

Review

# A Review of an Artificial Intelligence Framework for Identifying the Most Effective Palm Oil Prediction

Fatini Nadhirah Mohd Nain <sup>1</sup>, Nurul Hashimah Ahamed Hassain Malim <sup>1,\*</sup>, Rosni Abdullah <sup>1</sup>,  
Muhamad Farid Abdul Rahim <sup>2</sup>, Mohd Azinuddin Ahmad Mokhtar <sup>2</sup> and Nurul Syafika Mohamad Fauzi <sup>2</sup>

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

<sup>2</sup> 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.);

azinuddin.am@fgvholdings.com (M.A.A.M.); syafika.mf@fgvholdings.com (N.S.M.F.)

\* Correspondence: nurulhashimah@usm.my

**Abstract:** Machine Learning (ML) offers new precision technologies with intelligent algorithms and robust computation. This technology benefits various agricultural industries, such as the palm oil sector, which possesses one of the most sustainable industries worldwide. Hence, an in-depth analysis was conducted, which is derived from previous research on ML utilisation in the palm oil industry. The study provided a brief overview of widely used features and prediction algorithms and critically analysed current the state of ML-based palm oil prediction. This analysis is extended to the ML application in the palm oil industry and a comparison of related studies. The analysis was predicated on thoroughly examining the advantages and disadvantages of ML-based palm oil prediction and the proper identification of current and future agricultural industry challenges. Potential solutions for palm oil prediction were added to this list. Artificial intelligence and machine vision were used to develop intelligent systems, revolutionising the palm oil industry. Overall, this article provided a framework for future research in the palm oil agricultural industry by highlighting the importance of ML.

**Keywords:** agriculture; deep learning; machine learning; palm oil; palm oil breeding



**Citation:** Mohd Nain, F.N.; Ahamed Hassain Malim, N.H.; Abdullah, R.; Abdul Rahim, M.F.; Ahmad Mokhtar, M.A.; Mohamad Fauzi, N.S. A Review of an Artificial Intelligence Framework for Identifying the Most Effective Palm Oil Prediction. *Algorithms* **2022**, *15*, 218. <https://doi.org/10.3390/a15060218>

Academic Editor: Frank Werner

Received: 1 May 2022

Accepted: 17 June 2022

Published: 20 June 2022

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



**Copyright:** © 2022 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

Palm oil plantations cover millions of hectares worldwide, which encompass a significant portion of global trade. Palm oil trees, or *Arecaceae*, are a genus of stemless, tree-like monocot plants that thrive in the tropics and are extremely valuable to humans and the ecosystem [1]. The African oil palm, or *Elaeis guineensis*, is the most prominent palm species native to West Africa, cultivated for its oil-rich fruit as a semiwild food source for over 7000 years. The tree produces a profusion of fruit bunches yearly with each containing between 1000 and 3000 fruits [2]. The processed oil palm fruits are a significant source of oil for society and an integral industrial derivative, i.e., soaps, detergents, and cosmetics. Hence, the industry significantly impacts locals and broader biodiversity in their native regions [1,3,4].

Malaysia and Indonesia are the largest exporters of palm oil in Southeast Asia [5]. The development rate of the palm oil industry is the primary contributor to the agriculture sector's value, thus indicating a 37.1% share. The agricultural industry accounted for 7.4% of Malaysia's GDP in 2020, though its growth rate fell to 2.2% from 2.0% in the previous year. The commodity subsector has seen a diminished growth of 3.6% due to the palm oil shortage (2019: 1.5 %) [6]. This issue is escalated with the rise in crude palm oil prices, directly resulting from the market disequilibrium in India and China. Furthermore, palm oil production is increasingly threatened by environmental, economic, and political factors [7]. Success in the industry is determined by the estate managers' ability to strategically adapt

to administrative changes in a process-oriented manner [8]. A plantation manager must make rapid decisions on various issues, i.e., personnel, strategy, area, and input, which are made under duress, thus resulting in a substandard outcome. The immediate and long-term risks must be addressed after placing substantial faith in their intuition, which leads to rash decisions and perfunctory efforts in testing the data.

Decision support systems can assist managers by summarising data-driven analysis and providing objective and rational viewpoints on complex production systems. These systems encompass models for palm oil computer simulation, i.e., FGV Integrated Breeding System (FIBS) [9], Palm Oil Soil Monitoring System for Smart Agriculture [10], and others; however, it is challenging to create a comprehensive palm oil computer model for several reasons. Firstly, the predictions are based exclusively on the breeding genome, which neglects the environmental or phenotype parameters. Moreover, parameter quantification is an expensive and time-consuming process. Generalisations can also be tricky when applied to geographical and environmental settings that are vastly different. Finally, the output is commonly a single predicted yield based on environmental and breeding variables.

Notably, “Big data” has recently become a standard paradigm in agricultural research [11], typically described based on the five Vs: Volume, Velocity, Variety, Veracity, and Valorisation [12–14]. These acronyms illustrate the amount of data, how quickly it can be accessed, how diverse it is, and how easily it can be used to generate new insights and discoveries. Notably, there are approximately \$20 billion in annual global benefits from the efficient exploitation of agricultural resources, thus making it critical to utilise agricultural data efficiently. However, the data analysis in this sector is underdeveloped compared to other industries [15,16]. The raw data must first be transformed into high-value knowledge to construct actionable management [17,18] and make better decisions. Big data analysis in agriculture is incompatible with the conventional experimental and statistical methods used in the past [19,20]. The foundation of Fishers’ feature selection methods and associated experimental designs necessitates small samples drawn from large populations. However, big data analysis frequently includes extensive samples and, in some cases, the entire population. Thus, it is commonly associated with excessive noise, heterogeneity, spurious correlations, and unexpected endogenous data sources. Hence, it can be improved using Machine Learning (ML) [19,20], where human-like intelligence can be simulated using algorithms. Artificial intelligence learns to recognise patterns and structures in datasets before predicting future events based on the model learned. Moreover, ML algorithms can discover specific rules for the system under investigation as it does not rely on user-specified models to analyse big data.

ML has been increasingly utilised to analyse big agricultural data, including crop type prediction from satellite data, crop yields, irrigation needs, pest and disease attacks, and weed identification [15,16]. This approach is adopted in crop yield in crops such as corn [21,22], grape [23], soybean [24], wheat [25,26], chilli [27], and paddy [28], and especially in the palm oil industry, which automates various tasks such as tree counting and plant health assessment due to machine learning’s incredible learning and significant computational power [29–31]. Modern agriculture is motivated by examining all factors contributing to crop yield gaps. Nevertheless, various factors threaten agricultural productivity, including biotic and abiotic ones, amplified by climate change, which affects its long-term viability.

The yield prediction of tree crops, such as oil palms, is a significantly challenging task. Multilayered and large datasets are required to understand and mitigate these risk factors. Conventional methods are inapplicable to map their connections which explicitly deal with interdependent and erratic elements. Furthermore, advanced analytics must be integrated with the highly heterogeneous datasets, generating insights into the critical constraints of yields at the tree and field scales. Thus, ML serves as an efficient technique to obtain accurate results and appropriate solutions for complex problems. This “divide and conquer” ML strategy was used to study the overall yield gap using microcomponents analysis [32,33]. The studies concluded that investigating individual yield-reducing factors could shed light on the overall impact.

Machine learning offers a quantitative perspective in assessing different variables such as soil moisture, rainfall, yield, solar radiation, and plant growth. On the other hand, breeding a bunch offers a qualitative angle on the matter [34–36]. The preceding statement entails making such research accessible as scientifically structured reviews can help new researchers with information based on accumulated data. The scientific community’s strategic concerns can be illustrated, and data from cluster analyses can be assessed using this method. Moreover, researchers can identify research gaps by referring to previously described systems, concepts, and propositions. The review supports an effective research strategy predicated on previous research and current techniques [37].

Multiple studies were conducted on practical tools and ML techniques related to the palm oil industry. For instance, studies investigated remote sensing applications, followed by breeding applications and technologies in monitoring palm oil plantations [38,39]. Other studies assessed the biofuel-processing technologies in dealing with fruit and palm oil waste [40]. Meanwhile, Barbedo et al. [41] and Khosrokhani et al. [42] investigated ML to detect nutritional deficits in palm oil using proximal images. Studies have also reviewed the ML features in automated fruit grading using image processing and predicting crop yields, including palm oil [43]. Nevertheless, most of the current studies did not conduct a thorough literature review. Table 1 lists the objectives of the most recent related review articles, their critical evaluations, and publication dates.

**Table 1.** An analysis of recent literature reviews on algorithms in palm oil yield prediction.

Reference	Objectives of the Review	Lack of the Review
[38]	Applications of remote sensing technology for palm plantation monitoring. A list of the existing knowledge gaps is compiled. The findings are followed up with recommendations for further investigation.	This review omits any mention related to the algorithm prediction of palm oil yield prediction.
[39]	An overview of the biotechnology used in palm oil breeding. This paper only reviews the application of molecular methods.	It does not employ ML techniques. These methods are insufficient to forecast specific breeding.
[41]	An overview of the techniques used for detecting palm oil nutrition deficiencies using proximal images and application examples.	This review does not mention algorithms for predicting palm oil yields and breeding.
[44]	An overview of existing ML models for predicting crop yields. Various performance metrics are employed to assess the effectiveness of various strategies.	Excludes other factors such as breeding. It only uses ML models to predict crop yield and few predictions on palm oil.
[45]	An overview of several existing ML models in forecasting palm oil yields. Multiple performance metrics were used to determine the effectiveness of various strategies.	Excludes other factors such as breeding and only uses ML models to predict palm oil yield.

Based on Table 1, only several aspects of palm oil using ML are discussed, i.e., yield prediction, crop monitoring, and nutrient deficits. Over the last decade, unidimensional reviews have yet to present the full scope of palm oil cultivation involving ML and multiple artificial intelligence (AI) aspects. Thus, this review article provides an in-depth perspective on how ML is used in palm oil cultivation, summarising everything that has been conducted to date. Unlike others, this review employed a systematic protocol to retrieve data from databases, ensuring objectivity. The contributions of this review paper are summarised as follows:

- Outlining the background of the palm oil breeding programmes.
- Introducing the factor that affects the palm oil growth and the fruit quality.
- Comprehensive critical assessment of ML-based palm oil prediction algorithms, critical evaluation of feature sets used, and comparison of relevant research.

- A thorough examination of the advantages and drawbacks of ML algorithms in predicting palm oil yield.

This review outlined current types of research in predicting palm oil production using ML by extensively collecting various forecasting models for the palm oil ML framework. The rest of the paper is organised as follows: Section 2 presents a detailed review of related studies on the background of palm oil and factors affecting its growth and quality. Section 3 introduces the existing research on predicting palm oil yield using ML. Meanwhile, Section 4 discusses issues of palm oil predictions, future directions, and the selection of the optimum prediction algorithm. Section 5 summarises the framework prediction of palm oil, finalising with the conclusion in Section 6.

## 2. The PRISMA Strategy Article Selection

The search is narrowed down to the basic concepts relevant to the scope of this review. Various published studies are likely outside this review article's scope due to the substantial applications in ML. Figure 1 depicts the article selection process following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) strategies [46]. Google Scholar, Scopus, Web of Sciences, IEEE Explore, Science Direct, SpringerLink, and Taylor Francis were the databases used to find the relevant articles. Searching for the most basic information was performed by an automated search engine. We consulted several reputable websites in order to obtain specific information. The scope of this review is primarily determined by two keywords: (“machine learning” OR “deep learning”) AND “oil palm”. As a result, in the majority of the searches, we used these two keywords in conjunction with other keywords such as “crop yield”, “palm oil”, “genomic prediction”, “yield forecasting”, “yield estimation”, “prediction model”, “*Elaeis guineensis*”, “machine learning”, “deep learning”, “artificial intelligence”. Among the documents we have collected are original articles, review articles, book chapters, conference papers, lecture notes, and reports.

We downloaded, read the title, abstract, conclusion, keywords datasets, models used, techniques, and metric evaluations in the first round of review and selected only the relevant articles manually and then entered them into excel. Articles that have been included in excel will be refiltered to be analyzed with data that may be important in this study to facilitate information retrieval to improve contextual understanding. This selection is due to constraints such as the difficulty to find palm oil study factors, especially in phenotypic data, because they are rarely studied because they have a long plant life span to bear fruit. After removing duplicates, the remaining articles are classified according to their application. During the first round of screening, machine learning-based palm oil and crop prediction articles written in English are used as selection criteria.

The selected manuscripts are scrutinised in the second round of screening. The detailed methodology was the final article selection criteria in this case. After following the above procedures, only “163” articles are chosen for this review.

The following information was meticulously extracted from each article: publication year, dataset information, detailed information about features, prediction algorithms, and system performance. The number of selected documents is divided into four categories: journal articles (82%), conference papers (11%), book chapters (5%), and websites (2%), as shown in Figure 2. In terms of the annual distribution of published work, trends show that a greater number of research articles were published in 2017, 2019, and 2021, while the number of articles dropped too low in 2015. Figure 3 depicts the annual distribution of research publications.

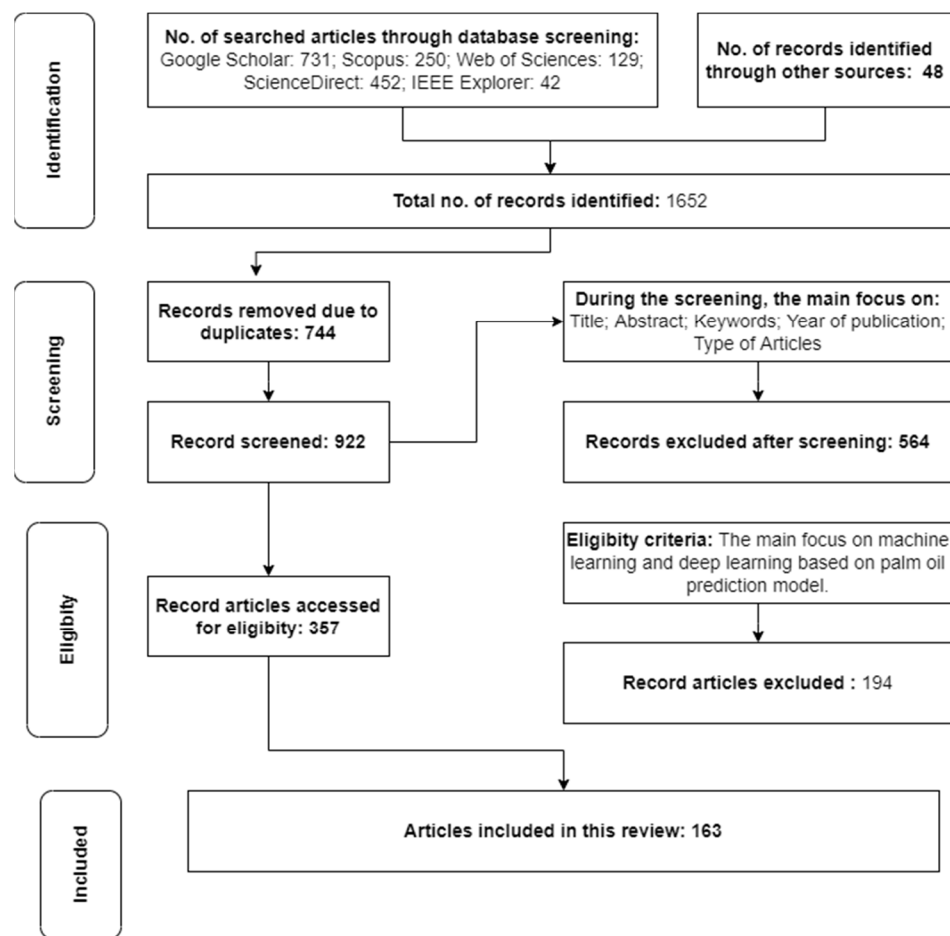


Figure 1. The PRISMA flow diagram.

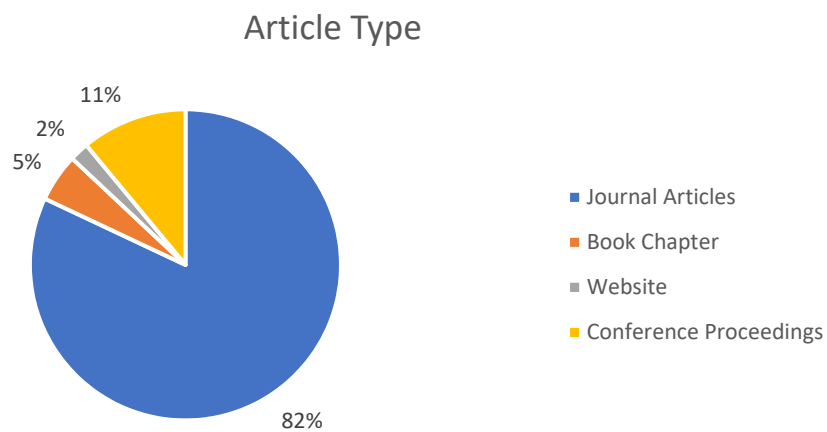
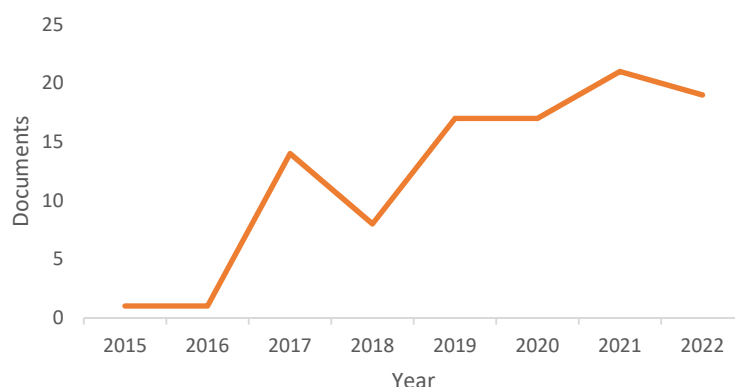


Figure 2. Type of Articles.



**Figure 3.** The annual publication of articles that have been reviewed.

### 3. Palm Oil Background

The Arecaceae family includes palm oil (*Elaeis* spp.), encompassing two species: African palm oil (*E. guineensis*) and American palm oil (*E. occidentalis*) (*E. oleifera*). The male and female inflorescences can be detected on the same palm, and on rare occasions, inflorescences of hermaphrodites can be found as well. A cross-pollinated crop is created when the male and female inflorescences are produced alternately. However, artificial pollination is necessary to produce specific hybrids [47,48]. The drupe, or the oil palm's fruit, matures approximately six months after pollination. Within the drupe, a pericarp is formed from the exocarp (the outer layer) and the husk. The outer layer protects the kernel containing the endosperm and embryo located within the endocarp (the inner layers).

The *E. guineensis* species are classified into three oil palm fruit forms (pisifera, dura, and tenera) based on the thickness of their shell, an essential factor that will be used for breeding [49]. The pisifera genotype of the recessive homozygous ( $Sh^- Sh^-$ ) alleles is shell-less. It is believed that most palms are sterile; however, several were reported to bear fruit and have varying degrees of sterility. The dura genotype comprises a thick shell consisting of the dominant ( $Sh^+ Sh^+$ ) alleles. Meanwhile, the tenera of the heterozygous ( $Sh^+ Sh^-$ ) alleles exhibit a mesocarp with a thinner outer shell and a thicker inner ring of fibres. Notably, the tenera genotype is the only form of oil palm fruit used for commercial planting because of its higher mesocarp content. The three types of oil palm fruits are the dura (D), pisifera (P) and tenera (T), identified based on the thickness of their shells. Alleles  $Sh^+$  and  $Sh^-$  are codominantly expressed at a single locus that controls shell thickness [50]. The thick-shell dura is controlled by a dominant homozygote gene ( $Sh^+ Sh^+$ ), whereas the shell-less pisifera is controlled by a recessive homozygote gene ( $Sh^- Sh^-$ ). The cross between the dura and pisifera would result in a homozygote tenera hybrid ( $Sh^+ Sh^-$ ) with a thin shell. Despite having a high mesocarp content of 95 per cent, pisifera is usually female sterile or semifertile and does not produce bunches. Hence, pisifera is only used as the male parent in the tenera hybrid. Notably, the tenera genotype is the only form of oil palm fruit used for commercial planting due to its higher mesocarp content.

#### 3.1. Breeding Programmes

Tenera is a commercial term for palm oil, a cross between dura  $\times$  pisifera, responsible for various new varieties. The germplasm of these materials is used to produce hybrid seeds. Vegetative characteristics and the bunch performance of four seedlings introduced to Indonesia in 1848 are consistent [51]. In 1848, Indonesia introduced the vegetative characteristics and consistent performance of four oil palm seedlings [51,52]. Private companies proposed various independent, breed-specific variations. Deli subpopulations and breeding materials are currently available for purchase or trade worldwide. Accordingly, the following populations were used in effective breeding programmes [2,53].

1. Deli: The thick-shelled dura is a descendant of the original Bogor palms from Java. The subsequent progeny and local selection distribution to other countries resulted in



the development of subpopulations in Malaysia. The regions include Elmina, Serdang, Avenue, and Ulu Remis Deli Dura, followed by the Ivory Coast of Dabou and Le Mé Dura. This idea led to speculation that all four Bogor palms were descended from the same ancestor. All major commercial hybrid seed production programmes utilise the mother Deli (dura) palm. The Dumpy and Gunung Melayu palms are short variants of the longer Deli palms.

2. AVROS: AVROS seeds were collected from the Eala Botanical Garden (Jardin Botanique d'Eala) in Zaire (now the Democratic Republic of Congo) in 1923. SP540 is a common name for this pisifera, known for its vigorous growth, precocious bearing, thin shell, thick mesocarp, and high-yielding traits. Notably, the Deli (dura) × AVROS (pisifera) is the basis for effective seed production programmes in Indonesia, Malaysia, Colombia, Papua New Guinea, and Costa Rica.
3. Yangambi: The seeds are acquired from the INEAC in Yangambi, Democratic Republic of the Congo. The population of the Dejongo palm and Yawenda tenera was developed using open-pollinated seeds, distinguished from their large fruits and high oil content.
4. La Mé: Twenty-one tenera palm seeds were collected from the wild groves of the Ivory Coast by IRHO, creating the La Mé populations. The tenera palms are used in the seed production industry in West Africa and Indonesia. The La Mé progenies (pisifera) are smaller and bear fewer fruits per bunch but are resilient in less-ideal growing conditions.
5. Binga: The pisifera subpopulation was derived from Yangambi progenies from F2 and F3 generations. They are planted in the Binga plantation in Yangambi, Democratic Republic of Congo. The Bg 312/3 and Bg 312/3 are two-parent palm varieties of interest for breeding purposes.
6. Ekona: Wild palms were used from the Ekona region to create the Ekona population. The regions include Unilever's Crown Estate, Ndian Estate, and Lobe Estate plantations in Cameroon. Its high bunch yield, excellent oil content, and wilt resistance make it a sought-after crop.
7. Calabar: Aba, Calabar, Ufuma, and Umuabi are all represented in Nigeria Institute for Oil Palm Research's (NIFOR) breeders, which are more diverse than their predecessors. Hence, many seed-production programmes make use of this pisifera.

### 3.2. Factors Affecting Palm Oil Growth and Quality

Palm oil forecasting has been critically shown in studies, particularly in early warning of potential problems. The current state of palm oil modelling indicates a lack of knowledge on palm oil growth and the ability to improve the fruit quality [54,55]; thus, an in-depth analysis is required. It is necessary to examine the heterogeneity of its production to better comprehend and exert control over the emergence of palm oil. An initial step can be taken before establishing predictive modelling to gain insight and understand the factors (environmental factors, phenotypes, and genotypes).

This step includes determining the relationships between palm oil and the method used in the growth analysis. An analysis method is essential to gain insight into the factors and inter-relationships in palm oil production. Palm oil analysis provides a comprehensive understanding of the uncertainty and nonlinearity of its forecast. The most common method of analysing the palm oil data is through, for example, using a preprocessing algorithm or feature selection algorithm. However, previous studies only needed the analysis with no attempt at making predictions.

Oil palms require light and nutrients such as nitrogen and phosphorus as a starting point for photosynthesis. Temperature and water turbidity are other environmental factors that affect palm oil production. Five environmental parameters facilitate the oil palm stress tolerance [56]: rainfall, temperature, relative humidity, light intensity, and wind speed. The inefficient production of palm oil may occur in low humidity levels, especially in the dry seasons when the watering holes and rivers have dried up. Oil palms may be more stressed during this season, affecting their yields.

Humidity concentration in palm oil cultivation can be predicted [57], for example, a study recognised several environmental effects influencing inflorescence abortion and sex determination. These factors include rainfall, monthly rain, sunshine hour, and evapotranspiration, followed by the minimum and maximum temperature [58]. Notably, in various studies, rainfall is positively correlated with crop output but is weakly associated with crop prices [59]. However, a cointegration analysis from 2018 to 2050 reported that rainfall changes affect future and spot prices with different time lags [60]. Furthermore, a study utilised monthly temperature anomalies to successfully predict palm oil yields [61].

Soil type and texture potentially improve the accuracy of forecasts and climate. Malaysian yields of over 30T fruit bunches per hectare were reported on all soil types, excluding shallow ones. This type exhibits issues such as reduced root proliferation, increased sensitivity to drought and flooding, and a higher risk of palms toppling. The most common soil type based on the Asian soil taxonomy is ultisols and oxisols [62,63]. One study used 14 different soil textures and chemical variables on six palm oil plantations to measure the effects on palm oil and the physicochemical properties of soil [64]. Anaba et al. [65] mentioned that palm oil physiological and behavioural adaptations to survive were defined as early tenera hybrid seedling stages. Sandy soils with macropores exhibit low resistance to penetration, producing excellent root growth in length and ramification. Meanwhile, the variability of soil in the water column is recognised as a more effective measurement approach than conventional techniques.

Phenotyping is possible at all levels of an organism's organisation, including the subcellular, cellular, tissue, organismic, and agrophytocenosis levels. This approach is used to identify productivity determinants, abiotic stressors, and plantation planning, especially on lands. This list can be extended to include the determining of the critical mechanisms of oil palm's resistance to pathogens [66]. In oil palm phenotyping, the fruit's size, shape, and physiological and biochemical characteristics are considered. These factors are then evaluated under specific environmental conditions and the oil palm genome [67,68]. Essentially, modern phenotyping methods enable the collection of real-time data and information analysis on the entirety of the phenotypic features. Hence, palm oil growth, development, and reproduction processes can now be comprehensively investigated [69,70]. One common phenotype in palm oil is the identification of ripe fruit. One of the prevalent phenotypes used in research is identifying the types of fruits, whether ripe or unripe.

Genotype has received significant attention, exhibiting the potential to improve the environment, phenotype, and forecast accuracy. The individual genotype can be determined using genotypic assaying, a method used in genotyping technology. Previous research used molecular markers (known DNA markers) breeding by deciphering genetics based on deoxyribonucleic acid (DNA) to enhance palm oil fruits. These markers include restriction fragment length polymorphism markers (RFLPs), amplified fragment length polymorphism markers (AFLPs), and short tandem repeat or simple sequence repeat markers (SSRs). Furthermore, the markers are used in the early stages of oil palm genetic mapping. For instance, the SNP's marker entails linkage and linkage disequilibrium (LD) mapping. This marker is favoured because of its abundance, low mutation rates, and amenability to high-throughput analysis.

Automated and high-throughput genotyping is well-suited to the binary SNPs at the genomewide scale. This genotyping technique is currently possible using the array [71,72] or sequencing-based technologies [73,74]. The SNP arrays serve as an alternative to the laborious cloning and primer design, though it lacks the discovery process and favours genotyping new populations. Hence, new sequencing techniques have emerged, including next-generation sequencing, i.e., restriction-site-associated DNA sequencing (RAD-seq) [75] and genotyping by sequencing (GBS) [76]. Currently, genome-wide markers can be discovered in a model of palm oil. Table 2 presents several categories in which the variables listed above, and others, can be sorted, and additional variables present a list of categories that influence oil palm growth along with input features used in previous studies.



**Table 2.** Categorical variables.

Type of Components	Features Data	Factor Category
Soil and Fertilisation	Levels of potassium (K), nitrogen (N), phosphorus (P), magnesium (Mg), calcium (Ca), nutrient supplements of groundwater, topography and slope, consumption of chemical fertilisers, consumption of agricultural pesticide, type and texture of the soil, and planting density.	Environment (E)
Climatic	Precipitation, humidity, maximum temperature, minimum temperature, mean temperature, rainfall, irrigation carbon dioxide concentration, solar radiation, water measurements, and waterlogging.	
UAV for Canopy Surveillance	Tree crown, stems length, tree colours, crown size, bare land, and water.	Phenotype (P)
Yield	Low lipase, high stearic and high oleic, low free fatty acids content and high iodine value, oil quality, quality, yield, oil to-dry mesocarp (O/DM), total oil yield per palm (O/P), oil content in mesocarp, oil-to-wet mesocarp (O/WM), palm oil yield (PO), FFB, yield, height, and high carotene.	
Fruit	Mesocarp-to-Fruit (M/F), Shell-to-Fruit (S/F), Kernel-to-Fruit (K/F), oil-bearing mesocarp, shell thickness, reasonable stalk length, good fruit set, and low parthenocarpy, fruit form, fruit quality, fruit colour virescent (VIR), and fruit ripeness.	
Bunch	Annual cumulative bunch number (BN), annual average bunch weight (ABW), annual cumulative bunch production (FFB), bunch production, bunch index, pulp-to-fruit ratio (PF), fruit-to-bunch (F/B), oil to bunch (O/B) oil extraction rate (OER), and oil-to-pulp ratio (OP), average bunch number per hectare (BUNCH_HA), and bunch weight (BW).	
Crop Samples	Flavonoid, anthocyanin content, fruit colour, fruit size, hue, saturation, intensity, contour lines, and blue-to-red fluorescence ratio (BRR_FRF).	
Tree Detection	Positive/negative histogram of oriented gradients (HOG), crown size, images of palm oil, built-up, bare land, water, and forest.	Genotype
Vegetative	Total dry matter, plant architecture, leaf length, number of leaves per plant, plant height, height increment (HT), and frond length (FL).	
Estimation	Nutrient content and plant height, shell thickness, mantled somaclonal variation together with gene networks for oil biosynthesis, drought and cold tolerance, in vitro regeneration potential, lipase activity, carotene and vitamin E contents, FA composition and iodine value, fruit shell thickness, fatty acid (FA) profile, fruit age, plant life, age factor, average value, daily CPO prices, monthly closing prices of oils, benefits, plant scale index, sex determination, inflorescence abortion, foliar nutrient composition, FFB yield, growth, and respiration.	
Disease detection	Bud and spear rot, sudden wilt, red ring disease and basal stem rot disease spectral reflectance, leaves and stem colours, freckle ( <i>Cercospora elaeidis</i> ), blast ( <i>Pythium splendens</i> and <i>Rhizoctonia lamellifera</i> ), vascular wilt ( <i>Fusarium oxysporum</i> f. sp. <i>elaeidis</i> ), and ganoderma trunk rot ( <i>Ganoderma</i> spp.).	
Others	Tree crowns and categorical features, thermal images, quantitative features, oil per palm, chlorophyll-sensitive wavelengths, and electrical properties of leaves.	
Breeding	DNA and tissues of the dura, pisifera, tenera; breeding-Deli, AVROS, Yangambi, Calabar, Ekona, Binga, and La Mé.	

The review results based on this review will be highly significant given that essential factors are incorporated into the prediction process via big data. However, more research is required to determine if big data can improve the prediction of palm oil performance.

#### 4. Data-Driven Prediction Model

Data and process-driven models are the most frequently employed algorithms in palm oil forecasting. Several parameters, such as initial conditions and agricultural variables, are required for process-driven models, which entail the extensive accuracy of their systems [9]. Nevertheless, the model exhibits multiple problems due to the uncertainty of breeding analysis and difficulty obtaining necessary data during the simulation. Meanwhile, empirical studies have reported proven success in using Data-Driven Models (DDMs) based on AI and ML [10]. The process requires a training dataset representing the system's behaviour to be fed into a machine-learning algorithm.

The data is then utilised to learn the relationship between the inputs and outputs. Eventually, the trained model can be tested using independent data to determine the generalisability of the new datasets. Accordingly, palm oil forecasts are more accurate if the optimum parameter level is acquired from previous data, though the correct features must first be selected. For instance, clustering can support data discovery and insight rather than relying on domain knowledge-based and unsupervised ML methods. Consequently, we can build a palm oil prediction model using ML, unsupervised and supervised, explained in the following sections.

##### 4.1. Prediction of Palm Oil Using Unsupervised ML

Conventional clustering methods are unsupervised, signifying the absence of outcome variables or relationships between the dataset observations [77]. Generally, it is easier to spot patterns and features when the palm oil data are clustered. There are numerous cluster approaches, e.g., hierarchical clustering, comprising Unweighted Pair Group Methods with Arithmetic Mean (UPGMA) and Neighbour-Joining (NJ). These two methods are widely utilised in genotype predictions of palm oil. The UPGMA is a unique high-dimensional data visualisation technique with a quick hierarchical clustering system, which provides a simple approach to constructing the distance matrix of the phylogenetic tree [78]. The method implicitly assumes a constant substitution rate with time and phylogenetic lines. Meanwhile, the NJ method is expressed as an algorithm involving knowledge of the distance between each pair of taxa to shape a tree [79]. The bottom-up (agglomerative) clustering approach creates phylogenetic trees based on DNA or protein sequence data [80]. A recent review of hierarchical clustering in palm oil genotype prediction proposed UPGMA and NJ to cluster the DNA of palm oil [81].

Bayesian network (BN) clustering is a prediction model where the division of items into subsets becomes a probability model parameter for the data. However, this model is subjected to presumptions and clustering references derived from post distribution properties. Notably, NJ and Bayesian methods were effective clustering and model-prediction tools. A study proposed these methods to predict the leaf genotype of the *Acrocomia aculeata* species [74]. Meanwhile, other clustering algorithms, such as k-means, are used in phenotype predictions of palm oil. K-means clustering is an unsupervised classification technique derived from signal processing. This model seeks to divide  $n$  observations into  $k$  clusters, where each observation is a cluster prototype of the cluster of the nearest mean.

Table 3 indicates that the clustering of phenotype and environment datasets is less, compared to genotype. This result demonstrates that k-means were applied less often to the phenotype dataset.

**Table 3.** Unsupervised prediction of palm oil (2016–2021).

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[71]	Breeding prediction	Three hundred twelve tenera palms, three palm oil breeding	S/F	NJ, chi-square, SNPs	Genomic Estimated Breeding Value (GEBV) S/F = 0.65%, 4425 SNP loci			/
[82]	Breeding prediction	Twenty-six leaves of <i>E. guineensis</i> palms	O/B, M/F	UPGMA, SSRs	Average polymorphic information content (PIC) = 0.7325, per centage of polymorphic loci = 83.3%			/
[83]	Breeding prediction	DxP cross produced 300 dura, 25 pisifera, 80 tenera, and 100 T×T progeny lines	BW, F/B, M/F; K/F, O/DM S/F, O/WM, O/B	Bayesian, SSRs	QTLs on seven linkage groups affecting F/B and O/B			/
[84]	Breeding prediction	Deli Dura and AVROS	F/B	NJ, PCA, AMOVA (Analysis of molecular variance)	230 alleles in 17 palm progenies			/
[85]	Breeding prediction	Eighty-one palm oil leaves	O/B, polyprenols. and dolichols	UPGMA	27 <i>E. guineensis</i> sites were grouped into the dura, pisifera, and tenera types			/
[86]	Yield prediction	Four hundred forty-seven blocks with a planted area of 19,809 ha	FFB, BUNCH_HA, ABW	Bayesian networks—ANN	FFB accuracy = 5%, BUNCH_HA accuracy = 85%, and ABW accuracy = 90%		/	
[87]	Breeding prediction	Two hundred and thirty-six oil palm leaf tissues	-	NJ Bayesian, SNPs	9% of SNP loci, 8 189 SNPs			/
[88]	Disease prediction	Thirty-nine oil palm trees	Soil moisture	k-Means-ANOVA (Analysis of variance)	Accuracy = 82%		/	
[89]	FFB ripeness prediction	Twenty-seven palm oils Tenera	FFB maturity levels	k-Means	Three ripeness centroids of 0, 1, and 2		/	
[76]	Breeding population prediction	Five hundred and fifty-three palm oil	-	UPGMA Bayesian network, DArTseq (Diversity Array Technology Sequencing), AMOVA	245 SNPs on all 16 chromosomes			/
[90]	Biomass palm oil boiler	Biopower boiler historical	Environment, deaerator, and economiser	Bayesian network	Probability of boiler = 54%	/		
[81]	Breeding prediction	Two hundred and fifty-one Dura cross Pisifera (DxP)	BN, BW, VIR, M/F, O/WM, O/B, S/F, PO, HT, and FL	NJUPGMA, SSRs, AMOVA	Shared alleles = 68.5% and specific alleles = 31.5%			/

\* Factor Category: E = environment; P = phenotype; G = genotype.

#### 4.2. Palm Oil Prediction Using Supervised ML

Function learning based on examples of input–output pairs is known as supervised ML. From a labelled set of training examples, it deduces the function from labelled training data, consisting of a set of training examples. Regression analyses, i.e., classification analysis and artificial intelligence, were used to analyse a historical dataset in predicting palm oil

yield [36,38]. The prediction models of palm oil growth were previously studied. Hence, this review will focus on these models and other methods such as Genetic Algorithm (GA), Naive Bayes (NB), Random Forest (RF), and Regression and Support Vector Machine (SVM).

#### 4.2.1. Types of Regression

Learning can be achieved using data and regression algorithms in an ML environment to minimise the observed loss or error. Regression is a supervised learning model to predict an output variable based on the known data variables. The objective of the regression process is to predict a continuous quantity from a set of input variables. In other words, a function that accounts for the observed quantity is approximated using relevant variables. Linear Regression (LR) and Multiple Linear Regression (MLR) are standard palm oil yield prediction algorithms. Accordingly, various complex regression algorithms were developed, such as support vector regression (SVR), pairwise regression (PR), principal component regression (PCR), and others.

A commonality among regression-based methods exists because the goal entails the determination of the optimum fit between the two variables. However, the most significant differences between the two methods are the type of image and the input variables used. The LR model depicts the link between one or more independent and dependent variable [91]. Meanwhile, multi-independent variables are more reliable than single ones using the MLR model [92,93], commonly conducted using the least-squares method (LSM).

Furthermore, the generalised cross-validation (GCV) index, which uses the MLR method to account for nonlinearity, can be used to analyse the MARS model [94,95]. A study proposed LR to predict the vegetative components of oil palms [96]. Similarly, Solichin [97] suggested a prediction method for LR, SVR, and MLP combined with a time series model and intelligent nonlinear model. Notably, this approach improves the error caused by time series analysis. Table 4 comprehensively illustrates regression predictions derived from previous studies.

**Table 4.** Regression prediction of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[98]	Crops yield prediction	Historical data	WOFOST crop model outputs, weather, crop areas, remote sensing, irrigated area, elevation, slope, soil, field size yield	Ridge regression kNN regression SVR GBDT regression	Wilcoxon <i>p</i> -value of 0.9	/	/	
[94]	Lard in palm olein oil prediction	Fourier-transform infrared (FTIR) data	Lard, pure palm olein oil, and the adulterated olein oil at 20% and 50% different temperatures	PLS SLR MLR	RMSE = 13.26 RMSE = 174.86 RMSE = 14.84	/		
[99]	Deforestation population prediction	Tree plantation in 2014	Environment and historical	LR	Accuracy = 98.25%	/		
[100]	Metisa plana population prediction	Twenty-five palm oils	Humidity, frond number	LR PR ANN	$R^2 = 0.01$ $R^2 = 0.01$ $R^2 = 0.45$		/	
[101]	Leaf nutrient content prediction	Leaf reflectance data	K, N, and Mg	PCR PLSR (Partial least squares regression)	$0.55 < R^2 < 0.94$ $0.86 < R^2 < 0.94$	/		
[102]	Drought prediction	Weather data Malaysian Meteorology Department and DID Malaysia (MMD)	Precipitation and temperature	SVR F-SVR BS-SVR	SPEI-1 RMSE = 0.644 SPEI-3 RMSE = 0.202 SPEI-6 RMSE = 0.187 SPEI-1 RMSE = 0.372 SPEI-3 RMSE = 0.159 SPEI-6 RMSE = 0.137 SPEI-1 RMSE = 0.626 SPEI-3 RMSE = 0.172 SPEI-6 RMSE = 0.146	/		
[103]	Bio-oil Production FFB prediction	Palm oil empty fruit bunch (OPEFB) data	Temperature, BW	LR	Error = 0.3%.	/		
[104]	Soil fertility prediction	Thirty-six samples of palm oil	Macronutrient soil	PCR	Accuracy = 91.67%	/		
[105]	FFB ripeness prediction	FFB palm oil thermal imaging data	Moisture, temperature, humidity	MLR	$R^2 = 0.8122$ .		/	
[106]	Soil nutrients prediction	Near-infrared spectroscopy data	Total organic carbon, total nitrogen, and soil pH	PLSR	RMSE = $0.44 \text{ g } 100 \text{ g}^{-1}$	/		
[107]	Dynamic Viscosity of MXene/palm Oil Nanofluid prediction	MXene/palm oil nanofluid data	Concentration and temperature	SVR-Grid Search	MAE = $7.9 \times 10^{-3}$	/		

\* Factor category: E = environment; P = phenotype; G = genotype.



#### 4.2.2. Support Vector Machine (SVM)

A support vector machine is a binary classification system that produces a linear hyperplane for data classification [108]. The hyperplane is intuitively separated, exhibiting the most significant distance to the closest data point in a reasonable margin. Generally, the larger the margin, the less the classification's generalised error. SVM differs from SVR, used to address regression issues through the SVR method [109]. The SVR algorithm aims to find a hyperplane in a space in N dimensions that defines the data points.

Concerning the palm oil growth point, a model of classification of SAR pictures is proposed for the L-band by hybridising SVM and CNN [110]. Subsequently, CNN removes unnecessary images in the first step, and SVM is then used as a preliminary basis for extracting plantation pixels. Chen and Liao [111] suggested a binary SVM classification algorithm using aerial UAV photos to identify the palm oil growth rate. Resultantly, SVM requires fewer samples but produces high prediction accuracy. Table 5 describes SVM prediction based on the previous studies.

**Table 5.** SVM prediction of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[23]	Grape content estimation	UAV and Sentinel-2 image data	NDVI, pH	SVM	$R^2 = 0.52 \pm 0.12$			
				Adaboost	$R^2 = 0.44 \pm 0.09$			
				RF	$R^2 = 0.41 \pm 0.09$			
				Decision Tree	$R^2 = 0.45 \pm 0.11$			
				Extra Trees	$R^2 = 0.43 \pm 0.08$			
				Huber Regression	$R^2 = 0.52 \pm 0.12$			
				OLS (Ordinary Least Square)	$R^2 = 0.51 \pm 0.09$			
				ARD (Automatic Relevance Determination)	$R^2 = 0.53 \pm 0.09$			
			Theil-Sen Regression	$R^2 = 0.51 \pm 0.12$				
[112]	Rice yield prediction	Sentinel-2 monthly image	Vegetative	SVM	MAPE = 4.4%			
				RF	MAPE = 4.5%			
				ANN	MAPE = 4.5%			
[113]	Wheat yield prediction	Satellite image data	Cropland information, crop yield data, satellite-based SIF data, vegetation index, climatic information	SVM	$R^2 = 0.75$			
				LASSO				
				RF				
			NN					
[114]	Nutrient deficiencies and leaf prediction	Three leaf images	Healthy, potassium deficiency, nitrogen deficiency, and magnesium deficiency	SVM-RBF	Accuracy = 100%			
				RBF-ANN	Accuracy = 97.92%			
[115]	Nitrogen status in mature palm oil prediction	Tenera from MPOB (Malaysian Palm Oil Board) data	Vegetative, NIR spectroscopy colours, soil	SVM	Accuracy = 81.82%			
[116]	Palm oil trees detection and enumeration	Images from palm oil tree data	Vegetative and nonvegetative	HOG (Histogram Of Oriented Gradient) – SVM	Accuracy = 99.21%			
[117]	Adulteration of palm oil fraud detection	Twenty samples of palm oil	NIR (Near-infrared) spectroscopy wavelength and different types of Sudan dyes	MSC-PCA (Principal Component Analysis) + SVM	Accuracy = 95.20%			
				LDA (Linear Discriminant Analysis)	Accuracy = 91.70%			

Table 5. Cont.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[118]	Female Inflorescences anthesis stages prediction	Anthesis thermal images	Temperature, humidity, rainfall	SVM	Exogenous + Endogenous	Accuracy = 40.45%	/	
				kNN (k-Nearest Neighbor)		Accuracy = 81.85%		
				RF		Accuracy = 88.5%		
[119]	Landsat Capability prediction	Geospatial Remote sensing data	SRTM (Vegetative and Shuttle Radar Topographic Mission), DEM (Digital Elevation Model)	SVM	Accuracy = 93.16%		/	
[120]	Bud root patterns	Aerial UAV (Unmanned Aerial Vehicle) images	Healthy and bud root presence	SVM	Accuracy = 93.53%		/	

\* Factor category: E = environment; P = phenotype; G = genotype.

#### 4.2.3. Random Forest (RF)

Random forest is widely used in agricultural research to examine organisms' spread and model ecosystem suitability [121,122]. It is a supervised learning algorithm commonly trained with bagging methods and an ensemble of the decision tree. The bagging approach is based on the premise that combining learning models improves the final result. Several studies have investigated the RF algorithm's potential for random agricultural forests [123,124]. The advantages of using this approach include resolving vector collinearity issues, often encountered while using traditional LR models.

Simultaneous and discrete variables could be used in the RF model, exhibiting an advantage over linear regression models [125]. Nevertheless, the RF variants present several drawbacks. For instance, the model predictions will over-ride other than the range of training outcomes. Hence, it can be challenging and unpredictable to estimate palm oil prediction due to the harsh environmental conditions. This restriction is necessary for RF regression in extrapolating the results and the absence of adequate dataset preparation. Furthermore, the constructed RF models presented a below-average yield. This phenomenon may be mitigated by the measurement number and the required training predictors. Recent studies demonstrated the potential of RF regression with long-term algometric variables to forecast palm oil. Table 6 presents RF predictions derived from previous studies.

Table 6. RF prediction of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[25]	Wheat yield prediction	Landsat data	Different regions of Punjab and Pakistan	ACO (Ant colony optimisation)-RF	201.27 < RMSE < 215.79	/		
				ACO-OSELM (Online-sequential extreme learning machine)	353.55 < RMSE < 381.57			
				ACO-ELM	352.46 < RMSE < 386.57			
[24]	Soybean yield prediction	Historical data	Water, early-season weed control, late-season weed control, season-long weed control, temperature, and crop management	RF	Variability = 88%	/		
				CART	Variability = 70%			
[126]	Cloud remote sensing tool	Satellite data	Spectral, environment, topographic	RF	Accuracy = 80.34%	/		
[127]	Commodity maps prediction across Indonesia	Landsat data	Vegetative, topographic, environment	RF	Accuracy = 95%	/		

Table 6. Cont.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[125]	Palm oil detection mapping	Landsat data	Spectral bands, SAR (Synthetic Aperture Radar) backscatter, vegetative, and texture	IGSO-RF (Improved Grid Search Optimisation)	Accuracy = 96.08%		/	
[128]	BSR diseases detection	ALOS PALSAR-2 images	Backscatter, types of polarisation	MLP (Multilayer Perceptron)	Accuracy = 95.65%		/	
				RF	Accuracy = 92.70%			

\* Factor category: E = environment; P = phenotype; G = genotype.

### 4.3. Deep Learning

Artificial Intelligence (AI) depicts the human brain structure through deep learning, a fundamental part of AI. The primary structure of deep learning, the neural network, is used to process hidden layers to improve learning. However, the role of AI in agriculture is vague, given that it is essentially a transformation of knowledge and labour. The primary reason is that the world is significantly dependent on the diversity of organisms. The agricultural sector plays the most crucial role in preserving biodiversity. The change or difficulty in agriculture's ecological balance directly impacts the human race and its ability to maintain the balance of the ecosystem. As such, the combination of ML and deep learning supports palm oil forecasting, which utilises two methods: artificial neural network (ANN) and time-series.

#### 4.3.1. Artificial Neural Network (ANN)

An ANN is an ML algorithm that can model the nonlinear dynamic input–output relationship [129], comprising three layers: an input layer, a hidden layer, and an output layer [130]. Several variables influence ANN success, including the number of nodes in the hidden layer, the learning intensity, and the training tolerance [95]. In a sequence of iterations, the study rate specifies the sum at which weight changes to obtain the expected value within a reasonable range of the observed value. The conventional ANN is a minimal local issue, whereby an optimised mechanism frequently stops in a local state rather than a global state.

The standard ML models frequently face difficulty in overfitting. The reverse and forward optimisation process, which maximises performance, is conducted within the back-propagation backward propagation learning algorithm. The removal of loss functions that could take place during the reverse propagation can be handled with effective activation functions such as sigmoid. Table 7 describes ANN predictions based on previous studies.

Table 7. ANN prediction of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[131]	Cotton yield prediction	Historical data	Drought index, precipitation, vegetative index	ANN	$R^2 > 0.80$		/	
[132]	Paddy yield prediction	Historical data	Cropland information, climate information, soil properties, agricultural production data, irrigation information	MLR-ANN	RMSE = 0.051, MAE = 0.041, R = 0.99		/	
[133]	Nutrient content prediction	Ninety leaf samples	Froned numbers and nutrients	ANN	Accuracy = 85.32%		/	
[134]	The ripeness of FFB prediction	Real-time FFB system data	Colour, texture, and thorn	ANN	Accuracy = 93%		/	
[135]	Metisa plana prediction	Twenty-five palms	LST (Land Surface Temperature), environment and NDVI (Normalised Difference Vegetation Index)	ANN	Accuracy = 95.42%		/	
[136]	Palm oil production planning	Historical data	FFB weight, harvesting time, yield	ANN	RMSE = 0.1290		/	
[137]	FFB maturity prediction	MPOB FFB images data	Colour features of palm oil FFB image	ANN	Accuracy = 94%		/	

\* Factor category: E = environment; P = phenotype; G = genotype.

### 4.3.2. Time-Series

A Recurrent Neural Network (RNN), or LSTM, is a feed-forward network with a backpropagation loop. LSTM provides additional benefits by holding onto the value of the previous output for short periods, serving as a small part of a network's memory that supports feedback analysis. Dynamic time-series models, such as the Autoregressive Integrated Moving Average (ARIMA), are the most widely used time-series approach. Other models, such as the Autoregressive (AR), Moving Average (MA), and Autoregressive Moving Average (ARMA), are subclasses of the ARIMA model. Box and Jenkins suggested a popular version of the ARIMA paradigm, the Seasonal ARIMA (SARIMA), for seasonal time-series forecasting. The ARIMA model's popularity is primarily due to the associated Box–Jenkins methodology for constructing the best model, in addition to its ability to represent a wide range of time-series effortlessly.

The complex variables and combinational inputs are the most common utilisation examples. Vegetative index, weather, and climatic data are used to track the progress of palm oils. Accordingly, LSTM is used to analyse the data for the attribute variables to better understand the time-series data. It yields an estimated 83 per cent of the crop [77]. It produces an estimated accuracy of 83 per cent of palm oil age detection using the Landsat time-series [77]. The LSTM model possesses two layers, one for processing genotypes and another for environmental factors. The two-layered approach provided a more effective process and straightforward suggestions with a value of 8.52 of RMSE [138].

The palm oil's satellite-based SIF was analysed to determine the vegetative index, exhibiting 0.69 of  $r^2$  [139]. The LSTM method combined price and time-series data to better estimate yield. The one feedback connection of the RNN can be directly processed by the one feedback connection data [140]. Furthermore, a low training time indicates that the processing time and storage space efficiency were due to this data type. The predicted palm oil yield suggested 2.7098% of MAPE. Table 8 describes the time-series forecasting based on previous literature.

**Table 8.** Time-series prediction of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[141]	Wheat yield prediction	Sentinel-2 image data	Minimum and maximum temperature, integrated solar radiation, cumulative precipitation, soil texture, soil chemical parameters, hydrological properties	LSTM	MRE (Mean relative error) = 9.70%		/	
[59]	CPO price prediction	One hundred thirty-six units of data of monthly prices	Temperature, rainfall, radiation	LSTM (Long short-term memory)	RMSE = 280.463		/	
[142]	CPO price prediction	Eighty-one complete sets of monthly observations	Monthly prices, total imports, and exports	ARIMA (Autoregressive Integrated Moving Average) ARAR (Autoregressive Autoregressive) ARFIMA (Autoregressive Fractional Integral Moving Average)	RMSE = 293,016.94 RMSE = 19,161.84 RMSE = 43,333.98		/	
[143]	Adverse side-effects of shaded agroforests prediction	GPS devise of plot data	Temperature, wind, drought, and soil erosion	Time series	95% confidence interval		/	
[144]	Age and biophysical qualities of yield prediction	MODIS (Moderate Resolution Imaging Spectroradiometer) data	Vegetative and yield	Time-series-CART (Classification And Regression Tree)	RMSE = 4.7		/	

\* Factor category: E = environment; P = phenotype; G = genotype.

#### 4.4. Other Approaches in Palm Oil Prediction

A hybrid solution incorporates more than one algorithm to boost efficiency. Integrating several ML algorithms will vastly increase the overall outcome of each algorithm by tuning, generalising, or adapting to new tasks. Table 9 shows other approaches to prediction derived from previous findings. Metaheuristic algorithms, such as the Genetic Algorithm (GA), can be used to solve a wide range of optimisation problems. The GA structures comprise two types based on the assumption of a finite solution space: local and global search [145]. The GA strings' populations, or chromosomes, are defined according to an optimisation problem.

**Table 9.** Other approaches' predictions of crops and palm oil.

Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances	Factor Category *		
						E	P	G
[27]	Chilli leaf disease	Images data	Types of chilli leaf diseases, up curl, down curl, Cer-cospora leaf spot, Geminivirus	SECNN (Squeeze-and-excitation-based convolutional neural network)	Accuracy = 99.54%			/
[26]	Wheat yield prediction	Historical data	Weather	CNN	RMSE = 0.66			/
				DNN	RMSE = 0.80			
				SVR	RMSE = 0.86			
				Regression tree	RMSE = 0.85			
				Ridge	RMSE = 0.91			
				Lasso	RMSE = 1.19			
				XGBoost	RMSE = 0.73			
				RF	RMSE = 0.79			
			kNN	RMSE = 0.88				
[21]	Corn yield prediction	Historical data	Time, corn variety, SOM (Soil organic matter) content, NFAR (Nitrogen fertiliser application rate), PFAR (Phosphorus fertiliser application rate), KFAR (Potassium fertiliser application rate), seeding rate	GBDT (Gradient boosting decision tree)	$R_{cv}^2 = 0.799$			/
				RF	$R_{cv}^2 = 0.749$			
[22]	Corn estimation	UAV image data	Different unique colors marked with different density	ResNet 18-CNN	0.73 < Accuracy < 0.97			/
[146]	Palm oil FFB production	Historical data	FFB (Fresh Fruit Bunch)	Exponential smoothing method	RMSE = 0.1			/
[147]	Optimisation yield production	Historical data	Environment	GA (Genetic Algorithm)	MSE = 0.022			/
[148]	Mapping palm oil plantation	Level-18 Google Earth images	Vegetative and impervious	RCANet (Residual Channel Attention Network)	Accuracy = 96.88%			/
[149]	FFB maturity prediction	106 FFB palm oil	Mesocarp colours and bunch grading	Lazy KStar	Accuracy = 63%			/
[150]	Predicting and diagnosing the quality of refined palm oil	Twenty-five samples of palm oil	FA (Fatty Acid), moisture content, and environment	PCorrA (Partial Correlation Analysis)	MSE < 0.01			/
[151]	CPO prediction	One hundred fifty-three months of CPO (Crude Palm Oil) production data	CPO	Exponential smoothing method	MAPE = 16.06			/
[152]	Classify BSR (Basal Stem Root) prediction	Eighty oil palm trees.	FronD number, frond angle, crown area, and crown significance	NB (Naïve Bayes)	Accuracy = 85%			/



Table 9. Cont.

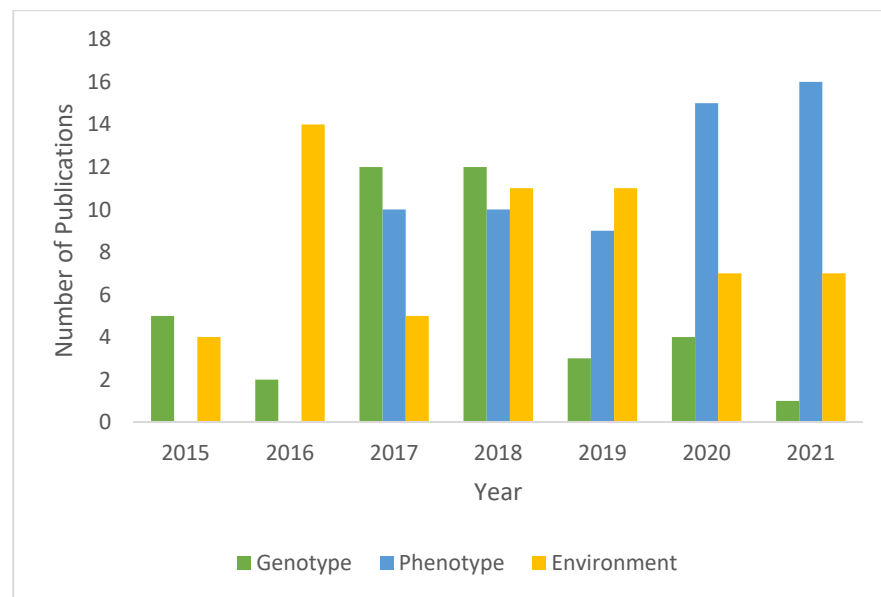
Author(s)	Objective	Dataset	Feature	Methods	Relevant Findings/Performances		Factor Category *		
							E	P	G
[153]	Classification of nutrient contents on the leaf	Types of fronds with different ages data	Ca, K, Mg, N, and P	LMT-SMOTE (Logistic Model Tree (LMT)-Synthetic Minority Over-sampling Technique (SMOTE)) + AdaBoost	Accuracy = 76.13 100.00%		/		
[154]	Movement CPO prediction	Data from the Malaysian palm oil industry in Malaysia	Monthly prices, total imports, and exports	RBFNN (Radial Basis Function Neural Network)-2SATRAAIS (Satisfiability Reverse Analysis with Artificial Immune System Algorithm)	Accuracy = 90.46%		/		
[155]	Palm oil detection mapping	Landsat data	Vegetation, humidity (water)	XGBoost	RMSE = 0.1512		/		
				LASSO (LASSO regression)	RMSE = 0.3487				
				RPART (Recursive Partitioning and Regression Trees)	RMSE = 0.1894				
				RF	RMSE = 0.1655				
			NN (Neural Network)	RMSE = 0.3925					
[156]	Palm oil detection	Aerial UAV images	Crown size, crown colour, and crown density	Faster-RCNN	ResNet 50	Precision = 96.34%	/		
					VGG-16	Precision = 95.1%			

\* Factor category: E = environment; P = phenotype; G = genotype.

A study employed GA to improve the generalisation ability of the Radial Basis Function Neural Network (RBFNN and Fuzzy Radial Basis Function Neural Network (FRBFNN) [157]. Meanwhile, Ibrahim et al. [158] proposed a prediction method of fatty acid and an intelligent nonlinear model to improve the error. Ishola et al. [159] suggested an investigation to determine the palm oil kernel process by improving the error of response surface methodology (RSM). This method includes the adaptive neuro-fuzzy inference system (ANFIS) accuracy when combined with GA.

## 5. Analysis and Discussion

The data-driven projects of various categories in a single dataset are a difficult task; thus, they are frequently overlooked. The issue with these projects includes the differing data complexity and its update frequency; hence, distinct preprocessing methods will be used to clean up the data because of the frequency difference. Various factors must be considered in future research. Table 2 shows how various critical factors, such as environment (E), phenotype (P), and genotype (G), can be divided into different categories. These factors presented evidence of unfinished work based on the literature. As of 2011, researchers have been focusing on these factors, which are used to summarise trends between 2015 and 2021, as shown in Figure 4.



**Figure 4.** The trend of using categorical factors between 2015 and 2021.

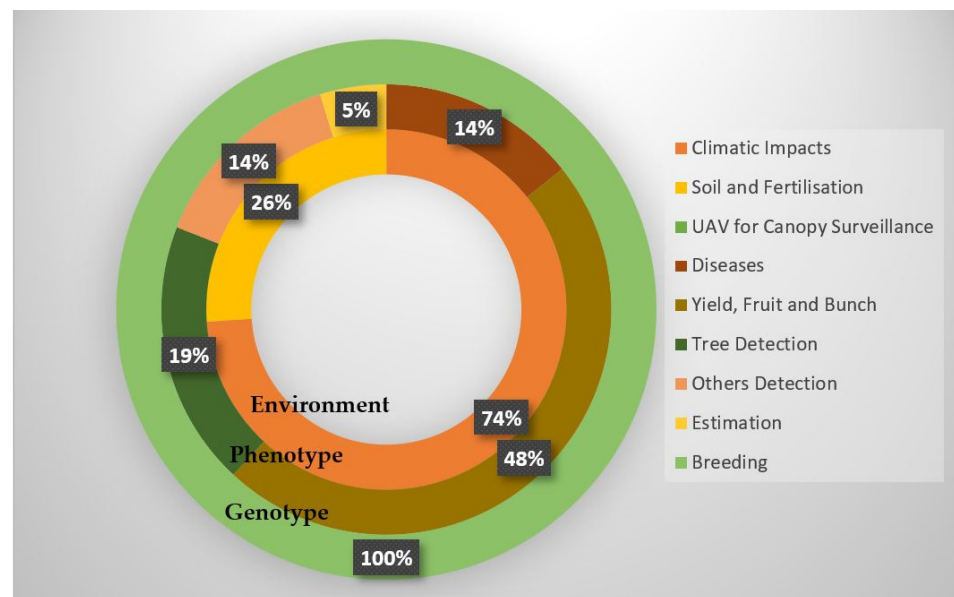
Numerous factors, including genetics, climate, and disease detection, were found to affect palm oil yield in this study. The term “phenotypes” refers to the traits that can be seen and measured in a breeding program, such as increased genetic variation in palm oil yield. Another factor that affected the study’s results was the selection of reliable data. Genotype, phenotype, and environmental data from oil palm crops are all difficult to analyse because of the high level of complexity. Most researchers use genotype data to gather breed-specific DNA for selection purposes. Additional data such as bunch characteristics, yield component, and vegetative measurement can be used for relative selection in oil palm. Environmental and historical data are interdependent. Data from both sources can be used to improve the process and save money.

Figure 5 shows most of the data features from previous studies utilised to predict palm oil. Seventy-four per cent of studies that rely on the climate and soil and fertilisation (twenty-six per cent) are in the environment category. However, it is impossible to indicate the most effective subfeatures for phenotype factors under each feature group. For instance, there are nine subfeatures listed under “phenotype factors” which are utilised extensively, e.g., bunch, fruit, and diseases. Nevertheless, these studies did not specify the optimal set of sub-features under the bunch and yield of palm oil, requiring further research.

The ideal phenotype subfeatures based on various regions should also be investigated. Because of the wide range of existing factors, such as diseases, tree detection, and physical characteristics of palm oil yield, predictions in similar fields should be made over several years [45,105]. Palm oil yield predictions in similar fields should be recorded over several years due to variations in existing factors such as diseases, tree detection, and physical characteristics [45,105]. Palm oil forecasting does not emphasise phenotype data similarly to genotype representation; therefore, redundant features must be eliminated to ensure the predictive model accuracy.

The insufficient data due to the frequency update would hinder the prediction process. Specific data are updated every minute, hour, or day, depending on the volume. However, sporadic data limit the finding of patterns, which occurs when data are collected only once a month. This phenomenon can also occur in the event of substantial missing data and when dealing with small-sized data. Hence, larger-sized datasets are the suggested means of solving this issue. Oil palm phenotype data are challenging to come by due to a lack of available information. Only a few factors reviewed thus far have fully utilised all three, even though this paper has learned that climates under the meteorological factor are critical. However, when it comes to the category of factors, there have been very few that have used

all of them. Besides, genotype environmental and historical information can be found in general datasets, whereas phenotype data come from the statutory body of the Malaysian Palm Oil Board (MPOB) or Felra Global Ventures (FGV), which have their own procedures. Because of the high cost of capturing data in the field of study, phenotypic data are difficult to communicate. Data storage and analysis tools, phenotype reduction, and other laboratory costs are among the challenges to developing an optimal digital phenotype platform [160]. As a result of these concerns, several gaps must be addressed, such as dealing with a wide range of data sizes and implementing the appropriate data-driven method.



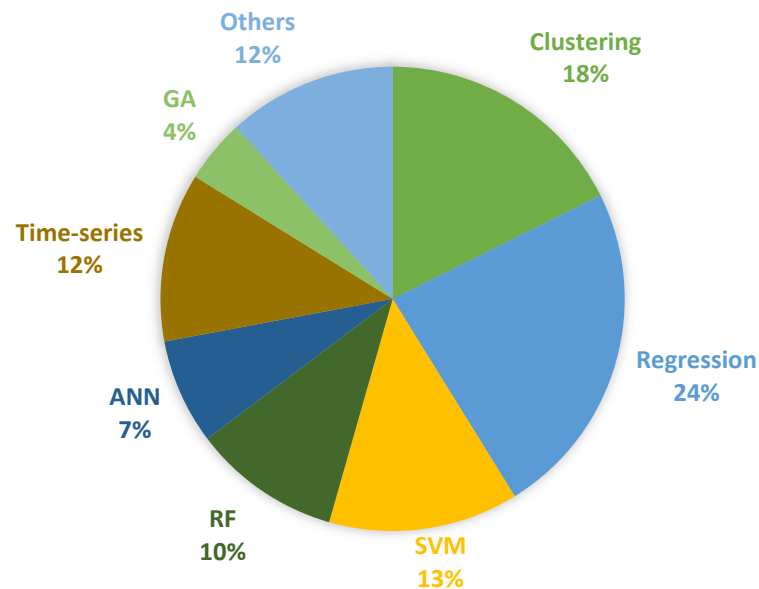
**Figure 5.** Factors category.

When parameter selection is required, techniques for managing and conducting data experiments become complex. In addition, parameter tuning must be performed. More experiments conducted on the data yield superior results. The difficulty in developing techniques for future research is minimising the impact of data inaccuracies within datasets. During testing and training, the first step recommended by this study is to identify and discard low-quality data points [86]. Due to the random selection of features, varying quantities of features deemed too high or too low may result in model-fitting issues and performance fluctuations. According to previous studies, the randomness in the dataset has caused the model to be overfit and forecasting errors to be made [115,126]. Another possible cause of forecasting failure is a lack of suitable feature selection methods for ML and DL.

Another challenge is model selection, which requires data. Accuracy in oil palm breeding models is achieved through a hybrid method in this study. A model selection problem is presented because it transitions from traditional to modern model [147]. Rather than using inefficient and time-consuming traditional modelling methods to optimise crop yields, farmers are now turning to AI models that are more accurate and efficient. Oil palm model selection is an example of how knowledge can be applied in various ways. Evaluating the impact of different parameters on oil palm yields is possible. The artificial intelligence (AI) method is a form of optimisation based on natural selection and inherited principles [159].

Previous studies have employed various classification and regression algorithms to predict palm oil. Figure 6 shows the prediction algorithms based on the previous findings. The extracted data indicate that regression (24 per cent) is the most used palm oil prediction algorithm, followed by clustering, SVM, time-series, and RF. The second most popular algorithms for prediction the clustering approaches, i.e., NJ, UPGMA, and Bayesian network, focusing on the breeding palm oil prediction. Meanwhile, K-means

perceive behaviour on nonlinear variables, though more studies are required to test this claim. The third most preferred algorithm for prediction is SVM, though several flaws must be resolved. The SVM results were more significant than the MSC-PCA hybrid with SVM, SVM-RBF, and SVM-ANN [104,107]. This result implies that the increased SVM performance is attributed to the improved optimisation techniques for a wide range of parameters [114].



**Figure 6.** Widely used palm oil prediction algorithms.

The SVM provides additional kernel functionality, improving the prediction models through feature recognition. This model possesses a more compact management potential than ANN, specifically in its distribution, geometry, and data overflow. Furthermore, the SVM is theorised based on the principle of reducing structural instability, diminishing the high bond error instead of training error [115]. Moreover, this study found various issues in ANN related to DL algorithms. For example, the ANN output involves several hidden layers and types of activation functions. Alternatively, a different optimisation algorithm can be utilised, optimising hidden layers and the ideal activation function, i.e., a genetic algorithm [161,162]. This substitute procedure may not be the highest performance substitute. The terms “much used” and “best output” are not similar definitions (See Figure 6). The regression and time-series predictably exceeded the ML and DL techniques as per the anticipated palm oil observations. The parameters and goal parameters could be isolated from regression and time-series [147,163].

To date, there have been very few investigations into the RF algorithm’s potential for classification and regression analysis in agriculture. Palm oil yield estimation may be difficult and unreliable due to varying environmental conditions in different fields. This constraint may be critical in extrapolating the results for RF regression as well as a result of the lack of sufficient training datasets. The developed RF models overestimated the average yield and underestimated the yield below it. As the number of observations and appropriate predictors for training increase, this issue may be mitigated. According to this study, long-term agrometric variables can be predicted using RF regression. It is clear that the ML and DL techniques, as well as LASSO’s ability to isolate dynamic correlations between the variables and the target predictor, outperformed LR in palm oil yield prediction [155].

Both clustering and classifier frameworks are used in the selected articles. Since images are used for clustering in certain articles, the article utilises a numerical dataset in conjunction with machine vision rather than ML. The use of clustering architectures may be investigated in depth in order to identify various opportunities regarding this issue. In time

series prediction, determining the window size, lag value, and even the features selected as input for the prediction algorithm is standard practice (i.e., LSTM). It is also common for LSTM to use past data to learn from all the features or input it receives [77]. Due to the prevalence of LSTM algorithms in palm oil forecasting, a focus on DL architectures should be investigated. All DL architectures such as ANN and LSTM were found to be the most frequently used in our review articles that used DL approaches. However, LSTM–CART hybrid DL algorithms [140,144] have also been applied to this problem.

In previous crop and palm oil yield prediction research, a wide variety of performance evaluation metrics have been employed. Accuracy is the most utilised metric in the algorithm for predicting crop and palm oil yield. When palm oil prediction models are evaluated using different performance metrics, it is nearly impossible to compare them. A standard and systematic approach, or one single measure, should be used to quantify the model's palm oil prediction accuracy. It is possible to compare crops and palm oil prediction algorithms if the performance metrics used are the same for all of the algorithms.

Only Hilal et al. [147] used a maximum number of features under the historical yield data, cropland information, and climatic information groups in other previous studies related to palm oil. Palm oil yield has been predicted using only one feature in other studies [30,56,110,140,151]. Using a large number of features in a crop yield prediction model may be more accurate than using a smaller number of features. Studies, such as [98], show that by combining information from WOFOST crop model outputs data, weather data, crop data, soil data, and yield statistics, they were able to accurately predict soft wheat, spring barley, sunflower, grain maize, sugar beets, and potato yields. Data on water, early-season weed control, late-season weed control, season-long weed control, temperature, and crop management were used to predict soybean yield by Landau et al. [24]. Soil properties, different types of fertiliser application rate, and yield data from historical yields were used to predict corn yields by Du et al. [21]. As a result, the palm oil yield prediction research should also take into account a large number of features groups.

The efficiency with which palm trees absorb nutrients is directly related to their soil properties. Nutrients like these have a significant impact on palm yield effectiveness [38]. It is possible to estimate palm oil yield by looking at a wide range of soil properties, such as soil electrical conductivity and conductivity to clay and silt, organic carbon content, pH and cation exchange capacity, bulk density, and the percentage of clay, silt, and sand in the soil. The use of remote sensing to examine soil properties in a palm grove could be a promising avenue for future investigation. It is possible to monitor the soil's characteristics from afar because the reflected light has different responses. For example, soil moisture content may be taken into consideration when determining the dielectric characteristics of the spreading wave. It is possible to investigate the relationships between recorded signals and a variety of soil characteristics, including soil texture, soil form, and overall soil structure, in order to develop analytical relationships.

However, the estimated parameters, such as soil moisture, LAI, greenness of the palm canopy, and height, that contribute significantly to the yield prediction should be taken into account. A precise yield forecasting model can be created once the connection is defined. Including historical yield data, various climatic data, vegetation data, fertiliser application information, and other types of satellite data could improve yield forecasting performance. Remote sensing makes it possible to quickly gather data on the ground from a large area. For example, it has been used to monitor the environment and topographic conditions [98,126]. It collects spatial data without any direct contact. According to an increase in demand for oil palm products, numerous studies have been conducted on oil palm plantation mapping using remote sensing.

Mapping a palm plantation accurately can have significant economic and environmental advantages. Traditional ML techniques, classical image processing methods, and deep learning methods can all be used to identify the crowns of trees. RF [125,126] and SVM [111,116] are two of the most commonly used classifiers for tree crown detection in traditional ML approaches. When compared to traditional image processing methods,



ML techniques have made significant progress. Image binarization and segmentation are commonly obtained using high-resolution UAV images such as local maximum filter [111]. However, complex situations like overlapping tree crowns may degrade detection results due to the method's unnecessary labels. Researchers using remote sensing images have increasingly turned to DL-based classifiers, which employ multiscale computational methods, in recent studies to detect trees from satellite images; the most advanced studies use deep learning-based classification [144]. Features can be extracted using DL, which is known for its impressive capacity. With Faster RCNN [156], advances in object detection based on pre-trained models have been made to detect palm trees in the plantation area. It is possible to improve the predictive accuracy of classifiers by optimising their input hyper-parameters, which are highly dependent on their accuracy. In order to improve accuracy, hyperparameter optimisation must be used. Its accuracy was superior to other classifiers, such as SVM, CART, NN, and IGSO-RF classifiers that used the IGSO algorithm to fine-tune the parameters of the traditional RF model [125].

Disease image recognition plays a critical role in the development of innovative agriculture. Oil palm diseases have been studied using a wide variety of diagnostic methods. Categories such as SVM [120], as well as ANN [135] and NB [152], are included. In this regard, traditional ML methods have some drawbacks when it comes to monitoring palm oil disease. Furthermore, existing methods are heavily reliant on disease images that are of high quality. As a result of these methods, which include image preprocessing, image segmentation, feature extraction, and classification, as well as significant operations that add complexity and delay implementation, the implementation can be significantly delayed. Training is challenging using traditional ML methods, especially when the training dataset is large. Deep learning and transfer learning, two of the most recent advanced ML techniques, have the potential to aid in the development of oil palm disease recognition.

Oil palm plantations require a lot of monitoring, which takes time and effort. RF [118], ANN [137], and hybrid CNN-SVM [110] are just a few of the ML approaches that have been used to track palm growth via remote sensing in the past. Nevertheless, landowners must extract useful information from remote sensing data. There is a chance that this will lead to solutions via deep learning, transfer learning, and recognition of objects.

Palm oil yield prediction is only indirectly addressed in a few ML-based studies conducted in the industry. So far, there have been five studies examining how well scientists can predict phenotype palm oil yields [56,87,141,150,151]. In light of these studies, it is difficult to determine which algorithm is best for predicting palm oil yields. It is possible to predict palm oil yield with the help of some regression algorithms based on ML (ML). Multiple algorithms should be considered instead of a single algorithm to improve prediction model robustness.

It is a common practice in the palm oil industry to employ SVM models and algorithms. Perhaps the most critical part of this success is regularisation. Outliers and noise are common in agricultural data, making it challenging to analyse. The classifier's generalisation capabilities may be improved through regularisation, which could help solve this issue [114]. The regularised linear SVM classifier outperformed the RBF classifier in this study. For example, the regularisation parameter  $C$  and the RBF width using kernel 2 require manual adjustment in SVM. These benefits occurred as a result of the low execution speed. In addition, because the SVM decision rule is a simple linear function in the kernel space, it is stable and has low variance [114]. It is critical to obtain low variability in agricultural data because features are highly dynamic and change over time. SVM's ability to withstand the curse of dimensionality may be the final possible explanation for the problem. Small training sets and high-dimensional feature vectors have allowed SVM to achieve excellent results [114]. In terms of finding a nonlinear pattern in data, the key advantages of RF are its generalisation performance, high speed, and the use of an ensemble of tree-structured classifiers. It has low data preprocessing requirements in training because it is robust regarding to unit differences and can make accurate predictions on

sparsely annotated data. The RF algorithm yields better results in the selection of features. Meanwhile, the input hyperparameters play a vital role in RF.

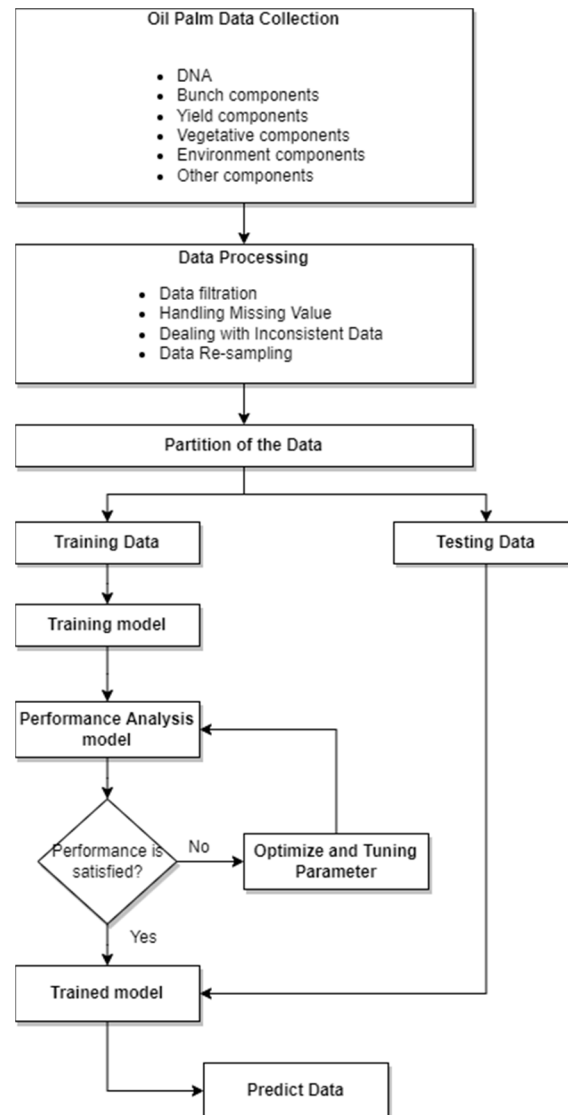
High-dimensional feature vectors can benefit from SVM, one of the best classification algorithms. To solve the problem of high dimensionality, dynamic classifiers can also consider a sequence of feature vectors rather than just one high-dimensional feature vector. Another benefit of using these methods is that they can deal with large datasets. A combination of classifiers may be used to solve this problem because it reduces the variance, especially when there is no stationarity in the dataset. This method is still useful, but it may be outperformed by a classifier that uses a combination of RBF and SVM. Due to the limited number of training datasets, simpler techniques such as LDA may be used. The k-NN algorithms are rarely employed in the palm oil industry due to the curse of dimension. However, k-NN may be superior if only low-dimensional feature vectors are taken into account in agriculture and the palm oil industry, or if no feature vectors are needed. In remote sensing, the ANN is widely used to predict vegetation parameters and crop yields because it can retrieve complex, dynamic, and non-linear patterns from the data. Many libraries and software tools are readily available, making them the most basic form of ML. A few drawbacks exist in practical applications, such as the rate at which neurons are learned, how many neurons are chosen for the hidden layers, and the overfitting issue when using a large training dataset. Large datasets cause the process to slow down. As the number of epochs required increases, backpropagation networks become slower in training. Using a CNN for image data with a large training set is a good option. A leaf image can identify a disease, map an oil palm plantation using remote sensing-based spectral images, and so on. The use of advanced CNN architectures, such as Faster R-CNN, is also highly effective for object recognition-based tasks. It is possible to predict palm oil yield and oil palm price using regression-based algorithms such as RF, ANN, and SVR. In addition, to improve the predictability of the model, an ensemble of multiple algorithms should be investigated rather than a single algorithm.

In DL, even though LSTM appears to be the best method in general, LSTM has primarily been applied to data pertaining to palm oil age detection, environment, as well as price prediction. Palm oil yield requires additional research. It is also worth noting that there are several possible reasons for the LSTM's fluctuating performance, including the number of indicators used and the method's limitation. According to [59,77], while the sequence-to-sequence architecture of LSTM can store long-term memories, its internal representation is limited to a fixed length. Internal representations or input sequences of LSTMs have a fixed length because information is divided into small pieces for easier recall. Various random weight initialisations impact LSTMs, and as a result, they behave in a manner similar to that of a feed-forward neural network. As a result, many unresolved issues must be addressed to improve or resolve the issue at hand. LSTM feature engineering and a basic grasp of the phenotypic factors of palm oil data can enhance performance in this area. The current performance of LSTM is expected to improve with additional parameter tuning and additional learning methods.

The DL is a subset of ML, commonly believed to be significantly persuasive for palm oil prediction. However, the only difference between ML and DL is that the latter is inefficient for a limited training dataset; thus, fundamental components are automatically removed from this dataset. The attributes of other samples must be manually removed to preclude the use of successful DL. Hence, detailed studies must be conducted on the use of DL strategies in palm oil, and there is still room to delve into the performance of algorithms. The research has shown that DL usage in forecast time-series is superior to basic ML in average performance; hence, future studies can focus on this idea. ML data-driven models were unable to extract features from multifactor timing data efficiently. According to this study, most models did not reflect the data's phenotypic characteristics. Time-series has shown to outperform other forecasting methods with minor forecasting errors.

## 6. Identifying Palm Oil Prediction Framework

In the light of previous research on palm oil prediction, we have proposed a framework for predicting palm oil yields in the future. Figure 7 depicts the prediction of palm oil yield as a prospective framework. In an effort to accurately predict palm oil, a diverse range of data must be gathered, including genotype, breeding, fruit, yield, and bunch. Others include soil properties, climate, vegetation indexes, diseases, previous yield records, fertilisation, and UAV monitoring data. Furthermore, data must be preprocessed after collection to further analyse the information. Once the data has been preprocessed, the entire dataset is divided into a training and testing set.



**Figure 7.** Identifying palm oil prediction flowchart.

A training dataset is used to build a prediction model, which is trained using a variety of ML-based regression and classification algorithms. Moreover, parameter optimisation is used to improve trained models when they fail to meet expectations, and the testing of the models is conducted after achieving the required performance. The critical factors affecting palm oil prediction include disease recognition and management, fruit and bunch, palm oil breeding, growth, and nutrition level monitoring. Accordingly, this study combines the prediction model output with palm oil variables to accurately predict palm oil yield.

## 7. Conclusions

This study examined existing research on ML in palm oil conducted over the last decade, categorised into unsupervised, supervised, and deep learning analyses. The critical objective of this study aimed to identify the breeding prediction of oil palms derived from the clustering category. Furthermore, the study incorporated regression analyses, including determining FFB yield, tree detection, and climate. The findings from the research were primarily focused on tree detection, the observation of palm oil plantation areas, and identifying the ripeness of fruit bunches. However, models for forecasting disease were not widely used, i.e., in fruit yield, oil production, and breeding.

Notably, SVM, RF, LR, and SVR are the most promising conventional machine learning architectures. Additionally, DL models such as LSTM and ANN are used to estimate crop yields in the field. This selection has been made based on most of the algorithms used in previous studies, hybrid modelling, comparison, and performance evaluation. Some feature selection algorithms and performing ensemble methods identify and treat missing and outliers' values, computational complexity, and model fit, as well as outperforming models, and should be examined comprehensively in order to determine the best model. Research on palm oil yield prediction is scarce, according to the findings of the current review. In addition, very few feature sets and ensemble methods have been used in the existing palm oil yield prediction studies, resulting in large discrepancies between the predicted and actual palm oil yield. When it comes to palm oil yield predictions, it is too early to speculate on the best set of features, models to use, and the appropriate ensemble models. As a result, more studies involving a wide range of features and prediction algorithms are needed. Extensive research on crop yield prediction and palm oil yield prediction is expected to be based on this paper.

Nevertheless, most of the data used in palm oil research were climatic from various sources. For example, identifying suitable land requires various factors that must be considered, e.g., soil classification, automated pest and weed detection, and optimising fertiliser use. This list of factors can be extended to identifying the symptoms of sunlight and water limitations in palm oil crops and seed assessment.

The findings from this review highlight current palm oil and ML research from various angles. Innovative ML practices, such as big data analytics and automated information extraction, support the advancement of knowledge-based palm oil. However, research trends indicate that existing MLAs and techniques were not adequately coupled to support effective decision-making systems compared to other ML application domains. Thus, the current research is insufficient to design practical tools that increase yields, quality, and plantation sustainability. A vivid idea could be presented on how ML can enhance palm oil development and inspire researchers to find relevant solutions in this area. Overall, this article may support the palm oil industry's automation and intelligence development. However, future work should include a technical review of other ML models used in palm oil agriculture.

**Author Contributions:** Conceptualization, F.N.M.N. and N.H.A.H.M.; writing—original draft preparation, F.N.M.N.; visualization, F.N.M.N.; writing—review and editing, F.N.M.N., N.H.A.H.M., R.A., M.F.A.R., M.A.A.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 was 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. Cosiaux, A.; Gardiner, L.M.; Stauffer, F.W.; Bachman, S.P.; Sonké, B.; Baker, W.J.; Couvreur, T.L.P. Low extinction risk for an important plant resource: Conservation assessments of continental African palms (Arecaceae/Palmae). *Biol. Conserv.* **2018**, *221*, 323–333. [[CrossRef](#)]
2. Corley, R.H.V.; Tinker, P.B. Selection and Breeding. In *The Oil Palm*; John Wiley & Sons: Hoboken, NJ, USA, 2015; pp. 138–207. [[CrossRef](#)]
3. Reddy, M.T.; Kalpana, M.; Sivaraj, N.; Kamala, V.; Pandravada, S.R.; Sunil, N. Indigenous Traditional Knowledge on Health and Equitable Benefits of Oil Palm (*Elaeis spp.*). *OALib* **2019**, *6*, 90022. [[CrossRef](#)]
4. Okolo, C.C.; Okolo, E.C.; Nnadi, A.L.; Obikwelu, F.E.; Obalum, S.E.; Igwe, C.A. The oil palm (*Elaeis guineensis* Jacq): Nature's ecological endowment to eastern Nigeria. *Agro-Sci.* **2019**, *18*, 48–57. [[CrossRef](#)]
5. 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.
6. Razak, M.Y. Bin A. In *Selected Agricultural Indicators, Malaysia, 2021*; Department of Statistics Malaysia: Kuala Lumpur, Malaysia, 2021.
7. Mark, B.; Steve, J.; Will, S.; Richard, S.; Sam, R.; James, F.; Yu, L.K.; Julian, M. *Study on the Environmental Impact of Palm Oil Consumption and on Existing Sustainability Standards*; Publications Office of the European Union: Luxembourg, 2018; ISBN 978-927-980-226-3.
8. Zabid, M.F.M.; Abidin, N.Z.; Applanaidu, S.D. MYPOBDEX: An interactive decision support system for palm-based biodiesel investors. *Int. J. Econ. Perspect.* **2017**, *11*, 260–272.
9. Mohamad Fauzi, N.S.; Abd Rahim, M.F.; Hj Mohamad, M.N. Implementation of information system in oil palm breeding research: FGV's experiences. *SSRG Int. J. Eng. Trends Technol.* **2020**, 104–108. [[CrossRef](#)]
10. Rawi, R.; Hasnan, M.S.I.; Sajak, A.A.B. Palm Oil Soil Monitoring System for Smart Agriculture. *Int. J. Integr. Eng.* **2020**, *12*, 189–199. [[CrossRef](#)]
11. Ang, K.L.M.; Seng, J.K.P. Big data and machine learning with hyperspectral information in agriculture. *IEEE Access* **2021**, *9*, 36699–36718. [[CrossRef](#)]
12. Chi, M.; Plaza, A.; Benediktsson, J.A.; Sun, Z.; Shen, J.; Zhu, Y. Big Data for Remote Sensing: Challenges and Opportunities. *Proc. IEEE* **2016**, *104*, 2207–2219. [[CrossRef](#)]
13. Dakir, A.; Barramou, F.; Alami, O.B. Opportunities for Artificial Intelligence in Precision Agriculture Using Satellite Remote Sensing. In *Advances in Science, Technology and Innovation*; Springer: Berlin/Heidelberg, Germany, 2022.
14. Zhang, X.; Zhou, Y.; Luo, J. Deep learning for processing and analysis of remote sensing big data: A technical review. *Big Earth Data* **2021**, *5*, 1964879. [[CrossRef](#)]
15. Balasubramanian, V.N.; Guo, W.; Chandra, A.L.; Desai, S.V. Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. *Adv. Comput. Commun.* **2020**. [[CrossRef](#)]
16. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
17. Morota, G.; Ventura, R.V.; Silva, F.F.; Koyama, M.; Fernando, S.C. Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *J. Anim. Sci.* **2018**, *96*, 1540–1550. [[CrossRef](#)] [[PubMed](#)]
18. Antle, J.M.; Jones, J.W.; Rosenzweig, C.E. Next generation agricultural system data, models and knowledge products: Introduction. *Agric. Syst.* **2017**, *155*, 186–190. [[CrossRef](#)]
19. Coble, K.H.; Mishra, A.K.; Ferrell, S.; Griffin, T. Big data in agriculture: A challenge for the future. *Appl. Econ. Perspect. Policy* **2018**, *40*, 79–96. [[CrossRef](#)]
20. Gopal, M.P.S.; Chintala, B.R. Big data challenges and opportunities in agriculture. *Int. J. Agric. Environ. Inf. Syst.* **2020**, *11*, 48–66. [[CrossRef](#)]
21. Du, Z.; Yang, L.; Zhang, D.; Cui, T.; He, X.; Xiao, T.; Xie, C.; Li, H. Corn variable-rate seeding decision based on gradient boosting decision tree model. *Comput. Electron. Agric.* **2022**, *198*, 107025. [[CrossRef](#)]
22. Vong, C.N.; Conway, L.S.; Feng, A.; Zhou, J.; Kitchen, N.R.; Sudduth, K.A. Corn emergence uniformity estimation and mapping using UAV imagery and deep learning. *Comput. Electron. Agric.* **2022**, *198*, 107008. [[CrossRef](#)]
23. Kasimati, A.; Espejo-García, B.; Darra, N.; Fountas, S. Predicting Grape Sugar Content under Quality Attributes Using Normalized Difference Vegetation Index Data and Automated Machine Learning. *Sensors* **2022**, *22*, 3249. [[CrossRef](#)]
24. Landau, C.A.; Hager, A.G.; Williams, M.M. Deteriorating weed control and variable weather portends greater soybean yield losses in the future. *Sci. Total Environ.* **2022**, *830*, 154764. [[CrossRef](#)]
25. Ali, M.; Deo, R.C.; Xiang, Y.; Prasad, R.; Li, J.; Farooque, A.; Mundher Yaseen, Z.; Mara, T.; Al-Khawarizmi, K.; Alam, S. Coupled online sequential extreme learning machine model with ant colony optimization algorithm for wheat yield prediction. *Sci. Rep.* **2022**, *12*, 5488. [[CrossRef](#)] [[PubMed](#)]
26. Srivastava, A.K.; Safaei, N.; Khaki, S.; Lopez, G.; Zeng, W.; Ewert, F.; Gaiser, T.; Rahimi, J. Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Sci. Rep.* **2022**, *12*, 3215. [[CrossRef](#)] [[PubMed](#)]



27. Naik, B.N.; Malmathanraj, R.; Palanisamy, P. Detection and classification of chilli leaf disease using a squeeze-and-excitation-based CNN model. *Ecol. Inform.* **2022**, *69*, 101663. [[CrossRef](#)]
28. Yan, J.; Tian, H.; Wang, S.; Wang, Z.; Xu, H. Paddy moisture on-line detection based on ensemble preprocessing and modeling for combine harvester. *Comput. Electron. Agric.* **2022**, *198*, 107050. [[CrossRef](#)]
29. Santoso, H.; Tani, H.; Wang, X. Random Forest classification model of basal stem rot disease caused by *Ganoderma boninense* in oil palm plantations. *Int. J. Remote Sens.* **2017**, *38*, 4683–4699. [[CrossRef](#)]
30. Zheng, J.; Li, W.; Xia, M.; Dong, R.; Fu, H.; Yuan, S. Large-Scale Oil Palm Tree Detection from High-Resolution Remote Sensing Images Using Faster-RCNN. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 28 July–2 August 2019.
31. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* **2017**, *9*, 22. [[CrossRef](#)]
32. van Dijk, M.; Morley, T.; Jongeneel, R.; van Ittersum, M.; Reidsma, P.; Ruben, R. Disentangling agronomic and economic yield gaps: An integrated framework and application. *Agric. Syst.* **2017**, *154*, 90–99. [[CrossRef](#)]
33. Khalili, E.; Kouchaki, S.; Ramazi, S.; Ghanati, F. Machine Learning Techniques for Soybean Charcoal Rot Disease Prediction. *Front. Plant Sci.* **2020**, *11*, 2009. [[CrossRef](#)]
34. Whetton, R.; Zhao, Y.; Shaddad, S.; Mouazen, A.M. Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI. *Comput. Electron. Agric.* **2017**, *138*, 127–136. [[CrossRef](#)]
35. Dash, Y.; Mishra, S.K.; Panigrahi, B.K. Rainfall prediction for the Kerala state of India using artificial intelligence approaches. *Comput. Electr. Eng.* **2018**, *70*, 66–73. [[CrossRef](#)]
36. Cardona, C.C.C.; Coronado, Y.M.; Coronado, A.C.M.; Ochoa, I. Genetic diversity in oil palm (*Elaeis guineensis* jacq) using RAM (random amplified microsatellites). *Bragantia* **2018**, *77*, 546–556. [[CrossRef](#)]
37. Paul, J.; Criado, A.R. The art of writing literature review: What do we know and what do we need to know? *Int. Bus. Rev.* **2020**, *29*, 101717. [[CrossRef](#)]
38. Chong, K.L.; Kanniah, K.D.; Pohl, C.; Tan, K.P. A review of remote sensing applications for oil palm studies. *Geo-Spatial Inf. Sci.* **2017**, *20*, 184–200. [[CrossRef](#)]
39. Soh, A.C. Applications and challenges of biotechnology in oil palm breeding. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *183*, 012002. [[CrossRef](#)]
40. Kurnia, J.C.; Jangam, S.V.; Akhtar, S.; Sasmito, A.P.; Mujumdar, A.S. Advances in biofuel production from oil palm and palm oil processing wastes: A review. *Biofuel Res. J.* **2016**, *3*, 332–346. [[CrossRef](#)]
41. Barbedo, J.G.A. Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Comput. Electron. Agric.* **2019**, *162*, 482–492. [[CrossRef](#)]
42. Khosrokhani, M.; Khairunniza-Bejo, S.; Pradhan, B. Geospatial technologies for detection and monitoring of *Ganoderma* basal stem rot infection in oil palm plantations: A review on sensors and techniques. *Geocarto Int.* **2018**, *33*, 260–276. [[CrossRef](#)]
43. Pandey, R.; Naik, S.; Marfatia, R. Image Processing and Machine Learning for Automated Fruit Grading System: A Technical Review. *Int. J. Comput. Appl.* **2013**, *81*, 29–39. [[CrossRef](#)]
44. 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](#)]
45. Khan, N.; Kamaruddin, M.A.; Sheikh, U.U.; Yusup, Y.; Bakht, M.P. Oil Palm and Machine Learning: Reviewing One Decade of Ideas, Innovations, Applications, and Gaps. *Agriculture* **2021**, *11*, 832. [[CrossRef](#)]
46. 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**, *88*, 105906. [[CrossRef](#)]
47. Youmbi, E.; Tabi, K.; Ebongue, N.; Frank, G.; Tonfack, L.B.; Ntsomboh, G. Oil palm (*elaeis guineensis* jacq.) improvement: Pollen assessment for better conservation and germination. *J. Oil Palm Res.* **2015**, *27*, 212–219.
48. Tandon, R.; Manohara, T.N.; Nijalingappa, B.H.M.; Shivanna, K.R. Pollination and pollen-pistil interaction in oil palm, *Elaeis guineensis*. *Ann. Bot.* **2001**, *87*, 831–838. [[CrossRef](#)]
49. Soh, A.C.; Mayes, S.; Roberts, J. Introduction to the oil palm crop. In *Oil Palm Breeding: Genetics and Genomics*; CRC Press: Boca Raton, FL, USA, 2017.
50. Malike, F.A.; Amiruddin, M.D.; Yaakub, Z.; Marjuni, M.; Abdullah, N.; Abu Bakar, N.A.; Mustafa, S.; Mohamad, M.M.; Hassan, M.Y.; Abdullah, M.O.; et al. Oil Palm (*Elaeis* spp.) Breeding in Malaysia. In *Advances in Plant Breeding Strategies: Industrial and Food Crops*; Springer International Publishing: Cham, Switzerland, 2019; Volume 6, pp. 489–535. ISBN 978-303-023-265-8.
51. Soh, A.C. *Breeding and Genetics of the Oil Palm*; AOCS Press: Cambridge, MA, USA, 2012; ISBN 9780128043462.
52. Wahid, M.B.; Abdullah, S.N.A.; Henson, I.E. Oil palm—Achievements and potential. *Plant Prod. Sci.* **2005**, *8*, 288–297. [[CrossRef](#)]
53. Soh, A.C.; Mayes, S.; Roberts, J.; Barcelos, E.; Amblard, P.; Alvarado, A.; Alvarado, J.H.; Escobar, R.; Sritharan, K.; Subramaniam, M.; et al. Breeding Plans and Selection Methods. In *Oil Palm Breeding: Genetics and Genomics*; CRC Press: Boca Raton, FL, USA, 2017; pp. 143–164. ISBN 978-149-871-544-7.
54. Rival, A. Breeding the oil palm (*Elaeis guineensis* Jacq.) for climate change. *OCL—Oilseeds Fats Crop. Lipids* **2017**, *24*, D107. [[CrossRef](#)]
55. Yue, G.H.; Ye, B.Q.; Lee, M. Molecular approaches for improving oil palm for oil. *Mol. Breed.* **2021**, *41*, 22. [[CrossRef](#)]

56. Kartika, N.D.; Astika, I.W.; Santosa, E. Oil palm yield forecasting based on weather variables using artificial neural network. *Indones. J. Electr. Eng. Comput. Sci.* **2016**, *3*, 626–633. [[CrossRef](#)]
57. Siang, C.S.; Sung, C.T.B.; Ismail, M.R.; Yusop, M.R. Modelling hourly AIR temperature, relative humidity and solar irradiance over several major oil palm growing areas in Malaysia. *J. Oil Palm Res.* **2020**, *32*, 34–49. [[CrossRef](#)]
58. Keong, Y.K.; Keng, W.M. Statistical Modeling of Weather-based Yield Forecasting for Young Mature Oil Palm. *APCBEE Procedia* **2012**, *4*, 58–65. [[CrossRef](#)]
59. Kanchymalay, K.; Salim, N.; Krishnan, R. Time series based crude palm oil price forecasting model with weather elements using LSTM network. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 3188–3192. [[CrossRef](#)]
60. Oktarina, S.D.; Nurkhoiry, R.; Pradiko, I. The effect of climate change to palm oil price dynamics: A supply and demand model. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *782*, 032062. [[CrossRef](#)]
61. Wardhani, R.; Rahadian, Y. Sustainability strategy of Indonesian and Malaysian palm oil industry: A qualitative analysis. *Sustain. Account. Manag. Policy J.* **2021**, *12*, 1077–1107. [[CrossRef](#)]
62. USDA Soil Taxonomy, A Basic System of Soil Classification for Making and Interpreting Soil Surveys. *Soil Sci.* **1977**, *123*, 270. [[CrossRef](#)]
63. Nachtergaele, F. Soil taxonomy—A basic system of soil classification for making and interpreting soil surveys. *Geoderma* **2001**, *99*, 336–337. [[CrossRef](#)]
64. Wong, M.K.; Selliah, P.; Ng, T.F.; Amir Hassan, M.H.; Van Ranst, E.; Inubushi, K. Impact of agricultural land use on physicochemical properties of soils derived from sedimentary rocks in Malaysia. *Soil Sci. Plant Nutr.* **2020**, *66*, 214–224. [[CrossRef](#)]
65. Anaba, B.D.; Yemefack, M.; Abossolo-Angue, M.; Ntsomboh-Ntsefong, G.; Bilong, E.G.; Ngando Ebongue, G.F.; Bell, J.M. Soil texture and watering impact on pot recovery of soil-stripped oil palm (*Elaeis guineensis* Jacq.) seedlings. *Heliyon* **2020**, *6*, e05310. [[CrossRef](#)]
66. Arias, D.; González, M.; Prada, F.; Ayala-Díaz, I.; Montoya, C.; Daza, E.; Romero, H.M. Genetic and phenotypic diversity of natural American oil palm (*Elaeis oleifera* (H.B.K.) Cortés) accessions. *Tree Genet. Genomes* **2015**, *11*, 122. [[CrossRef](#)]
67. Low, E.T.L.; Jayanthi, N.; Chan, K.L.; Sanusi, N.S.N.M.; Halim, M.A.A.; Rosli, R.; Azizi, N.; Amiruddin, N.; Angel, L.P.L.; Ong-Abdullah, M.; et al. The oil palm genome revolution. *J. Oil Palm Res.* **2017**, *29*, 456–468. [[CrossRef](#)]
68. Chan, K.L.; Tatarinova, T.V.; Rosli, R.; Amiruddin, N.; Azizi, N.; Halim, M.A.A.; Sanusi, N.S.N.M.; Jayanthi, N.; Ponomarenko, P.; Triska, M.; et al. Evidence-based gene models for structural and functional annotations of the oil palm genome. *Biol. Direct* **2017**, *12*, 1–23. [[CrossRef](#)]
69. Ajeng, A.A.; Abdullah, R.; Malek, M.A.; Chew, K.W.; Ho, Y.C.; Ling, T.C.; Lau, B.F.; Show, P.L. The effects of biofertilizers on growth, soil fertility, and nutrients uptake of oil palm (*Elaeis guineensis*) under greenhouse conditions. *Processes* **2020**, *8*, 1681. [[CrossRef](#)]
70. Carolita, I.; Sitorus, J.; Manalu, J.; Wiratmoko, D. GROWTH PROFILE ANALYSIS OF OIL PALM BY USING SPOT 6 THE CASE OF NORTH SUMATRA. *Int. J. Remote Sens. Earth Sci.* **2017**, *12*, 21–26. [[CrossRef](#)]
71. Kwong, Q.B.; 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](#)]
72. Ong, A.L.; Teh, C.K.; Kwong, Q.B.; Tangaya, P.; Appleton, D.R.; Massawe, F.; Mayes, S. Linkage-based genome assembly improvement of oil palm (*Elaeis guineensis*). *Sci. Rep.* **2019**, *9*, 6619. [[CrossRef](#)] [[PubMed](#)]
73. Xia, W.; Luo, T.; Dou, Y.; Zhang, W.; Mason, A.S.; Huang, D.; Huang, X.; Tang, W.; Wang, J.; Zhang, C.; et al. Identification and Validation of Candidate Genes Involved in Fatty Acid Content in Oil Palm by Genome-Wide Association Analysis. *Front. Plant Sci.* **2019**, *10*, 1263. [[CrossRef](#)]
74. Díaz, B.G.; Zucchi, M.I.; Alves-Pereira, A.; de Almeida, C.P.; Moraes, A.C.L.; Vianna, S.A.; Azevedo-Filho, J.; Colombo, C.A. Genome-wide SNP analysis to assess the genetic population structure and diversity of *Acrocomia* species. *PLoS ONE* **2021**, *16*, e0241025. [[CrossRef](#)]
75. Leaché, A.D.; Oaks, J.R. The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* **2017**, *48*, 69–84. [[CrossRef](#)]
76. Pereira, V.M.; Filho, J.A.F.; Leão, A.P.; Vargas, L.H.G.; de Farias, M.P.; de Rios, S.A.; da Cunha, R.N.V.; Formighieri, E.F.; Alves, A.A.; Souza, M.T. American oil palm from Brazil: Genetic diversity, population structure, and core collection. *Crop Sci.* **2020**, *60*, 3212–3227. [[CrossRef](#)]
77. Danylo, O.; Pirker, J.; Lemoine, G.; Ceccherini, G.; See, L.; McCallum, I.; Hadi; Kraxner, F.; Achard, F.; Fritz, S. A map of the extent and year of detection of oil palm plantations in Indonesia, Malaysia and Thailand. *Sci. Data* **2021**, *8*, 96. [[CrossRef](#)]
78. Sokal, R.R. *A statistical Method for Evaluating Systematic Relationships*; University of Kansas Science Bulletin; University of Kansas: Lawrence, KS, USA, 1958; Volume 38.
79. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [[CrossRef](#)]
80. Hollingsworth, P.M.; Ennos, R.A. Neighbour joining trees, dominant markers and population genetic structure. *Heredity (Edinb.)* **2004**, *92*, 490–498. [[CrossRef](#)]

81. Sarimana, U.; Herrero, J.; Erika, P.; Indarto, N.; Wendra, F.; Santika, B.; Ritter, E.; Sembiring, Z.; Asmono, D. Analysis of genetic diversity and discrimination of Oil Palm DxP populations based on the origins of pisifera elite parents. *Breed. Sci.* **2021**, *71*, 134–143. [[CrossRef](#)] [[PubMed](#)]
82. Okoye, M.N.; Bakoumé, C.; Uguru, M.I.; Singh, R.; Okwuagwu, C.O. Genetic Relationships between Elite Oil Palms from Nigeria and Selected Breeding and Germplasm Materials from Malaysia via Simple Sequence Repeat (SSR) Markers. *J. Agric. Sci.* **2016**, *8*, 159. [[CrossRef](#)]
83. Kalyana Babu, B.; Mathur, R.K.; Naveen Kumar, P.; Ramajayam, D.; Ravichandran, G.; Venu, M.V.B.; SparjanBabu, S. Development, identification & validation of CAPS marker for SHELL trait which governs dura, pisifera & tenera fruit forms in oil palm (*Elaeis guineensis* Jacq.). *PLoS ONE* **2017**, *12*, e0171933. [[CrossRef](#)]
84. Chun, T.C.; Ling, C.K.; Hee, L.C.; Choo, C.S.; Barin, J.; Fong, A.W.; Abdullah, M.P.; Guan, T.S. Genetic diversity and inbreeding level in deli dura and avros advanced breeding materials in oil palm (*Elaeis guineensis* jacq.) using microsatellite markers. *J. Oil Palm Res.* **2018**, *30*, 366–379. [[CrossRef](#)]
85. Basyuni, M.; Wati, R.; Deni, I.; Tia, A.R.; Slamet, B.; Siregar, E.S.; Syahputra, I. Cluster analysis of polyisoprenoid in oil palm (*Elaeis guineensis*) leaves in different land-uses to find the possible cause of yield gap from planting materials. *Biodiversitas* **2018**, *19*, 1492–1501. [[CrossRef](#)]
86. 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](#)]
87. Nugroho, Y.A.; Tanjung, Z.A.; Yono, D.; Mulyana, A.S.; Simbolon, H.M.; Ardi, A.S.; Yong, Y.Y.; Utomo, C.; Liwang, T. Genome-wide SNP-discovery and analysis of genetic diversity in oil palm using double digest restriction site associated DNA sequencing. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *293*, 012041. [[CrossRef](#)]
88. Aziz, M.H.A.; Bejo, S.K.; Hashim, F.; Ramli, N.H.; Ahmad, D. Evaluations of soil resistivity in relation to basal stem rot incidences using soil moisture sensor. *Pertanika J. Sci. Technol.* **2019**, *27*, 225.
89. Ishak, H.; Shiddiq, M.; Fitra, R.H.; Yasmin, N.Z. Ripeness Level Classification of Oil Palm Fresh Fruit Bunch Using Laser Induced Fluorescence Imaging. *J. Aceh Phys. Soc.* **2019**, *8*, 84–89. [[CrossRef](#)]
90. Alaw, F.A.; Sulaiman, N.S. Development of risk assessment model for biomass plant boiler using bayesian network. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *991*, 012136. [[CrossRef](#)]
91. Abbas, F.; Afzaal, H.; Farooque, A.A.; Tang, S. Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* **2020**, *10*, 1046. [[CrossRef](#)]
92. Sousa, S.I.V.; Martins, F.G.; Alvim-Ferraz, M.C.M.; Pereira, M.C. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ. Model. Softw.* **2007**, *22*, 97–103. [[CrossRef](#)]
93. Guo, Y.; Fu, Y.; Hao, F.; Zhang, X.; Wu, W.; Jin, X.; Robin Bryant, C.; Senthilnath, J. Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Indic.* **2021**, *120*, 106935. [[CrossRef](#)]
94. Sim, S.F.; Chai, M.X.L.; Jeffrey Kimura, A.L. Prediction of lard in palm olein oil using simple linear regression (SLR), multiple linear regression (MLR), and partial least squares regression (PLSR) based on fourier-transform infrared (FTIR). *J. Chem.* **2018**, *2018*, 7182801. [[CrossRef](#)]
95. Kim, N.; Ha, K.J.; Park, N.W.; Cho, J.; Hong, S.; Lee, Y.W. A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 240. [[CrossRef](#)]
96. Ostfeld, R.; Howarth, D.; Reiner, D.; Krasny, P. Peeling back the label—Exploring sustainable palm oil ecolabelling and consumption in the United Kingdom. *Environ. Res. Lett.* **2019**, *14*, 014001. [[CrossRef](#)]
97. Solichin, A.; Hasanah, U. Jayanta Development of Prediction System for Crude Palm Oil (CPO) Production with Time Series Data Mining Approach. In Proceedings of the 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020, Jakarta, Indonesia, 19–20 November 2020; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2020; pp. 147–152.
98. Paudel, D.; Boogaard, H.; de Wit, A.; van der Velde, M.; Claverie, M.; Nisini, L.; Janssen, S.; Osinga, S.; Athanasiadis, I.N. Machine learning for regional crop yield forecasting in Europe. *F. Crop. Res.* **2022**, *276*, 108377. [[CrossRef](#)]
99. Shevade, V.S.; Loboda, T.V. Oil palm plantations in Peninsular Malaysia: Determinants and constraints on expansion. *PLoS ONE* **2019**, *14*, e0210628. [[CrossRef](#)]
100. Ruslan, S.A.; Muharam, F.M.; Zulkafli, Z.; Omar, D.; Zambri, M.P. Using satellite-measured relative humidity for prediction of Metisa plana's population in oil palm plantations: A comparative assessment of regression and artificial neural network models. *PLoS ONE* **2019**, *14*, e0223968. [[CrossRef](#)]
101. Santoso, H.; Tani, H.; Wang, X.; Segah, H. Predicting oil palm leaf nutrient contents in kalimantan, indonesia by measuring reflectance with a spectroradiometer. *Int. J. Remote Sens.* **2019**, *40*, 7581–7602. [[CrossRef](#)]
102. Fung, K.F.; Huang, Y.F.; Koo, C.H.; Mirzaei, M. Improved svr machine learning models for agricultural drought prediction at downstream of langat river basin, Malaysia. *J. Water Clim. Chang.* **2020**, *11*, 1383–1398. [[CrossRef](#)]
103. Kusworo, T.D.; Pratama, B.A.; Safira, D.P. Optimization of Bio-oil Production from Empty Palm Fruit Bunches by Pyrolysis using Response Surface Methodology. *Reaktor* **2020**, *20*, 1–9. [[CrossRef](#)]
104. Pasaribu, U.S.; Nurhayati, N.; Ilmi, N.F.F.; Sari, K.N. Determining soil fertility using principal componen regression analysis of oil palm plantation in West Sulawesi, Indonesia. *J. Phys. Conf.Ser.* **2020**, *1494*, 012013. [[CrossRef](#)]

105. Makky, M.; Cherie, D. Pre-harvest oil palm FFB nondestructive evaluation technique using thermal-imaging device. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *757*, 12003. [[CrossRef](#)]
106. Ahmad Jani, H.F.; Meder, R.; Hamid, H.A.; Razali, S.M.; Yusoff, K.H.M. Near infrared spectroscopy of plantation forest soil nutrients in Sabah, Malaysia, and the potential for microsite assessment. *J. Near Infrared Spectrosc.* **2021**, *29*, 148–157. [[CrossRef](#)]
107. Parashar, N.; Khan, J.; Aslfattahi, N.; Saidur, R.; Yahya, S.M. Prediction of the Dynamic Viscosity of MXene/palm Oil Nanofluid Using Support Vector Regression. Recent Trends in Thermal Engineering. In *Lecture Notes in Mechanical Engineering*; Das, L.M., Sharma, A., Hagos, F.Y., Tiwari, S., Eds.; Springer: Singapore, 2022; pp. 49–55. [[CrossRef](#)]
108. Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873. [[CrossRef](#)]
109. Curtin, R.R.; Cline, J.R.; Slagle, N.P.; March, W.B.; Ram, P.; Mehta, N.A.; Gray, A.G. MLPACK: A scalable C++ machine learning library. *J. Mach. Learn. Res.* **2013**, *14*, 801–805.
110. Toh, C.M.; Tey, S.H.; Ewe, H.T.; Vetharatnam, G. Classification of oil palm growth status with L band microwave satellite imagery. In Proceedings of the 2019 Photonics and Electromagnetics Research Symposium—Fall, PIERS—Fall 2019, Xiamen, China, 17–20 December 2019.
111. Chen, Z.Y.; Liao, I.Y. Evaluation of feature extraction methods for classification of palm trees in UAV images. In Proceedings of the 2019 International Conference on Computer and Drone Applications, IConDA 2019, Kuching, Malaysia, 19–21 December 2019.
112. Son, N.T.; Chen, C.F.; Cheng, Y.S.; Toscano, P.; Chen, C.R.; Chen, S.L.; Tseng, K.H.; Syu, C.H.; Guo, H.Y.; Zhang, Y.T. Field-scale rice yield prediction from Sentinel-2 monthly image composites using machine learning algorithms. *Ecol. Inform.* **2022**, *69*, 101618. [[CrossRef](#)]
113. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
114. Muhammad Asraf, H.; Tahir, N.M.; Nur Dalila, K.A.; Hussain, A. An agricultural tele-monitoring method in detecting nutrient deficiencies of oil palm leaf. *Int. J. Eng. Technol.* **2018**, *7*, 227–230. [[CrossRef](#)]
115. Amirruddin, A.D.; Muharam, F.M. Evaluation of linear discriminant and support vector machine classifiers for classification of nitrogen status in mature oil palm from SPOT-6 satellite images: Analysis of raw spectral bands and spectral indices. *Geocarto Int.* **2019**, *34*, 735–749. [[CrossRef](#)]
116. Wang, Y.; Zhu, X.; Wu, B. Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier. *Int. J. Remote Sens.* **2019**, *40*, 7356–7370. [[CrossRef](#)]
117. Teye, E.; Elliott, C.; Sam-Amoah, L.K.; Mingle, C. Rapid and nondestructive fraud detection of palm oil adulteration with Sudan dyes using portable NIR spectroscopic techniques. *Food Addit. Contam. Part A Chem. Anal. Control. Expo. Risk Assess.* **2019**, *36*, 1589–1596. [[CrossRef](#)] [[PubMed](#)]
118. Yousefi, D.B.M.; Mohd Rafie, A.S.S.; Abd Aziz, S.; Azrad, S.; Mazmira Mohd Masri, M.; Shahi, A.; Marzuki, O.F.F. Classification of oil palm female inflorescences anthesis stages using machine learning approaches. *Inf. Process. Agric.* **2020**, *8*, 537–549. [[CrossRef](#)]
119. Shaharum, N.S.N.; Shafri, H.Z.M.; Ghani, W.A.W.A.K.; Samsatli, S.; Al-Habshi, M.M.A.; Yusuf, B. Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms. *Remote Sens. Appl. Soc. Environ.* **2020**, *17*, 100287. [[CrossRef](#)]
120. Montero, D.; Arenas, W.; Salinas, S.; Rueda, C. Development of a system based on aerial images for the morphological patterns classification using support vector machine. *J. Phys. Conf. Ser.* **2020**, *1702*, 012010. [[CrossRef](#)]
121. Sega, M.; Xiao, Y. Multivariate random forests. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 80–87. [[CrossRef](#)]
122. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [[CrossRef](#)]
123. Nguyen, T.D.; Kappas, M. Estimating the aboveground biomass of an evergreen broadleaf forest in Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, using SPOT-6 data and the random forest algorithm. *Int. J. For. Res.* **2020**, *2020*, 4216160. [[CrossRef](#)]
124. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [[CrossRef](#)]
125. Xu, K.; Qian, J.; Hu, Z.; Duan, Z.; Chen, C.; Liu, J.; Sun, J.; Wei, S.; Xing, X. A new machine learning approach in detecting the oil palm plantations using remote sensing data. *Remote Sens.* **2021**, *13*, 236. [[CrossRef](#)]
126. Shaharum, N.S.N.; Shafri, H.Z.M.; Ghani, W.A.W.A.K.; Samsatli, S.; Prince, H.M.; Yusuf, B.; Hamud, A.M. Mapping the spatial distribution and changes of oil palm land cover using an open source cloud-based mapping platform. *Int. J. Remote Sens.* **2019**, *40*, 7459–7476. [[CrossRef](#)]
127. Condro, A.A.; Setiawan, Y.; Prasetyo, L.B.; Pramulya, R.; Siahaan, L. Retrieving the national main commodity maps in indonesia based on high-resolution remotely sensed data using cloud computing platform. *Land* **2020**, *9*, 377. [[CrossRef](#)]
128. Hashim, I.C.; Shariff, A.R.M.; Bejo, S.K.; Muharam, F.M.; Ahmad, K. Machine-Learning Approach Using SAR Data for the Classification of Oil Palm Trees That Are Non-Infected and Infected with the Basal Stem Rot Disease. *Agronomy* **2021**, *11*, 532. [[CrossRef](#)]



129. Zaied, B.K.; Rashid, M.; Nasrullah, M.; Bari, B.S.; Zularisam, A.W.; Singh, L.; Kumar, D.; Krishnan, S. Prediction and optimization of biogas production from POME co-digestion in solar bioreactor using artificial neural network coupled with particle swarm optimization (ANN-PSO). *Biomass Convers. Biorefinery* **2020**, *10*, 1–16. [CrossRef]
130. Marr, B. Deep Learning vs. Neural Networks—What’s The Difference? Available online: <https://bernardmarr.com/default.asp?contentID=1789> (accessed on 19 May 2021).
131. Yildirim, T.; Moriasi, D.N.; Starks, P.J.; Chakraborty, D. Using Artificial Neural Network (ANN) for Short-Range Prediction of Cotton Yield in Data-Scarce Regions. *Agronomy* **2022**, *12*, 828. [CrossRef]
132. Maya Gopal, P.S.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* **2019**, *165*, 104968. [CrossRef]
133. Jayaselan, H.A.J.; Nawi, N.M.; Ismail, W.I.W.; Mehdizadeh, S.A. Application of artificial neural network classification to determine nutrient content in oil palm leaves. *Appl. Eng. Agric.* **2018**, *34*, 497–504. [CrossRef]
134. Alfatni, M.S.M.; Mohamed Shariff, A.R.; Bejo, S.K.; Ben Saaed, O.M.; Mustapha, A. Real-time oil palm FFB ripeness grading system based on ANN, KNN and SVM classifiers. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *169*, 012067. [CrossRef]
135. Ruslan, S.A.; Muharam, F.M.; Omar, D.; Zulkafli, Z.D.; Zambri, M.P. Development of geospatial model for predicting *Metisa plana*’s prevalence in Malaysian oil palm plantation. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *230*, 012110. [CrossRef]
136. Adizue, U.L.; Nwanya, S.C.; Ozor, P.A. Artificial neural network application to a process time planning problem for palm oil production. *Eng. Appl. Sci. Res.* **2020**, *47*, 161–169. [CrossRef]
137. Alfatni, M.S.M.; Mohamed Shariff, A.R.; Ben Saaed, O.M.; Albhah, A.M.; Mustapha, A. Colour Feature Extraction Techniques for Real Time System of Oil Palm Fresh Fruit Bunch Maturity Grading. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *540*, 012092. [CrossRef]
138. Shook, J.; Wu, L.; Gangopadhyay, T.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Crop yield prediction integrating genotype and weather variables using deep learning. *PLOS ONE* **2021**, *16*, e0252402. [CrossRef] [PubMed]
139. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Tao, F. Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. *Remote Sens.* **2020**, *12*, 21. [CrossRef]
140. Sugiyarto, A.W.; Abadi, A.M. Prediction of Indonesian palm oil production using long short-term memory recurrent neural network (LSTM-RNN). In Proceedings of the 2019 1st International Conference on Artificial Intelligence and Data Sciences, AiDAS, Ipoh, Malaysia, 19 September 2019.
141. Xie, Y. Combining CERES-Wheat model, Sentinel-2 data, and deep learning method for winter wheat yield estimation. *Int. J. Remote Sens.* **2022**, *43*, 630–648. [CrossRef]
142. Mah, P.J.W.; Nanyan, N.N. A comparative study between univariate and bivariate time series models for crude palm oil industry in peninsular Malaysia. *Malays. J. Comput.* **2020**, *5*, 374. [CrossRef]
143. Tschora, H.; Cherubini, F. Co-benefits and trade-offs of agroforestry for climate change mitigation and other sustainability goals in West Africa. *Glob. Ecol. Conserv.* **2020**, *22*, e00919. [CrossRef]
144. Hashemvand Khiabani, P.; Takeuchi, W. Assessment of oil palm yield and biophysical suitability in Indonesia and Malaysia. *Int. J. Remote Sens.* **2020**, *41*, 8520–8546. [CrossRef]
145. Yang, X.-S. Genetic Algorithms. In *Nature-Inspired Optimization Algorithms*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2021; pp. 91–100. [CrossRef]
146. Siregar, B.; Butar-Butar, I.A.; Rahmat, R.F.; Andayani, U.; Fahmi, F. Comparison of Exponential Smoothing Methods in Forecasting Palm Oil Real Production. *J. Phys. Conf. Ser.* **2017**, *801*, 12004. [CrossRef]
147. 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]
148. Dong, R.; Li, W.; Fu, H.; Gan, L.; Yu, L.; Zheng, J.; Xia, M. Oil palm plantation mapping from high-resolution remote sensing images using deep learning. *Int. J. Remote Sens.* **2020**, *41*, 2022–2046. [CrossRef]
149. Tuerxun, A.; Mohamed Shariff, A.R.; Janius, R.; Abbas, Z.; Mahdiraji, G.A. Oil Palm Fresh Fruit Bunches Maturity Prediction by Using Optical Spectrometer. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *540*, 012085. [CrossRef]
150. Rashid, N.A.; Hoong, K.W.; Shamsuddin, A.; Rosely, N.A.M.; Noor, M.A.M.; Jin, K.W.; Lee, M.H.; Hamid, M.K.A. Quality prediction and diagnosis of refined palm oil using partial correlation analysis. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *884*, 012018. [CrossRef]
151. Suppalakpanya, K.; Nikhom, R.; Booranawong, T.; Booranawong, A. A comparison of mhw and ahw methods for forecasting crude palm oil productions in Thailand. *Sci. Technol. Asia* **2020**, *25*, 25–35. [CrossRef]
152. Husin, N.A.; Khairunniza-Bejo, S.; Abdullah, A.F.; Kassim, M.S.M.; Ahmad, D.; Aziz, M.H.A. Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning. *Agronomy* **2020**, *10*, 1624. [CrossRef]
153. Amirruddin, A.D.; Muharam, F.M.; Ismail, M.F.H.; Tan, N.P.; Ismail, M.F.H. Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm’s macronutrients observed from frond 9 and 17. *Comput. Electron. Agric.* **2020**, *178*, 105768. [CrossRef]
154. Alzaeemi, S.A.S.; Sathasivam, S. Examining the Forecasting Movement of Palm Oil Price Using RBFNN-2SATRA Metaheuristic Algorithms for Logic Mining. *IEEE Access* **2021**, *9*, 22542–22557. [CrossRef]
155. Watson-Hernández, F.; Gómez-Calderón, N.; Silva, R.P. da Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques. *AgriEngineering* **2022**, *4*, 279–291. [CrossRef]

156. Yarak, K.; Witayangkurn, A.; Kritiyutanont, K.; Arunplod, C.; Shibasaki, R. Oil palm tree detection and health classification on high-resolution imagery using deep learning. *Agriculture* **2021**, *11*, 183. [[CrossRef](#)]
157. Rifa'i, A. Optimasi Fuzzy Artificial Neural Network dengan Algoritma Genetika untuk Prediksi Harga Crude Palm Oil. *J. Tek. Inform. Dan Sist. Inf.* **2020**, *6*, 234–241. [[CrossRef](#)]
158. Ibrahim, N.U.A.; Abd Aziz, S.; Hashim, N.; Jamaludin, D.; Khaled, A.Y. Dielectric Spectroscopy of Palm Olein During Batch Deep Frying and Their Relation with Degradation Parameters. *J. Food Sci.* **2019**, *84*, 792–797. [[CrossRef](#)] [[PubMed](#)]
159. Ishola, N.B.; Adeyemi, O.O.; Adesina, A.J.; Odude, V.O.; Oyetunde, O.O.; Okeleye, A.A.; Soji-Adekunle, A.R.; Betiku, E. Adaptive neuro-fuzzy inference system-genetic algorithm vs. response surface methodology: A case of optimization of ferric sulfate-catalyzed esterification of palm kernel oil. *Process Saf. Environ. Prot.* **2017**, *111*, 211–220. [[CrossRef](#)]
160. 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](#)] [[PubMed](#)]
161. Salman, N.; Lawi, A.; Syarif, S. Artificial Neural Network Backpropagation with Particle Swarm Optimization for Crude Palm Oil Price Prediction. *J. Phys. Conf. Ser.* **2018**, *1114*, 012088. [[CrossRef](#)]
162. Liu, Z.; Loo, C.K.; Pasupa, K. A novel error-output recurrent two-layer extreme learning machine for multi-step time series prediction. *Sustain. Cities Soc.* **2021**, *66*, 102613. [[CrossRef](#)]
163. Crane-Droesch, A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. [[CrossRef](#)]