

# 농지 휴경면적 효율적 예측에 관한 연구

임청룡\*, 최돈우\*\*, 이항아\*\*

\*한국농어촌공사 농어촌연구원

\*\*경상북도 농업기술원

e-mail: chrylims@ekr.or.kr

## A Study on Efficient Prediction Area of Idle Farmland

Cheong-Ryong Lim\*, Don-Woo Choi\*\*, Hang-Ah Lee\*\*

\*Rural Reserch Institute Korea Rural Community Corporation

\*\*Gyeongsangbuk-do Agricultural Research & Extension Services.

### 요약

이 연구에서는 지역단위 농지 휴경면적 예측모형을 구축함에 있어서 토양DB, 농업총조사 원자료 및 기상DB 등으로 구성된 농업빅데이터를 활용하였다. 예측모형 추정은 3단계로 진행하였으며, 1단계는 기본모형인 회귀분석, 의사결정나무, 인공신경망, 서포트벡터머신 등 기법을 활용하였고, 2단계로는 머신러닝 기법 중 앙상블모형인 배깅, 부스팅, 랜덤포레스트 등 모형들을 활용하였으며, 3단계에서는 딥러닝 기법 중 심층신경망을 활용하였다. 여러 모형 추정결과를 활용하여 도출한 평균오차제곱을 기준으로 모형 간 예측력을 비교하여 최적 모형을 선택하였다. 분석결과 2단계 모형들이 1단계 모형들에 비해 예측력이 우수하고, 3단계 모형이 2단계 모형에 비해 예측력이 우수한 것을 확인하였다. 따라서 예측력 향상을 위해 딥러닝 기법들의 적극적인 활용이 필요하다.

### 1. 서론

경제 발전으로 통한 국민 소득의 지속적인 상승과 도시화 수준의 심화는 농촌지역은 인구 감소를 유발하게 되었고, 농촌지역 인력 감소와 농산업의 완만한 성장은 농지의 휴경 등 문제를 초래할 수 있다. 휴경면적의 발생은 지역생산성 저하와 농가소득 감소 등의 주요 요인으로 작용하여 농촌의 지속 가능한 발전에 악영향을 미칠 수 있으며, 토지 이용의 비효율성을 야기하게 된다.

농지 휴경 관련하여 자료의 유형과 특징에 따라 다양한 방법이 시도되었다. 종속변수가 휴경여부(박인환 외, 2006; Xie et al, 2017)이거나 용도별 토지 이용 변화(Thomas et al, 1999; Ahn et al, 2001)인 경우, 농업인의 상대적인 토지 이용 의사결정에 관한 행태방정식을 확률로써 나타낼 수 있으며 주로 로짓모형으로 모형화하여 분석하였다. 또한 특정 지역 및 규모별 경지 면적 변화를 추정하기 위하여 변화할당효과를 고려하였고(장우석 외, 2008) 정부의 관련 정책에 대한 농업인의 입장 및 태도를 조사하여 반영하기도 하였다(Xie et al, 2017). 아울러 농업빅데이터와 머신러닝 기법인 앙상블 모형을 활용하여 휴경면적과 폐경면적 추정 및 활용에 관한 연구도 진행되었다(최돈우 외, 2019).

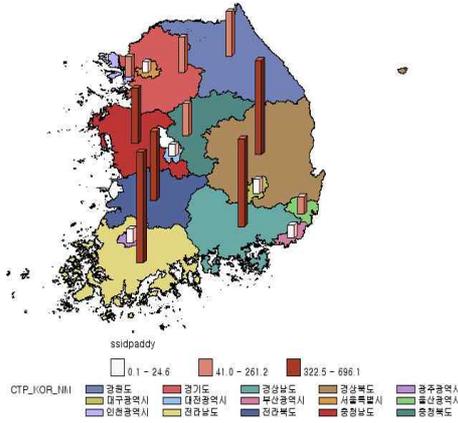
선행연구를 살펴본 결과, 농지이용 또는 휴경에 영향을 미

치는 요인을 확인하는 것에 주안점을 두고 연구를 수행하였으며, 휴경면적을 예측 또는 정확도 제고를 목적으로 수행된 연구들은 미비하였다. 따라서 이 연구에서는 읍·면단위 휴경면적을 예측함에 있어서 휴경에 영향을 미칠 수 있는 요인들을 발굴하고 다양한 모형에 적용하여 최종적으로 예측력이 가장 우수한 모형을 선정하고자 한다.

### 2. 지역별 휴경지 현황 및 분석자료

#### 2.1 시도별 농 휴경 현황

2015년 논에 대한 휴경면적은 0~40ha미만, 40~300ha미만 및 300ha이상 등 3개 그룹으로 구분하면 다음과 같다. 상위그룹에는 전라남도, 경상북도, 경상남도, 전라북도, 충청남도 등 5개 도가 포함되고, 중간그룹에는 강원도, 충청북도, 경기도, 인천광역시, 울산광역시가 포함되며, 하위그룹에는 대구광역시, 부산광역시, 광주광역시, 대전광역시 및 서울특별시도 포함된다.



[그림 1] 시·도별 논 휴경면적

## 2.2 자료

휴경 농지 예측을 위해 읍·면단위 논 면적을 경사도와 배수 등급 수준에 따라 구분하였으며, 농가가구유형을 농작형태, 가구규모, 연령대, 주요업종 등으로 구분하였다. 지역별 차별성을 나타내기 위해 1인당 평균재배면적 및 시군별 GRDP를 활용하였고, 농작물에 영향을 미치는 기상상황을 반영하기 위해 온도, 풍속 및 습도 등 변수들을 활용하였으며, 2010년과 2015년 변수들에 대한 기초통계량은 다음과 같이 나타났다.

[표 1] 분석 자료에 대한 기초통계량 (단위: ha, 가구, 십억원)

구분	2010년		2015년		
	평균	표준편차	평균	표준편차	
논 휴경지	2.11	3.05	2.10	3.05	
경사도	2-7%	293.80	265.74	298.09	283.80
	7-15%	264.78	250.98	263.93	252.02
	15-30%	111.86	329.44	112.19	325.86
	30-60%	105.43	295.04	103.42	291.14
	60-100%	218.93	1846.85	212.70	1820.47
배수등급	매우불량	53.24	144.85	56.75	156.00
	불량	29.83	119.28	32.14	125.55
	약간불량	299.23	372.34	302.26	376.10
	약간양호	256.17	245.06	253.86	245.78
농작형태	양호	431.70	857.77	431.80	849.49
	논벼	100.7	78.8	101.0	78.8
	식량	230.8	145.4	231.0	146.5
	노지채소	278.3	167.3	278.4	169.0
	기타노지	345.1	209.4	344.9	211.9
	시설채소	807.1	483.3	806.6	488.6
	시설기타	1.0	4.0	0.9	3.9
가구	주요과수	7.9	30.0	8.0	30.3
	특작	17.8	27.9	17.9	27.7
	1인가구	12.5	13.8	12.6	13.8
연령대	다인가구	14.4	18.3	14.4	18.1
	40대	17.9	19.5	17.9	19.4
	50대	114.0	116.4	115.1	116.6
	60대	104.8	72.1	104.0	72.1
주요업종	70대이상	407.8	246.8	408.2	249.2
	농업	433.2	432.7	387.6	415.9
	임업	84.4	119.4	78.7	113.6
	어업	72.8	130.1	65.8	118.4
	제조	7,287.3	84,823.7	9,211.3	76,765.8
건설	24.1	62.3	19.4	49.5	

	도소매	9.9	34.9	8.4	27.1
숙박음식	81.0	143.6	48.0	100.5	
가타	22.5	42.1	25.3	37.5	
1인당 경지면적	1.21	0.65	1.21	0.65	
시·군GRDP	2,297	0.003	2,886	0.003	
평균온도	12.91	1.11	12.90	1.11	
평균풍속	1.93	0.58	1.93	0.58	
평균습도	67.35	3.98	67.36	4.01	

자료: 통계청, 농촌진흥청

## 3. 분석모형 및 추정방법

### 3.1 분석모형

이 연구에서는 휴경농지 예측모형을 구축함에 있어서 방법론적 측면에서 다양한 연구들에 대한 고찰을 수행하였으며, 농가의 측면에서 내부적 요인과 외부적 요인들로 변수들을 구성하였다. 또한 분석에 활용될 데이터의 성격에 부합하고 논의 휴경 결정에 영향을 미칠 수 있는 변수들로 분석모형을 다음과 같은 함수형태로 설정하였다.

$$PL_i = F(land_i, agri_i, farm_i, region_i, climate_i) + \epsilon_i \quad (1)$$

식(1)에서  $PL_i$ 는  $i$ 읍·면의 논 휴경면적이고,  $land_i$ 는 토양 특성 변수들에 대한 벡터이고,  $agri_i$ 는 농업특성 변수들에 대한 벡터이며,  $farm_i$ 는 농가특성 변수들에 대한 벡터이다.  $region_i$ 은 지역변수들에 대한 벡터이고,  $climate_i$ 는 기상변수들에 대한 벡터이며, 잔차항은  $\epsilon_i$ 로 설정할 수 있다.

### 3.2 추정방법

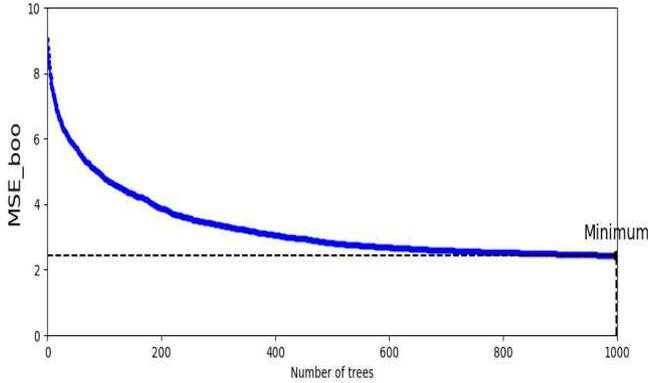
이 연구에서 읍·면별 논 휴경면적 예측모형을 추정함에 있어서 3단계로 거쳐서 분석을 진행하고자 한다. 1단계는 기본모형인 회귀분석, 의사결정나무(Breiman et al., 1984), 인공신경망(Hastie et al., 2009; Izenman, 2008), 서포트벡터머신(Cortes, 1995; Vapnik, 2000) 등 기법을 활용하고, 2단계로는 머신러닝 기법 중 앙상블모형인 배깅(Breiman, 1996), 부스팅(Breiman, 1997; Friedman, 1997), 랜덤포레스트(Breiman, 2001) 등 모형들을 활용하며, 3단계에서는 딥러닝 기법 중 심층신경망(Bengio, 2013; Schmidhuber, 2014)을 활용하고자 한다.

## 4. 휴경면적 예측모형 분석결과 비교

### 4.1 적정 반복회수 설정

앙상블 모형 중 배깅, 부스팅 및 랜덤포레스트 등을 추정함

에 있어서 트리모형에 대한 반복 회수 설정이 필요하며, 이 연구에서는 반복회수를 설정함에 있어서 충분한 크기의 반복회수 설정이후 평균오차제곱(Mean Square Error)이 안정화 되는 상황을 살펴보고 결정하였다. 부스팅일 경우 최대반복회수 1,000회에 가까워짐에 따라 MSE가 안정됨을 알 수 있다(그림 2).



[그림 2] 부스팅 적용 트리 개수에 따른 예측오차 변화

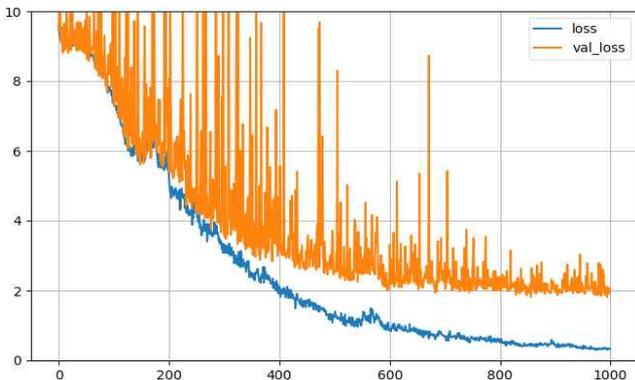
심층신경망 구조를 설정함에 있어서 입력층과 출력층을 제외한 은닉층을 4개로 구성하였으며, 은닉층1-3까지는 200개의 뉴런을 가지고, 은닉층4에는 100개의 뉴런으로 설정하였다(표 2).

[표 2] 심층신경망 추정 형태

구분	출력 형태	파라미터 개수
입력층	37	0
은닉층1	200	40,200
은닉층2	200	40,200
은닉층3	200	40,200
은닉층4	100	20,100
출력층	1	101

총 파라미터: 108,201

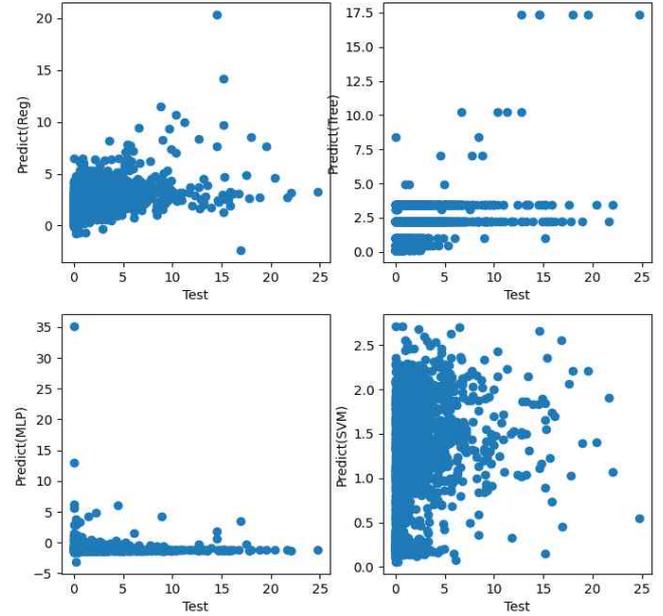
심층신경망 모형의 충분한 학습을 위해 적정 에포크 수가 필요하며 이 연구에서는 1,000회로 설정하였으며, 2010년 자료를 활용하여 학습하고, 2015년 자료를 활용하여 검정하였다. 분석결과 에포크 증가에 따라 학습데이터 MSE(loss)는 지속적으로 감소하지만, 검정데이터 MSE(val\_loss)는 일정 수준에 수렴하는 것을 알 수 있었다(그림 3).



[그림 3] 심층신경망 에포크 회수에 따른 MSE변화

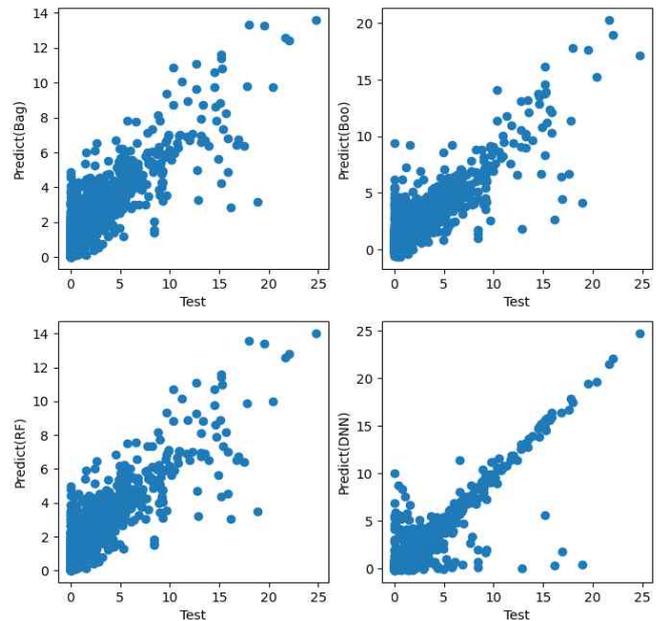
#### 4.2 모형간 예측력 비교

모든 모형들에 있어서 2010년 자료를 활용하여 분석한 후, 2015년 자료를 활용하여 검정을 하였으며, 실측치와 예측값에 대한 산점도를 통해 시각화하였으며, 우선 기본모형들에 대한 산점도를 살펴보면 대각선 산으로 수렴하지 않아서 예측력이 좋지 않음을 알 수 있다(그림 4).



[그림 4] 기초모형에 대한 추정결과 비교

양상블 모형들과 심층신경망 모형에 대한 실측값과 예측값에 대한 산점도를 보면 기본모형 보다 대각선에 잘 분포되어 있으며, 그 중 심층신경망이 가장 집중적으로 나타남을 알 수 있다(그림 5).



[그림 5] 양상블모형과 딥러닝 추정결과 비교

여러 추정모형들에 대한 예측력을 통계량 기준으로 살펴보면, 전체모형 중 심층신경망 모형의 MSE가 가장 작게 나타나 예측력이 가장 우수한 것을 알 수 있다. 또한 예측값과 실측값 평균 비율이나 변이계수비율을 살펴봐도 심층신경망이 가장 우수한 것을 알 수 있다(표 3).

[표 3] 추정모형들에 대한 예측결과 비교

구분	MSE	평균	변이계수	평균비율	변이계수비율	
휴경지	-	2.10	0.69	-	-	
기본모형	회귀	7.63	2.28	1.54	108.44	48.51
	의사나무	7.33	2.07	1.40	98.60	48.43
	신경망	20.52	-1.06	-0.96	-50.29	36.20
	SVM	9.51	1.15	1.76	54.63	21.36
앙상블모형	배깅	3.40	2.21	1.24	105.14	58.48
	부스팅	2.44	2.17	0.91	103.13	77.93
	랜덤포레스트	3.35	2.20	1.23	104.90	58.83
답타닝	심층신경망	1.81	2.05	0.70	97.73	95.83

### 5. 요약 및 결론

이 연구에서는 토양DB, 농업총조사 원자료 등으로 구성된 농업 빅데이터로 회귀분석, 의사결정나무, 인공신경망, 배깅, 부스팅, 랜덤포레스트 및 심층신경망 등 모형들을 활용하여 읍·면단위 휴경면적 예측모형을 추정하였으며, 분석결과를 요약하면 다음과 같다.

첫째, 딥러닝 모형의 예측력이 앙상블모형보다 우수하였으며, 앙상블모형들이 기본모형들보다 예측력이 우수한 것을 알 수 있었다.

둘째, 실측값과 예측값에 대한 평균과 변이계수 비교를 통해 심층신경망 추정결과가 실측값에 가장 근접할뿐만 아니라 변동성도 작아서 안정적인 것을 알 수 있다.

셋째, 앙상블모형에 있어서는 부스팅의 예측력이 가장 우수한 것으로 나타났다.

이러한 분석결과로부터 다음과 같은 시사점을 제시하고자 한다.

첫째, 종속변수에 대한 변수들의 영향력이 아닌 종속변수 예측결과가 분석목표일 경우 딥러닝 모형이 훨씬 효과적임을 알 수 있다.

둘째, 지역 내 휴경농지에 대한 정확한 예측은 휴경농지 규모를 객관적이고 과학적으로 파악하여 지역단위 농지이용계획의 효율적인 수립에 기여할 수 있다.

셋째, 지역별 휴경농지를 적정하게 활용할 경우 농업소득 향상과 지역경제 활성화를 도모할 수 있다.

이 연구에서는 심층신경망 분석을 통해 전통적인 모형들에 비해 예측력이 가장 우수함을 확인하였으나, 기타 딥러닝 모형 간 비교는 수행이 되지 못하였으므로, 추후 딥러닝 모형 간 비교를 통해 더 나은 모형을 구축하고자 한다.

이 연구는 농촌진흥청 연구과제(과제번호: PJ015331062021)의 지원으로 수행되었음.

### 참고문헌

- [1] 박인환·장갑수·서동조, 「다중 로지스틱 모형에 의한 농경지 휴경잠재성 분석」, <환경영향평가>, 15(1): pp53-65, 2006.
- [2] 장우석·정남수·박기욱, 「변화할당효과를 고려한 논 면적 예측 모형의 개발」, <한국농공학회논문집>, 50(3):pp83-89, 2008.
- [3] 최돈우·김동춘·이항아·임청룡, 「휴·폐경 농지 효율적 활용에 대한 연구 -성주 참외와 경산 복숭아를 중심으로-」, <농업생명과학연구>, 53(4):pp137-144 2019.
- [4] Bengio, Y. A. Courville, and P. Vincent., "Representation Learning: A Review and New Perspectives," IEEE Trans. PAMI, special issue Learning Deep Architectures, 2013
- [5] Breiman L, Fredman J, Olshen R and Stone C., Classification and regression trees, Wadsworth, 1984.
- [6] Breiman L, Bagging predictors. Machine Learning, 26: 123-140, 1996.
- [7] Breiman, L. Arcing The Edge, Technical Report 486. Statistics Department, University of California, Berkeley. 1997.
- [8] Breiman L, Random Forests. Machine Learning. 45: 5-32, 2001.
- [9] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". 《Machine Learning》 20 (3): 273. , 1995.
- [10] Hastie T, Tibshirani R, and Friedman J. The element of statistical learning: data mining, inference, and prediction(2nd edition). Springer New York. 2009.
- [11] Izenman A.J. Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer New York. 2008
- [12] Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine, Springer New York. 1999.
- [13] Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine.. 1999.
- [14] Schmidhuber, J. "Deep Learning in Neural Networks: An Overview" <http://arxiv.org/abs/1404.7828>, 2014.
- [15] SoEun Ahn, Andrew J. Plantinga, Ralph J. Alig., Historical Trends and Projections of Land Use for the South-Central United States, U.S. Department of Agriculture Forest Service, Pacific Northwest Research Station, 2001.
- [16] Tomas E. Mauldin, Andrew J. Plantinga, Ralph J. Alig., Determinant of Land Use in Maine with Projections to 2050, NJAF, 16(2):82-88, 1999.
- [17] Vapnik, V. Support Vector Machines. The nature of statistical learning theory. New York: Springer-Verlag New York. 2000.
- [18] Xie Hualin, Cheng Lingjuan, Lv Tianguai, Factors Influencing Farmer Willingness to Fallow Winter Wheat and Ecological Compensation Standards in a Groundwater Funnel Area in Hengshui, Hebei Province, China, Sustainability, 9:1-18, 2017.