

Opportunities to Improve Alfalfa Forage Yield and Quality in the Southwest USA by Machine Learning Predictive Analytics

Dr. Md. Abdullah Al Bari

Department of Plant and Environmental Sciences

Email: mabari@nmsu.edu

Date: 07/15/2026

Alfalfa at a Glance

Alfalfa (*Medicago sativa* subsp. *sativa* L.)

- Drought-tolerance
 - $2n = 4x = 32$
- Salinity-tolerance
 - Genome size = 793.2 Mb
 - Heterozygous
 - Perennial legume forage
- Extensive root system
- Forage yield
 - 4th Largest field crop
 - Biomass productivity
 - Dry matter yield
- Quality
 - Low lignin, high digestibility
 - High-protein, total feed values (TFV)
 - More leaf to stem ratio
- Persistence
 - Long-lasting, productive stand
 - Recover crop cutting
 - Fall dormancy
- Disease resistance
 - Phytophthora seeding root rot
 - Anthracnose
 - Fusarium wilt
 - Bacterial wilt
 - Root and crown rot
- Insects
 - Lygus
 - Alfalfa weevil
- Aphids
 - Blue alfalfa aphid
 - Spotted alfalfa aphid
 - Pea aphid
- Nematodes
 - Stem nematode
 - Southern root-knot nematode

National long-term average yield is 3.4 tons per acre



New Mexico State University

College of Agricultural, Consumer, and Environmental Sciences | ACES

Plant and Environmental Sciences

BE BOLD. Shape the Future.[®]
New Mexico State University

New Mexico Alfalfa:

○ Alfalfa in 2025

- 125,000 acres
- 738,000 tons
- \$231 per ton
- Value \$170 million

○ Hay is number one cash crop

- Alfalfa is predominant
- Pasture
- Hay
- Silage
- Greenchop

○ Tremendous contribution to growers

- Dairy, beef industry
- Wool market
- Horse racing

○ The relative humidity

- Country's best hay producer



○ Recurrent drought and growing salinity is top priority

- New Mexico acreage is decreasing

○ Efficient Alfalfa Breeding meaning

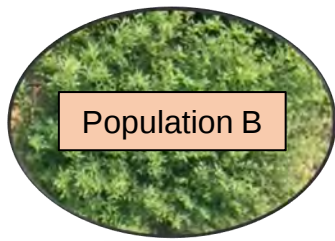
- Long lasting impact of the used dollars

Alfalfa Cultivar

Development ~12 Years



Population A



Population B

Population selection



+

Population C

Crossing in the greenhouse



Evaluation and selection



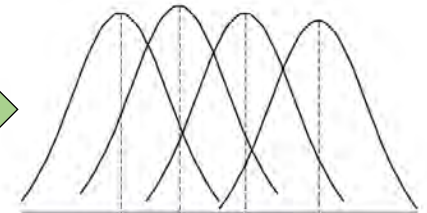
Intermating in cages: Leaf-cutter bees



Multilocation, Multiyear trials



Seed multiplication in cages



Several cycle of RS

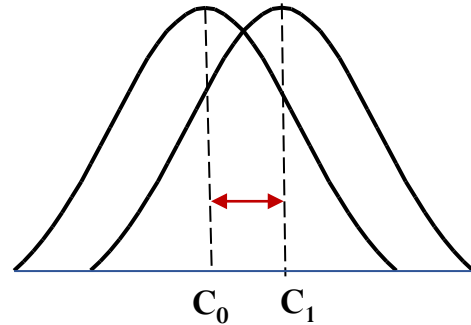
Genomic selection impacting genetic gain

- Genomic selection (GS) uses genome wide markers to predict breeding values (Meuwissen et al. 2001)
- Complex traits can be predicted early in the breeding cycle using marker data
- Early generation and accurate selection saves time and resources, increase selection intensity, thus, finally enhance genetic gain

Genetic gain:

$$R = \frac{ir\sigma_A}{y}$$

R = Genetic gain over time
i = Selection intensity
r = Selection accuracy
 σ_A = Additive genetic variances
y = Years per cycle



Lush (1937)

Genomic selection models

Ridge Regression BLUP (RR-BLUP): All markers have an equal contribution to the genetic variance

Support Vector Regression (SVR): Maps markers into a higher-dimensional feature space using kernel functions to capture non-linear relationships between markers and phenotypes while minimizing prediction

Least Absolute Shrinkage and Selection Operator (LASSO): Applies an L1 penalty to shrink many marker effects toward zero, performing variable selection by retaining only the most influential markers for prediction

Random Forest (RF): An average of multiple decision trees are used to determine predicted values

Chrom	Position	1	10	100	101
Chr1	307185	4	3	4	3
Chr1	307192	0	1	1	2
Chr1	511160	2	3	1	3
Chr1	511216	1	3	3	3
Chr1	723238	4	4	4	4
Chr1	723376	3	4	4	4
Chr1	743298	3	3	3	3
Chr1	1057549	3	3	3	3
Chr1	1056469	1	2	2	2
Chr1	1056490	1	2	2	2



Genotype

Phenotype

Cross validation(80/20)

ZhongmuNo.	1:Chr1_307185	Chr1	307185.0
4	3	2	3
4	4	3	3
2	4	3	3
4	3	4	2
4	3	4	4
4	4	3	4
4	4	3	4
2	4	4	2
2	3	4	2

$$y = \mu + Zu + e$$



Fig. Univariate genomic selection model

y = trait of interest, μ = intercept, Z = marker matrix, u = marker effect

New Mexico Genomic Study (NMGS) Population

- Phenotyping: Forage yield and 15 forage quality traits were measured by NIRS from the 3rd regrowth cycle in April 2018, 2019 and 2020 on **215** Half-sib family; trait Best Linear Unbiased Estimates (BLUEs) were used for analysis.
- Genotyping: Genotyping-by-sequencing derived and filtered 6,267 SNPs makers were used, using the Zhongmu No. 1 reference genome for SNP discovery.
- Support Vector Regression (SVR), Random Forest (RF), LASSO were tested as a predictive model.

Data Courtesy: Harpreet Kaur and Ian Ray

Predictive ability using genomic prediction models

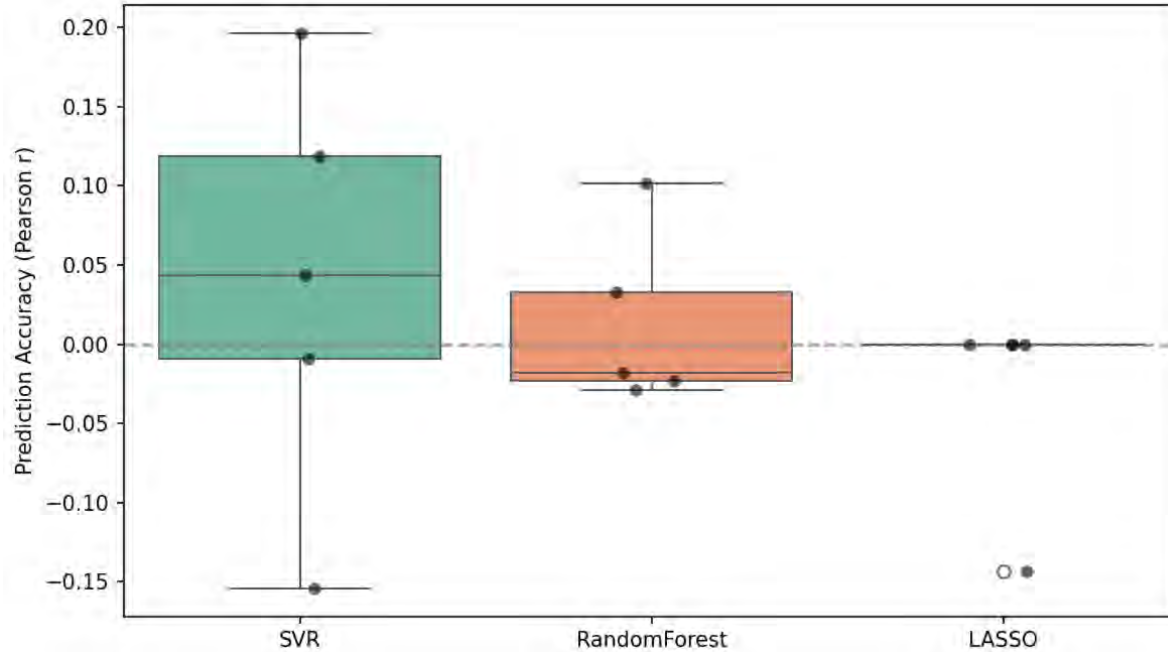
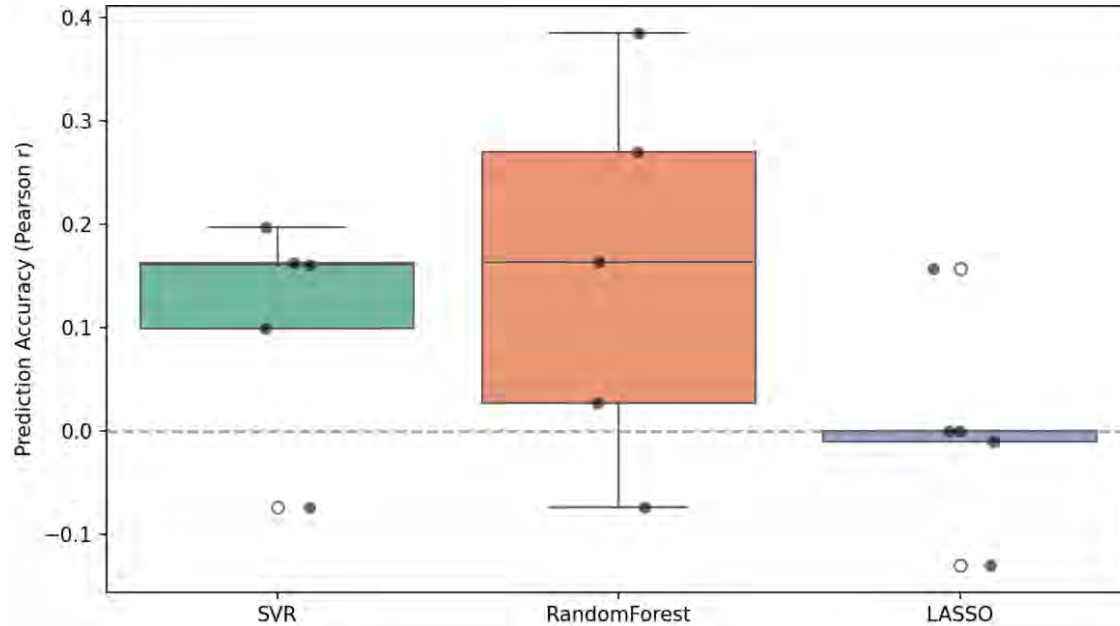


Figure. Forage biomass yield 2018 data on NMGS Population

Predictive ability using genomic prediction models



- Try a few more models to capture complex patterns
- Fully Connected Neural Network (FCNN), Deep Neural Network
- Transformer encoder
- Call SNP using Alfalfa Pangenome

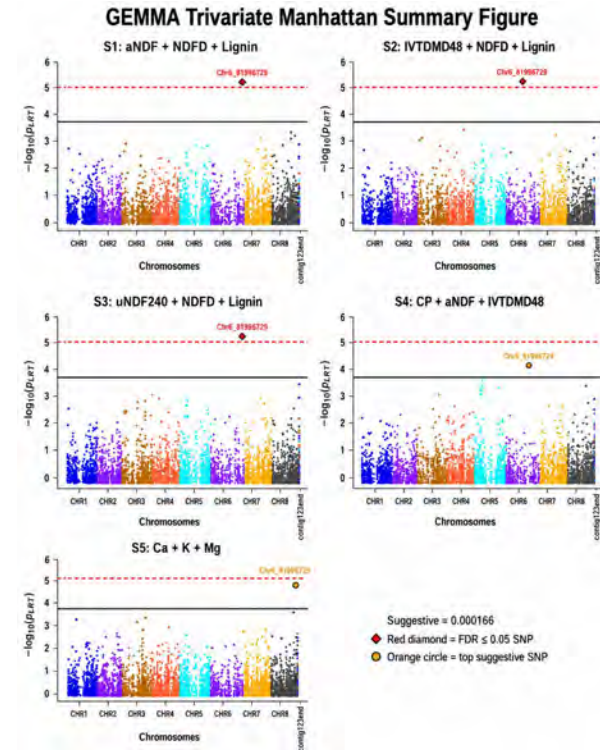
Figure. Forage nutritional trait crude protein (CP) 2018 data on NMGS Population

Augmenting Association Study on Alfalfa Nutritional Traits

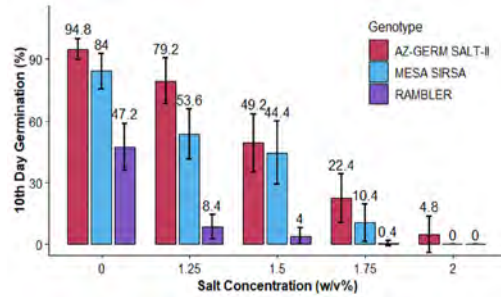


Amna Sajjad, Poster 15

- Multivariate GWAS: GEMMA (mvLMM) 2 evaluated biologically correlated bivariate and trivariate trait combinations using pLRT & FDR correction.
- Machine learning: Random Forest, LASSO, and XGBoost 3 were used to rank informative SNPs & compare signals across biomass yield and forage-quality traits.
- Few loci identified, Chr6_81996729 passed FDR correction in five trivariate and bivariate GEMMA models

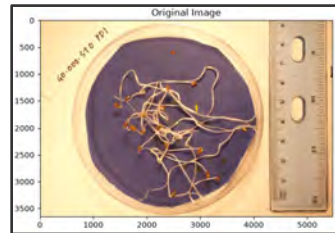


Optimizing Salinity Screening Protocol in Alfalfa and Digitizing Root Phenotyping

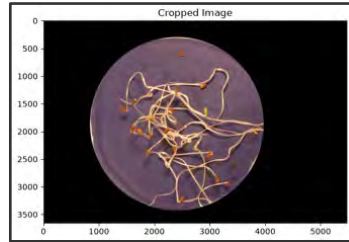


Biswajit Das, Poster 6

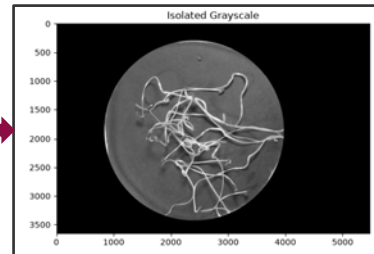
Figure. Germination at 10 Day with different salinity concentration



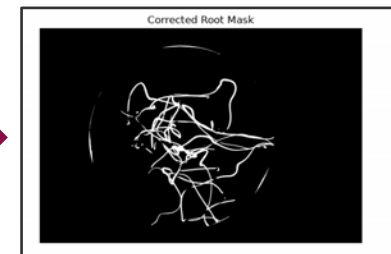
Original image



Cropped image



Isolated gray scale



Corrected root mask to extract pixel features

Figure. Root Segmentation and feature extraction

Phenomic Prediction to Forage Yield Prediction in Alfalfa



Figure : A. NMSU Alfalfa Breeding drought study with water smart alfalfa germplasm, B. UAV preparation for capturing reflectance from the water smart alfalfa germplasm; UAV is DJI Matrice 210 mounted with Micasense altum sensor



Figure: Shape polygon on top of generated orthomosaic

- Calculated indices are: GEMMA (mvLMM) 2 evaluated biologically correlated bivariate and trivariate trait combinations using pLRT & FDR correction.
- Machine learning: Random Forest, LASSO, and XGBoost 3 were used to rank informative SNPs & compare signals across biomass yield and forage-quality traits.

Atikul Hoque, Orthomosaic generation, Indices calculations

Indices used

Indices were used NGDRI, NDWI, NDVI, NDRE, GLI, WSI; Flight April 27, Harvesting April 29, 2026

Normalized Green-Red Difference Index (NGRDI):

$$NGRDI = \frac{Green - Red}{Green + Red}$$

Normalized Difference Water Index (NDWI):

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

Normalized Difference Vegetation Index (NDVI):

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

Normalized Difference Red Edge Index (NDRE):

$$NDRE = \frac{NIR - RedEdge}{NIR + RedEdge}$$

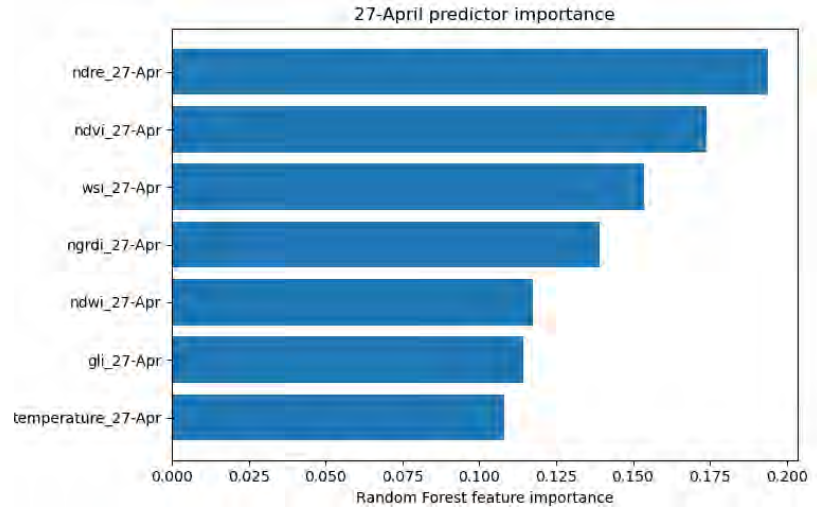
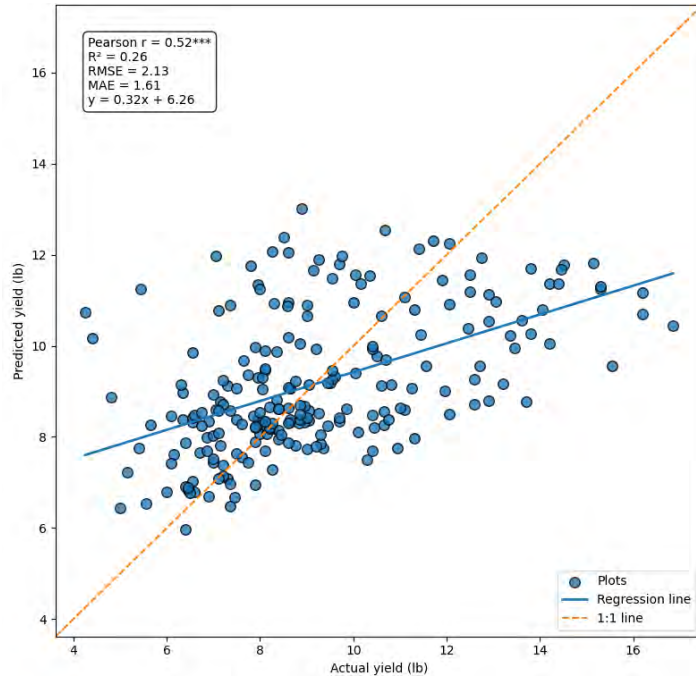
Green Leaf Index (GLI):

$$GLI = \frac{2G - R - B}{2G + R + B}$$

Water Stress Index (WSI):

$$WSI = 5 \times NDVI - NDRE$$

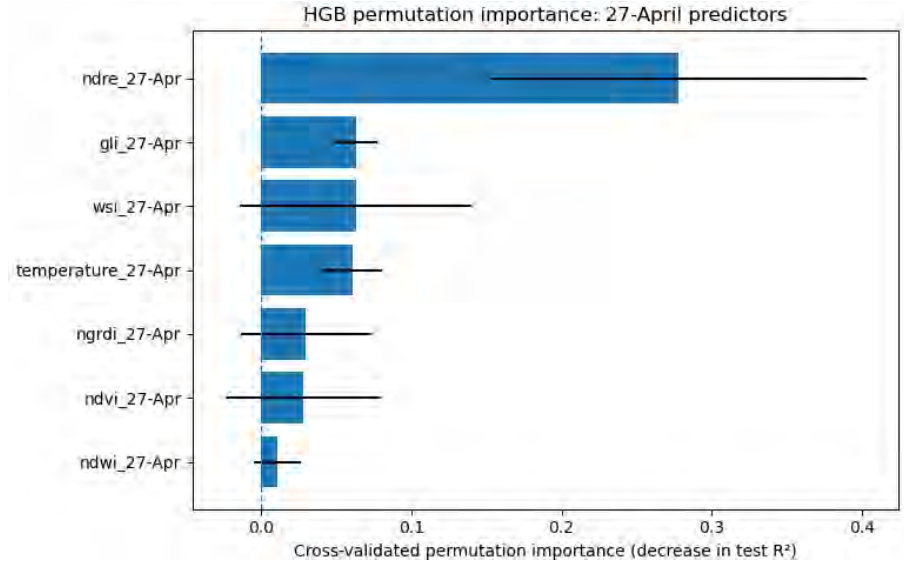
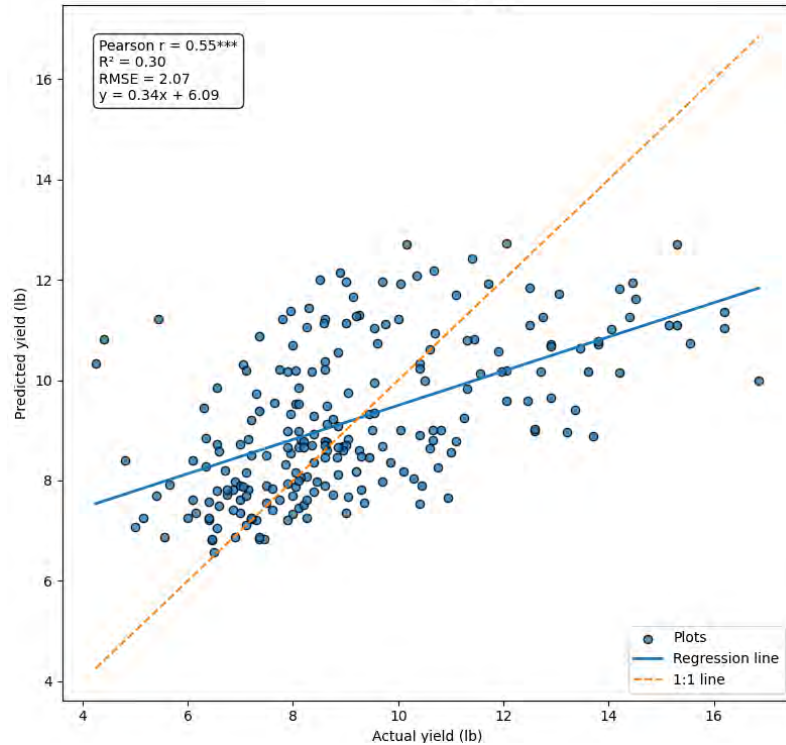
Phenomics Random Forest (RF)



Relative contribution of indices in the RF

Prediction accuracy with Random Forest (Random Forest)

Prediction with XGradient Boosting

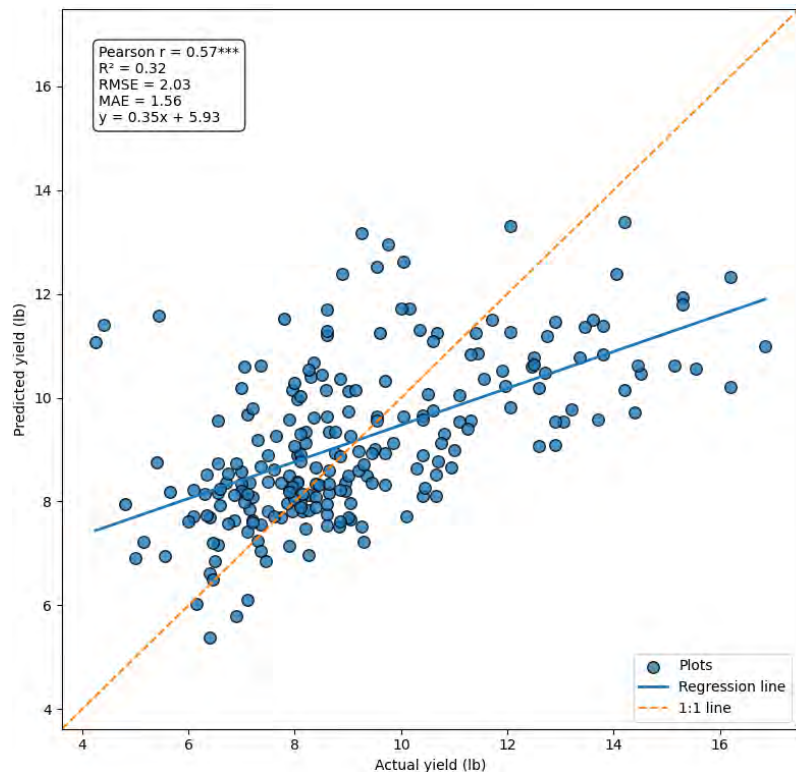


Relative contribution of indices in XGBoosting

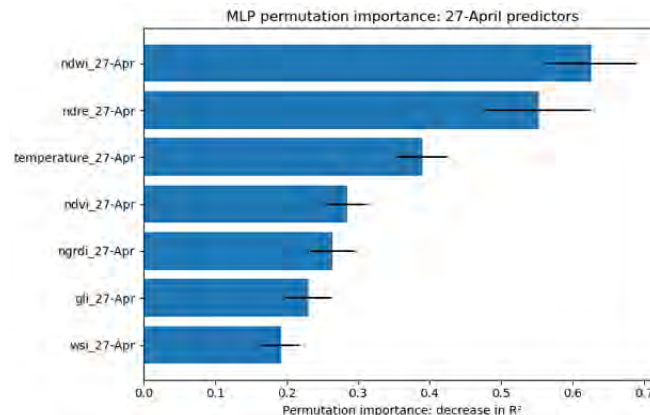
Prediction accuracy with XGBoosting

Sequentially builds decision trees that correct the errors of previous trees

Prediction with Deep Learning Model Multilayer Perceptron



Prediction accuracy with Multilayer Perceptron



Relative contribution of indices in Multilayer Perceptron

Multilayer Perceptron Parameters

Three hidden layers: 64, 32, 16 (Compression > Noise Reduction > Generalization)

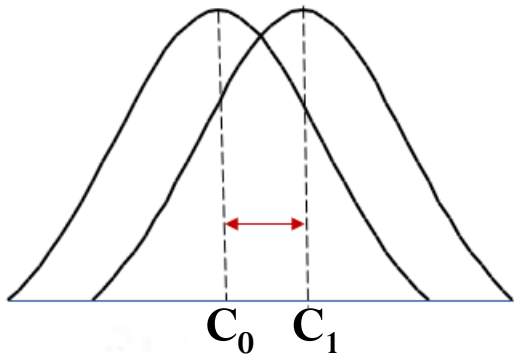
ReLU activation (Computation, non-linearity)

Adam optimizer

Five-fold cross-validation

Harnessing Genetic Gain in Crop Breeding

Genetic gain:



$$R = \frac{ir\sigma_A}{y}$$

R = Genetic gain over time

i = Selection intensity

r = Selection accuracy

σ_A = Additive genetic variances

y = Years per cycle

Potential ML application:

- Germplasm characterization
- Parental selection
- Resource allocation
- Skip early-stage trial and advance to second stage trial
- Discard less performed genotypes
- Evaluating promising genotypes
- Increase selection accuracy
- Increase selection intensity
- Reduce breeding cycle time

More to explore and use but these are tools to support mainstream breeding....

NMSU Forage Team:

Md. Abdullah Al Bari

Biswajit Das, PhD Student

Amna Sajjad, MS Student

Hector Lion, Undergraduate

Atikul Haque, Data Scientists

Collaborators and Contributors:

Ian Ray, NMSU

Harpreet Kaur, University of Minnesota

Jahid Chowdhury Choton, Kansas State University

Cesar Medina, University of Minnesota

Christopher Pierce, NMSU

Michaela Buenemann, Geography, NMSU

Clint Loest, Animal Nutrition, NMSU

Don Miller, Mountain View Seeds

Funding:

NMSU AI Educational & Training Grants

Agricultural Experiment Station, NMSU

Center of Excellence in Sustainable Food and Agricultural Systems, NMSU



New Mexico State University
Plant and Environmental Sciences

Continue the legacy of Alfalfa



Courtesy: Extension Forage Program, NMSU, KSU Animal Science