

한우 도체형질의 합성곱신경망을 이용한 유전체 예측 정확도 추정

장명진^{1,2}, 임다정¹, 박원철¹, 박종은^{1*}

¹농촌진흥청, 국립축산과학원 동물유전체과

²전북대학교 공과대학 정보통신학과

Estimation of Genetic Prediction Accuracy Using Convolutional Neural Network in Hanwoo

Myoungjin Jang^{1,2}, Dajeong Lim¹, Woncheoul Park¹, Jong-Eun Park^{1*}

¹Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, RDA

²Information and Communication, College of Engineering, Jeonbuk National University

요약 본 연구는 기계학습 기법을 활용하여 유전체 예측 기법을 시험하고 기존의 기법과 비교해보고자 실시되었다. 본 연구에서는 기계학습의 일종인 딥러닝 (DL) 기법, 유전체 최적 선형 불편 예측 (GBLUP) 기법과 통합기법인 앙상블 (Ensemble) 기법이 사용되었다. 데이터는 한우 7,324 두의 유전자형 자료 (37,712 SNPs)와 4가지 도체형질 자료 (등지방두께, 도체중, 등심단면적, 근내지방도)가 이용하여 예측 분석을 진행하였다. 5배수 교차 검증을 사용하여 정확도를 예측하였고, 이를 시험 데이터에 적용하여 상관도를 계산하였다. 제한최대우도법 (REML)을 이용하여 추정된 유전력은 근내지방도가 0.44 ± 0.02 로 가장 높게 나타났다. 계산된 상관도는 도체중과 등심단면적이 가장 높게 나타났으며, 유전상관계수는 0.79 ± 0.01 , 표현형상관계수는 0.52 ± 0.02 이다. 유전체 예측 정확도는, 단일 분석에 있어서 GBLUP 기법이 등지방두께 0.34 ± 0.01 , 도체중 0.41 ± 0.01 , 등심단면적 0.37 ± 0.01 , 근내지방도 0.41 ± 0.01 로 높게 나타났다. 또한, 예측 성능을 향상시키기 위해 가중치를 주는 통합기법인 앙상블 기법을 사용하였을 때 등지방두께 0.35 ± 0.01 , 도체중 0.42 ± 0.01 , 등심단면적 0.38 ± 0.01 로 더 높게 나타났다. 따라서 다양한 기법들을 통해 유전체 예측의 정확도를 높일 수 있을 것으로 사료되며, 정확도가 높은 예측 모델을 개발하여 육종산업에 많은 기여를 하길 기대한다.

Abstract This study was conducted to test genomic prediction using machine learning and to compare predictions with those of existing techniques. In this study, DL, which is a type of machine learning, and GBLUP and Ensemble, which are integrated techniques, were used. Data were predicted and analyzed using 7,324 genotype data (37,712 SNPs) of Korean cattle and data of four carcass traits. Accuracies were predicted using 5-fold cross-validation and correlations were calculated using test data. Heritability estimated using REML was highest at 0.44 ± 0.02 for MS. Regarding calculated correlations, the strongest relationship was observed between CWT and EMA, which was 0.79 ± 0.01 for genetic and 0.52 ± 0.02 for phenotypic correlations. Single analysis showed prediction accuracies were 0.34 ± 0.01 for BFT, 0.41 ± 0.01 for CWT, 0.37 ± 0.01 for EMA, and 0.41 ± 0.01 for MS, but Ensemble produced significantly better coefficients, that is, 0.35 ± 0.01 for BFT, 0.42 ± 0.01 for CWT, and 0.38 ± 0.01 for EMA. We conclude that the accuracy of genomic prediction can be improved by using various techniques, and consider that the findings of this study may be of considerable value to the breeding industry, as they demonstrate the feasibility of developing highly accurate prediction models.

Keywords : Hanwoo, Carcass Weight, Deep Learning, Genetic Prediction, Accuracy

본 논문은 농촌진흥청 기관고유사업(세부과제번호: PJ01476401, 한우 멀티오믹스 데이터 활용 플랫폼 구축 연구)의 지원에 의해 이루어진 것입니다.

*Corresponding Author : Jong-Eun Park(National Institute of Animal Science)

email: jepark0105@korea.kr

Received January 27, 2022

Revised March 7, 2022

Accepted April 1, 2022

Published April 30, 2022

1. 서론

유전체는 유전자와 염색체의 합성어로, A, C, G 및 T로 표시되는 단일 DNA 염기서열로 구성되며 한 개체의 모든 유전정보를 포함한다. 현재, 유전체 데이터를 기반으로 특정 표현형에 대한 마커를 개발하고 예측에 활용하는 연구들은 농축산 유전체연구의 중요한 분야이다. FAANG (Functional Annotation of Animal Genome) 프로젝트는 다양한 유전정보를 표현형에 연계하기 위한 목적으로 2015년에 처음 시작되어, 북미와 유럽 외 약 15개 국가에서 참여하고 있으며, 유전자의 기능 연구를 진행함에 따라 표현형·유전적 다양성 연구에 기여를 하고 있다. 유전체 예측 (또는 선발, GP or GS: Genomic Prediction or Selection, 이하 GS)은 전장 유전체 변이마커를 이용하여, 개체의 유전적 잠재성을 예측하는 방법으로 육종 및 의학 분야에서 널리 사용되고 있다[1,2]. 지금까지 유전체 예측을 위한 다양한 방법론들이 제안되었으며, 마커 효과 방법[1,3,4]과 유전체 최적 선형 불변 추정법 (GBLUP: Genomic Best Linear Unbiased Prediction, 이하 GBLUP)[5]으로 구분할 수 있다[2]. 최근에는, 유전체 예측에서 표현형과 연관된 유전적 구조, 유전자 기능, 조절 패턴, 상호작용 유전자 네트워크 등의 생물학적인 지식을 접목하여, 유전체 예측의 정확도를 개선하는 연구가 이루어졌다 [6-8]. 이러한 연구를 통해 육종프로그램에 적용하여 세대 간격을 단축시키거나, 체중, 사료효율 같은 경제형질을 높이는 이점이 작용할 수 있다. 또한, 비모수 기반의 기계학습 기법을 적용하는 유전체 예측 연구들도 발표되었다[9-11].

기계학습은 관측치보다 모수의 수가 많은 빅데이터를 분석 방법으로 널리 사용되고 있다. 특히, 이미지, 음성 인식, 텍스트 마이닝 (Text Mining) 등에 주로 사용되어 왔으며, 모델을 설정하지 않아도 되는 기계학습 알고리즘은 대용량 데이터로부터 숨어있는 정보를 추출해내는 장점을 갖고 있다[12]. 기계학습의 일종인 딥러닝 (DL: Deep Learning)은 최근 개발된 머신러닝 기술로, 합성곱신경망으로 이루어진 구조이다. 합성곱신경망은 수많은 뉴런이 대용량 데이터로부터 복잡한 비선형 관계를 추출한다[13]. 이러한 기계학습 기법이 적용된 최근 연구로는 딥러닝 기반 항생제 내성균 감염을 예측하는 연구 [14], 약물에 노출된 암세포주의 생존력을 예측하는 연구 [15], 단백질의 3차원 구조에서 중앙 억제 유전자 및 원 중앙유전자를 예측하는 연구[16] 등이 있다.

본 연구에서는 한우의 4가지 도체형질에 대하여 기계 학습 기법을 활용하여 유전체 예측 기법을 시험하고 기존의 GBLUP 기반의 기법과 비교 분석을 진행하고자 한다.

2. 재료 및 방법

2.1 공시재료

본 연구에서 사용된 데이터는 국립축산과학원에서 보유 중인 한우 7,324 두 (수컷: 7,147 두, 암컷: 536 두)의 정보를 사용하였다. 이 개체들은 2014년 5월부터 2017년 6월까지 145 개의 농장에서 태어났다. 수집된 도체 형질은 총 4가지; 등지방두께 (BFT: Back Fat Thickness), 도체중 (CWT: Carcass Weight), 등심단면적 (EMA: Eye Muscle Area), 근내지방도 (MS: Marbling Score)이다. 등지방두께는 mm 단위로 피하 지방 두께를 측정하였다. 도체중은 kg 단위로 도축 후 냉장된 도체의 무게를 측정하였다. 등심단면적은 cm^2 단위로 등심 부분의 절개면 면적을 측정하였다. 근내지방도는 등심 단면의 지방 침착도를 1부터 9까지 score로 구분하여 평가하였다. 이들은 평균 30 개월령에 도축되었으며, 축산물 품질평가원 (KAPE: Korea Institute for Animal Products Quality Evaluation)에서 제시한 지침에 따라 도체 형질을 측정하였다. 이 연구에 대한 윤리 승인은 대한민국 농촌진흥청 국립축산과학원 동물관리 및 사용 위원회에서 승인 (2018-293)되었다.

각 계놈의 DNA는 DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, United States)를 사용하여 조직 샘플에서 추출되었다. 샘플은 맞춤형 Hanwoo 50K SNP 칩 (58,990 SNPs: Single Nucleotide Polymorphism)을 사용하여 유전자형을 분석하였다. 이중 대립형 SNP 유전자형은 동형 및 이형 접합체로 대립 유전자 쌍으로 나타낼 수 있으며, 동형 접합 유전자형의 경우는 AA와 BB가 있으며, 이형 접합 유전자형은 AB의 경우이다. PLINK 소프트웨어[17]를 사용하여 이러한 유전자형을 AA는 0, AB는 1, BB는 2로 각각 표현할 수 있다. 품질 관리를 위해 다음 임계 값 수준이 적용되었다. MAF < 0.01 또는 call rate < 0.90 및 Hardy-Weinberg 불균형 < 0.000001 인 경우 SNP가 제거되었다. 0번 염색체와 성 염색체에 위치한 유전자형도 제외되었다. 또한 10 % 이상의 유전자형이 누락 된 개체는 제거되었다. 이러한 전처리 기준을 적용한 후 37,712 개의 SNP가 분석에 사용되었다.

2.2 유전상관

본 연구에서는 MTG2 소프트웨어[18]를 사용하여 유전분산을 다음과 같이 추정하였다.

$$r_g = \frac{cov_g}{\sqrt{V_{g1}V_{g2}}} \quad (1)$$

r_g 는 유전상관, cov_g 는 형질1과 형질2 사이의 유전적 공분산, V_{g1} 은 형질1의 유전적 분산, V_{g2} 는 형질2의 유전적 분산이다.

2.3 표현형 보정 및 유전력

연구에 사용된 표현형은 보정된 변수로써, PREDICTF90의 혈통 기반 모델을 사용하여 단일 특성 분석에서 고정효과에 대해 보정되었다[19]. 분석 시 고정 효과에 사용된 형질은 출생날짜 (birthdate), 도축날짜 (slaughter date), 나이 (age), 성별 (sex), 도축장소 (slaughter place)이다.

$$y = Xb + Za + e \quad (2)$$

y 는 해당 도체형질의 관측치에 대한 벡터이며, b 는 모든 고정효과에 대한 추정치 벡터이고, a 는 모든 임의효과 벡터이다. X 와 Z 는 각각 고정효과와 임의효과에 관한 계수행렬이고, e 는 랜덤 잔차 효과의 벡터이다.

본 연구에서는 단일 형질에 대해 상가적 유전효과와 유전모수 및 육종가 추정을 위해 제한최대우도법 (REML: Restricted Maximum Likelihood)[19]을 사용하였다. GCTA 소프트웨어[20]를 통해 각 형질에 대한 분산 값과 유전력을 추정하였다.

$$h^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \quad (3)$$

h^2 는 유전력이며, σ_a^2 는 상가적 유전분산, σ_e^2 는 임의환경 효과의 분산이다.

2.4 유전체 예측

유전체 예측 분석을 위해 GBLUP, Deep Learning, Ensemble 3가지 기법을 사용하였다.

2.4.1 유전체 최적선형불편추정 (GBLUP)

$$y = \mu + s_i\alpha_i + Zg + e \quad (4)$$

y 는 관찰된 표현형의 벡터이고, μ 는 전체 평균, s_i 는 SNP에 대한 유전자형 벡터 (0, 1, 2로 인코딩됨), α_i 는 마커 효과 (대립유전자 치환 효과), g 는 모든 개체의 유전자 육종 값의 벡터, Z 는 g 와 y 를 연결하는 결합 행렬, e 는 무작위 잔차 효과의 벡터이다. MTG2 소프트웨어 (ver 2.21)를 통해 분산성분을 추정하고, 유전체 육종가를 계산하였다.

2.4.2 딥러닝 (DL)

DL 기법을 사용한 유전체 예측은 합성곱신경망 (CNN: Convolutional Neural Network)을 기반으로 한 DeepGS[22]를 사용하였다. DeepGS에 적용된 합성곱신경망은 8-32-1구조; 하나의 입력 층 (코딩된 개체별 유전자형), 8개의 뉴런으로 구성된 하나의 합성곱 층, 하나의 샘플링 층, 3개의 드롭아웃 층, 32개와 1개의 뉴런으로 구성된 두 개의 완전 연결 층 그리고 하나의 출력 층으로 구성되어 있다. 합성곱 층과 첫 번째, 완전 연결 층에서는, 정류 선형 유닛 (ReLU) 함수를 적용하였다. 첫 번째와 두 번째 완전 연결 층 간에는 0.05의 dropout 비율을 설정하였다. 두 번째 완전 연결 층의 출력 값은 선형 회귀 모델을 통해 개체의 최종 예측 표현형으로 출력된다.

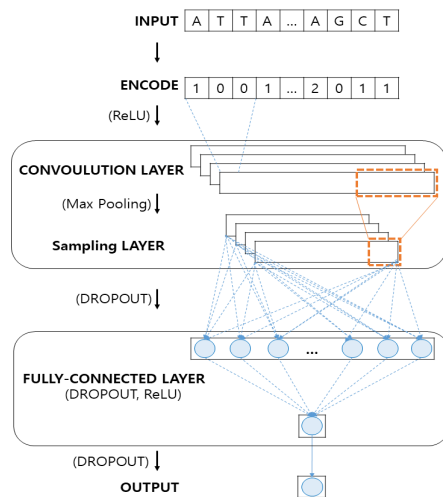


Fig. 1. Deep convolutional neural network architecture in Deep Learning

2.4.3 GBLUP와 DL기반의 앙상블

DL, GBLUP 성능을 평가하는 것 이외에도 예측 성능을 향상시키기 위해 DL과 GBLUP의 선형결합형태의 통합 (Ensemble) 학습 모델을 통하여 분석을 진행하였다.

$$predict_l = (W_D \times predict_D + W_R \times predict_R) / (W_D + W_R) \quad (5)$$

predict는 각 기법에 의한 추정치이며, W는 각 기법에 대한 가중치이다. 아래 첨자 D는 딥러닝에 의한 결과, R은 GBLUP 기법에 의한 결과이다. 매개변수(W)는 PSO (particle swarm optimization) 알고리즘[23]을 사용하여 교차검증을 통해 해당 검증 집단에 최적화하였다.

2.5 예측 정확도

GS 모델의 예측 성능을 평가하기 위해 5배수 교차검증을 사용하였다. 전체 데이터 (7,324 두)를 대략적으로 동일한 크기의 그룹으로 무작위 분할하는 5배 교차검증을 실행하였다. 데이터 세트는 4 그룹 (~6,860 두)의 유전자형 및 표현형 데이터를 사용하여 훈련 세트로 사용하고 나머지 1 그룹 (1,468 두)은 유전자형만 사용하여 표현형 값을 예측하는 모델을 만들었다. 이 과정을 5번 반복하여 예측된 표현형 값과 비교 검증하였다.

한우의 4 가지 도체형질의 실제 표현형 값과 검증집단 (test set) 나온 예측값의 상관도 (correlation, r^2)와 표준오차를 요약하였다.

3. 결과 및 고찰

3.1 표현형 분석

국립축산과학원에서 보유 중인 한우의 4 가지 도체형질에 대한 평균과 표준편차는 Table 1과 같다. 등지방두께 (BFT)는 14.23 ± 13.36 mm, 도체중 (CWT)은 439 ± 879.78 kg, 등심단면적 (EMA)은 96.15 ± 29.21 cm²이고 근내지방도 (MS)는 5.99 ± 1.98 이다.

도체성적에 대한 이전 연구결과와 비교해 보았을 때 먼저 등지방두께는 한우 93,992 두[24]는 13.5 ± 5.27 mm, 한우 5,843 두[25]는 12.36 ± 4.79 mm로 보고하였고, 경남지역의 한우 9,592 두[26]는 13.57 ± 4.61 mm로 본 연구보다 높게 보고되었으며, 강원도 지역의 한우 181,283 두[27]는 12.85 ± 5.25 mm로 본 연구보

Table 1. Summary statistics of carcass traits in Hanwoo

	Min.	Median	Mean	Max.	SD*
BFT	1	14	14.23	45	13.36
CWT	159	440	439	645	879.78
EMA	35	95	96.15	155	29.21
MS	1	6	5.99	9	1.98

*SD: standard deviation
 *BFT: Back Fat Thickness
 *CWT: Carcass Weight
 *EMA: Eye Muscle Area
 *MS: Marbling Score

다 낮은 결과를 나타내었다. 도체중은 한우 93,992 두 [24]는 437.71 ± 52.52 kg, 한우 5,843 두 [25]는 423.37 ± 42.52 kg로 보고하였고, 경남지역의 한우 9,592 두[26]는 452.50 ± 46.60 kg으로 본 연구보다 높게 보고되었으며, 강원도 지역의 한우 181,283 두[27]는 399.99 ± 67.38 kg으로 본 연구보다 낮은 결과를 나타내었다. 등심단면적은 한우 93,992 두[24]는 94.34 ± 13.85 cm², 한우 5,843 두[25]는 90.39 ± 9.45 cm²로 보고하였고, 경남지역의 한우 9,592 두[26]는 95.42 ± 11.50 cm²로 본 연구보다 높게 보고되었으며, 강원도 지역의 한우 181,283 두[27]는 88.27 ± 11.76 cm²로 본 연구보다 낮게 보고되었다. 근내지방도는 한우 93,992 두 [24]는 5.94 ± 2.01 , 한우 5,843 두 [25]는 5.34 ± 1.91 로 보고되었고, 경남지역의 한우 9,592 두[26]는 6.38 ± 1.78 로 본 연구보다 높게 보고되었으며, 강원도 지역의 한우 181,283 두[27]는 5.02 ± 2.06 으로 전국 한우 도체기록보다 더 낮게 보고되었다. 본 연구의 자료와 차이가 나는 원인으로는 1)이전 연구보다 본 연구의 한우 도체성적이 향상되었음을 확인하였고, 2)지역의 차이에 따라 한우의 개량 효과가 다르다는 것으로 사료된다.

3.2 유전모수 추정

Table 2는 한우의 도체형질을 이용한 유전력, 유전분산, 환경 분산의 결과이다. 본 연구에서 추정된 유전력은 등지방두께가 0.35 ± 0.02 , 도체중이 0.39 ± 0.02 , 등심단면적이 0.36 ± 0.02 , 근내지방도가 0.44 ± 0.02 로 나타났다.

한우 970,141 두[28]의 등지방두께와 근내지방도의 유전력의 평균이 순서대로 0.35 ± 0.03 , 0.48 ± 0.07 로 본 연구보다 높게 추정되었고, 도체중, 등심단면적의 유전력 평균은 순서대로 0.28 ± 0.02 , 0.28 ± 0.03 로 다른 형질에 비해 낮게 추정되었다. 강원도 지역의 한우

Table 2. Variance component and genetic heritability for carcass traits by REML in Hanwoo

	V(G)	V(e)	h ²
BFT	7.71±0.57	14.57±0.42	0.35±0.02
CWT	824.95±54.1	1270.27±37.34	0.39±0.02
EMA	45.89±3.25	83±2.36	0.36±0.02
MS	1.33±0.08	1.71±0.05	0.44±0.02

*BFT: Back Fat Thickness
 *CWT: Carcass Weight
 *EMA: Eye Muscle Area
 *MS: Marbling Score

181,283 두[27]의 각 형질별 추정된 유전력은 도체중이 0.53, 등지방두께가 0.43, 등심단면적이 0.38, 그리고 근내지방도가 0.54로 본 연구보다 높은 추정치를 보고하였다. 사전에 연구된 결과와 비교해 보면 형질에 따라 추정된 유전력은 다소 차이를 보였으나 지역적 차이가 있는 것으로 보아 환경적 영향이 큰 것으로 사료된다.

Table 3은 한우의 도체형질 유전적 및 표현형 상관관계를 추정한 것이다. 유전적 상관계수는 0.281 ~ 0.794 범위 내였으며, 등지방두께와 도체중, 등심단면적과 근내지방도, 도체중과 등심단면적이 0.699, 0.778, 0.794 순으로 높은 상관관계를 보였다. 표현형 상관계수는 0.091 ~ 0.515 범위 내였으며, 마찬가지로 등지방두께와 도체중, 등심단면적과 근내지방도, 도체중과 등심단면적이 0.383, 0.451, 0.515 순으로 높은 상관관계를 보였다.

Table 3. Correlation for carcass traits in Hanwoo

	BFT	CWT	EMA	MS
BFT		0.699 ±0.016	0.356 ±0.022	0.281 ±0.024
CWT	0.383 ±0.02		0.794 ±0.01	0.495 ±0.019
EMA	0.07 ±0.023	0.515 ±0.017		0.778 ±0.013
MS	0.091 ±0.023	0.204 ±0.022	0.451 ±0.018	

*upper triangular matrix: genetic correlation,
 *lower triangular matrix: phenotypic correlation
 *BFT: Back Fat Thickness
 *CWT: Carcass Weight
 *EMA: Eye Muscle Area
 *MS: Marbling Score

사전에 연구된 한우 970,141 두[28]의 연구를 살펴보면 추정된 유전상관계수는 -0.29 ~ 0.58 범위 내였으며, 등지방두께와 도체중, 등심단면적과 근내지방도, 도체중과 등심단면적의 상관관계가 높게 추정된 경향을 보였다.

며, 표현형 상관계수도 0.11 ~ 0.5 범위 내로, 등지방두께와 도체중, 등심단면적과 근내지방도, 도체중과 등심단면적의 상관관계가 높게 추정되는 본 연구와 비슷한 경향을 보였다.

3.3 예측 정확도 비교

등지방두께의 경우, 예측 정확도가 DL, GBLUP 순으로 0.26, 0.35로 GBLUP의 값이 높게 나왔으며 Ensemble의 예측 정확도는 0.35로 단일 모델보다 성능이 개선되었다. 도체중의 경우, DL, GBLUP순으로 0.35, 0.41로 나왔으며, Ensemble의 예측 정확도는 0.42로 단일 모델보다 성능이 개선되었다. 등심단면적의 경우, DL, GBLUP 순으로 0.33, 0.38로 GBLUP의 값이 높게 나왔으며 Ensemble의 예측 정확도는 0.38로 단일 모델보다 성능이 개선되었다. 근내지방도의 경우, DL, GBLUP 순으로 0.36, 0.42로 GBLUP의 값이 높게 나왔으며 Ensemble의 예측 정확도는 0.38로 단일 모델보다 성능이 개선되었다.

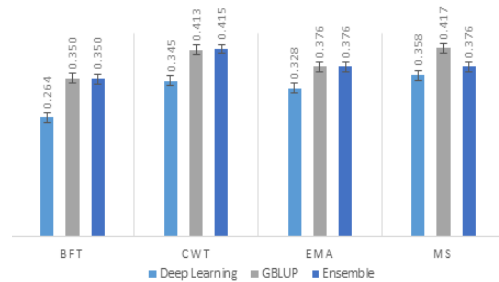


Fig. 2. Predictive accuracy for carcass traits in Hanwoo

미국 홀스타인 황소 11,790 두[29] 58K SNPs에 대하여 합성곱신경망 (CNN), GBLUP 등의 방법으로 형질 예측 성능을 비교하였다. 그 결과 GBLUP은 0.33, CNN은 0.29 순으로 예측 정확도가 GBLUP이 더 높게 측정되었다. GBLUP의 또 다른 기법인 rrBLUP을 사용하여 CNN 등의 성능 예측을 비교한 650개 라인의 밀 데이터 [30]에 대하여, 곡물수확량에 대한 예측 정확도는 GBLUP, CNN 순으로 0.39 0.39이고, 곡물 단백질 함량에 대한 예측 정확도는 GBLUP, CNN 순으로 0.48 0.48이다. 테스트 무게에 대한 예측 정확도는 GBLUP, CNN 순으로 0.45 0.47이고, 식물 높이에 대한 예측 정확도는 GBLUP, CNN 순으로 0.52, 0.55이며, 표제 날짜에 대한 예측 정확도는 GBLUP, CNN 순으로 0.46, 0.49로 본 연구와 반대로 CNN의 성능이 더 높게 측정

되었다. 데이터의 조건 및 분석 기법에 따라 예측 정확성의 차이가 있으며, 이러한 경우 통합 기법을 이용하는 등의 추가 연구를 통해 기능을 더 개선시킬 수 있을 것으로 사료된다.

4. 결론

본 연구는, 한우의 유전체 정보를 활용하여, 도체형질의 분산 성분과 유전모수 (유전력 및 유전상관)를 추정했다. 추정된 유전력은 0.35에서 0.44 사이였으며, 유전적 및 표현형 상관관계는 등지방두께와 도체중, 등심단면적과 근내지방도, 도체중과 등심단면적의 상관관계가 높게 추정된 경향을 보였다. 또한, 네 가지 도체형질에 대한 예측 정확도를 여러 가지 기법을 통해 조사 및 비교 해본 결과, DL보다 GBLUP의 예측 정확도가 높게 나타났다. 개별 분석을 조합한 Ensemble 기법을 통해서 예측 정확도를 높일 수 있었다. 이처럼, GS 분야의 기존 기법과 더불어 새로운 기법을 적용시키고 비교 분석을 통해 성능을 확인하여 보다 정확한 예측이 이루어질 것으로 사료된다. 앞으로도, 정확도가 높은 한우 능력 예측 모델을 개발하여 한우 육종산업에 많은 기여를 하길 기대한다.

References

- [1] Meuwissen, T. H., Hayes, B. J., Goddard, M. E., "Prediction of total genetic value using genome-wide dense marker maps.", *Genetics*, Vol.157, No.4, pp.1819-1829, 2001.
DOI: <https://doi.org/10.1093/genetics/157.4.1819>
- [2] N. Gao, J. Teng, S. Ye, X. Yuan, S. Huang, "Genomic prediction of complex phenotypes using genic similarity based relatedness matrix.", *Frontiers in genetics*, Vol.9, pp.364, Aug. 2018.
DOI: <https://doi.org/10.3389/fgene.2018.00364>
- [3] D. Habier, R. L. Fernando, K. Kizilkaya, D. J. Garrick, "Extension of the bayesian alphabet for genomic selection.", *BMC Bioinformatics*, Vol.12, pp.186, May. 2011.
DOI: <https://doi.org/10.1186/1471-2105-12-186>
- [4] D. Gianola, "Priors in whole-genome regression: the Bayesian alphabet returns.", *Genetics*, Vol.194, pp.573-596, Jul. 2013.
DOI: <https://doi.org/10.1534/genetics.113.151753>
- [5] P. M. VanRaden, "Efficient methods to compute genomic predictions.", *Journal of Dairy Science*, Vol.91, issue.11, pp.4414-4423, Nov. 2008.
DOI: <https://doi.org/10.3168/jds.2007-0980>
- [6] Z. Zhang, U. Ober, M. Erbe, H. Zhang, N. Gao, "Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies.", *PLoS one*, Vol.9, No.3, e93017, Mar. 2014.
DOI: <https://doi.org/10.1371/journal.pone.0093017>
- [7] S. M. Edwards, I. F. Sørensen, P. Sarup, T. F. C. Mackay, P. Sørensen, "Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*.", *Genetics*, Vol.203, pp.1871-1883, May 2016.
DOI: <https://doi.org/10.1534/genetics.116.187161>
- [8] N. Gao, J. W. R. Martini, Z. Zhang, X. Yuan, H. Zhang, "Incorporating gene annotation into genomic prediction of complex phenotypes.", *Genetics*, Vol.207, pp.489-501, Aug. 2017.
DOI: <https://doi.org/10.1534/genetics.117.300198>
- [9] R. M. Nelson, M. Kierczak, Ö. Carlborg, "Higher order interactions: detection of epistasis using machine learning and evolutionary computation.", *Methods Mol Biol.*, Vol.1019, pp.499-518, May 2013.
DOI: https://doi.org/10.1007/978-1-62703-447-0_24
- [10] D. Gianola, R. L. Fernando, A. Stella, "Genomic-assisted prediction of genetic value with semiparametric procedures." *Genetics*, Vol.173, No.3, pp.1761-1776. Jul. 2006.
DOI: <https://doi.org/10.1534/genetics.105.049510>
- [11] D. Gianola, J. B. C. H. M. van Kaam, "Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits." *Genetics*, Vol.178, No.4, pp.2289-2303, Apr. 2008.
DOI: <https://doi.org/10.1534/genetics.107.084285>
- [12] O. Gonzalez-Recio, S. Forni, "Genome-wide prediction of discrete traits using Bayesian regressions and machine learning." *Genet Sel Evol.*, Vol.43, No.1, pp.7, Feb. 2011.
DOI: <https://doi.org/10.1186/1297-9686-43-7>
- [13] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning.", *Nature*, Vol.521, No.7553, pp.436-444. May 2015.
DOI: <http://doi.org/10.1038/nature14539>
- [14] S. W. Oh, H. K. Lee, J. Y. Shin, J. H. Lee, "Antibiotics-Resistant Bacteria Infection Prediction Based on Deep Learning.", *The Journal of Society for e-Business Studies*, Vol.24, No.1, pp.105-120. Feb. 2019.
DOI: <http://doi.org/10.7838/jsebs.2019.24.1.105>
- [15] M. P. Menden, F. Iorio, M. Garnett, U. McDermott, C. H. Benes, "Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties.", *PLoS one*, Vol.8, No.4, e61318, Apr. 2013.
DOI: <http://doi.org/10.1371/journal.pone.0061318>
- [16] A. Tavanaei, N. Anandanadarajah, A. Maida, R. Loganatharaj, "A Deep Learning Model for Prediction of Tumor Suppressor Genes and Oncogenes from PDB

- Structure.”, *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp.613-617. Oct. 2017.
DOI: <http://doi.org/10.1101/177378>
- [17] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.”, *The American Journal of Human Genetic*, Vol.81, No.3, pp.559-575. Sep. 2007.
DOI: <https://doi.org/10.1086/519795>
- [18] S. H. Lee, Van der Werf J. H., “MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information.” *Bioinformatics*, Vol.32, No.9, pp.1420-1422. 2016.
DOI: <https://doi.org/10.1093/bioinformatics/btw012>
- [19] R. R. Corbeil, S. R. Searle, “Restricted maximum likelihood (REML) estimation of variance components in the mixed model.”, *Technometrics*, Vol.18, No.1, pp.31-38. 1976.
DOI: <http://doi.org/10.2307/1267913>
- [20] J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, “GCTA: A Tool for Genome-wide Complex Trait Analysis.”, *American journal of human genetics*, Vol.88, No.1, pp.76-82, Jan. 2010.
DOI: <http://doi.org/10.1016/j.ajhg.2010.11.011>
- [21] S. Srivastava, B. I. Lopez, H. Kumar, M. Jang, H. Chai, “Prediction of Hanwoo Cattle Phenotypes from Genotypes using Machine Learning Methods.”, *Animals*, Vol.11, No.7, pp.2066. Jul. 2021.
DOI: <https://doi.org/10.3390/ani11072066>
- [22] W. Ma, Z. Qiu, J. Song, J. Li, Q. Cheng, “A deep convolutional neural network approach for predicting phenotypes from genotypes.”, *Planta*, Vol.248, No.5, pp.1307-1318, Aug. 2018
DOI: <https://doi.org/10.1007/s00425-018-2976-9>
- [23] J. Kennedy, R. Eberhart, “Particle swarm optimization.”, *Proceedings of ICNN'95 - international conference on neural networks*, Vol.4, pp.1942-1948, Nov. 1995.
DOI: <http://doi.org/10.1109/ICNN.1995.488968>
- [24] J. M. Han, Y. S. Lee, H. S. Kong, “The effect of the evaluation factor for genetic improvement associated with economic traits of Hanwoo.”, *Journal of the Korean Data And Information Science Society*, Vol.29, No.5, pp.1279-1286, Sep. 2018.
DOI: <http://dx.doi.org/10.7465/jkdi.2018.29.5.1279>
- [25] Y. S. Lee, J. Y. Lee, “Estimation of genetic parameter for carcass traits in commercial Hanwoo steer.” *Journal of the Korean Data and Information Science Society*, Vol.27, No.3, pp.741-747, 2016.
DOI: <https://dx.doi.org/10.7465/jkdi.2016.27.3.741>
- [26] J. M. Han, Y. S. Lee, H. S. Kong, “The effect of the evaluation factor for genetic improvement associated with economic traits of Hanwoo.” *Journal of the Korean Data and Information Science Society*, Vol.29, No.5, pp.1279-1286, Sep. 2018.
DOI: <https://dx.doi.org/10.7465/jkdi.2018.29.5.1279>
- [27] C. W. Lee, J. W. Choi, H. J. Shin, J. B. Kim, “Genetic Prediction of Hanwoo Carcass Traits in Kangwon Regional Hanwoo Cow Test Farms.” *Annals of Animal Resource Sci.*, Vol.31, No.1, pp.1-12, 2020.
DOI: <https://doi.org/10.12718/AARS.2020.31.1.1>
- [28] J. H. Son, Y. M. Gu, Y. H. Jung, D. H. Cha, S. J. Kim, “Estimation of Genetic Parameters on Carcass Traits for Estimated Breeding Value in Hanwoo Cow.” *Journal of Agriculture & Life Science*, Vol.54, No.6, pp.73-80, Dec. 2020.
DOI: <https://doi.org/10.14397/jals.2020.54.6.73>
- [29] R. Abdollahi-Arpanahi, D. Gianola, F. Penagaricano, “Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes”, *Genetics Selection Evolution*, Vol.52, No.12, pp.1-15, Feb. 2020.
DOI: <http://doi.org/10.1186/s12711-020-00531-z>
- [30] K. S. Sandhu, D. N. Lozada, Z. Zhang, M. O. Pumphrey, A. H. Carter, “Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program”, *Frontiers in Plant Science*, Vol.11, No.613325, pp.1-13, Jan. 2021.
DOI: <https://doi.org/10.3389/fpls.2020.613325>

장 명 진(Myoungjin Jang)

[정회원]



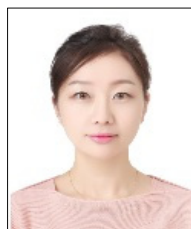
• 2015년 6월 ~ 현재 : 농촌진흥청 국립축산과학원 연구원

<관심분야>

유전체, 정보통신

임 다 정(Dajeong Lim)

[정회원]



• 2007년 11월 ~ 현재 : 농촌진흥청 국립축산과학원 농업연구관

<관심분야>

가축 집단 유전체, 생명정보 적용 유전체 선발

박 원 철(Woncheoul Park)

[정회원]



• 2019년 2월 ~ 현재 : 농촌진흥청
국립축산과학원 농업연구사

<관심분야>

전사체, 유전체

박 종 은(Jong-Eun Park)

[정회원]



• 2015년 7월 ~ 현재 : 농촌진흥청
국립축산과학원 농업연구사

<관심분야>

통계유전, 생물통계