

## 기계학습을 이용한 벼 수발아율 예측

반호영<sup>1</sup>, 정재혁<sup>\*</sup>, 황운하<sup>1</sup>, 이현석<sup>1</sup>, 양서영<sup>1</sup>, 최명구<sup>1</sup>, 이충근<sup>1</sup>, 이지우<sup>2</sup>,  
이채영<sup>3</sup>, 윤여태<sup>4</sup>, 한채민<sup>5</sup>, 신서호<sup>6</sup>, 이성태<sup>7</sup>

<sup>1</sup>농촌진흥청 국립식량과학원 작물재배생리과, <sup>2</sup>강원도농업기술원 작물연구과, <sup>3</sup>충청북도농업기술원 작물연구과,  
<sup>4</sup>충청남도농업기술원 작물연구과, <sup>5</sup>경상북도농업기술원 작물연구과, <sup>6</sup>전라남도농업기술원 식량작물연구소,  
<sup>7</sup>경상남도농업기술원 작물연구과

(2020년 7월 20일 접수; 2020년 10월 8일 수정; 2020년 11월 9일 수락)

## Predicting the Pre-Harvest Sprouting Rate in Rice Using Machine Learning

Ho-Young Ban<sup>1</sup>, Jae-Hyeok Jeong<sup>1\*</sup>, Woon-Ha Hwang<sup>1</sup>, Hyeon-Seok Lee<sup>1</sup>, Seo-Yeong Yang<sup>1</sup>,  
Myong-Goo Choi<sup>1</sup>, Chung-Keun Lee<sup>1</sup>, Ji-U Lee<sup>2</sup>, Chae Young Lee<sup>3</sup>, Yeo-Tae Yun<sup>4</sup>,  
Chae Min Han<sup>5</sup>, Seo Ho Shin<sup>6</sup>, Seong-Tae Lee<sup>7</sup>

<sup>1</sup>Division of Crop Physiology and Production, National Institute of Crop Science, Rural development  
Administrarion, Hyeoksin-ro 181, Iseo-myeon, Wanju-gun, Jeollabuk-do, Republic of Korea

<sup>2</sup>Grain Research Division, Gangwondo Agricultural Research and Extension Services, 83, Chungyeol-ro,  
Chuncheon-si, Gangwon-do, Republic of Korea

<sup>3</sup>Grain Research Division, Chungcheongbuk-do Agricultural Research and Extension Services, 46, Gagok-gil,  
Ochang-eup, Cheongwon-gu, Cheongju, Cheongju-si, Chungcheongbuk-do, Republic of Korea

<sup>4</sup>Grain Research Division, Chungcheongnam-do Agricultural Research & Extension Services, 167, Chusa-ro,  
Sinam-myeon, Yesan-gun, Chungcheongnam-do, Republic of Korea

<sup>5</sup>Grain Research Division, Gyeongsangbuk-do Agricultural Research & Extension Services, 47,  
Chilgokjungang-daero 136-gil, Buk-gu, Daegu, Republic of Korea

<sup>6</sup>Food crop research center, Jeollanamdo Agricultural Research & Extension Services, 1508, Senam-ro,  
Sanpo-myeon, Naju-si, Jeollanam-do, Republic of Korea

<sup>7</sup>Grain Research Division, Gyeongsangnam-do Agricultural Research & Extension Services, 570, Daesin-Ro,  
Jinju-si, Gyeongsangnam-do, Republic of Korea

(Received July 20, 2020; Revised October 8, 2020; Accepted November 9, 2020)

### ABSTRACT

Rice flour varieties have been developed to replace wheat, and consumption of rice flour has been encouraged. damage related to pre-harvest sprouting was occurring due to a weather disaster during the ripening period. Thus, it is necessary to develop pre-harvest sprouting rate prediction system to minimize damage for pre-harvest sprouting. Rice cultivation experiments from 2017 to 2019 were conducted with three rice flour varieties at six regions in Gangwon-do, Chungcheongbuk-do, and Gyeongsangbuk-do. Survey components were the heading date and pre-harvest sprouting at the harvest date. The weather data were collected daily mean temperature, relative humidity, and rainfall using Automated Synoptic Observing



\* Corresponding Author : Jae-Hyeok Jeong  
(rodnf2010@korea.kr)

System (ASOS) with the same region name. Gradient Boosting Machine (GBM) which is a machine learning model, was used to predict the pre-harvest sprouting rate, and the training input variables were mean temperature, relative humidity, and total rainfall. Also, the experiment for the period from days after the heading date (DAH) to the subsequent period (DA2H) was conducted to establish the period related to pre-harvest sprouting. The data were divided into training-set and vali-set for calibration of period related to pre-harvest sprouting, and test-set for validation. The result for training-set and vali-set showed the highest score for a period of 22 DAH and 24 DA2H. The result for test-set tended to overpredict pre-harvest sprouting rate on a section smaller than 3.0%. However, the result showed a high prediction performance ( $R^2=0.76$ ). Therefore, it is expected that the pre-harvest sprouting rate could be able to easily predict with weather components for a specific period using machine learning.

**Key words:** Pre-harvest sprouting rate, Machine learning, Rice, Weather components, Prediction

## I. 서 언

벼는 우리나라의 주식으로 가장 중요한 식량작물이며(Kim, 2007; Park and Kim, 2016), 최근 핵가족화 및 서구화된 식습관으로 쌀 소비가 계속해서 줄어들고 있는 실정이다(Kang, 2017). 쌀 소비를 촉진시키기 위해 음식의 주요 재료인 밀가루를 쌀가루로 대체하여 국수, 빵, 전 등 다양한 종류의 음식에 적용하기 위한 노력을 많이 기울이고 있다(Jeong *et al.*, 2019; Kim and Chung, 2018; Won *et al.*, 2020). 하지만 밀가루에 비해 비용이 비싸고 가공이 용이하지 않은 단점이 있다(Kim, 2013). 이 같은 문제를 해결하기 위해 쌀가루용 벼 품종이 개발되고 있다(Won *et al.*, 2019).

우리나라는 온대 기후에 속하며, 벼의 이삭이 차는 시기인 등숙기(8~9월)에 잦은 강우와 고온다습으로 인해 병해충 및 수발아 등 여러 피해들이 발생하고 있다(Park and Kim, 2009; Park and Park, 1984). 특히 지구 온난화로 인해 고온다습, 잦은 강우 등 평년에는 없었던 기상 사건들이 불시에 발생하고 있으며(Im *et al.*, 2019), 태풍으로 인해 수확 전 벼 종자가 발아하는 수발아 및 도복 피해가 발생하고 있다. 최근 수발아와 도복 피해규모를 보면, 2019년 태풍 ‘링링’과 ‘타파’로 인해 전남 10,097 ha, 전북 5,882 ha, 충남 4,789 ha 등 전체 26,798 ha이였다(MAFRA, 2019).

벼 수발아 피해에 관한 여러 연구들이 진행되었다(Baek and Jung, 2014; Kang *et al.*, 2018; Lee *et al.*, 2015; Park and Kim, 2009). Kang *et al.*(2018)은 2013년부터 2017년까지 다양한 생태형의 품종들을 재배하여 출수 후 적산온도에 따른 품종 및 연도별로 수발아

율을 조사하였으며, Baek and Jung(2014)은 호품벼와 신동진벼를 이용하여 품종 별 종자의 휴면성을 검정하고 이삭 착생 부위별로 수발아 양상을 조사하였다. Park and Kim(2009)은 2000년 이후 육성된 중부지역 적응 신품종 11품종과 재배면적이 많은 품종 9개를 더해 총 20개 품종을 대상으로 재배 실험을 수행하였으며, 실내 및 포장에서 출수후 경과일수에 따른 수발아 분포를 조사하였다. 이러한 연구들은 여러 품종들의 수발아 발생 시기와 발생 기상 조건을 알 수 있는 중요한 연구들이지만, 수발아 관련 피해를 조사하기 위해 실내 및 포장에서 인위적으로 조건을 설정하였으며, 자연조건에서 수발아에 관한 연구는 미비한 실정이다. 따라서 최근 많이 활용하고 있는 기계학습을 이용하여 실제 자연조건에서 관련 기상요인들로 벼 수발아율을 예측하는 연구가 필요하며, 그 중에서도 수발아 발생 위험이 높은 쌀가루용 벼에 대한 연구가 필요하다.

본 연구의 목적은 자연 조건에서 쌀가루용 벼의 수발아율을 예측하기 위한 것으로 기계학습을 이용하여 관련 기상요소들로 수발아율을 간단하게 예측할 수 있는 초기 시스템을 개발하는 것이다.

## II. 재료 및 방법

### 2.1. 벼 재배 정보

본 연구는 강원도농업기술원(춘천), 철원군농업기술센터(철원), 충청북도농업기술원(청주), 경상북도농업기술원(대구), 경상북도농업기술원 생물자원연구소(안동)와 충북 보은군 보은읍 금굴리 소재 농가(보은) 총 6군데 시험포장에서 벼 재배실험을 수행하였다. 품종은

쌀가루용 벼 품종인 신길(중생), 설갱(중만생), 한가루(중만생) 3품종이었다. 이앙 날짜는 연도 및 품종 그리고 지역에 맞게 날짜를 달리하였으며, 약 10일 간격으로 4번에 걸쳐 1주 3본으로 손이앙하였다(Table 1).

재식 간격은 30cm x 15cm이었으며, 비료는 농촌진흥청 권장 표준시비량에 따른 N-P-K= 9.0-4.5-5.7 kg/10a 수준으로 기바-분얼바-수비를 질소 5:2:3, 인산 10:0:0, 그리고 칼리 7:0:3으로 각각 나누어 시비하였다. 기타

재배관리는 농촌진흥청 표준재배법에 의거 수행하였다.

### 2.2. 기상 수집

기상자료 수집을 위해 기상자료개발 포털(<https://data.kma.go.kr/cmmn/main.do>)에서 제공하는 종관기상관측자료(ASOS; Automated Synoptic Observing

**Table 1.** Transplanting dates by year, variety, and region

Region	Variety	Transplanting date (m.d)			Region	Variety	Transplanting date (m.d)		
		2017	2018	2019			2017	2018	2019
Chuncheon	Seolgaeng (Mid-late)	5.10	5.10	5.20	Boeun	Seolgaeng	5.20	5.10	5.20
		5.20	5.20	5.30			5.30	5.20	5.30
		5.30	5.30	6.10			6.10	5.30	6.10
		6.10	6.10	6.20			6.20	6.10	6.20
	Hangaru (Mid-late)	5.10	5.10	5.20		Hangaru	5.20	5.10	5.20
		5.20	5.20	5.30			5.30	5.20	5.30
		5.30	5.30	6.10			6.10	5.30	6.10
		6.10	6.10	6.20			6.20	6.10	6.20
	Shingil (Medium)	5.10	5.10	5.20		Shingil	5.20	5.10	5.20
		5.20	5.20	5.30			5.30	5.20	5.30
		5.30	5.30	6.10			6.10	5.30	6.10
		6.10	6.10	6.20			6.20	6.10	6.20
Cheorwon	Seolgaeng	5.10	4.30	5.10	Daegu	Seolgaeng	5.20	5.20	5.30
		5.20	5.10	5.20			5.30	5.30	6.10
		5.30	5.20	5.30			6.10	6.10	6.20
		6.10	5.30	6.10			6.20	6.20	6.30
	Hangaru	5.10	4.30	5.10		Hangaru	5.20	5.20	5.30
		5.20	5.10	5.20			5.30	5.30	6.10
		5.30	5.20	5.30			6.10	6.10	6.20
		6.10	5.30	6.10			6.20	6.20	6.30
	Shingil	5.10	4.30	5.10		Shingil	5.20	5.20	5.30
		5.20	5.10	5.20			5.30	5.30	6.10
		5.30	5.20	5.30			6.10	6.10	6.20
		6.10	5.30	6.10			6.20	6.20	6.30
Cheongju	Seolgaeng	5.20	5.20	5.30	Andong	Seolgaeng	5.20	5.20	5.30
		5.30	5.30	6.10			5.30	5.30	6.10
		6.10	6.10	6.20			6.10	6.10	6.20
		6.20	6.20	6.30			6.20	6.20	6.30
	Hangaru	5.20	5.20	5.30		Hangaru	5.20	5.20	5.30
		5.30	5.30	6.10			5.30	5.30	6.10
		6.10	6.10	6.20			6.10	6.10	6.20
		6.20	6.20	6.30			6.20	6.20	6.30
	Shingil	5.20	5.20	5.30		Shingil	5.20	5.20	5.30
		5.30	5.30	6.10			5.30	5.30	6.10
		6.10	6.10	6.20			6.10	6.10	6.20
		6.20	6.20	6.30			6.20	6.20	6.30

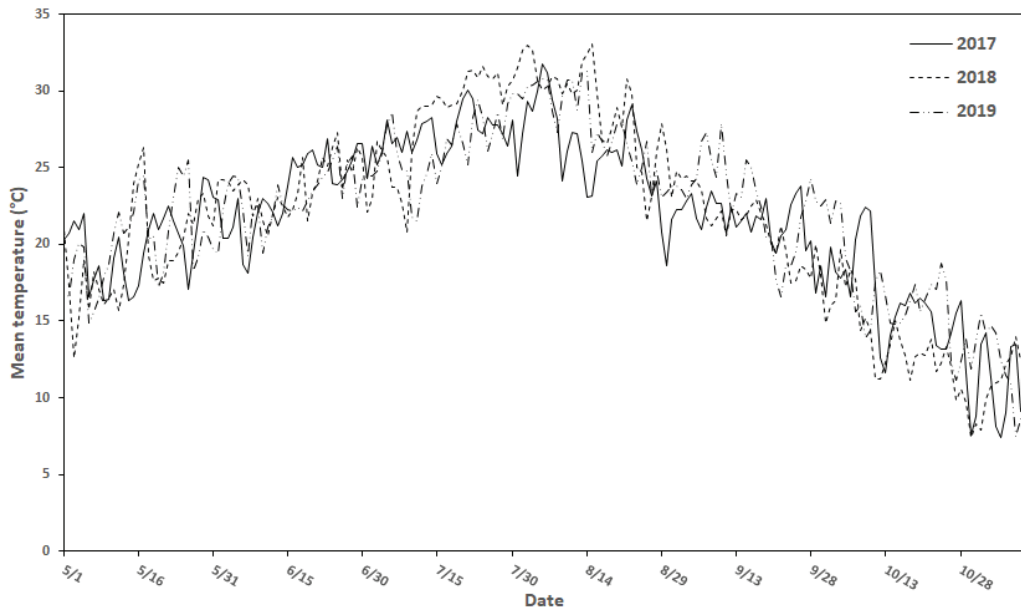


Fig. 1. Daily mean temperatures at Cheongju weather station from 2017 to 2019.

Table 2. The mean temperature, relative humidity, and total rainfall during September and October by region

Weather component	Year	Chuncheon	Cheorwon	Cheongju	Boeun	Daegu	Andong
Mean temperature (°C)	2017	16.6	15.7	18.6	16.1	18.8	17.3
	2018	15.2	14.3	17.3	14.9	17.7	16.1
	2019	18.0	16.8	19.7	17.4	19.7	18.0
Mean Relative humidity (%)	2017	79.0	70.7	62.2	79.9	69.8	77.1
	2018	76.3	76.7	65.4	78.2	66.9	77.2
	2019	77.2	83.7	69.1	80.8	74.9	77.2
Total Rainfall (mm)	2017	86.0	37.4	107.3	151.5	154.3	170.6
	2018	197.4	130.2	319.0	247.5	274.6	202.5
	2019	192.9	229.9	239.4	286.8	340.9	289.4

System)를 활용하였고, 재배 실험 지역과 동일한 이름을 가진 6개 지역의 기상을 2017년부터 2019년까지 수집하였다. 수집한 기상요소들은 일 평균 기온(°C)과 평균 상대 습도(%), 그리고 강수량(mm)이었다(Fig. 1).

본 연구에서는 일사가 없는 지역이 있어 제외되었으며, 벼 등숙기간에 해당하는 9월과 10월의 평균 기온과 상대 습도 그리고 총 강수량을 지역 별로 정리하였다(Table 2).

### 2.3. 수발아율과 출수기 조사

수발아율은 수확시기에 임의의 3개의 지점에서 이

삭을 10개씩 랜덤으로 수집하여 정조상태에서 전체 영화수에 대한 수발아 영화수의 비율로 수발아율을 조사하였다. 출수는 포장의 50%가 출수하였을 때를 출수일로 기록하였다.

### 2.4. 기계학습 모델

본 연구에서 기계학습 모델로 정확도가 높은 Gradient Boosting Machine (GBM) (Friedman, 2001)을 python 언어로 코딩하였으며, Scikit-learn (version 0.23.1)이라는 python 라이브러리를 이용하여 GBM을 구동하였다.

GBM은 효율적이고 정확하여 많은 분야에서 널리 이용되고 있으며(Feng *et al.*, 2018), Boosting 알고리즘을 기반으로 여러 개의 약한 모델을 순차적으로 학습하고 예측하면서 예측 오류가 높은 모델에 가중치를 부여하여 오류를 개선하는 모델이다(Natekin and Knoll, 2013). GBM은 score 값으로 모델 정확도를 판단하며, score값은 0.0에서 1.0사이로 1.0에 가까울수록 정확도가 높다.

**2.5. 기계학습 투입 변수 설정**

수발아 발생에 여러 기상요인들이 복잡하게 작용을 하며, 기상 요소들과 수발아 관련 기간이 특정되어 지지 않았다. 따라서 특정기간동안의 평균 기온과 상대 습도, 그리고 총 강수량을 투입 변수로 설정하였으며, 또한 수발아율 조사에서의 오차를 생각한 지역 변수와 품종 특성을 고려한 품종 변수를 추가 설정하였다.

**2.6. 수발아 피해 관련 기간 설정 및 자료 분류**

수발아 피해와 관련된 기간을 설정하기 위해 출수 후 일정기간 이후(DAH)부터 일정기간동안(DA2H)에 대한 기간 설정 실험을 수행하였다. 기간은 출수 후 (DAH) 15~25일 후부터 그 이후 1~25일동안의 기간 (DA2H)을 설정하였고(Fig. 2), 그 기간들에 대해 교정

및 검증을 수행하였다.

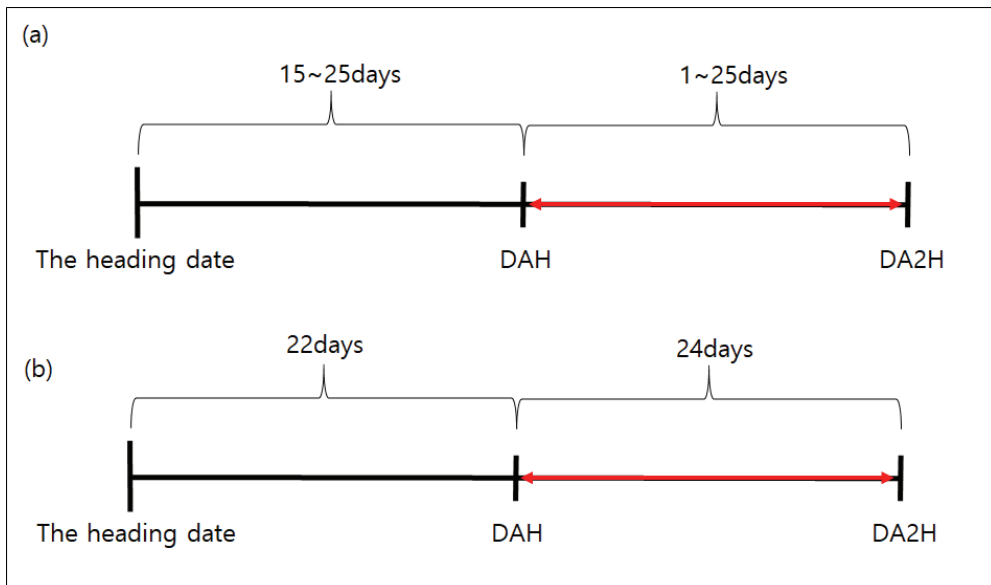
수집된 자료는 기계학습을 위한 training-set과 실험을 위한 vali-set, 즉 수발아 관련 기간을 교정하는 자료들과 수발아 관련 기간을 검증하기 위한 test-set 총 세 부분으로 분류하였다. 자료 분류는 랜덤으로 자료를 분류해주는 scikit-learn 라이브러리에 포함된 train\_test\_split 함수를 이용하였으며, 전체 자료에서 training-set 75% 그리고 vali-set과 test-set 나머지 25%로 나누었다. vali-set과 test-set은 다시 각각 75%와 25% 비율로 나누었으며, 최종적으로 training-set과 vali-set, 그리고 test-set은 각각 162개와 41개, 그리고 13개이었다.

**III. 결과 및 고찰**

**3.1. 수발아율과 출수일**

Table 3은 재배 실험 지역 및 품종, 그리고 이앙시기 별 조사한 수발아율과 출수일이다.

출수일은 모든 실험에서 8월초부터 9월초 사이로 나타났으며, 중생종인 신길이 중만생종인 다른 두 품종에 비해 출수가 빨랐다(Park *et al.*, 2013). 2018년의 출수일이 대부분의 지역에서 다른 두 연도에 비해 좀 더 빨랐다. 청주 지역의 2017년부터 2019년까지의 일



**Fig. 2.** The description of period for days after the heading date(DAH) and days after DAH(DA2H) (a) and illustration for 22 DAH and 24 DA2H (b).

**Table 3.** The pre-harvest sprouting rate and heading date by region, variety, and transplanting date

Region	Variety	Transplanting date	Heading date (m.d)			Pre-harvest sprouting rate (%)		
			2017	2018	2019	2017	2018	2019
Chuncheon	Seolgaeng	1st	8.21	8.17	8.22	6.2	3.0	3.9
		2nd	8.27	8.23	8.29	5.5	1.3	3.5
		3rd	9.01	8.24	9.03	2.7	0.7	3.7
		4th	9.05	8.31	9.06	1.4	0.7	3.0
	Hangaru	1st	8.08	8.11	8.18	1.8	4.8	3.7
		2nd	8.14	8.13	8.22	3.6	2.2	1.9
		3rd	8.21	8.15	8.28	1.1	1.5	1.3
		4th	8.26	8.25	9.01	1.1	0.3	1.3
	Shingil	1st	8.06	8.01	8.16	4.0	11.3	2.9
		2nd	8.10	8.06	8.20	6.9	7.3	1.7
		3rd	8.15	8.12	8.25	6.5	3.8	1.1
		4th	8.18	8.22	8.30	2.6	1.3	0.4
Cheorwon	Seolgaeng	1st	8.16	8.16	8.12	2.4	1.0	3.1
		2nd	8.18	8.16	8.20	2.2	1.3	1.8
		3rd	8.22	8.17	8.26	0.8	0.3	0.8
		4th	8.25	8.25	9.02	0.3	0.2	0.2
	Hangaru	1st	8.15	8.15	8.10	2.0	0.3	0.7
		2nd	8.18	8.15	8.16	2.4	1.5	1.7
		3rd	8.22	8.16	8.22	0.9	1.2	0.6
		4th	8.24	8.19	8.29	1.0	0.2	0.0
	Shingil	1st	8.13	8.02	8.06	3.3	7.5	4.9
		2nd	8.15	8.03	8.11	3.0	6.3	2.9
		3rd	8.17	8.08	8.19	2.3	5.0	0.6
		4th	8.20	8.12	8.28	1.6	3.5	0.5
Cheongju	Seolgaeng	1st	8.19	8.16	8.16	0.0	0.7	0.0
		2nd	8.25	8.18	8.24	1.0	1.0	1.8
		3rd	8.28	8.20	8.29	2.3	2.3	3.8
		4th	9.03	8.31	9.02	2.3	0.0	2.5
	Hangaru	1st	8.10	8.07	8.11	4.3	10.3	5.3
		2nd	8.21	8.11	8.17	4.7	3.0	3.3
		3rd	8.24	8.19	8.24	4.3	1.3	4.7
		4th	8.28	8.25	9.01	2.0	1.3	1.4
	Shingil	1st	8.13	8.09	8.16	0.7	1.0	0.5
		2nd	8.18	8.13	8.27	0.0	4.3	0.3
		3rd	8.24	8.20	9.05	1.7	0.0	1.5
		4th	9.04	8.26	9.13	0.3	0.0	2.7
Boeun	Seolgaeng	1st	8.16	8.05	8.13	1.3	3.3	1.0
		2nd	8.23	8.11	8.16	3.0	2.3	3.7
		3rd	8.27	8.15	8.23	3.0	0.7	1.3
		4th	8.31	8.23	8.29	0.0	0.7	2.0
	Hangaru	1st	8.05	8.01	8.07	2.3	24.0	5.0
		2nd	8.17	8.05	8.11	1.3	8.7	6.7
		3rd	8.23	8.09	8.18	1.7	2.7	2.9
		4th	8.26	8.18	8.24	1.3	2.0	2.7
	Shingil	1st	8.08	8.01	8.09	0.3	1.7	1.3
		2nd	8.16	8.06	8.14	0.3	0.3	0.3
		3rd	8.23	8.10	8.24	0.3	0.7	0.7
		4th	8.31	8.21	9.04	1.7	0.0	1.9

Region	Variety	Transplanting date	Heading date (m.d)			Pre-harvest sprouting rate (%)		
			2017	2018	2019	2017	2018	2019
Daegu	Seolgaeng	1st	8.14	8.07	8.16	3.7	3.6	3.5
		2nd	8.20	8.15	8.19	2.5	1.2	2.8
		3rd	8.24	8.22	8.25	2.0	3.6	1.1
		4th	8.26	8.25	8.31	6.1	3.4	3.2
	Hangaru	1st	8.04	8.02	8.13	5.8	6.7	8.2
		2nd	8.14	8.08	8.16	1.2	7.0	3.0
		3rd	8.18	8.16	8.20	8.5	5.1	1.9
		4th	8.23	8.21	8.27	4.5	3.9	4.5
	Shingil	1st	8.08	8.05	8.17	0.0	0.3	1.3
		2nd	8.14	8.10	8.21	0.2	0.1	0.4
		3rd	8.19	8.18	9.01	0.0	0.0	0.3
		4th	8.26	8.29	9.07	0.0	0.0	0.7
	Seolgaeng	1st	8.18	8.06	8.17	6.8	3.7	8.2
		2nd	8.22	8.18	8.21	9.7	0.6	3.2
		3rd	8.26	8.25	8.30	5.7	9.7	5.1
		4th	8.29	8.31	9.05	7.1	9.5	4.4
Andong	Hangaru	1st	8.07	8.03	8.15	8.8	7.4	20.1
		2nd	8.18	8.10	8.20	10.4	3.5	13.1
		3rd	8.23	8.23	8.27	6.8	5.1	7.9
		4th	8.26	8.28	9.03	5.1	1.3	7.0
	Shingil	1st	8.14	8.06	8.20	1.0	2.5	5.2
		2nd	8.18	8.12	8.24	0.6	0.4	1.1
		3rd	8.23	8.25	9.03	0.1	0.0	4.9
		4th	9.02	9.03	9.11	0.7	0.1	3.3

\*1st: first transplanting date, 2nd: second transplanting date, 3rd: third transplanting date, 4th: Fourth transplanting date.

평균기온을 살펴보면 2018년의 일 평균 기온이 다른 두 연도에 비해 높았으며(Fig. 1), 높은 일 평균 기온으로 인해 출수가 빠른 것을 판단된다(Lee *et al.*, 2019; Park *et al.*, 2013).

수발아율이 가장 높은 지역은 안동이였으며, 가장 낮은 지역은 철원이었다. Table 3에서 대구와 청주가 등숙기간에 해당하는 9월과 10월의 평균 기온이 안동보다 높지만, 두 지역에 비해 안동이 평균 상대 습도가 많이 높았다. 수발아는 고온다습한 조건에서 발생하며 (Bewley, 1997; Hu *et al.*, 2017), 고온다습으로 인해 안동에서 수발아율이 가장 높으며, 철원의 경우 평균 기온이 제일 낮아 수발아율이 가장 낮은 것으로 판단된다. 강원도 지역인 춘천과 철원을 제외한 4개 지역 모두 한가루의 수발아율이 가장 높았다. 철원은 신길의 수발아율이 높았으며, 춘천은 2017년과 2018년은 신길, 그리고 2019년은 설갱의 수발아율이 높았다. 품종 간 수발아에 대한 발생 조건에 정도 차이가 있으며 (Ju *et al.*, 2000; Lee *et al.*, 2018), 이양시기를 보면

대부분 일찍 이양한 경우 출수 이 후 등숙기간동안 기온과 상대 습도가 높아 늦게 이양한 것보다 좀 더 높은 수발아율을 나타낸 것으로 판단된다.

### 3.2. 수발아율 예측 성능 교정

기계학습을 이용한 수발아율 예측 성능 교정에 대한 DAH와 DA2H 별 training-set과 vali-set의 score 값의 합들을 나타내었다(Table 4).

수발아율 예측과 관련이 높은 기간은 training-set과 vali-set의 score 합이 1.584로 가장 크게 나타낸 출수 후(DAH) 22일 후부터 24일(DA2H)동안이었다. 이 기간은 황숙기부터 완숙기까지의 기간으로 이 시기의 기상요인들이 수발아와 관련이 높다는 연구 결과와 일치한다(Park and Park, 1984). 기계학습 투입 변수들 중 어떤 변수들이 중요하게 활용되는지 확인하기 위해 변수 중요도(Feature Importance, FI)를 확인하였다 (Table 5).

투입 변수들 중 평균기온과 총 강수량 그리고 품종

**Table 4.** The sum of scores for training-set and vali-set by days after the heading date(DAH) and days after DAH(DA2H)

DA2H	DAH										
	15	16	17	18	19	20	21	22	23	24	25
1	1.266	0.996	1.267	1.264	1.113	1.290	1.360	1.469	1.273	1.142	1.128
2	1.182	1.123	1.307	1.361	0.970	1.335	1.451	1.409	1.118	1.347	0.738
3	1.309	1.234	0.637	1.251	1.107	1.417	1.441	0.977	1.280	1.242	1.175
4	1.359	1.206	0.933	1.179	1.397	1.433	1.245	1.246	1.233	1.232	1.220
5	1.263	0.816	1.336	1.311	1.181	1.356	1.427	1.399	1.378	1.311	1.298
6	1.391	1.043	1.433	0.982	1.174	1.210	1.470	1.343	1.419	1.472	1.456
7	1.342	1.235	0.681	1.203	1.264	1.384	1.450	1.396	1.273	1.409	1.315
8	1.201	1.417	1.063	1.153	1.429	1.452	1.478	1.444	1.365	1.229	1.196
9	1.463	1.310	1.177	1.363	1.382	1.379	1.523	1.392	1.416	1.297	1.281
10	1.434	1.308	1.492	1.491	1.416	1.477	1.536	1.192	1.175	1.349	1.187
11	1.220	1.353	1.440	1.134	1.505	1.447	1.482	1.151	1.130	1.397	1.387
12	1.427	1.464	1.455	1.292	1.468	1.448	1.371	1.163	1.379	1.427	1.317
13	1.473	1.429	1.323	1.433	1.460	1.409	1.297	1.385	1.394	1.365	1.185
14	1.226	1.178	1.111	1.441	1.392	1.383	1.384	1.308	1.244	1.314	1.194
15	1.468	1.456	1.359	1.418	1.506	1.456	1.469	1.381	1.346	1.266	1.277
16	1.532	1.388	1.453	1.380	1.407	1.496	1.427	1.168	1.366	1.108	1.149
17	1.419	1.423	1.561	1.410	1.454	1.368	1.477	1.116	1.328	1.489	1.391
18	1.327	1.557	1.505	1.460	1.432	1.196	1.171	1.376	1.098	1.378	1.385
19	1.307	1.559	1.510	1.379	1.384	1.486	1.221	1.428	1.456	1.407	1.500
20	1.578	1.443	1.345	1.394	1.485	1.484	1.263	1.536	1.410	1.453	1.442
21	1.490	1.268	1.491	1.354	1.423	1.470	1.462	1.418	1.365	1.487	1.478
22	1.412	1.463	1.428	1.375	1.447	1.476	1.481	1.344	1.494	1.390	1.414
23	1.534	1.426	1.371	1.442	1.495	1.575	1.368	1.516	1.498	1.374	1.475
24	1.526	1.374	1.487	1.386	1.435	1.446	1.495	1.584	1.440	1.450	0.935
25	1.480	1.449	1.401	1.478	1.502	1.541	1.489	1.522	1.440	1.306	1.400

**Table 5.** Feature importance for predicting pre-harvest sprouting rate based on Gradient Boosted Machines(GBM)

Mean relative Humidity (%)	Mean Temperature (°C)	Total rainfall (mm)	Region	Variety
0.088	0.243	0.236	0.188	0.245

이 높게 나타났으며, 그 중에서도 근소하지만 품종 변수가 가장 높았다. 가장 낮은 변수는 평균 상대 습도였다. 상대 습도의 경우 강수량과 직접적으로 관련이 있으며(Huang *et al.*, 2019), 기계학습에서 강수량이 높은 중요도를 나타내었기 때문에 평균 상대 습도는 기계학습에서 낮게 활용된 것으로 판단된다. 높은 중요도를 가진 평균 기온과 총 강수량은 수발아와 직접적

으로 높은 관련이 있다는 연구와 일치하였으며(Im *et al.*, 2019; Park and Kim, 2009; Park and Park, 1984), 품종 변수는 수발아에 대한 발생 시기 및 조건의 차이로 인해 가장 높은 중요도를 나타낸 것으로 판단된다. 지역에 대한 변수는 각 지역마다 수발아 피해 조사과정에서 약간의 차이가 발생할 수 있으며, 그 차이로 인해 약간 높은 변수 중요도를 나타낸 것으로 판단된다.



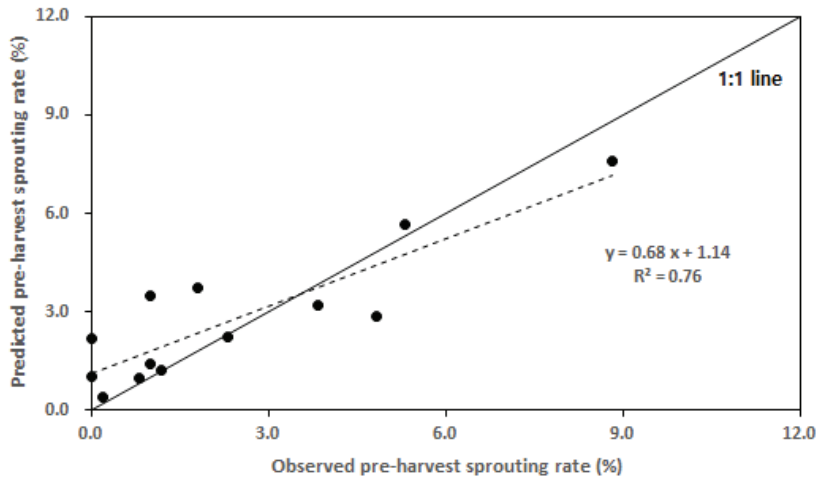


Fig. 3. Comparison of the observed and predicted pre-harvest sprouting rate for test-set of 22 days after the heading date(DAH) and 24 days after DAH(DA2H).

### 3.3. 수발아율 예측 성능 검증

가장 높은 score를 나타낸 DAH가 22일, 그리고 DA2H가 24일일 때의 기상요소들로 test-set을 검증한 결과를 비교하였다(Fig. 3).

DAH가 22일, 그리고 DA2H가 24일일 때의 기상요소들로 test-set을 검증한 결과, 3.0%보다 낮은 구간에서 예측 수발아율이 관측 수발아율에 비해 약간 높은 경향이 있지만, 높은 예측력을 보였다( $R^2=0.76$ ). 본 연구에서 기계 학습 투입 변수로 품종 별 특성을 고려하여 품종 관련 변수를 설정하였지만 단순하게 품종만 구별하기 위해 사용되었다. 하지만, 품종마다 종자의 휴면기간이 다르며, 그 기간으로 인해 수발아에 미치는 기간이 품종마다 차이가 있다(Ju et al., 2000; Lee et al., 2018). 따라서, 품종 별로 수발아 관련 기간을 달리 설정하여 예측을 한다면 좀 더 높은 예측력을 보일 것으로 판단된다.

## 적 요

본 연구는 자연 조건에서 쌀가루용 벼의 수발아율을 예측하기 위한 것으로 기계학습을 이용하여 기상요소들에 따른 수발아율을 간단히 예측할 수 있는 초기 시스템을 개발하기 위해 수행되었다. 이를 위하여 강원도, 충청북도, 경상북도에 위치한 6개 지역에서 쌀가루용 벼 3품종을 재배하였다. 수확 후 수발아율과 출수일을 조사하였으며, 각 지역의 종관기상대의 일

평균 기온과 상대 습도, 그리고 강수량 정보를 이용하여 기계학습 모델 중 하나이며, 정확도가 높은 GBM 모델로 수발아율을 예측하였다. 2017년부터 2019년까지 강원과 충북, 그리고 경북의 6개 지역에서 쌀가루용 벼 3품종에 대해 재배 실험을 수행하였다. 조사 항목은 출수일과 수발아율이었다. 기상자료는 동일한 지역명의 종관기상대를 이용하여 일 평균 기온 및 상대 습도, 그리고 강수량 자료를 수집하였다. 수발아율 예측을 위해 기계학습 모델인 Gradient Boosting Machine (GBM)을 이용하였으며, 학습 투입 변수로는 평균 기온과 상대 습도, 그리고 총 강수량이었다. 또한 수발아 피해 관련 기간을 설정하기 위해 출수 후 몇 일 후부터 그 이후의 기간에 대한 실험도 수행하였다. 자료는 수발아 피해 관련 기간의 교정을 위한 training-set과 vali-set, 검증을 위한 test-set으로 구분하였다. training-set과 vali-set으로 교정한 결과, 출수 후 22일 후부터 24일동안에서 가장 높은 score를 나타내었다. test-set으로 검증한 결과는 3.0%보다 낮은 구간에서 수발아율을 약간 높게 예측한 경향이 있었지만, 높은 예측력을 보였다( $R^2=0.76$ ). 따라서, 기계학습을 이용하여 특정기간동안의 기상요소들로 수발아율을 간단하게 예측할 수 있을 것으로 예상된다. 본 연구의 결과를 종합해 볼 때, 기계학습을 이용하여 특정 기간 동안에 평균 기온과 상대 습도, 그리고 총 강수량으로 높은 수발아율 예측 성능을 보였으며, 이 시스템을 이용하여 일반 농가들을 대상으로 수발아에 관한

피해를 예방할 수 있는 조기 수발아 예측 시스템으로 이용가능 할 것으로 판단된다. 하지만 품종마다 휴면 정도 차이로 인한 수발아 관련 기간에 차이가 있으므로, 다른 쌀가루용 벼 품종에 대해서도 추가로 조사하고, 개별 품종으로 세분화하여 분석한다면 좀 더 정확도 높은 예측 시스템을 개발할 수 있을 것으로 판단된다.

## 감사의 글

본 연구는 농촌진흥청 작물시험연구사업(과제 번호: PJ01296001)의 지원에 의해 수행되었다.

## REFERENCES

- Baek, J., and N. Chung, 2014: Pre-Harvest Sprouting Variation of Rice Seeds Located on Each Panicle Position According to Grain Filling Days. *Korean Journal of Crop Science* **59**(1), 22-26. <http://dx.doi.org/10.7740/kjcs.2014.59.1.022>
- Bewley, J. D., 1997: Seed germination and dormancy. *Plant Cell* **9**, 1055-1066.
- Feng, Z., C. Xu, and D. Tao, 2018: Historical gradient boosting machine. *EPiC Series in Computing* **55**, 68-80. <http://doi.org/10.29007/2sdc>
- Friedman, J. H., 2001: Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29**(5), 1189-1232.
- Hu, Q., C. Lin, Y. Guan, M. S. Sheteiwiy, W. Hu, and J. Hu, 2017: Inhibitory effect of eugenol on seed germination and pre-harvest sprouting of hybrid rice (*Oryza sativa* L.). *Scientific Reports* **7**, 5295.
- Huang, D., P. Yan, X. Xiao, J. Zhu, X. Tang, A. Huang, and J. Cheng, 2019: The tri-pole relation among daily mean temperature, atmospheric moisture and precipitation intensity over China. *Global and Planetary Change* **179**, 1-9. <https://doi.org/10.1016/j.gloplacha.2019.04.016>
- Im, E., N. Thanh, Y. Kim, and J. Ahn, 2019: 2018 summer extreme temperatures in South Korea and their intensification under 3 °C global warming. *Environmental Research Letters* **14**, 094020. <https://doi.org/10.1088/1748-9326/ab3b8f>
- Jeong, G. A., J. Y. Park, Y. L. Shin, S. J. Lee, and C. J. Lee, 2019: Quality characteristics of noodles supplemented with rice flour and alkaline reagent. *Korean journal of Food and Nutrition* **51**(3), 237-242. <https://doi.org/10.9721/KJFST.2019.51.3.237>
- Ju, Y. C., S. W. Han, J. S. Park, and K. Y. Park, 2000: Effective Screening Method for Viviparous Germination of Rice. *Korean Journal of Crop Science* **45**(2), 103-107.
- Kang, H., 2017: An analysis on the characteristics of rice consumption in Korea. *Korean Journal of Agricultural Economics* **58**(3), 205-214.
- Kang, S., J. Shon, H. Kim, S. Kim, J. Choi, J. Park, Y. Yoon, J. Sim, and W. Yang, 2018: Analysis of genetic variation in pre-harvest sprouting at different cumulative temperatures after heading of rice. *Korean Journal of Crop Science* **63**(1), 8-17. <https://doi.org/10.7740/kjcs.2018.63.1.008>
- Kim, M., 2013: Review on rice flour manufacturing and utilization. *Journal of Biosystems Engineering* **38**(2), 103-112. <http://dx.doi.org/10.5307/JBE.2013.38.2.103>
- Kim, S. H., 2007: Cultural perspectives and current consumption changes of cooked rice in Korean diet. *Nutrition Research and practice* **1**, 8-13.
- Kim, S. S., and H. Y. Chung, 2018: Effects of the type and percentage of a lipid on the quality characteristics of rice bread depending on the use of wet and dry rice flour. *Korean journal of Food and Nutrition* **31**(4), 511-519.
- Lee, G., Y. Jeon, H. Lee, and D. Hyun, 2018: Variation in pre-harvest sprouting resistance, seed germination and changes in abscisic acid levels during grain development in diverse rice genetic resources. *Plant Genetic Resources* **16**(1), 18-27. <https://doi.org/10.1017/S1479262116000319>
- Lee, H., W. Hwang, J. Jeong, S. Yang, Y. Lim, M. Choi, N. Jeong, C. Lee, and K. Choi, 2019: Studies on the temperature response and critical day-length affecting the heading date of major cultivating rice varieties in recent Korean paddy field. *Korean Journal of Crop Science* **64**(4), 323-335. <https://doi.org/10.7740/kjcs.2019.64.4.323>
- Lee, J., E. Ahn, W. Ha, E. Jeong, Y. Won, Y. Jeon, S. Lee, K. Jung, J. Chang, and J. Lee, 2015: Viviparous characteristics of four Korean Japonica Rice cultivars adapted to tropical region. *Journal of the Korean Society of international Agriculture* **27**(2), 168-171. <https://doi.org/10.12719/KSIA.2015.27.2.168>
- Ministry of Agriculture, Food and rural Affairs (MAFRA), 2019: "The damaged rice by typhoon, purchasing all purchases of desired production", Retrieved from <https://www.mafra.go.kr/mafra/293/subview.do?enc=Zm5jdDF8QEB8JTJGYmJzJTJGbwFmcmEIMkY2OCUyRjMyMTU1OSUyRmFydGNvSv>

- mllldy5kbyUzRmJic0NsU2VxJTNEJTl2cmdzRW5kZ  
GVTdHlIM0QIMjZiYnNPcGVuV3JkU2VxJTNEJTl2  
cGFzc3dvcmQIM0QIMjZzcmNoQ29sdW1uJTNEc2ol  
MjZwYWdlJTNEODEIMjZyZ3NCZ25kZVN0ciUzR  
DIwMTkuMDkuMjYIMjZyb3clM0QxMCUyNmlzV  
mllld01pbmUIM0RmYWxzZSUyNnNyY2hXcmQIM0  
QIMjY%3D
- Natekin, A., and A. Knoll, 2013: Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* **7**, 21.
- Park, H., B. Ku, S. Kang, M. Choi, W. Sang, T. Park, Y. Kim, K. Nam, and B. Kim, 2013: Estimate of translating Dates for Three Maturity Groups of Rice Cultivars to harvest before Chuseok holiday in Korea. *Journal of the Korean society of international Agriculture* **25**(3), 307-313. <https://doi.org/10.12719/KSIA.2013.25.3.307>
- Park, J. S., and H. D. Kim, 2009: Viviparous germination characteristics of rice varieties adaptable to central region of Korea. *Korean Journal of Crop Science* **54**(3), 241-248.
- Park, S., and H. Kim, 2016: Effect of black rice powder levels on quality properties of emulsion-type sausage. *Food Science of Animal Resources* **36**(6), 737-743. <https://doi.org/10.5851/kosfa.2016.36.6.737>
- Park, K. B., and R. K. Park, 1984: Studies on the viviparous germination of Indica × Japonica type varieties in paddy rice. *Korean Journal of Crop Science* **29**(1), 15-18.
- Won, Y., E. Ahn, E. Jeong, J. Chang, J. Lee, K. Jung, U. Hyun, Y. Cho, S. Oh, M. Yoon, B. Kim, and B. Kim, 2019; An opaque endosperm rice cultivar, ‘Hangaru’, suitable for exclusive dry-milling rice flour production. *Korean Journal of Breeding Science* **51**(2), 134-139.
- Won, Y., E. Ahn, K. Jung, H. Hong, U. Hyun, H. Park, M. Yoon, J. Lee, J. Seo, E. Jeong, J. Chang, J. Kwak, and S. Oh, 2020: Agricultural characteristics of ‘Misiru’ with improved dry-milling rice flour trait. *Korean Journal of Breeding Science* **52**(2), 151-157.