0.0	0.0	7.42
성산	2017-08-01	19.2	99.3	0.0	16.3	0.3	99.7	99.7	0.0	9.9	0.3	...	0.0	0.0	0.0	0.0	0.0	6.36
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
성산	2017-11-15	19.2	99.3	0.0	16.3	0.3	99.7	99.7	0.0	9.9	0.3	...	13.9	62.6	0.0	13.9	0.4	9.08

곱 오차(Mean Square Error, MSE)를 최적화(optimizer)는 수식 2의 Adam[15]을 가지고 개발했다.

$$\begin{aligned}
 M(t) &= \beta_1 M(t-1) + (1-\beta_1) \frac{\partial}{\partial w(t)} Cost(w(t)) \\
 V(t) &= \beta_2 V(t-1) + (1-\beta_2) \left( \frac{\partial}{\partial w(t)} Cost(w(t)) \right)^2 \\
 \hat{M}(t) &= \frac{M(t)}{1-\beta_1^t} \quad \hat{V}(t) = \frac{V(t)}{1-\beta_2^t} \\
 W(t+1) &= W(t) - \alpha * \frac{\hat{M}(t)}{\sqrt{\hat{V}(t) + \epsilon}}
 \end{aligned}
 \tag{2}$$

#### IV. 성능 평가

농촌진흥청 국립원예특작과학원 과수 생육·품질 관리 시스템에서는 2017년부터의 기상 데이터만을 제공하고 있기에 2017년부터 2020년까지의 기상 데이터와 당도 데이터만을 가지고 개발한 예측 모델의 성능 평가를 수행했다. 성능 평가는 [표 2]와 같이 다양한 환경에서 다양한 모델(hidden layer 노드 수, hidden layer 깊이)을 구축 후 테스트에 사용된 모델의 손실(Loss)과 절대 평균 오차(Mean Absolute Error, MAE)를 통해 성능을 측정했다. 참고로 총 990개의 입력 변수가 사용되기에 첫 hidden layer의 노드 수는 입력 변수 총 개수의 1/2 근사 값인 500개로 설정 했고, 이후부터는 평가에 의미 있게 변경했다.

표 2. AI 기반 감귤 당도 예측 모델의 성능 평가 방법

구분	설명
평가 1	성산 지역 2017년부터 2019년까지의 전체 기상 데이터와 당도 데이터로 학습 후 2020년 데이터로 테스트 수행
평가 2	평가 1에서 가장 좋은 모델로 온도 데이터와 당도 데이터만을 가지고 평가 1의 테스트 수행
평가 3	성산 지역과 하원 지역 2017년부터 2019년까지의 전체 데이터로 학습 후 성산과 하원 지역 2020년 데이터로 테스트 수행
평가 4	하원 지역 2017년부터 2019년까지의 전체 기상 데이터와 당도 데이터로 학습 후 2020년 데이터로 테스트 수행

[표 3]은 평가 1에서 hidden layer 노드 수를 변경하면서 측정된 성능 평가 결과를 보여준다. 첫 hidden layer의 노드 수가 4,000일 때 가장 좋은 성능을 보인다. [그림 7]은 [표 3]에서 첫 hidden layer의 노드 수가 4,000일 때 당도 예측 결과를 보여준다.

표 3. 평가 1에서 hidden layer 노드 수 변경에 대한 평가 결과

Relu 모델 구성	Loss	MAE
500 - 250 - 124 - 1	4.1255	1.7351
1000 - 500 - 250 - 1	0.2539	0.3541
2000 - 1000 - 500 - 1	0.5271	0.5872
3000 - 1500 - 750 - 1	0.2031	0.4087
<b>4000 - 2000 - 1000 - 1</b>	<b>0.0682</b>	<b>0.2188</b>
5000 - 2500 - 1250 - 1	0.1251	0.3181
6000 - 3000 - 1500 - 1	0.2068	0.4359

검사날짜: 2020-08-01, 실제당도: 6.510, 예상당도: 6.783
검사날짜: 2020-08-15, 실제당도: 7.030, 예상당도: 7.003
검사날짜: 2020-09-01, 실제당도: 7.090, 예상당도: 7.113
검사날짜: 2020-09-15, 실제당도: 7.010, 예상당도: 7.456
검사날짜: 2020-10-01, 실제당도: 7.890, 예상당도: 8.209
검사날짜: 2020-10-15, 실제당도: 8.360, 예상당도: 8.658
검사날짜: 2020-11-01, 실제당도: 9.190, 예상당도: 9.456
검사날짜: 2020-11-15, 실제당도: 9.520, 예상당도: 9.619

그림 7. [표 3]의 당도 예측 결과

[표 4]는 평가 1에서 hidden layer 깊이를 변경하면서 측정된 성능 평가 결과를 보여준다. 앞의 실에서 첫 hidden layer의 노드 수가 4,000일 가장 좋은 성능을 보이기 때문에 기준 첫 hidden layer의 노드 수를 4,000으로 설정했다. Hidden layer 깊이가 6일 때 가장 좋은 성능을 보인다. [그림 8]은 [표 4]에서 hidden layer의 깊이가 6일 때 당도 예측 결과를 보여준다.

표 4. 평가 1에서 hidden layer 깊이 변경에 대한 평가 결과

Relu 모델 구성	Loss	MAE
4000 - 1	5.9710	2.0011
4000 - 2000 - 1	0.2608	0.4561
4000 - 2000 - 1000 - 1	0.0682	0.2188
4000 - 2000 - 1000 - 500 - 1	0.1207	0.2974
4000 - 2000 - 1000 - 500 - 250 - 1	2.4094	1.5159
<b>4000 - 2000 - 1000 - 500 - 250 - 125 - 1</b>	<b>0.0148</b>	<b>0.1154</b>
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 1	0.0160	0.1073
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 1	0.0283	0.1485
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 15 - 1	0.0204	0.1033

검사날짜: 2020-08-01, 실제당도: 6.510, 예상당도: 6.613
검사날짜: 2020-08-15, 실제당도: 7.030, 예상당도: 6.855
검사날짜: 2020-09-01, 실제당도: 7.090, 예상당도: 6.987
검사날짜: 2020-09-15, 실제당도: 7.010, 예상당도: 7.135
검사날짜: 2020-10-01, 실제당도: 7.890, 예상당도: 7.946
검사날짜: 2020-10-15, 실제당도: 8.360, 예상당도: 8.425
검사날짜: 2020-11-01, 실제당도: 9.190, 예상당도: 9.039
검사날짜: 2020-11-15, 실제당도: 9.520, 예상당도: 9.375

그림 8. [표 4]의 당도 예측 결과

즉, 평가 1의 평가 결과를 통해 성산 지역은 hidden layer의 노드 수가 4,000이고 hidden layer의 깊이가 6일 때 가장 좋은 예측 성능(Loss: 0.0148, MAE: 0.1154)을 보이는 것을 확인했다.

[표 5]는 평가 2를 평가 1에서 가장 좋은 성능을 보여준 hidden layer의 깊이가 6인 모델을 가지고 온도 데이터와 당도 데이터만을 가지고 측정된 성능 평가 결과를 보여준다. 온도 데이터와 당도 데이터만을 사용할 경우 입력 변수의 총 수는 199개이다. 평가 1에서 입력 변수의 4배가 되는 첫 hidden layer의 노드 수가 가장 좋은 성능을 보였다. 이에 평가 2의 첫 hidden layer의 노드 수를 800개로 설정했다. [그림 9]는 [표 5]의 당도 예측 결과를 보여준다.

표 5. 평가 2의 평가 결과

Relu 모델 구성	Loss	MAE
800 - 400 - 200 - 100 - 50 - 25 - 1	<b>0.2846</b>	<b>0.5089</b>

검사날짜: 2020-08-01, 실제당도: 6.510, 예상당도: 6.092
검사날짜: 2020-08-15, 실제당도: 7.030, 예상당도: 6.365
검사날짜: 2020-09-01, 실제당도: 7.090, 예상당도: 6.624
검사날짜: 2020-09-15, 실제당도: 7.010, 예상당도: 6.759
검사날짜: 2020-10-01, 실제당도: 7.890, 예상당도: 7.542
검사날짜: 2020-10-15, 실제당도: 8.360, 예상당도: 7.798
검사날짜: 2020-11-01, 실제당도: 9.190, 예상당도: 8.596
검사날짜: 2020-11-15, 실제당도: 9.520, 예상당도: 8.752

그림 9. [표 5]의 당도 예측 결과

평가 1과 평가 2를 통해 온도 데이터만 활용하기 보다는 다른 기상 데이터를 사용할 때 예측 모델의 정확도가 향상되는 것을 확인했다.

[표 6]은 평가 3에서 첫 hidden layer 노드 수를 4,000으로 하고 hidden layer 깊이를 변경하면서 측정된 성능 평가 결과를 보여준다. Hidden layer 깊이가 8일 때 가장 좋은 성능을 보인다. 하지만, 평가 1의 결과와 비교하면 예측 결과의 정확도가 떨어지는 것을 확인할 수 있다. 이를 통해, 전체 지역을 예측하는 한 개의 모델 사용보다는 각 지역별 최적화된 모델 사용이 필요하다는 것을 확인했다. [그림 10]은 [표 6]에서 hidden layer의 깊이가 8일 때 당도 예측 결과를 보여준다.

표 6. 평가 3에서 hidden layer 깊이 변경에 대한 평가 결과

Relu 모델 구성	Loss	MAE
4000 - 1	15.8298	3.4649
4000 - 2000 - 1	0.5527	0.6100
4000 - 2000 - 1000 - 1	0.2226	0.3780
4000 - 2000 - 1000 - 500 - 1	0.2903	0.4564
4000 - 2000 - 1000 - 500 - 250 - 1	0.3236	0.4239
4000 - 2000 - 1000 - 500 - 250 - 125 - 1	0.2413	0.4238
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 1	0.3552	0.4906
<b>4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 1</b>	<b>0.2144</b>	<b>0.4076</b>
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 15 - 1	0.2242	0.4145

(성산) 검사날짜: 2020-08-01, 실제당도: 6.510, 예상당도: 7.064
(성산) 검사날짜: 2020-08-15, 실제당도: 7.030, 예상당도: 7.282
(성산) 검사날짜: 2020-09-01, 실제당도: 7.090, 예상당도: 7.534
(성산) 검사날짜: 2020-09-15, 실제당도: 7.010, 예상당도: 7.489
(성산) 검사날짜: 2020-10-01, 실제당도: 7.890, 예상당도: 8.536
(성산) 검사날짜: 2020-10-15, 실제당도: 8.360, 예상당도: 8.961
(성산) 검사날짜: 2020-11-01, 실제당도: 9.190, 예상당도: 9.781
(성산) 검사날짜: 2020-11-15, 실제당도: 9.520, 예상당도: 9.650
(하원) 검사날짜: 2020-08-01, 실제당도: 6.920, 예상당도: 7.025
(하원) 검사날짜: 2020-08-15, 실제당도: 7.140, 예상당도: 7.108
(하원) 검사날짜: 2020-09-01, 실제당도: 7.570, 예상당도: 7.370
(하원) 검사날짜: 2020-09-15, 실제당도: 7.610, 예상당도: 7.220
(하원) 검사날짜: 2020-10-01, 실제당도: 8.580, 예상당도: 8.218
(하원) 검사날짜: 2020-10-15, 실제당도: 9.070, 예상당도: 8.620
(하원) 검사날짜: 2020-11-01, 실제당도: 9.770, 예상당도: 9.378
(하원) 검사날짜: 2020-11-15, 실제당도: 10.140, 예상당도: 9.247

그림 10. [표 6]의 당도 예측 결과

표 7. 평가 4에서 hidden layer 깊이 변경에 대한 평가 결과

Relu 모델 구성	Loss	MAE
4000 - 1	17.9837	3.6237
4000 - 2000 - 1	0.3591	0.5682
4000 - 2000 - 1000 - 1	0.2914	0.4947
4000 - 2000 - 1000 - 500 - 1	0.0592	0.1853
<b>4000 - 2000 - 1000 - 500 - 250 - 1</b>	<b>0.0526</b>	<b>0.1983</b>
4000 - 2000 - 1000 - 500 - 250 - 125 - 1	0.0610	0.2219
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 1	0.1073	0.2903
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 1	0.3360	0.5319
4000 - 2000 - 1000 - 500 - 250 - 125 - 62 - 31 - 15 - 1	0.1474	0.3670

검사날짜: 2020-08-01, 실제당도: 6.920, 예상당도: 7.240
검사날짜: 2020-08-15, 실제당도: 7.140, 예상당도: 7.121
검사날짜: 2020-09-01, 실제당도: 7.570, 예상당도: 7.716
검사날짜: 2020-09-15, 실제당도: 7.610, 예상당도: 7.955
검사날짜: 2020-10-01, 실제당도: 8.580, 예상당도: 8.759
검사날짜: 2020-10-15, 실제당도: 9.070, 예상당도: 9.281
검사날짜: 2020-11-01, 실제당도: 9.770, 예상당도: 9.823
검사날짜: 2020-11-15, 실제당도: 10.140, 예상당도: 9.826

그림 11. [표 7]의 당도 예측 결과

[표 7]은 평가 4에서 첫 hidden layer 노드 수를 4,000으로 하고 hidden layer 깊이를 변경하면서 측정된 성능 평가 결과를 보여준다. Hidden layer 깊이가 5일 때 가장 좋은 성능을 보인다. 평가 1, 평가 3, 평가 4를 통해 각 지역별 예측 모델을 구성하는 것이 가장 좋은 성능을 도출하는 것을 확인했다. [그림 11]은 [표 7]에서 hidden layer의 깊이가 5일 때 당도 예측 결과를 보여준다. 즉, 평가 4의 평가 결과를 통해 하원 지역은 hidden layer의 노드 수가 4,000이고 hidden layer의 깊이가 5일 때 가장 좋은 예측 성능(Loss:

0.0526, MAE: 0.1983)을 보이는 것을 확인했다.

## V. 결론 및 향후 연구

본 논문에서는 기존 수집된 감귤 당도 값과 기상 데이터인 평균 온도, 습도, 강수량, 일사량, 평균 풍속을 기반으로 측정되지 않은 날짜의 감귤 당도를 예측하는 AI 예측 모델을 개발했다. 또한, 다양한 성능 평가를 통해 각 지역별 예측 모델을 구성하는 것이 가장 좋은 성능을 도출하는 것을 확인했고, 개발한 예측 모델이 실제 감귤의 당도와 예측된 감귤의 당도 오차가 0.5 Brix 이하인 예측 정확도를 제공하는 것을 확인했다. 향후 연구로는 기상 데이터 이외의 CO<sub>2</sub>와 같이 공기 중에 포함된 요소 및 토질 데이터를 사용한 당도 예측 모델 개발을 계획하고 있고, 감귤의 상품성을 결정하는 당도 외에 산도를 예측할 수 있는 모델 개발을 계획하고 있다. 또한, 개발한 당도 예측 모델을 기반으로 최상품 감귤을 재배할 수 있도록 감귤 재배에 도움을 주는 스마트팜 서비스 개발을 계획하고 있다.

## 참고 문헌

- [1] <https://ko.wikipedia.org/wiki/인공지능>
- [2] 이정우, “인공지능(빅데이터) 시장 및 기술 동향,” S&T Market Report, Vol.71, pp.1-26, 2019.
- [3] <https://www.public.harvestai.com/>
- [4] 서동민, “감귤 상품성 예측 인공지능 모델 개발,” 한국콘텐츠학회 종합학술대회 논문집, 2021.
- [5] <https://www.nihhs.go.kr/usr/main/mainPage.do>
- [6] <http://www.citrus.or.kr/>
- [7] 김창길, 2018 *식품소비행태조사 통계보고서*, 한국농촌경제연구원, 2018.
- [8] <https://fruit.nihhs.go.kr/>
- [9] Kurtulmus, Ferhat, W. S. Lee, and Ali Vardar, “Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network,” J. of Precision Agriculture, Vol.15, No.1, pp.57-79, 2014.



[10] 이희준, 이원석, 최인혁, 이충권, “YOLOv3을 이용한 과일표피 불량검출 모델: 복숭아 사례,” J. of Information Systems Review, Vol.22, No.1, pp.113-124, 2020.

[11] 정결, 최영수, “VIS/NIR 스펙트럼을 이용한 배 당도 예측 인공지능경망 개발,” 한국농업기계학회 2014 추계 학술대회 논문집, Vol.19, No.2, pp.226-227, 2014.

[12] 장인국, 김세한, 박현, 김명은, 은지숙, *스마트팜을 위한 인공지능경망 기술 및 발전 방향*, 네이버 블로그, 2017.

[13] <https://selenium-python.readthedocs.io/>

[14] <https://dataon.kisti.re.kr/>, DOI: 10.22711/0101LB017494156611.0

[15] <https://twinw.tistory.com/247>

저 자 소 개

서 동 민(Dongmin Seo)

정회원



- 2002년 : 충북대학교 정보통신공학과(공학사)
  - 2004년 : 충북대학교 정보통신공학과(공학석사)
  - 2008년 : 충북대학교 정보통신공학과(공학박사)
  - 2010년 ~ 현재 : 한국과학기술정보연구원 연구데이터공유센터 센터장
  - 2015년 ~ 현재 : 과학기술연합대학원대학교 데이터 & HPC과 학 교수
- 〈관심분야〉 : 빅데이터 분석, XML, 이동객체 데이터베이스, 센서 네트워크, 네트워크 분석, 인공지능