EJONS

International Journal on Mathematic, Engineering and Natural Sciences

(Ulusalararası Fen,Mühendislik ve Doğa Bilimleri Dergisi)

https://ejons.org/index.php/ejons

e-ISSN: 2602 - 4136

Review Article

Doi: https://doi.org/10.5281/zenodo.14253740

Bitki Islahında Modelleme Stratejileri

Murat TUNC[1], Süreyya Betül RUFAIOGLU[2],*

[1] Harran University, Agriculture Faculty, Department of Field Crops, Sanliurfa, Türkiye.
[2] Harran University, Agriculture Faculty, Department of Soil Science and Plant Nutrition, Sanliurfa, Türkiye.

*Corresponding Author e-mail: sureyyarufaioglu@harran.edu.tr

Makale Tarihçesi

Geliş: 25.10.2024
Kabul: 21.11.2024

Anahtar Kelimeler

Bitki Islahı,
Modelleme,
Makine öğrenmesi,
Yapay sinir ağları,
Algoritmalar

Öz: Bitki ıslahı, tarımsal verimliliği ve mahsul kalitesini artırmak için daha hızlı ve daha doğru tahminler sağlayan modelleme stratejilerinden yararlanmak üzere geleneksel yöntemlerin ötesine geçmektedir. Geleneksel yöntemler uzun vadeli tarla denemelerine dayanırken, yeni ortaya çıkan dijital tarım uygulamaları ve modelleme stratejileri genomik seçim, fenotip tahmini ve çevre-genotip etkileşim analizi gibi alanlarda verimliliği artırmak için önem kazanmıştır.Makine öğrenimi algoritmaları, genetik ve çevresel verileri değerlendirerek bitki ıslahında karmaşık tarımsal sistemlerin daha iyi anlaşılmasını sağlar. Destek vektör makineleri (DVM), rastgele ormanlar (RF) ve yapay sinir ağları (YSA) gibi algoritmalar, fenotipik özelliklerin ve genotip-çevre etkileşimlerinin tahmininde yaygın olarak kullanılmaktadır. Bu yöntemler, genetik potansiyelin en iyi şekilde kullanılmasını ve çevresel değişkenliğe adaptasyonu sağlayarak sürdürülebilir tarımsal üretime katkıda bulunmaktadır. Bu derleme, bitki ıslahında çevresel stres faktörlerinin modellenmesinde ve genotiplerin bu koşullara tepkisinin tahmin edilmesinde derin öğrenme ve diğer yapay zeka tabanlı tekniklerin önemini vurgulamaktadır. Modelleme stratejileri, bitki ıslahı süreçlerini daha verimli hale getirme ve gıda güvenliğine katkıda bulunma potansiyeline sahiptir. Bu bağlamda, gelecekte bitki ıslahında modelleme stratejilerinin etkin bir şekilde kullanılması, tarımsal biyoteknoloji alanında önemli ilerlemelere yol açacaktır.

Atıf Künyesi: Tunç, M. ve Rufaıoğlu S.B. (2024). Bitki Islahında Modelleme Stratejileri, EJONS International Journal on Mathematic, Engineering and Natural 8 (4) 464-474. **How to Cite:** Tunç, M. ve Rufaıoğlu S.B. (2024). Modelling Strategies in Plant Breeding, EJONS International Journal on Mathematic, Engineering and Natural 8 (4) 464-474

Modelling Strategies in Plant Breeding

Article Info

Received: 25.10.2024
Accepted: 21.11.2024

Abstract: Plant breeding is moving beyond traditional methods to leverage modelling strategies that provide faster and more accurate predictions to improve agricultural productivity and crop quality. While traditional methods rely on long-term field trials, emerging digital agriculture applications and modelling strategies have gained importance to improve efficiency in areas such as genomic selection, phenotype prediction and environment-genotype interaction analysis. Machine learning algorithms enable a better understanding of complex agricultural systems in plant breeding by evaluating genetic and environmental data. Algorithms such as support vector machines (SVM), random forests (RF) and artificial neural networks (ANN) are

Keywords

Plant Breeding,
Modelling,
Machine Learning,
Artificial neural network,
Algorithms

widely used in the prediction of phenotypic traits and genotype-environment interactions. These methods contribute to sustainable agricultural production by enabling the optimal utilization of genetic potential and adaptation to environmental variability. This review highlights the importance of deep learning and other artificial intelligence-based techniques in modelling environmental stressors in plant breeding and predicting the response of genotypes to these conditions. Modelling strategies have the potential to make plant breeding processes more efficient and contribute to food security. In this context, the effective use of modelling strategies in plant breeding in the future will lead to significant advances in the field of agricultural biotechnology. .

1.Introduction

Plant breeding is crucial for improving agricultural efficiency and crop quality in the face of a growing world population, climate change and the need to sustainably manage limited resources in agricultural production. Traditional plant breeding methods are based on the evaluation of phenotypic traits of the crops through long-term field experiments. This process requires a lot of time and resources and therefore the need for new approaches is becoming increasingly important (Xu, 2010). In the last years, emerging modelling strategies and digital agriculture practices have started to play an important role in plant breeding studies to increase efficiency and speed up decision-making processes (Hickey et al., 2017).

In plant breeding studies, modelling techniques are especially common, particularly in the areas such as genomic selection, phenotype prediction, and environment-genotype interaction analysis (Crossa et al., 2017). The use of genomic data enables a better understanding of complex agricultural systems and more accurate predictions in plant breeding research. For this purpose, machine learning-based algorithms have become a vital instrument, especially for the evaluation of complicated environmental and genetic data (González-Camacho et al., 2018). The application of these methods, combined with precision agriculture technologies and big data analysis in plant breeding, offers significant opportunities to enhance productivity (Bramley & Ouzman, 2019).

Genomic selection (GS) is a method that allows plant breeders to predict traits using a large number of molecular markers to identify target phenotypes. This method offers significant time and cost savings by reducing the need to evaluate a large number of candidates in breedings programs (Meuwissen et al., 2001). Furthermore, machine learning techniques have been extensively used to improve the efficiency of genomic selection (Heslot et al., 2012). Machine learning algorithms such as support vector machines (SVM), random forests (RF), and artificial neural networks (ANN) have been widely used in predicting phenotypic traits and evaluating genotype-environment interactions (Morota & Gianola, 2014).

In plant breeding, the analysis of environment-genotype interactions is crucial for predicting crop performance, particularly under different environmental conditions. In this regard, modelling approaches such as multiple regression models, principal component analysis (PCA) and AMMI (Additive Main Effects and Multiplicative Interaction) analysis are frequently used in plant breeding studies (Piepho, 2000; Gauch, 2013). Recent advancements in artificial intelligence and machine learning techniques have significantly improved the analysis of these interactions, enabling faster and more effective outcomes in breeding programs (Spindel et al., 2015). For instance, deep learning approaches have emerged as powerful tools for uncovering complex relationships between genetic and environmental data (Montesinos-López et al., 2021).

Modelling strategies also contribute to ensuring sustainability in plant breeding. These strategies have the potential to enhance agricultural productivity by addressing environmental conditions and genetic factors together (Jannink et al., 2010). Specifically in the framework of climate change, genetic breeding studies are becoming more significant in order to increase the resilience of plants against environmental stresses (Fisher et al., 2015). In this scope, models that model environmental stress factors and predict how plants respond to these conditions will play an important role in the future of plant breeding.

As a conclusion, the use of modelling strategies in plant breeding goes beyond traditional methods and offers an approach that is faster, more accurate and can effectively use multi-dimensional input data. These strategies have great potential to provide sustainability of agricultural production, to adapt to environmental variability and to make it possible to make the best use of genetic potential (Jannink et al., 2010).

2. Historical Development of Modelling Strategies

Plant breeding is a science based on various strategies that have been developed throughout history to increase agricultural productivity and ensure food security. Modelling strategies in plant breeding, especially in the last century, have played an important role in modern agricultural practices and brought many innovations. The first modelling strategies in plant breeding date back to humanity's introduction to agriculture. Traditional plant breeding started with farmers selecting the best-performing plants to produce crops. This selection method was based purely on observational data and did not involve any numerical modelling strategies. However, it is known that during this period, people carefully monitored various factors in order to improve plant traits.

The work by Gregor Mendel on pea plants founded the genetic principles in plant breeding, allowing the development of scientific bases for modelling approaches. Mendel's laws demonstrated that mathematical models are essential for understanding the inheritance of genetic diversity. This later led to advanced prediction and selection models being developed in the field of plant breeding. In the course of the 20th century, quantitative genetics was developed fully, and much of the fundamental theoretical framework that enabled the use of mathematical approaches in modelling complex inherited traits got its place in plant breeding—mainly by the work of Fisher and Wright. Fisher's idea of 'Genetic Analysis of Variance' made it possible to apply statistical modelling methods in plant breeding and helped breeders in estimating various environmental and genetic effects simultaneously.

By the middle of the 20th century, advances in computer technologies enabled a more rapid and advanced modelling approaches in the plant breeding context. In this regard, computer-assisted modelling in forecasting genetic algorithms and selection patterns across the quadruplicate came into being. By around the same period, computer-assisted simulations became a very crucial aspect in analysing genetic information and creating new varieties.

Starting from the 2000s, GS has revolutionized the art of plant breeding altogether. Genomic selection is based on the use of abundant genetic markers for predicting an organism's phenotypic performance. This process has been advanced substantially by integrating artificial intelligence along with machine learning techniques in agriculture.

Of these machine-learning techniques, applications in the prediction of complex phenotypic traits have especially been done using artificial neural networks (ANNs), support-vector machines (SVMs), and random forests. These modern approaches allow for faster and more accurate results when compared to traditional breeding methods. Big data has become very important in plant breeding nowadays. New modelling methods bring together and mix phenotypic, genotypic, and environmental data to make better predictions. Combined models such as these reduce risks in plant breeding and help to breed more efficiently productive crops. Cloud computing technologies for big data analysis have lowered the time used in analysing complicated datasets within plant breeding and provided breeders with much more flexibility. This technology is used for such purposes as developing predictive models and optimizing crop quality, among others, in plant breeding.

Plant breeding strategies are changing quickly. Applications of artificial intelligence and machine learning will be very important for the solution of increasingly complex farming problems of the future. But issues of quality in data used in such models and uncertainty surrounding environmental factors in modelling remain huge unsolved issues. The historical development of modelling strategies in plant breeding ranges from old observational approaches to up-to-date AI-based methods. Knowledge of genetic information, aided by computer-aided analysis and integration of big data, has helped produce more productive and resistant varieties, usually in agriculture. This area is only going to see more substantial impacts of AI and data analytics in the future.

3.Basic Modelling Techniques Used in Plant Breeding

Plant breeding uses modern technologies and scientific methods to increase genetic potential and improve productivity, disease resistance and environmental adaptation of plants. Basic modelling techniques are an integral part of this process, enabling genetic engineers and researchers to better understand plant traits and make effective selection.

3.1. Genomic Selection (GS): Genomic selection utilizes statistical models to predict phenotypic traits using genotypic data of individuals. This technique is particularly used in large data sets and in the selection of complex traits. Genomic selection is an effective method to accelerate genetic gain and shorten the breeding process by analyzing comprehensive genomic and phenotype data (Meuwissen et al., 2001).

3.2. Marker Assisted Selection (MAS): Marker-assisted selection is a technique that accelerates genetic selection by using the association between a specific genetic marker and one or more traits. It is particularly used in the selection of traits such as single gene disease resistance or productivity. MAS facilitates the elimination of unwanted alleles by enabling direct selection of targeted genes (Hospital, 2009).

3.3. Phenotyping and High-Resolution Imaging: High-resolution phenotyping using advanced imaging technologies and sensors enables precise and scalable assessment of morphological and physiological characteristics of plants. This technique is particularly used to understand genotype-environment interactions and to monitor plant responses under stress. High-resolution phenotyping increases the accuracy and speed of genetic improvements (Furbank and Tester, 2011).

3.4. Machine Learning and Artificial Intelligence: Machine learning techniques are used to predict plant traits by learning from large and complex data sets. Algorithms such as neural networks, support vector machines and decision trees have been effective in modelling phenotypic traits and the consequences of genetic engineering interventions. Artificial intelligence enables rapid and accurate analysis of traits in plant breeding, especially yield, disease resistance and response to environmental stresses (Crossa et al., 2017).

3.4.1. Regression Models

Regression analysis constitutes one of the primary modelling methodologies frequently employed in the domain of plant breeding. The application of simple and multiple regression analyses is deemed appropriate for investigations aimed at elucidating the impacts of genetic and environmental variables on plant yield (Montgomery and Peck, 2012). This analytical technique facilitates the formulation of predictions by establishing linear correlations between genotype and phenotype throughout the breeding process. In an investigation conducted by Smith et al. (2015), regression models were employed to forecast yield enhancements of legumes across varying planting densities. Within the context of this research, regression analysis proved to be a potent instrument for comprehending the responses of diverse genotypes to environmental conditions.

3.4.2. Decision Trees and Random Forest

Decision trees are effective methods for determining the relationships between genetic and environmental data and for selecting suitable plant varieties. Decision trees allow data to be classified by branching through simple rules (Breiman, 2001). The Random Forest algorithm is a method created by combining a large number of decision trees and provides genotype prediction with higher accuracy. In a study conducted by Martin et al. (2018) on legumes, it was aimed to identify genotypes resistant to drought conditions using the Random Forest algorithm. This technique has provided great success in understanding the responses of legumes to environmental stress factors.

3.4.3. Support Vector Machines (SVM) and Support Vector Regression (SVR)

Support Vector Machines (SVM) is one of the powerful modelling techniques used in plant breeding for classification and regression purposes. SVM is especially preferred for its high accuracy in complex data sets (Vapnik, 1998). This method allows the data to be divided into certain classes and the boundaries to be determined optimally. Jones et al. (2019) used SVM models to predict yield traits of chickpea plants under environmental stress conditions. The study showed that SVM has high accuracy in yield prediction and resilience to environmental conditions in legume species. SVR is a version of support vector machines used for regression and is widely used for yield prediction in plant breeding. This method is used to achieve high accuracy, especially in data sets, and to minimize the bias of predictions (Smola and Schölkopf, 2004). This technique allows for precise yield estimation in plant breeding processes.

3.4.5. Artificial Neural Networks (ANN)

Artificial neural networks have been used as a powerful modelling tool for predicting plant traits. Thanks to its ability to process and learn from large data sets, ANN is an effective tool for analyzing the complexity of gene-environment interactions. For example, ANN models used to predict water stress tolerance in wheat have been used to identify genotypes that will maximize yield under water stress conditions (Al-Hashimi et al., 2019). In addition, in legume breeding, artificial neural networks are effectively used to analyze genetic data under different environmental conditions. Singh et al. (2020) obtained high accuracy predictions using ANN model for yield prediction in lentil plant.

3.4.6. Genetic Algorithms

Genetic algorithms are an optimization method inspired by the principles of biological evolution and are used in plant breeding to determine the most suitable genotypes (Goldberg, 1989). This method allows increasing genetic diversity in plant breeding and selecting individuals with desired traits. Wang et al. (2017) conducted a study in which genetic algorithms were used to identify individuals with the highest yield potential among chickpea genotypes. This study showed that genetic algorithms are an effective tool for selecting suitable candidates in plant breeding.

3.4.7. Bayesian Networks

Bayesian networks are a technique used in plant breeding to model probabilistic relationships between genetic and environmental factors. This method is suitable for exploring dependencies between different variables and making probability predictions using these dependencies (Heckerman, 1997). Especially in cases where environmental conditions and genetic factors need to be evaluated together, Bayesian networks make a great contribution to the breeding process.

3.4.8. K-Nearest Neighbor (KNN)

The K-nearest neighbor algorithm (KNN) is a simple but effective modelling technique that performs classification and regression based on the proximity of data (Cover and Hart, 1967). In plant breeding, the KNN algorithm is widely used to evaluate genetic similarities and phenotypic traits. Patel et al. (2019) used KNN algorithm to determine the tolerance of chickpea varieties to various environmental conditions. The results of the study showed that KNN was successful in analyzing the responses of similar genotypes to environmental conditions.

3.4.9. Deep Learning

Deep learning consists of multilayer artificial neural networks and is highly effective at learning complex patterns and relationships from large data sets. In plant breeding, deep learning models are used to automatically extract phenotypic traits, especially from high-resolution image data, and to identify differences between genotypes (Lee et al., 2018).

3.4.10. Genetic Algorithms and Optimization Methods

Genetic algorithms and other optimization methods are used in plant breeding to select the most suitable genotypes (Goldberg, 1989). These methods optimize genetic diversity in breeding studies and ensure the selection of the most suitable individuals. Wang et al. (2017) conducted a study in which genetic algorithms were used to identify individuals with the highest yield potential among chickpea genotypes. This study showed that genetic algorithms are an effective tool for selecting suitable candidates in plant breeding.

These technologies play a critical role in solving complex problems in the plant breeding process, increasing the speed and accuracy of genetic improvements. Machine learning and artificial intelligence enable genetic engineers to make more informed decisions and allow for large-scale improvement of plant traits.

5.Simulation Models: Simulation models are used to evaluate the genetic gain and demographic effects of specific breeding strategies. These models predict the genetic structure and evolution of plant populations under various scenarios, providing critical information for the design of breeding programs. Simulation techniques also provide important insights into issues such as genetic diversity and conservation of gene flow (Wang et al., 2012).

4. Big Data and Machine Learning Applications

Big data and machine learning have revolutionized plant breeding (Bishop, 2006). In this section, the applications of big data analysis and machine learning models in plant breeding will be explained in detail. Especially in legume species, the analysis of large data sets and the evaluation of these data are carried out with machine learning techniques. Patel et al. (2021) used the Gradient Boosting method on a large data set to estimate growth and yield parameters of legumes. The results of the study revealed that this technique is an effective tool for big data analysis in plant breeding.

4.1. Model Validation and Performance Evaluation in Plant Breeding

Evaluating the effectiveness of modelling strategies is very important in plant breeding because the accuracy and reliability of the models developed directly affect the success of breeding efforts. The process of model validation and performance evaluation aims to test how well the model predicts and performs on different data sets. This process is critical to determine whether the model is over- or under-learning and its generalizability.

4.2. Model Validation and its Importance

Model validation is the process of evaluating the success of the model on different data sets. Usually, the data set is divided into two: training and test. The training set is used for the model to learn, while the test set is used to evaluate the model's success on data it has not seen before. This method is important to determine whether the model has overfitting or underfitting problems (Bishop, 2006).

4.3. Performance Evaluation Metrics

Performance evaluation metrics are used to measure the accuracy and error rate of the model. The main performance evaluation metrics used in plant breeding studies are.

4.3.1. Mean Squared Error (MSE)

MSE measures the error rate by averaging the squares of the differences between the model's predicted values and the actual values. MSE is a sensitive measure of the magnitude of error values, as it penalizes large errors in the predictions more (Chai and Draxler, 2014).

4.3.2. Root Mean Squared Error (RMSE)

RMSE is the square root of MSE and presents the magnitude of prediction errors on a more understandable scale. RMSE is widely used to assess the overall performance of the model and indicates the accuracy of the model in yield predictions (Chai and Draxler, 2014).

4.3.3. Mean Absolute Error (MAE)

MAE averages the absolute values of the differences between predicted values and actual values. This metric helps to understand the performance of the model by directly measuring the error rate of the model and is particularly useful in the absence of large errors (Willmott and Matsuura, 2005).

4.3.5. R Square (R^2) Value

R^2 refers to the proportion of variance of the model explained by the independent variables. The closer the R^2 value is to 1, the better the fit of the model. In plant breeding, it is used to evaluate the effects of genetic and environmental variables on plant yield (Montgomery and Peck, 2012).

4.3.6. Mean Absolute Percentage Error (MAPE)

MAPE gives the percentage error of the predicted values compared to the actual values. This metric expresses the error rate of the model as a percentage and enables comparisons to be made on data at different scales (Makridakis and Hibon, 2000).

5. Model Validation Methods

5.1. K-Fold Cross Validation

K-fold cross-validation divides the data set into K equal parts and one part at a time is used as the test set and the remaining K-1 parts are used as the training set. This method is used to increase the generalizability of the model and to understand how it performs in different data sets (Kohavi, 1995).

5.2. Bootstrap Method

Bootstrap allows us to create many different training and test sets by taking random samples from the data set. This method is used to assess model accuracy and improve model performance on small data sets (Efron and Tibshirani, 1993).

Lee et al. (2016) evaluated the accuracy of ANN models to model the response of legumes to environmental stress conditions and proved the effectiveness of the model using criteria such as RMSE and MAE. These metrics are critical to evaluate the performance of the model developed in plant breeding and to understand how well the model can generalize to different conditions.

6. Application Areas and Success Cases

Modelling strategies are of increasing significance in the field of plant breeding with the integration of technological innovations and scientific progress. These strategies in modern agronomy are making genetic improvement processes faster, scalable and efficient, increasing the capacity to provide solutions to various challenges. This chapter presents practical applications of genetic modelling, artificial intelligence, and advanced imaging techniques and provides successful examples of how these technologies have been integrated into real-world agricultural practices. Modelling strategies used in plant breeding have been influential in many different areas, from genetic diversity conservation to disease management. Advanced statistical methods and computer simulations allow breeders to more rapidly develop plant varieties with specific traits. These technologies also predict plant performance under environmental stress conditions, facilitating the selection of more resistant and adaptive varieties.

Concrete examples of how these methods have been successfully applied in agriculture are highlighted, illustrating the breadth of applications of modelling strategies in plant breeding and the many advantages of these strategies. The integration of innovative technologies into agricultural practices increases the speed and efficiency of genetic enhancements and contributes to the maintenance of global agricultural production.

6.1. Genomic Selection Applications: Genomic selection techniques have been used for genetic improvement of legumes, especially species such as lentils and chickpeas. This method is important to better understand genetic variation and to improve productivity traits. For example, one study used genomic selection in chickpea to develop high-yielding and disease-resistant varieties in a short time (Roorkiwal et al., 2018). In addition, especially in large-scale crops such as maize and wheat, genomic selection has accelerated selection processes and maximized genetic gain compared to traditional breeding methods. For example, in one study, selection on yield traits of maize using genomic selection accelerated the improvement process by 25% and increased yields by up to 20% (Bernardo, 2016).

6.2. Phenotyping and High-Resolution Imaging: AI-enabled high-resolution imaging techniques are used to monitor plant growth and disease and stress responses in real time under field conditions. High-resolution phenotyping techniques are used to study traits such as adaptation to stress conditions and disease resistance in legume crops. For example, high-resolution imaging techniques applied to lentil plants have been used to understand physiological responses under water stress and to identify genotypes that are best adapted to these stress conditions (Kumar et al., 2020). In another study, they reported that AI-assisted imaging in potato plants detects disease symptoms at an early stage and thus increases the effectiveness of disease management strategies (Quirós Vargas et al., 2019).

6.3. Disease Prediction with Machine Learning: Machine learning models have been used as an effective tool in legume disease management. They have been used for early detection and spread prediction of common legume diseases such as fusarium wilt and root rot. These models analyze the relationships between disease agents and environmental factors over large data sets and contribute to the development of management strategies (Srivastava et al., 2019).

6.4. Identification of Optimal Genotypes under Salt Stress: Predictive modelling approaches used to identify wheat genotypes resistant to salt stress have enabled the selection of more resistant varieties by predetermining the response of genotypes to environmental stresses. This approach contributes to global food security by increasing the feasibility of agriculture in saline soils (Zhang et al., 2021).

6.5. Modelling Legume Performance under Drought Conditions: Artificial intelligence and machine learning methods have been used to predict the resilience of legume varieties to abiotic stress conditions such as drought. Artificial neural networks and support vector machines have been used to model genetic adaptations to water stress in species such as lentils and beans and to optimize their efficient use of water resources (Gupta et al., 2017).

6.6. Simulation and Management of Disease Spread: Simulation models are used to predict the spread of legume diseases and develop effective management strategies. For example, stochastic simulation models have been used to assess the resistance of chickpea varieties to Ascochyta blight (Ascochyta rabiei). These models have helped to optimize disease control strategies by assessing disease development in combination with environmental factors (Singh et al., 2018).

6.7. Assessment of Genetic Diversity and Improvement Potential: Large-scale genomic data analysis for the assessment of genetic diversity and improvement potential is an important area of research in legumes. Multi-locus genomic selection models have been used to understand the genetic basis of complex traits such as disease resistance and yield traits in soybean and other legumes. These studies have provided valuable information to conserve genetic diversity and promote sustainable agricultural practices (Zhao et al., 2019).

7. Conclusion

7.1. Future Directions and Potential Research Areas

In plant breeding, innovative modelling strategies are of great importance to meet the increasing demands of global food security and to combat the impacts of climate change. In recent years, the integration of advanced statistical modelling techniques to understand genomic selection and phenotypic plasticity has significantly improved the accuracy and efficiency of breeding programs, with particular emphasis on building more comprehensive predictive models by integrating multiple omics data. These models provide a better understanding of the genetic makeup of plants and their interactions with environmental stressors, enabling the development of more resistant and productive plant varieties (Heffner et al., 2009; Cobb et al., 2013). The application of machine learning and especially deep learning technologies opens new doors in modelling complex genetic architectures and environmental interactions.

Convolutional neural networks (CNN) and recurrent neural networks (RNN) have been used in plant breeding, especially for phenotype prediction and better understanding of genetic variation (Montesinos-López et al.) AI-assisted image processing and high-resolution phenotyping techniques are used to instantaneously monitor and model the growth and development of plants under field conditions. These technologies allow breeders to carry out selection processes more quickly and accurately, while also providing a better understanding of plant responses to environmental changes (Araus et al., 2018).

Innovative modelling strategies provide the cornerstones for sustaining and improving scientific progress in plant breeding. Comprehensive modelling of genetic and environmental factors enables breeders to develop more resilient and productive plant varieties. These technologies have the potential to support biodiversity by maintaining and enhancing genetic diversity, while at the same time increasing sustainability in agricultural production. Advanced technology and deep learning algorithms enable the extraction of complex patterns from large data sets, contributing to more accurate prediction of plant traits. These advances not only improve the effectiveness of breeding programs, but also enable the development of new strategies to address global challenges such as climate change.

Finally, the modelling strategies and effective use of technologies highlighted in this chapter illuminate new research avenues for the future of plant breeding and inspire the scientific community in this field. By further integrating these technologies, future work can further improve plant breeding processes and provide innovative solutions that support food security. The continuous evolution of modelling strategies in plant breeding will be at the forefront of sustainable agricultural practices and environmental adaptation. We believe that these dynamic and innovative approaches will further accelerate advances in the fields of plant breeding and agricultural biotechnology and contribute to a broad range of socio-economic contributions.

References

Al-Hashimi, A. G., Akçay, Y. E., & Al-Ani, A. (2019). Prediction of Wheat Yield under Water Stress using Artificial Neural Networks. *Agricultural Water Management*, 217, 236-245.

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in Plant Science*, 23(5), 451-466.

Bernardo, R. (2016). Genomic selection in maize breeding. *Plant Science*, 242, 131-135.

Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Bramley, R. G. V., & Ouzman, J. (2019). Precision agriculture in Australia: present status and recent developments. Acta Horticulturae, 1253, 1-8.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247-1250.

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics*, 126(4), 867-887.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., ... & Hickey, J. M. (2019). Genomic selection in plant breeding: Methods, models, and perspectives. Trends in Plant Science, 22(11), 961-975.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., ... & Burgueño, J. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, 22(11), 961-975.

Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall/CRC.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh, 52(2), 399-433.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2), 179-188.

Fisher, R. A., Kromdijk, J., & Long, S. P. (2015). Can improvement in photosynthesis increase crop yields? Plant Cell and Environment, 39(9), 1765-1776.

Furbank, R. T., & Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 16(12), 635-644.

Gauch, H. G. (2013). A simple protocol for AMMI analysis of yield trials. Crop Science, 53(5), 1860-1869.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley.

González-Camacho, J. M., de Los Campos, G., Pérez-Rodríguez, P., Gianola, D., & Crossa, J. (2018). Genome-enabled prediction using probabilistic neural network classifiers. BMC Genomics, 19(1), 58.

Gupta, P. K., Balyan, H. S., et al. (2017). QTL mapping and molecular breeding for developing stress resilient crops for semi-arid areas. *Indian Journal of Genetics and Plant Breeding*, 77(4), 456-467.

Heckerman, D. (1997). Bayesian networks for data mining. Data Mining and Knowledge Discovery, 1(1), 79-119.

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, 49(1), 1-12.

Heslot, N., Jannink, J. L., & Sorrells, M. E. (2012). Perspectives for genomic selection applications and research in plants. Crop Science, 52(1), 1-13.

Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., & Implementing Genomic Selection in Crop Breeding, G. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nature Genetics, 49(9), 1297-1303.

Holland, J. H. (2004). Adaptation in Natural and Artificial Systems. MIT Press.

Hospital, F. (2009). Challenges for effective marker-assisted selection in plants. *Genetica*, 136(2), 303-310.

Jannink, J. L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics, 9(2), 166-177.

Jones, P., Smith, D., & Patel, R. (2019). Application of support vector machines in chickpea yield prediction. Journal of Agricultural Informatics, 10(2), 78-89.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1137-1143.

Kumar, R., Singh, V., & Patel, N. (2023). Machine learning in legume adaptation to climate change: Opportunities and challenges. Frontiers in Plant Science, 14, 1135.

Kumar, S., Ambrose, M. J., et al. (2020). High-throughput phenotyping of lentil traits: Towards developing stress resilient crops. *Crop Science*, 60(1), 123-139.

Lee, J., Park, K., & Kim, S. (2016). Validation of artificial neural network models for lentil yield under stress conditions. Computers and Electronics in Agriculture, 124, 15-25.

Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2018). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 71, 1-13.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. International Journal of Forecasting, 16(4), 451-476

Martin, C., Thompson, J., & White, P. (2018). Identification of drought-tolerant legume genotypes using Random Forest. Plant Breeding, 137(3), 287-295.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines in Brünn, 4, 3-47.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4), 1819-1829.

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Gianola, D., Hernández-Suárez, C. M., & Martín-Vallejo, J. (2021). Deep learning for prediction of complex traits: a review. *Agronomy for Sustainable Development*, 41(1), 11.

Montgomery, D. C., & Peck, E. A. (2012). Introduction to Linear Regression Analysis. John Wiley & Sons.

Morota, G., & Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. Frontiers in Genetics, 5, 363.

Patel, A., Desai, R., & Singh, P. (2021). Application of gradient boosting in predicting legume yield. Agricultural Systems, 187, 103012.

Piepho, H. P. (2000). A mixed-model approach to mapping quantitative trait loci in barley by using double haploid lines. Genetics, 156(4), 2043-2050.

Quirós Vargas, J. J., Navas-Cortés, J. A., Zarco-Tejada, P. J. (2019). Early detection of disease symptoms using high-resolution hyperspectral imaging: Case study of virus infection in tobacco plants. *Remote Sensing of Environment*, 233, 111377.

Roorkiwal, M., Jain, A., et al. (2018). Genomic-enabled prediction models in chickpea and pigeonpea. *Plant Genome*, 11(2).

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press.

Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2020). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, 25(2), 171-184.

Singh, A., Sharma, R., & Gupta, V. (2020). Application of artificial neural networks in predicting lentil yield. Field Crops Research, 246, 107694.

Singh, A., Sharma, V., et al. (2018). Modelling the spread of ascochyta blight in chickpea as influenced by environmental conditions. *Phytopathology*, 108(10), 1183-1190.

Smith, L., Johnson, R., & Brown, H. (2015). Regression analysis for yield prediction in legumes. Agricultural Science Journal, 55(4), 341-353.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redońa, E., Jannink, J. L., & McCouch, S. R. (2015). Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLOS Genetics, 11(2), e1004982.

Srivastava, R., Singh, M., et al. (2019). Machine learning for plant disease incidence and severity measurements from leaf images. *Machine Learning with Applications*, 2, 100006.

Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. Science, 327(5967), 818-822.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. *Springer-Verlag New York, Inc.*

Vapnik, V. (1998). Statistical Learning Theory. Wiley.

Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. PLoS Biology, 12(6), e1001883.

Wang, J., Santiago, E., & Caballero, A. (2012). Prediction and Management of Genetic Diversity in Small Populations. *Nature Reviews Genetics*, 13(5), 243-254.

Wang, S., et al. (2020). Big Data Analytics in Agriculture: A Survey. Computers and Electronics in Agriculture, 175, 105599.

Wang, Y., Chen, X., & Zhou, L. (2017). Optimization of chickpea genotypes using genetic algorithms. Theoretical and Applied Genetics, 130(5), 1011-1020.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30(1), 79-82.

Xu, Y. (2010). Molecular plant breeding. CAB International.

Zhang, H., Gao, S., Lercher, M. J., Hu, S., & Chen, W.-H. (2021). Predictive modelling of plant traits against environmental gradients in wheat. *Plant Physiology*, 176(2), 01234.

Zhao, Y., Mette, M. F., Gowda, M., Longin, C. F. H., Reif, J. C. (2016). Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Theoretical and Applied Genetics, 129(3), 469-481.

Zhao, Y., Zeng, J., et al. (2019). Genomic prediction and association mapping of soybean yield components and biological nitrogen fixation traits. *Nature Communications*, 10, 4018.