



기계학습을 활용한 오리사 바닥재 수분 발생량 분석

Estimation of Duck House Litter Evaporation Rate Using Machine Learning

김다인^a · 이인복^{b,†} · 여옥현^c · 이상연^c · 박세준^d · 크리스티나^e · 김준규^d · 최영배^e · 조정화^e · 정효혁^f · 강솔뢰^f
Kim, Dain · Lee, In-bok · Yeo, Uk-hyeon · Lee, Sang-yeon · Park, Sejun · Cristina Decano · Kim, Jun-gyu ·
Choi, Young-bae · Cho, Jeong-hwa · Jeong, Hyo-hyeog · Kang, Solmoe

ABSTRACT

Duck industry had a rapid growth in recent years. Nevertheless, researches to improve duck house environment are still not sufficient enough. Moisture generation of duck house litter is an important factor because it may cause severe illness and low productivity. However, the measuring process is difficult because it could be disturbed with animal excrements and other factors. Therefore, it has to be calculated according to the environmental data around the duck house litter. To cut through all these procedures, we built several machine learning regression model forecasting moisture generation of litter by measured environment data (air temperature, relative humidity, wind velocity and water contents). 5 models (Multi Linear Regression, k-Nearest Neighbors, Support Vector Regression, Random Forest and Deep Neural Network). have been selected for regression. By using R-Square, RMSE and MAE as evaluation metrics, the best accurate model was estimated according to the variables for each machine learning model. In addition, to address the small amount of data acquired through lab experiments, bootstrapping method, a technique utilized in statistics, was used. As a result, the most accurate model selected was Random Forest, with parameters of n-estimator 200 by bootstrapping the original data nine times.

Keywords: Duck house; litter; machine learning; regression model; water generation

1. 서론

국내 축산업 총생산액은 지난 54년간 연평균 12.2% 증가하여, 2019년 기준 19조 7,819억 원으로 농림업 총생산액의 37.9%를 차지하고 있다 (Statistics Korea, 2020). 그중 오리 산업은 2000년대 중반 오리고기의 영양학적 효능이 알려지면서 생산과 소비가 늘어나 2020년 오리 사육 가구는 566가구, 전

체 마릿수는 929만 마리로 연간 약 1조 원의 생산액을 기록하고 있다. 하지만 이후 주기적으로 발생하는 AI (Artificial Intelligence, 인공지능)에 따른 수급 불균형과 가격 하락 등의 불황을 겪으며 추락세를 보이게 되었다. 또한, 오리 축사는 방역에 취약하고 시설의 단열, 보온 등이 열악한 비닐하우스형이 76.3%를 차지하여 생육 환경을 개선하기 위한 연구가 필요한 실정이다 (Korean duct society, 2019).

국내 오리사 관련 연구는 오리사 바닥 형태와 난방 방법에 따른 오리의 생산성 및 사육환경에 관한 연구 (Bang et al., 2010; Bang et al., 2013), CFD를 이용한 오리사 표준 설계도 초안에 대한 환기 분석 (Yeo et al., 2015; Kim et al., 2017), 사육 밀도에 따른 생산성 및 균일도 평가 (Hong et al., 2019), 오리사 바닥재 주변 공기의 기상조건에 따른 바닥재 수분 증발량 분석 (Lee et al., 2019), 기계환기식 오리사 온습도 환경 분석을 위한 동적 에너지 모델개발 (Lee et al., 2020), ICT 접목을 위한 오리사 바닥 깔짚 살포 자동화 장치 개발 (Kwon et al., 2021) 등으로, 축산 규모와 비교하면 연구의 진행 속도가 매우 더딘 편이다.

가금류 시설 관리의 주요 요인 중 하나인 바닥재 수분 환경은 시설 내부의 습도에 영향을 미치며, 난방 시에도 잠열로 이용되는 등 난방 부하를 증가시킬 수 있다. 또한, 생산성 및 면역력, 질병 발생 등에 큰 영향을 줄 수 있는 요인임에도 오리사

^a Undergraduate Student, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

^b Full professor, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

^c Research assistant professor, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

^d Researcher assistant professor, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

^e Ph.D. Student, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

^f M.S. Student, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University

[†] Corresponding author

Tel.: +82-2-880-4586, Fax: +82-2-873-2087

E-mail: iblee@snu.ac.kr

Received: October 07, 2021

Revised: October 29, 2021

Accepted: November 08, 2021

바닥재에 관한 연구는 국내뿐 아니라 국외 연구의 경우도 미미한 실정이다. 오리사와 유사한 환경인 육계사를 대상으로 진행된 연구는 깔짚의 보관, 취급, 유동 특성이 바닥재의 습도 변화에 미치는 영향 분석 (Bernhart and Fasina, 2009), 육계사 내부 급수기의 위치에 따른 함수량에 대한 샘플링 및 각 측정 지점에서의 함수비와 미생물량 산정 (Wadud et al., 2012), 계사 바닥재의 수분 첨가, 증발, 수분 보유량 정량적 분석 (Dunlop et al., 2015), 환경 요인에만 초점을 맞추어 젖은 바닥재가 육계 생산량이 미치는 다차원적 요인 분석 (Dunlop et al., 2016) 등이 수행되었다. 하지만 오리사 바닥재는 육계사와 비교하였을 때, 물을 좋아하는 오리의 생체 특성이 반영된 생육 환경을 필요로 하므로 생육 환경관리에 명백한 차이가 존재한다. 또한, 국내 기후를 고려한 오리사 내부 환경에 따른 바닥재 수분 발생량 연구가 전무하여 이에 관한 연구가 필요한 실정이다.

이에, 최근 농업 분야에서는 실시간 관리 감독의 한계를 극복하고 원격으로 현장을 감시·관리하는 스마트팜에 관한 연구가 활발히 진행 중이다. Lee et al. (2017); Lee et al. (2018)은 수집한 돈사의 환경 및 음수량 데이터를 사용하여 급수량을 예측하는 기계학습 회귀모델들의 정확도를 비교·분석하였고, 축사에서 목걸이형 및 목 부착형 센서를 이용하여 젖소의 생체정보를 수집하고 축사 환경을 종합하여 다양한 기계학습 모델에 적용하였다. 국외 연구의 경우, Gorczyca et al. (2018)은 환경 데이터를 수집해 돼지의 심부, 표피 및 모피 온도를 예측하였고, Gorczyca (2019)는 젖소의 생리학적 반응에 각 환경 조건이 영향을 미치는 정도를 분석하였다. Wang et al. (2018)은 기계학습을 이용하여 돈사 내부 기상을 실시간 모니터링하여 관리·제어하고 생육 상태를 분석하는 모델의 정확도를 분석하였다.

오리사 바닥재에서 발생하는 수분량은 가축 생육 시 배출하는 분뇨의 수분과 섞여 실시간으로 모니터링하는 것이 까다롭다. 따라서 기계학습 회귀분석 모델을 이용해 오리사 환경 조건과의 관계를 파악하면 오리사 바닥재 수분 발생량을 실시간으로 예측할 수 있다. 본 연구의 선행 연구인 Lee et al. (2019)에서 활용한 다중 선형 회귀분석은 독립변수와 종속변수가 선형 관계에 가까울수록 정확도가 높고, 이상치 (outlier)에 취약하다는 한계점을 가지고 있다. 따라서 본 연구에서는 다양한 기계학습 회귀분석 모델을 이용하여 오리사 바닥재 주변 공기의 환경 조건에 따른 바닥재 수분 발생량 데이터를 분석하고자 한다.

본 연구에서는 여러 기계학습 회귀분석을 이용하여 오리사 바닥재 주변 공기의 온도, 습도, 풍속 및 함수비 등 실시간 모니터링이 가능한 변수들을 바탕으로 바닥재 수분 발생량을 예

측하는 모델을 제안했다. 오리의 수나 오리과 관련된 변수는 바닥재 함수비에 영향을 미치고, 이를 통해 바닥재 수분 발생량에 영향을 미칠 것으로 판단되어 조금 더 직접적인 인자인 바닥재 함수비를 활용하였다. 우선, 오리사 내부 환경 조건에 따른 바닥재의 수분 발생량을 측정하기 위한 실험 챔버를 제작하여 회귀분석에 사용할 기초적인 실험데이터를 얻었다. 다음으로 이 데이터를 기계학습에 적용하였을 때 신뢰할 수 있을 만한 정확도를 가지도록 데이터를 증폭시켰고, 최종적으로 대표적인 회귀분석 기계학습 모델인 MLR (Multi Linear Regression), kNN (k Nearest Neighbor regression), SVR (Support Vector machine Regression), RF (Random Forest), DNN (Deep Neural Network)에 학습시켜 모델별 정확도를 비교·분석하였다.

II. 재료 및 방법

1. 오리사 바닥재 수분 발생량 데이터

본 연구에서는 오리사 바닥재 주변 공기의 온도, 습도, 풍속, 바닥재 함수비에 따른 오리사 바닥재 수분 발생량을 측정한다. Lee et al. (2019)에서 수집한 데이터를 기초로 오리사 바닥재 수분 발생량을 회귀 분석하였다.

실험대상 오리사 바닥재를 채취한 오리사는 전라남도 영암군 신북면 월지리에 위치한 농장으로, 음압식 강제환기를 하는 무창 오리사이다. 오리사의 규격은 폭 12 m, 동고 4 m, 측고 3 m, 길이 45 m이다. 채취한 오리사 바닥재의 수분 발생량을 측정하기 위해 실험 챔버를 제작하고, 실험을 진행하였다. 제작된 챔버를 온도와 습도 조절이 가능한 항온항습조 안에 두고 오리사 바닥재에서 발생하는 수분을 측정하였다. Fig. 1에 선정된 오리사와 제작한 실험 챔버의 모식도를 나타내었다.

각 실험 조건은 오리사 생육환경을 고려하여 Table 1과 같이 온도 3가지 (15, 25, 35 °C), 습도 3가지 (40, 60, 80%), 풍속 2가지 (1.8, 2.3 m/s), 바닥재 함수비 3가지 (10, 35, 60%)이다. 측정된 결과를 환산하여 최종적으로 바닥재에서 하루 동안 발생하는 단위면적당 수분량 ($L/m^2 \cdot day$) 결과를 도출하였다.

Table 1 Experimental conditions for measuring litter water generation according to environmental conditions

Condition	Values	Number of cases
Air temperature	15, 25, 35°C	3
Relative humidity	40, 60, 80%	3
Wind velocity	1.8, 2.3 m/s	2
Litter water contents	10, 35, 60%	3

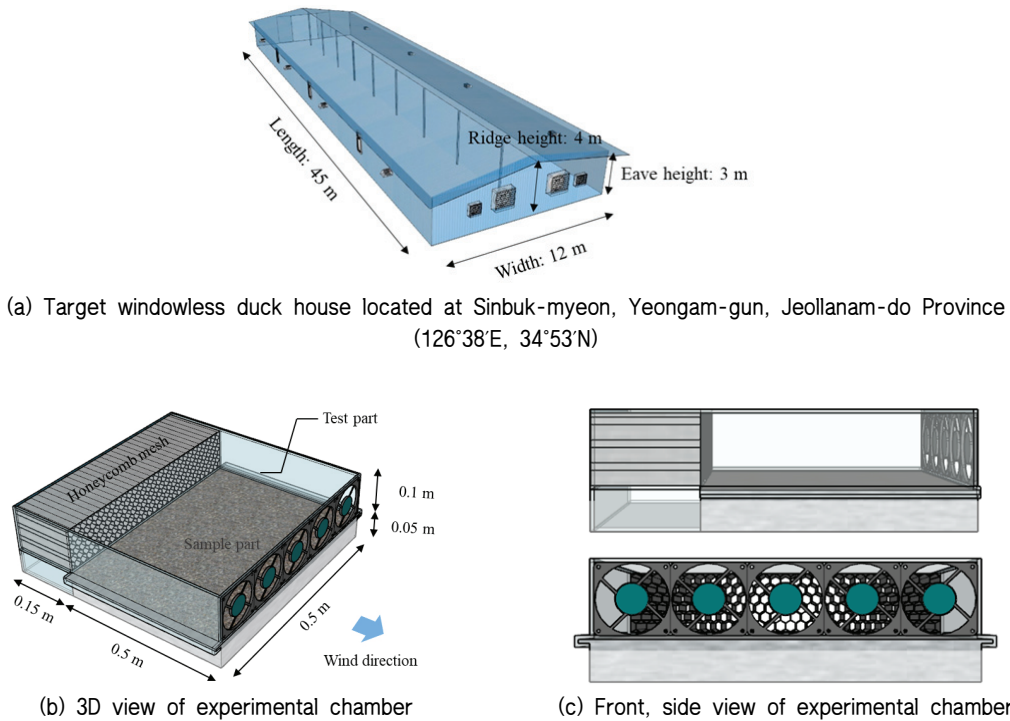


Fig. 1 Schematic view of target duck house and experimental chamber for litter moisture generation measurement

2. 오리사 바닥재 수분 발생량 데이터 분석 모델

농업 분야에서 다양한 환경 조건에 따라 내부 생육 환경을 관리, 예측하는 기계학습 회귀분석 모델을 활용한 선행연구를 참고하여 모델별 장단점, 활용 사례와 특성을 분석하였다. 오리사 바닥재 주변 공기의 환경 조건인 온도, 습도, 풍속, 바닥재 함수비에 따른 바닥재 수분 발생량을 예측하기 위해 선정된 모델은 MLR, kNN, SVR, RF, DNN이다. Table 2에 각 모델의 장단점을 나타내었다.

단순 회귀분석을 확장한 MLR과 다수의 의사결정 나무가 모여 복원추출을 통해 각각의 단일 모델을 형성하고 예측된 결과를 투표나 평균하는 RF 등은 토마토 스마트팜 생육량 및

생산량의 예측 정확도를 비교 (Hong et al., 2020)하기 위해 사용되었다. kNN은 주어진 표본 집합 내에서 특정 데이터와 가장 가까이 위치한 k개의 값을 이용하여 해당 값을 예측하고 분류하는 알고리즘으로, 평균 기온과 강수량에 따른 작물 수확 시기 예측을 위해 활용되었다 (Shakoor et al., 2017). 두 부류 사이에 존재하는 여백을 최대화하여 일반화 능력을 극대화하는 SVM을 회귀분석에 적용할 수 있도록 확장한 SVR은 온실 내부 환경에 따른 토마토 생장과 수확율을 예측하였고 (Alhnaity et al., 2020), 입력층, 은닉층, 출력층에 존재하는 노드의 변화하는 값과 가중치의 곱을 합산하여 활성화 함수로 값을 도출하는 DNN은 창문 자동 개폐, 온도 예측을 통해 온실

Table 2 Properties of each machine learning regression model

Model	Advantage	Disadvantage
MLR	Calculates the correlation coefficient between variables Usable in variable fields	Reliability concerns of relationship between each independent variable
kNN	Intuitive and relatively robust Simple algorithm	Time consuming with large data Sensitive to the scale of the data
SVR	Able to avoid overfitting Insensitive to outliers	Time consuming with large data Sensitive to outliers
RF	Avoid over-fitting of decision tree Flexible to both classification and regression problems	Time consuming with a lot of decision trees
DNN	High accuracy with large amount of data	Relies on the selection of model's various hyper-parameters

내 온도 제어 단점을 해결하였다 (Shin et al., 2018).

3. 연구 방법

본 연구에서는 Lee et al. (2019)에서 수집한 데이터를 기계 학습에 적용하기 위해 데이터를 증폭하는 부트스트래핑과 변수 간 독립성을 확인하는 다중 공선성을 이용해 가공하는 과정을 거친 후, 앞서 선정된 다섯 가지 기계학습 모델을 이용하여 회귀 분석하였다. 분석의 정확도를 높이기 위해 공통 변수인 test size, bootstrapping amount와 size, 및 모델별 변수를 수정해가며 세 가지 평가 지표 (R^2 , RMSE (Root Mean Square Error), MAE (Mean Absolute Error))를 종합적으로 비교한 뒤

최적 모델을 설계하였다.

가. 학습 데이터 가공

1) Bootstrapping

선행연구의 랩 실험에서 얻은 데이터는 54개로, 정확도 높은 학습 모델 구현을 위해서 충분하지 않다. 따라서 부트스트래핑 기법을 이용하여 데이터를 증폭한 후 회귀분석 모델에 적용하였다. 부트스트래핑은 통계학에서 사용되는 기법으로, 가설 검증이나 평가 지표 (metric)를 계산하기 전에 복원추출 (random sampling)을 적용하는 방법이다 (Efron et al., 1998). 기존 데이터에서 N개의 데이터를 복원추출하여 평균한 값을

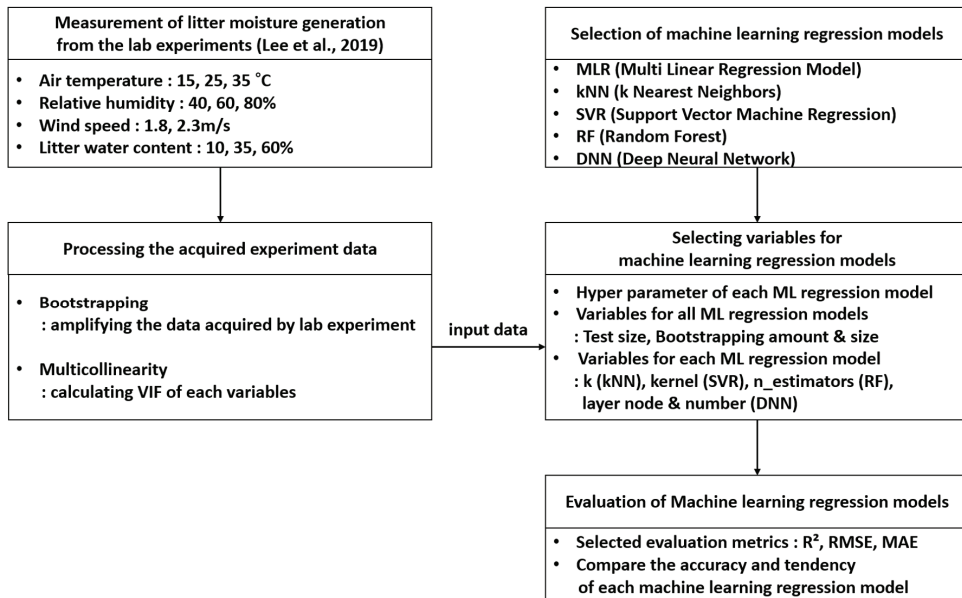


Fig. 2 Research flow to estimate duck litter moisture generation using machine learning regression model

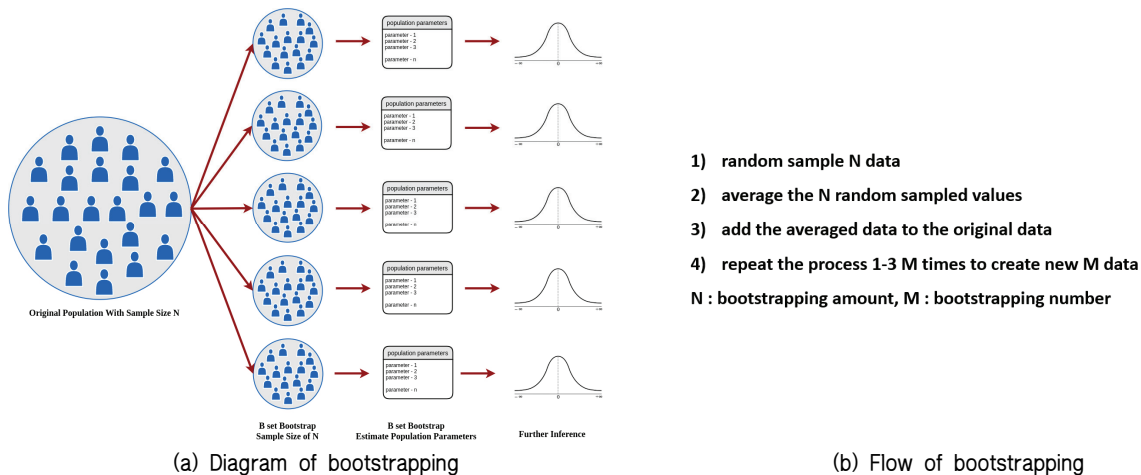


Fig. 3 Schematic processes of bootstrapping to amplify the number of data

새로운 데이터로 쌓고, 이 과정을 여러 번 반복하여 기존의 데이터를 M배 증폭시킨다. Fig. 3은 부트스트래핑의 과정을 나타낸다. 평균을 이용한 부트스트래핑은 수집한 데이터의 독립 변수와 종속변수가 선형적일 때 회귀분석에서 가장 유의미한 결과를 도출할 수 있으므로 수집한 데이터가 선형 관계를 갖는지 확인하는 과정을 거쳤다. 본 연구의 기계학습 회귀분석 모델의 정확도를 높이기 위해, 부트스트래핑 시 사용하는 복원추출 데이터의 개수 및 증폭한 새로운 데이터의 개수를 일정 범위에서 바꾸어가며 학습 모델의 정확도를 비교하였다.

2) 다중 공선성

회귀분석은 각 변수가 서로 독립이라는 전제를 바탕으로 시행되어야 하므로 변수 간 독립성을 확인하기 위하여 다중 공선성을 활용하였다. 다중 공선성은 회귀분석에서 사용된 모형의 일부 예측 변수가 다른 예측 변수와 상관 정도가 높아, 데이터 분석 시 부정적인 영향을 미치는 현상을 분석한다. 다중 공선성 평가에 가장 많이 활용되는 인자는 VIF (분산 팽창 인수, Variance Inflation Factor)로, 1과 유사할수록 모형을 신뢰할 수 있으며, 10 이상이면 변수 선택에 대하여 재고려할 필요성이 대두된다. VIF의 계산식은 식 (1)과 같다. R²은 실측치와 예측치 사이의 선형 관계를 정량적으로 나타내는 지표인 결정계수이다.

$$VIF = \frac{1}{1 - R^2} \tag{1}$$

나. 회귀분석 모델 변수 선정

기계학습 회귀분석에서는 각 하이퍼 파라미터를 적절하게 선정하여 모델의 정확도를 높이는 것이 중요하다. 각 회귀분석 모델은 모델링 과정에서 설정할 수 있는 하이퍼 파라미터가 존재하는데, 이를 수정함으로써 모델의 정확도를 개선할 수 있다. 데이터의 형태에 따라 최적 파라미터가 다르므로, 시행착오를 거쳐 하이퍼 파라미터를 선정해야 한다.

SVR의 경우 대표적인 하이퍼 파라미터인 ε (epsilon)은 데이터와 경계 함수 사이의 거리로, 0과 1 사이의 값을 가진다. DNN의 학습률, 최적화 함수에 사용되는 beta 1, 2, 정규화 과정에서 분모가 0이 되는 것을 방지하기 위한 ε은 0과 1 사이의 값을 가진다. 해당 하이퍼 파라미터들은 시행착오를 통해 가장 높은 정확도를 가지는 값으로 선정하였다. 활성화 함수 (ReLU), 최적화 함수 (Adam), 평가 지표 (MSE, Mean Square Error)는 최근 연구에 가장 많이 사용되는 파라미터를 선정하였다. 기계학습 회귀분석 모델에 최적화된 하이퍼 파라미터는 Table 3과 같다.

Table 3 Hyper-parameter of each ML regression model

Model	Parameter	Value
SVR	Epsilon	0.01
	Activation function	ReLU
	Loss	MSE
	Optimizer	Adam
DNN	Learning rate	0.01
	Beta_1	0.9
	Beta_2	0.999
	Epsilon	1e-8
	Metrics	MSE
	Epochs	100

최적화된 하이퍼 파라미터 외에도 여러 가지 변수를 선정하여 정해진 범위에 따라 바꾸어가며 모델별 정확도를 비교하였다. 전체 데이터 중 예측에 활용한 데이터의 비율을 나타내는 test size를 10, 20, 30%로 분할하였다. 부트스트래핑의 경우 복원추출 하는 데이터의 개수인 bootstrapping amount를 5개 부터 100개까지 5씩 증가시켰고, 증폭한 데이터의 개수인 bootstrapping size를 기존 데이터 (1배)부터 10배까지 설정하였다. Table 4, 5에 모든 모델에 동일하게 설정한 변수 (test size, bootstrapping amount, bootstrapping size)와 모델별 변수 및 설정 범위를 나타내었다.

모델별 변수의 경우 각각의 대표적인 하이퍼 파라미터인 kNN의 가까이 위치한 데이터의 개수를 의미하는 k를 3에서 12 범위로 조절하였고, SVR의 예측값을 근사하는 추정함수를 설정하는 네 가지 kernel인 linear, radial, sigmoid, poly를 선정하였다. RF는 의사결정 나무의 개수를 결정하는 n_estimators를 5개에서 250개 범위에서 5씩 증가시켰고, DNN의 은닉층은

Table 4 Variables for all machine learning regression models

Variable	Range
Test size	10, 20, 30%
Bootstrap amount	5-100 by 5
Bootstrap size	origin-10 times

Table 5 Variables for each machine learning regression model

Model	Variable	Range
kNN	k	3-12
SVR	Kernel	linear, radial, sigmoid, poly
RF	n-estimators	5-250 by 5
DNN	layer node	2 ⁿ (n=1-6)
	layer num	1-6

1개에서 6개, 노드 개수는 2개부터 64개까지 두 배씩 조절하며 각 모델을 학습하는 과정을 거쳤다.

다. 예측 모델 성능평가

기계학습 회귀분석 예측 모델의 정확도를 평가하기 위해 대표적인 평가 지표인 R² (determination coefficient), RMSE (Root Mean Square Error) 및 MAE (Mean Absolute Error)를 선정하였다. R²은 결정계수라고 불리는데, 실측치와 예측치 사이의 선형 관계를 정량적으로 나타내는 지표로서 여러 방면에서 널리 이용된다. RMSE는 예측치에 대한 단위를 포함하고 있는 평가 지표로서, 오차의 정량적 평가가 가능하다. MAE의 경우 오차의 절대적인 정도의 정량적 평가를 나타내는 지표로, RMSE보다 사용 빈도가 적지만 비교적으로 이상치의 영향을 덜 받는다.

각 평가 지표에 관한 설명과 식은 Table 6에 나타내었다 (Akinoglu et al., 1991; Bakay et al., 2021; Bakirci, 2009; Gouda et al., 2019; Wang et al., 2018). x_i와 y_i는 각각 측정 및 예측 데이터값이고, \bar{x}_i 는 측정 데이터의 평균, n은 데이터 개수이다.

III. 결과 및 고찰

1. LAB 실험을 통한 데이터 구축

챔버 실험을 통한 다양한 환경 조건에 따른 오리사 바닥재

수분 발생량 측정 결과, 공기 온도, 바닥재 함수비, 풍속이 높을수록, 상대 습도가 낮을수록 오리사 바닥재 수분 발생량이 선형적으로 증가하는 경향이 나타났다. 따라서 평균을 이용한 부트스트래핑을 적용할 수 있음을 확인하였다. 오리의 성장에 따라 바닥재에 니플 음수 유출이나 오리 분뇨 누적에 의해 일반적으로 함수비가 지속적으로 증가하게 되며, 이에 따라 바닥재의 수분 증발량도 증가한 것으로 판단된다.

또한, 측정값의 타당성을 판단하기 위해 계사 바닥재 수분 발생량 측정 선행연구와 비교하였다 (Table 7). 공기 온도 25°C, 상대 습도 50%, 바닥재 함수비 60%, 풍속 2 m/s 조건에서 계사 바닥재의 수분 발생량은 약 10 L m² day⁻¹이고 (Dunlop et al., 2015), 본 연구에서 기타 조건은 동일하고, 상대 습도가 60%, 풍속 2.3 m/s인 오리사 바닥재 수분 발생량은 약 13.4 L m² day⁻¹로 측정되었다. 주변 풍속과 상대 습도 및 바닥재 종류의 차이 등을 고려하였을 때 타당한 결과가 도출되었다고 판단하였다.

2. 다중 공선성 평가

모델별로 최적값을 보인 데이터는 test size, 부트스트래핑 변수 등에 의해 조금씩 차이가 있다. 모델별 선정한 데이터의 분산 팽창 인수 (VIF) 계산 결과를 Fig. 4에 나타내었다. 모든 데이터의 변수별 VIF 값이 모두 1과 유사한 값으로 산정되었고, 이에 따라 모형의 예측 변수 간 상관도가 데이터 분석에 부정적인 영향을 주지 않아 오리사 바닥재 수분 발생량 예측

Table 6 Equation, description and evaluation of metrics

Metrics	Description	Performance criteria
$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2}$	R ² is the most used metric for the researchers all time. It gives an idea about how well a model can forecast a set of actual data.	The closer R ² is to 1, the better it reflects the forecast results.
$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$	RMSE is often preferred for evaluating short-term forecasting results. It represents the difference between actual and forecasted data.	RMSE takes positive values in all cases and being closer to zero, the more successful the forecast results gets.
$MAE = \frac{1}{n} \sum_{i=1}^n y_i - x_i $	MAE measures the goodness of the fit used to obtain the models.	The closer the MAE value to zero, the more successful the forecast results.

Table 7 Comparison between result of Dunlop et al., 2015 and Lab experiment

Condition	Air temperature (°C)	Relative humidity (%)	Wind velocity (m/s)	Litter water contents (%)	Litter water generation (L m ⁻² day ⁻¹)
Dunlop et al., 2015	25	50	2	50	10
Lab experiment	25	60	2.3	50	13.4

Features	VIF	Features	VIF	Features	VIF
Air temperature	1.0505	Air temperature	1.0308	Air temperature	1.0257
Relative humidity	1.2210	Relative humidity	1.0418	Relative humidity	1.0437
Wind velocity	1.0404	Wind velocity	1.0232	Wind velocity	1.0241
Water contents	1.0012	Water contents	1.0015	Water contents	1.0016

(a) Original data

Features	VIF	Features	VIF	Features	VIF
Air temperature	1.0481	Air temperature	1.0535	Air temperature	1.0470
Relative humidity	1.0235	Relative humidity	1.0226	Relative humidity	1.0257
Wind velocity	1.0373	Wind velocity	1.0416	Wind velocity	1.0405
Water contents	1.0012	Water contents	1.0012	Water contents	1.0012

(d) SVR

Features	VIF	Features	VIF	Features	VIF
Air temperature	1.0481	Air temperature	1.0535	Air temperature	1.0470
Relative humidity	1.0235	Relative humidity	1.0226	Relative humidity	1.0257
Wind velocity	1.0373	Wind velocity	1.0416	Wind velocity	1.0405
Water contents	1.0012	Water contents	1.0012	Water contents	1.0012

(e) RF

Features	VIF	Features	VIF	Features	VIF
Air temperature	1.0505	Air temperature	1.0308	Air temperature	1.0257
Relative humidity	1.2210	Relative humidity	1.0418	Relative humidity	1.0437
Wind velocity	1.0404	Wind velocity	1.0232	Wind velocity	1.0241
Water contents	1.0012	Water contents	1.0015	Water contents	1.0016

(c) kNN

Features	VIF	Features	VIF	Features	VIF
Air temperature	1.0505	Air temperature	1.0308	Air temperature	1.0257
Relative humidity	1.2210	Relative humidity	1.0418	Relative humidity	1.0437
Wind velocity	1.0404	Wind velocity	1.0232	Wind velocity	1.0241
Water contents	1.0012	Water contents	1.0015	Water contents	1.0016

(f) DNN

Fig. 4 VIF factor of each machine learning regression model

을 위해 선택한 오리사 바닥재 주변 공기의 기상자료 선택이 적절함을 확인할 수 있었다. 즉, 독립변수인 온도, 습도, 바닥재 함수비, 풍속이 모두 오리사 바닥재 수분 발생량 예측을 위해 적절함을 검증하였다.

3. 오리사 바닥재 수분 발생량 예측 모델 평가

앞서 설정한 변수를 토대로 각 모델의 test size, 부트스트랩핑 및 모델별 변수를 수정해가며 각 회귀모델을 구동하고, 모델별 정확도가 가장 높게 나온 오리사 바닥재 수분 발생량 회귀분석 모델의 평가 지표를 변수와 함께 Table 8에 나타냈다.

5개 모델의 성능을 비교한 결과, RF 모델의 R²이 1에 가장 가까웠고, MLR, SVR, kNN, DNN 순서로 나타났다. RMSE와 MAE의 경우 RF, MLR, kNN, SVR, DNN 순으로 0에 가까운 값을 가졌다. 이를 토대로 오리사 바닥재 주변 공기 데이터를 바탕으로 RF를 활용하여 오리사 바닥재 수분 발생량을 예측하였을 때의 정확도가 가장 높고, 다음은 MLR인 것을 확인할 수 있다. 분석에 사용된 오리사 수분 발생량 데이터의 개수가 적었기 때문에 bagging 기법을 이용하여 데이터를 증폭하는 RF가 가장 좋은 성능을 보인 것으로 판단된다.

가. 학습 데이터 수와 부트스트랩핑 최적화

모든 모델에 대해 공통적으로 적용한 변수인 test size, bootstrapping amount, bootstrapping size에 따른 모델별 예측 정확도 평가 지표를 비교하였다. Fig. 5에서 볼 수 있듯, test size는 10, 20, 30% 중 10%를 선정하였을 때 모든 모델에서

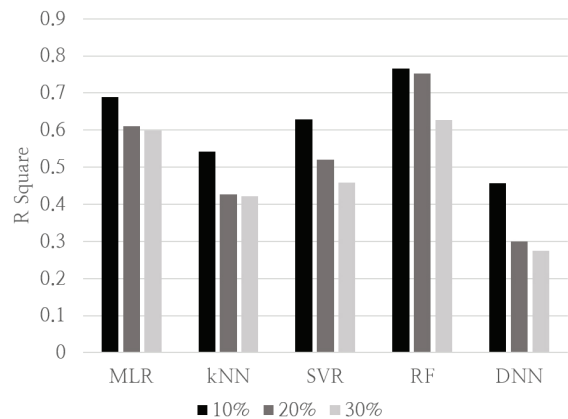


Fig. 5 R² of each machine learning regression model according to test size

Table 8 Performance of Machine Learning Regression Models

Factors	MLR	kNN	SVR	RF	DNN
Test size (%)	10	10	10	10	10
Bootstrap amount	90	65	10	80	95
Bootstrap size	X5	X5	X8	X9	X5
ML variable	-	7	linear	200	5 layer (16 node)
R ²	0.6689	0.5411	0.6284	0.7657	0.4561
RMSE	0.0079	0.0082	0.0130	0.0046	0.0195
MAE	0.0056	0.0059	0.0065	0.0031	0.0129

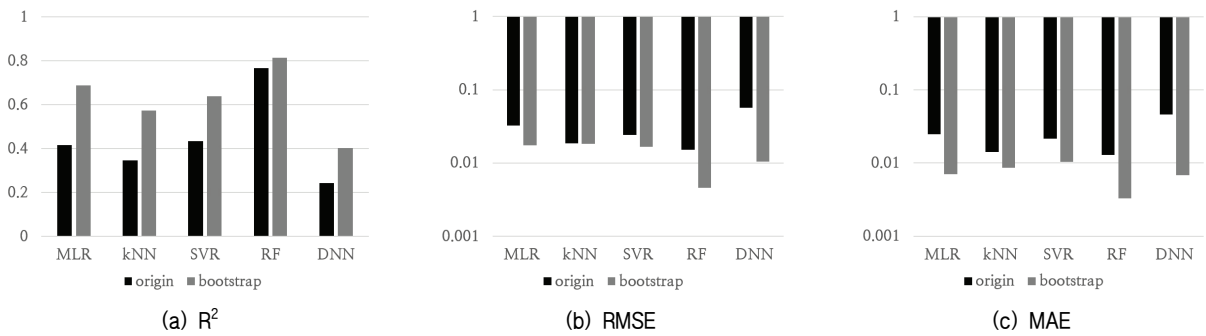


Fig. 6 Evaluation metrics (R^2 , RMSE, MAE) of each machine learning regression model according to bootstrapping status

가장 예측 정확도가 높았다. 랩 실험을 통해 수집한 데이터 개수가 적어 훈련 데이터가 많을수록 정확도가 높은 것을 확인할 수 있었다.

Fig. 6에서 부트스트래핑 여부에 따른 모델별 예측 정확도 지표를 비교하였는데, 모든 모델과 지표에서 부트스트래핑을 사용한 모델이 신뢰도가 높다는 것을 확인할 수 있었다. 또한, 부트스트래핑의 추출 데이터 개수와 데이터의 크기에 대한 명확한 경향은 나타나지 않았지만, 5개의 모델의 예측 결과 모두에서 부트스트래핑을 5~9배 규모로 시행하였을 때의 정확도가 높은 것으로 확인되어 부트스트래핑의 필요성과 효율성을 확인할 수 있었다. 따라서 랩 실험을 통해 얻은 50개의 데이터를 가공하기 위해 부트스트래핑을 활용하는 것이 효율적이라는 것을 확인할 수 있었다.

나. 모델별 시험·예측 데이터 산포도

모델별로 입력한 데이터 값의 차이로 인한 오차 발생 여부를 확인하기 위해 공통 변수인 test size, bootstrapping amount 와 size를 모든 모델에 동일하게 적용하여 모델별 산포도 분포를 비교하였다. 산포도 분포표의 경우 실측값과 예측값의 경향성을 한눈에 확인할 수 있으며, 본 데이터는 Lab 실험을 거쳐 구축하였기 때문에 현장에 적용하면 오리 여부 등으로 인해 나타나는 비교적 균일한 오차값을 확인할 수 있다. 이때, 오차가 발생한 데이터의 독립변수와 유사한 범위에 존재하는 데이터를 재구성할 필요성이 대두된다.

해당 데이터는 test size 10%, bootstrap size 5, bootstrap amount 90을 적용하여 구축하였다. 각 기계학습 회귀분석 모델별 시험 데이터와 예측 데이터의 산포도 분포표를 Fig. 7을 통해 확인할 수 있다. 해당 그래프의 x축은 랩 실험을 통해 얻

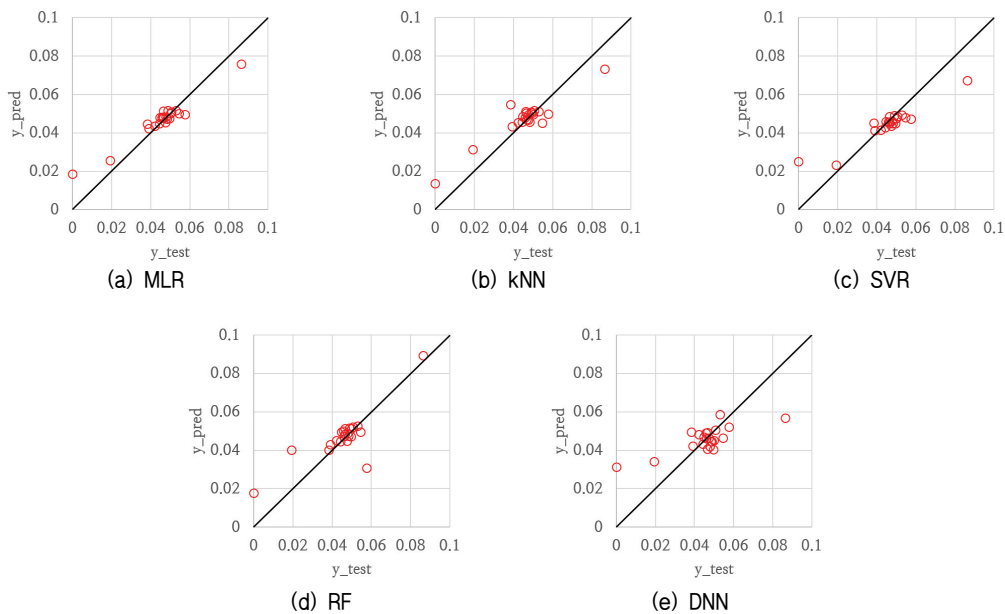


Fig. 7 Dispersion of test and prediction values of each model

Table 9 Evaluation metrics of SVR and kNN

Regression model	R ²	RMSE	MAE
kNN	0.5411	0.0082	0.0059
SVR	0.6284	0.0130	0.0065

은 오리사 바닥재 수분 발생량의 측정값이고, y축은 각 회귀분석 모델을 이용해 예측한 값을 나타낸다. 선형 관계에서 멀어질수록 시험 데이터와 예측 데이터 간의 오차가 커 모델의 예측 정확도가 떨어지게 된다.

해당 산포도 표를 통해 앞서 산정한 모델별 예측 정확도 평가 지표의 경향에 대하여 설명할 수 있다. 특히, Table 9에서 평가 지표 R²과 RMSE, MAE의 예측 정확도가 상이하게 나타난 SVR, kNN을 비교하면 SVR의 예측치와 실측치의 전체적인 선형 관계는 비슷하였지만, 일부 값들에서 예측치와 실측치의 오차가 커 R²이 높은 반면 RMSE, MAE가 큰 값을 보였고, kNN은 타 모델과 비교해 오차는 작고 선형 관계는 낮아 R²이 낮고 RMSE, MAE가 낮게 산출되었다.

해당 결과를 바탕으로 오리사 바닥재 수분 발생량 예측에 필요한 모델을 선정할 때 전체적인 경향성이 좋은 모델과 오차가 적은 모델을 비교할 수 있다. 따라서 예측 모델의 성능 비교를 위해서는 다양한 지표를 이용하여 정확도를 비교한 뒤 가장 정확한 모델을 선정해야 한다.

다. 모델별 변수 최적화

모델별 변수에 따른 회귀분석 모델의 정확도를 비교하였다.

kNN의 경우 k = 7, 즉 예측값과 이웃한 데이터의 개수를 7로 선정하였을 때, SVM의 경우 kernel = linear, 즉 예측값을 근사하는 커널 함수가 선형일 때 가장 높은 정확도를 보였다. RF의 경우 n_estimator = 210, 즉 의사결정 나무의 개수가 210개일 때, DNN의 경우 16 node의 5 layer, 즉 16개의 노드를 가지는 5개의 은닉층을 설계하였을 때 가장 높은 정확도를 보였다. 노드 개수가 32, 64인 경우에 R²이 음수인 결과가 나왔는데, 이는 편차의 제공보다 오차의 제공이 더 큰 상황으로, 회귀분석 모델이 평균을 통해 예측하는 모델보다 예측 성능이 낮음을 나타낸다. Fig. 8은 각 모델의 변수에 따른 모델의 정확도를 나타낸 그래프이다.

공통 변수, 모델별 변수, 평가 지표 및 산포도 분포를 종합적으로 살펴보았을 때, 랩 실험을 통해 구축한 오리사 바닥재 주변의 기상자료에 따른 바닥재 수분 발생량 데이터 회귀분석을 위해서는 RF (Random Forest, 랜덤 포레스트)가 가장 적절한 모델로 선정되었다. 이때, 시험 데이터 10%를 적용하고, bootstrap size 5배 이상, n_estimator 200개 이상으로 설정하였을 때 상관계수 (R²)가 가장 1과 가까워 실측치와 예측치의 선형 관계가 가장 유사하였고, RMSE와 MAE가 모두 0과 유사해 실측치와 예측치의 오차가 가장 적음을 확인할 수 있었다.

본 연구에서 오리사 바닥재 주변 환경 조건에 따른 수분 발생량 분석을 위해 수집한 데이터들이 대개 선형 관계를 나타내고 있어 선행연구 Lee et al. (2019)에서 수행한 다중 선형 회귀분석과 유사한 상관계수 (R²)를 보였다. 구축한 기초 데이터가 많아 이상치 (outlier)에 민감한 다중 선형 회귀분석의 한

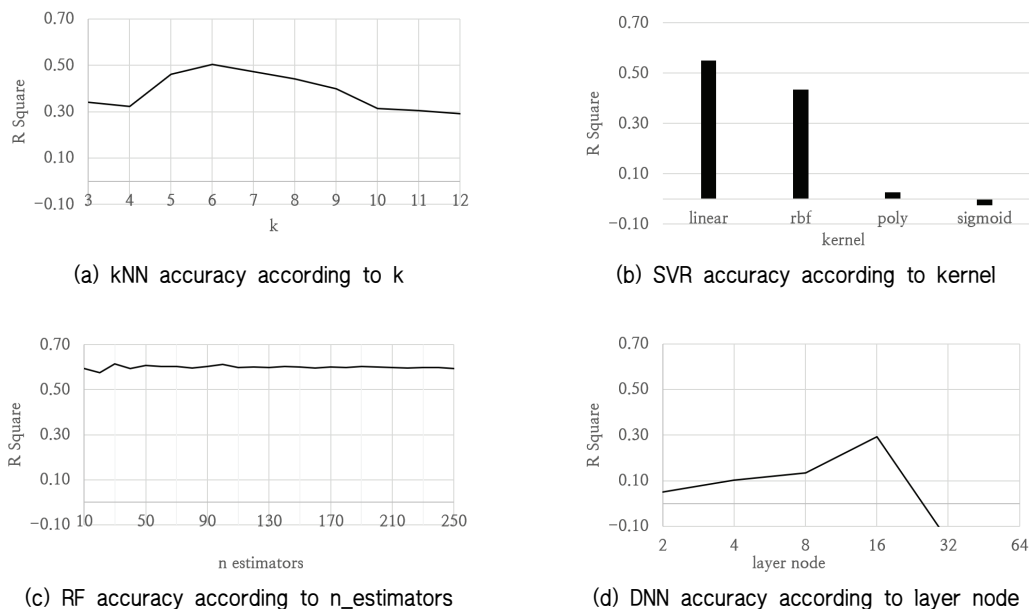


Fig. 8 Accuracy changes according to variables of each model

계점이 나타나지 않았음이 우려된다. 특히, RMSE와 MAE를 함께 고려하는 과정을 거쳐 실측치와 예측치의 경향성뿐만 아니라 이상치에 대한 오차를 확인할 수 있었다. 추가적인 실험과 현장 실측을 통해 데이터의 범위를 확장하여 데이터의 독립변수와 종속변수 간의 선형 관계에서 발생하는 오차가 증가하게 되어 본 연구에서 가장 정확도가 높았던 RF를 적용하는 것이 더 정확한 회귀분석 결과를 도출할 수 있으므로 예상된다.

IV. 결론

본 논문에서는 랩 실험을 통해 구축한 오리사 바닥재 주변 공기의 환경 데이터 (공기 온도, 습도, 풍속, 함수비)에 따른 바닥재 수분 발생량 데이터를 회귀 분석하였다. 오리사 바닥재 수분 발생량은 오리의 생육 관리에 가장 큰 영향을 주는 요인인데, 관련 연구는 국내뿐 아니라 국외 연구도 미미해 오리 농가에서는 계사 연구를 참고하여 설계·관리하는 실정이다. 따라서 실시간 측정이 불가능한 바닥재 수분 발생량을 예측하기 위하여 상대적으로 측정이 쉬운 주변 공기의 환경 자료를 이용하여 바닥재 수분 발생량을 회귀 분석하고자 하였다. 선정한 회귀분석 모델인 MLR, kNN, SVR, RF, DNN을 이용하여 회귀분석을 진행한 후, 세 가지 평가 지표 (R^2 , RMSE, MAE)를 이용하여 정확도를 평가하였다. 또한, 랩 실험을 통해 구축한 데이터의 양이 적다는 점을 해결하기 위해 부트스트랩핑 기법을 이용하여 데이터를 확장하고, 다양한 변수 (test size, bootstrapping variables, 기계학습 모델별 하이퍼 파라미터)에 따른 정확도가 높은 모델을 선정하였다. 분석 결과, R^2 0.7657, RMSE 0.0046, MAE가 0.0031로 RF의 정확도가 가장 높은 것을 확인하였다.

본 논문에서 연구한 내용을 바탕으로 다양한 변수를 적용하여 오리사 바닥재 수분 발생량 데이터에 최적화한다면 더욱 정확도가 높은 모델을 개발할 수 있을 것이다. 또한, 뚜렷한 선형 관계를 보이는 학습 데이터를 회귀 분석하였다는 한계점을 고려하여 추가적인 실험과 현장 실측을 통해 빅데이터를 구축한다면 이상치 (outlier)에 대응이 가능할 것이다. 이후, 개발된 모델을 실제 오리사에 적용하여 현장에서의 적용성을 확인하고, 실시간 오리사 기상자료를 통해 바닥재 수분 발생량을 모니터링한다면 오리사 생육 환경관리 현황을 개선할 수 있을 것이다.

감사의 글

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 농촌현안해결리빙랩프로젝트의 지원을 받아 연구

되었음 (120099-03).

REFERENCES

1. Korean Duck Association, 2019. A report of research on the improvement of duck breeding facilities I. *Duck Village* 195: 24-35 (in Korean).
2. Korean Duck Association, 2019. A report of research on the improvement of duck breeding facilities II. *Duck Village* 196: 20-40 (in Korean).
3. Korean Duck Association, 2019. A report of research on the improvement of duck breeding facilities III *Duck Village* 197: 30-51 (in Korean).
4. Statistics Korea, 2020. Results of the farm and fishery household economy survey in 2020.
5. Statistics Korea, 2020. 2020 Livestock production cost survey.
6. Bang, H. T., H. C. Choi, H. S. Chae, J. C. Na, H. G. Kang, M. J. Kim, D. W. Kim, S. B. Park, S. H. Jung, and O. S. Seo, 2010. A study on the productivity and environmental change according to duck house litter. *Korean Society of Poultry Science* 27: 132-134 (in Korean).
7. Bang, H. T., D. W. Kim, H. B. Jong, J. C. Na, H. K. Kang, M. J. Kim, M. M. H. Mushtaq, H. C. Choi, S. B. Lee, M. Kang, and J. H. Kim, 2013. Effect of various forms of floor system on performance of meat-type duck and environments of duck house. *Korean Journal Society of Poultry Science* 40(3): 253-262 (in Korean). doi:10.5536/KJPS.2013.40.3.253.
8. Yeo, U. H., Y. S. Cho, K. S. Kwon, T. H. Ha, S. J. Park, R. W. Kim, S. Y. Lee, S. N. Lee, I. B. Lee, and I. H. Seo, 2015. Analysis of ventilation efficiency for the summer about a design plan of standard duck house using CFD. *Korean Journal Society of Poultry Science* 57(5): 51-60 (in Korean). doi:10.5389/KSAE.2015.57.5.051.
9. Kim, Y. H., 2017. Analysis of ventilation efficiency of standard duck house using computational fluid dynamics. *Seoul National University*: 1-91 (in Korean).
10. Hong, E. C., B. S. Kang, W. K. Kang, J. J. Jeon, H. S. Kim, J. S. Son, and C. H. Kim, 2019. Effect of different stocking densities in plastic wired-floor house on performance and uniformity of Korean native commercial ducks. *Korean Journal Society of Poultry Science* 46(4): 215-221 (in Korean). doi:10.5336/KJPS.2019.46.4.215.
11. Lee, S. Y., I. B. Lee, R. W. Kim, U. H. Yeo, C. Decano,

- J. G. Kim, Y. B. Choi, Y. M. Park, and H. H. Jeong, 2019. Assessment of evaporation rates from litter of duck house, *Journal of the Korean Society of Agricultural Engineers* 61(5): 101-108 (in Korean). doi:10.5389/KSAE.2019.61.5.101.
12. Lee, S. Y., I. B. Lee, R. W. Kim, U. H. Yeo, J. G. Kim, and K. S. Kwon, 2020. Dynamic energy modeling for analysis of the thermal and hygroscopic environment in a mechanically ventilated duck house, *Biosystems Engineering*, 200: 431-449. doi:10.1016/j.biosystemseng.2020.10.015
13. Kwon, K. S., J. S. Woo, J. H. Noh, S. I. Oh, J. B. Kim, J. K. Kim, K. Y. Yang, D. H. Jang, and S. M. Choi, 2021. Development and field-evaluation of automatic spreader for bedding materials in duck houses. *Journal of the Korean Society of Agricultural Engineers* 63(1): 37-48 (in Korean). doi:10.5389/KSAE.2021.63.1.037.
14. Bernhart, M., and O. O. Fasina, 2009. Moisture effect on the storage, handling and flow properties of poultry litter. *Waste Management* 29: 1392-1398. doi:10.1016/j.wasman.2008.09.005.
15. Brye, K. R., N. A. Slaton, R. J. Norman, and M. C. Savin, 2005. Short-term effects of poultry litter form and rate on soil bulk density and water content. *Communications in Soil Science and Plant Analysis* 35(15 & 16): 2311-2325. doi:10.1081/LCSS-200030655.
16. Choi, I. H., and Nahm K. H., 2004. Effects of applying two different chemical additives to the litter on broiler performance and the carbon dioxide gas production in poultry houses. *Korean Society of Poultry Science* 31(3): 171-176 (in Korean).
17. Wadud, S., A. Michaelsen, E. Gallagher, G. Parcsi, O. Zemb, R. Stuetz, and M. Manefield, 2012. Bacterial and fungal community composition over time in chicken litter with high or low moisture content. *British Poultry Science* 53(5): 561-569. doi:10.1080/00071668.2012.723802.
18. Dunlop, M. W., P. J. Blackall, and R. M. Stuetz, 2015. Water addition, evaporation and water holding capacity of poultry litter. *Science of the Total Environment* 538: 979-985. doi:10.1016/j.scitotenv.2015.08.092.
19. Dunlop, M. W., J. McAuley, P. J. Blackall, and R. M. Stuetz, 2016. Water activity of poultry litter: relationship to moisture content during a grow-out. *Journal of Environmental Management* 172: 201-206. doi:10.1016/j.jenvman.2016.02.036.
20. Kim, S. H., S. Y. Park, S. J. Lee, and K. S. Ryu, 2003. Effect of feeding lactobacillus reuteri to broiler on growing performance, intestinal microflora and environmental factor. *Korean Society of Poultry Science* 30(1): 17-28 (in Korean).
21. Miles, D. M., D. E. Rowe, and T. C. Cathcart, 2011. Litter ammonia generation: moisture content and organic versus inorganic bedding materials. *Poultry Science* 90: 1162-1169. doi:10.3382/ps.2010-01113.
22. Lee, W. S., J. Y. Ryu, T. W. Ban, S. H. Kim, and H. C. Choi, 2017. Prediction of water usage in pig farm based on machine learning. *Journal of the Korea Institute of Information and Communication Engineering* 21(8): 1560-1566 (in Korean). doi:10.6109/jkiice.2017.21.8.1560.
23. Lee, W. S., S. M. Park, T. W. Ban, S. H. Kim, J. Y. Ryu, and K. Y. Sung, 2018. Health monitoring of livestock using neck sensor based on machine learning. *Journal of the Korea Institute of Information and Communication Engineering* 22(11): 1421-1427 (in Korean). doi:10.6109/jkiice.2018.22.11.1421.
24. Lee, W. S., J. Y. Ryu, T. W. Ban, S. H. Kim, S. K. Kang, Y. H. Ham, and H. J. Lee, 2018. Estimation of body core temperature of cow using neck sensor based on machine learning. *Journal of the Korea Institute of Information and Communication Engineering* 22(12): 1611-1617 (in Korean). doi:10.6109/jkiice.2018.22.12.1611.
25. Lee, W. S., K. Y. Sung, T. W. Ban, and Y. H. Ham, 2020. Production performance prediction of pig farming using machine learning. *Journal of the Korea Institute of Information and Communication Engineering* 24(1): 130-133 (in Korean). doi:10.6109/jkiice.2020.24.1.130.
26. Gorczyca, M. T., H. F. M. Milan, A. S. C. Maia, and K. G. Gerbredhin, 2018. Machine learning algorithms to predict core, skin and hair-coat temperatures of piglets. *Computers and Electronics in Agriculture* 151: 286-294. doi:10.1016/j.compag.2018.06.028.
27. Gorczyca, M. T., 2019. Machine learning applications for monitoring heat stress in livestock. *Faculty of the Graduate School of Cornell University*: 1-64.
28. Wang, Y., X. Yong, Z. F. Chen, H. Y. Zheng, J. Y. Zhuang, and J. J. Liu, 2018. The design of an intelligent livestock production monitoring and management system. *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*: 1-5 doi:10.1109/DDCLS.2018.8516021.
29. ASTM, Standard test methods for laboratory determination of water (moisture) content of soil and rock by mass (Designation: D2216-10).

30. Hong, S. E., T. J. Park, J. I. Bang, and H. J. Kim, 2020. A study on the prediction model for tomato production and growth using ConvLSTM. *The Journal of Korean Institute of Information Technology* 18(1): 1-10 (in Korean). doi: 10.14801/jkiit.2020.18.1.1.
31. Shakoor, M. T, K. Rahman, S. N. Rayta, and A. Chakrabarty, 2017. Agricultural production output prediction using supervised machine learning techniques. *2017 1st International Conference on Next Generation Computing Applications (NextComp)*:182-187. doi:10.1109/NEXTCOMP.2017.8016196.
32. Alhnaity, B., S. Pearson, G. Leontidis, and S. Kollias, 2020. Using deep learning to predict plant growth and yield in greenhouse environments. *Acta Horti* 1296: 425-432. doi: 10.17660/ActaHortic.2020.1296.55.
33. Cho, Y. H., Y. D. Seo, D. J. Park, and J. C. Jeong, 2016. Study on the activation functions for efficient learning in DNN. *The Institute of Electronics and Information Engineers*: 800-803 (in Korean).
34. Shin, H. Y., H. K. Yim, and W. T. Kim, 2018. Intelligent green house control system based on deep learning for saving electric power consumption. *Institute of Korean Electrical and Electronics Engineers*: 53-60 (in Korean). doi:10.7471/ikeee.2018.22.1.53.
35. Efron B., and R. J. Tibshirani, 1998. *An Introduction to the bootstrap. Monographs on statistics and applied probability* 57. London: Chapman & Hall/CHC.
36. Akinoglu, B. G., 1991. A review of sunshine-based models used to estimate monthly average global solar radiation. *Renewable Energy* 1(3): 479-497.
37. Bakay, M. S., and U. Agbulut, 2021. Electricity production based forecasting of greenhouse gas emissions in turkey with deep learning, Support Vector Machine and Artificial Neural Network algorithms. *Journal of Cleaner Production* 285: 1-18.
38. Gouda, S. G., Z. Hussein, S. Luo, and Q. X. Yuan, 2019. Model selection for accurate daily global solar radiation prediction in China. *Journal of Cleaner Production* 221: 132-144. doi:10.1016/j.jclepro.2019.02.211.
39. Bakirci, K. 2009. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. *Energy* 34: 485-501. doi:10.1016/j.energy.2009.02.005.
40. Wang, H. K., L. Li, W. Yong, M. Fanjia, H. H. Wang, and N. A. Sigrimis, 2018. Recurrent Neural Network model for prediction of microclimate in solar greenhouse. *IFAC PapersOnLine* 51(17): 790-795. doi:10.1016/j.ifacol.2018.08.099.