

## 다중 입력 딥러닝을 이용한 서리 발생 추정

김용석, 허지나\*, 김응섭, 심교문, 조세라, 강민구

국립농업과학원 기후변화평가과

(2023년 12월 12일 접수; 2024년 01월 11일 수정; 2024년 02월 23일 수락)

## Estimation of Frost Occurrence using Multi-Input Deep Learning

Yongseok Kim, Jina Hur\*, Eung-Sup Kim, Kyo-Moon Shim, Sera Jo, Min-Gu Kang

*Climate Change Assessment Division, National Institute of Agricultural Sciences*

(Received December 12, 2023; Revised January 11, 2024; Accepted February 23, 2024)

### ABSTRACT

In this study, we built a model to estimate frost occurrence in South Korea using single-input deep learning and multi-input deep learning. Meteorological factors used as learning data included minimum temperature, wind speed, relative humidity, cloud cover, and precipitation. As a result of statistical analysis for each factor on days when frost occurred and days when frost did not occur, significant differences were found. When evaluating the frost occurrence models based on single-input deep learning and multi-input deep learning model, the model using both GRU and MLP was highest accuracy at 0.8774 on average. As a result, it was found that frost occurrence model adopting multi-input deep learning improved performance more than using MLP, LSTM, GRU respectively.

**Key words:** Frost, Deep learning, MLP, LSTM, GRU, Multi-input Model



\* Corresponding Author : Jina Hur  
(hjn586@korea.kr)

## 1. 서 론

일반적으로 서리는 맑고 바람이 없으며, 기온이 어느점 이하로 떨어질 때 자주 발생한다고 알려져 있다. Kwon *et al.*(2008)은 안정된 대기 조건에서 이동성 고기압의 영향을 받은 맑고 바람이 없는 야간에, 지표층의 기온이 0℃ 이하로 떨어질 때 수증기가 승화하여 서리가 발생한다고 하였다. 또한, Temeyer *et al.*(2003)은 역전층이 형성되어 지표층이 상층보다 차가워졌을 때, 수증기가 유입되거나 이류현상에 의해 차가워진 지역에 습윤한 공기가 유입될 때 서리가 발생한다고 하였다. 우리나라에 서리가 발생하기 쉬운 기간은 가을부터 봄철 기간이며, 발생 시기에 따라 다양한 농작물에 피해를 입힌다. 개화기 전후 발생하는 봄철 늦서리는 과수 및 채소 작물의 어린 잎, 꽃에 냉각·결빙 현상을 야기하고, 수확시기에 발생하는 가을철 이른 서리는 작물(고구마, 땅콩 등)들의 생육 지연 및 상품 가치 하락을 야기한다. 서리 피해를 최소화하기 위해서는 서리 발생 위험성에 대한 사전 정보를 기반으로 사전 방지 대책(송풍법, 연소법, 미세살수법 등)을 마련해야 한다. 따라서 서리 발생 가능성을 예측하는 것은 서리 피해 경감을 위한 준비 기간을 지원하는데 있어 매우 중요하다.

국내에서는 지난 20년동안 서리 발생 예측 모형에 대한 연구가 수행되었다. 2000년대 서리 발생 분석 연구는 일부지역에 대한 서리발생 전날의 24시 기온, 구름량, 5일간의 강수량 등 서리발생 전날의 기상특성을 이용한 판별분석(DA, Discriminant Analysis) 연구 및 상세 최저기온 공간정보에 특정기온(예: 3℃)을 기준으로 지역별 서리피해 경보의 살펴보는 연구가 수행되었다(Han *et al.*, 2009; Chung *et al.*, 2004). 최근 10년 서리 발생 연구는 전국의 서리 발생자료를 기반으로 서리발생과 밀접한 서리발생 전날의 기상요인을 분석하고 이 자료를 기계학습법(예: 랜덤포레스트, 다층퍼셉트론, 서포트벡터머신)의 학습데이터로 활용하여 기계학습 기반의 서리발생 예측 가능성을 살펴보았다(Kim *et al.*, 2017; Kim and Kim, 2022; Chun *et al.*, 2021; Noh *et al.*, 2021). 또한, 기상청도 농림수산부문의 기상융합서비스의 일환으로 주요 기상데이터를 바탕으로 기계학습기법(XGBoost)을 적용하여 5km 격자형의 서리발생 가능성을 예측하고 있다(<https://bd.kma.go.kr/kma2020/fs/agricultureInfo.do>). 그 밖에 Noh *et al.*(2021)는 농업에서 서리나 저온 피해를 예측하기 위

해 20m 고해상도의 기온과 풍속자료를 생산하는 연구를 수행하였다.

또한, 딥러닝 모형의 구축함에 있어 Dua *et al.*(2021)은 웨어러블 센서를 이용하여 인간 활동을 인식하는 모형을 구축하기 하기 위해 다중 입력 딥러닝 모형을 구축하여 단일 입력 딥러닝 모형에 비해 우수한 연구 결과를 도출하였으며, Chen *et al.* (2022)은 인간의 감정을 인식하기 위해 다중 입력 딥러닝 모형을 구축하여 우수한 성능을 발휘하였다.

따라서 본 연구에서는 기존에 서리 발생 예측에 많이 연구된 단일 입력 딥러닝 모형에 보다 더 다양한 특성을 반영할 수 있는 다중 입력 딥러닝 모형의 우수한 특징을 분석하기 위해 수행하였다. 이에 따라 1일 단위 자료를 이용하는 다층퍼셉트론(MLP; Multi Layer Perceptron)과 3일의 연속된 자료를 이용하는 장단기 메모리(LSTM; Long Short Time Memory)와 게이트 순환 유닛(GRU; Gated Recurrent Unit)을 함께 적용한 다중 입력 딥러닝 모형을 구축하고 서리 발생 예측 가능성을 살펴보았으며, 다중 입력 딥러닝 모형의 서리 발생 예측에 대한 정확도를 단일 입력 딥러닝 모형의 정확도와 비교 평가해 보았다.

## 2. 재료 및 방법

### 2.1. 기상자료 수집 및 데이터 세트 구축

서리 관측 자료는 국립기상과학원으로부터 수집한 기상청 종관기상관측지점 24지점(Table 1)에 대한 2014년부터 2022년까지의 관측 정보를 이용하였다. 서리 발생과 관련된 기상인자들은 서리가 발생한 날과 발생하지 않은 날로 분류하여 정리하였다. 기상인자는 Han *et al.*(2009), Robinson and Mort(1996), Temeyer *et al.*(2003)에서 제시된 인자 중 기상청 동네예보에서 예보하는 최저기온, 평균풍속, 구름량, 강수량, 상대습도로 선정하였으며, 일반적으로 서리와 관련성이 많으나 기상청 예보가 제공되지 않는 초상온도나 지면온도는 사용하지 않았다. 연구를 수행하기 위해 서리가 발생한 날에 대한 데이터 6,431 세트와 서리가 발생하지 않은 날에 대한 데이터 59,719 세트를 수집하여 총 66,150 세트의 데이터를 수집하였지만 학습데이터의 균형을 맞추기 위해 서리발생한 날의 데이터 세트에 맞춰서 서리가 발생한 날의 데이터 세트와 서리가 발생하지 않은 날의 데이터 세트를 6,431세트로 동일하게 맞춰서 총 12,862세트로 구성하였다. 그리고 이중

20%인 2,573세트는 Test Set로 정확도를 검증하는데 사용하였고, 나머지 10,289세트 중에서 모형을 구축하기 위해 80%는 8,231세트는 Training Set로, 20%인 2,050세트는 Validation Set로 사용하였다. 또한, 모형의 재현성을 분석하기 위해 학습과 검증을 5번 반복하여 정확도를 비교하였다.

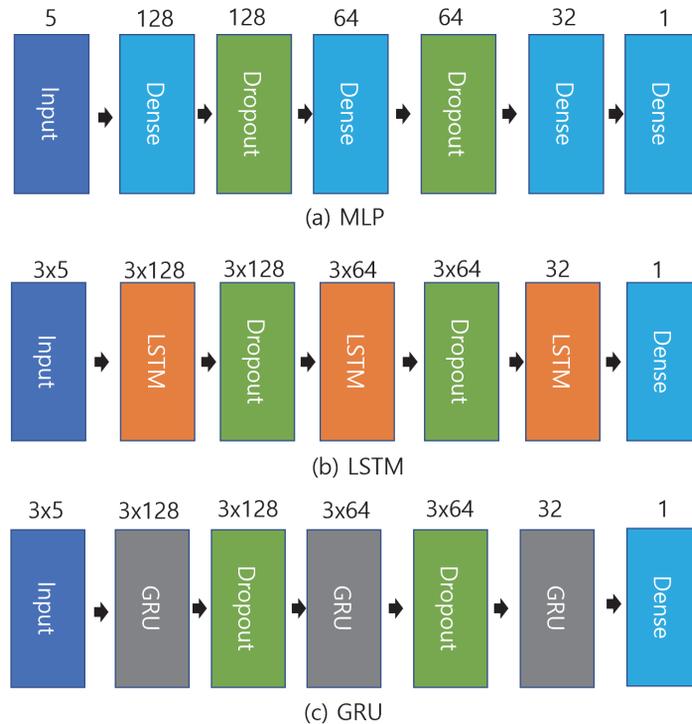
**Table 1.** Synoptic weather stations collected the frost data

Region	Station
Gangwondo	Chuncheon, Bukchuncheon, Bukgangneung
Gyeonggido	Suwon
Chungcheongdo	Cheongju, Seosan, Hongseong
Gyeongsangdo	Ulsan, Pohang, Andong, Changwon
Jeollado	Jeonju, Mokpo, Yeosu
Metropolitan city	Daejeon, Daegu, Incheon, Seoul, Gwangju, Busan
Island	Baengnyeongdo, Heuksando, Jeju, Ulleungdo

## 2.2. 딥러닝을 이용한 서리발생추정

기상자료를 이용하여 서리발생 유무를 추정하기 위해 딥러닝 중 Fig. 1과 같이 MLP, LSTM, GRU를 활용하였으며, MLP의 5가지의 설명변수를 1일 단위로 입력 받아 128개의 노드의 Dense 층과 64개, 32개의 Dense 층을 각각 거쳐 마지막 Dense 층에서 시그모이드(Sigmoid) 함수에 의한 서리 발생 확률 값이 출력되며, 확률 값이 0.5가 초과하면 서리가 발생한 것으로 추정하였다. 그리고, 과적합을 피하기 위해 20%의 Dropout을 2번 수행하였다. LSTM과 GRU는 5가지의 설명변수를 3일 단위로 입력 받기 때문에 3x5의 입력을 받으며, MLP의 3개의 Dense 층을 각각 LSTM 층과 GRU 층으로 교체하여 구성하였으며, 마지막은 Dense 층을 두어 서리 발생 확률을 결과 값으로 출력하게 구성하였다.

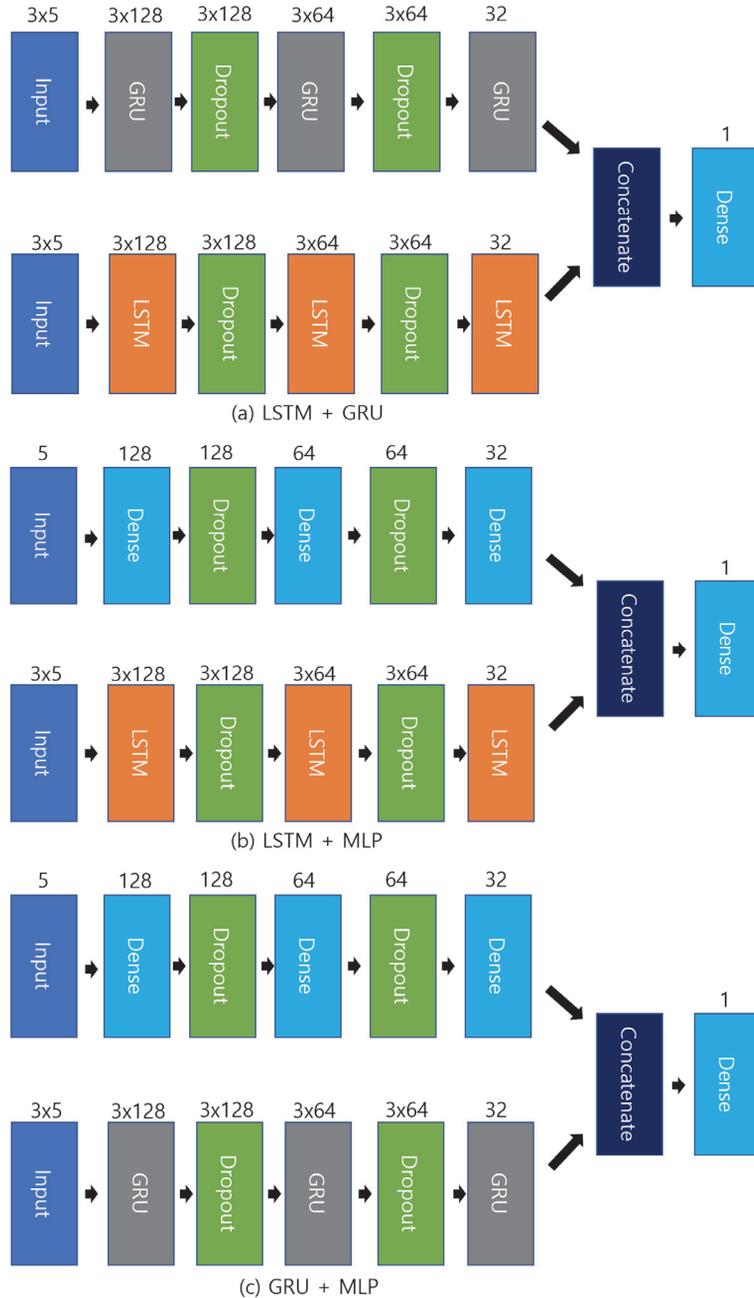
Fig. 2와 같이 MLP, LSTM, GRU를 함께 학습시킨 다중 입력 모형을 만들어 각각의 단일 딥러닝으로 만든 모형과 정확도를 비교하였다. 에포크(epoch)는 200으로 설정하였으며, 손실함수(loss function)은 binary\_crossentropy로 설정하였다. Early-stopping을 위해



**Fig. 1.** Schematic diagram of single-input deep learning frost occurrence estimation model.

Validation Accuracy가 가장 높은 값을 기준으로 5회 동안 변화가 없으면 학습을 멈추게 설정하였다. MLP는 1일(당일)의 기상자료로 서리발생 유무를 추정하였으며, LSTM과 GRU는 3일(당일~2일전)의 시계열 기상자료를 이용하여 서리발생 유무를 추정하였다. 프로

그램은 Google의 Colaboratory (python version: 3.10.12)에서 딥러닝 프레임워크 중 TensorFlow에서 제공하는 Keras의 Functional API를 활용하여 모형을 구축하였다.



**Fig. 2.** Schematic diagram of multi-input deep learning frost occurrence estimation model.

### 2.2.1. MLP

MLP는 SLP (Single-Layer Preceptron)의 XOR 게이트 같은 배타적인 논리합의 문제를 해결한 신경망으로 입력층, 은닉층, 출력층 방향으로 연결되어 있고 한번 지나간 층으로 다시 연결되는 피드백(feedback)이 없는 순방향 인공신경망(Feed-Forward Deep Neural Network, FFDNN) 중 하나이다.

입력층의  $x_i$ 는 은닉층의  $j$ 번째 노드에 입력되어 다음과 같은 식(1)과 같은 계산과정을 거쳐 최종적으로 활성화 함수를 거쳐 출력층으로 결과값을 출력하며, 손실함수를 통해 결과값과 실제값의 오차를 계산하여 가중치( $w_{ji}$ )를 계속 업데이트 하여 최적의 결과값을 계산한다.

$$Z_j = \sum_{i=1}^n w_{ji}x_i \quad (1)$$

### 2.2.2. LSTM

LSTM은 은닉층의 노드에서 활성화 함수를 통해 나온 결과값을 출력층 방향으로 보내면서, 다시 은닉층 노드의 다음 계산의 입력으로 보내는 특징을 가진 RNN(Recurrent Neural Network)의 일종으로 RNN에서 나타나는 입력 데이터가 많아 질수록 은닉층에서 과거의 정보가 마지막 레이어까지 충분히 전달될지 못하는 현상인 장기 의존성 문제를 해결하기 위해 고안되었다.

LSTM은 이전 단계 정보를 Memory Cell에 저장하여 다음 단계로 전달하는 것으로 이전 정보를 적절히 기억하고 그 결과에 새로운 정보를 갱신하여 다음 시점으로 전달한다. LSTM의 구성은 망각 Forget Gate, Input Gate, Output Gate의 3가지 게이트와 Cell State로 이루어져 있다.

Cell State는 일종의 정보의 컨베이어 벨트 역할을 하여 오래된 정보를 기억하고 새로운 정보를 적절하게 갱신할 수 있게 해준다.

Forget gate는 과거의 정보를 얼마나 반영할지를 결정하는 게이트이며, 식(2)와 같이  $x_t$ (현재 데이터)와  $h_{t-1}$ (과거 은닉층 값)에 가중치  $W_{xf}$ ,  $W_{hf}$  각각 곱하고 더하여 활성화 함수(sigmoid)를 적용한다. 그리고 최종 값이 1에 가까울수록 과거 정보를 많이 반영하고 0에 가까울수록 과거 정보는 낮게 반영된다.

$$F_t = \text{sigmoid}(W_{hf}h_{t-1} + W_{xf}x_t + b_f) \quad (2)$$

Input Gate는 새로운 정보를 어떻게 반영할지 결정하는 게이트이며, 식(3)과 같이  $x_t$ (현재 데이터)와  $h_{t-1}$ (과거 은닉층 값)에 가중치  $W_{xi}$ ,  $W_{hi}$  각각 곱하고 더하여 sigmoid 함수를 적용하여 어떤 정보를 업데이트 할지 결정하고, 식(4)와 같이  $x_t$ (현재 데이터)와  $h_{t-1}$ (과거 은닉층 값)에 가중치  $W_{xc}$ ,  $W_{hc}$  각각 곱하고 더하여 활성화 함수(tanh)를 적용하여 현재의 새로운 정보를 생성한다.

$$I_t = \text{sigmoid}(W_{hi}h_{t-1} + W_{xi}x_t + b_i) \quad (3)$$

$$\hat{C}_t = \tanh(W_{hc}h_{t-1} + W_{xc}x_t + b_c) \quad (4)$$

다음의 Cell state는 과거 Cell State를 새로운 정보로 업데이트 하는 과정으로, 식(5)와 같이 Forget Gate의  $F_t$ 와 Input Gate의  $I_t$  및  $C_t$ 를 이용하여 저장하는 단계이며, 과거의 정보를 Forget Gate을 통해 잊고, 현재의 정보를 Input Gate을 이용해 대처하여 현재 시점 기준의 Memory Cell 값을 계산한다.

$$C_t = F_t C_{t-1} + I_t \hat{C}_t \quad (5)$$

Output Gate는 식(6)과 식(7)에서와 같이 Forget Gate와 Input Gate에 의해서 구해진  $C_t$ 값을 얼마나 다음 레이어로 전달할지 결정한다.

$$O_t = \text{sigmoid}(W_{ho}h_{t-1} + W_{xo}x_t + b_o) \quad (6)$$

$$H_t = O_t * \tanh(C_t) \quad (7)$$

### 2.2.3. GRU

GRU는 LSTM에서 사용되는 cell state가 따로 없고 hidden state가 그 역할을 함께 하며, 게이트 또한 Reset Gate과 Update Gate 두가지만 존재하여 이 과정에서 LSTM에 비해 연산량이 줄어든다.

GRU 구조에서 Reset gate는 과거의 정보를 얼마나 반영할지를 결정하는 게이트이며, 식(8)와 같이  $x_t$ (현재 데이터)와  $h_{t-1}$ (과거 은닉층 값)에 가중치  $W_{xr}$ ,  $W_{hr}$  각각 곱하고 더하여 활성화 함수(sigmoid)를 적용한다. 그리고 최종 값이 1에 가까울수록 과거 정보를 많이 반영하고 0에 가까울수록 과거 정보는 사라지게 된다.

$$R_t = \text{sigmoid}(W_{hr}h_{t-1} + W_{xr}x_t + b_r) \quad (8)$$

Update Gate는 과거와 현재 정보를 각각 어느정도 반영할지 결정하는 게이트이며, 식(9)과 같이  $x_t$ (현재 데이터)와  $h_{t-1}$ (과거 은닉층 값)에 가중치  $W_{xz}$ ,  $W_{hz}$  각각 곱하고 더하여 활성화 함수(sigmoid)를 적용한다.

$$Z_t = \text{sigmoid}(W_{hz}h_{t-1} + W_{xz}x_t + b_z) \quad (9)$$

또한, 다음 시점으로 전달해 줄 hidden state 데이터  $H_t$ 를 만들기 위해, 과거 출력 값  $H_{t-1}$  과  $R_t$ 값을 이용함으로써, 현시점의 데이터  $x_t$ 와 과거 데이터를 가지고  $\tilde{H}_t$  데이터 생성한다.

$$\tilde{H}_t = \text{tanh}(W_{hc}h_{t-1} + W_{xc}x_t * R_t + b_c) \quad (10)$$

즉, 식(11)과 같이  $(1-Z_t) * H_t$  은 과거의 출력 값  $H_{t-1}$ 를 얼마나 반영할지 나타내며,  $(Z_t) * \tilde{H}_t$ 는 현 시점의 데이터 가운데 얼마만큼 다음 단계로 반영할지 계산하는 것이므로, 현재와 과거데이터 가운데 얼마만큼 다음 단계로 반영할지를 정하기 위해 현재와 과거 데이터를 학습된 비율로 모두 합한 값을 최종 출력 값  $H_t$ 로 계산하게 된다.

$$H_t = (1 - Z_t) * H_{t-1} + Z_t * \tilde{H}_t \quad (11)$$

**2.2.4. 다중 입력 모형(Multi-Input Model)**

데이터 특성에 따른 서로 다른 모델이 input으로 사용되어 하나의 output을 도출하는 네트워크를 구현하기 위해 LSTM과 GRU, LSTM과 MLP, GRU와 MLP의 각각에 대한 결과를 Concatenate 층을 통해 결합하여 학습하도록 모형을 구축하고 서리발생 유무를 추정하였다.

**3. 결 과**

먼저, 딥러닝의 학습데이터로 활용될 기상인자들의 특징을 살펴보았다. 수집된 관측 자료를 이용하여 서리가 발생한 날과 서리가 발생하지 않은 날의 기상특징을 비교한 결과 Table 2과 같았다.

최저기온의 경우 서리가 발생한 날의 평균은 -3.2℃로 서리가 발생하지 않은 날의 11.2℃보다 13.4℃가 낮았으며, 평균풍속은 서리가 발생한 날은 1.6m/s로 서리가 발생하지 않은 날의 2.6m/s보다 1.0m/s 약한 것으로 나타났다. 구름량은 서리가 발생한 날이 3.2할로 서리가 발생하지 않은 날의 5.3할보다 2.1할 낮았으며, 강수량은 0.2mm로 서리가 발생하지 않은 날의 3.6mm보다 3.4mm 적었다. 상대습도는 서리가 발생한 날은 62.1%로 서리가 발생하지 않은 날의 68.8%보다 6.7% 낮았다.

서리가 발생한 날의 기상인자와 서리가 발생하지 않은 날의 기상인자 간에 T-test를 실시한 결과 0.05의 유의수준에서 p-value가 0.05이하로 나타나 모든 인자에서 유의하였다. 이것으로 통계적으로 서리가 내린 날이 서리가 내리지 않은 날에 비해 기온이 낮고 구름이 적었으며, 바람이 약한 맑은 날인 많은 것을 알 수 있었다.

분석된 기상인자들을 이용하여 6가지의 서리발생 예측 모형을 구축하고, 각 모형의 5회에 걸쳐 랜덤으로 샘플을 선택하여 모형을 구축하고 서리발생 예측 정확도를 분석하였다. 모형을 구축하면서 먼저 단일 딥러닝을 이용하여 서리발생을 추정해 보았을 때, MLP를 이용하여 서리발생을 추정한 결과는 서리가 발생한다고 추정한 것과 발생하지 않는다고 추정한 것이 얼마나 정확하게 맞는지를 나타내는 정확도(Accuracy)는 5회 평균 0.8586으로 나타났다. 그리고 정밀도(Precision)는 서리가 발생했다고 추정한 것 중

**Table 2.** Characteristic of climate factors of frost occurrence days and frost free days

		Tmin (°C)	WS (m/s)	Cloud (0.1%)	Pre (mm)	RH (%)
FOD	Average (SD)	-3.2 (4.2)	1.6 (0.9)	3.2 (2.5)	0.2 (1.3)	62.1 (12.4)
FFD	Average (SD)	11.2 (9.1)	2.6 (1.6)	5.3 (3.1)	3.6 (12.9)	68.8 (17.0)

Tmin: Minimum temperature, WS: Average wind speed, Cloud: Amount of cloud, Pre: Amount of precipitation, RH: Average relative humidity, SD: Standard deviation, FOD: Frost occurrence day, FFD: Frost free day

실제 서리가 발생했는지 나타내는 지표로 0.8080로 나타났으며, 재현율(Recall)은 실제 서리가 발생했을 때 서리가 발생한다고 추정했는지 나타내는 지표로 0.9438로 나타났다. 정밀도와 재현율을 함께 고려한 지표인 F1-score는 0.8704로 나타났다(Table 3).

LSTM를 이용하여 서리발생을 추정한 결과는 정확도(Accuracy)가 5회 평균 0.8662으로 나타났다. 그리고 정밀도(Precision)는 0.8363로 나타났으며, 재현율(Recall)은 0.9123로 나타났고, F1-score는 0.8724로 나타났다(Table 4).

GRU를 이용하여 서리발생을 추정한 결과는 정확도(Accuracy)는 5회 평균 0.8741으로 나타났다. 그리

고 정밀도(Precision)는 0.8347로 나타났으며, 재현율(Recall)은 0.9348로 나타났고, F1-score는 0.8817로 나타났다(Table 5).

LSTM과 GRU를 함께 학습시킨 다중 입력 딥러닝 모형의 경우, 5회 평균 정확도가 0.8771, 정밀도가 0.8372, 재현율이 0.9374, F1-score가 0.8844로 나타났다 (Table 6).

LSTM과 MLP를 함께 학습시킨 다중 입력 딥러닝 모형의 경우, 5회 평균 정확도가 0.8763, 정밀도가 0.8351, 재현율이 0.9391, F1-score가 0.8840로 나타났다 (Table 7).

GRU과 MLP를 함께 학습시킨 다중 입력 딥러닝

**Table 3.** The accuracy comparison of frost occurrence classification estimation using MLP

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8480	0.8574	0.8752	0.8562	0.8562	0.8586
Precision	0.7844	0.8006	0.8415	0.8054	0.8079	0.8080
Recall	0.9621	0.9536	0.9256	0.9412	0.9366	0.9438
F1-score	0.8642	0.8705	0.8816	0.8680	0.8675	0.8704

**Table 4.** The accuracy comparison of frost occurrence classification estimation using LSTM

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8632	0.8570	0.8729	0.8733	0.8647	0.8662
Precision	0.8332	0.8510	0.8371	0.8276	0.8327	0.8363
Recall	0.9094	0.8668	0.9272	0.9442	0.9140	0.9123
F1-score	0.8696	0.8588	0.8798	0.8821	0.8715	0.8724

**Table 5.** The accuracy comparison of frost occurrence classification estimation using GRU

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8799	0.8706	0.8644	0.8803	0.8752	0.8741
Precision	0.8309	0.8461	0.8118	0.8378	0.8469	0.8347
Recall	0.9551	0.9070	0.9505	0.9442	0.9171	0.9348
F1-score	0.8886	0.8755	0.8757	0.8878	0.8806	0.8817

**Table 6.** The accuracy comparison of frost occurrence classification estimation using multi-input LSTM-GRU model

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8811	0.8760	0.8752	0.8772	0.8760	0.8771
Precision	0.8316	0.8370	0.8415	0.8426	0.8333	0.8372
Recall	0.9566	0.9349	0.9256	0.9287	0.9411	0.9374
F1-score	0.8898	0.8833	0.8816	0.8836	0.8840	0.8844

**Table 7.** The accuracy comparison of frost occurrence classification estimation using multi-input LSTM-MLP model

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8799	0.8768	0.8756	0.8737	0.8756	0.8763
Precision	0.8349	0.8449	0.8265	0.8326	0.8365	0.8351
Recall	0.9481	0.9241	0.9520	0.9365	0.9349	0.9391
F1-score	0.8879	0.8827	0.8848	0.8815	0.8830	0.8840

**Table 8.** The accuracy comparison of frost occurrence classification estimation using multi-input GRU-MLP model

	1st	2nd	3rd	4th	5th	Average
Accuracy	0.8776	0.8791	0.8752	0.8768	0.8784	0.8774
Precision	0.8280	0.8254	0.8509	0.8179	0.8488	0.8342
Recall	0.9543	0.9628	0.9109	0.9706	0.9218	0.9441
F1-score	0.8866	0.8888	0.8799	0.8877	0.8838	0.8854

모형의 경우, 5회 평균 정확도가 0.8774, 정밀도가 0.8342, 재현율이 0.9441, F1-score가 0.8854로 나타났다 (Table 8).

단일 딥러닝의 경우 평균적으로 MLP가 가장 정확도가 낮았으며, LSTM, GRU 순으로 정확도가 높게 나타났고, 다중 입력 딥러닝 모형의 경우 3가지 모형이 거의 비슷한 결과가 나타났지만 그 중 평균적으로 GRU와 MLP를 이용한 모형이 가장 정확도가 높았다.

정확도면에서 단일 딥러닝이 다중 입력 딥러닝에 비해 샘플에 따라 정확도 편차가 더 컸으며, 전반적으로 다중 입력 딥러닝 모형이 정확도 면에서 더 우수했다. 다만, 서리 발생 예측 모형의 경우 서리가 발생할 때를 더 정확하게 서리 발생을 예측 것이 중요하기 때문에 재현율도 중요한 지표 중에 하나이다. 재현율을 봤을 때 정확도가 가장 낮은 MLP가 0.9438로 상대적으로 높게 나타났다.

6개의 딥러닝을 이용한 서리 발생 예측 모형을 종합적으로 판단할 때 GRU와 MLP의 다중 입력 딥러닝 모형이 정확도와 재현율 면에서 가장 우수하게 나타났다.

#### 4. 결 론

단일 딥러닝 모형과 두 개의 딥러닝 모형을 연결한 다중 입력 딥러닝 기법을 이용하여 남한지역의 서리 발생 예측 모형을 개발하였다. 서리 발생 예측 모형을 평가하기에 앞서, 딥러닝의 학습데이터로 이용된 기상 인자들(최저기온, 평균풍속, 상대습도, 구름량, 강수량)의 특성을 살펴보았을 때, 서리가 발생하지 않은 날에 비해 서리가 발생한 날은 상대적으로 더 맑고 기온이 낮았으며, 바람이 약한 특성이 나타났다. 또한 이러한 차이는 통계적으로도 유의하게 나타났다. 분석

**Table 9.** Climate factors characteristic of correct prediction data and incorrect prediction data of frost occurrence classification estimation using multi-input GRU-MLP model

		Tmin (°C)	WS (m/s)	Cloud (0.1%)	Pre (mm)	RH (%)
Correct Prediction	Average (SD)	6.2 (8.8)	2.6 (1.3)	4.3 (3.0)	2.3 (12.0)	63.2 (15.8)
Incorrect Prediction	Average (SD)	6.1 (8.4)	2.5 (1.1)	4.1 (2.8)	1.8 (9.6)	62.8 (15.0)

Tmin: Minimum temperature, WS: Average wind speed, Cloud: Amount of cloud, Pre: Amount of precipitation, RH: Average relative humidity, SD: Standard deviation

된 기상인자들을 이용하여 6가지 딥러닝을 이용한 서리 발생 예측 모형을 구축하고, 5회 반복 실험을 통해 평균적인 정확도를 비교하였다. 평균적으로 MLP가 가장 정확도가 낮았으며, LSTM, GRU 순으로 정확도가 높게 나타났고, 다중 입력 딥러닝 모형의 경우 3가지 모형이 거의 비슷한 결과가 나타났지만 그 중 평균적으로 GRU와 MLP를 이용한 모형이 가장 정확도가 높아 6가지 모형 중에서도 가장 정확도가 높았다. 다중 입력 딥러닝의 정확도 향상은 단일 모형의 정보들을 결합함으로써 정보 손실을 줄일 수 있기 때문인 것으로 추정되지만 단일 모형의 입력 정보 형태와 다중 입력 모형의 입력 정보 형태가 수치 정보로 동일하기 때문에 다양한 형태의 정보의 결합을 통한 정확도 향상만큼은 나오지 못했다. 또한 Table 9에서 GRU와 MLP를 이용한 다중 입력 딥러닝 모형이 정확히 서리 발생을 예측한 기상데이터와 서리 발생을 잘못 예측한 기상데이터의 특징과 Table 1에서의 기상특징 및 모형의 정밀도와 재현율을 종합적으로 분석하였을 때 모형은 다소 서리가 발생할 것이라고 예측하는 경향이 강하여 낮은 최저기온과 상대습도, 약한 풍속, 적은 구름량 및 강수량 뿐만 아니라 일부 상대적으로 높은 최저기온과 상대습도, 강한 풍속, 많은 구름량과 강수량이 발생한 날에도 서리가 발생할 것으로 예측하였기 때문인 것으로 예상된다. 그러므로, 앞으로의 연구에서 추가적인 계절적 특징과 다양한 형태(위성 정보, 격자형 재분석 자료 등)의 입력정보를 함께 고려할 수 있다면 다중 입력 딥러닝 기반 서리 발생 예측 모형의 성능이 더 크게 향상될 수 있을 것으로 생각된다.

## 적 요

본 연구에서는 딥러닝을 이용한 모형을 이용해서 우리나라 지역에 대한 서리 발생 예측 모형을 구축하였다. 딥러닝 모형의 학습 데이터로 다양한 기상인자들(최저기온, 풍속, 상대습도, 구름량, 강수량)을 사용하였으며, 기상인자들에 대한 통계적 분석 결과, 서리가 발생한 날과 서리가 발생하지 않은 날에 대해 각 요소별로 유의한 차이가 있는 것을 볼 수 있었다. 단일 딥러닝 모형 3가지와 다중 입력 딥러닝 모형 3가지를 이용하여 서리발생을 추정한 결과, 평균적으로 MLP가 가장 정확도가 낮았으며, LSTM, GRU 순으로 정확도가 높게 나타났고, 다중 입력 딥러닝 모형의 경우 3가지 모형이 거의 비슷한 결과가 나타났지만 그 중

평균적으로 GRU와 MLP를 이용한 모형이 가장 정확도가 높았다. 또한, 단일 딥러닝이 다중 입력 딥러닝에 비해 샘플에 따라 정확도 편차도 더 컸다. 이에 따라 결과적으로 단일 딥러닝 기반의 서리발생 예측 모형보다 다중 입력 딥러닝 기반의 서리발생 예측 모형이 안정성과 정확도와 재현율 측면에서 다소 우수한 것을 확인할 수 있었다.

## 감사의 글

본 연구는 농촌진흥청 “신농업기후변화대응체계구축사업(과제번호: RS-2020-RD009396)”의 지원으로 수행되었습니다.

## REFERENCES

- Chen, J. L., H. B. Liu, W. Wu, and D. T. Xie, 2010: Estimation of monthly solar radiation from measured temperatures using support vector machines - A case study. *Renewable Energy* **36**(2), 413-420.
- Chen, P., B. Zou, A. N. Belkacem, X. Lyu, X. Zhao, W. Yi, Z. Huang, J. Liang and C. Chen, 2022: An improved multi-input deep convolutional neural network for automatic emotion recognition. *Frontiers in Neuroscience* **16**.
- Chun, J. A., H.-J. Lee, S.-H. Im, D. Kim, and S. Baek, 2021: Comparative assessment of frost event prediction models using logistic regression, random forest, and LSTM networks. *Journal of Korea Water Resources Association* **54**(9), 667-680.
- Dua, N., S. N. Singh, and V. B. Semwal, 2021: Multiinput CNNGRU based human activity recognition using wearable sensors. *Computing* **103**, 1461-1478.
- Han, J. H., J. J. Choi, U. Chung, K. S. Cho, and J. P. Chun, 2009: Frostfall forecasting in the naju pear production area based on discriminant analysis of climate data. *Korean Journal of Agricultural and Forest Meteorology* **11**(1), 13-18. (in Korean with English abstract)
- Kim, H. K., and S. Kim, 2022: A study on frost prediction model using machine learning. *The Korean Journal of Applied Statistics* **35**(4), 543-552. (in Korean with English abstract)
- Kim, Y., K.-M. Shim, M.-P. Jung, and I.-T. Choi, 2017: Study on the Estimation of Frost Occurrence Classification Using Machine Learning

- Methods. *Korean Journal of Agricultural and Forest Meteorology* **19**(3), 86-92. (in Korean with English abstract)
- Kwon, Y. A., H. S. Lee, W. T. Kwon, and K. O. Boo, 2008: The weather characteristics of frost occurrence days for protecting crops against frost damage. *Journal of the Korean Geographical Society* **43**(6), 824-842. (in Korean with English abstract)
- Noh, I.S., H.W. Doh, S.O. Kim, S.H. Kim, S.E. Shin, S.J. Lee, 2021: Machine Learning-Based Hourly Frost-Prediction System Optimized for Orchards Using Automatic Weather Station and Digital Camera Image Data. *Atmosphere* **12**(7), 846.
- Noh, I., S.-J. Lee, S. Lee, S.-J. Kim, and S.-D. Yang, 2021: A High-Resolution (20 m) Simulation of Nighttime Low Temperature Inducing Agricultural Crop Damage with the WRF-LES Modeling System. *Atmosphere* **12**(12), 1562.
- Robinson, C. and N. Mort, 1996: A neural network solution to the problem of frost Prediction. *UKACC International Conference on Control. Control.* 136-139.
- Temeyer, B. R., W. A. Gallus Jr, K. A. Jungbluth, D. Burkheimer, and D. McCauley, 2003: Using an artificial neural network to predict parameters for frost deposition on Iowa bridgeways. *Proceedings of the 2003 Mid-Continent Transportation Research Symposium*, Iowa state University, Ames, IA, 2003.