RESEARCH ARTICLE https://doi.org/10.7235/HORT.20220051

# LSTM 네트워크를 이용한 한라봉 가격 예측 모델

정대호[1] · 조영열[2,3*]

[1]연암대학교 스마트원예계열, [2]제주대학교 아열대농업생명과학연구소, 친환경연구소, [3]제주대학교 원예환경전공

# A Prediction Model for Hallabong Tangor Product Prices using LSTM (Long Short-term Memory) Network

Dae Ho Jung[1] and Young-Yeol Cho[2,3*]

[1]Division of Smart Horticulture, Yonam College, Cheonan 31005, Korea
[2]Research Institute for Subtropical Agriculture and Animal Biotechnology, SARI, Jeju National University, Jeju 63243, Korea
[3]Horticultural Science Major, Jeju National University, Jeju 63243, Korea

*Corresponding author: ycho@jejunu.ac.kr

## Abstract

The predicted prices of agricultural products have a great influence on market supply and demand control decisions. The purpose of this study was to predict the average price of Hallabong tangor, which is distributed in 9 wholesale markets nationwide, using a deep learning-based LSTM network. Daily data (2,398 data points) from 2007 to January 2021 based on the average price of 3 kg of Hallabong tangor sold in 9 wholesale markets were used in this study. To train the model, 80% of the dataset was used as training data and 20% was used as test data. For accurate price predictions using an LSTM network, it is necessary to determine the epoch and batch size. An epoch of 700 and batch size of 128 were judged to be suitable for the model (RMSE = 1040). The resulting model could accurately predict the average price of Hallabong tangor. The deep learning-based Hallabong tangor price prediction model provides useful information for agricultural product price trends, government policy establishment, and consumer demands.

**Additional key words:** artificial intelligence, deep learning, long-term dependency, recurrent neural network

**Received:** February 3, 2022
**Revised:** April 7, 2022
**Accepted:** June 9, 2022

OPEN ACCESS

HORTICULTURAL SCIENCE and TECHNOLOGY
40(5):571-577, 2022
URL: http://www.hst-j.org

pISSN : 1226-8763
eISSN : 2465-8588

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright©2022 Korean Society for Horticultural Science.

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음(IITP-2022-2020-0-01489).

## 서 언

RNN(recurrent neural network)는 출력값의 일부가 입력값에 다시 포함되는 연속성 개념의 신경회로망을 의미한다(Shin et al., 2018). 이 신경망은 시계열 데이터를 처리하기 위해 제안되었지만, 장기간에 걸쳐 발생되는 패턴을 인식하지 못하는 문제점이 발견되었다. 이러한 문제점을 해결하기 위한 방법으로 LSTM(long short-term memory) 네트워크가 개발되었다. LSTM는 RNN의 한 구조로, 가중치뿐 아니라 메모리에 대한 추가 정보를 셀 상태에 저장하고 시계열 패턴의 길이도 조정할 수 있다(Hochreiter and

Schmidhuber, 1997; Olah, 2015; Shin et al., 2018). LSTM은 여러 분야의 문제를 굉장히 잘 해결했고, 지금도 널리 사용되고 있다(Olah, 2015). 특히, 실내 환경을 예측하는 방법으로 LSTM 예측 모델이 선형 회귀보다 700%, RNN 모델보다는 140% 성능 개선되었다(Kim and Oh, 2018).

만감류인 한라봉은 부지화로 불리는데, 일본 과수연구소 감귤부 구찌노쯔에서 1972년 청견에 폰칸을 교배해서 육성된 품종으로 우리나라에서는 1990년 초부터 재배되기 시작하여 매년 재배면적과 생산량이 증가되고 있다(Han and Kang, 2011). 재배면적과 생산량 증가로 과잉 생산체계가 도래하면서 가격 하락 변동이 큰 경향을 보여주고 있다(Ko, 2017). 한라봉 수급 및 가격전망을 위해 통상최소자승법을 이용한 수급모델을 개발하였다(Ko and Kim, 2012). 배추와 양파, 마늘, 무를 대상으로 가락시장에서 월별 가격 예측에 대한 ARIMA 모형 연구(Kim, 2005)와 청과물 14품목의 가격 계절성을 고려하여 가락시장의 평균가격을 GARCH(generalized autoregressive conditional heteroscedasticity)모형과 VAR(vector autoregression) 모형을 이용하여 예측한 연구(Lee et al., 2006), 자기회기시차 모형으로 양파의 재배면적과 단수를 예측하고, 양파 출하시기 도매가격을 예측한 연구(Nam and Choe, 2015), LSTM 네트워크를 활용하여 쌀 양파, 대파, 애호박, 시금치 등의 가격(Shin et al., 2018), 사과 과일 가격(Im et al., 2018), 양배추, 토마토와 상추 가격(Kurumatani, 2020), 양배추, 콜리플라워와 Brinjal 가격(Banerjee et al., 2022)과 온실내의 환경 변화 추정을 예측(Lee et al., 2020)한 연구 등이 있다. ARIMA는 시계열 알고리즘에서 가격 예측에 가장 널리 사용되는 선형 모델로 비선형 패턴을 포착하지 못하는 단점이 있는데, 이러한 단점을 보완할 수 있는 방법이 LSTM 모델이다(Prakash and Sumaiya Farzana, 2019).

좋은 예측 모델을 얻으려면 LSTM 네트워크의 파라미터들의 값 설정이 중요하다. Epoch는 학습의 반복을 의미하며 가장 뛰어난 예측 모델을 만든 시점에서 중지해야 한다. 그 이유는 컴퓨팅 리소스가 제한적이고 학습 시간에 대한 제약 때문이다. 모델에 의한 예측 가격과 실제 가격의 차이를 통해 모델의 성능을 가늠해 볼 수 있다. 두 값의 오차를 측정하는 가장 보편적인 평가도구는 RMSE(root mean squared error)이다.

본 연구는 한라봉 농산물을 대상으로 9개 도매시장에서 거래되고 있는 평균가격을 바탕으로, LSTM 네트워크 기법을 이용하여 학습한 후 2달 동안의 평균가격을 예측해 보았다.

## 재료 및 방법

한라봉 가격 경매가는 제주특별자치도 감귤출하연합회 자료를 이용하였으며, 모델 학습을 위해 2007년 10월 31일부터 2021년 1월 31일까지의 기간에 2,398개의 데이터(훈련/검증/테스트)를 사용하였다. 경매가 이루어지지 않는 날은 입력 데이터에서 삭제하였다. 2,398개의 데이터 중 80%가 훈련/검증 테스트셋으로 사용하였다. 그리고 학습한 학습 모델을 기반으로 하여 추후 2021년 2월 동안(28일간)의 한라봉 평균 가격을 예측해 보았다. 입력 데이터의 값은 9개 도매시장의 평균가격으로, 서울 가락, 부산 엄궁, 대구 북부, 인천 남촌, 광주 각화, 대전 오정, 경기 구리, 서울 강서와 경기 수원 도매시장이었다. 각 상자 규격(소포장, 1kg, 3kg, 5kg, 10kg 등)은 모두 3kg으로 환산한 가격을 이용하였다.

LSTM의 첫 단계로는 cell state로부터 어떤 정보를 버릴 것인지를 정하는 것으로, sigmoid layer에 의해 결정된다(Olah, 2015). 이 단계를 forget gate layer(Fig. 1B, Eq. 1)라고 한다. 다음 단계는 앞으로 들어오는 새로운 정보 중 어떤 것을 cell state에 저장할 지를 결정하는 것으로 input gate layer(Fig. 1C, Eq. 2-1, 2)라 하며, sigmoid layer가 어떤 값을 업데이트할 것인지 정한다(Eq. 2-1). 그 다음에 tanh layer가 새로운 후보 값들인 vector를 만들고(Eq. 2-2), cell state에 더할 준비를 한다. 이렇게 두 단계에서 나온 정보를 합쳐서 state를 업데이트할 재료를 만든다. 다은 단계는 과거의 cell state를 업데이트해서 새로운 cell state를 만드는 단계이다(Fig. 1D, Eq. 3). 마지막 단계는 무엇을 output으로 내보낼 지를 정하는 것이다. 먼저, sigmoid layer에 input 데이터를 태워서 cell state의 어느 부분을 output으로 내보낼 지를 정한다(Fig. 1E, Eq. 4). 그리고 나서 cell state를 tanh layer에 태워서 −1과 1사이의 값을 받은 뒤에 방금 전에 계산한 sigmoid gate의 output과 곱해준다(Eq. 5). 그 결과 우리

가 output으로 보내고자 하는 부분만 내보낼 수 있게 된다.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2-1}$$

$$\widetilde{C}_t = \tan h(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{2-2}$$

$$C_t = f_t \cdot C_{t-1} + i_i \cdot \widetilde{C}_t \tag{3}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

$$h_t = O_t \cdot \tan h(C_t) \tag{5}$$

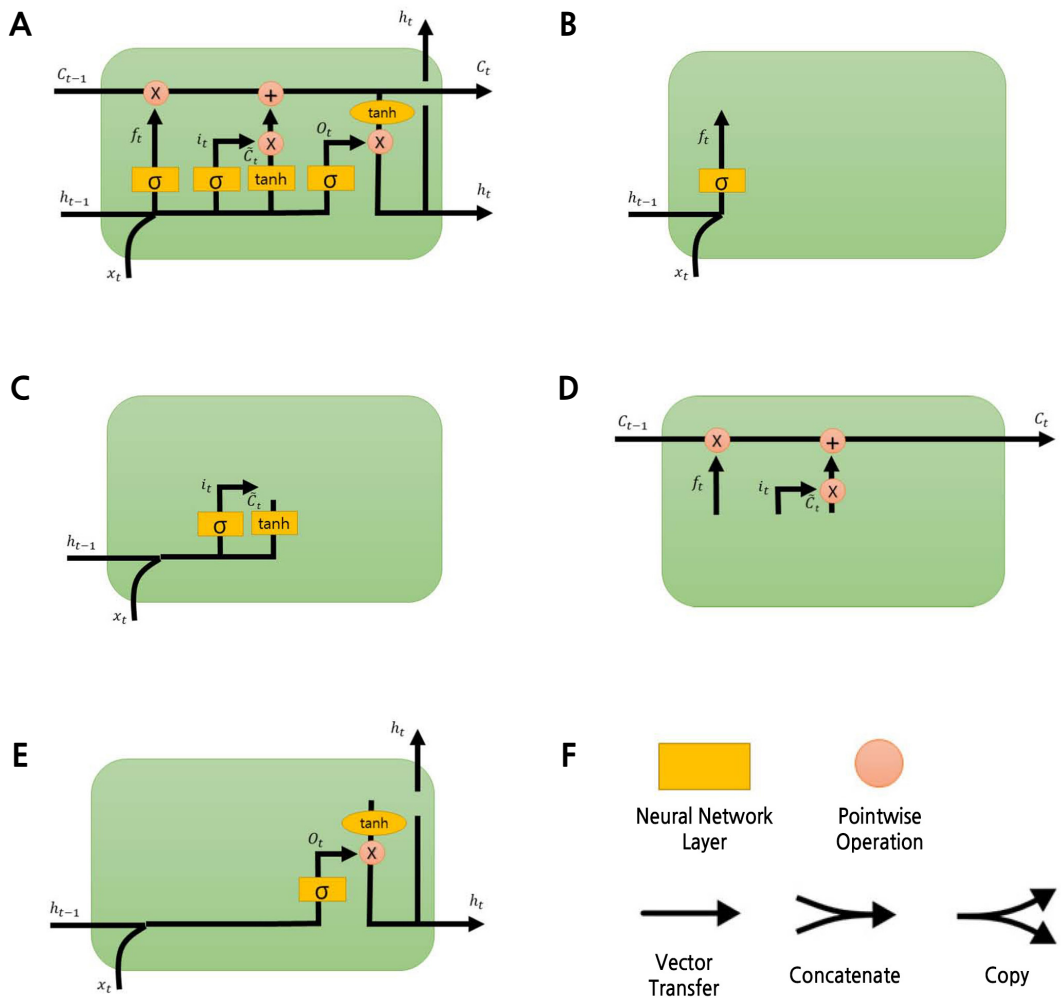


**Fig. 1.** The repeating module in an LSTM contains four interacting layers (Olah, 2015). (A) LSTM architecture, (B) Forget gate layer, (C) Input gate layer, (D) Cell state, (E) Output gate layer, and (F) Legend.

일반적인 정규화 식을 통해 모든 학습 데이터에 대하여 0 – 1 사이의 값으로 변환해서 정규화된 값으로 학습과 테스트를 진행하였다. LSTM 네트워크 기법을 사용하여 최적의 epoch와 batch size를 구했다. 최적화하는 방법으로는 일반적으로 많이 사용되는 아담최적화(AdamOtimizer)를 이용해서 추정값과 실측값의 제곱근 평균오차(RMSE)를 최소화하는 방법을 수행하였다. 실험에 사용한 언어는 Python 언어로 작성하여 PyCharm을 사용하였으며, 사용된 라이브러리로는 numpy, pandas와 keras 등 이었다.

## 결과 및 고찰

9개 도매시장의 한라봉 가격을 살펴보면, 3kg 기준 최대 평균 55,000원(2013년 10월 3일), 최소 평균 4,100원(2010년 6월 17일)으로 평균 13,370원 정도 형성되고 있었다(Fig. 2). 9개 도매시장에서 가장 높게 받는 가격은 80,000원(2009년 9월 30일)이었으며, 가장 낮게 받는 가격은 1,000원(2016년 12월 22일)이었다. 대략 년중 5월에서 10월까지는 한라봉이 생산되지 않았으며, 주로 한라봉이 출하되는 시점부터 11월 초까지 가격이 높게 형성되는 경향을 보였다.

경매 수량이 적으면 적을수록 평균 가격은 높게 형성되는 경향을 보였다(Table 1). 경매수량과 가격 간에는 부의 상관, 하한가와 평균가와는 고도로 유의한 부의 상관분석을 보였다. Ko and Kim(2012)의 결과에 따르면, 3가지 시나리오별 시뮬레이션 결과, 시나리오 Ⅱ의 실질농가수취가격은 2021년 kg당 2,417원으로 전망했는데, 본 연구에서는 1월말까지의 평균 가격을 보면 kg당 최대 5,300원에서 최저 3,600원, 평균 4,020원으로 경매가 이루어져, 본 연구와는 차이가 났다.

한라봉의 평균 가격은 하락하는 추세를 보였다(Fig. 3). 상한가와 하한가의 차이를 보면, 0 – 76,000원의 차이로 평균 17,720원의 차이를 보였다. 그 이유를 살펴보면, 한라봉 품질에 차이가 클 것으로 생각되기 때문이다. Han and Kang(2011)는 우리나라에서 한라봉 재배에서 주로 탱자를 대목으로 이용했는데, 이 탱자가 수세저하 등을 초래하여 과실 품질을 저하시킨다고 한다고 하였다. 이러한 과실 품질 저하 차이는 재배적인 방법에 기인한 경향이 많으므로, 품질적인 측면에서 가격 차이가 발생한

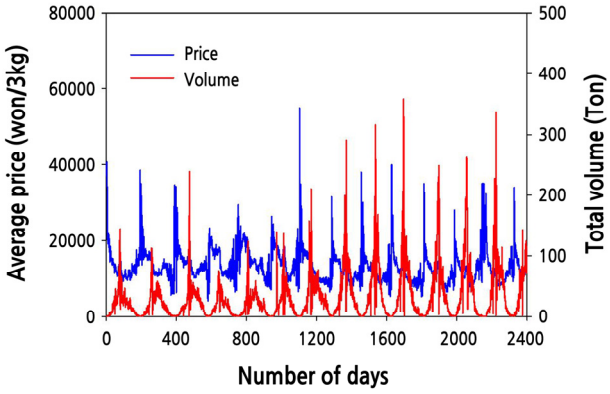


**Fig. 2.** Changes in the average price and total volume of Hallabong tangor fruits from October 31, 2007 to January 31, 2021.

**Table 1.** Correlations among total volume and the high, low and average prices

| | Total volume | High | Low | Average |
|---|---|---|---|---|
| Total volume | 1 | −0.042* | −0.319*** | −0.198*** |
| High | | 1 | 0.273*** | 0.601*** |
| Low | | | 1 | 0.829*** |
| Average | | | | 1 |

것으로 판단된다.

초기 학습결과에 따라 모델을 구성하는 여러 파라미터를 변경하여 모델을 최적화하는데, 이러한 파라미터를 하이퍼 파라미터(hyper parameter)라고 부르는데, 이들의 조합에 따라 모델의 성능이 좌우된다. 본 연구에서의 하이퍼 파라미터는 학습반복 횟수(Epochs)와 배치사이즈(batch size)였다. 학습용 자료를 이용하여 학습을 한번 마치는 것을 epoch라 하며, 어떤 크기로 dataset를 나눌지 결정하는 크기를 batch size라 한다. 본 연구에서 사용한 하이퍼 파라미터들을 조정하여 RMSE를 고려하여 최적 모델값을 결정하였다. RMSE의 변화를 살펴본 결과, 700 epochs와 128 batch size에서 가장 낮았다(Figs. 4 and 5).

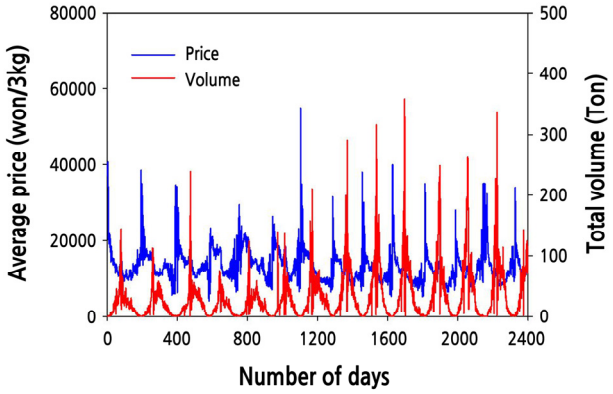


**Fig. 3.** Changes in the price (high-low prices) of Hallabong tangor fruits from October 31, 2007 to January 31, 2021.

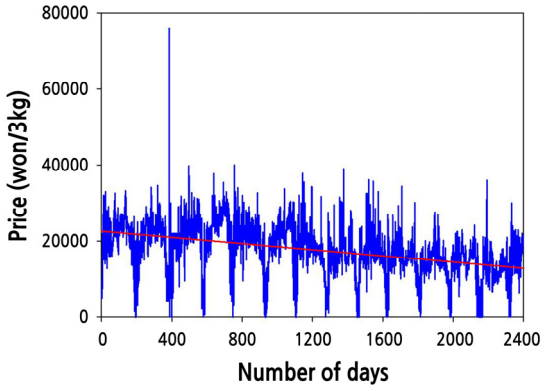


**Fig. 4.** The RMSE of the LSTM network using different epochs (Batch size = 128).

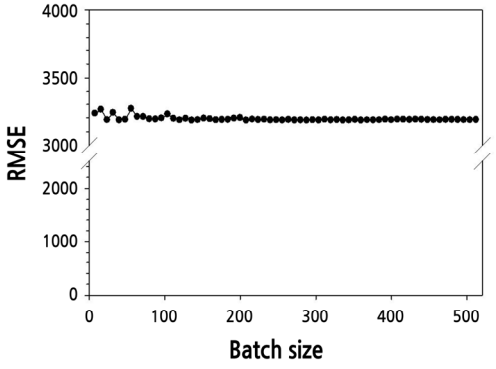


**Fig. 5.** The RMSE of the LSTM network using different batch sizes (Epochs = 700).

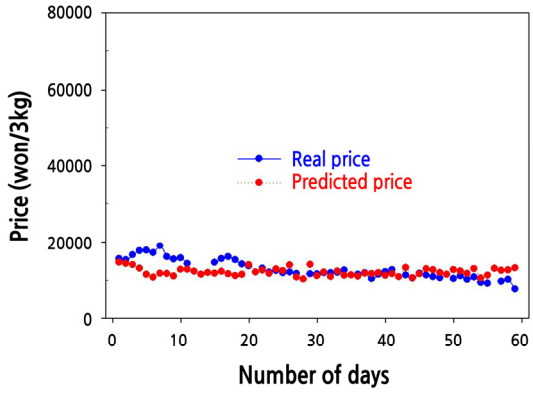


**Fig. 6.** Predicted Hallabong tangor fruit price for two months (February 1-March 31, 2021, Batch size = 128 and Epochs = 700).

LSTM 네트워크 최적 모델값 결과를 바탕으로 2달 동안(2021년 2월 – 3월) 시장의 가격 동향을 예측해 보았다(Fig. 6). 초기 예측값에는 약간의 변화가 있었지만, 전체적으로 감소하는 경향을 보였다. 예측값보다는 실측값의 가격 변화폭이 컸다. 실측값과 예측값의 최대와 최소 절대값 편차는 각각 7,260원과 90원이었으며, 평균 절대값 편차는 2,060원이었다. 평균 제곱근 오차는 1040이었다. 예측하는 시간이 길수록 예측 성능이 감소하는 경향을 보인다고 했는데(Lee et al., 2020), 본 연구에서는 예측 성능이 감소하는 경향은 보이지 않았지만, 60일 이후에는 예측 성능이 감소될 것으로 판단하였다.

농산물의 과잉 공급은 가격을 폭락시켜 농가에 재정적인 손실을 초래하는 반면에, 농산물의 공급 부족은 가격을 상승시켜 소비자에게 부담을 준다. 이러한 농산물의 수요와 공급의 불균형은 농민과 소비자 모두에게 영향을 미친다. 특히, 농산물 가격을 예측하면, 재배자가 파종, 수확 및 거래 시기를 결정하는 데 도움이 된다. 따라서, 가격 동향이나 물량 동향을 예측하는 방법을 딥러닝 기반으로 보다 더 정확하게 예측할 수만 있다면, 농산물 수급을 위한 정부 정책 수립에 사용될 수 있으며, 농가 소득 향상과 소비자들에게도 유용한 정보로 활용될 수 있을 것이다. 또한 이러한 예측 정보를 이용하여 농산물 재배에 영향을 끼칠 수 있는 요인들을 고려하여 공급량을 안정적을 유지할 수 있다면 좋을 것이다.

## 초 록

농산물의 가격 예측은 시장 수급 조절 결정에 큰 영향을 끼친다. 본 연구의 목적은 전국 9개 도매시장에서 유통되고 있는 한라봉의 평균가격을 딥러닝 기반의 LSTM 네트워크를 이용하여 한라봉 가격을 예측하고자 하였다. 9개의 도매시장에서 거래되고 있는 한라봉의 3kg 단위 평균가격을 바탕으로 2007년부터 2021년 1월까지 2,398개의 일일 데이터를 사용하였다. Training 데이터는 전체 데이터의 80%, test 데이터는 전체 데이터의 20%를 사용하여 모델을 학습시켰다. LSTM 네트워크를 위해 보다 더 정확한 가격을 예측하기 위해서는 epoch와 batch size를 결정해야 하는데, epoch는 700과 batch size 128가 적합한 모델이라 판단하였다. 이 값을 LSTM 네트워크를 적용한 결과, 비교적 한라봉의 평균 가격을 정확하게 예측할 수 있었다 (RMSE = 1040). 딥러닝 기반 한라봉 가격 예측 모델은 농산물 가격 동향과 정부 정책 수립 및 소비자들 요구에게도 유용한 정보로 활용될 수 있을 것이다.

**추가 주요어 :** 딥러닝, 인공지능, 순환신경망, 장기 의존성

## Literature Cited

**Banerjee T, Sinha S, Choudhury P** (2022) Long term and short term forecasting of horticultural produce based on the LSTM network model. Appl Intell 52:9117-9147. doi:10.1007/s10489-021-02845-x

**Han SH, Kang JH** (2011) Effect of water states of fruit vesicle and leaf on fruit quality in 'Trifoliate' orange and 'Swingle citrumelo' rootstock of 'Shiranuhi' [(*Citrus unshiu* × *C. sinensis*) × *C. reticulata*] mandarin Hybrid, M16A line in plastic film house cultivation. J Bio-Env Con 20:204-210

**Hochreiter S, Schmidhuber J** (1997) Long short-term memory. Neural Comput 9:1735-1780. doi:10.1162/neco.1997.9.8.1735

**Im JM, Kim WY, Byoun WJ, Shin SJ** (2018) Fruit price prediction study using artificial intelligence. JCCT 4:197-204. doi:10.17703/JCCT.2018.4.2.197

**Kim BS** (2005) A comparison on forecasting performance of the application models for Forecasting of vegetable prices. KJAT 46:89-113

**Kim SK, Oh TI** (2018) Real-time PM10 concentration prediction LSTM model based on IoT streaming sensor data. JKAIS 19:310-318. doi:10.5762/KAIS.2018.19.11.310

**Ko SB** (2017) A study on the analysis of shipment structure of hallabong tangor in Korea. JKAIS 18:670-676. doi:10.5762/KAIS.2017.18.7.670

**Ko SB, Kim BS** (2012) A study on the model specification for supply-demand forecast of hallabong tangor in Korea. JKAIS 13:5163-5168. doi:10.5762/KAIS.2012.13.11.5163

**Kurumatani K** (2020) Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. SN Appl Sci 2:1434. doi:10.1007/s42452-020-03225-9

**Lee JK, Oh JW, Cho YJ, Lee DH** (2020) A research about time domain estimation method for greenhouse environmental factors based on artificial intelligence. Protected Hortic Plant Fac 29:277-284. doi:10.12791/KSBEC.2020.29.3.277

**Lee YS, Choi BO, Shim SB** (2006) A time series analysis on prices of fruits and vegetables. Agri-Policy Focus R537, KREI 53-58

**Nam KH, Choe YC** (2015) A study on onion wholesale price forecasting model. JAECD 22:423-434. doi:10.12653/jecd.2015.22.4.0423

**Olah C** (2015) Understanding LSTM networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs. 2021.03.17

**Prakash N, Sumaiya Farzana G** (2019) Short term price forecasting of horticulture crops using long short term memory neural network. Adv Comput Bio-Eng 111-118. doi:10.1007/978-3-030-46943-6_12

**Shin SH, Lee MK, Song SK** (2018) A prediction model for agricultural products price with LSTM network. Jour of KoCon a 18:416-429. doi:10.5392/JKCA.2018.18.11.416