ScienceDirect

Computers and Electronics in Agriculture
Volume 211, August 2023, 107966

A comprehensive investigation of Genotype-Environment interaction effects on seed cotton yield contributing traits in *Gossypium hirsutum L.* Using multivariate analysis and artificial neural network

Amol E. Patil [a], D.B. Deosarkar [a], Narendra Khatri [b], Ankush B. Ubale [c]

Show more

https://doi.org/10.1016/j.compag.2023.107966

Highlights

- ANN model outperforms MLR model in predicting seed cotton yield with an RMSE of 6.63 g/plant and an R^2 of 0.888.
- Number of bolls per plant has the most significant impact on yield estimates, while 100 seed weight has the least impact.
- Sensitivity analysis reveals the comprehensive effects of genotype-environment interactions on seed cotton yield.
- The ANN model is less influenced by environmental factors than the MLR model, as indicated by R^2 values.
- The study contributes to the development of effective cotton breeding strategies.

Abstract

The prediction of seed cotton yield is a critical aspect of cotton breeding. In the present study, an artificial neural network (ANN) and a multiple linear regression (MLR) model were used to predict seed cotton yield based on experimental data obtained from quantitative traits measured under different environmental conditions, including number of bolls per plant (NB), boll weight (BW), 100 seed weight (SI), number of sympodia per plant (NS), lint index (LI), internode length (IL), and seed cotton yield per plant (SCY). The experimental data underwent ANOVA and correlation analysis across different environments. The selected features were utilized for ANN and MLR modeling. The results demonstrated that the ANN model provided precise predictions of SCY, with a root mean square error (RMSE) of 6.63 g/plant and a determination coefficient (R^2) of 0.888, which outperformed the MLR model, which showed an RMSE of 8.613 g/plant and an R^2 of 0.816. Sensitivity analysis revealed that the number of bolls per plant had the most significant impact on yield estimates, while 100 seed weight had the least impact, as determined by both ANN and MLR models. Furthermore, the ANN model was less influenced by environmental factors than the MLR model, as indicated by R^2 values. Overall, this study provides a comprehensive analysis of genotype-environment interaction effects on seed cotton yield and contributes to the development of effective cotton breeding strategies.

Introduction

Cotton (*Gossypium hirsutum L.*) is one of the most important fibre crops which is mostly used as fibre, animal feed and oil purpose. Amongst the major cultivated cotton species, hirsutum cotton acquires prime position in terms of area, production and productivity due to its better fibre quality and high yield. However, the productivity of cotton is declining day by day due to numerous factors viz. pests like aphids, thrips, whiteflies, bollworms and heavy or low rainfall, etc. The per hectare yield in India in the year 2021–22 stands at 469 kg/ha as against 491 kg/ha during the year 2020–21 (India, 2022). Hence to increase yield is the biggest challenge in the current breeding objectives. Seed cotton yield is a polygenic trait and affected by the environment. Nonetheless, a critical genotype-by-environment (G×E) component exists that diminishes the association between genotype and phenotype, thereby impeding genetic progress in breeding programs and influencing the expression of quantitative traits in crops (Patil, 2018, Zeng et al., 2014). Furthermore, the G×E interaction is indicative of the type of cultivars used, including pure lines, single or double-cross hybrids, top crosses, or other variants tailored to a specific breeder. Typically, lower G×E interaction is preferred due to greater adaptability and stability in breeding programs (Killi and Harem, 2006, Ligarreto-Moreno and Pimentel-Ladino, 2022). Indeed, the cotton crop is a very sensitive to the change in the weather conditions, results in higher magnitude of G×E interaction (Campbell and Jones, 2005). However, the direct selection of yield from a singular environment may not provide accurate information because environmental factors influence estimates of combining ability (Dwivedi and Dwivedi, 2020). The environment has a significant influence on the quantitative nature of yield. Consequently, response to the direct selection for yield may result in minimal output. Therefore, the genetic modification of such difficult-to-exploit traits could be enhanced by indirect selection utilizing highly correlated traits with high heritability, simple selection, and evaluation (Liu et al., 2022).

Many researchers explored the association between the dependent variable i.e., seed cotton yield with independent variables viz. boll number plant^{-1}, boll weight, number of sympodia plant^{-1}, plant height etc. by employing multivariate analytical methods such as correlation, path analysis, multiple linear regressions, and other numerous statistical methods. Researchers applied correlation and regression studies to find the degree and direction of association between different plant variables with seed cotton yield (Awais et al., 2021, Salahuddin et al., 2010). They observed positive correlation between seed cotton yield with boll weight, number of sympodia plant^{-1} and boll number. In the study of Shastry et al. (2017), they used multiple linear regression (MLR) to decipher the relationship among seed cotton yield and total dry weight, nitrogen uptake, number of bolls per plant, leaf area index etc. (Corrales et al., 2022, Shastry et al., 2017).

All of the above-mentioned traits studied by the various researchers are quantitative in nature and hence their expression is altered if evaluated in multiple environments. In such situations the linear and nonlinear relationships between the plant traits with yield cannot be established correctly by applying regression-based models and it could lead to the misinterpretation while assigning weights to the traits contributing to yield. However, the major limitation of regression-based models is that they are only applicable when the relationship between the response and predictor variable is linear. In such situations the application of artificial intelligence algorithms like ANN could be advantageous over the conventional models as it can overcome the limitations of normality, linearity, trait independence etc. Many researchers are nowadays applying ANN to reveal complex problems in crop improvement.

Aboukarima et al. (2015) applied ANN for prediction of cotton leaf area to determine crop growth and productivity (Aboukarima et al., 2015). Golhani et al. (2018) carried out disease detection in cotton through ANN image processing technique (Golhani et al., 2018). Brasileiro et al. (2015) applied ANN for the selection of sugarcane families (Brasileiro et al., 2015). Anna et al. (2018) verified superiority of ANN for genetic classification (Chlingaryan et al., 2018). Zaefizadeh et al. (2011) highlighted superiority of ANN over multiple linear regressions in predicting the yield in Hulles barley (Zaefizadeh et al., 2011). The seed yield in sesame is predicted by Emamgholozadeh et al. (2015) by using ANN and MLR (Emamgholizadeh et al., 2015). Safa et al. (2015) also predicted yield by using ANN (Safa et al., 2015). Other researchers also employed ANN for indirect selection and plant identification in lettuce and tea genotypes (Azevedo et al., 2015, Li and He, 2008). Nascimento et al. (2013) evaluated adaptability and stability in alfalfa using ANN (Nascimento et al., 2013). Peixoto et al. (2015) used ANN for genetic value prediction (Peixoto et al., 2015). Silva et al. (2014) developed ANN models for the estimation of the breeding values along with the associated genetic gains (Silva et al., 2014). Yildirim et al. (2022) has applied ANN for cotton yield prediction in short range, whereas the regions are data scarce (Yildirim et al., 2022). The ease of utilizing readily available input data for training an ANN model to forecast cotton yield several months in advance has garnered attention. In another study, Silva et al. (2017), explored use of ANN for genotypic selection in cotton for fiber improvement (Silva Júnior et al., 2017). They reported that the ANNs are recognizing the pattern effectively and classifying the cotton genotypes. It has also been reported that ANN-based predictions are utilised for the selection of fibre quality (FQ) in cotton genotypes.

The existing literature suggests that the decline in cotton productivity is a major challenge for breeders. Seed cotton yield, a polygenic trait affected by the environment, is a crucial factor for the genetic improvement of cotton. However, direct selection for yield in a single environment may not provide accurate information due to environmental effects on combining ability estimates. Therefore, genetic improvement could be observed through the indirect selection of the highly correlated traits with heritability. Regression-based models have limitations in accurately establishing linear and nonlinear relationships between quantitative traits and yield, especially when evaluated in multiple environments. Indeed, ANNs have been applied for multiple crop improvement studies, including cotton, to overcome the limitations of traditional models. ANN has shown potential for indirect selection, plant identification, adaptability and stability evaluation, genetic value prediction, and yield prediction in data-scarce regions. However, to the best of our knowledge, no study has yet investigated the potential of ANN for predicting cotton yield while taking into account the effect of different environmental factors.

The objectives of this investigation are: (i) To investigate the effect of genotype-environment interaction on seed cotton yield and determine the most stable genotypes across different locations in Maharashtra state. (ii) To determine the association between different plant variables and seed cotton yield using correlation and regression analyses. (iii) To identify the most important plant variables contributing to seed cotton yield and evaluate their performance in predicting yield using ANN and MLR models. (iv) To compare the efficacy of the developed ANN and MLR models using selected plant variables to predict seed cotton yield. (v) To provide recommendations for cotton growers in Maharashtra state regarding the selection of genotypes and plant traits to maximize seed cotton yield.

Section snippets

Multienvironment field evaluation

Nonetheless, in this study, the selection of parents was based on their representation of diversity with respect to fiber properties, sucking and drought tolerance, earliness, and other relevant traits.

Based on the characteristics of the parents various crossing design combinations were formulated. There were nine testers, and seven lines were deployed, presented in Table 1.

Although the primary goal was to select diverse parents, the choice of testers and lines was also influenced by their…

Results and discussion

This investigation was conducted across three different sites located in the Maharashtra state of India, where the data collected on various plant variables was used to develop regression models based on Eberhart and Russell, correlation, ANN, and MLR analyses. The combined analysis of variance results derived from the Eberhart and Russell model indicated that a significant genotype-environment interaction was present for the traits under investigation, as outlined in Table 4. Notably, the…

Conclusions

Cotton breeding heavily relies on accurate prediction of seed cotton yield. This study deployed ANN and MLR models for the prediction of the seed cotton yield using data on quantitative traits obtained under different environmental conditions, including number of bolls per plant (NB), boll weight (BW), 100 seed weight (SI), number of sympodia per plant (NS), lint index (LI), internode length (IL), and seed cotton yield per plant (SCY). The ANN model significantly outperformed in comparison with …

CRediT authorship contribution statement

Amol E. Patil: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – original draft. **D.B. Deosarkar:** Supervision, Writing – original draft. **Narendra Khatri:** Software, Writing – review & editing, Visualization. **Ankush B. Ubale:** Methodology, Formal analysis.…

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.…

Special issue articles Recommended articles

References (53)

M. Abedinpour *et al.*

Performance evaluation of AquaCrop model for maize crop in a semi-arid environment

Agric. Water Manag. (2012)

A.N. Amenaghawon *et al.*
Biotechnological conversion of yam peels for enhanced citric acid production: Data-driven machine learning modeling and global sensitivity analysis of the impact of metabolic stimulants
Ind. Crops Prod. (2023)

T. Ayoub Shaikh *et al.*
Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming
Comput. Electron. Agric. (2022)

D. Bastani *et al.*
Prediction of CO2 loading capacity of chemical absorbents using a multi-layer perceptron neural network
Fluid Phase Equilib. (2013)

M.H. Baziar *et al.*
Evaluation of lateral spreading using artificial neural networks
Soil Dyn. Earthq. Eng. (2005)

A. Chlingaryan *et al.*
Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review
Comput. Electron. Agric. (2018)

D.C. Corrales *et al.*
A surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern France
Comput. Electron. Agric. (2022)

A. Dubois *et al.*
Short term soil moisture forecasts for potato crop farming: A machine learning approach
Comput. Electron. Agric. (2021)

S. Emamgholizadeh *et al.*
Seed yield prediction of sesame using artificial neural network
Eur. J. Agron. (2015)

K. Golhani *et al.*
A review of neural networks in plant disease detection using hyperspectral data
Inf. Process. Agric. (2018)

View more references

Cited by (1)

Elucidating the phenotypic basis of multi-environment stability for fiber yield and quality traits of cotton (Gossypium hirsutum L.) using 498 recombinant inbred lines

2024, Industrial Crops and Products

Show abstract

View full text