Article

Predicting Heritability of Oil Palm Breeding Using Phenotypic Traits and Machine Learning

Najihah Ahmad Latif ¹, Fatini Nadhirah Mohd Nain ¹ , Nurul Hashimah Ahamed Hassain Malim ^{1,*} ,
Rosni Abdullah ¹ , Muhammad Farid Abdul Rahim ², Mohd Nasruddin Mohamad ²
and Nurul Syafika Mohamad Fauzi ²

¹ School of Computer Sciences, Universiti Sains Malaysia (USM), Gelugor 11800, Pulau Pinang, Malaysia; najihahlatif97@student.usm.my (N.A.L.); fatininadhirahnain@student.usm.my (F.N.M.N.); rosni@usm.my (R.A.)

² FGV Research and Development (R&D) Sdn Bhd, Unit Biak Baka Sawit, Pusat Penyelidikan Pertanian Tun Razak, Jengka 26400, Pahang, Malaysia; mfarid.ar@fgvholdings.com (M.F.A.R.); mnasruddin.m@fgvholdings.com (M.N.M.); syafika.mf@fgvholdings.com (N.S.M.F.)

* Correspondence: nurulhashimah@usm.my

Abstract: Oil palm is one of the main crops grown to help achieve sustainability in Malaysia. The selection of the best breeds will produce quality crops and increase crop yields. This study aimed to examine machine learning (ML) in oil palm breeding (OPB) using factors other than genetic data. A new conceptual framework to adopt the ML in OPB will be presented at the end of this paper. At first, data types, phenotype traits, current ML models, and evaluation technique will be identified through a literature survey. This study found that the phenotype and genotype data are widely used in oil palm breeding programs. The average bunch weight, bunch number, and fresh fruit bunch are the most important characteristics that can influence the genetic improvement of progenies. Although machine learning approaches have been applied to increase the productivity of the crop, most studies focus on molecular markers or genotypes for plant breeding, rather than on phenotype. Theoretically, the use of phenotypic data related to offspring should predict high breeding values by using ML. Therefore, a new ML conceptual framework to study the phenotype and progeny data of oil palm breeds will be discussed in relation to achieving the Sustainable Development Goals (SDGs).

Keywords: oil palm breeding; phenotype; machine learning; framework; sustainable

Citation: Ahmad Latif, N.; Mohd Nain, F.N.; Ahamed Hassain Malim, N.H.; Abdullah, R.; Abdul Rahim, M.F.; Mohamad, M.N.; Mohamad Fauzi, N.S. Predicting Heritability of Oil Palm Breeding Using Phenotypic Traits and Machine Learning.

Sustainability **2021**, *13*, 12613.

<https://doi.org/10.3390/su132212613>

Academic Editor: Amir Mosavi

Received: 3 August 2021

Accepted: 26 October 2021

Published: 15 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Oil palm is one of the main plantation crops grown to help the economy and maintain sustainability in Malaysia. Sustainable palm oil is in line with pre-determined global standards, which is better for the environment than unsustainable palm oil, which involves deforestation, resulting in landslides, floods, and other disasters. The crop is vulnerable to environmental changes, weather, temperature, and insect and disease attacks during its 25-year lifespan. Palm oil products include cooking ingredients, cosmetics, detergents, food and beverages, personal care products, and more. Malaysia and Indonesia are among the largest exporters and producers of palm oil globally, accounting for 27% of exported trade and 11% of world oil and fat production. Currently, 2.13 tons of palm kernel oil and 17.73 million tons of palm oil are produced from 4.49 million hectares of land planted with oil palm in Malaysia [1]. The industry employs more than half a million people and provides livelihoods for approximately 1 million people to run the industry. This contrasts with annuals and perennials that have a shorter lifespan and years of observation. Therefore, selecting oil palm breeding for high-yielding and stress-tolerant crops is necessary for avoiding huge losses and maintaining sustainability. The maintained high yield and sustainability emphasize benefit all parties for food security, which is the basis of livelihoods, without depleting natural resources and human energy [2].

The introduction of artificial intelligence (AI) in computer science with the advancement of agricultural technology, including robotics, drone technology, cameras, censored machine technology, and geographical information systems, has been useful in capturing data. Phenotypic data such as weather, soil, humidity, historical crop performance, and temperature can be analyzed using machine learning (ML). The analysis can predict the optimal date to plant or harvest crops and can analyze which seeds will be suitable. Moreover, ML can make the decision on where the seeds are to be planted, when it is time to reduce the use of fertilizers and water, etc. With AI technology, worker safety can be improved, the impact on natural ecosystems can be reduced, food prices can decrease, and a large population should have access to the food [3], achieving sustainability goals.

In recent years, many researchers have studied oil palm breeding by contributing different and novel solutions on how to increase the yield of oil palm crops. This study includes combining the best crosses of progenies to obtain quality planting material. This study has reviewed various papers in plant breeding research, such as Rahaman in [4]. Rahaman conducted a review of data phenotypes on plant growth, and Dijk [5] also published a review on machine learning concerning plant breeding. Both papers give the reason for this study to conduct further research on phenotypes using machine learning. This study may provide different results for molecular studies that will affect the overall plant breeding program, where phenotypes are closely interrelated in molecular or genetic studies. However, there is no specific study of ML application on OPB specifically. This study intends to fill this research gap by identifying, classifying, and synthesizing related studies in this domain.

The main aim of this study is to examine the use of ML on OPB using factors other than genetic data. This paper also includes knowledge regarding the dataset required for the survival of plants, especially concerning new techniques of variations in OPB, phenotypic traits, diversity, and machine learning [6]. The paper is structured as follows: Section 2 explains materials and methods; Section 3 provides a compilation of ML techniques applied to oil palm crops; Section 4 discusses the framework of the study; and lastly, this paper is concluded in Section 5.

2. Materials and Methods

Many research innovations in breeding use machine learning, such as for maize [7] or corn [8], soybean [9], oil palm [10–13], wheat [14], yeast, and rice, [15] in order to understand the basic, most essential traits, and the flow of the study to fill the knowledge gap in the improvement of oil palm breeding studies. Therefore, this study aims to examine machine learning to analyze phenotypic data related to oil palm progenies data, which is likely to give different results from previous studies focused on genotypic data.

This study is conducted in a conceptual framework in which the concept of breeding will adapt using machine learning. Many requirements must be emphasized, such as data selection, essential traits, techniques, and even the type of evaluation, which presents a set of research questions. In this study, we will elaborate one by one the questions that are closely related and relevant for this study.

2.1. Understanding Oil Palm Breeding

The oil palms, *Dura*, *Pisifera*, and *Tenera*, are categorized into three breed groups. *Tenera* (T) is a hybrid seed with commercial plants that cross between the barren female seed named *Pisifera* (P) and the male seed called *Dura* (D) [16]. The difference between *Dura* and *Pisifera* is that *Dura* has a thick shell (sh+) while *Pisifera* does not have a shell (sh−). In general, *Tenera* and *Dura* have no kernel difference in fruit, oil to wet mesocarp, and fruit to bunch ratio. However, *Tenera* has a higher proportion of oil to bunches, as mesocarp to fruit is higher, while shell to fruit is lower [17].

Fruit traits can be categorized into three types: *Dura* (sh+sh+) is homozygous for the major gene, with a small proportion of oil-bearing mesocarp, and characterized by the production of large fruits with a thick shell. While *Pisifera* (sh−sh−) is not grown

as a commercial crop due to shell-less but essentially female-sterile palms. Lastly, *Tenera* (sh+sh−) has a larger proportion of oil-bearing mesocarp than *Dura* due to smaller kernels and produces smaller fruits with a thinner shell [18].

Every good tree is produced through the best breeding program. Plant breeding is the manipulation of plant species to create desired phenotypes and genotypes for a particular purpose. This manipulation involves genetic engineering, controlled pollination, or both. Subsequently, at the end of the study, the selection of artificial offspring will be made. Currently, studies focus on molecular markers or genotypes for plant breeding purposes but lack research on phenotypes. Breeding studies can be improved by combining phenotypes and genotypes to reduce time, space, and energy in order to produce new cultivars or populations [19]. In the past decade, research has stated that phenotype selection within crosses is not efficient, due to the requirements of the evaluation phase in field trials to identify high-yielding clones. However, this phase had been achieved by several institutions: Felda, ASD, and the Center for International Research Cooperation (CI-RAD). Therefore, this study needs to be reviewed, since Felda Global Venture (FGV) obtained enough phenotypic data for further research [20]. Nonetheless, this study aims to gain relevant knowledge in this domain.

Plant breeding, in general, is to change the nature of the plant to produce the desired traits to improve the nutritional quality of products for animals and humans [21,22]. Plant breeding is practiced all over the world and includes oil palm. The oil palm breeding system is female and male inflorescences with the same palm size, overlapping different cycles. Male flowering occurs when the condition of the *Dura*'s palm is pressed. However, *Pisifera* palms produce female inflorescences in good growing condition but are usually void because *Pisifera* does not produce oil to the bunch [19]. The importance of oil palm breeding is to improve the quality of palm oil, choose drought tolerance [23], enable the expansion of planting areas, extend the economic lifespan of crops, reduce vertical growth, and maximize oil palm and yields [24].

For oil palm breeding, there are several breeding schemes in Table 1 for which hybrid pollination is conducted to produce the desired crop.

Table 1. OPB scheme.

Breeding Scheme	Purpose	Crosses
[24–28] Reciprocal Recurrent Selection (RRS)	<ul style="list-style-type: none"> Exploits the heterosis caused by the reproductive depression between the crosses performed between specific origins and constraints. Populations of the <i>Dura</i> and <i>Pisifera</i> types as a starting point for crossbreeding and progeny tests in which the best parents are selected and <i>Tenera</i> type seeds are produced. The hybrid improvement phase refines specific combinations, while the recombination phase involves circulation that maintains genetic variability for long-term genetic improvement. Two complementary population groups, A and B, use strict criteria to examine elite trees as a basis for interbreeding. Integrated at different levels into the parent RRS consolidated to varying levels into the RRS breeding scheme after three selections. 	<i>Deli Dura</i> and <i>Pisifera</i> (DxP) <i>Deli Dura</i> and <i>Tenera</i> (DxT) <i>Deli Dura</i> and <i>Deli Dura</i> (DxD) <i>Tenera</i> and <i>Tenera</i> (TxT)

Table 1. Cont.

Breeding Scheme	Purpose	Crosses
[25,27–29] Modified Recurrent Selection (MRS)	<ul style="list-style-type: none"> Parents select for commercial hybrid seed production palms and further breeding based on (FIPS). Involves crossing between selected parents. Selected offspring can carry over to the next cycle. New introductions from other populations to increase genetic variability. Has proven that oil yields have increased. 	Deli Dura and <i>Pisifera</i> (DxP) <i>Tenera</i> and <i>Pisifera</i> (TxP) <i>Pisifera</i> and <i>Pisifera</i> (PxP) Deli Dura and Deli Dura (DxD)
[19,20] Family and Individual Palm Performances (FIPS)	<ul style="list-style-type: none"> To combine common, family, and phenotypic values. Selection of T for subsequent P breeding based on T performance in the T × T family followed by D × P progeny testing. 	Deli Dura and <i>Pisifera</i> (DxP) <i>Tenera</i> and <i>Pisifera</i> (TxP) <i>Tenera</i> and <i>Tenera</i> (TxT)
[6,30] Reciprocal Recurrent Genomic Selection (RRGS)	<ul style="list-style-type: none"> Decreases the generation interval by increasing the intensity of selection. Genome selection occurs among parental populations to identify hybrids with the highest genetic values and conducting the offspring test before making final choices on all traits. 	LaMé × Sibiti/Yangambi Yangambi × Nigeria Deli × Angola Lisombe Kinshasa

In addition to the breeding scheme, Cros [31] further explains the history of the RRS scheme established in the 1950s, which was the basis for the genetic improvement of oil palm. It depends on two populations, namely, Deli (Asia) and mixed populations (Africa), used as the parents of commercial hybrids. Phenotypically, these two groups have different characteristics or traits in bunch production, and the number of founders is then passed on to selection, genetic, and breeding generations. The mass selection was also applied to both populations regarding the nature of interest, intensity, and more. Mass selection is conventional selection, in which mixed seeds are collected to form new varieties of breeds. Usually, selection is based on its phenotype. At the same time, modern mass selection compares their progenies after harvesting the best plants separately.

2.2. Research Questions

To gain a more detailed knowledge of this topic, the Research Questions (RQ) help us clarify the OPB using machine learning, determine the direction of future research, and identify obstacles in the field. The RQs are listed in Table 2 to guide the process.

Table 2. Research questions.

RQ #	Research Question	Motivation
RQ 1	What kind of data involving the OPB program have been applied to machine learning?	To highlight the type of datasets that are widely used in oil palm breeding programs.
RQ 2	What are the most influenced phenotypic traits of OPB?	To highlight the characteristics, features, or traits which influenced the genetic improvement of progenies.
RQ 3	How can machine learning benefit from the phenotype, environment, and other than genetic data?	To highlight machine learning approaches that had been applied to crops and the impact of that crop.
RQ 4	What are the evaluation metrics used in the model?	To highlight the most used evaluation metrics in the precision model and the results from it.

2.3. PRISMA

According to PRISMA guidelines, this section provides a way to obtain papers relevant to this study [32]. Four research databases were used in this study: Google Scholar, Scopus,

Web of Science, and DOAJ. The databases are the primary data source of potentially relevant studies affecting this research.

Relevant search terms are essential to finding adequate and appropriate studies. Applicable search terms have suggested a PICO structure consisting of population, interventions, context, and outcome in this case. A generic search string using a PICO structure is essential to maintain search consistency across multiple databases. Therefore, the following search strings to perform automatic searches in a database such as: ((Oil Palm Breeding OR Plant Breeding) AND (Artificial Intelligence OR Machine Learning OR Deep Learning) AND (Phenotype OR Genotype OR Environment)). The following search terms have critical terms identified from the Oil Palm Breeding research domain:

- Population: Oil Palm Breeding and Phenotype;
- Intervention: Machine Learning and Deep Learning;
- Context: Artificial Intelligence;
- Outcome: Identification and Prediction.

Searching for candidate papers through several stages in Figure 1. includes paper identification, screening, and contents, resulting in 36 articles. Exceptions to research papers are made based on three widely used selection policies: namely, title, abstract, and full paper access.

This section lists the candidate criteria of entry papers as the study focuses on datasets, models used, techniques, and metric evaluations for any crop breeding. This selection is due to the constraints of phenotype data that are difficult to find. Some papers apply bioinformatic techniques that use commonly used data and genotypes in the last five years. The articles that have been saved may be helpful in the construction work of the prediction model to be applied for future oil palm breeding. Meanwhile, articles that do not have complete content were excluded, whereas papers after 2011 onwards are included due to oil palm study factors that are rarely studied because they have a long crop life to bear fruit.

In the first stage, 4189 papers were counted through string searches from the research databases, and 2749 were removed based on titles and duplicates among other research databases. A total of 1444 papers were screened based on abstracts in the second stage, and 662 articles were eliminated. A total of 778 reports were issued, but only 129 were fully accessed and extracted into Excel for analysis with data that may be important in this study to facilitate information retrieval to improve the understanding of context. Finally, a total of 36 papers were used in this research to create a comparison on the performance of techniques used in crop breeding.

The frequency distribution of selected papers over the past 11 years is shown in Figure 2. A timespan of 11 years was chosen because of the long lifespan of oil palm, causing this field to be rarely studied except by recognized research bodies. The graph shows an increase due to the technological shift from conventional to modern crops. The paper in 2011 was selected because it looks at the phenotypic characteristics of interest in the study, which identifies the technological change in processing and analyzing phenotypic, genotypic, and environmental data.

This study has used the Cochrane Collaboration's tool to conduct a biased risk analysis of traits of interest through phenotypic observations in oil palm breeding studies. For the results of the investigation, refer to Figure 3. Most papers focus more on yield traits, fresh fruit bunch traits, oil traits, and stress resistance traits. There are three biased indicators which are unclear, high, and low. Ambiguous decisions usually occur if there are insufficient details about what happened in the trial. Bias refers to a magnitude sufficient to affect the conclusion or outcome of a problem significantly. Finally, low bias will be classified because the information provided covers almost everything within the scope.

Figure 1. PRISMA Taxonomy.

Figure 2. Number of selected papers per year.

Figure 3. Risk of bias for traits of interest.

3. Results

This section presents the study results to answer the research questions (RQ) which are presented in Table 2.

3.1. RQ1. What Kind of Data Involving the OPB Program Has Been Applied to ML?

Several types of data other than genetics can be analyzed to make space for further study. Among the data involved in OPB are phenotypic, environmental, historical, and management. In addition, genetic data influenced by progeny, climate, humidity, soil, and temperature are significant for further studies to be conducted using ML [4]. There are advantages from it that might give a different result to change the plant breeders' perspective.

Phenotypes are physical traits that result from the effects of interactions between the genetics of an organism (genotype) [33,34] and the environment [15,35]. Physical traits are observed and are key features such as maturity, resistance to pests, or other observable traits [35]. Phenotypic data were obtained for several years before biochemical analysis was made by recording results such as mesocarp mass, the number of bunches produced per year, and mesocarp oil content [36]. Selection based on phenotype is called mass selection, i.e., selecting the desired type to include better adaptation. This mass selection includes quantitative phenotypes related to complex yield traits for annual crops, resistance to pests, and drought [15]. Combination seeds based on phenotypic selection will undoubtedly be helpful in the future [35,36].

Various imaging methodologies use phenotyping platforms to obtain non-environmentally damaging yield phenotypic data and quantitative studies of complex traits such as growth, tolerance, resistance, architecture, physiology, and basis [37,38]. The selection of stress-resistant and high-yielding crops is necessary to ensure production in line with population growth. Rahaman states there is an increase in production when genotype and phenotype relationships are established [39]. Thus, phenotype and genotype are equally crucial in the production of traits and genes. Phenotypes are an obstacle in genetic analysis and prediction because the tools do not achieve the goal of sustainability, which is expensive, time-consuming, requires a big workforce, and the destruction of crops that occurs at certain times. The purpose of current crop phenotypes at all levels of biological organization is to increase phenotypic inferential throughput, accuracy, and precision of phenotype while achieving the goal of sustainability for profit without destroying the environment. It reduces workforces through mechanization, remote sensing via drones, improving data integration through big data technologies, and experimental design [4].

3.2. RQ2. What Are the Most Influenced Phenotypic Traits of OPB?

There are many phenotypic traits for each data type and each trait, with more components involved, such as the bunch component, yield component, fruit component, and other components. A phenotypic trait is a trait that can be seen clearly, even measurable,

and is also a visual gene expression. An example of a phenotypic trait is a particular fruit's ripeness. The underlying genes that make up the genotype determine fruit ripeness, but the fruit ripeness observed is phenotypic. This is also applied to the whole component of an oil palm. Plant traits in biology refer to the physiology of plant character. Planting materials that help high and sustainable oil yields are designed to possess useful traits that contribute directly to a crop's economic products. For example, plant-pathogen interactions, ability to withstand stress, response to fertilizers, nutrient uptake, ease of harvesting, and of course, high Fresh Fruit Bunch (FFB) [2,11,26] are essential for breeding for seed production [40].

Many base populations of plants are cultivated, and many of the most desired phenotypes are grown as individual crops. Only selected plants were mated randomly to produce better new populations. Among the desired phenotypic traits are as follows in Table 3.

Table 3. Phenotypic traits.

Components	Types of Traits	Traits
Oil Yield Component	Oil Quality Traits	Low lipase, high stearic and high oleic [19], low free fatty acids content and high iodine value [24], oil quality [13,17,20,23,41,42], quality, and yield [27].
	Oil Yield Related Traits	Oil weight-to-dry mesocarp (O/DM) [41], total oil yield per palm (O/P), mesocarp-to-fruit weight (M/F), kernel-to-fruit weight (K/F), shell-to-fruit weight (S/F), and fruit-to-bunch (F/B) [19,20,24], oil to bunch [25,42], oil content [42] in mesocarp [12,43] and oil yield [12,14,20,23,44].
	Oil Palm Traits	FFB [12], Yield [44] [17,20,42] [45], E. Guineensis for traits such as height and high carotene [9,13].
Bunch Component	Bunch Production Traits	Annual cumulative bunch number (BN) and annual average bunch weight (ABW, in kg) [46] and annual cumulative bunch production (FFB, in kg) [30], Bunch production [47], Bunch index [25,29,45,46].
	Bunch Quality Traits	Pulp-to-fruit ratio (PF, in %), fruit-to-bunch ratio (FB, in %), oil extraction rate (OER, in %), and oil-to-pulp ratio (OP, in %) [30].
Fruit Component	Fruit Traits	The fruit traits, including Oil-to-Dry Mesocarp (O/DM), Mesocarp-to-Fruit (M/F), Shell-to-Fruit (S/F), and Kernel-to-Fruit (K/F) [12,48]. Oil-bearing mesocarp [18], Shell-to-Fruit (S/F%) [47], Shell Thickness [39], Fruit traits such as high mesocarp to fruit, reasonable stalk length, good fruit set, and low parthenocarp [25], fruit form, fruit quality, and fruit ripeness [49].
	Agronomic/Agricultural Traits	Short stems, midribs, high and fast production for commercial production [50], Efficiency in nutrient utilization, stress tolerance, and disease resistance are under the control of many genes [27], Nutrient content and plant height [45], Shell thickness [51], fruit color [52], or the mantled soma clonal variation [53,54] together with gene networks for oil biosynthesis [55], (1) drought and cold tolerance, (2) in vitro regeneration potential, (3) lipase activity, (4) carotene and vitamin E contents, (5) FA composition and iodine value, (6) Fusarium wilt disease tolerance, (7) fruit shell thickness, (8) fruit and kernel size, i.e., fatty acid (FA) profile, (9) total and vegetative dry matter production, (10) fresh fruit bunch and crude palm oil yield, (11) bunch number, weight and production, i.e., oil yield, (12) increment in the growth in height, i.e., breeding for shorter palms and (13) leaf petiole, rachis length, i.e., breeding for the compact palms [56,57].
Other Components	Physiological/Vegetative Traits	Total dry matter [25], plant architecture [25], vegetative parameters such as leaf length, number of leaves per plant, and plant height [40,58].
	Environmental Stress Resistance/Adaptation Traits	Dwarfness [13,16]. Disease resistance [14,17,20,23,26,46], Environmental stresses (storms, global warming, rainfall fluctuations, and extreme temperature) [44].

3.3. RQ3. How Can Machine Learning Benefit from the Phenotype, Environment, and Other Than Genetic Data?

The types of Machine Learning (ML) algorithms that may lead to future employment in the OPB domain are divided into supervised and unsupervised learning. Both ML types are differentiated with data that have a label or no label. Identification and recognition of unlabeled data will be made through clustering. Data with labels are easier to classify by using prediction algorithms. In the past decade, the most widely applied ML algorithms towards oil palm programs are Support Vector Machines (SVM), Artificial Neural Network

(ANN), Random Forest (RF), Regression, Classification and Regression Tree (CART), and Convolutional Neural Network (CNN) [59].

Singh et al., in [60], advocate using ML in plant phenotypes for image data to overcome the morphological features of plants such as uncontrolled environment, chaos, and noise. Coping with phenotypic and environmental variations requires considerable effort in method development. ML may be able to model genotype–environment interactions to link genotypes to phenotypes so that decomposition of complex traits, such as yield or growth, can be made [7,49].

Chapman et al. [10] predict the yield functions using machine learning (Bayesian Network Learning compared with Artificial Neural Network) to extract complex data. The research explains that machine learning algorithms have evolved in oil palm research, including the identification of recording yield errors [61], female classification inflorescence [11], oil palm plantations detection [13], fruit maturation [62], forecast fruit ripeness [63], and genome selection [48] in oil palm breeding programs. Despite the availability of big data, machine learning applications at the plantation level have yet to be explored [10]. These indicate that ML has a big potential to be applied in the oil palm breeding program.

3.4. RQ4. What Are the Evaluation Metrics Used in the Model?

The evaluation metrics for the plant breeding prediction framework will determine whether the model is built efficiently or not. Every output that comes from the prediction framework must be justified whether it has a significant impact or not. For example, insignificant results will allow researchers to enhance the framework and optimize the breeding work.

Oil palm pollination requires a process of detection, inspection, accuracy, and validation in the field. Humans performed those processes manually with uncertainty, pollination latency, and erroneous assumptions. However, ML provides a non-destructive, efficient, and cost-effective solution for determining future pollination breeding levels. The selection of appropriate assessment metrics will contribute to obtaining optimal classification results during the training session.

The most widely used assessment metric for classification is the accuracy, precision, recall, and F1-score, while, for the regression, there are three evaluation metrics, which are the Root Mean Square Error (RMSE), R-Squared, and the Mean Absolute Error (MAE) [11,64].

4. Discussion

Table 4 tabulates a lot of useful information that helps to analyze data from various studies, including phenotype data, environmental data, heritability, and parental data. In the meantime, the table provides an overview of the appropriate methods to analyze the data. The table can anticipate the results of future studies so that the achievement of sustainability will continue.

Table 4. Literature review matrix.

Author	Purpose	Datasets/Features	Methods	Description	Results
[65]	To identify the relevant variables and search for the best hybrid for modeling to investigate the potential of oil palm productions.	Mature and immature area, rainfall, radiation, temperature, wind speed, evaporation, cloud cover, Index (API) calculations are ozone (O ₃), carbon monoxide (CO), nitrogen dioxide (NO ₂), sulfur dioxide (SO ₂), and particulate matter (PM).	Features selection and sensitivity test.	Genetic Algorithm/Correlation Analysis (GA/CA) was developed to select the quality of land, kinds of oil palm areas, climate, and air pollutant factors that affect oil palm production predictions.	Effect of climate change, air pollution, and quality of land of oil palm areas significantly contribute to land productivity.
[10]	Yield function prediction from three commercial oil palm estates.	Daily rainfall, basic site information, and soil survey data. Fruit yield (FFB), average bunch number per hectare (BUNCH-HA), and the average weight of fruit bunches (ABW).	Prediction of fruit yield by using machine learning.	Bayesian Network: Stage 1: To find the relationship between variables; Stage 2: To compute the probability of each parameter.	Bayesian networks can be used to successfully predict yield functions based on climate data, soil conditions, and management factors.
[63]	To predict grades of oil palm fresh fruit ripeness.	Neural Network parameters Architecture, learning algorithm, activation function, and repetition run. Genetic Algorithm parameters The number of variables input. The number of observations used for training and validation data. The number of bits for encoding one gene and one chromosome. Lower and upper bound interval. The number of chromosomes in the population. Crossover and Mutation probability. Maximum fitness threshold and the number of generations.	Multi-layer perceptron (MLP) Feed-Forward Network used for pattern classification. Genetic Algorithm (GA) for the optimization process. Genetic Algorithm Neural Network (GANN) to reduce the computational complexity and improve convergence speed.	MLP is used to receive independent computations and results to real numbers. Chromosomes as parameters that are an alternative to the optimal solution for the GA process. GANN was used to analyze data.	GANN is adequate for solving the multi-class classification problem.
[48]	To select individuals with the desired overall hybrid of breeding traits.	Shell-to-fruit ratio (S/F, %), kernel-to-fruit ratio (K/F, %), fruit per bunch (F/B, %), mesocarp-to-fruit ratio (M/F, %), oil per bunch (O/B, %) and oil per palm (O/P, kg/palm/year).	Ridge Regression Best Linear Unbiased Predictor (RR-BLUP), Bayes A (BA), C π (BC), Bayes Lasso (BL), and Bayes Ridge Regression (BRR)	RR-BLUP and all Bayesian methods assist in predicting the traits.	Heritability estimation correlated well with the conventional heritability of phenotypes. Slight increase in accuracy using RR-BLUP as compared to Bayesian methods in traits controlled.

Table 4. Cont.

Author	Purpose	Datasets/Features	Methods	Description	Results
[46]	To study the performance of biparental <i>Dura</i> × <i>Pisifera</i> (D × P) progenies and their parental genetic origins on yield components and fruit set.	Yield, parental, fresh fruit bunch quality, vegetative, physiological, bunch number, bunch weight traits, and nonsexual and functional traits.	Statistical Analysis (SAS 9.4). Analysis of variance (ANOVA). The mean and standard error (Stderr). Restricted Maximum Likelihood (REML)	SAS was used to analyze data. ANOVA calculated using a general linear model to solve the unequal distribution of progenies. Stderr was used to get the level of probability for the comparison of progenies. REML is used for variance components estimation.	ANOVA exhibited a highly significant effect for yield and bunch traits. FFB is lower than BNO, and ABW is a high variance in variance estimates for yield components.
[26]	To analyze the level of diversity expressed by several categorical variables using diversity index.	Genitors' population. Cross schemes of candidate genitors. The genitors, their parents, and grandparents.	Graphical and univariate method analysis.	Categorical variables and descriptors of parental diversity for classification based on origin, genealogy criteria, and production groups—indicators of diversity and indices to obtain the modality between classes of genitors.	Two genitors <i>Dura</i> Deli and <i>Tenera</i> Lame, slightly different for a given agronomical trait, can still be used in recombination for another categorical trait discrimination them.
[66]	To discover the optical and physical traits of the oil palm uniqueness.	Oil palm FFB physical, surface color information, grading standard, size, width, length, weight, shape, density, circumference, volume, and all related axial dimensions of oil palm FFB.	Determination of Grading Standard, Physical Properties, Size, Shape, Density Ratio, Porosity, Volume, Density, Mass, RGB Properties, Image Processing, Optical Sensor, Mean and Percentage of Intensity.	The process was evaluated using a multi-band portable, active optical sensor system comprising four spectral bands to detect oil palm FFB physical characteristics.	The advantage of the optical sensor is that it can identify and measure the characteristics of the palm FFB.
[67]	Identification of DNA markers associated with oil content traits for Marker-Assisted Selection (MAS).	Oil to bunch (O/B) and oil to dry mesocarp (O/DM).	ANOVA and Genotyping Analysis.	ANOVA and mean used to analyze oil content traits for phenotypic traits. Genotype analysis for palm progenies.	Statistical analysis of the phenotypic data showed that the two traits were normally distributed ($p > 0.05$), indicating polygenic variation.

Table 4. Cont.

Author	Purpose	Datasets/Features	Methods	Description	Results
[12]	To evaluate the effect of different marker systems and modeling methods for implementing Genomic Selection (GS).	Fruit-to-bunch (F/B), shell-to-fruit (S/F), kernel-to-fruit (K/F), mesocarp-to-fruit (M/F), oil per palm (O/P), and oil-to-dry mesocarp (O/DM).	Support Vector Machine (SVM), Ridge Regression Best Linear Unbiased Prediction (RR-BLUP), LASSO, Random Forest (RF), Bayesian A, B, and C π .	A modified chart function was used for phenotypic analysis from the library “Performance Analytics” under R version 3.0.0. RR-BLUP assumes that the marker effects of Bayes A and B allow different variance across different markers to obtain the probability of π . π at C π follows a uniform distribution. Bayesian Ridge Regression uses a Gaussian prior. Lasso uses a Double-Exponential prior. A 5-fold cross-validation was carried out for all methods across all traits. SVM, 10-fold cross-validation was used on the training set to estimate the optimal parameters. The values of the prediction properties of the confirmation set and the observed yield the Pearson Correlation Coefficient.	The performance of the machine learning method has a marginal advantage over (0.32) over other methods
[13]	To detect oil-palm plantation from remotely sensed imagery	Spectral Data (Image)	Artificial Neural Network (ANN), Maximum Likelihood (ML), Support Vector Machine (SVM), Classification and Regression Tree (CART), Random Forest Tree (RFT), Mislabelled Discarded (MD), Grey-Level Co-occurrence Matrix (GLCM), Multi-Layer Perceptron (MLP)	ANN, ML, SVM, CART, RFT, and MD were used for pixel classification. GLCM was used for texture analysis then classified by SVM. MLP was used as a classifier to obtain the combination of fractal dimension and local binary pattern (LBP). Gabor filter was used for Gabor texture descriptor. SVM, K-Means, RFT, and CART were used for the classification of texture extracted by the Gabor filter. References ground truth are used for oil palm plantation verification to obtain accuracy values.	RFT was preferred for the precision of oil-palm detection.

Table 4. Cont.

Author	Purpose	Datasets/Features	Methods	Description	Results
[68]	Progeny selection application for selection dwarf oil palm planting materials and high fresh fruit bunch (FFB).	Fresh Fruit Bunch (FFB), Bunch Number (BNO), Average Bunch Weight (ABW), Mean Fruit Weight (MFW), Mean Nut Weight (MNW), Mesocarp to Fruit ratio (MTF), Kernel to Fruit ratio (KTF), Shell to Fruit ratio (STF), Oil to Dry Mesocarp (OTDM), Fruit to Bunch ratio (FTB), Oil to Bunch ratio (OTB), Kernel to Bunch ratio (KTB), Oil to Fruit ratio (OTF), Oil Yield (OY), Kernel Yield (KY), Frond Production (FP), Petiole Cross Section (PCS), Rachis Length (RL), Leaflet Length (LL), Leaflet Width (LW), Leaflet Number (LN), Palm Height (HT), Leaf Area (LA), Leaf Area Index (LAI), Diameter of Palm Trunk (DIAM)	Statistical Analysis	ANOVA was used for bi-parental analysis. A general linear model (GLM) is used to handle missing data. Duncan's New Multiple Range Test (DNMRT) was used to separate the mean for mean comparison. The intra-class correlation coefficient was used for heritable estimation and components variance for all traits. Unweighted Pair Group Method with Arithmetic mean (UPGMA) was used for cluster analysis for progeny quantitative traits.	Clusters (I and II) are outstanding in terms of the bunch weight, oil to bunch, fruit size, and oil yield comprised of DP3, DP5, and DP8 that belong to Ulu Remis Deli <i>Dura</i> grant parent UR 515/316 and UR 515/3018.

Predicting the heritability of oil palm breeding using machine learning aims to embrace all relevant knowledge about it. This study produces a framework that provides knowledge of appropriate oil palm breeding programs conducted on data other than genetic data using ML. The problem is solved by elaborating the research question that helps in producing the framework. The framework in this study was produced due to the low heritability value, which resulted in the low breeding value in the Kwong [12] study. Several genetic studies (GS) were conducted to evaluate the effects of different marker systems using the ML model to implement GS in derivative families [12]. For accuracy, empirical estimates of GS were calculated through the value of genome breeding in oil palm (*Elaeis guineensis*) [31]. Kwong, in [12], has clarified the limitations that occur in ML, such as inconsistent feature results which require adjustments of different parameters. The offspring with low genomics depends on phenotypic variation as in the study. Kwong also stated that Kernel to Fruit (K/F) has the lowest genomics, making the most insufficient accuracy due to low K/F phenotypic variation in the families assessed. Thus, it is the reason this study was conducted to embrace all the knowledge that may be relevant to the problem. A part of the problem can be solved by using the ML framework, as illustrated in Figure 4. The use of statistics in heritability estimation result in low breeding value. The increase in heritability value can contribute to the increase in breeding value for genomic selection.

Figure 4. OPB ML framework.

Theoretically, metric values as phenotypic values for a particular population are a result of environmental factors, genetic factors, and environmental factors that interact with genetic factors. Phenotypic variance (V_p) is a variance observed and measured according to the nature of interest. For example, in this study on oil yield, then, the variation of oil yield observed among its descendants is V_p . These factors in the population that separate the quantitative traits contribute to the differences in those populations. The

total variance can be divided in the following Phenotypic Variation, (1) and Heritability Estimation, (2) [69]. Using the genomic selection [58] approach of previously conventional inheritance calculations is usually calculated from the deterioration of parental offspring regression [70]. For a full study of genomic selection, please go to the reference section [48].

$$V_p = V_g + V_e + V_{ge} \quad (1)$$

$$H^2 = V_g/V_p \quad (2)$$

V_p = number of population phenotypic variations;

V_g = genetic variation;

V_e = environmental contribution;

V_{ge} = variation in the interaction of genetic and environmental factors;

H^2 = Estimated heritability.

A high heritability value of 0.5 means that half of the differences between plant phenotypes are genetic. In contrast, a low value of 0.1 means that most of the differences are not genetic. The measurement of the rate of phenotypic variation is the result of genetic and heritability factors. (3) can predict heritability phenotypes. (4) is used to narrow population values [71]:

$$PH = T + h^2 (T^* - T) \quad (3)$$

$$T^* = \text{midparent value } [(T_f + T_m)/2] \quad (4)$$

PH = heritability phenotype;

T = minimum population;

h^2 = narrow-sense heritability.

Heritability can be estimated from the observed selection response (R) in proportion to the practical selection difference (S) in the artificial selection experiment. Heritability is essential for determining how a population will respond to selection. Usually, parents with phenotypic interests and values are selected from the baseline population. These parents crossed over the following (5) and (6) for a new population to be developed [69]:

$$S = \vec{T} - \vec{T}_s \quad (5)$$

$$R = \vec{T}' - \vec{T} \quad (6)$$

S = Selection Differential;

R = Selection Response.

$$R = h^2 S \quad (7)$$

The breeding (7) summarizes this relationship. The narrow-sense heritability estimates are illustrated by decreasing regression lines. The selection difference is between the mean of the baseline population and the minimum of selected parents. The selection response calculated how much profit could be made while mating the selected parents. Thus, the response to selection is performed with different choices. This multiplier of heritability is intended to predict the response to choose [69]. When a high genetic value can be obtained, a quality tree offspring can also be obtained, such as high yield, weather-resistance, and pest and wild animal resistance. This high genetic value can maintain the sustainability of crops and achieve high profitability. There is no felling of trees infected with the disease or damage from wild animals such as wild boars and elephants. Based on the calculation theory of convincing to perform breeding heritability prediction on oil palm crops using ML, the presence of ML will further facilitate calculating the estimate of heritability and be able to make predictions about it.

The first research question in predicting heritability using ML highlights data related to oil palm breeding other than genetic data. Knowledge of the relevant data can ease the burden early before the genomic selection process is carried out. Genomic selection involves the relationship between genetics and the environment for a particular offspring.

Findings of the Kwong [48] study, conventional phenotypic heritability such as Mesocarp to Fruit (M/F), Kernel to Fruit (K/F), Fruit to Bunch (F/B), and Oil to Bunch (O/B) correlated well with each other. Offspring with high genetic and phenotypic value will get the high breeding value that will produce quality breeds. Therefore, an ML framework can perform heritability prediction with high breeding value without involving genetic data, which is a conflict here: Can an approach that does not include genetic data be carried out? Heritability prediction results, if not involving genetics, will be very different from the molecular approach. The data collected or received must be pre-processed by solving problems related to missing values as well as noisy and invalid data, resulting in increased errors at the end of the study [30,71–73]. In addition, understanding the relationship between features and parameters is needed to see if they are interrelated by using relational analysis methods to compare values from different tables that represent the same thing. Data-driven methods are typically used to identify, classify, and analyze data from plants using ML techniques in plant breeding. Thus, machine learning techniques help solve complex problems with big data, such as using multiple data sources while providing higher predictive accuracy. In the past, the fresh fruit ratio (TBS), the oil-to-Bunch (OB) ratio, and the weight were the main parameters used as selection criteria in assessing offspring. Now, many other parameters are included in the selection of heritability parents, including FFB yield, oil quality, disease tolerance, male rescue sex ratio, and desired vegetative traits such as increased frond length. Each heritability iteration over many years was used in the calculations. The calculation is due to the unbalanced and inaccurate analysis of group components to estimate variance and heritability. Experimental designs that are analyzed in one section will produce unbalanced datasets, and biases will appear in the estimation of non-genetic effects. Typically, in real life, oil palm farmers use the entire result of an experimental design to calibrate a model that benefits parental relationships, cross-volume balance, and more [17]. On the other hand, ML requires a lot of time to process when data is complex and tuning parameters but always depends on the purpose of the study. Prediction models have provided an opportunity to be applied by extending and calibrating more to reduce the prediction error value [51,68,74–76].

The second research question is to highlight traits or characteristics that influence genetic improvement from offspring. The selection of traits in the data using the feature selection method is required to reduce the input in the process phase to improve the predictive model's performance for a framework. Phenotypic selection has always had a place in plant breeding research because its effectiveness depends on selecting the traits under consideration and minimizing environmental impacts. Phenotypic selection based on breeding value becomes more effective because it has an experimental and analytical design to estimate the environmental effects, inherited genetics, and parental breeding value. Based on experimental phenotypic data, year and location became the target environment for collecting and making a replicated heritability analysis (to obtain the variance component) [22]. The data available for processing can be stored for a certain period to enable future research work to be easily carried out. Indirectly, crop sustainability will be achieved because there is no need to collect samples in the field, such as cutting down trees or branches for data collection. Therefore, the data that has been processed will be divided between training and testing. With that, the data can learn through the training process. Data can be divided by data size. There are studies using a five-fold cross-validation method and some also use a 10-fold cross-validation method [4,11]. This will give different results. However, studies can make comparisons of differences between different folds.

The third research question is to highlight the use of the ML approach. ML was used to analyze the data using training data in the process phase. Each data point processed will generate a metric value for model performance observation. However, the crisis encountered is during the adaptation of heritability values to ML through metric values because it involves other values that influence phenotypic values. Technique classification, clustering, and regression can be used for data analysis using ML. In each technique,

many methods can be used for problem solving. For example, the breeding study using ML by Frouin et al. [77] uses regression techniques because its numerical data can obtain breeding values. It is challenging to consider the best ML algorithm for making predictions. However, ML-based regression algorithms, including ANN, SVR, and RF, are very effective for predicting palm oil prices and palm oil yields. Neither a single algorithm nor a mixture of algorithms should be investigated to increase the *Dura*'s ability of the predictive model. In addition, many papers present Linear Regression (LR) as a benchmark in many cases to confirm whether the implemented method is better than LR or not [64]. In the findings of Rashid [64], constraints will be faced by RF regression due to lack of data set training and will underestimate the average results. Therefore, increasing the number of observations with appropriate predictors for training is essential in each study to demonstrate its ability to make predictions and improve model performance. In addition, regression techniques can make predictions to achieve the objectives of this study. Among the frequently performed methods, Ridge Regression, Bayesian, RF, BLUP, SVM, LR, and LASSO are among the best prediction models [8,12,15,50]. Noise phenotype data were compared with the original data, and the proportion of noise data was added to explain the variance ratio. The accuracy results show a decline with best-performing Random forests (RF), followed by GBM, BLUP, SVM, and LASSO, with regression gaps behind [15].

Khan [59] discusses the regression analysis in ML to perform prediction and estimation in achieving objectives such as predicting oil palm growth, FFB yield, seasonal variation, palm oil price, oil quantity, harvest time, and seasonal effect on production. Khan states that ML is used less to predict and make descriptive analysis in the oil palm industry because ML is still early, and some problems must be addressed. One day ML will contribute and grow in the oil palm industry to solve more challenging issues in agriculture and plantations, especially oil palm [59,78]. From the point of view of this study, many predictions can be made using regression analysis. However, there are still gaps for estimation studies that use regression analysis in ML on oil palm breeding, such as heritability estimation that often uses statistical methods to obtain breeding values (BV) to aid in phenotypic variation. Frouin [77] applied ML in human studies. The study stated the problem of heritability estimation using an inferential statistical approach in genetics. Then, the heritability estimation problem is solved from a machine learning perspective using ridge regression and mixed models. Significantly, estimating the BV of offspring using ML can improve genetics in performing GS. Offspring with high genetics will produce high yields and accuracy to achieve the sustainability of oil palm crops.

The fourth research question aims to highlight the evaluation of appropriate metrics to use in the framework of this study. The evaluation of the metrics of each data point resulting from the training data is intended to obtain the accuracy, precision, and error of a model that occurs in the post-processing phase. When the results of the metric evaluation are carried out satisfactorily, the study should rerun the prediction model using the test data for the preparation of the new data prediction. When still unsatisfactory, the study should be repeated with feature tuning until satisfactory metrics are obtained. Techniques that can be used to evaluate metrics are Accuracy [10,12,50], Recall, Precision, and F1-Score [13]. Metric evaluation techniques can measure the performance of a prediction model. Prediction models that have high accuracy, high precision, and low error are more reliable. Regardless of the results received, there is still room for improvement in data selection, properties, and the appropriate algorithms. The metrics provide estimates of the error (RMSE, MBE, RRMSE) and the variance explained by the models (R^2) [8]. However, no evidence testing has been conducted based on this framework on oil palm breeding. This framework is based on a concept adapted from Rashid and Khan's study [59,64] on oil palm and ML.

This further research may be able to explain the relationship of genetics to oil palm breeding. The limitations of this research are as follows: no test evidence was obtained on heritability prediction because this study uses the concept of machine learning black box testing. It refers to testing without algorithm details and features to ensure the quality

of the constructed framework. This study has given some overview of the data and the traits produced by the method to analyze the data and to evaluate the model that will be implemented for more in-depth study in the prediction heritability of oil palm breeding using ML and without using genetic data.

The study has gained relevant knowledge on heritability prediction of the phenotypic properties of oil palm breeding using ML. There is an alternative way to predict heritability without using genetic data to obtain breeding values. Heritability value indeed plays a vital role in genomic selection to obtain a high breeding value. This study has produced a framework for an oil palm breeding program using ML. However, no evidence testing has been conducted. Therefore, this study will be continued by conducting experiments on phenotypic data based on the framework that has been introduced. This study suggests that the use of ML in oil palm studies be expanded to continue the sustainability of oil palm.

For the oil palm breeding framework, secondary data received from FGV in phenotype and numbered were used to predict offspring with high variability. These secondary data have eight different year groups. This convinced the authors that hereditary predictions could be made based on the group. The group with high variation is the offspring with a high breeding value. FFB, BNO, ABW in (kg/per/year), and O/B in value (%) are the parameters. These parameters make it difficult for the authors to identify whether those features played a significant role in the study. Studies focus only on FFB, BNO, and ABW [10] and analyze all of these traits [12,28,50]. Next, data cleaning should be performed to avoid biased data affecting the model's performance. Blank data are filled with an average from [74] according to the surrounding data points. The deletion method is also performed if many empty cells are in a row or column to remove noise [15]. The selection of features is also significant because features have less impact on heritability, such as Block, Trial, Represent, Program, and more. Once the data are cleaned, the data need to be divided into exercises and tests to be analyzed using algorithms. Data division usually uses folding techniques.

The k-fold cross-validation method is used to measure model performance, five-fold cross validation where data are divided into five subsets without overlapping, four subsets for training, and one subset for evaluation. The study used 20% of the data as a test set and the rest as a training set [79]. In comparison, 10-fold cross validation forms eight training sets, two folds as validation sets, and testing for model performance [71]. The Pearson correlation (PCCC) is a metric used to measure the prediction accuracy by changing ordinal and binary variables. Each test set and all folds were calculated from the properties of the combination to measure the predictive performance. The tuning process used a set of exercises and 20% of each set as a validation set to avoid biased results at each fold. A new study [80] produced automated validation to evaluate model performance on validation data sets to avoid the implementation of manual k-fold cross-validation that requires more computational resources.

This research shows that for the prediction of phenotypic properties, a ML method is preferred because the traditional method requires labor and is time-consuming. Since the phenotypic trait data are too large, more attention will be paid to the fruit yield component to build training and testing models. Data sets tend to produce bias against the majority sample, and some studies combine under-sampling and oversampling algorithms. The goal is to render a trained model with a balanced data set, improving model generalization, accuracy, F1, recall, and AUC for test sets. In a quantitative property prediction study of the genome [81], the best SVR performance was achieved when using LS-SVR with an RBF kernel. The study found that the SVR prediction correlation was similar to Bayesian Lasso, while SVR Predictive Mean Squared Error (PMSE) was slightly larger than Bayesian Lasso.

Other studies use algorithms, Regression [82], Random Forest (RF) [83], Support Vector Machine (SVM), Support Vector Regression (SVR) [84,85] and Boosting to compare the evaluation results. SVM parameter optimization was performed for each iteration during cross-validation using the function in package R e1071. The 10-fold cross-validation in the training set was used to calculate the optimal parameters in Genomic Selection of oil

palm [12]. SVM in the female inflorescence classification of oil palm study did not show good performance. This poor performance may be due to kernel dependencies, multi-class classification, and data measurement because SVM works well on binary classification problems. The study found that SVM had difficulty interpreting high-dimensional data [11].

The Bayesian methods used in the study [12] are $C\pi$, Ridge Regression [82], and Lasso: $C\pi$ assumes that π follows a uniform distribution [33], Bayesian Ridge Regression, in contrast, uses the previous Gaussian, and Lasso [35] uses the previous Double-Exponential [34]. The machine learning model had the highest accuracy among the methods evaluated, followed by the Bayesian approach and Ridge Regression Best Linear Unbiased Prediction (RR-BLUP). The study found that machine learning methods gave inconsistent results between features. However, the computational time cost of the Bayesian approach is more extended than RR-BLUP to increase the accuracy. In addition, the study found that machine learning methods gave inconsistent results between properties. Different parameters are required for various traits to maximize the accuracy of predictions using machine learning methods. Regression is commonly applied in oil palm breeding programs to make predictions about genomic selection. Phenotypic prediction using ridge regression provides higher accuracy than support vector regression [84,85]. Ridge regression shows a blue curve applied to support vector regression that typically achieves high accuracy in genetic variant studies.

Random forest (RF) [86] is an ensemble of machine learning that uses many decision trees. Each decision tree predicts a phenotype separately, and the average of all decision tree predictions is the result. To distinguish one tree from another, RF used markers, and the suspension of each tree was selected from a subgroup of random markers. For example, use 100 number trees as the basic model and calculate the permutation significance of the input properties with 10 iterations and consider the 10-fold cross-validation scheme at random. However, it should be noted the 10-fold cross-validation is consistent with the number of hyperparameters of the RF model tree [8]. In genome selection studies optimizing parameters for RF using grid search, parameters such as the number of trees and features need to be considered when finding the best fraction. RF can also be used in Scikit-learn [79].

The RF method is capable of identifying female inflorescences of oil palm using real-world data. RF can recognize nonlinear patterns in tree-structured data. The classification proposed by RF shows that RF's independent learning ability is not very sensitive to sample size. Thus, the RF model obtains more reliable information against redundant installations [11]. They were using the library for each cross-validation step to optimize the random forest parameters. Similar to SVM, 10-fold cross-validation is used on training sets. The optimization range, node size, and tree are set. Accuracy is defined as the predictive ability for each iteration calculated as the Pearson correlation coefficient of the predicted property value of the confirmation set versus its observed property value. The overall accuracy is the average of the prediction abilities from all iterations [12].

The use of RF for prescriptive breeding in high-dimensional phenotypic studies using ML could yield a framework for the development of prescriptive cultivars. Data were partitioned into 80% of training and 20% of test data sets, and models were trained to measure training performance using 10-fold cross-validation. Meanwhile, a test set was used for the validation of a model consisting of independent data and not trained data [83]. Model performance was evaluated with Root Mean Squared Error (RMSE) and coefficient of determination (R^2) for both training and test data. For model stability and accurate performance estimation, the study repeated the training and testing procedures 10 times. The study found an average of performance metrics across iterations. The average of several metrics is classified as a classification model and the average of R^2 is a regression model prediction capability. Model training, testing, and hyperparameter optimization were performed with the Caret and Forest package implemented in R by [83].

A recent study [87] used ensemble learning methods to identify orphan genes. In the study, it was found that XGBoost has better performance than RF and SVM. Therefore, this

method is particularly suitable for unbalanced data sets. Even so, XGBoost needs extensive parameter tuning to achieve a good performance. The study continued by comparing five types of SMOTE algorithms with the XGBoost model and found that SMOTE-ENN-XGBoost could predict a value of 0 or 1, and this method could also support species evolution. In line with that, a “boosting heritability” framework [88] used a multiple sample splitting strategy to employ the variable selection step. The selection is to remove irrelevant covariates that do not contribute to the variability of a trait and thus produce a reliable estimate of heritability. The study found that boosting heritability applicable for random-effect models to make estimates has a simple procedure. However, these two methods have not yet been implemented into oil palm breeding programs. It can be applied in this study by making hereditary predictions about whether they have a high reproductive value or not. The high reproductive value will aid in genetic improvement.

Fifth, model testing uses black boxes to test models where the internal works are unknown. Furthermore, this focuses on the incoming inputs and the outputs produced by the model being tested. The testing does not cover interior details such as server logic, expansion methods, and code. This model describes the overall output and performance of the model. This form of testing can improve model quality, reduce the risk of failure, and save time.

5. Conclusions

Sustainability ensures that the development of the oil palm industry can improve the national economy, the social well-being of the workers and communities involved, and not cause environmental pollution. As a producing country, palm oil can meet the demand for oils and fats, which are expected to develop sustainability with an increase in the global population to 9.7 billion by 2050. These reviews indicate that various studies have been conducted related to data other than genetic data. Several studies suggest that existing or new frameworks ensure that the framework’s accuracy can be trusted and applied in the oil palm breeding research industry to produce the best breed crosses. However, studies focusing on ML are still lacking in agriculture and plantation. This technology was newly formed in 2001, which can be adapted in all fields in today’s modern technological age.

Given the dynamics and differences in conventional breed selection and its implementation, further research on phenotypic variances and heritability estimation using ML should be conducted to further increase the breeding values and prediction models that will produce strong genes in the breed. While there are many results of the effectiveness of crop breeding models in achieving various sustainability objectives, ML breeding models have become a practical mechanism in studying and improving genetics in a particular offspring. Implementing machine learning in oil palm breeding data can play an essential role in achieving the SDGs.

Author Contributions: Conceptualization, N.A.L., F.N.M.N., and N.H.A.H.M.; writing—original draft preparation, N.A.L.; writing—review and editing, F.N.M.N., N.H.A.H.M., R.A., M.F.A.R., M.N.M., and N.S.M.F.; supervision, project administration, and funding acquisition, N.H.A.H.M., and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded via the Long-term Research Grant (LRGS) by Malaysian Research University Network (MRUN)—203.PKOMP.6777002.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful for the support provided by Universiti Sains Malaysia, our collaborator FGV Research and Development (R & D) Sdn Bhd, and anyone significant.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferdous Alam, A.S.A.; Er, A.C.; Begum, H. Malaysian oil palm industry: Prospect and problem. *J. Food Agric. Environ.* **2015**, *13*, 143–148.
2. Teo, C.J.; Chin, S.Y.; Wong, C.K.; Tan, C.C.; Goh, K.J. Planting Materials for High Sustainable Oil Palm Yields. In *Proceedings of the Malaysian Oil Science and Technology (MOST)*; Malaysian Oil Scientists' and Technologists' Association: Petaling Jaya, Selangor, Malaysia, 2017; Volume 26, pp. 58–119.
3. Eli-Chukwu, N.C. Applications of Artificial Intelligence in Agriculture: A Review. *Eng. Technol. Appl. Sci. Res.* **2019**, *9*, 4377–4383. [\[CrossRef\]](#)
4. Rahaman, M.M.; Chen, D.; Gillani, Z.; Klukas, C.; Chen, M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* **2015**, *6*, 619. [\[CrossRef\]](#)
5. Dijk, A.D.J.v.; Kootstra, G.; Kruijer, W.; Ridder, D.d.; van Dijk, A.D.J.; Kootstra, G.; Kruijer, W.; de Ridder, D.; Dijk, A.D.J.v.; Kootstra, G.; et al. Machine learning in plant science and plant breeding. *iScience* **2021**, *24*, 101890. [\[CrossRef\]](#)
6. Rival, A. Breeding the oil palm (*Elaeis guineensis* Jacq.) for climate change. *Oilseeds Fats Crop. Lipids* **2017**, *24*, 107. [\[CrossRef\]](#)
7. Washburn, J.D.; Burch, M.B.; Franco, J.A.V. Predictive breeding for maize: Making use of molecular phenotypes, machine learning, and physiological crop models. *Crop Sci.* **2020**, *60*, 622–638. [\[CrossRef\]](#)
8. Shahhosseini, M.; Hu, G.; Huber, I.; Archontoulis, S.V. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Sci. Rep.* **2021**, *11*, 1606. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Yoosefzadeh-Najafabadi, M.; Earl, H.J.; Tulpan, D.; Sulik, J.; Eskandari, M. Application of Machine Learning Algorithms in Plant Breeding: Predicting Yield From Hyperspectral Reflectance in Soybean. *Front. Plant Sci.* **2021**, *11*, 2169. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Chapman, R.; Cook, S.; Donough, C.; Lim, Y.L.; Vun Vui Ho, P.; Lo, K.W.; Oberthür, T. Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: A proof of concept analysis. *Comput. Electron. Agric.* **2018**, *151*, 338–348. [\[CrossRef\]](#)
11. Yousefi, D.B.M.; Rafie, A.S.M.; Aziz, S.; Azrad, S.; Masri, M.M.; Shahi, A.; Marzuki, O.F.M. Classification of oil palm female inflorescences anthesis stages using machine learning approaches. *Inf. Process. Agric.* **2020**, 1–13. [\[CrossRef\]](#)
12. Bin Kwong, Q.; Teh, C.K.; Ong, A.L.; Chew, F.T.; Mayes, S.; Kulaveerasingam, H.; Tammi, M.; Yeoh, S.H.; Appleton, D.R.; Harikrishna, J.A. Evaluation of methods and marker Systems in Genomic Selection of oil palm (*Elaeis guineensis* Jacq.). *BMC Genet.* **2017**, *18*, 107. [\[CrossRef\]](#)
13. Puttinaovarat, S.; Horkaew, P. Deep and machine learnings of remotely sensed imagery and its multi-band visual features for detecting oil palm plantation. *Earth Sci. Inform.* **2019**, *12*, 429–446. [\[CrossRef\]](#)
14. González-Camacho, J.M.; Ornella, L.; Pérez-Rodríguez, P.; Gianola, D.; Dreisigacker, S.; Crossa, J. Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. *Plant Genome* **2018**, *11*, 170104. [\[CrossRef\]](#)
15. Grinberg, N.F.; Orhobor, O.I.; King, R.D. An evaluation of machine-learning for predicting phenotype: Studies in yeast, rice, and wheat. *Mach. Learn.* **2020**, *109*, 251–277. [\[CrossRef\]](#)
16. Kushairi, A.; Basri, B. Senario Pasaran Bahan Tanaman dan Industri Sawit Negara. *Res. Inst. Malays. Palm Oil Board* **1996**, 1–12.
17. Amiruddin, M.D. Bahan Tanaman Berkualiti. *War. Sawit* **2013**, *54*, 4–5, 27.
18. Moretsohn, M.C.; Nunes, C.D.M.; Ferreira, M.E.; Grattapaglia, D. RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* **2000**, *100*, 63–70. [\[CrossRef\]](#)
19. Soh, A.C. *Breeding and Genetics of the Oil Palm*; AOCS Press: Urbana, IL, USA, 2012; ISBN 9780128043462.
20. Durand-Gassel, T.; de Franqueville, H.; Breton, F.; Amblard, P.; Jacquemard, J.; Syaputra, I.; Cochard, B.; Louise, C.; Nouy, B. Breeding for sustainable palm oil. In *Proceedings of the International Seminar on Breeding for Sustainability in Oil Palm*, Kuala Lumpur, Malaysia, 18 November 2011; pp. 178–193.
21. Hartung, F.; Schiemann, J. Precise plant breeding using new genome editing techniques: Opportunities, safety and regulation in the EU. *Plant J.* **2014**, *78*, 742–752. [\[CrossRef\]](#)
22. Sleper, D.A.; Poehlman, J.M. *Breeding Field Crops*; Sleper, D.A., Poehlman, J.M., Eds.; Blackwell Publishing: Oxford, UK, 2006; ISBN 9780813824284.
23. Murphy, D.J.; Goggin, K.; Paterson, R.R.M. Oil palm in the 2020s and beyond: Challenges and solutions. *CABI Agric. Biosci.* **2020**, *19*, 1–22. [\[CrossRef\]](#)
24. Bakoumé, C.; Ebongué, G.N.; Ajambang, W.; Ataga, C.; Okoye, M.; Enaberue, L.; Konan, J.E.C.; Allou, D.; Diabaté, S.; Konan, E.; et al. Oil Palm Breeding and Seed Production in Africa. In *Proceedings of the International Seminar on Oil Palm Breeding and Seed Production and Field Visits*, Kisaran, Indonesia, 29–30 September 2016; pp. 21–38.
25. Rajanaidu, N.; Ainul, M.M.; Kushairi, A.; Din, A. Historical Review of Oil Palm Breeding for the Past 50 Years. *J. Oil Palm Res.* **2010**, 11–28.
26. Allou, D.; Nan, O.A.N.; Guetta, P.A.N.; Breeding, P. Parental Diversity in Improved Populations of Oil Palm (*Elaeis guineensis* Jacq.) Jacq After Three Cycle of o Reciprocal Recurrent Selection. *Int. J. Agric. Innov. Res.* **2014**, *3*, 592–595.
27. Sambanthamurthi, R.; Singh, R.; Kadir, A.P.G.; Abdullah, M.O.; Kushairi, A. Opportunities for the Oil Palm via Breeding and Biotechnology Breeding Plantation Tree Crops: Tropical Species. In *Breeding Plantation Tree Crops: Tropical Species*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 377–421. ISBN 978-0-387-71201-7.
28. Sritharan, K.; Subramaniam, M.; Arulandoo, X.; Yusop, M.R. Yield and bunch quality component comparison between two-way crosses and multi-way crosses of dxp oil palm Progenies. *Sains Malays.* **2017**, *46*, 1587–1595. [\[CrossRef\]](#)

29. Amiruddin, M.D.; Kushairi, A. Development of New Oil Palm Cultivars in Malaysia. *J. Oil Palm Res.* **2020**, *32*, 420–426.
30. Cros, D.; Bocs, S.; Riou, V.; Ortega-Abboud, E.; Tisné, S.; Argout, X.; Pomiès, V.; Nodichao, L.; Lubis, Z.; Cochard, B.; et al. Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. *BMC Genom.* **2017**, *18*, 1–17. [[CrossRef](#)] [[PubMed](#)]
31. Cros, D.; Denis, M.; Sánchez, L.; Cochard, B.; Flori, A.; Durand-Gasselin, T.; Nouy, B.; Omoré, A.; Pomiès, V.; Riou, V.; et al. Genomic selection prediction accuracy in a perennial crop: Case study of oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* **2015**, *128*, 397–410. [[CrossRef](#)] [[PubMed](#)]
32. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, 372. [[CrossRef](#)]
33. Wei, W.-H.; Hemani, G.; Haley, C.S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* **2014**, *15*, 722–733. [[CrossRef](#)]
34. Mackay, T.F.C. Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nat. Rev. Genet.* **2014**, *15*, 22–33. [[CrossRef](#)]
35. Hallauer, A.R. Evolution of plant breeding. *Crop Breed. Appl. Biotechnol.* **2011**, *11*, 197–206. [[CrossRef](#)]
36. Teh, H.F.; Neoh, B.K.; Ithnin, N.; Daim, L.D.J.; Ooi, T.E.K.; Appleton, D.R. Review: Omics and Strategic Yield Improvement in Oil Crops. *J. Am. Oil Chem. Soc.* **2017**, *94*, 1225–1244. [[CrossRef](#)]
37. Chen, D.; Neumann, K.; Friedel, S.; Kilian, B.; Chen, M.; Altmann, T.; Klukas, C. Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis. *Plant Cell* **2015**, *26*, 4636–4655. [[CrossRef](#)] [[PubMed](#)]
38. Li, L.; Zhang, Q.; Huang, D. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* **2014**, *14*, 20078–20111. [[CrossRef](#)] [[PubMed](#)]
39. Rahaman, M.M.; Ahsan, M.A.; Chen, M. Data-mining Techniques for Image-based Plant Phenotypic Traits Identification and Classification. *Sci. Rep.* **2019**, *9*, 19526. [[CrossRef](#)]
40. Rethinam, P.; Murugesan, P. Global perspective of germplasm and breeding for seed production in oil palm. *Int. J. Oil Palm* **2018**, *10*, 17–34.
41. Morcillo, F.; Cros, D.; Billotte, N.; Ngando-Ebongue, G.-F.; Domonhédou, H.; Pizot, M.; Cuéllar, T.; Espéout, S.; Dhouib, R.; Bourgis, F.; et al. Improving palm oil quality through identification and mapping of the lipase gene causing oil deterioration. *Nat. Commun.* **2013**, *4*, 2160. [[CrossRef](#)]
42. Herrero, J.; Santika, B.; Herrán, A.; Erika, P.; Sarimana, U.; Wendra, F.; Sembiring, Z.; Asmono, D.; Ritter, E. Construction of a high density linkage map in Oil Palm using SPET markers. *Sci. Rep.* **2020**, *10*, 9998. [[CrossRef](#)]
43. Teh, C.K.; Ong, A.L.; Bin Kwong, Q.; Apparow, S.; Chew, F.T.; Mayes, S.; Mohamed, M.; Appleton, D.; Kulaveerasingam, H. Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. *Sci. Rep.* **2016**, *6*, 19075. [[CrossRef](#)]
44. Kainer, D.; Lanfear, R.; Foley, W.J.; Külheim, C. Genomic approaches to selection in outcrossing perennials: Focus on essential oil crops. *Theor. Appl. Genet.* **2015**, *128*, 2351–2365. [[CrossRef](#)]
45. Dislich, C.; Keyel, A.C.; Salecker, J.; Kisel, Y.; Meyer, K.M.; Auliya, M.; Barnes, A.D.; Corre, M.D.; Darras, K.; Faust, H.; et al. A review of the ecosystem functions in oil palm plantations, using forests as a reference system. *Biol. Rev.* **2017**, *92*, 1539–1569. [[CrossRef](#)]
46. Soh, A.C. Applications and challenges of biotechnology in oil palm breeding. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, 183. [[CrossRef](#)]
47. RAMLI, U.S. OMICS Platform Technologies for Discovery and Understanding the Systems Biology of Oil Palm. *J. Oil Palm Res.* **2020**, 141–157. [[CrossRef](#)]
48. Swaray, S.; Amiruddin, M.D.; Rafii, M.Y.; Jamian, S.; Ismail, M.F.; Jalloh, M.; Marjuni, M.; Mohamad, M.M.; Yusuff, O. Influence of parental dura and pisifera genetic origins on oil palm fruit set ratio and yield components in their D × P Progenies. *Agronomy* **2020**, *10*, 1793. [[CrossRef](#)]
49. Kalyana Babu, B.; Mary Rani, K.L.; Sahu, S.; Mathur, R.K.; Naveen Kumar, P.; Ravichandran, G.; Anitha, P.; Bhagya, H.P. Development and validation of whole genome-wide and genic microsatellite markers in oil palm (*Elaeis guineensis* Jacq.): First microsatellite database (OpSatdb). *Sci. Rep.* **2019**, *9*, 1899.
50. Bin Kwong, Q.; Ong, A.L.; Teh, C.K.; Chew, F.T.; Tammi, M.; Mayes, S.; Kulaveerasingam, H.; Yeoh, S.H.; Harikrishna, J.A.; Appleton, D.R. Genomic selection in commercial perennial crops: Applicability and improvement in oil palm (*Elaeis guineensis* Jacq.). *Sci. Rep.* **2017**, *7*, 2872. [[CrossRef](#)]
51. Rao, B.B.; Surya, M.C.; Kumar, T.K.; Naresh, S.; Madhuri, M.L. Oil Palm breeding strategies through molecular and genomics technologies: Status and way forward. *Int. J. Oil Palm* **2017**, *9*, 25–30.
52. Pangaribuan, I.F.; Yenni, Y. Evaluasi Karakter Kompak Hasil Pengujian Keturunan Siklus Ketiga Program Pemuliaan Kelapa Sawit Pusat Evaluation for Compact Character of Third Cycle Progeny Test in Iopri's Oil Palm Breeding Program. *Pus. Penelit. Kelapa Sawit* **2019**, *27*, 149–162. [[CrossRef](#)]
53. Singh, R.; Low, E.T.L.; Ooi, L.C.L.; Ong-Abdullah, M.; Ting, N.C.; Nagappan, J.; Nookiah, R.; Amiruddin, M.D.; Rosli, R.; Manaf, M.A.A.; et al. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **2013**, *500*, 340–344. [[CrossRef](#)]

54. Singh, R.; Low, E.-T.L.; Ooi, L.C.-L.; Ong-Abdullah, M.; Nookiah, R.; Ting, N.-C.; Marjuni, M.; Chan, P.-L.; Ithnin, M.; Manaf, M.A.A.; et al. The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. *Nat. Commun.* **2014**, *5*, 4106. [\[CrossRef\]](#)
55. Jalignot, E.; Hooi, W.Y.; Debladis, E.; Richaud, F.; Beulé, T.; Collin, M.; Agbessi, M.D.T.T.; Sabot, F.; Garsmeur, O.; D'Hont, A.; et al. DNA methylation and expression of the EgDEF1 gene and neighboring retrotransposons in mantled somaclonal variants of oil palm. *PLoS ONE* **2014**, *9*, e91896. [\[CrossRef\]](#)
56. Ong-Abdullah, M.; Ordway, J.M.; Jiang, N.; Ooi, S.-E.; Kok, S.-Y.; Sarpan, N.; Azimi, N.; Hashim, A.T.; Ishak, Z.; Rosli, S.K.; et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **2015**, *525*, 533–537. [\[CrossRef\]](#)
57. Guerin, C.; Joët, T.; Serret, J.; Lashermes, P.; Vaissayre, V.; Agbessi, M.D.T.T.; Beulé, T.; Severac, D.; Amblard, P.; Tregear, J.; et al. Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *Plant J.* **2016**, *87*, 423–441. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Corley, R.H.V.; Tinker, P.B. *The Oil Palm*; Sons, J.W., Ed.; Blackwell Science Ltd.: Oxford, UK, 2003; ISBN 9780470750971.
59. Barcelos, E.; De Almeida Rios, S.; Cunha, R.N.V.; Lopes, R.; Motoike, S.Y.; Babiychuk, E.; Skirycz, A.; Kushnir, S. Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.* **2015**, *6*, 190. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Bin Kwong, Q.; Teh, C.K.; Ong, A.L.; Heng, H.Y.; Lee, H.L.; Mohamed, M.; Low, J.Z.B.; Apparow, S.; Chew, F.T.; Mayes, S.; et al. Development and Validation of a High-Density SNP Genotyping Array for African Oil Palm. *Mol. Plant* **2016**, *9*, 1132–1141. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Khan, N.; Kamaruddin, M.A.; Sheikh, U.U.; Yusup, Y. Oil Palm and Machine Learning: Reviewing One Decade of Ideas, Innovations, Applications, and Gaps. *Agriculture* **2021**, *11*, 832. [\[CrossRef\]](#)
62. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [\[CrossRef\]](#)
63. Pushparani, M.; Sagaya, A.; Ravan, S. Big data analytics using weight estimation algorithm for oil palm plantation domain. *Int. J. Adv. Soft Comput. Its Appl.* **2018**, *10*, 71–89.
64. Bensaeed, O.M.; Shariff, A.M.; Mahmud, A.B.; Shafri, H.; Alfatni, M. Oil palm fruit grading using a hyperspectral device and machine learning algorithm. *IOP Conf. Ser. Earth Environ. Sci.* **2014**, *20*, 012017. [\[CrossRef\]](#)
65. Silalahi, D.D.; Reaño, C.E.; Lansigan, F.P.; Panopio, R.G.; Bantayan, N.C. Using Genetic Algorithm Neural Network on Near Infrared Spectral Data for Ripeness Grading of Oil Palm (*Elaeis guineensis* Jacq.) Fresh Fruit. *Inf. Process. Agric.* **2016**, *3*, 252–261. [\[CrossRef\]](#)
66. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. *IEEE Access* **2021**, *9*, 63406–63439. [\[CrossRef\]](#)
67. Hilal, Y.Y.; Ishak, W.; Yahya, A.; Asha'ari, Z.H. Development of genetic algorithm for optimization of yield models in oil palm production. *Chil. J. Agric. Res.* **2018**, *78*, 228–237. [\[CrossRef\]](#)
68. Hazir, M.H.M.; Shariff, A.R.M. Oil palm physical and optical characteristics from two different: Planting materials. *Res. J. Appl. Sci. Eng. Technol.* **2011**, *3*, 953–962.
69. Bai, B.; Wang, L.; Lee, M.; Zhang, Y.; Alfiko, Y.; Ye, B.Q.; Wan, Z.Y.; Lim, C.H.; Suwanto, A.; Chua, N.H.; et al. Genome-wide identification of markers for selecting higher oil content in oil palm. *BMC Plant Biol.* **2017**, *17*, 1–11. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Arolu, I.W.; Rafii, M.Y.; Marjuni, M.; Hanafi, M.M.; Sulaiman, Z.; Rahim, H.A.; Abidin, M.I.Z.; Amiruddin, M.D.; Din, A.K.; Nookiah, R. Breeding of high yielding and dwarf oil palm planting materials using Deli *dura* × Nigerian *pisifera* population. *Euphytica* **2017**, *213*, 154. [\[CrossRef\]](#)
71. Hartwell, L.; Goldberg, M.L.; Fischer, J.A.; Hood, L.E.; Aquadro, C.F. *Genetics: From Genes to Genomes*; McGraw-Hill Education: New York, NY, USA, 2018.
72. Falconer, D.S. Introduction to quantitative genetics. In *Pearson Education India*; Pearson: London, UK, 1996.
73. Liu, Y.; Wang, D.; He, F.; Wang, J.; Joshi, T.; Xu, D. Phenotype Prediction and Genome-Wide Association Study Using Deep Convolutional Neural Network of Soybean. *Front. Genet.* **2019**, *10*, 1091. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Rastin, N.; Aminafshar, M.; Honarvar, M.; Jomeh, N.E. Imputation of Ungenotyped Individuals Based on Genotyped Relatives Using Machine Learning Methodology. *J. Epigenet.* **2021**. [\[CrossRef\]](#)
75. Bai, B.; Wang, L.; Zhang, Y.J.; Lee, M.; Rahmadsyah, R.; Alfiko, Y.; Ye, B.Q.; Purwantomo, S.; Suwanto, A.; Chua, N.H.; et al. Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Sci. Rep.* **2018**, *8*, 691. [\[CrossRef\]](#)
76. Ali, M.; Zhang, L.; DeLacy, I.; Arief, V.; Dieters, M.; Pfeiffer, W.H.; Wang, J.; Li, H. Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis. *Crop J.* **2020**, *8*, 866–877. [\[CrossRef\]](#)
77. Harahap, I.Y.; Lubis, M.E.S. Penggunaan Model Jaringan Saraf Tiruan (Artificial Neuron Network) Untuk Memprediksi Hasil Tandan Buah Segar (Tbs) Kelapa Sawit Berdasar Curah Hujan Dan Hasil Tbs Sebelumnya. *J. Penelit. Kelapa Sawit* **2018**, *26*, 59–70. [\[CrossRef\]](#)
78. Xavier, A.; Muir, W.M.; Craig, B.; Rainey, K.M. Walking through the statistical black boxes of plant breeding. *Theor. Appl. Genet.* **2016**, *129*, 1933–1949. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Frouin, A.; Dandine-Roulland, C.; Pierre-Jean, M.; Deleuze, J.F.; Ambroise, C.; Le Floch, E. Exploring the Link Between Additive Heritability and Prediction Accuracy From a Ridge Regression Perspective. *Front. Genet.* **2020**, *11*, 1–15. [\[CrossRef\]](#)
80. Sandhu, K.S.; Lozada, D.N.; Zhang, Z.; Pumphrey, M.O.; Carter, A.H. Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program. *Front. Plant Sci.* **2021**, *11*, 2084. [\[CrossRef\]](#) [\[PubMed\]](#)

81. Saiz-Rubio, V.; Rovira-Más, F. From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management. *Agronomy* **2020**, *10*, 207. [[CrossRef](#)]
82. Jubair, S.; Domaratzki, M. Ensemble supervised learning for genomic selection. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, San Diego, CA, USA, 18–21 November 2019; pp. 1993–2000. [[CrossRef](#)]
83. Montesinos-López, O.A.; Martín-Vallejo, J.; Crossa, J.; Gianola, D.; Hernández-Suárez, C.M.; Montesinos-López, A.; Juliana, P.; Singh, R. A benchmarking between deep learning, support vector machine and Bayesian threshold best linear unbiased prediction for predicting ordinal traits in plant breeding. *G3 Genes Genomes Genet.* **2019**, *9*, 601–618. [[CrossRef](#)] [[PubMed](#)]
84. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Application of support vector regression to genome-assisted prediction of quantitative traits. *Theor. Appl. Genet.* **2011**, *123*, 1065–1074. [[CrossRef](#)] [[PubMed](#)]
85. Aljouie, A.; Roshan, U. Prediction of continuous phenotypes in mouse, fly, and rice genome wide association studies with support vector regression SNPs and ridge regression classifier. In Proceedings of the IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, 9–11 December 2015; pp. 1246–1250. [[CrossRef](#)]
86. Parmley, K.A.; Higgins, R.H.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Machine Learning Approach for Prescriptive Plant Breeding. *Sci. Rep.* **2019**, *9*, 17132. [[CrossRef](#)]
87. Gao, Q.; Jin, X.; Xia, E.; Wu, X.; Gu, L.; Yan, H.; Xia, Y.; Li, S. Identification of Orphan Genes in Unbalanced Datasets Based on Ensemble Learning. *Front. Genet.* **2020**, *11*. [[CrossRef](#)]
88. Mai, T.T.; Turner, P.; Corander, J. Boosting heritability: Estimating the genetic component of phenotypic variation with multiple sample splitting. *BMC Bioinform.* **2021**, *22*, 1–16. [[CrossRef](#)]