

http://dx.doi.org/10.17703/JCCT.2021.7.4.779

JCCT 2021-11-95

파프리카 생산성 추정을 위한 선형 회귀모형 개발 시 외부광량 활용 적합성을 높이기 위한 방법

How to Improve Suitability of Irradiation Utilization in Development of Linear Regression Model for Estimating Paprika Productivity

우승미*, 김가영**, 김호철***

Seung Mi Woo*, Ga Yeong Kim**, Ho Cheol Kim***

요약 파프리카 생산성에는 광량이 매우 중요한 요인으로 작용하나 광량을 독립변인으로 한 파프리카 생산성 추정을 위한 표준 모델 개발에 있어 어려움을 겪고 있다. 본 연구는 이러한 표준 모델을 개발할 시 독립변인으로서 외부 광량의 활용 적합성을 높이기 위한 방법을 알아보기 위해서 수행되었다. 선형회귀 분석 시 독립변인(누적 외부광량)과 종속변인(누적 생산성)을 전체 농가 생산성의 평균값, 그리고 그 평균값을 기준으로 이상과 미만(MHFP, MLFP)으로 분류하여 각각 평균값을 활용하였다. 도출된 선형회귀모형의 RMSE 값은 MHFP의 모델에서 $0.9418\text{kg}\cdot\text{m}^{-2}$ 로 MTFP 모델의 $1.5468\text{kg}\cdot\text{m}^{-2}$, MLFP 모델의 $1.3812\text{kg}\cdot\text{m}^{-2}$ 보다 크게 낮았다. 그리고 시간(월)이 경과함에 따른 RMSE 값도 MHFP에서는 모든 월에 $1.0\text{kg}\cdot\text{m}^{-2}$ 이하로 가장 낮았다. 따라서 외부 광량을 활용한 파프리카 생산성 추정을 위한 회귀모형 개발 시 농가들의 생산성 차이를 적합한 방법으로 분류하여 분석하는 것이 추정 모델의 적합성을 향상시킬 것으로 판단된다.

주요어 : 외부광량, 생산성, 선형회귀모형, 환경요인

Abstract The amount of sunlight (irradiation) acts as a very important factor for paprika (*Capsicum annuum* L.) productivity, but there are difficulties in developing a standard model for estimating paprika productivity using irradiation factors. This study was conducted to investigate how to increase the suitability of using irradiation as an independent variable when developing a standard model. In the linear regression analysis using the independent variable (cumulative irradiation) and the dependent variable (cumulative productivity) were classified as the average value of the total farm productivity (MTFP), and above and below (MHFP, MLFP) based on the average value, respectively. The RMSE value of the estimated linear regression model was $0.9418\text{kg}\cdot\text{m}^{-2}$ in the MHFP, which was significantly lower than $1.5468\text{kg}\cdot\text{m}^{-2}$ in the MTFP and $1.3812\text{kg}\cdot\text{m}^{-2}$ in the MLFP. And in due course of time (month), RMSE value was also the lowest in MHFP, below $1.0\text{kg}\cdot\text{m}^{-2}$ in all months. Therefore, when developing a regression model for estimating paprika productivity using irradiation, it is judged that it will improve the suitability of the estimation model by classifying and analyzing the difference in productivity of farms with an appropriate method.

Key words : Irradiation, Productivity, Linear regression model, Environmental factors

*준회원, 원광대학교 원예산업학부 학부생 (제1저자)
**준회원, 원광대학교 원예학과 석사과정 (참여저자)
***정회원, 원광대학교 원예산업학부 조교수 (교신저자)
접수일: 2021년 10월 18일, 수정완료일: 2021년 10월 27일
게재확정일: 2021년 11월 2일

Received: October 18, 2021 / Revised: October 27, 2021

Accepted: November 2, 2021

*Corresponding Author: go-hc@hanmail.net

Divi. of Horticulture Industry, Wonkwang Univ, Korea

I. 서론

국내 신선농산물 수출 작물 가운데 파프리카는 매우 큰 비중을 차지하고 있고, 일본을 중심으로 대부분 수출되고 있다[1]. 파프리카의 생산성 추정 모델 개발은 아직까지 많은 농가의 추정 모델로 대표성을 갖지 못하고 있다. 최근 모든 산업 분야에서 대두되고 있는 빅데이터, AI 기술의 도입되고 있는데[2][3] 파프리카 산업에서도 활용되고 있지만 아직 다양한 문제로 어려움이 많은 실정이다. 파프리카 생산성에는 광합성의 필수 요소인 광량이 매우 중요하여 외부 광량 데이터를 생산성을 예측하기 위한 독립변인으로 활용 가치가 매우 높고 중요하다[4]. 대부분 온실은 외부 광량에 의존적이어서 내부의 수광량을 정확히 예측하기 위한 광 추적 시뮬레이션도 연구되었다[5]. 하지만 아직까지 현장에서 외부 광량을 통한 생산성 추정 연구들 결과가 정확도가 높게 활용되지 못하고 있는 실정이다. 따라서 본 연구는 파프리카 생산성 추정을 위한 회귀 모형 개발 시 독립변인으로서 외부 광량의 활용 적합성을 높이기 위한 방법을 탐색하고자 수행되었다.

II. 재료 및 방법

1. 환경 및 생산량 데이터 수집

전북 지역에서 파프리카를 재배하고 11개 농가를 무작위로 선정한 후 농가에 축적된 2020년 작기의 외부 일평균 광량 데이터와 매 출하된 박스(10kg) 수 데이터를 수집하였다. 그리고 농가의 온실 면적을 수집하였다.

2. 환경 및 생산량 데이터 가공

일별 생산성(kg·m⁻²)은 출하 박스 수×10kg×농가면적(m²)로 하여 계산되었다. 그리고 이를 월별 합으로 하여 월별 생산성으로 변환하였고, 월별 생산성을 시간의 경과함에 따른 누적 생산성으로 변환하였다. 또한 외부의 일별 광량도 시간의 경과함에 따른 누적 광량으로 변환하였다.

3. 생산성에 따른 농가 그룹핑

생산성 차이에 따른 그룹핑은 최종 전체 생산성을 평균하여 도출된 15.2kg·m⁻²을 기준으로 초과 농가, 이하 농가로 하였다. 이에 따라 초과 농가는 7개 농가, 이하

농가는 4개로 구분되었다(그림 1).

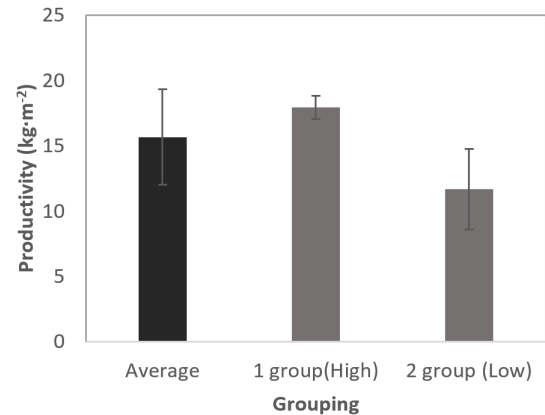


그림 1. 파프리카 재배 농가의 생산성 차이에 따른 그룹핑.
Figure 1. Grouping according to difference in productivity of paprika farms

4. 광량과 생산성 회귀 분석 및 적합성

회귀분석은 선형회귀분석으로 수행하였으며, 독립변인을 누적 광량, 종속변인을 누적 생산성으로 하였다. 독립 및 종속 변인을 각 그룹별 평균값으로 한 회귀모형 그룹내 각 농가를 적용하여 평균값 회귀모형에 대한 분포정도를 알아보았다. 적합성 검정은 평균제곱근오차 (root mean square error, RMSE)로 검정하였다. RMSE 산출식은 아래와 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

각 그룹별 도출된 회귀모형에 대한 RMSE와 함께 시간의 경과함에 따른 월별 RMSE값을 구하여 재배기간의 후반부로 갈수록 회귀모형의 적합성도 알아보았다.

III. 결과 및 고찰

1. 파프리카 재배 11개 농가의 생산성

파프리카 재배 11개 농가의 월별 평균 생산량을 조사하였다(그림 2). 월별 생산량의 추세는 1월-2월에 각각 1.0과 1.2kg·m⁻²로 낮았고, 4-5월에는 1-2월의 약 2배인 2.1과 2.6kg·m⁻²로 높았다. 농가간의 편차는 0.34-0.78kg·m⁻² 범위로 생산성이 가장 낮은 1-2월에

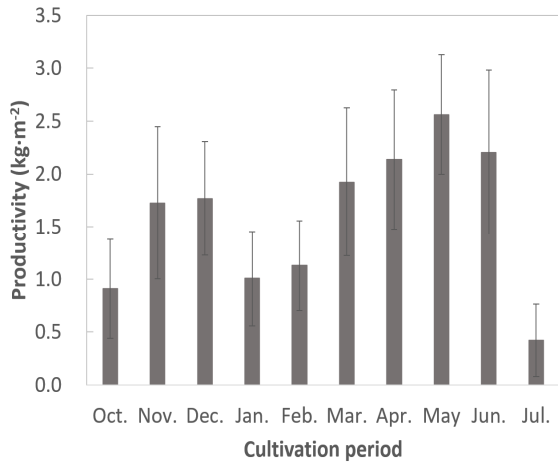


그림 2. 파프리카 재배 11개 농가의 월별 평균 생산성.
 Figure 2. Monthly mean productivity of 11 paprika farms.
 bars are standard deviation

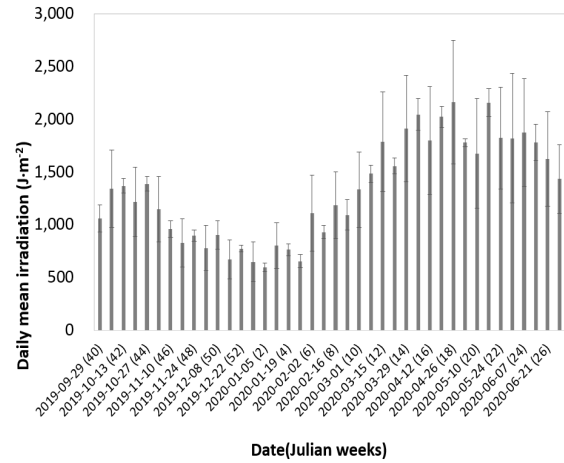


그림 3. 파프리카 재배 11개 농가의 월별 평균 광량.
 Figure 3. Monthly mean irradiation of 11 paprika farms.
 bars are standard deviation.

가장 낮았다. 즉, 1-2월의 생산성 편차가 낮은 것은 모든 농가에서 생산성이 가장 낮은 뿐만 아니라 농가 간 편차가 작아 이때의 특정 환경요인이 매우 큰 영향을 준 것으로 판단된다.

파프리카 재배 11개 농가의 월별 일평균 광량을 조사하였다(그림 3). 전체적인 변화 패턴이 그림 2와 매우 유사하였다. 따라서 1-2월에 생산성 저하는 광량에 의해 발생한 것으로 판단되고, 1-2월 농가간 광량 편차도 낮은 경향으로 대부분의 농가에서 이 시기에 광량이 매우 적고 이로 인해 생산성에 크게 영향을 준 것으로 판단된다. 하지만 파프리카의 화아 형성 후 생산까지의 약 2개월 소요 기간을 고려하면 1-2월 생산성이 낮은 원인은 11-12월 동안 지속적으로 저하되는 광량이 화아의 형성 수 및 충실도에 영향을 주었고, 남아 있는 꽃 수정 후 과실 비대 시기에도 낮은 광량으로 광합성 동화산물 량이 적어 과실 크기는 다소 평균과중을 나타내나 착과수가 적어 생산성이 낮은 것으로 판단된다.

그림 2와 3의 결과로 본 조사 농가는 생산성에 광량의 영향이 있는 것으로 판단되지만, 광량이 충분한 시에는 이외의 지상부 및 지하부 요인이 작용할 수 있다는 것을 고려해야 할 것으로 판단된다.

우리나라는 사계절이 뚜렷하고, 10월 이후 일조시간이 짧아질 뿐만 아니라 이에 따라 광량이 부족해진다. 이러한 광량 부족은 착과율을 저하시켜 생산성도 저하시키는 것으로 보고되었다[4]. 이러한 겨울철 광량 부족을 해결하기 위하여 형광등(fluorescent lamp, FL), 메탈할라이드(metal halide lamp, MH), 수은등(mercury

lamp, MC), 할로겐 램프(halogen lamp, HL), 백열등(incandescent lamp, IL), 고압나트륨등(high pressure sodium lamp, HPS) 등 보광등에 대한 연구들도 진행되었다[6].

2. 외부 누적광량을 이용한 누적생산성 분석

외부 누적광량(X)에 따른 누적생산성(Y)에 대한 선형 회귀분석을 실시하였다(그림 4). 회귀모형은 $Y = 0.0389X - 0.5688$ ($R^2=0.9989$)으로 나타났다. 누적값을 활용한 관계로 결정계수(R^2)은 높은 값을 나타내었지만, RMSE는 $1.5468\text{kg}\cdot\text{m}^{-2}$ 로 다소 높아 전체 농가의 평균값을 활용한 회귀모형의 적합도는 다소 낮은 것으로 판단된다.

그림 4에 따라 추정된 각 월별 생산량(y')와 실제 월별 생산성(y) 차이($y-y'$)는 월별 모두 음의 값을 나타내어 y'값은 y값보다 $1.11-1.97\text{kg}\cdot\text{m}^{-2}$ 까지 높게 추정되었다. 이런 차이 값은 실제 생산성(y)의 15.19-32.43% 범위로 매우 컸고 이로 RMSE가 $1.5468\text{kg}\cdot\text{m}^{-2}$ 로 다소 큰 값을 나타낸 것으로 판단된다(자료 미제시).

전체 농가의 평균 생산성과 광량을 이용한 Figure 4의 회귀모형의 적합성이 낮아 이를 다시 전체 평균 생산성 값을 기준으로 상위 7개 농가(1 그룹, high farms), 하위 4개 농가(2 그룹, low farms)로 그룹핑을 한 후 각각 평균값을 활용하여 다시 회귀모형을 도출하였다(그림 5).

그룹별 누적광량(X)에 따른 누적생산성(Y)의 선형회귀식은 1그룹에서 $Y = 0.0429X - 0.4882$ ($R^2 = 0.9988$),

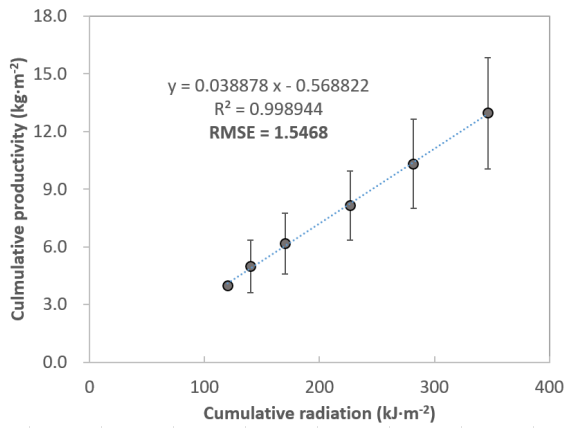


그림 4. 파프리카 재배 11개 농가의 평균값을 활용한 누적 광량과 누적 생산성 간 회귀모형.
Figure 4. Linear regression between cumulative irradiation and productivity using mean values of 11 paprika farms. Bars are standard error

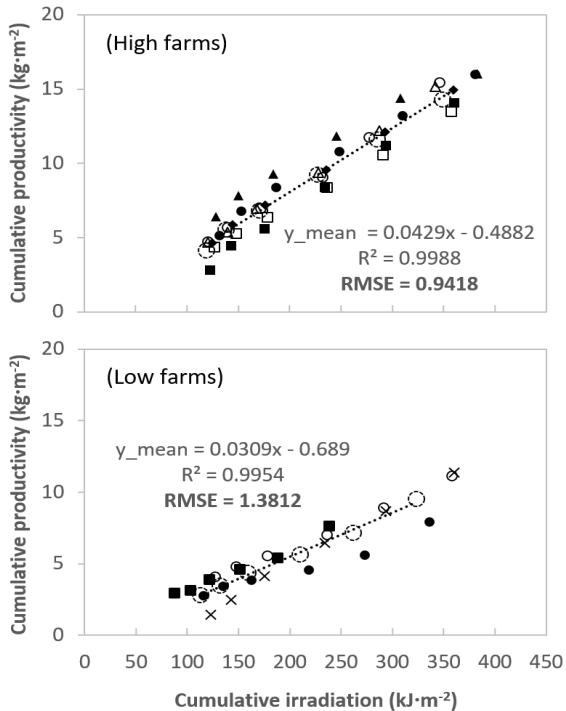


그림 5. 생산성에 따른 파프리카 재배 농가의 그룹별 평균값 활용 회귀모형(○)과 모형에 대한 각 농가의 생산성 분포.
Figure 5. Linear regression model using mean value for each group according to productivity of paprika, and distribution in productivity of each farm for the each model

2그룹에서는 $Y = 0.0309X - 0.6890$ ($R^2 = 0.9954$)로 도출되었다. 단위 누적광량 당 누적생산성은 1그룹에서 2그룹의 1.39배로 매우 높았다. 그리고 각각 평균값을 활용한 회귀선에 각각의 농가별 분포(그림 5의 작은 점들)도

1그룹에서 대체로 가까웠다. 이에 따라 RMSE값도 각각 0.9418과 1.3812로 나타났다. 따라서 그림 4와 5의 결과로 보았을 때 전체 농가의 평균값보다 생산성 고저에 따른 그룹핑 후 평균값을 활용하는 것이 적합할 것으로 보고, 이중에서도 전체 평균보다 생산성이 높은 농가들의 회귀모형만이 실제 활용하기에 적합할 것으로 판단된다.

3. 파프리카 생산성 추정 회귀모형의 적합성

그림 6에서 시간(재배기간)이 경과함에 따라서 월별로 회귀모형을 만들 시 전체 평균값을 활용한 회귀모형의 RMSE가 $1.76\text{kg}\cdot\text{m}^{-2}$ 에서 $2.90\text{kg}\cdot\text{m}^{-2}$ 으로 지속적으로 높아져 적합성이 더욱 낮아졌다. 생산성이 낮은 2그룹에서는 초기에 $1.49\text{kg}\cdot\text{m}^{-2}$ 에서 중반부 $0.96\text{kg}\cdot\text{m}^{-2}$ 으로 낮아지다가 다시 $1.68\text{kg}\cdot\text{m}^{-2}$ 까지 지속적인 증가추세를 보였다. 하지만 생산성이 높은 1그룹에서는 $1.0\text{kg}\cdot\text{m}^{-2}$ 이하를 지속적으로 유지하여 적합성도 3 방법 중 가장 높고 시간이 경과함에 따라서도 적합성이 유지가 되었다.

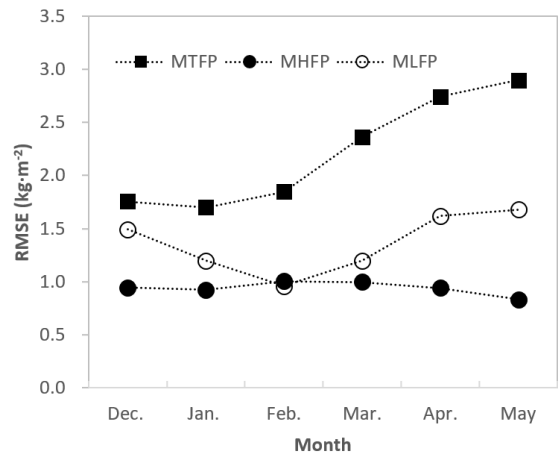


그림 6. 누적 광량을 이용한 파프리카 생산성 추정 시 시간의 경과함에 따른 선형회귀모형의 RMSE 변화. MTFP, MHFP, 및 MLFP: 각각 총 농가들, 생산성이 높은 농가들, 생산성이 낮은 농가들.

Figure 6. Change in RMSE of linear regression in estimation of cumulative productivity of paprika using cumulative irradiation of greenhouse over time. MTFP, MHFP, and MLFP: total farms, farms with high, and low productivity, respectively.

토마토에서 생산량 예측을 위한 방법으로 다중선형 회귀, 랜덤 포레스트, 딥러닝(ConvLSTM) 알고리즘을 비교하였을 시 딥러닝 방법에서 결정계수(R^2)가 가장

높았다는 연구가 있으나[7] 아직까지 현장에서 적용되지 못한 것을 고려하면 누적값을 활용한 회귀분석뿐만 아니라 다양한 방법을 활용한 비교 연구가 더욱 필요해 보인다.

모든 결과를 종합하여 볼 때 많은 파프리카 농가를 대상으로 외부 누적광량을 활용하여 생산성을 예측할 수 있는 표준 모형을 개발하고자 할 시 생산성을 기준으로 고저에 따라 분류하여 모형을 도출하는 것이 효율적이다. 특히, 생산성이 높은 농가를 위한 표준 회귀모형 개발에는 생산성 예측을 위한 독립변인으로 외부 광량을 누적하여 활용하여도 적합할 것으로 판단된다. 그러나 생산성이 낮은 농가들에서는 새로운 독립변인을 선별하거나 농가별 회귀모형을 도출 및 활용하는 것이 효율적일 것으로 판단된다.

IV. 결 론

본 연구는 파프리카 생산성을 예측할 시 외부광량의 활용에 대한 적합성을 검토하고자 수행되었다. 선형회귀 분석 시 독립변인(누적 외부광량)과 종속변인(누적 생산성)을 총 농가 평균값(MTFP), 그리고 전체 농가의 평균 생산성 기준으로 이상과 이하(MHFP, MLFP) 그룹으로 분류하여 각각 평균값을 활용하였다. 도출된 선형회귀 모형의 RMSE 값은 MTFP에서 $1.5468\text{kg}\cdot\text{m}^{-2}$, MHFP에서 $0.9418\text{kg}\cdot\text{m}^{-2}$, 그리고 MLFP에서 $1.3812\text{kg}\cdot\text{m}^{-2}$ 로 나타났다. 시간이 경과함에 따라 도출된 월별 선형회귀 모형의 RMSE 값은 MTFP와 MLFP에서 각각 $1.5\text{-}3.0\text{kg}\cdot\text{m}^{-2}$ 와 $1.0\text{-}1.7\text{kg}\cdot\text{m}^{-2}$ 범위를 나타내었고, 후반부로 갈수록 값이 커졌다. 하지만 MHFP에서는 모든 월에 $1.0\text{kg}\cdot\text{m}^{-2}$ 이하였으며, 후반부로 갈수록 다소 낮아지는 경향이였다. 따라서 생산성 예측을 위한 표준 회귀모델 도출 시 독립변인으로서 외부광량의 활용 적합도는 생산성이 높은 농가 그룹에서 높은 것으로 판단 된다.

References

- [1] Korea Agro-Fisheries and Food Trade Corporation (KAFFTC) "Kati 'export and import information of paprika," (in Korean) <https://www.kati.net/product/basisInfo.do?lcdCode=MD147>, 2020
- [2] J.Y. Lee, K.D. Jung, K.H. Ryu, and S.Y. Park,

"A Study on the Analysis of Regional Tourism in Uijeongbu Using Big Data," J. Convergence on Culture Technology, Vol. 6, No. 1, pp. 413-418, DOI: 10.17703/JCCT.2020.6.1.413. doi.org/10.17703/JCCT.2020.6.1.413

- [3] S.W. Lee and M.Y. Han, "Utilization and Analysis of Big-data," International Journal of Advanced Culture Technology, Vol.7, No.4, pp.255-259, 2019. doi.org/10.17703/IJACT.2019.7.4.255
- [4] W.J. Jeong, D.J. Myoung, and J.H. Lee, "Comparison of climatic conditions of sweet pepper's greenhouse between Korea and Netherlands," J. Bio-Environ. Control, Vol. 18, No. 3, pp. 244-252, 2009.
- [5] W.H. Kang, I.H. Hwang, D.H. Jung, D.P. Kim, J.W. Kim, J.H. Kim, K.S. Park, and J.E. Son, "Time change in spatial distributions of light interception and photosynthetic rate of paprika estimated by ray-tracing simulation," Protected Hort. Plant Factory, Vol. 28, No. 4, pp. 279-285, 2019.
- [6] J.W. Lee, H.C. Kim, P.H. Jeong, Y.G. Ku, and J.H. Bae, "Effects of supplemental lighting of high pressure sodium and lighting emitting plasma on growth and productivity of paprika during low radiation period of winter season," Kor. J. Hort. Sci. Technol., Vol. 32, No. 3, pp. 346-352, 2014. doi.org/10.7235/hort.2014.14029
- [7] S.E. Song, T.J. Park, J.I. Bang, and H.J. Kim. "A study on the prediction model for tomato production and growth using ConvLSTM," J. KIIT, Vol. 18, No. 1, pp. 1-10, 2020. doi.org/10.14801/jkiit.2020.18.1.1

※ 본 연구는 2021년 농림축산식품부 연구사업
(과제번호: 320101-03-2-SB020) 지원
에 의해 수행되었음.