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

Mohammad Abdullah Bari, New Mexico State University
Harpreet Kaur, University of Minnesota
Amna Sajjad, New Mexico State University
Biswajit Das, New Mexico State University
Jahid Chowdhur Choton, Kansas State University
Cesar Medina, University of Minnesota
Christopher Pierce, New Mexico State University
Ian Ray, New Mexico State University

National yield averages for alfalfa (*Medicago sativa* L.) have remained stagnant for the past 40 years. The New Mexico State University (NMSU) Alfalfa Breeding team aims to accelerate genetic gain and optimize breeding strategies to develop drought resilient varieties for the water-stressed Southwest USA. With these goals in mind, a New Mexico Genomic Study (NMGS) population of 215 parental genotypes was GBS-profiled with 13K curated and filtered markers, and their maternal half-sib families were assessed for forage yield over three years (2018-2020) across multiple cuts, and forage nutritional traits from a single cut in 2018 and 2019, under deficit-irrigation management at Las Cruces, NM. Conventional genomic prediction models provided low to moderate (-0.04 to 0.26) prediction accuracies across traits and models. Our objective is to leverage modern machine learning approaches including regularized linear models (e.g., LASSO, Elastic Net), kernel-based methods (e.g., support vector regression), ensemble learning models (e.g., Random Forest and XGBoost), and deep learning architectures (e.g., fully connected neural networks and attention-based models), to improve prediction accuracy by capturing complex, nonlinear relationships between DNA markers and phenotypic traits. We will also investigate association mapping to identify loci regulating forage yield and forage nutritional traits. Based on phenotypic superiority, 15 populations were selected from the NMGS base population. These entries are being evaluated for forage yield, persistence, and drought resilience under normal irrigation (NI), and early irrigation termination (ET) management field conditions at Las Cruces, NM, during 2023 to 2026. Several germplasms have performed well and persisted with only four months of supplemental irrigation (March to June) provided annually in each of three years. Under NI, mean yield was 7.66 tons/acre across six cuts, compared to 4.73 tons/acre across three cuts under ET, over 2023-2025. Our next objective then is to leverage phenomics to understand drought tolerance mechanisms and improve screening germplasm. The refined Artificial Intelligence and Machine Learning predictive analytics with improved prediction accuracy will allow us to screen more genotypes in less time, reduce the breeding cycle, and thus assist in increasing genetic gain. Predictive analytics has the potential to redesign alfalfa breeding to benefit yield, nutritional value, and abiotic stress resilience.

Keywords: Alfalfa, Association Studies, Genetic Gain, Machine Learning, Predictive Analytics, Water Stress Tolerance, Germplasm Selection