IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

Using Machine Learning tools and Techniques, Genetic Disease Analysis Prediction on agriculture crops using AI & ML

Adlin Jebakumari S[1], Dr. A. Jayanthiladevi[2]

Research Scholar, College of Computer Science and Information Science, Srinivas University, Mangalore, India.[1]

Professor, College of Computer Science & Information Science, Srinivas University, Mangalore, India.[2]

Abstract:
Purpose: The most often published international journal articles are shown to discover the study fields that are most interested in ML in agriculture. In all, 129 journals were pertinent to the topic of the course. Remote sensing, as a general point, was important because satellite data, for example, provides significant input data for ML systems. Agricultural sustainability, smart farming, and the environment were also major concerns. There was also a high frequency of journals related to computational approaches. Methods that combine various stimuli to make a choice are referred to as EL in this context. One kind of inducer receives labelled samples and builds a model capable of generalizing these examples. As a result, more unlabelled cases may be predicted. Weed identification methods relied heavily on images as input data. Photos derived from in-situ measurements and multispectral images from the aforementioned sources were used to create these images. Regression-based multivariate statistical approaches are the most efficient for indirect selection. Artificial neural networks (ANN) may be used in plant tissue culture research to perform pattern recognition, nonlinear regression, and classification due to their ability to process binary information, continuous information, categorical information, and fuzzy information. A target function is used to establish a starting population and an evaluation of an individual's fitness at the beginning of the process. Accuracy necessitates the acquisition of a big dataset. Each picture in each category has its own set of characteristics that we've extracted. GLCM texture extraction and edge detection are used to extract the features, and the degree of moisture content is measured using these two techniques. Using pixel values, these characteristics define the plant's current state. As an analytical tool, moisture content is one of the most important factors.

Methodology: In the third part, we examine the applied technique, inclusive criteria, and search engines. This section also includes the metrics that were utilized in the chosen articles. It makes use of cutting-edge sensor and actuator technology to meet the problems and potentials of suitable approaches in agriculture, as outlined in this contribution. Reduction approaches are available in the Statistics and Machine Learning Toolbox to help you uncover variables or features that influence your model. Data-driven methods, like ML, function better when more data is available. This is analogous to the improvement in performance that comes with practice and repetition in real life. However, SVM is a widely accepted classification tool in a wide range of data-driven fields, including agricultural research.

Findings/Result: Agricultural firms may save money by decreasing the cost of illness diagnosis as a consequence of the findings. Crop farming might see a reduction in expenses as a consequence of these findings. The neural network prediction results in the detection of late blight disease on the leaves of a plant. It is essential to keep in mind that each branch in the tree represents the same comparison and that each node in the tree implies a pairwise comparison about a certain feature. An essential consequence of the study being reviewed right now was the presentation of the input data that was used in the Machine Language (ML) algorithms and the sensors that were associated with them.

Originality: Using the five-fold k-fold cross-validation, the initial dataset of 250 observations was randomly partitioned into training sets and validation sets to perform the analysis.

Paper Type: The authors of this study provide an in-depth analysis of precision agriculture's use of Machine Language (ML). An overview of the research community's views on the implementation of digital methods in the farm management system will be provided in this review paper. Agricultural crop disease diagnostics might benefit through a genetic approach, according to this article. Consider the diagnosis of potato leaf late blight as an example. According to this paper's findings, 20% of the 10% for each category was accounted for by the general categories dealing with water and soil management.

Keywords: Artificial Intelligence; Convolutional Neural Network (CNN) & K-Mean Clustering mechanism.

DOI: 10.17148/IARJSET.2023.101224

I. INTRODUCTION

Modern agriculture needs to deal with several issues, such as the rising demand for food due to the world's growing climate change [1], natural resource depletion [2,] dietary shifts [3, 4], and worries about food safety and health. An urgent demand arises for agricultural techniques to be optimized to alleviate the strain on the agricultural sector and at the same time reduce the environmental impact. Precision agriculture has been propelled in large part by the development of these two key technologies. Agricultural modernization has the potential to provide long-term sustainability, maximum production, and a safe environment [5]. A potential multidisciplinary path at the intersection of agriculture and information technology is to automate and speed up the process of looking for the best solutions for these jobs, which reduces the expenses of agricultural operations. It is possible to efficiently identify gene copy number variation, genome structural variation, and single-nucleotide polymorphism (SNP) variation, which enables the establishment of population-wide genotyping collections for a broad variety of species [8]. We are now capable of quantifying changes throughout the whole genome at the biochemical level. These modifications could include the amounts of components including RNA, proteins, and metabolites, as well as their interactions with one another. The term "-omics" is used to refer to these collections of data. [7].

ML has a slew of potential uses in the agricultural sector. A recent literature review by Lakes et al. [12] found four general groups for the years 2004 to 2018. (Figure 1). Crop, water, soil, and livestock management are all included in these areas. When it came to articles on agriculture, crop management accounted for 61% of all content and was further broken down into the following subcategories:

a) Yield prediction;
b) Disease recognition;
c) Weed recognition;
d) Crop detection;
e) Crop quality.

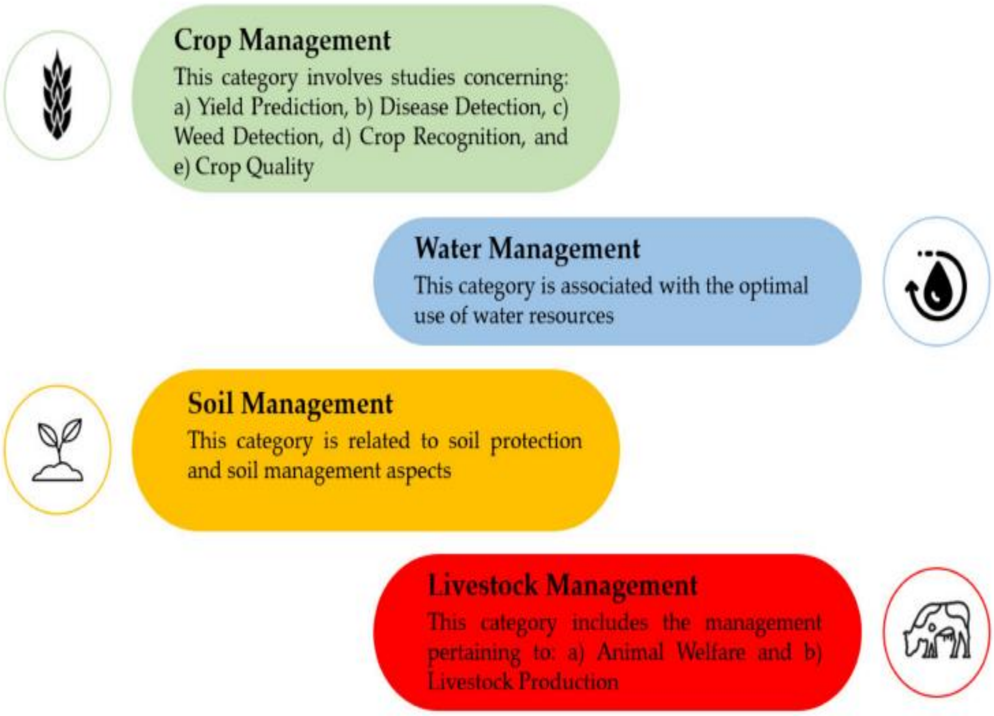

Figure 1. The four categories in agriculture

The development of farms and agro-based enterprises relies heavily on technology. Due to advancements in technology, it is now feasible to cultivate crops in arid regions. Technology has changed the agriculture style. In agriculture, automation technology is now the most sought-after tool.

DOI: 10.17148/IARJSET.2023.101224

Farming is being transformed into Digital Agriculture thanks to the newest innovations in Machine Learning and Artificial Intelligence. Many studies have shown that putting technology to use on farms increases agricultural yields and, therefore, profits for farmers. Component characteristics of yield are the most important determinants of plant production [10–12].

Numerous characteristics, such as the number of nodes on a plant, NRNP, RNP, and PP, which are together referred to as important yield components, all have a role in determining the yield of soybeans [13]. As a consequence of this, selecting soybeans that have superior performance in their yield components might potentially limit the rise in yield output [13, 14]. Several research [15–18] have shown how boosting yield components improves yield in plants.

II. OBJECTIVE

Most machine learning techniques are used to improve task performance by using examples or prior experience. Furthermore, ML can establish efficient links between data inputs and reconstruct a knowledge architecture, which is very useful. Data-driven methods, like ML, function better when more data is available. This is analogous to the improvement in performance that occurs when one's level of experience increases [22].

Machine learning (ML) is all about generalizability, which measures how well ML algorithms can predict new data based on previously learned rules based on their previous exposure to comparable data. [23] As a result of the goal function's work, the Model class and the Algorithm class were created to provide numerical techniques for computing the characteristics of composites with complex variables. The main objective in plant research and breeding is to use genetic, phenotypic, and environmental data to explain a complicated attribute like yield. The following are the goals of machine learning and other techniques in this regard:

Determine how much of a trait's variation can be explained by a given locus and how much can be explained by the combined effects of all the loci. Similarly, genetic correlations between phenotypes are interesting because they measure the degree of overlap between genetic signals.

new genotypes, where only marker data is provided to infer anticipated trait values (genomic prediction [GP]).

On the other hand, it will be some time before data sets that are sufficiently big and rich are made available. Active learning, and transfer learning are all examples of developments in ML that may help to alleviate this problem, but they are only a partial solution. These will offer more data, more objectively and correctly. However, these are only a few examples of developments in ML that may help to alleviate this problem.

III. EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

When John McCarthy first used the phrase "artificial intelligence" at the Dartmouth conference in 1956, he described it as a branch of science and engineering that focuses on creating intelligent machines, notably computer programs. Computational intelligence is provided to machines by AI technology, allowing them to learn, comprehend, and respond to their environment. Machine learning and deep learning are shown in Figure 2.

This field of study has a wide range of potential applications, which may be found in several spheres of human endeavor. Intelligent artificial intelligence systems are now undergoing testing in a variety of industries, including healthcare, agriculture, banking, robotics, e-commerce, and the automation industry. Several of the world's largest electronics firms, including Samsung, Apple, and others, have declared plans to include this technology in all of their next products.

Another new technology is the Internet of Things (IoT), which connects smart sensors and gadgets via the internet. For example, these smart sensors may be used to collect data across a variety of domains, from solar plants to agricultural fields to disaster-prone locations, and the manufacturing business, using the AI methodologies indicated in Fig. 2.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

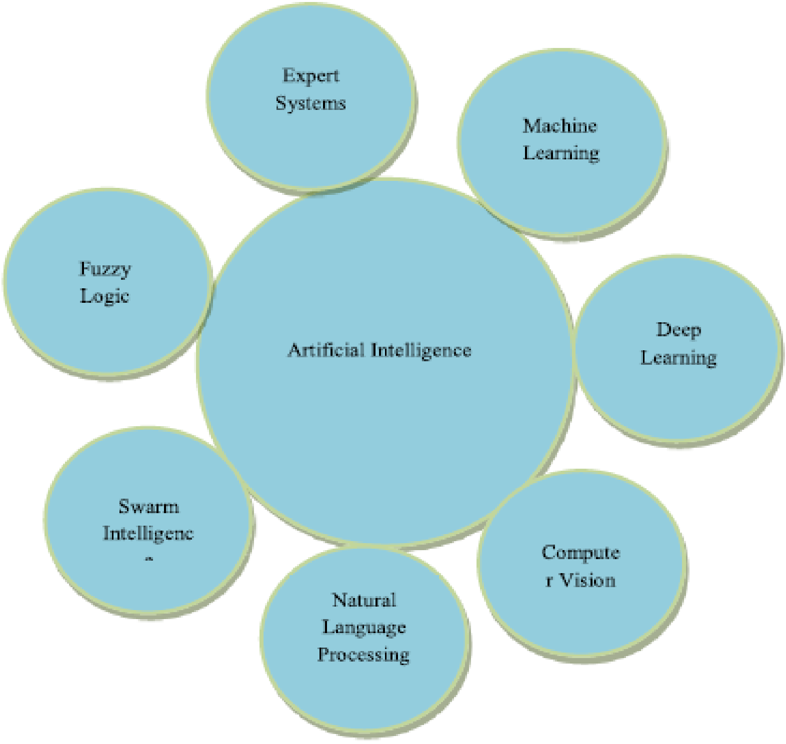

FIGURE 2. Artificial intelligence techniques

Genetic |Disease Prediction using AI & ML		
S.No.	**Author & Year**	**Findings & Research**
1	C. Wang, ,Q. Zhao, G. Jin, and X. Yin, 2013	Using KNN and SVM, a two-fold surfaced representation is employed to restrict the curves in the overlapping arrangement of layer. More than 96 percent of the infection area is confined by a combination of morphological division, design coordination, and tint coordination. Figure 8 displays photographs of cotton leaves that were taken during the first phases of the improvement process.
2	S. Jeevan Prasad, C. H. Usha Kumari, and G. Mounika, (2019).	The majority of cotton plant illnesses are found on the plant's leaves. The main objective of disease recognition technologies that have been used in the past has been to identify plant diseases at an early stage to maximize crop output.
3	3. T. Suman, S. Kumar, K. M. V. V. Prasad, A. Srilekha, B. Pranav Rao, and J. Naga Vamshi Krishna, (2020).	Root exudates from the plant play a vital function in enhancing the soil's nutritional content. When compared to their wild relatives, plants that have been cultivated are always more resistant to disease. What we're talking about here is a very large number of organisms of the same type or many kinds that share a gene and develop together at a distance of several kilometres from one another.
4	Z. Y. Wang Diao, and H. Wang, Y. Song,	To maintain a healthy ecosystem, plant diseases are an important ecological element. Pathways inside plant cells that strengthen defences against animals and insects, as well as infections, are primarily responsible.

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 12, December 2023
DOI: 10.17148/IARJSET.2023.101224

5	Windham, G. L., Williams, W. P., Buckley, P. M., & Daves, C. A. (2002).	Many fungi, bacteria, viruses, mycoplasma, and nematodes are included in this group. When a plant is infected with a disease, it loses much of its normal health and ability to function. Seasonal variations in weather conditions, crop types and cultivars, and pathogen populations may all influence the frequency and severity of plant diseases. Some plant species are more prone to disease outbreaks than others, and this may be shown in their genetic makeup.
6	Shan, Q. Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., ... & Gao, C. (2013)	The pathogen's genetic diversity is critical to studies of disease resistance and to disease management via host resistance. Genetically engineering crops might be a way to increase productivity and minimize the use of pesticides in current agriculture, which may represent a risk to human health.
7	Holme, IT., & Holm. B., Wendt, P. B. (2013).	Crop plants should be genetically modified to increase output and decrease the need for pesticides, which may constitute a health hazard to humans. A better knowledge of the genetic foundation for plant health and novel ways to improve disease resistance and crop output will be gained via the identification of the genetic basis.
8	Balint-Kurti PJ, Dixon MS, Jones DA, Norcott KA and Jones JDG (1994)	Plants that are resistant to disease are an essential part of sustainable farming. To comprehend it and use it in the most effective manner possible, the genetic foundation of disease resistance that occurs naturally has been the subject of intensive research. Two different genetic pathways might confer resistance to disease. These investigations provided the evidence to support that.
9	Agrawal, R.L. 1996.	It has been important to preserve a balance between the key components of genetics and plant breeding, such as fundamental genetics and biotechnology in the development of post-graduate courses.
10	Wilson RF , 2012	Many of the beneficial qualities of polyunsaturated oils may be acquired from oils with high natural oleic acid content.
11	Ramasubramanian V, Beavis, 2020	Five approaches were used for 40 cycles of recurrent selection, and interactions between these parameters were analyzed.
12	Rebetzke G, Jimenez-Berni J, Fischer R, Deery D, Smith D. 2019	Many crop breeding efforts will limit their genetic basis via controlled genomic selection and higher heritability.
13	Yuan J, Njiti V, Meksem K, Iqbal M, Triwitayakorn K, Kassem MA, et al. 2002	The purpose of this research was to recognize genetic markers linked with yield quantitative trait loci (QTL) in two RIL populations.
14	Tester M, Langridge P. 2010	In order to speed up breeding, new technologies must be developed to improve genotyping and phenotyping methodologies and to increase the genetic variety of breeding material.
15	Araus JL, Cairns JE. 2014	High-throughput phenotyping platforms (HTPPs) in the field are still a major challenge for future breeding advancements.

DOI: 10.17148/IARJSET.2023.101224

16	Qiu R, Wei S, Zhang M, Li H, Sun H, Liu G, et al. 2018	This limits the advancement of breeding since the database is inadequate to fulfil the needs of plant breeders, making traditional phenotyping time-consuming and tedious.
17	Robbins MD, Staub JE. 2009	There have only been a few empirical studies that are studied to analyze these selection processes within the framework of a plant breeding effort.
18	Richards R. 2000	Leaf area, and daily photosynthesis duration, have all contributed to an increase in total photosynthesis.
19	Kumudini S, Hume DJ, Chu G. 2001	Analysis on the impact of water deficit stress on maize lines that are non-drought tolerant and drought-resistant was conducted.
20	Zeng Q, Huang H, Pei X, Wong S, Gao M. 2016	As part of this research, a neural network (NN) model was developed to investigate the nonlinear connection between accident frequency and risk variables.

IV. LITERATURE REVIEW

To avoid overfitting and local minima, an artificial neural network (ANN) requires a significant amount of training samples, and optimization utilising back-propagation methods cannot be carried out effectively [24]. To detect diseased plants, ANN-based approaches are superior to SVM-based methods because they employ more samples and characteristics [25]. Understanding the transfer of information from the DNA sequence to the visible phenotypes of plants has been the primary focus of research on plant breeding conducted both traditionally and via in vitro-based methods. Machine learning, which is the discipline of programming computers to learn from data, has been used in both types of research. Three different classifications may be used to organize the many approaches that are used in machine learning and are shown in Figure 3.

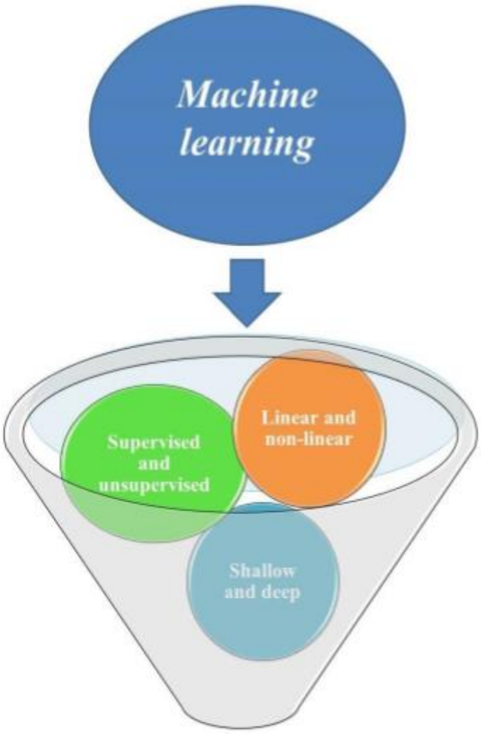

Figure 3. Different categories of machine learning algorithms.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

Because they don't need previous data structure or extensive knowledge about physical processes, artificial neural networks (ANNs) are nonlinear nonparametric models that can cope with data loss and include several hidden layers [26, 27]. DNNs outperform ANNs in terms of predictive capability thanks to the additional layers they include. The DCNN takes its cues from the natural visual perception process of live beings, which serves as their source of inspiration [28].
Because they can automatically extract features, CNNs are well-suited for classification research [29]. Images can be classified, objects can be tracked, and poses can be estimated using CNNs [28]. Algorithms may be tested against each other to identify the most effective method for solving a given problem. PLS-DA stands for partial least squares discriminative analysis [23]. Table 1 shows the advantages and disadvantages of several nonlinear machine learning approaches in comparable situations.

Table 1. Machine learning algorithms comparison

Leaning Algorithm	Advantages	Disadvantages
ANNs	• Good learning capabilities	• Lack the interpretation capability • Overfitting and local minima in small number of training data • Implementing very small number of hidden neurons
CNNs	• Ability of automatic feature extraction	• Lack the interpretation capability • Require large amounts of data for training • Require considerable skill and experience to select suitable hyperparameters
SVMs	• Uses a large number of hidden units • Quadratic optimization task in the formulation of the learning problem	• Shallow architecture
RF	• Ability to handle noise • Prevent overfitting • Ability to manage a large number of features	• Shallow architecture

ANNs—artificial neural networks; CNN—convolutional neural networks; RF—random forest; SVMs—support vector machines.

Figure 4 illustrates various application areas for nonlinear machine learning technologies that may be used in traditional and in vitro-based plant breeding research.

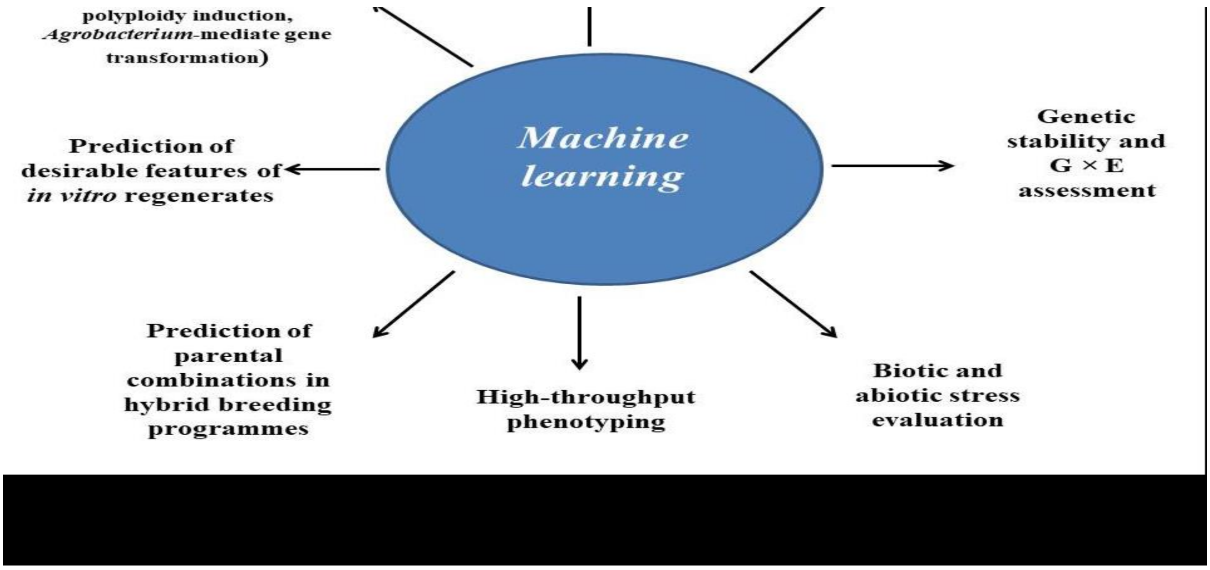

Figure 4. Machine learning applications

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

Table 2 is a listing of several utilised relatively recently in either traditional or in vitro-based plant breeding research.

Table 2. Nonlinear machine learning models

Plant Species	Type of Machine Learning	Techniques	Purpose(s)
	ANN	MLR	Modeling and predicting of seed yield
Ajowan (*Trachyspermum ammi* L.)	ANN	MLR	Modeling and predicting of essential oil content
	ANN	MLR, IP	Predicting physical properties of embryogenic callus and number of somatic embryos
Arabidopsis thaliana	DT, SVMs, NB	Gaussian kernel	Predict the plant abiotic stresses response through the miRNAs' concentration
Carrot (*Daucus carota*)	RF	-	Precision agriculture-yield mapping
	ANN	GA	Modeling and optimizing of in vitro sterilization
Chrysanthemum	ANFIS	GA	Modeling and optimizing of somatic embryogenesis
	ANN, SVMs	MLP	Modeling effect of plant growth regulators on somatic embryogenesis
Cucumber (*Cucumis sativus*)	CNN	IP	Segmentation and quantification of powdery mildew disease
Garnem (G x N15) *Prunus* rootstock	ANN	GA	Prediction and optimization of mineral salts of in vitro culture medium
	ANN	GA	Modeling and optimizing of in vitro hormonal combination
	ANN	GA	Modeling and optimizing of new in vitro culture medium
Grapevine rootstock	ANN	Principal coordinate analysis, UPGMA	Genetic diversity assessment through molecular markers (RAPD-SSR) dataset
Maize (*Zea mays* L.)	CNN	IP	Identification of haploid and diploid maize seeds
	CNN	IP	Classification model to identify the infected and healthy leaves
	CNN	IP	Plant diseases recognition
	CNN	IP	Identification and classification of drought stress
Okra (*Abelmoschus esculentus* L.)	DNN	IP	High-throughput salt-stress phenotyping
Pearl millet (*Pennisetum glaucum*)	DNN	IP	Identification of mildew disease
Potato (*Solanum tuberosum*)	ANN	IP	Identification and discrimination of potato varieties
	RF		Classification of *Phytophthora infestans* infected cultivars
Rapeseed (*Brassica napus*)	ANN	MLP	Seed yield modeling
	CNN	IP	Stand count estimation
	ANN	MLP	Multicriteria yield prediction based on meteorological data and mineral fertilization data
	ANN	MLP	Early prediction and simulation of seed yield based on meteorological and mineral fertilization data
Rice (*Oryza sativa*)	CNN		Plant diseases and pest recognition
Safflower (*Carthamus tinctorius* L.)	ANN	MLR	Seed yield modeling
Sesame (*Sesamum indicum* L.)	ANN	MLR	Oil content modeling
	ANN, SVMs	RBF, ERBF, GRNN, M5-Rule, M5-Tree, MLR	Estimation of oil and protein content
Soybean (*Glycine max*)	CNN	IP	Estimation of seeds per pod
	DNN	IP	Evaluation of stomatal density diversity
Tomato (*Lycopersicon esculentum* L.)	ANN	MLR, IP	Modeling of callus induction and regeneration in anther culture
	CNN	IP	Evaluation of disease severity
	ANN	MLP	Estimation of salinity tolerance
	ANN	MLP	Prediction of seed yield based on meteorological data and information on mineral fertilization
	ANN	MLP	Prediction and simulation of seed yield with qualitative and quantitative data sets
Wheat (*Triticum aestivum* L.)	CNN	IP	Quantification of spikes
	DNN	LSTM	Production forecasting
	CNN	-	Genomic selection
	ANN, GRNN	MLP	Modeling in vitro shoot regeneration
	ANN	MLP	Analysis of concentration of ferulic acid, deoxynivalenol, and nivalenol
White ginger (*Hedychium coronarium*)	ANN	MLP	Prediction and optimization of coronarin D content

4.1 Types of Cotton.

Perennial shrubs in the Malvaceae family are produced for their soft fabric under the collective term "cotton." Figure 3 depicts the four species of plants in the genus that are capable of preserving plant seeds. Nearly 90 percent of the planet is made up of Gossypium hirsutum. Many stems branch out from the primary stem of these plants. There are three to five triangular lobes and lengthy petioles on the branches of the plant. "Boll" is the name given to the leathery, oval seed capsule that occurs on each auxiliary branch of the plant, which may be red or purple, yellowish-white in color, and 2 to 6 centimetres long. One to 1.5 meters (3.3 to 4.9 feet) is the maximum height of a cotton plant.

4.2. Diseases of Cotton and its Symptoms

Lack of nutrition and chemical variables that may lead to imbalances in the cotton plant are the most prevalent factors that contribute to the development of diseases in cotton. During the growth of the cotton crop, several variables influence the amount of cotton produced. There is a stark contrast between modifying a plant and exposing it to potentially hazardous conditions, such as those caused by diseases. Cotton productivity is severely constrained by verticillium wilt and cotton leaf curl [30, 31].

DOI: 10.17148/IARJSET.2023.101224

4.3. Disease Analysis - Vascular Wilt Disease.

An organism that is responsible for causing harm. Aspergillus neoformans and Aspergillus necroticus (Atk.) Snyder and Hansen are the authors of this piece. Symptoms. This disease is more likely to spread in soils with a pH between 6 and 8.00. It affects almost every component of the plant. This yellowing, as well as the sautéing and filling of the dried leaves, occurs in seedlings. During this part of the process.

The following three diseases affected in cotton plants are shown in fig. 5.

(a) (b) (c)

Figure 5: (a) Angular leaf spot, (b) vascular wilt, (c) grey mildew

4.4 Anthracnose Disease.

An organism that is responsible for causing harm. **Symptoms.** Spots of reddish color emerge on the cotyledons and first leaves of seedlings. Seedlings may wilt and perish if the lesions are concentrated on the top portion of the stem. After infecting a mature plant, the fungus will divide the stem and cause the bark to be shredded off of it. Bolls show water-soaked, round, somewhat sunken, reddish-brown patches that eventually become black as symptoms. They open early because of a blood infection. Consequently, the lint is discoloured, hard, and compacted as a consequence. Detection Techniques That Work Best. Fuzzy logic and decision trees.

4.5 Root Rot Disease.

An organism that is responsible for causing harm. Symptoms of Rhizoctonia bataticola (Taub.) Brown patches on the cotyledons may be the first sign of the disease in the seedling stage. There is a black shade around the collar, which may extend to the lower sections. The roots of the sinewy stems are decaying. The bark of the roots shows signs of root decay and shredding. It is simple to remove plants that have been influenced. The field is showing signs of infection. Detection Techniques That Work Best. A network of neurons.

(d) (e) (f)

(g)

Figure. 6 (d) anthracnose, (e) root rot, (f) boll rot, and (g) leaf spot or blight.

DOI: 10.17148/IARJSET.2023.101224

4.6 Boll Rot disease

An organism that is responsible for causing harm. The species Macrophomina phaseolina Goid It's a disease caused by a variety of different fungi: Anthracnose-causing fungi include Fusarium moniliforme, Botryodiplodia species, Colletotrichum capsicum and anthracnose-causing A. Niger. Symptoms. Small, brown, or black dots occur at the outset of the illness and spread to encompass the whole boll. The infected stem or lint is the consequence of contamination spreading to internal tissues. In the early stages, the bolls do not break open or fall off. When the stem is decaying, the infection can spread to the pericarp, allowing the interior tissues to be exposed. The diseased bolls may have a huge quantity of fungus on them. Detection Methods Used. A network of neurons.is shown in Fig. 6 above.

4.7 Leaf Spot or Blight Disease.

An organism that is responsible for causing harm. Alternaria macrospora Zimm. Symptoms. It is possible for 'e illness to take many different forms. After 45 to 60 days after planting, the leaves begin to suffer the most. A core lesion is in the centre of each area, which is encircled by concentric rings of lesions. Blighted regions are formed when several locations come together. When leaves are infected, they weaken and wither. Occasionally, stem lesions are seen. Occasionally, these patches may occur on the bracts and the bolls of a flower. Detection Techniques that Work the Best. In the image above, the defective illness leaves are displayed by image processing and a support vector machine.

V. SUMMARY OF RELATED WORKS

The quality of the crop is extremely important for the market, and it is generally tied to the soil conditions, climate, cultivation practices, and crop characteristics Products of a higher grade in the agricultural industry often fetch higher prices, which in turn brings in more money for the farmers that produce them. When it comes to fruit quality, for example, the flesh hardness, the quantity of soluble solids, and the color of the skin are some of the most common maturity indicators that are used for harvesting [32]. When high-value crops as well as arable crops are harvested, the date of harvest has a significant influence on the qualitative characteristics of the items that are obtained from the harvest. Consequently, the development of decision support systems may provide farmers with the ability to choose appropriate management options, resulting in improved product quality. One type of management that, for instance, has the potential to significantly improve quality is selective harvesting. An examination of the published research in the relevant fields reveals that there is a sizeable window of opportunity for the use of machine learning strategies in the performance of agricultural duties. These responsibilities include the creation of novel technologies for the automated processing of crops. Machine learning is a cutting-edge technology that is now being used to tackle complex problems in the agricultural industry and assist farmers in cutting their financial losses. In this analysis, it was shown that machine learning algorithms have received outstanding results when applied to the problem of overcoming problems associated with agriculture.

VI. RESEARCH GAP

'The Turing Test for Machine Intelligence' was written by Alan Turing in the year 1950, when he suggested the idea of learning machines. He conducted a test to see whether the computer might exhibit human-like intelligence. Learning and extracting information from the input, an intelligent computer programme constructs a framework for making intelligent judgments. As seen in Figure 3, the ML process is broken down into three main components: data input, model development, and generalization. Predicting an algorithm's result based on data that it has never seen before is known as generalization. Weather forecasting, spam filtering, disease detection in plants, and pattern recognition are all examples of complicated issues where ML algorithms may be used. RF has piqued the interest of researchers in a variety of domains, due to its ability to function adequately.

Overfitting in single prediction algorithms is one of the main drawbacks of employing individual ML algorithms [34]. Ensemble approaches [35] may be used to get around this problem. The use of ensemble approaches, which integrate many algorithms to increase prediction accuracy by minimizing overfitting rates is considered an important factor in Machine Learning. Stacking, boosting, and bagging are three popular ensemble techniques that are utilized depending on the dataset and the different machine learning algorithms used. To evaluate and comprehend the huge amounts of phenotyping data that we can currently collect, machine learning (ML) is being used in plant research. The importance of ML is only going to grow in the next years, as researchers get easier access to high-throughput measurement data generation and storage technologies. As a result of inexpensive technology and the extensive availability of data on the Internet, tremendous advances have been fostered. Academic and data-centric industries (Google, Facebook, Amazon, etc.) are constantly developing new methodologies and tools and making them accessible to the general public, frequently for free.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

VII. RESEARCH AGENDA

Many researchers have encountered difficulties related to data, such as a lack of data, a lack of data in the correct format, a lack of quality data, and the presence of additional characteristics. Data sources like Kaggel, Meandly, and IEEE Data Port are used by several researchers to develop models from their findings in this survey Researchers must create their dataset if the necessary data is not readily accessible [36, 38]. Create your dataset and make it available to other academics using open platforms such as Kaggel, Meandly, IEEE Data port, etc. to train your model. Use publicly accessible datasets for testing and validation of the models. Use "Transfer Learning" approaches to speed up model training. In the future, additional agricultural researchers will benefit from this in-depth look at the various machine-learning methods already in use.

7.1 Analysis of Research Agenda:

In addition to climate reanalysis [39], the Alberta Climate Information Service (ACIS) provides spatially interpolated weather station network data (on an hourly time scale) (1961-2016) 1. Station-based air temperature estimations from up to eight nearby stations are weighted by inverse distance within a 60-kilometer correlation radius in the ACIS interpolation technique. Rainfall totals are redistributed proportionately to the closest station that has a full monthly record by weighting precipitation inversely by distance. Among the most advanced climate reanalysis datasets presently available, the high-resolution JRA-Year Reanalysis (JRA-55) was chosen. There is also a 3-hourly (and 6-hourly) temporal resolution for the years 1990-2015, as well as a geographical resolution of 0.55° X 0.55° (55 km).

7.2 Advantages

It is anticipated that this study will help raise awareness of the potential benefits of machine learning in agriculture and contribute to a more systematic investigation of this field as a guide for all stakeholders. Additionally, it is anticipated that this study will help raise awareness of the potential benefits of machine learning in other industries. EL, on the other hand, has piqued attention due to its ability to aggregate predictions from several models, in contrast to [42]. SVM is the third and final member of the set of three machine learning models in agriculture that are considered to be the most accurate due to its excellent performance when dealing with visual data[41]. It is feasible to predict the results of treatments by using these approaches; for instance, it is conceivable to predict what will occur if a gene is silenced. This is a benefit of using these methodologies. There is a possibility that ML techniques provide advantages over other, more stringent statistical methods; nevertheless, the increased flexibility may come at a cost. For starters, ML approaches don't do a good job of explicitly inferring and so don't produce accurate confidence estimations (bounds or p values). There are a few well-known ML algorithms that don't readily provide consumers with information about how a certain prediction was achieved. When it comes to analyzing the complicated function of gene activity in response to environmental perturbations and forming plant phenotypes, ML provides distinct benefits over other approaches. [43] used differentially expressed genes to classify abiotic and biotic challenges, advancing our knowledge of plants' numerous stress responses, which are fundamentally complicated.

7.3 Benefits

To name just one, ANN models can handle noisy data [44], which is a regular occurrence in agricultural measurements. Unsupervised, semi-supervised, or supervised DNNs are all valid options for DNN implementations in machine learning. Different from other types of neural networks, the layers of ConvNets may build up neurons in three dimensions, making them a typical DNN type [45]. In addition to wasting fertilizer and polluting the environment, these methods failed to provide the optimal results from fertilizer application. Because of this, there is a need to find better fertilizers and more scientific fertilization techniques while also ensuring that an adequate nutrient intake is made. If you compare genetically modified crops to their conventional counterparts, you'll find that their yield, stress tolerance (such as resistance to disease and insect pests), as well as their nutrient content, will be greatly improved. The commercialization of genetically modified crops lowers agriculture's costs considerably and provides a steady stream of economic, environmental, and social advantages throughout the globe. More food security, improved nutritional content and quality and resistance to pests and disease are only some of the advantages of genetic engineering in agriculture. To create new and better plants, plant breeders may now use the basis of Mendelian genetics to do very precise crossbreeding. Plants resistant to pests or disease have also been developed using these strategies to increase yields.

7.4 Constraints

The lack of timely and hassle-free availability of agricultural inputs, such as HYV seeds, fertilizers, pesticides, farm implements, etc., is one of the most serious limitations on agricultural productivity. People can use genetic information to observe or monitor the growth status of crops and provide guidelines for the management of fields.

DOI: 10.17148/IARJSET.2023.101224

This allows us to increase the efficiency of agricultural practices such as fertilization and irrigation, as well as regulate the maturity and growth habits of other significant growth processes. All of these factors combine to make it difficult for people to start their farms.

7.5 Disadvantages

"increasing levels of recognised naturally existing toxicants, the emergence of novel, not previously detected toxicants [and] greater potential of concentrating toxic compounds from the environment" were some of the specific warnings from the experts (e.g., pesticides or heavy metals).

The following are the disadvantages of hereditary diseases:
As a result, the nutritious content of food may be lower.
It is possible that pathogens can adapt to the new genetic profiles
• There may be unforeseen side effects
• The quantity of variety created can be less advantageous
• Copyrighted genetic engineering can have significant repercussions. As a result, they may be a factor in an increase in allergic responses, as well as antibiotic resistance and cancer
• Genetic food may provoke allergic reactions from diverse foods
• GMOs may contribute to antibiotic resistance
• Some studies have connected GMOs to cancer.

7.6 RESEARCH PROPOSAL:

"Discovery Science," "Science to Practice," and "Infrastructures" are the major categories for these purposes. Among the expanded and revised aims in this plan are additional goals to better characterize the micro biome, to increase the use of gene editing and other biotechnologies, and to preserve genetic diversity. Animal genomics research topics are broken down into sections, each with its own set of deliverables and collaborations that must be developed to reach their full potential. Brazil1, India2, and China3 are boosting their spending on agricultural research so that they can contribute more to the global market. New techniques for animal production that fit the needs and values of consumers and the expanding demands of the global population must be developed by the United States if it is to stay competitive. This country's history of innovation in animal husbandry necessitates the development of new technologies that boost the efficiency of the production systems Some crops (e.g. wheat, rice, and soybeans) may have mineral concentrations up to 8% lower than usual in the presence of increased carbon dioxide levels. Carbohydrates may be greater and protein concentrations lower (FAO, 2015). Additionally, climate change is projected to lead to a rise in water-borne infections, such as diarrhea, which hurt a person's capacity to absorb the nutrients from food. Temperatures are expected to rise, and rainfall is expected to decrease, reducing the availability of clean water and increasing the risk of water-borne infections spreading.

VIII. CONCLUSION

It is clear from the geographical distribution and the wide range of study topics that ML applications for agricultural management are a key problem on a worldwide level. Its adaptability makes it ideal for studying convergence. Because it relies on the cross-fertilization of findings from several disciplines, a study of convergence research may be beneficial to society as a whole in the future. There are various sides to this, such as reducing one's impact on the environment and ensuring one's health. ML has a lot of potential to provide value in agriculture if it is used in this way. Agricultural crop disease diagnostics might benefit from the use of a genetic approach, according to this article. Consider the diagnosis of potato leaf late blight as an example. Using genetic algorithms and a recurrent form of the neural network, which can identify damaged spots on potato leaves, it is suggested that the forecasting and diagnostic issue be solved.

Crop farming might see a reduction in expenses as a consequence of these findings. For example, in genetic modification (GM) procedures, ML may be used to forecast which portions of the genome should be modified to obtain a particular phenotype, or it can be used to ensure optimum local growth circumstances by assessing crop performance in vivo in the greenhouse or field. Even though these are technological challenges, strong machine learning techniques will provide researchers with useful tools, particularly when the researchers are better equipped to allow the interpretation of the findings that they reach. As a result, machine learning can help us meet the problems of ensuring food security in fast-changing circumstances for an increasing population. Soybean seed production may be accurately predicted using RBF alone, according to findings from this research. For soybean seed yield prediction, the RBF metaClassifier-based E-B algorithm surpassed all individual ML techniques and is thus recommended. Using real-world field data, we combined the E-B algorithm with GA for the first time to determine the optimal values of yield component characteristics in a theoretical genotype.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

REFERENCES

[1] Thayer, A.; Vargas, A.; Castellanos, A.; Lafon, C.; McCarl, B.; Roelke, D.; Winemiller, K.; Lacher, T. Integrating Agriculture and Ecosystems to Find Suitable Adaptations to Climate Change. Climate 2020, 8, 10.

[2] Nassani, A.A.; Awan, U.; Zaman, K.; Hyder, S.; Aldakhil, A.M.; Abro, M.M.Q. Management of natural resources and material pricing: Global evidence. Resour. Policy 2019, 64, 101500.

[3] Conrad, Z.; Niles, M.T.; Neher, D.A.; Roy, E.D.; Tichenor, N.E.; Jahns, L. Relationship between food waste, diet quality, and environmental sustainability. PLoS ONE 2018, 13.

[4] Benos, L.; Bechar, A.; Bochtis, D. Safety and ergonomics in human-robot interactive agricultural operations. Biosyst. Eng. 2020, 200, 55–72.

[5] Lampridi, M.; Sørensen, C.; Bochtis, D. Agricultural Sustainability: A Review of Concepts and Methods. Sustainability 2019, 11, 5120.

[6] Torkamaneh, D., Boyle, B., and Belzile, F. (2018). ‘‘Efficient genome-wide genotyping strategies and data integration in crop plants.’’ TAG. Theoretical and applied genetics. Theor. Appl. Genet. 131, 499–511.

[7] Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., Wang, J., and Fan, J. (2019). Crop phenomics: current status and Perspectives. Front. Plant Sci. 10, 71.

[8] Onishi Y , Yoshida T , Kurita H , Fukao T , Arihara H , Iwai A . An auto- mated fruit harvesting robot by using deep learning. Robomech J 2019; 6:13 https://doi.org/10.1186/s40648-019-0141-2.

[9] Vieira, S.; Lopez Pinaya, W.H.; Mechelli, A. Introduction to Machine Learning; Mechelli, A., Vieira, S.B.T.-M.L., Eds.; Academic Press: Cambridge, MA, USA, 2020; Chapter 1; pp. 1–20. ISBN 978-0-12-815739-8.

[10] Kenga R, Tenkouano A, Gupta S, Alabi S. Genetic and phenotypic association between yield components in hybrid sorghum (Sorghum bicolor (L.) Moench) populations. Euphytica. 2006; 150(3):319–26.

[11] Robbins MD, Staub JE. Comparative analysis of marker-assisted and phenotypic selection for yield components in cucumber. Theoretical and applied genetics. 2009; 119(4):621–34. https://doi.org/10. 1007/s00122-009-1072-8 PMID: 19484431

[12] Richards R. Selectable traits to increase crop photosynthesis and yield of grain crops. Journal of experimental botany. 2000; 51(suppl_1):447–58. https://doi.org/10.1093/jexbot/51.suppl_1.447 PMID: 10938853

[13] Specht J, Hume D, Kumudini S. Soybean yield potential—a genetic and physiological perspective. Crop science. 1999; 39(6):1560–70.

[14] Kumudini S, Hume DJ, Chu G. Genetic improvement in short season soybeans. Crop science. 2001; 41(2):391–8.

[15] Xavier A, Rainey KM. Quantitative Genomic Dissection of Soybean Yield Components. G3: Genes, Genomes, Genetics. 2020; 10(2):665–75.

[16] Sah R, Chakraborty M, Prasad K, Pandit M, Tudu V, Chakravarty M, et al. Impact of water deficit stress in maize: Phenology and yield components. Scientific reports. 2020; 10(1):1–15. https://doi.org/10. 1038/s41598-019-56847-4 PMID: 31913322

[17] Majhi PK, Mogali SC, Abhisheka L. Genetic variability, heritability, genetic advance and correlation studies for seed yield and yield components in early segregating lines (F3) of greengram [Vigna radiate (L.) Wilczek]. International Journal of Chemical Studies. 2020; 8(4):1283–8.

[18] Jiang Y, Lindsay DL, Davis AR, Wang Z, MacLean DE, Warkentin TD, et al. Impact of heat stress on pod-based yield components in field pea (Pisum sativum L.). Journal of Agronomy and Crop Science. 2020; 206(1):76–89.

[19] Pisner, D.A.; Schnyer, D.M. Support Vector Machine; Mechelli, A., Vieira, S.B.T.-M.L., Eds.; Academic Press: Cambridge, MA, USA, 2020; Chapter 6; pp. 101–121. ISBN 978-0-12-815739-8.

[20] Siegmann B, Jarmer T. Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. International journal of remote sensing. 2015; 36(18):4519–34.

[21] Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. Sensors **2018**, 18, 2674.

[22] Vieira, S.; Lopez Pinaya, W.H.; Mechelli, A. Introduction to Machine Learning; Mechelli, A., Vieira, S.B.T.-M.L., Eds.; AcademicPress: Cambridge, MA, USA, 2020; Chapter 1; pp. 1–20. ISBN 978-0-12-815739-8.

[23] Asefpour Vakilian, K. Machine learning improves our knowledge about miRNA functions towards plant abiotic stresses. Sci. Rep. **2020**, 10, 3041.

[24] Domingos, P. A few useful things to know about machine learning. Commun. ACM **2012**, 55, 78–87.

[25] Iniyan, S.; Jebakumar, R.; Mangalraj, P.; Mohit, M.; Nanda, A. Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks. In Artificial Intelligence and Evolutionary Computations in Engineering Systems; Advances in Intelligent Systems and Computing; Dash, S., Lakshmi, C., Das, S., Panigrahi, B., Eds.; Springer: Singapore, 2020; pp. 15–27.

[26] Alvarez, R. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. Eur. J. Agron. 2009, 30, 70–77.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 10, Issue 12, December 2023

DOI: 10.17148/IARJSET.2023.101224

[27] Azevedo, A.M.; Andrade Júnior, V.C.D.; Pedrosa, C.E.; Oliveira, C.M.D.; Dornas, M.F.S.; Cruz, C.D.; Valadares, N.R. Application of artificial neural networks in indirect selection: A case study on the breeding of lettuce. Bragantia **2015**, 74, 387–393.

[28] Gu, J.;Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.;Wang, X.;Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. Pattern Recognit. **2018**, 77, 354–377.

[29] Hesami, M.; Naderi, R.; Tohidfar, M.; Yoosefzadeh-Najafabadi, M. Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: E_ect of plant growth regulators on somatic embryogenesis of chrysanthemum, as a case study. Plant Methods **2020**, 16, 112.

[30] P. Chen, L. I. Shao-Kun, K.-R. Wang et al., "Spectrum character of cotton canopy infected with verticillium wilt and applications," *Agricultural Sciences in China*, vol. 7, no. 5, pp. 561–569, 2007.

[31] B. B. Kalbande and A. S. Patil, "Plant tissue culture independent agro bacterium tumefaciens medicated in-plant transformation strategy for upland cotton (*Gossypium hirsutum*)," *Journal of Genetic Engineering and Technology*, vol. 14, no. 1, pp. 9–18, 2016.

[32] Papageorgiou, E.I.; Aggelopoulou, K.; Gemtos, T.A.; Nanos, G.D. Development and Evaluation of a Fuzzy Inference System and a Neuro-Fuzzy Inference System for Grading Apple Quality. Appl. Artif. Intell. **2018**, 32, 253–280.

[33] A. M. Turing, ``Computing machinery and intelligence,'' in *Parsing the Turing Test*. Dordrecht, The Netherlands: Springer, 2009, pp. 23_65.

[34] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. Sensors. 2018; 18(8):2674. https://doi.org/10.3390/s18082674 PMID: 30110960

[35] Yoosefzadeh-Najafabadi M, Earl HJ, Tulpan D, Sulik J, Eskandari M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield from Hyperspectral Reflectance in Soybean. Frontiers in Plant Science. 2021; 11(2169). https://doi.org/10.3389/fpls.2020.624273. PMID: 33510761.

[36] Mihai Oltean . Fruits 360 dataset. Mendeley Data 2018;V1 10.17632/rp73yg93n8.1 .

[37] prabira Kumar sethy . Indian Fruits-40. Mendeley Data 2020;V1 10.17632/bg3js4z2xt.1 .

[38] Meshram VA , Thanomliang K , Ruangkan S , Chumchu P , Patil K . FruitsGB: top indian fruits with quality. IEEE Dataport 2020 https://dx.doi.org/10.21227/gzkn-f379.

[39] Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorol. Soc. Jpn. Ser. II*. 93, 5–48. doi: 10.2151/jmsj.2015-001.

[40] Bebber, D. P., Castillo, A. D., and Gurr, S. J. (2016). Modeling coffee leaf rust in Columbia with climate reanalysis data. *Philos. Trans. R. Soc. B* 371:20150458. doi: 10.1098/rstb.2015.0458

[41] Chandra, M.A.; Bedi, S.S. Survey on SVM and their application in image classification. Int. J. Inf. Technol. **2018**.

[42] Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. Sensors **2018**, 18, 2674.

[43] Shaik, R., and Ramakrishna, W. (2014). Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant Physiol. 164, 481–495, https://doi.org/ 10.1104/pp.113.225862.

[44] Sadiq, R.; Rodriguez, M.J.; Mian, H.R. Empirical Models to Predict Disinfection By-Products (DBPs) in Drinking Water: An Updated Review, 2nd ed.; Nriagu, J.B.T.-E., Ed.; Elsevier: Oxford, UK, 2019; pp. 324–338. ISBN 978-0-444-63952-3.

[45] De Oliveira, M.A.; Monteiro, A.V.; Vieira Filho, J. A New Structural Health Monitoring Strategy Based on PZT Sensors and Convolutional Neural Network. Sensors **2018**, 18, 2955.

[46] Wilson RF. The role of genomics and biotechnology in achieving global food security for high-oleic vegetable oil. Journal of Oleo science. 2012;61(7):357–67. pmid:22790166.

[47] Ramasubramanian V, Beavis WD. Factors affecting Response to Recurrent Genomic Selection in Soybeans. BioRxiv. 2020.

[48]Rebetzke G, Jimenez-Berni J, Fischer R, Deery D, Smith D. High-throughput phenotyping to enhance the use of crop genetic resources. Plant Science. 2019;282:40–8. pmid:31003610

[49] Yuan J, Njiti V, Meksem K, Iqbal M, Triwitayakorn K, Kassem MA, et al. Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop science. 2002;42(1):271–7. pmid:11756285

[50] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010;327(5967):818–22. pmid:20150489

[51] Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. Trends in plant science. 2014;19(1):52–61. pmid:24139902

[52] Qiu R, Wei S, Zhang M, Li H, Sun H, Liu G, et al. Sensors for measuring plant phenotyping: A review. International Journal of Agricultural and Biological Engineering. 2018;11(2):1–17.

[53] Robbins MD, Staub JE. Comparative analysis of marker-assisted and phenotypic selection for yield components in cucumber. Theoretical and applied genetics. 2009;119(4):621–34. pmid:19484431

[54] Richards R. Selectable traits to increase crop photosynthesis and yield of grain crops. Journal of experimental botany. 2000;51(suppl_1):447–58. pmid:10938853

[55] Kumudini S, Hume DJ, Chu G. Genetic improvement in short season soybeans. Crop science. 2001;41(2):391–8.