IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received March 16, 2021, accepted April 19, 2021, date of publication April 22, 2021, date of current version May 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075159

A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction

MAMUNUR RASHID[1], BIFTA SAMA BARI[1], YUSRI YUSUP[2], MOHAMAD ANUAR KAMARUDDIN[2], AND NUZHAT KHAN[2]

[1]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan 26600, Malaysia
[2]Environmental Technology Program, School of Industrial Technology, Universiti Sains Malaysia, Penang 11800, Malaysia

Corresponding author: Mohamad Anuar Kamaruddin (anuarkamaruddin@usm.my)

This work was supported by the School of Industrial Technology, Universiti Sains Malaysia, under Grant 203.PTEKIND.6777007.

ABSTRACT An early and reliable estimation of crop yield is essential in quantitative and financial evaluation at the field level for determining strategic plans in agricultural commodities for import-export policies and doubling farmer's incomes. Crop yield predictions are carried out to estimate higher crop yield through the use of machine learning algorithms which are one of the challenging issues in the agricultural sector. Due to this developing significance of crop yield prediction, this article provides an exhaustive review on the use of machine learning algorithms to predict crop yield with special emphasis on palm oil yield prediction. Initially, the current status of palm oil yield around the world is presented, along with a brief discussion on the overview of widely used features and prediction algorithms. Then, the critical evaluation of the state-of-the-art machine learning-based crop yield prediction, machine learning application in the palm oil industry and comparative analysis of related studies are presented. Consequently, a detailed study of the advantages and difficulties related to machine learning-based crop yield prediction and proper identification of current and future challenges to the agricultural industry is presented. The potential solutions are additionally prescribed in order to alleviate existing problems in crop yield prediction. Since one of the major objectives of this study is to explore the future perspectives of machine learning-based palm oil yield prediction, the areas including application of remote sensing, plant's growth and disease recognition, mapping and tree counting, optimum features and algorithms have been broadly discussed. Finally, a prospective architecture of machine learning-based palm oil yield prediction has been proposed based on the critical evaluation of existing related studies. This technology will fulfill its promise by performing new research challenges in the analysis of crop yield prediction and the development of an extremely effective model for the prediction of palm oil yields with the most minimal computational difficulty.

INDEX TERMS Artificial intelligence, crop yield prediction, deep learning, machine learning, palm oil yield.

NOMENCLATURE

ABW	Average Weight of Fruit Bunches
AEZs	Agro-Ecological Zones
ANN	Artificial Neural Network
AWIFS	Advanced Wide Field Sensor
BAN	Big Ass Number
BNFY	Bayesian Networks Predicted Fruit Yield
BPNN	Back Propagation Neural Network
BSR	Basal Stem Rot
$BUNCH_HA$	Average Bunch Number Per Hectare
CART	Classification and Regression Trees
CC	Climate Change
CDL	Cropland Data Layer
CMFI	Cropping Management Factor Index
CNN	Convolutional Neural Network

The associate editor coordinating the review of this manuscript and approving it for publication was Hiram Ponce.

CorrCoef	Correlation Coefficient
CPO	Crude Palm Oil
CROPTD	Cross-Regional Oil Palm Tree Detection
CSF	Correlation-Based Filter
CWSI	Crop Water Stress Index
DBN	Deep Belief Network
DES	Double Exponential Smoothing
DFNN	Deep Fully Connected Neural Networks
DL	Deep Learning
DNN	Deep Neural Network
DRL	Deep Reinforcement Learning
DT	Decision Tree
EAHW	Extended Additive Holt-Winters
ES	Exponential Smoothing
EVI	Enhanced Vegetation Index
ExtraTree	Extremely Randomized Trees
FFB	Fresh Fruit Bunch
FLEX	FLuorescence EXplorer
FPAR	Fraction of Photosynthetically Active Radiation
FRBS	Fuzzy Rule Based Systems
FTS	Fuzzy Time Series
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GDD	Growing Degree Days
GDI	Gini Diversity Index
GDP	Gross Domestic Product
GEE	Google Earth Engine
GESAVI	Generalized Soil-Adjusted Vegetation Index
GI	Greenness Index
GLAI	Green Leaf Area Index
GNDVI	Green Normalized Difference Vegetation Index
$GOME-2$	Global Ozone Monitoring Experiment-2
GPP	Gross Primary Production
GPR	Gaussian Process Regression
GRVI	Green-Red Vegetation Index
GVI	Green Vegetation Index
GWR	Geographically Weighted Regression
HOG	Histogram Of Oriented Gradient
IGSO	Improved Grid Search Optimization
IPVI	Infrared Percentage Vegetation Index
$K-NN$	K-nearest Neighbors
KDE	Kernel Density Estimation
KRR	Kernel Ridge Regression
LAI	Leaf Area Index
LASSO	Least Absolute Shrinkage and Selection Operator Regression
LR	Linear Regression
LSM	The Least Squares Method
LST	Land Surface Temperature
LSTN	Long Short-Term Network
MADAN	Multi-level Attention Domain Adaptation Network
MAE	Mean Absolute Error
MANNs	Modular Artificial Neural Networks
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Splines
MHDA	Multi-source Heterogeneous Data Analysis
ML	Machine Learning
MLR	Multiple Linear Regression
MSAVI	Modified Soil Adjust Vegetation Index
MSE	Mean Squared Error
NDVI	Normalized Difference Vegetation Index
NDVIre	Red Edge Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-Infrared Reflectance
$OCO-2$	Orbiting Carbon Observatory 2
OLS	Ordinary Least Squares
OP	Oil Palm
OSAVI	Optimized Soil-Adjusted Vegetation Index
PCA	Partial Correlation Analysis
PLSR	Partial Least Square Regression
PSO	Particle Swarm Optimization
PVI	Perpendicular Vegetation Index
RAE	Regularized Autoencoder
RBFNN	Radial Basis Function Neural Network
RCANet	Residual Channel Attention Network
ReLU	Rectified Linear Units
REMAP	Remote Ecosystem Monitoring Assessment Pipeline
RF	Random Forest
RFE	Recursive Feature Elimination
RIDGE	Ridge Regression
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RSPO	Roundtable on Sustainable Palm Oil
RVI	Ratio Vegetation Index
SAR	Synthetic Aperture Radar
SARIMA	Seasonal Autoregressive Integrated Moving Average
SATRAAIS	Satisfiability Reverse Analysis with Artificial Immune System Algorithm
SAVI	Soil Adjusted Vegetation Index
SIF	Solar-Induced Chlorophyll Fluorescence
SNN	Semiparametric Neural Networks
SR	Simple Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression
TLS	Terrestrial Laser Scanner
TLS	Terrestrial Laser Scanning
TM	Thematic Mapper
TROPOMI	Tropospheric Monitoring Instrument
TSAVI	Transformed Soil-Adjusted Vegetation Index
UAVs	Unmanned Aerial Vehicle
VOD	Vegetation Optical Depth
WDRVI	Wide Dynamic Range Vegetation Index

XGBoost Extreme Gradient Boosting
Yap Years After Planting

I. INTRODUCTION

Agriculture is one of the main sectors of social concern since it provides a significant amount of food. At present, numerous nations are still hungry due to the shortage or lack of food with a rising population [1]. The consolidated impacts of a rising population, natural weather variability, soil loss and climate-changing demand techniques to ensure crop growth and production in a timely and reliable manner. It also requires to contribute to expanding agricultural food production sustainability [2]. These requirements indicate that land assessment, the protection of crops and crop yield prediction is of greater importance to worldwide food production [3]. Thus, an accurate crop yield prediction is mandatory to rely on by the nation's policymaker to obtain convenient export and import evaluations for enhancing national food security.

However, due to numerous complex factors, the prediction of crop yield is challenging. Basically, the crop yield is dependent on numerous factors, including landscapes, soil quality, pest infestations, genotype, quality and accessibility of water, climatic conditions, harvest planning and so on [4]–[6]. Crop yield processes and approaches are time-specific and fundamentally nonlinear [7]. These strategies are also complex because of the incorporation of a large range of interrelated factors which are described and affected by non-arbitration and external aspects [8], [9]. Previously, farmers depended on their experiences and authentic historical information for crop yield prediction and settle on the significant cultivation decisions according to the prediction. Nevertheless, in recent years, the advancement of new innovations, including crop model simulation and machine learning, has appeared to predict yield more precisely, along with the capacity to analyze a huge amount of data using high-performance computing [10]–[13]. Numerous investigations are presently exhibiting relatively higher potential in the use of machine learning algorithms than traditional statistics [4], [14], [15]. Machine learning is an area of artificial intelligence where computers can be taught without certain programming. Such approaches overcome agricultural structures, which are both non-linear or linear, by ensuring a notable prediction capacity [16]. The strategies are obtained from the learning method in the machine learning agricultural system. These methods involve on carrying out a particular task through the train with the training information. The model presumes that the information should be tested after successfully finishing the training phase [1].

There has been published several informative, original articles and reviews on the estimation of crop yield as well as palm oil yield prediction over the last 15 years. Chong *et al.* [17] presented a comprehensive review on remote sensing application for palm oil cultivations, including tree counting, change detection, age estimation, pest and disease identification, AGB and carbon production estimation as well as yield prediction. They have discovered the potential research gap and also recommended possible solutions. However, this review did not emphasize on the palm oil yield prediction algorithms. Young [18] analyzed the key methods that are employed recently in generating official statistics, remote sensing, surveys and their integration with meteorological, administrative or other data. The opportunities in research to enhance current crop yield prediction approaches and probable uncertainty related to the prediction were also emphasized in their study. Although the algorithms for all types of crop yield prediction in a large geographical area have been reviewed in this article, it lacks machine learning approaches that are extensively used to predict crop yield. Moreover, the readers who are searching for precise crop yield prediction models for a specific crop may not be benefitted from this article. An excellent systematic review has been introduced in [19] where a wide range of features and prediction algorithms were reviewed. However, the article is more about information extraction rather than critical evaluation, research gap investigation, and recommendation. A review on the estimation of nitrogen status utilizing machine learning was conducted by Chlingaryan *et al.* [10]. Their study concluded that the rapid advances in sensing and ML technologies could lead to economical solutions in agriculture. A survey on machine learning frameworks relevant to the prediction of crop yield was carried out by Elavarasan *et al.* [15] that was mainly focused on climate parameters. It suggests for seeking more criteria for crop yields in a broad-based manner. Another review was carried out in [20] in which the authors investigated existing knowledge about the productivity of palm oil from a physiological plant in order to obtain a clear image of factors that lead to the gaps in palm oil production. One more review was conducted by Liakos *et al.* [21] based on the implementation of machine learning in agriculture. The publications based on water management, livestock management, crop management and land management were included to perform their analyses. Li *et al.* [22] conducted a review study on assessing fruit maturity to perform the optimum prediction of harvest time and yield. The objective of the recent related review articles, together with critical evaluations, have been listed in Table 1.

The above discussion and Table 1 indicate that, there is still necessary of a comprehensive review article that addresses the lack of existing review articles. Hence, the foremost contributions of this paper are summarized as follows:

1) Providing the current status of palm oil production.
2) Describing the fundamentals aspects of crop yield prediction process.
3) Extensive critical review of the machine learning based crop yield prediction algorithms; critical evaluation of utilized feature sets; comparative analysis of related study.
4) Detailed investigation of benefits and challenges associated with features and machine learning algorithms in the prediction of crop yield.

TABLE 1. Critical evaluation of the recent review article regarding crop yield prediction Algorithms.

Reference	Objectives of the Review	Lack of the Review
[17]	Remote sensing applications for palm plantation monitoring. The existing knowledge gaps are identified. The recommendations for further research are given.	This review contains a lack of potential emphasis on palm oil yield prediction algorithms.
[19]	A wide range of features and machine learning algorithms have been listed. Performance evaluation metrics are also extracted.	This article has the following lacking:- critical evaluation, existing research gap investigation and recommendation and future direction.
[15]	It presents an overview of some of the existing machine learning models associated with the crop yield. It compares one approach with others using various performance measures.	It concentrates machine learning models for predicting the crop yield based on the climatic parameters only where other features including soil nutrients, crop diseases, and water salinity have been avoided.
[10]	It reviews machine learning-based techniques for accurate crop yield prediction and nitrogen status estimation. It recommended sensing technologies and hybrid systems combining ML and signal processing techniques to get an optimum solution.	This article lacks critical evaluation of existing machine learning strategies. Moreover, analysis of features that affect the crop yield prediction model is absent.
[21]	It presents a comprehensive review on machine learning applications in agricultural sectors including crop management, livestock management, water management and soil management.	Since the scope of this review is very broad, a concise section for crop yield prediction has been presented. The commonly used influential features have not reviewed.
[18]	Official survey methods for crop yield prediction in large geographical area are reviewed.	It avoided the machine learning approaches. The reviewed methods are not suitable to predict a specific crop yield.
[20]	This review presents an overview of the available data on yield determining, yield-limiting, and yield-reducing factors in oil palm. It presents current causes of yield gaps and future outlook.	It does not review any machine learning approach for palm oil yield prediction.

5) Expanding the areas of future research on machine learning based crop yield prediction as driven by proper identification of current and projected future technological challenges in the palm oil industry.

This paper differentiates from various recent review works as it emphasizes on palm oil yield prediction using machine learning approaches. This is done by gathering the required information from the latest research works and conceptualizing directions for future research work. This paper takes a step forward by comprehensively gathering a wide range of machine learning-based crop yield prediction algorithms and combining them with other experimental articles, i.e., ANN model [23], RF technique [24], CNN-RNN framework [25], hybrid MLR-ANN architecture [26], hybrid CNN-LSTM architecture [27], deep reinforcement learning [1], and so forth. Moreover, this work is anticipated to capture emerging and diversifying features and machine learning algorithms in the agriculture sector to serve as a foundation to advance the recent adoption of machine learning-based crop yield prediction algorithm in the palm oil industry.

The rest of this paper is structured as follows. Section II presents a systematic way to include and exclude articles for this review. Section III explores the current status of palm oil yield. Section IV describes the fundamentals aspects of the crop yield prediction process. Section V provides a critical evaluation of state-of-the-art crop yield prediction approaches. Next, Section VI recommends the directions for research challenges that seeking to improve the performance of the palm oil and other crop yield prediction model. Finally, the paper is concluded in Section VII.

II. ARTICLE SELECTION: PRISMA STRATEGY

The searching is done by narrowing down to the basic concepts that are relevant to the scope of this review. Machine learning has many application fields, which means that there are a lot of published studies that are probably not in the scope of this review article. The article selection process undertaken according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) [28] strategies is illustrated in Fig. 1. The most renowned databases, namely, Scopus, Web of Science, IEEE Explore, Science Direct, SpringerLink and Taylor & Francis have been used to find the relevant articles. In order to acquire some specific information, we have referred to a number of reputed websites. The scope of this review is selected mainly by two keywords, namely "machine learning" AND "crop yield prediction". Hence, in most of the search, we have used these two keywords together with other keywords including "palm oil", "yield prediction", "yield forecasting", "yield estimation", "prediction model' etc. We have collected various types of documents, including original articles, review articles, book chapters, conference papers, lecture notes and reports, amongst others.

In the first round of review, we read the title, abstract, conclusion, and keywords. After removing the duplicates, the remaining articles are categorized accordingly with respect to their application. Machine learning-based crop yield prediction articles which were written in English are regarded as the selection criteria during the first round of screening. In the second-round screening, the selected manuscripts are scrutinized. Here, the elaborate methodology was the final article selection criteria. After maintaining the above formalities, only "223" articles are selected for this particular review. The following facts have been thoroughly extracted from each article: publication year, information about the dataset, detailed information about features, prediction algorithms and performance of the system. The number of selected documents is categorized according to the journal articles (81%), conference papers (14%), book chapters (2%)

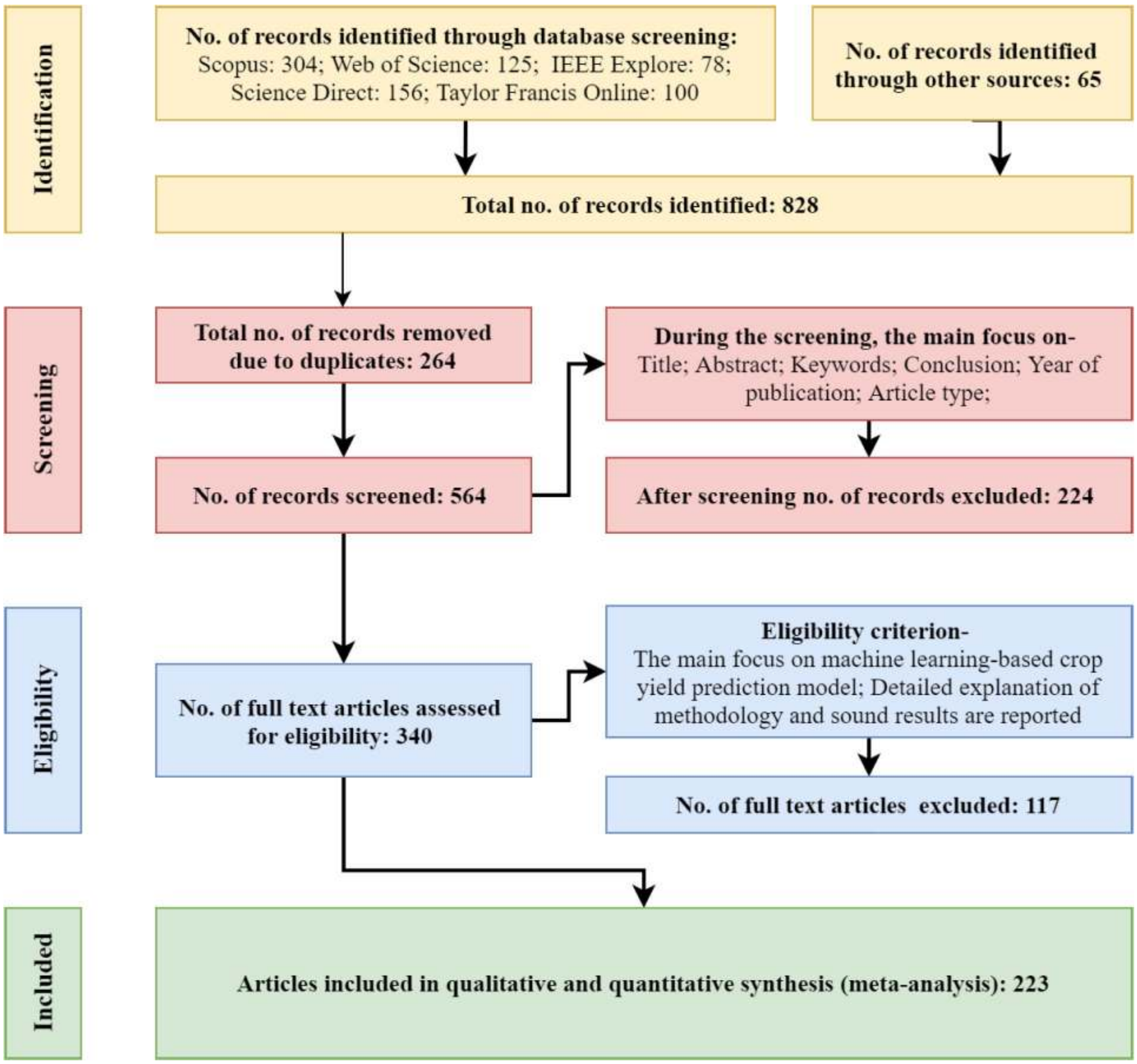

FIGURE 1. **Flow diagram of the article selection process.**

and website (3%) shown in Fig. 2. The number of publications of the selected documents by year for this review is depicted in Fig. 3. It could be viewed that the highest number of publications, i.e., 156, have been selected over the last five years (upon the omission of unrelated articles) for this particular review, suggesting the importance of this field of research.

III. CURRENT STATUS OF PALM OIL PRODUCTION

In the past decade, the oil palm (Elaeis guineensis) plant is cultivated for its oil-rich fruit that has gained popularity in South-East Asia, especially in Malaysia, Indonesia and Thailand. Among other vegetable oils crops, including rapeseed, sunflower and soybean [17], palm oil plays a leading role in the oil-yielding ability. Vegetable oils are usually derived from plants in different ways with several textures such as oily, fatty and liquid. When compared to animal fats for consuming as a food product, vegetable oils are obtained as the better option for a healthier alternative since they are composed of more unsaturated fatty acids than animal fats. Most of the vegetable oils are suitable to be utilized for the production of fuel and diesel or as cooking oil. Canola oil, soybean oil, sunflower seed oil and palm oil are considered as the most commonly utilized cooking oil. In the tropical climate of Africa, South America and South-East Asia, the flesh of the palm fruit were primarily found from which palm oil is produced. Palm oil has grown to be one of the largest consumed vegetable oils worldwide (see Fig. 4) [29].

Around 90 per cent of palm oil is consumed for food use, while approximately 10 per cent of palm oil is used in industrial goods, including cosmetic, fuel and diesel [29]. In the 2019/2020 crop year worldwide, the production of vegetable oil reached up to 200 million metric tons. Palm oil is a big

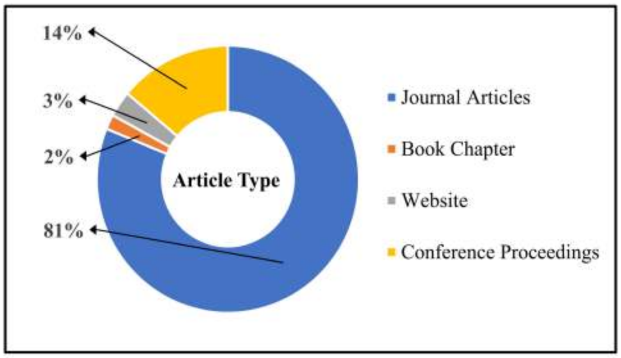

FIGURE 2. Number of journal articles, conference papers and book chapters amongst others.

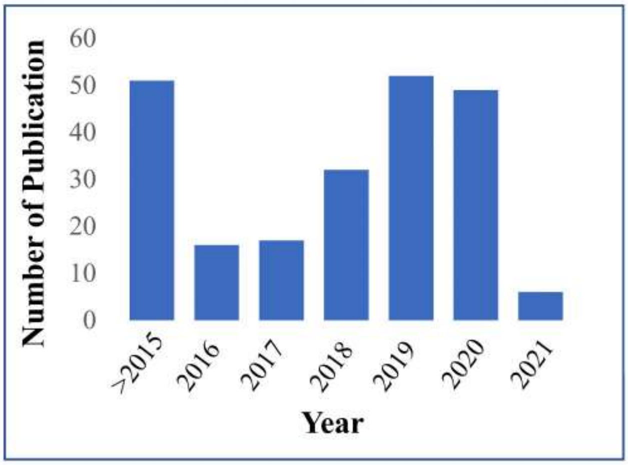

FIGURE 3. The number of reviewed articles with respect to the publication year.

commodity in that era by contributing the highest amount of oil production at 72.27 million metric tons compared to other numerous vegetable oils in the world (see Fig. 5) [30].

Palm oil is a widespread oil that is used in consumer goods as well as processed foods. Palm fruits are harvested from the oil palm tree, and then the oils are extracted from these fruits. The land for palm oil planting is very productive. With a high yield for a limited amount of land use, palm oil is a significant yielding crop over other sorts of oil crop. The amount of palm oil production has been continuously growing at a rapid pace for the last few years. Thus, there are huge as well as active exports marts for palm oil worldwide, whilst Malaysia and Indonesia are the largest palm oil exporters [30]. Fig. 6 presents the major producers and exporters of Palm oil [31].

From Fig. 6, it is clear that although palm oil production is high in Indonesia, the highest palm oil exportation occurs in Malaysia. Thus, Malaysia and Indonesia play the highest role in palm oil exportation, where Malaysia is a bit higher than Indonesia.

Palm oil has been produced in Malaysia at first in the year 1911 that has been predicted to become one of the leading industries in the future world economy [32]. There are great economic benefits in continuing the export of palm oil, including international interest, expertise, contributions of money and managing knowledge [33]. It is a cheaper and powerful biofuel that is preferred for the rising global population [34]. It boosts the Malaysian Gross Domestic Product (GDP) by providing a lot of job opportunities to millions of people that are generated by the palm oil industry. Agriculture plays a key role in Malaysia's economy in a way that helps by creating job opportunity for 16 per cent of the population and 12 per cent to the national GDP [35]. According to national statistics, the Malaysian government achieved 10.55 million metric tons of palm oil in 1999 by ranking one of the world's biggest palm oil producers. These estimates show that almost 85 per cent or 8.8 million metric tons of this palm oil was exported to a foreign market [35].

Palm oil is produced mainly in countries like Asia, Africa and Latin America. The major producers of palm oil are Indonesia, Malaysia, Thailand, Colombia, and Nigeria [36]. It is unsurprising that the largest exporters of palm oil in the world are Indonesia and Malaysia that has the certified area for a wide amount of palm trees plantation. It was reported by the Roundtable on Sustainable Palm Oil (RSPO) that the area planted with palm oil in Indonesia had risen to 1.7 million hectares in 2017 that was up from 1.54 million hectares in 2016. There has been an increase in the import of palm oil in recent years. The report issued by the World Bank that the average price for palm oil rose to the point that it was 649 U.S. dollars per metric ton in 2017. It is predicted that the price should hit about 744 USD per metric ton by 2025. As shown in Fig. 7, the worldwide consumption of palm oil is developing steadily.

Palm oil is the main traded edible oil in the world. For the year 2011, its export amounted to 39.04 million tonnes, where there was a 46 per cent share in Malaysia. Malaysian factories have spent huge amounts of money to refine palm oil for the benefit of the public and has made it suitable for the consumption of human. The export of processed palm oil had been most fruitful during the years of 1974 to 1999 when it rose from 0.9 to 8.9 million tonnes. Malaysia's export of palm oil hit an all-time high, amounting to 17.99 million tonnes in 2011 after setting a previous high of 16.66 million tonnes in 2010. This same export pattern was also evident in the same period, with the shipment volume reaching 1.17 million tons in 2011 [31].

IV. FUNDAMENTALS ASPECTS OF CROP YIELD PREDICTION PROCESS

The machine learning-based crop yield prediction method consists of some phases, namely data collection, data pre-processing, data partition and data analysis. Fig. 8 illustrates the architecture of the machine learning-based crop yield prediction method. In data analysis section, machine learning based regression or classification algorithm is employed.

A. POPULAR FEATURES USED IN CROP YIELD PREDICTION

There are various factors that influence crop yield and the uncertainties connected with cultivation. The feature lists are

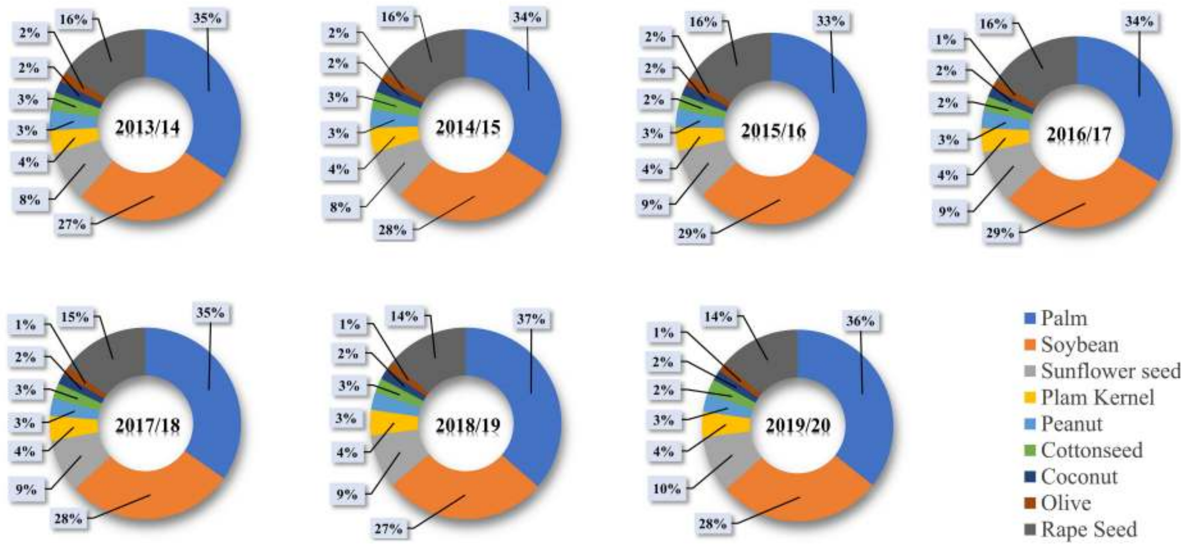

FIGURE 4. Vegetable oils consumption from 2013/14 to 2019/2020 by oil categories (in a million metric tons) globally [29].

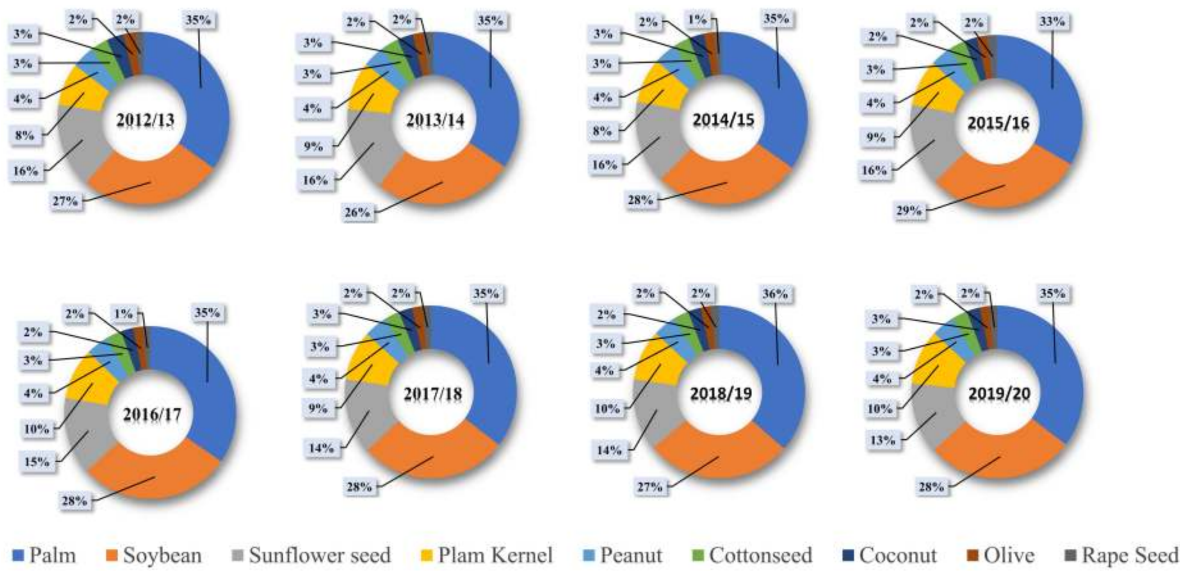

FIGURE 5. Major vegetable oils production from 2012/13 to 2019/2020 by categories (in a million metric tons) globally [30].

the most vital components to predict crop yield. According to the recent literature, the most significant factors that contribute to crop yield are the leave and fruit information, irrigation information, soil properties, climatic information, cropland information, vegetation index and satellite data, crop management data, historical yield data, groundwater characteristics, fertilization information, socio-economic factors, phenology data and others. Table 2 lists the most popular features utilized in the prediction of crop yield with data description. There are a wide variety of nutrient supplements used in crop yield, which play a crucial role in improving agricultural production. The most common utilized fertilizer or nutrient supplements are phosphorus, nitrogen, calcium, potassium, sulfur, magnesium, manganese, iron and so on [10]. These factors are equivalently important to the crops though they can differ by quantity to use. The lack of any of these essential nutrients would considerably degrade the output yield. Hence, soil properties greatly affect the spatial variability of the crop yield when the weather and the crop management are in the same condition. A study in [38] concluded that soil quality is important in predicting the productivity of crops, which would lead to provide complementary information and improve the accuracy of yield prediction. Remote sensing will help to cover a vast scale with non-invasive and efficient techniques to detect the spatial variability in plant status with high temporal

TABLE 2. List of Popular features with utilized data.

Feature	Data
Leave and fruit Information	Healthy and unhealthy image [44], leaf development [45], leaf area index [40], leaves and fruit mortality [40], number of leaves [40], number of fruits [40], number of stems [40].
Irrigation information	Irrigation ratio [46], number of open wells (OW) [26] [41], number of tanks (TK) [41] [26], canals length (CL) [26] [47] [41] , irrigation details [26] [48], net irrigated area [1] [49], gross irrigated area [1] [49] [50], tube wells number [41].
Soil properties	Clay [38] [46] [25] [51] [13] [49], silt [38] [49] [46] [51], sand [38] [46] [25] [51] [13] [49], organic carbon [38] [49] [46] [50] [52], PH [38] [46] [25] [53] [54] [51] [1] [13] [49] [52], cation exchange capacity [38] [49]] [46] [50], bulk density [38] [46] [53] [50], wet soil bulk density [25], dry bulk density [25] [13], upper limit of plant available water content [25], lower limit of plant available water content [25] [51], hydraulic conductivity [25], organic matter percentage [25] [51], and saturated volumetric water content at different depths [25], number of macronutrients in the soil [53] [26] [48], nutrient supplements [53] [1] [13], soil moisture [50] [48] [54] [52], SAVI [48], soil electrical conductivity [54] [1] [13] [49], soil slope [54], top soil depth [1] [13] [50].
Climatic information	Precipitation [38] [55] [52] [49] [50] [46] [25] [42]] [53] [56] [24] [51] [1] [13] [43] [57] [52], vapor pressure deficit [38] [46] [56] [57], humidity [58] [59] [38] [53] [40] [1] [43] [49], wind speed [59] [38] [53] [1] [43] [49], frost frequency of ground [53], solar radiation [59] [[38] [47] [57] [25] [26] [40] [51] [13] [49] [41], snow water equivalent [25], maximum temperature [60] [47] [52] [50] [25] [42] [26] [51] [1] [13] [43] [57] [49] [41], minimum temperature [60] [47] [52] [50] [25] [42] [26] [51] [1] [13] [43] [57] [49] [41], mean temperature [59] [38] [55] [47] [42] [53] [26] [40] [56] [1] [43] [57] [41], vapor pressure [59] [25] [53] [51] [1] [13] [50] [57], rainfall [58] [59] [47] [26] [61] [41], CO_2 concentration [40], GDD [24], growing degree days [46], killing degree days [46], day length [51] [13] [43], snow water [13], drought index (DI) [52]
Vegetation indices and satellite data	EVI [[39] [62] [55] [46] [42] [56] [50] [57], [46] [52] [63], NDVI [58] [39] [38] [64] [65] [63] [55] [52] [42] [48] [[56] [54] [57], [50] [66], GCVI [50], VOD [46] [62], RVI [63] [39], GNDVI [63] [39], GRVI [63] [39], EVI2 [63], OSAVI [63] [39], WDRVI [63] [39], NDVIre [64], TSAVI [39], IPVI [39], MSAVI [39], GI [39], PVI [39], SAVI [39]], GESAVI [39], GLAI [39], CWSI [39], NDWI [39], GVI [39] LAI [42] [40] [65], FPAR [42], GPP [42], NIRv [56], and LST [58] [56] [24] [62], VCI [24] [[55], surface reflectance [62] [64] [38], land cover type [62] SIF [46] [57] OCO-2 [56], TROPOMI [56], GOME-2 [56] [57], GVI [48].
Cropland information	Plantation area [41] [23] [26] [61] [1] [50] [13] [57] [52] [47], CDL [42], cropland census [42], satellite images from the landsat thematic mapper (TM) [42], satellite images from advanced wide field sensor (AWIFS) [42], empty-land [45], harrowed land [45], texture conditions [48], PVI [48].
Crop management data	Weekly cumulative percentage of planted fields [25], fertilizer usage [26], seed quantity [26] [45] [41], vegetative growth [45] [27], flowering [45], maturity [45], crop genotype [51], Plant population [13], planting progress [13] [59], power of agricultural machinery [50], the electricity consumed in rural areas [50].
Historical yield data	[42] [26] [45] [56] [27] [24] [50] [51] [57] [52] [[67] [59] [46] [25] [13].
Groundwater characteristics	Type of aquifer [42], transmissivity [42], rock layer permeability [42], water conductivity [42], and the number of micronutrients [42] and hydrochemical analysis [42].
Fertilization information	Potassium [1] [41], nitrogen [1] [41], phosphorus [1] [41], nutrient supplements of groundwater [1] [68], consumption of chemical fertilizers [50], consumption of agricultural pesticide [50].
Phenology data	Dates of sowing [43], emergence [43], three leaves [43], transplanting [43], turning green [43], tiller [43], booting [43], heading [43], flowering [43], milky [43] and mature [43].

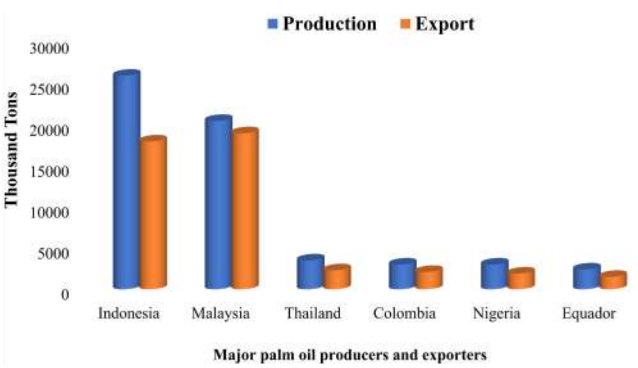

FIGURE 6. Major palm oil producers and exporters.

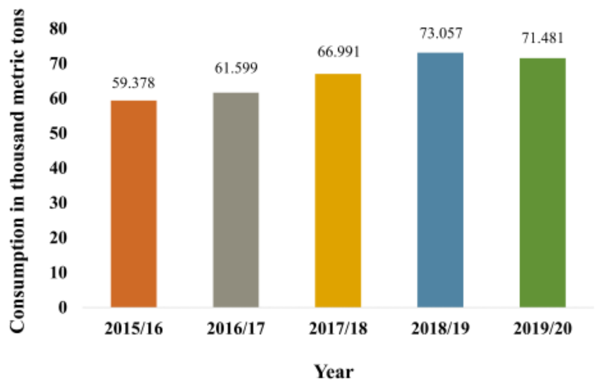

FIGURE 7. Palm oil consumption worldwide from 2015/2016 to 2019/2020 (in 1,000 metric tons) [37].

resolution [39]. There is a large implementation of remote sensing approaches based on infrared thermometry and spectral vegetation indices for crop yield prediction because of their non-destructive and not labour-or time-intensive characteristics. The feature group "Leave and Fruit Information" consists of the following variables: healthy and unhealthy image, leaf development, leaf area index, leaves and fruit mortality, number of leaves, number of fruits, number of stems [40]. Irrigation information refers to the number of open wells, irrigation ratio, net irrigated area, number of tanks, irrigation details, canals length, tube wells number and gross irrigated area [41]. Previous studies have confirmed that historical crop yield data significantly affects the crop yield prediction algorithms [13], [25]. The features that fall under the groundwater characteristics group include

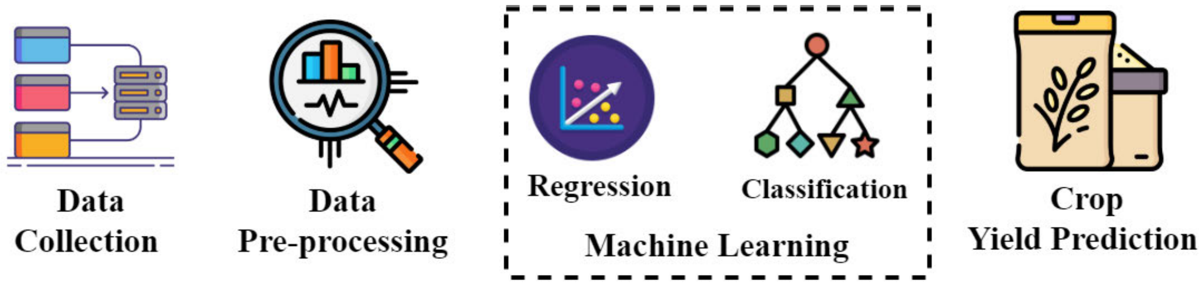

FIGURE 8. The general architecture of machine learning-based crop yield prediction.

transmissivity, water conductivity, aquifer type, rock layer permeability, as well as the number of micronutrients and hydrochemical analysis [42]. Other measurements that also significantly contribute to crop yield prediction are the cropland information [42], crop management data [25], phenology data [43].

B. PREDICTION ALGORITHMS

A wide range of regression and classification-based prediction algorithms have been utilized to forecast crop yield. In crop yield prediction, linear regression (LR) and multiple linear regression (MLR), multivariate adaptive regression splines (MARS), k-nearest neighbors (K-NN), support vector machine (SVM) and support vector regression (SVR), decision tree (DT), random forest (RF), extremely randomized trees (extra tree) (ERT), artificial neural network (ANN), deep neural network (DNN), convolution neural network (CNN) and long short-term memory (LSTM) have been employed.

1) MACHINE LEARNING ALGORITHMS

The LR model represents a relationship between independent and one or more dependent variables [54]. In a machine learning framework, learning can be done by using data and minimizing the loss or error (RMSE or MSE) that are experienced by using regression algorithms. The MLR analysis has been used in several applications [43] in which multi-independent variables was proved to be the most efficient and reliable compared to one independent variable [69]. The least-squares method (LSM) is widely used for regression estimation in MLR models. A MARS model can be evaluated by the well-known generalized cross-validation (GCV) index [42] which incorporates MLR methods to account for nonlinearity and correlations of different variables. The learning data is split into sets of adaptive splines with unique slopes [70]. The K-NN is used for classification and regression that provides more weightage to close neighbors in making the prediction so that they relate more to the average than distant ones [54]. Various distance formulas such as Euclidean, Manhattan, and Minkowski can be applied to compute the distance between two neighbors. SVM is a binary classifier class, which generates a linear separating hyperplane for the classification of data instances [21]. SVM is a little bit different from SVR where SVR method can be used to solve regression problems [71]. DT is a model of supervised machine learning model which can be applied to both regression and classification [72]. It consists of three nodes, namely root node (no incoming edge), decision node (both incoming and outgoing edges) and leaf node (no outgoing edges). In a decision tree, each attribute is divided by each outgoing node into two or more branches according to the splitting criteria. The most successful methods of DT induction is Classification and Regression Trees (CART) that was developed by Breiman *et al.* [73] which is supposed to nonparametric and generates binary trees from such data by employing the discrete and continuous features [74]. In CART, information gain, Gini Diversity Index (GDI) and gain ratio are used to split the attributes. RF is a powerful tool for the prediction of yield, which has been applied to agricultural research [24], [42], [46], [53], [56]. It generates a wide range of regression trees that are produced by a large set of decision trees for computing regression [75]. The RF is superior to any other decision tree since it performs more precisely, and the bias is compensated for by the single decision tree due to the randomness [50]. Extremely randomized trees (extra tree) (ERT) is an ensemble model as same as RF, but it utilizes unpruned decision trees. It splits the nodes by randomly chosen cut-points and incorporates the complete learning sample (without bootstrap copying) to grow the trees [76]. The number of trees and the number of variables utilized to divide the nodes are set to be the same as those of RF [42]. An ANN is the most commonly utilized machine learning technique to predict crop yield [21] by which the complex nonlinear relationship between input and output can be modeled [77]. It comprises of three layers, including the input layer, hidden layer and output layer. There are numerous factors that have an impact on the performance of ANN, including the number of nodes in the hidden layer, the learning rate, and the training tolerance [42]. The learning rate determines the amount by which the weights change during a series of iterations to bring the predicted value within an acceptable range of the observed value.

2) DEEP LEARNING ALGORITHMS

The classic ANN consists of a local minima problem in which an optimization process often stops at a local rather than the globally optimized state. Moreover, generic machine learning models sometimes have complexity with overfitting problem. The local minima and overfitting problems can be resolved by DNNs through an intensive optimization process in a deep network structure. In addition, in order to enhance the performance, backward and forward optimization is conducted in the back-propagation algorithm. The problem of vanishing gradients of loss functions, which may occur during the back-propagation process, can be managed by employing efficient activation functions, including sigmoid and rectified linear units (ReLU) [42]. A CNN consists of a series of basic units (convolutional, pooling, and activation layers) stacked between the input and output layers [78]. Several local filters are applied to perform convolution operation on the input data in a convolution layer, whereas a pooling layer can obtain the low-dimensional data from the input data through different operations (max-pooling and average-pooling). The nonlinear fitting capability of CNN can be enhanced by the nonlinear operations in the activation layer [78]. CNN generally updates weight through BP in the same way as BPNN. An LSTM network is effectively employed for classifying as well as predicting the sequence and time-series data [79] which consists of forgetting, input, and output gates that are employed to control the filtering of the previous status. This structure aims to acquire previous statuses that are influential to the present instead of the most recent ones. The detailed algorithm and network structure of the LSTM models can be referred to by Rashid *et al.* [79].

3) MISCELLANEOUS ALGORITHMS

Besides these, many classification and regression algorithms have successfully been utilized in crop yield prediction and these algorithms include [13], [46], [51], [56], [57], XGBoost [46], MARS [42], elastic net [54], Gradient boosting [52], [53], TOMGRO and Vanthoor [40], ridge regression (RIDGE) [46], [50], [56], [62], SNN [49], [51], DRL [1] and miscellaneous [13], [39], [49], [50], [52], [59]. Some hybrid approaches including CNN-RNN [25], CNN-LSTM [27], MLR-ANN [26] have also been employed in crop yield prediction. The key characteristics, benefits and drawbacks of widely utilized crop yield prediction algorithms [42] are briefly tabulated in Table 3.

C. PERFORMANCE EVALUATION METRICS

The performance of a model can be defined by evaluation metrics. Evaluation measure plays a significant role because of their capability in differentiating among the outcomes of different learning models [53]. There are various performance metrics applied to evaluate the performance of the regression model including mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), determination coefficient (R-squared) and mean absolute percentage error (MAPE). MAE is defined as an arithmetical mean of the absolute variation between the forecasted observation and the actual observation which is used to calculate the average importance of the errors with a given array of predictions [83]. MSE reflects the closeness of the regressor line to dataset points which is used to determine the performance of the estimator [84]. RMSE reflects how well the information is concentrated on the best fit line and utilized for the estimation of the residuals or forecasted error's standard deviation [85]. The determination coefficient is used for measuring the accuracy of the fit of the regression framework that reflects how the developed framework is superior to the baseline framework [86]. MAPE is used to calculate the average of the percentage errors that determines how far the prediction of the model deviates from its corresponding outcomes [53]. The machine learning based classification algorithms for crop yield prediction are evaluated by accuracy [87], [88], precision [89], recall [89], sensitivity [90], specificity [90], and F1 Score [91]. However, classification accuracy is the most widely used and effective metric for classification problems.

D. WIRELESS SENSOR NETWORKS IN PRECISION AGRICULTURE

Wireless sensor networks (WSNs) can be utilized to supply farmers with a lot of useful information for their crop production and quality decisions that has an enormous potential in agriculture. Thus, a number of WSNs have been found to different uses in agriculture, such as in climate and nutrient data monitoring, crop health forecasting as well as crop production monitoring [92]. Irrigation planning can be predicted using WSNs have a broad impact on the prediction of irrigation planning by considering different factors including weather conditions (such as temperature and humidity) and soil moisture. In prediction algorithms, a combination of AI and WSN can be used in agriculture fields for real time monitoring and intelligent farming decisions making. An Internet of Things (IoT) sensor network which consists of a soil moisture sensor, an electrochemical sensor, and an optical sensor measures the field data continuously that can be used as training data in ML and DL algorithms. The authors in [93] employed a sensor network using AI to evaluate land condition after every cultivation as a suitable, more suitable, moderately suitable and unsuitable land. Three experiments in palm trees were carried out by [94], to investigate a signal propagation model that is based on the received signal strength indicator (RSSI) data measurements from the ZigBee wireless protocol. To monitor and control essential parameters that affect crop growth such as soil and weather conditions in Florida, USA, a microcontroller-based WSN was employed which includes the use of the ZigBee module and one Arduino in [95]. In [96], the BT module was used in integrated control method for controlling the irrigation system in greenhouses based on soil and weather information. The number of leaves, height, dry weight, and fresh weight of red and romaine lettuce was increased in greenhouses by using their technology. In [97], the agricultural data, such as soil moisture, soil temperature, sunlight intensity, weather temperature and CO_2 humidity and were stored at first in a gateway and then transmitted to the server computer over a WiFi network. An automatic crop irrigation approach was

TABLE 3. Summary of benefits, Drawbacks, and key characteristic of widely used prediction algorithms.

Algorithm Name	Key Features	Advantages	Drawbacks
ANN	This network is inspired by the brain and essentially consists of input, output and hidden layers [21].	It can perform complex nonlinear relationships between dependent and independent variables [21].	It contains "black box" nature and greater computational burden [80].
DT	DT network consists of three nodes namely root node, decision node and leaf node [21].	DT can be used in both classification and regression problem. It does not require normalization and scaling of data [21].	This model has overfitting problem of the data which ultimately leads to wrong predictions [21].
RF	It consists of ensemble learning which utilizes the bootstrap and bagging process [81].	It is less prone to overfitting problem than decision tree [81].	It may change significantly by a small change in the data.
MARS	A method of non-parametric regression integrating a number of linear models to deal with nonlinearity and correlations between variables [21].	It provides a versatile model capable of dealing with nonlinearities and linearities [21].	It is restricted to handle large data and prone to overfitting issue [21].
SVM	The SVM constructs multi-dimensional boarders between data points in the feature space [82].	It supports effective data grouping by optimizing the margin across categories employing kernel functions [82].	Prone to over-fitting problems which depends on the kernel functions used for optimum classification [82].
DNN	It utilizes several layers to gradually extract upper level properties from the raw input [19].	It can address issues of over-fitting and local minima via an intensive optimization and activation process [19].	Requires a high-end computer.
CNN	It consists of one input, hidden and an output layer whereas the hidden layers contain convolutional layers, ReLU layers, pooling layers, fully connected layers [78].	A fully connected layer to convolution layer replacement indicates huge decrease in the number of parameters [78].	A convolution is a considerably slower operation than ANN.
LSTM	It is the feedforward NN with feedback from the input layer to the output layers of neurons [78].	The sequence of data is modeled through the RNN such that each sample can be presumed to be dependent on previous ones [78].	This network cannot process very long sequences when tanh or ReLU is used as an activation function [78].

implemented using the GPRS module and WSN by Gutiérrez *et al.* [98]. In their study, the data was captured by temperature and soil moisture sensors which were installed at the root zone of plants. This system was considered as a cost-effective and practical solution for the improvement of water quality in PA. The numerous sensors, microcontrollers, and the LoRa wireless protocol were used to monitor the soil moisture and temperature, air temperature and humidity, and light intensity inside greenhouses [99]. Because of its low power consumption and applicable communication range [100], the performances comparison of different wireless protocols showed that the ZigBee and LoRa wireless protocols are both good for agricultural use than other related existing systems.

E. USES OF IOT IN PRECISION AGRICULTURE

As the IoT is rapidly spreading, many industries such as smart agriculture, which were based on statistical methodologies for marketing, were revolutionized with the advancement of quantitative methodologies. The revolutionary changes are mitigating the existing agriculture systems and creating new opportunities along a range of challenges [101]. The advance innovative IoT and sensor reduce the traditional method's complexity and have become a potential solution of different agricultural issues, like soil sampling and mapping, fertilization, irrigation, disease monitoring, and forecasting, yield measurement, and forecasting. The authors in [102], carried out an approach using the MODIS sensor for the mapping of different soil functional properties to monitor the land degradation risk for sub-Saharan Africa. This continent-wide prediction model was created using soil maps and field survey data for all the major climate zones on the continent. It is predicted that the use of IoT-based techniques such as crop water stress index (CWSI)- based irrigation management will yield an enormous increase in crop efficiency [103]. In [104], with the help of IoT-based fertilizing techniques, the NDVI from aerial/satellite images were employed for the monitoring of crop nutrient status. A modern strawberry disease prediction technique was implemented in [105], to develop real-time monitoring, modeling and disease forecasting facilities. The authors in [106], proposed a yield monitoring technique that can be installed on any harvester and linked with the mobile app FarmRTX. This app has enormous potentials including the ability to display live harvest data, upload data to the manufacturer's web-based platform, generate high-quality yield maps and share these maps with an agronomist. In [107], authors implemented an approach to predict how well the crop is progressing with the help of IoT where the fruit growth was considered as the most basic and relevant parameter to estimate. Thus, the use of IoT-based sensors and communications will be a key in increasing crop yields [92].

V. CRITICAL EVALUATION OF RECENT STUDIES RELATED TO THE CROP YIELD PREDICTION

A. CORN YIELD PREDICTION

Corn yield prediction using different ML techniques have been investigated in study [12], [13], [48], [49], [65], [108]. Panda *et al.* [48] investigated the strength of key spectral vegetation indices and concluded that distance-based VI technique and PVI outperform over the other four techniques. Lykhovyd [65] investigated the robustness and reliability of LAI and NDVI and then concluded that LAI is superior and more efficient than the NDVI-based method. However, the prediction ability could be enhanced by including other agricultural features rather than only vegetation indices-based features. An experiment conducted by Ransom *et al.* [108] proposed predictive models utilizing ML algorithms for corn nitrogen recommendation systems by employing soil and weather data. Shahhosseini *et al.* [12] evaluated computer simulation models based on MLR and RF for the yield prediction and nitrate losses. Based on the analysis, Shahhosseini *et al.* [13] suggested that optimized weighted ensemble and the average ensemble are the most reliable and accurate techniques. However, the efficiency of ensemble modeling is affected by the selection of diverse ML models; therefore, it needs to be investigated. In reference to [49], the authors developed a framework for augmenting parametric statistical models, namely semiparametric neural networks (SNN) and claimed that the SNN achieved better performance in out-of-sample prediction than anything else yet published.

B. SOYABEAN YIELD PREDICTION

In studies [27], [55], [63], [64], [109], a wide range of ML based soyabean yield prediction have been investigated. Schwalberta *et al.* [55] utilized satellite image and weather data to perform real-time soybean yield forecasts using LSTM, OLS linear regression, RF models. The LSTM has an outstanding performance compared to RF and the multivariate OLS regressions. The proposed method can predict the soybean yield with only 16 days (normal duration is 70 days) with reasonable accuracy. The DNN model was more successful when the number of the input features was increased compared to other regression techniques, including SVR, RFR and PLSR. For example, Maimaitijianga *et al.* [63] investigated and summarized that UAV-based multimodal data fusion with RGB, multispectral and thermal sensors could be effectively estimated soybean yield using a DNN approach. Terliksiz and Altỳlar *et al.* [109] employed a 3D CNN method where LST and SR based spatiotemporal features from satellite data were utilized. Kross *et al.* [64] evaluated the relative importance of predictor set consists of NDVI, red edge NDVI and SR. Finally, ANN based prediction model was developed using the selected variables. Hybridization of prediction algorithm generally perform better when it is compared with the single prediction algorithm. For example, Sun *et al.* [27] claimed that the performance of CNN-LSTM based prediction model is better compared to the pure CNN or LSTM model to predict both end-of-season and in-season yield in CONUS at the county-level. However, the computational complexity due to the hybridization is comparatively higher than the single algorithms.

C. PADDY YIELD PREDICTION

A wide range of studies [1], [26], [39], [41], [43], [53] have been carried out to investigate the paddy yield prediction using ML architectures. Elavarasan and Vincent [1] proposed deep reinforcement learning and utilized 38 features to design the prediction model. The efficiency and precision of the proposed model is superior to other frameworks including LSTN, BAN, BDN, RAE and IDANN. However, the RNN-based DRL may cause those gradients to burst or vanish as the time series is very long. So, probabilistic boosting and bagging techniques can be applied to deal with this issue. Gopal and Bhargavi [26] developed a hybrid MLR-ANN model where the MLR's coefficients and bias are employed to initialize the input layer weights and bias of ANN instead of random weights and bias initialization. The proposed hybrid MLR-ANN model showed better prediction accuracy using same agricultural dataset over other ML algorithms, including SVR, K-NN and RF. The climate, phenology, geographical and preseason data of early mature rice in South China from 1981 to 2010 were applied to SVM, RF and BPNN separately to predict rice yield in [43]. The findings of this study showed that the performance of SVM was much better than that of the BPNN and RF. Shiu and Chuang [39] employed SVR, OLS and GWR models where GWR outperformed SVR, OLS model since numerous weight values were established with the GWR model for spatial grids. Gopal *et al.* [41] compared the performance of SVR, RF, ANN and K-NN and claimed that the RF algorithm is superior by achieving the highest accuracy. Elavarasan *et al.* [53] proposed the implementation of a hybrid-based feature selection technique by coupling of correlation type filter CFS along with a wrapper RF-RFE based on recursive feature elimination.

D. WHEAT YIELD PREDICTION

Wheat yield prediction is another popular research trend that have been carried out in many studies [23], [51], [57]. Guo and Xue [23] claimed that the spatial NN model was able to estimate the wheat yield providing high accuracy for a given plantation area as compared to the temporal NN models. However, they considered only one factor, namely, the plantation area, which makes the model less reliable. Khaki and Wang [51] achieved a better prediction accuracy utilizing a DNN technique with an RMSE of % of the average yield, whilst the standard deviation was 50% for the validation dataset utilizing predicted weather data. However, the framework was vulnerable to the drawbacks that it the property black box. Cai *et al.* [57] employed LASSO, SVM, RF and NN algorithms to design yield prediction model utilizing climate and satellite data. They achieved higher accuracy and concluded that it would be easier to generate an annual mapping of wheat for using the satellite data as it contributes to the reduction of errors. Han *et al.* [52] investigated a modeling approach for predicting winter wheat yield by integrating soil data, climate data and vegetation index data. The authors employed eight typical machine learning models, and the results showed that SVM, GPR and RF were the top three best techniques to predict yields. Wang *et al.* [38] proposed a framework using a two-branch DL model where the LSTM received inputs from remote sensing and meteorological data. On the other hand, CNN was employed to construct another branch to model static soil features. The finding of the experiment explored that the yield prediction could be obtained with a small loss of accuracy at least one month prior to harvest. Jiang *et al.* [110] claimed that the ANN is better than MLR when remote sensing and climate data are utilized to design a prediction model. Based on study in [111], authors claimed that the association rule mining technique outperformed DT across all locations. Bhojani and Bhatt [60] proposed a new activation function for MLP and revised random weights and bias values for estimating the crop yield with several weather parameter datasets. The outcomes indicated that that newly created activation functions outperformed compared to 'sigmoid', which is a default activation function NN. Nevavuori *et al.* [66] utilized NDVI and RGB data separately to the CNN model and claimed that the CNN architecture performed better with RGB images over NDVI.

E. CORN AND SOYABEAN YIELD PREDICTION

Khaki *et al.* [25] investigated a DL approach using hybrid CNN-RNN for crop yield prediction using environmental data and management data. In order to make the model black box less and more explainable, a backpropagation-based feature selection was conducted. The authors claimed that the proposed method's results are much higher than other approaches, including DFNN, RF and LASSO. However, the proposed model was validated with only two types of crops' data, including corn and soybean. Additionally, an extra feature selection algorithm makes the hybrid CNN-RNN model more computationally complex. Kim *et al.* [42] has developed a DNN model that has the potential to forecast the corn and soybean yields very reliably by utilizing satellite-based vegetation indices, meteorological and hydrological variable dataset in conformity with a high-resolution CDL. Prasad *et al.* [112] developed an approach for predicting corn and soybean yields using piecewise LR method with breakpoint. The remote sensing data and other surface variables were utilized in their study.

F. MAIZE YIELD PREDICTION

Zhang *et al.* [46] applied fluorescence, optical, environmental and thermal satellite data to RF, LASSO, LSTM and XGBoost algorithms for predicting county-level maize yield in China. Outcomes of the study indicated that SIF had better output compared to EVI due to coarse spatial resolution and low signal-to-noise ratio. However, the proposed model was validated with only one types of crop. A regularized LR and kernel ridge regression were employed by Zhang *et al.* [46] to relate EVI and VOD time series. Peng *et al.* [56] developed a successful approach to estimate maize and soybean yield in the U.S. Midwest using satellite-based high-resolution SIF products from gap-filled OCO-2 and TROPOMI. In the study in [50], different climate data, vegetation indices and socio-economic factors are utilized to predict wheat yield in China with the help of ridge regression, RF and LightGBM models. Because of the complexity of the proposed framework produced by the structure of RF and LightGBM model, it is hard to generate hypotheses that can provide biological insights into final crop yields.

G. MISCELLANEOUS CROPS YIELD PREDICTION

Abbas *et al.* [54] employed LR, SVR and k-NN algorithms for potato yield prediction using datasets containing NDVI and soil properties. The SVR models performed better in predicting the potato yield over the other three models. However, due to the impact of other climatic, chemical, weather and physical factors, the variation of prediction in similar fields in different years occurred. In 2019, a mathematical optimization model was developed by Awad [113] in order to predict potato yield that used the biomass calculated by the model. Prasad *et al.* [24] used VCI, GDD, SPI, LST, historical yield data, as the predictors for the RF algorithm to design cotton yield prediction. The outcomes of their study reflect a reliable and faster prediction technique with a performance estimation obtaining the value of *SSR* of 0.69, 0.60 and 0.39 in September, December and February, respectively. Yalcin [45] used plant images for the estimation of sunflower yield by employing pre-trained CNN model. Here, AlexNet was employed as the pre-trained model and the validation accuracy of the proposed system was 87.67%. Lin [40] proposed an integrated model consists of TOMGRO and Vanthoor model for greenhouse tomato yield prediction. However, the model does not result in sufficient modification of the greenhouse environment parameters. Fukuda *et al.* [114] found that RF was ideal for mango fruit yield prediction because of environmental factors of irrigation period and water management. Villanueva and Salenga [44] developed a model with CNN to classify good and bad leaves of bitter melon with very good prediction accuracy. Although the authors claimed this study as the yield prediction, it is very difficult to predict crop yield accurately through the classification of good and bad leaves only. Khosla *et al.* [61] proposed MANNs-SVR approach for the prediction of the number of major Kharif crops, namely bajra, maize, rice and ragi. Authors utilized the MANNs for estimating the amount of rainfall, and then the yielded kharif crops were predicted utilizing the rainfall data along with the area given to that particular crop by employing SVR. However, only rainfall and area attribute are not enough to design a strong model since the yield of particular crops varies with a variety of factors, most notably irrigation, fertilizers used and many more. Jeong *et al.* (2016) [115] proposed a framework for predicting maize, potato and wheat yield using MLR and RF. The RF performed significantly better in predicting crop yields than MLR. Various ML and LR models were compared for crop yield prediction over a variety of crop datasets by Sanchez *et al.* [47], where M5-Prime and K-NN approach performed the best. There are various reasons that influence the accuracy of the above-mentioned methods, such as knowledge representation, model structure, training time, missing data handling and implementation cost.

H. PALM OIL PRICE PREDICTION

Crude palm oil (CPO) price prediction plays a vital role in agricultural economic development. Because of the volatile pricing of CPO, several researchers have been conducted research to predict how the CPO price can affect the firms that rely on the agricultural commodity. With this demand, a CPO price forecasting approach was proposed by Rahim *et al.* [116] that had employed a weighted subsethood-based algorithm to produce fuzzy rules of forecasting. The idea of Fuzzy rule based systems (FRBS) was rooted in the implementation of Fuzzy time series (FTS) too. The main objective of the proposed approach was to improve the efficacy of time series prediction and provide more reliable performance. The forecasting of the palm oil price fluctuation was conducted using MHDA method by Chai *et al.* [117] in which investor comments and multiple-source information were combined. The accuracy, however, is only 64.15% which is not strong enough for reliable prediction. Several methods are utilized in order to predict the price of CPO, including Fuzzy time series with the proposed sliding window technique [33], LSTM [118] ANN [118], RF [119], SARIMA [120], and DES- EAHW [121] hybrid model consists of Sliding Window and GARCH [122] have been utilized.

I. PALM OIL GROWTH MONITORING

The overall palm oil yield prediction would significantly depend on the reliable palm oil growth prediction. Several research projects have been conducted on predicting as well as monitoring palm oil development. A scattering model was developed by Teng *et al.* [123] for oil palm based on the radiative transfer equations. It was also capable of solving iteratively up to the second-order equation for understanding the backscattering behavior of oil palm canopies clearly with SAR image interpretation. The model is developed for calculating the backscattering coefficients of oil palm canopies at several stages of growth and for different polarizations, incident angles and frequencies. The proposed model, in addition, can predict SAR images at L-band, which are more sensitive to changes in the structure of the canopies, especially that of the fronds. Thus, the model has become an efficient tool for oil palm growth monitoring and disease detection. Toh *et al.* [124] proposed a model for the classification of L band SAR image with respect to the oil palm growth stage by the hybridization of SVM and CNN. The author generated a large set of simulation data by employing RT model to supplement additional data. The hybrid classification model achieved 90% of training accuracy. A ML-based approach was proposed by Yousef *et al.* [81], which recognizes inflorescences anthesis stages of female oil palm by utilizing thermal features. The best classification outcomes occurred in RF algorithms than other ML approaches. An automated inspection system for the oil palm FFB was developed by Alfatni *et al.* [125] that utilizes color histogram technique based on the ANN classifier. Ellsäßer *et al.* [126] developed a prediction model for the sap flux and leaf stomatal conductance based on RF that employs drone-recorded and meteorological data. However, an encouraging outcome was achieved in their study. The genericity of this approach is limited since it was conducted by only one study region and only oil palm systems with a period of less than one-month of measurement.

J. NUTRIENT CONTENT OF OIL PALM PREDICTION

In order to point out the quality of fresh fruit bunch (FFB), the classification of oil palm nutrient level plays an important role based on leaf samples. The nutrient level in oil palm leaf samples is classified by Nawi *et al.* [68] that analyses the types of nutrient and fronds. It also used to determine the nutrient level in spectral reflectance of wave-length. The classification of spectral reflectance data gave very desirable outcomes. This data was collected utilizing a spectroradiometer. The ANN framework was conducted to determine the nutrient content level for oil palm leaf samples with a prediction accuracy of 85.32%. Future research is recommended by applying a larger sample size and consideration of other nutrient types for the development of a robust classification approach. Ashraf *et al.* [127] implemented an approach for monitoring the nutritional deficiencies such as magnesium, potassium and nitrogen deficiency using SVM by analyzing the oil palm's leaf surface. An RF-based chlorophyll level of mature oil palm monitoring system was developed by Amirruddin *et al.* [128] that utilizes hyperspectral remote sensing data. A promising potential outcome was achieved in the RF algorithm than the DT in assessing the chl content of mature oil palms utilizing hyperspectral data. A hybrid system of LMT-SMOTE+AdaBoost was developed by Amirruddin *et al.* [129] for monitoring the nutrients status (calcium (Ca), potassium (K), magnesium (Mg), nitrogen (N) and phosphorus (P) sufficiency levels) of mature oil palm using hyperspectral spectroscopy. The results showed that frond 9 was more reliable than frond 17 to monitor the oil palm nutrient status via remote sensing platforms.

K. MAPPING OF OIL PALM PLANTATION AND TREE COUNTING

The number and distribution of oil palm trees in a palm oil plantation are important in various ways, such as predicting palm yield, knowing the survival rate and age of palm trees after plantation, and making a projection on their yield or lifespan, as well as understanding the administration process of palm oil irrigation [130]. Different automatic palm tree detection algorithms have been implemented in response to the increased availability and variety of high-resolution remote sensing images. To detect oil palm tree utilizing large scale high- resolution satellite images, different deep learning approaches including Faster R-CNN [89], [91], CNN [131], [132] have been employed. Xu *et al.* [87] developed IGSO-RF model utilizing remote sensing data to classify young and mature palm plantation. Although the high performance was achieved in their study, different issues should be addressed including the experimental area under humid tropics, inadequate samples and data of high-resolution images. A novel automatic approach was implemented by Wang *et al.* [133] for palm tree detection with UAV images where the overall accuracy was achieved by 99.21%. Hence, for feature extraction, the HOG algorithm was utilized that were capable of describing the texture of palm trees. Moreover, SVM was employed for performing the binary classification task. Zheng *et al.* [134] developed a novel oil palm tree counting and detection algorithm (MADAN) utilizing remotely sensed images of large scale oil palm plantation areas. A hybrid CNN-SVM method was developed by Puttinaovarat and Horkaew [135] for oil-palm planation detection from THEOS images. Hence, CNN in the first step was intended to remove irrelevant images. In addition, SVM was utilized to extract planation pixels, provided a priori of their existence. Shaharum *et al.* [136] developed REMAP based cloud computing platform in which an open-source stacked Landsat data was used that allows land cover classification to be implemented using a built-in RF algorithm. However, the optical and radar data can be integrated to enhance image quality. Other ML algorithms should be tested to achieve optimum approach, including SVM, ANN and DL. Another study was carried out in [137] in which Landsat 8 data were processed utilizing a cloud computing platform of GEE to classify oil palm land cover employing SVM, CART and RF. Hence, the SVM outperformed the CART and RF. Dong *et al.* [138] implemented a novel deep learning-based semantic segmentation technique, namely RCANet for oil palm plantation mapping that uses remote sensing images. Hence, the great time consumption is still a restriction of RCANet for large-scale oil palm mapping.

L. OIL PALM DISEASE RECOGNITION

Intelligent agriculture requires extensive use of image recognition of agricultural disease. Several machine learning approaches along with more recent artificial intelligence (AI) techniques like deep learning and transfer learning, have begun to be applied to agricultural diagnostics. Fahy *et al.* [139] implemented a multi-agent for estimating the spread of Lethal Wilt disease that uses five years' historical data. This forecasting model incorporates of KDE and biased-walk, which does a much better job of predicting the high-risk areas of infection. Alaa *et al.* [88] suggested an image recognition system using CNN that detects Oil Palm Diseases, which include Leaf Spots, Blight Spots, and Blight Weevil. Husin *et al.* [140] implemented an identification method for palm oil yield with 85% detection performance. Montero *et al.* [90] proposed a classification algorithm with binary SVM using 798 aerial images from their UAV for the detection of Bud Rot. Although the system's accuracy is high, this study's dataset was insufficient. Husin *et al.* [141] employed TLS data to study crown shape measurements in order to construct a stratal model for BSR detection. It is possible to generate more detailed models that take into account the percentage of laser returns, leaf orientation, and occlusion in oil palm trees. Ruslan *et al.* [142] analyzed the relationship between humidity and Metisa plana's infestations. ANN outperformed traditional regression models when both were applied to predict Metisa plana's outbreak.

M. PALM OIL YIELD PREDICTION

The major goal of this article is to provide an in-depth analysis of palm oil palm oil yield prediction. Very few studies have been conducted to investigate the palm oil yield prediction. Siregar *et al.* [143] proposed exponential smoothing (ES) technique by utilizing dataset consists of the production of palm oil data, production of palm oil core data and production of FFB for the prediction of the palm oil yield. Hilal *et al.* [59] proposed a Genetic algorithm to find the best combinations for model variable and then utilized correlation analysis to predict the palm oil yield. A considerable number of studies have also been conducted to predict palm oil prediction utilizing LSTM [67], ANN [144], [145].

N. MISCELLANEOUS ML APPLICATION IN OIL PALM INDUSTRY

A significant number of factors influence the prediction of oil palm yield, including CC effect, LAI-oil relationship and process time. A small number of researches was conducted for investigating these factors. Shanmuganathan *et al.* [146] developed a model for discovering new insights that explores the effect of climate change on oil palm yield by hybridizing of data mining frameworks and statistical analyses. They employed an inadequate dataset for conventional analyses. The relationship between LAI of oil palm with microwave backscatter was explored by the authors in [147]. A four-layer DNN was used for the oil palm lands classification based on their LAI. An ANN model was implemented as a solution for the prediction of process time for palm oil production by Adizue *et al.* [148] where the number of staff and the computed time in each process was analyzed. However, the outcomes of their study were not compared to the related works because of not existing any known replica of these processes.

VI. CHALLENGES AND FUTURE DIRECTIONS TO THE CROP YIELD PREDICTION

A. LACK OF STANDARD COMBINATION OF FEATURE SET

In this review, the entire feature lists are categorized into thirteen groups so that the readers may easily comprehend the features of crop yield prediction. Extensive information is missed because of this judgment, but transparency has been preserved. The most used features, together with a number of studies where they were employed, have been illustrated in Fig. 9. According to Fig. 9, climatic information and historical crop yield data were utilized in 30 and 32 studies respectively to predict crop yield. The most utilized feature is vegetation index and satellite data, which is utilized in 35 studies. Besides these, the cropland information (16 studies), soil properties (17 studies), irrigation information (9 studies) and crop management data (8 studies) are moderately used in the prediction of crop yield. Although there is no doubt about the widely used and most efficient group of features for crop yield prediction, it is difficult to say the most efficient sub-feature list under each feature group. For example, under climatic information, there

TABLE 4. Different features and Prediction algorithms for crop yield prediction.

Reference	Objective	Feature	Prediction Algorithm	Performance
[25]	Corn and soybean yield prediction	Historical yield data, weather parameters, soil's characteristics data, crop management data	Hybrid CNN-RNN	Validation RMSE:- corn yield: 24.10, soybean yield: 6.35
[42]	Corn and soybean yield prediction	Cropland information, satellite images, meteorological data, hydrological data, crop yield statistics	MARS, SVM, RF, extra tree, ANN, and DNN	For optimized DNN model:- corr coef: 0.945 (corn), corr coef: 0.901 (soybean)
[53]	Paddy yield prediction	Climate information, soil productivity, groundwater characteristics and availability	RF, DT, gradient boosting	Model accuracy:-RF: 91.23%, DT: 82.58%, gradient boosting: 85.41%
[26]	Paddy yield prediction	Cropland information, climate information, soil properties, agricultural production data, irrigation information	Hybrid MLR-ANN model	RMSE:0.051, MAE:0.041, R: 0.99
[45]	Sunflower yield prediction	Field image, crop growth information	CNN	Validation accuracy: 87.67%
[40]	Tomato yield prediction	Climate information, crop growth information	Hybrid TOMGRO and vanthoor model	RMSE: 2.5974
[48]	Corn yield prediction	Irrigation information, soil properties	BPNN	Accuracy: 83.50% to 96.04% (PVI)
[56]	Corn and soybean yield prediction	Historical yield data, climate information, MODIS data, satellite-based SIF data	LASSO, RIDGE regression, SVR, RF, and ANN	Random forest climate + SIF (OCO2):- RMSE: 18.11 (corn), RMSE: 5.31 (soybean)
[44]	Bitter melon yield prediction	Leave images	CNN	Accuracy: 100%
[46]	Maize yield prediction	Irrigation information, soil properties, climatic information, satellite variables, cropland information, historical yield data, satellite-based SIF data	LASSO, RF, XGBoost, LSTM	R-squared value for SIF:- LASSO: 0.39, RF: 0.75, XGBoost: 0.77, LSTM: 0.68; R-squared value for EVI:- LASSO: 0.38, RF: 0.76, XGBoost: 0.75, LSTM: 0.69
[27]	Soybean yield prediction	Historical yield data, climate information, MODIS data, cropland information	CNN-LSTM	RMSE: 329.53, R-squared: 0.78
[23]	Wheat yield prediction	Cropland information	ANN (MLP)	MSE: 0.0001
[24]	Cotton yield prediction	VCI, GDD, SPI, LST, historical yield data	RF	R-squared: 0.69 for September, R-squared: 0.60 for December, R-squared: 0.39 for February
[54]	Potato yield prediction	Soil electrical conductivity, soil moisture, soil slope, NDVI, and soil chemistry	LR, elastic net, k-NN, and SVR	RMSE:- LR: 4.69, elastic net: 4.72, k-NN: 5.23, SVR: 4.62
[61]	Bajra, maize, rice and ragi yield prediction	Rainfall and field area	MANNs-SVR	N/A

TABLE 4. ***(Continued.)*** **Different features and Prediction algorithms for crop yield prediction.**

[51]	Maize yield prediction	Crop genotype, yield performance, weather and soil	DNN, LASSO, SNN, DT	RMSE:- DNN: 12.79, LASSO: 21.40, SNN: 18.04, DT: 15.03
[1]	Paddy yield prediction	Fertilization information, groundwater characteristics, crop management data, cropland information, historical yield data, irrigation information, soil properties, climatic information	Deep reinforcement learning	Accuracy: 93.7%.
[66]	Wheat and barley yield prediction	Spectral image from NDVI and RGB	CNN	MAPE: 8.8%
[62]	Corn, soybean and wheat yield prediction	Vegetation optical depth (VOD), enhanced vegetation index (EVI)	Regularized linear regression, Kernel ridge regression	R-squared:- RLR: 0.8 ± 0.03, KRR: 0.82 ± 0.03
[13]	Corn yield prediction	Crop management data, climatic information, soil properties, historical yield data	LR, LASSO, XGBoost, lightGBM, RF, optimized weighted ensemble	RRMSE:- optimized weighted ensemble: 9.56%
[50]	Wheat yield prediction	Cropland information, crop yield data, soil properties, socio-economic factors, climate information, vegetation indices	Ridge regression, random forest, light gradient boosting	R-squared: (0.68 0.75)
[43]	Rice yield prediction	Phenology data, climatic information and cropland information	BPNN, RF, SVM	RMSE (R-squared):- BP: 800 (0.24) kg/ha, SVM: 737 (0.33) kg/ha, RF: 744 (0.31) kg/ha
[57]	Wheat yield prediction	Cropland information, crop yield data, satellite-based SIF data, vegetation index, climatic information	LASSO, SVM, RF, NN	R-squared: 0.75
[49]	Corn yield prediction	Climate information, Soil properties, irrigation information	SNN, OLS regression	RMSE
[41]	Paddy yield prediction	Irrigation information, cropland information,crop yield data, crop management data, fertilization information	ANN, SVR, KNN and RF	RMSE:- ANN: 0.098, SVR: 0.099, K-NN: 0.127, RF:0.085
[52]	Wheat yield prediction	Climate information, soil properties, vegetation index, cropland information, crop yield data	SVM, GPR, RF, DT, NN, K-NN, boost trees (BST) and bagging trees (BGT)	R-squared:- GPR: 0.79, SVM: 0.77, RF: 0.81
[47]	Numerous crops yield prediction	Climate information, cropland information, irrigation information	MLR, DT, NN, SVR, K-NN	RMSE:- MLR: 5.41, DT: 5.60, NN: 5.14, SVR: 4.91, K-NN: 5.02
[55]	Soybean yield prediction	Temperature, precipitation, EVI, LST, NDVI	MLR, RF, LSTM	MAE, MSE, RMSE
[63]	Soybean yield prediction	Vegetation index	DNN, SVR, RF	R-squared:- SVR: 0.669, RF: 0.662, DNN: 0.720
[65]	Corn yield prediction	LAI, NDVI	LR	LAI:- R-squared: 0.92–0.94, NDVI:- R-squared: 0.85

TABLE 4. *(Continued.)* **Different features and Prediction algorithms for crop yield prediction.**

[109]	Soybean yield prediction	Surface reflectance, land surface temperature and land cover type	CNN	RMSE: 0.81
[64]	Soybean, corn yield prediction	Vegetation index, topographic attributes, structural variables	ANN	Overall RMAE for corn is smaller compared to soybean
[60]	Wheat yield prediction	Climate information	ANN	MAE, MSE, RAE, RMSE, MAPE and RRSE
[38]	Wheat yield prediction	Climate information, soil properties, vegetation index	DNN	R-squared: 0.77, RMSE: 721 kg/ha, MAPE: 15.38%;
[39]	Paddy yield prediction	Vegetation index	OLS, SVR and GWR	Error rate: 0.06% 13.22%.
[68]	Nutrient level in oil palm leaf prediction	Types of nutrient and types of fronds	ANN	Classification accuracy: 85.32%
[148]	Palm oil production time prediction	The number of staff and the computed time in each process	ANN	R: 0.99902, MSE: 0.01261
[124]	Oil palm growth stage monitoring	L band SAR image	Hybrid CNN-SVM	Accuracy: 90%
[143]	Forecasting of the real palm oil production	Production of FFB, production of palm oil data and palm oil core data	Exponential smoothing method	RMSE: 0.1
[89]	Oil palm tree detection	High-resolution satellite images	Faster R-CNN	Precision: 96.79%
[131]	Oil palm tree detection	High-resolution satellite images	CNN	Precision: 96.79%
[91]	Oil palm tree detection	High-resolution satellite images	Faster R-CNN	F1-score: 94.99%
[149]	Oil palm tree detection	High-resolution satellite images	CNN	Classification accuracy: 97.5%
[59]	Palm oil yield predictions	Historical yield data, cropland information, climatic information, air pollutants	Genetic algorithm/correlation analysis (GA/CA)	R-squared: 0.948, MSE: 0.022
[58]	Prediction of Metisa Plana outbreak in oil palm plantation	LST, rainfall, humidity, and NDVI	ANN	Accuracy: 95.42%
[144]	Palm oil yield predictions	Climatic information	ANN	MAE: 0.5346, MSE: 0.4707
[67]	Palm oil yield predictions	Historical yield data	LSTM	MAPE: 2.7098%
[87]	Detection of mature and young oil palm plantation	Spectral feature, SAR backscatter, vegetation indices, and texture attributes	IGSO-RF	Accuracy: 96.08%
[127]	Nutrient level monitoring	Red pixels, entropy and correlations	SVM	Accuracy: 100%
[88]	Oil palm disease recognition	Normal and thermal image	CNN	Accuracy: 97.9%

TABLE 4. *(Continued.)* **Different features and Prediction algorithms for crop yield prediction.**

[133]	Oil palm tree detection	UAV images	SVM	Accuracy: 99.21%
[140]	Oil palm disease recognition	Frond number, frond angle, crown area, crown significance	NB	Accuracy: 85%
[81]	Female inflorescences anthesis stages of oil palm recognition	Thermal image	RF	Accuracy: 88.70%
[125]	Monitoring oil palm FFB	Color features of oil palm FFB image	ANN	Accuracy: 94%
[134]	Oil palm tree counting	Satellite images of large-scale oil palm plantation area	MADAN	Accuracy: 84.81%
[135]	Detection of oil palm plantation	Satellite images of large-scale oil palm plantation area	CNN-SVM	Accuracy: 92.29%
[150]	Detection of oil palm plantation	High resolution images from UAV	SVM	Accuracy: 96.1%
[151]	Nitrogen status in mature oil palm	Satellite images	SVM	Accuracy: 81.82%
[152]	Oil palm price forecasting	Monthly manufacturing data of palm oil	RBFNN-2SATRAAIS	Accuracy: 90.46%
[128]	Chlorophyll levels monitoring in oil palm	Hyperspectral remote sensing data	RF	Accuracy: 92.79 98.77%
[129]	Monitoring of oil palm's macronutrients	Hyperspectral spectroscopy data	LMT-SMOTE+AdaBoost	Accuracy: 76.13 100.00%
[153]	Individual palm tree detection	Remote sensing image	CNN	MAP: 86%
[90]	Bud rot disease detection	Aerial images UAV	SVM	Accuracy: 93.53%
[154]	Oil Palm FFB maturity prediction	Optical spectrometer	Lazy KStar	Accuracy: 63%
[136]	Mapping of oil palm plantation	Optical remote sensing data	RF	Accuracy: 80.34%
[137]	Mapping of oil palm plantation	Optical remote sensing data	SVM	Accuracy: 93.16%
[138]	Mapping of oil palm plantation	High-resolution remote sensing images	RCANet	Accuracy: 96.88%
[155]	Quality prediction and diagnosis of refined palm oil	Raw CPO data	PCA	MSE less than 0.01

TABLE 4. *(Continued.)* **Different features and Prediction algorithms for crop yield prediction.**

[156]	Oil palm FFB ripeness grading	Oil palm FFB image	ANN	Accuracy: 93%
[157]	Retrieving the main commodity maps	GEE based remote sensing data	RF	Accuracy: 95%
[141]	Oil palm BSR disease recognition	TLS data	MLR	Accuracy: 92.5%
[158]	BUNCH-HA, ABW, BNFY	Crop management data	Bayesian networks	Accuracy: BNFY-75%, BUNCH-HA-85%, ABW-90%

are 18 sub-features recorded. Among these sub-features, precipitation, vapor pressure deficit, humidity, solar radiation, temperature, rainfall, and vapor pressure have been widely employed by researchers. However, the optimum set of sub-feature under climatic information for any specific crop is not clearly mentioned in the previous studies. Hence, more studies should be conducted to investigate the optimum set of sub-feature under climatic information for any specific crop. Moreover, the list of optimum climatic sub-features based on different regions should also be explored. The crop yield prediction in similar fields should be recorded in several years since there are variations of existed factors, including weather, climactic, chemical, and physical variables [54]. For example, the crop yield is partially affected by climate changes [159]. It was reported by Maqsood *et al.* [160] that extreme weather indices calculated for around 39% of the tuber production variation, whereas the remaining of the change in tuber production was explained by the other parameters, including efficient strategy, fertilization, pure seed, efficient agricultural technology, soil properties, field topography, supplement irrigation, hydrologic, physical and chemical properties, and so on. A wide range of variables are used to represent genotype, and environment data do not have similar importance in the forecasting of crop yield. Thus, it is essential to identify the most vital features and remove the other redundant ones that may lessen the performance of predictive models.

Satellite-based SIF features might be the potential feature for predicting crop yield. In the study [56], the authors reported that the implementation of SIF features of high-resolution from TROPOMI and OCO-2 could considerably enhance the performance of yield forecasting in relation to the coarse-resolution SIF features from GOME-2. There are potential ways for improving the performance of yield forecasting utilizing SIF. Initially, numerous ways of utilizing SIF data to generate crop yield forecasting algorithms that may indicate to performance variation. For instances, the conversion of SIF into a whole canopy was found to correlate better with the total emission of canopy

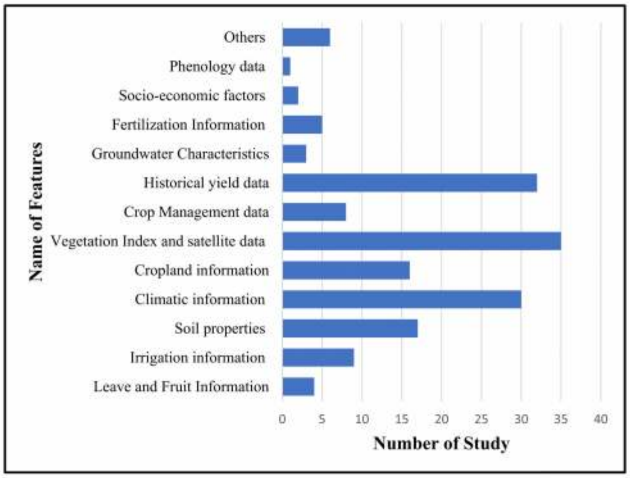

FIGURE 9. **The popular feature set for crop yield prediction.**

photosynthesis [161], [162], that could increase the crop yield prediction too. Secondly, by increasing the training data, the performance of yield prediction utilizing SIF products may be further enhanced. Finally, employing better quality of future SIF products may further enhance the yield prediction performance in new satellite missions such as FLuorescence Explorer (FLEX), which may supply SIF products with higher spatial resolutions than existing SIF products. Downscaling statistical knowledge also has the potential to further improve the spatial resolution of existing SIF products, while previous attempts have mostly concentrated on scaling up the coarse-resolution SIF products, including those from GOME-2 [163].

The results in [46] demonstrated that satellite information supplied huge data in details where the latest identified SIF had a marginally higher score than EVI, which was mostly due to the fPAR. It was also stated that the fluorescence production was subjected to the amount of moisture in the root zone soil that provided more knowledge about drought and heat stress [164]–[166]. Both LST-based KDD and GDD were positively correlated with yield and were more prone to plant water situations compared to Tair throughout AEZs, which agreed with studies showing that LST might be a useful predictor of crop water stress [167], [168]. The combination of SIF and static environmental factors were reasonable in producing yield projections that were more accurate than those from using static weather parameters only (i.e., climate information, soil information, and irrigation properties). Using the peak or late of climate information on the top of them could make models more accurate. Although satellite data indicate the advantages of monitoring crop biomass, in the end, yield weights will determine the final yield. The grain number in the period of flowering and the individual grain measurements in the period of grain filling are related [169]. Additionally, several agronomy experiments have demonstrated that there is a great impact of drought and heat in the above periods for crop growth [170]. The findings in this study have clarified why applying the silking or maturity stage of climate factors had such significant changes in crop yield prediction. Other factors such as soil properties should be applied beside satellite data in predicting crop yield at regional scales. The cause was that their findings regarding environmental stress on crop growth were unique and special [171]–[174]. Thus, it is suggested that the integration of satellite data from various spectral bands and multiple environmental variables should be employed for the large-scale crop yield prediction.

There is no linear relation between crop production and its environment, and the features collected by the image may not be adequately included. The study in [66] reported that there had been low levels of prediction error, and the resulting model error can be effectively constrained by incorporating multi/hyperspectral data, temporal image data, soil and environmental properties in the function matrix. Due to their relatively long sequence and high-spatial-resolution, visible and NIR-based VIs are commonly dominated [175], [176]. Nevertheless, other spectrum satellite data may supply extra information on plant growth and development [62], [177]. Additionally, other elements, including weather variables and soil information, historical yield data, cropland information, irrigation information and crop management data affecting crop yield significantly, contains abiotic information and cannot be captured via satellite data [178]. More studies should be conducted to integrate multi-band satellite data with weather parameters to forecast crop yields at the county level.

B. FINDING THE MOST SUITABLE PREDICTION ALGORITHM

A wide range of classification and regression algorithms have been employed in previous studies to predict crop yield. Fig. 10 illustrates the name of prediction algorithms, together with a number of studies where the algorithms were employed. According to the extracted data, the most utilized crop yield prediction algorithm is ANN (21 studies), and the second most used algorithm is RF (24 studies). The other popular algorithms, namely LR, CNN, SVM, SVR and LASSO, were utilized in 13, 16, 11, 8 and 5 studies, respectively. Although ANN is the most popular prediction algorithm, there are still some issues for further investigation. For example, a number of hidden layer and type of activation function play an important role in the performance of ANN. To get the optimum number of hidden layer and best activation function, different optimization algorithm, including a genetic algorithm and PSO, could be an effective alternative.

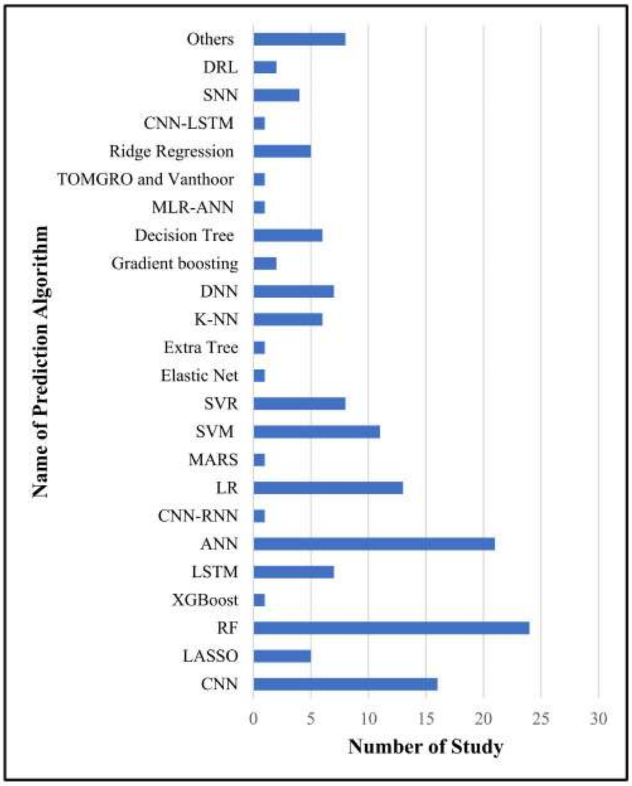

FIGURE 10. Widely used prediction algorithms for crop yield forecasting.

The second most employed algorithm, as per Fig. 10 is linear regression. LR is used to verify whether or not the implemented method is superior to LR as a benchmarking algorithm in many instances. Therefore, it does not imply that it is the highest performing method, although it is presented in several papers. Fig. 10 needs to be viewed carefully that the words "most used" and "best performing" don't mean the same.

In the field of ecological studies, RF is used extensively to investigate the distribution of species and habitat suitability modeling [179], [180]. Very few studies to date have studied the RF algorithm's potential for agricultural [10] and ecological [181] regression analysis. The benefits of utilizing the RF method are; variable collinearity issue can be fixed by utilizing RF regression models that are mostly attracted by using conventional LR models. In the RF model, which offers a benefit overline regression models, simultaneous and discrete variables could be utilized [24]. There are also some demerits of the RF models. There is a possibility of overfitting in the model's predictions outside of the range of training results. The estimation of crop yield might be complicated and unpredictable because of the extreme environmental conditions existing in the different fields. In extrapolating the findings, this constraint may be crucial for RF regression and also due to the absence of sufficient training of datasets. The built RF models have underestimated the yield below the average and overestimated the yield above the average. This issue could be mitigated by continuing to increase the number of observations together with the appropriate predictors for training. The outcomes of the current analysis showed the ability of RF regression to predict crop production with long-term agrometric variables.

In studies [54] and [182], the performance of SVR was much better than the K-NN, BNN and LR. The explanation for improved SVR efficiency due to improved optimization strategies for a large number of parameters [54]. SVR offers the kernel's additional feature, improving prediction model capabilities through the understanding of features. SVR has the potential to manage the distribution, geometry, and overfitting of data in a more compact manner than LR. The SVR theory is based on the concept of reducing the structural uncertainty that decreases the upper bond error as opposed to the training error [54]. The poor efficiency of k-NN is appeared due to an excessive number of attributes involved in the model. It is obvious that for less correlated parameters, the efficiency of the k-NN method is higher [54]. It appears that k-NN performed better on variables with the nonlinear relationship, although more research is needed to validate this assertion.

According to the findings from previous crop yield prediction [46], the ML and DL techniques certainly exceeded LR, primarily due to LASSO's ability to isolate the dynamic correlations between the variables and the target predictor [49], [57], [183]. The DL is a subsidiary of ML, was recently utilized for the issue of crop yield projections and is thought to be very convincing. The major difference between ML and DL is the low performance of the DL network with a small training sample. In addition, the DL approach can extract key functions from input data automatically; however, in other studies, the features are manually extracted, and effective DL might not be utilized appropriately [66], [184], [185]. However, it is recommended that extensive studies on the utilization of DL techniques in crop yield forecasting should be pursued since the DL approaches perform better in other problem areas.

Among the selected articles, both clustering and classifiers frameworks are employed. Since images are employed in certain articles for clustering, the article uses a numerical dataset in conjunction with machine vision instead of ML. In order to identify various prospects on this issue, the use of clustering architectures may be explored in detail. The DNN loss function is very broad and non-convex, which makes it harder to refine this function due to the fact that it contains many local optima and saddle points [25], [186]. Deeper networks can also have the disappearing gradient problem that can be minimized by residual shortcuts or loss functions [186], [187]. Additional methods for improving the performance of DL models such as batch normalization, dropout, and SGD are also developed [25], [186]. Since NN is the most commonly used algorithm, the goal should be to investigate the degree to which DL architectures have been employed to forecast crop yields. After investigating 30 articles where DL approaches were applied, we found the most favored profound DL architectures were CNN, LSTM, and DNN. Nevertheless, there are also some hybrid DL algorithms CNN-LSTM [25], [27] applied to this problem.

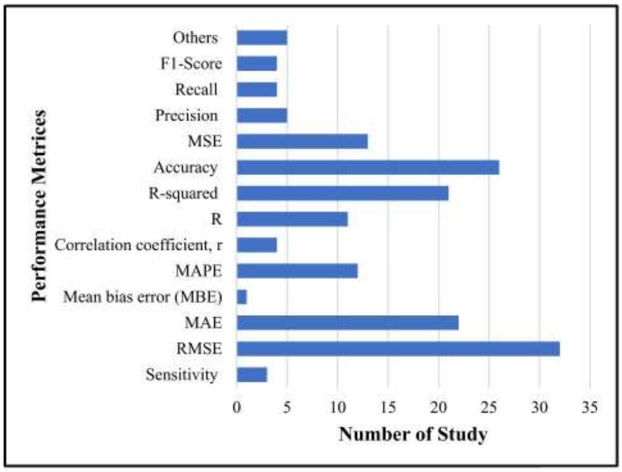

FIGURE 11. Popular performance evaluation metrics for crop yield prediction algorithms.

A wide range of performance evaluation metrics has been utilized in the previous crop yield prediction researches, shown in Fig. 11. According to Fig. 11, RMSE is the highest used metric in the crop yield prediction algorithm. More specifically, the RMSE, R-squared, MAE and classification accuracy have been utilized in 32, 21, 22 and 26 studies as the performance evaluation metrics.

It is almost impossible to compare the same types of crop yield prediction models when they are evaluated by dissimilar performance metrics. Hence, it should be recommended that a standard and systematic approach or a single metric quantify a specific crop yield prediction model. If the same performance metrics are used for the crop yield prediction algorithm, then a direct comparison between different crop yield prediction algorithms is possible.

C. PROSPECTIVE OF PALM OIL RESEARCH

1) APPLICATION OF REMOTE SENSING IN PALM OIL YIELD PREDICTION

The rapid expansion of oil palm plantations has also led to deforestation and a series of negative environmental impacts, such as forest estate losses, social costs, alternative revenue losses, reduced biodiversity, and diminished ecological connectivity [87]. Additionally, the accurate figure of oil palm plantations is necessary for precise palm oil yield prediction. Therefore, to scientifically manage and supervise this activity and to safeguard forests beneficial for the global climate and ecosystem services, it is necessary to precisely detect and monitor oil palm plantations. Remotely sensed information and relevant technologies have currently been applied in various land use analyses, such as plantation area extraction, urban and forest identification, etc. [135]. Particularly, radar satellite image and optical satellite image are predominant sources in oil-palm plantation identification [188] The detection of oil palm plantations using satellite remote sensing data has been carried out in many studies [189], [190]. These modalities, however, differ in their functional characteristics. Radar imaging operates on active sensors. Scene acquisition can thus be made day and night, regardless of weather conditions and unaffected by cloud [135]. Thanks to these advantages, there were a few works that employed radar imagery, such as PALSAR [191]. On the other hand, an optical-based satellite operating on passive sensors, such as THEOS (Thaichote) [192] etc., is unable to as effectively acquire images under varying atmospheric challenges. Nonetheless, for meticulous oil-palm extraction, the abilities to acquire images under such conditions in real-time are not so critical as their spatial resolution. In fact, if an optical image is partially occluded by cloud and/ or defected by other causes, it is typically replaced by that acquired at another available period. Therefore, a majority of recent oil-palm plantation studies has preferred optical-based to radar-based modality [193], [194].

Sometimes the optical method remains difficult to separate oil palm plantations from other spectrally similar vegetation such as forests and rubber trees, and the frequent presence of clouds in the humid tropics hinders image-based method analysis [87]. Due to the similar scattering values for palm trees of different ages, it is difficult to distinguish mature and young oil palm plantations using only SAR data [195]. To overcome the limitations of a single type of data, several recent studies have detected oil palm plantations by using data fusion techniques [130], [196], [197]. Several studies have shown that using the appropriate vegetation indices to analyze and select feature combinations can also yield improved results [198], [199]. For example, a study in [87] achieved promising performance utilizing spectral attributes, SAR backscatter values, vegetation indices, and texture features.

Besides palm oil yield prediction, there some additional applications of remote sensing in palm oil cultivation management. Mapping the distribution of oil palm is crucial in order to manage and plan the sustainable operations of oil palm plantations [137]. Remote sensing provides a means to detect and map oil palm from space effectively. Recent advances in cloud computing and big data allow rapid mapping to be performed over large a geographical scale [137]. Palm tree detection using remote sensing images has received increasing attention in recent years, concerning the issues of sustainability, productivity and profitability [150]. A common practice of chlorophyll (chl) determination has been using chemical analysis that is destructive and time-consuming. A current prospective alternative method such hyperspectral remote sensing offers a nondestructive measurement of chl, which provides a result in a rapid and real-time manner [128].

2) COMMON DISEASES IN OIL PALM PLANTATION AND THEIR IMPACTS

There are some insects, including basal stem rot, bud rot, oryctes, ganoderma, rat and some leaf-eating insects cause significant yield losses in oil palm plantations. Basal stem rot, caused by the pathogenic fungi ganoderma boninense, can devastate old plantations [200]. The onset of infection happens earlier at each replanting if no sanitation measures are taken and can occur as soon as 1–2 years after planting when oil palm is planted after oil palm or coconut [20]. There are some insects, including basal stem rot (BSR), bud rot, oryctes, ganoderma, rat and some leaf-eating in- sects which adversely impact on palm oil yields. The BSR especially infected by the pathogenic fungi ganoderma boni nense, can severely damage old crops [201]. The onsets of infection occur sooner in each replantation when no sanitation steps have been taken [20]. The introduction of a 1- year fallow will decrease the rate of infection slightly but enhance the immature to mature ratio from 0.12 to 0.15. The elimination of disease particles was proposed in mature farming as a crop management technique. However, there is no strong proof of evidence indicating the occurrence of infection is decreased [20]. In America, another deadly disease is extensively found known as bud rot. The outbreaks of this disease have devastated full stands over thousands of hectares over the last few decades [202]. With severe outbreaks, this pest causes up to 100% mortality in South America [20], [203]. Treatment and precautionary initiatives exist but tend to be costly and labor-intensive [20]. Leaf- eating pests are usual in all areas that may cause complete defoliation of palm clusters in case of severe infestation. As a result, after 1, 2 and 3 year of complete defoliation, the yield loss could be 50%, 25% and 15%, respectively [204]. Oryctes is the common pest in inexperienced farming in all areas that reduce the production of 50% in the first year and 20% in the second year through the extreme infection in child plants [205]. Rat is also a common pest in all regions that cause severe infestation when their populations reach >300 individuals per hectare [206]. Destruction of young palms, which results in a prolonged immature period resulting in a yield loss of 5% may be reached in mature farming with rat populations at 'saturation' level [204]. Ganoderma is another common disease in all areas, particularly southeast Asia. Its severe attack may cause up to 80% mortality at >15 years after planting (YAP) in Malaysia and Sumatra. Moreover, palm damages of up to 30–40% at 12 YAP and >50% at 25 YAP in the mild infected region [20]. As a result of faulty, non-standard methods, there is a lot of diversity in the literature, and that has increased confusion for research assessments [207]. Laboratory methods are suitable for detecting environmental changes in the early stages; however, they are expensive, difficult to implement in the field, and not suitable for observation. An innovative system that involves an advanced sensor system can distinguish between healthy and unhealthy oil palm trees with varying levels of accuracy. Unfortunately, these techniques are unable to identify the different levels of infection. Light detection and ranging (LiDAR) is an active ranging approach for measuring the distance or range to a target using pulsed laser light [140]. It can represent the structure and surface appearance of an object as the surface of a tree. Numerous studies [208], [209] have proven terrestrial LiDAR to obtain canopy vegetation profiles and other structural tree properties from an understory perspective. TLS describes as an active ranging method that utilizes laser light, which is an aid in detecting the external shape of the tree's infestation [140].

3) USE OF IOT TO MITIGATE LABOR DEFICIENCY

Another factor known as lack of labour in farming, which is particularly seen in Malaysia. It causes lengthy harvesting rounds, leading to lower rate of oil yield, unharvested bunches and fruits deficit [200]. Malaysian's palm oil industries are experiencing 20–30% shortage of labour force, leading to 15% of yield loss. Labor seems to be more costly and leads to a competitive disadvantage in Latin America and the South. Efforts are being made to create automated alternatives to distribute fertilizers, plant and spread pesticides, but they have not yet proved efficient enough to address the issue of manpower [20], [200]. Hence, the manpower in palm industries needs to be considered for accurate yield prediction. A degree of automation in oil palm plantations is necessary in order to offset that drawback which may lessen the human intervention for every infield inspection and monitoring tasks. Applications of machine learning let better decision making in a real-field environment without or with minimal human intervention. Thus, the utilization of internet of things (IoT) and artificial intelligence (AI) serves to alleviate these human dependency issues by automating responses to real-time issues. Furthermore, it controls labor costs, late pollination as well as any possible injuries, and the requirement for sensor mounting [81].

4) ESTIMATION OF NUTRIENT CONTENTS IN OIL PALM

There are a number of types of agronomic practices, however, which prefer foliar, and soil analyses which are used to assess the nutrients of all plants. Although this technique provides excellent accuracy, it has some limitation including costly, destructive, time-consuming and laborious, because it must be done every year. Moreover, it is not widely applicable to monitor spatial and temporal dynamics of nutrient content due to the vast size of palm oil plantations. In contrast to foliar and soil analyses, the spectroscopic approach offers non-destructive analysis and wide distribution. Moreover, the spectral feature of hyperspectral spectroscopy offers knowledge of the plant's nutrient status since that depends on the organic molecules (cellulose, lignin, protein), leaf pigment (chlorophyll, anthocyanin, carotenoid), leaf structure, water, nitrogen, and variables [129], [210]. Nutrient contents are also important factors that could affect the palm oil yield. Hence, different nutrient contents including phosphorus (P), nitrogen (N), magnesium (Mg), potassium (K) etc., should be considered in the palm oil yield prediction. Table 5 shows the effect and symptoms of deficiency of nutrient contents on palm oil growth and yield.

TABLE 5. **The effect and symptoms of deficiency of nutrient contents on palm oil growth and yield.**

Element	Impact of insufficiency on OP growth and yield	Visual symptoms
Phosphorus (P)	Decreased yield response to N and K fertiliser; Reduction of yield on certain soils	Conical trunk shape
Nitrogen (N)	Reduced vegetative dry object yield; Decreased the rate of assimilation; Reduced the number and weight of bunch; Higher time for phyllochrons	Chlorosis of child leaves
Magnesium (Mg)	Decreased yield response to N and K fertiliser; Reduced oil to bunch ratio; Reduction of yield on certain soils	The colour of old leaves is yellow/orange
Potassium (K)	Reduced vegetative dry object yield; Reduced the number and weight of bunch	Old leaves contain yellow spots

5) EFFECT OF CLIMATE CHANGE IN OIL PALM CULTIVATION

The adverse impact of climate change (CC) on the oil palm sector is becoming increasingly evident [211], [212]. The CC can (1) restrict the present crop areas in their entirety, (2) expand plantations to new zones and deal with problems like the destruction of biodiversity and (3) threaten growers' adaptability. The oil palm cultivation factors can most frequently be influenced by abiotic stresses including precipitation, temperature, soil salinity and carbon dioxide and biotic stresses such as pests, diseases and pollinators [213]. Tropical crops have always been at the limits of growing, in which minor climatic variations can instantly impact survival. Achieving innovative farming techniques and ensuring world-wide food safety in the palm oil industry, it is important to explore the CC impact on oil palm [211], [212]. The effects of CC on crop productivity and oil palm phenology have significant consequences at the regional and global levels. Heavy rains and extreme temperatures are highly beneficial to palm oil production with a lag period of three to four months. Natural disasters like massive flooding and drought are unfavorable in some instances. The higher rain or flooding reduces the CPO and production rate, leading to affect the fruit ripening stage. The fusion of data from multi-sensor is highly significant for oil palm classification. These data can further be utilized in automatic plant counting, age monitoring and change recognition [17], [214]. The open-source data from different sensors including ALOS PALSAR, Sentinel-1, Sentinel-2, Landsat or Google Earth images are very interesting research routes for multi- sensor data exploitation.

6) USE OF UAV IN PALM OIL YIELD PREDICTION

Due to the large cultivation areas of palm trees, such data has been collected using remote sensing. Past studies of

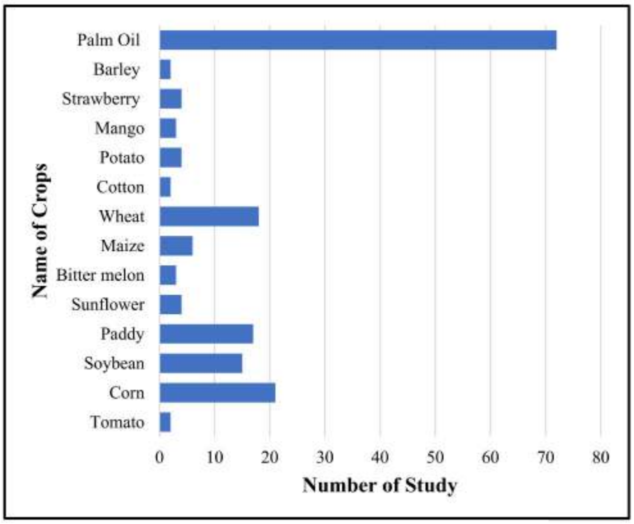

FIGURE 12. Widely used crop in the crop yield prediction research.

the palm trees detection have been mostly limited to commercial high-resolution satellite images [189], [133]. Satellite imagery is expensive and data availability is subjected to frequent delays because of weather conditions. Currently, unmanned aerial vehicles (UAVs) have been used as cost-effective remote-sensing data acquisition systems. It allows the land surface to be mapped and monitor land in high resolution within few minutes. It also allows interactive measurements according to a customer's specific needs. The price and flexibility of UAVs' has allowed for many solutions in the agricultural sector, such as precision agriculture [215] and vegetation monitoring [133], [216]. The higher spatial resolution of UAV images can acquire more detailed data about objects on the ground, which makes renders many techniques less applicable for traditional satellite imaging.

7) OPTIMUM FEATURE SET FOR PALM OIL YIELD PREDICTION

Oil palm is a growing source of vegetable oil that exceeds sunflower, soybean and rapeseed. Hence, the precise palm oil yield forecasting model is an appealing research topic that facilitates the ultimate oil production. The present review investigates the crop yield prediction of 14 different crops shown in Fig. 12. Since this review gives special emphasis on palm oil yield prediction, around 72 studies related to the machine learning-based palm oil are critically evaluated. Besides palm oil, the studies of yield prediction associated with other crops, including corn, wheat, soybean and paddy, are also critically reviewed.

Among 72 studies associated with palm oil, only Hilal *et al.* [59] utilized a maximum number of features under historical yield data, cropland information, climatic information groups. Other studies [67], [89], [91], [124], [131], [144], [149] utilized only one type of feature to predict palm oil yield. Crop yield prediction model using a large number of feature set could be more accurate than the model with a small group of features. For example, the study in [42], the corn and soybean yield has been predicted utilizing five features group, namely cropland information, satellite images, meteorological data, hydrological data, crop yield statistics. Elavarasan and Vincent [1] utilized fertilization information, groundwater characteristics, crop management data, cropland information, historical yield data, irrigation information, soil properties, climatic information to predict paddy yield prediction. Zhang *et al.* [46] employed irrigation information, soil properties, climatic information, satellite variables, cropland information, historical yield data, satellite-based SIF data to predict maize yield. Hence, a large number of features group should also be considered in the palm oil yield prediction research.

Soil properties are the crucial factors that lead to how effectively the nutrients are taken up by palm trees. These nutrients greatly influence palm yield efficiency [17]. Among a wide range of soil properties, the percentage of clay, silt, sand, organic carbon, PH, cation exchange capacity, bulk density, number of macronutrients in the soil, nutrient supplements, soil electrical conductivity, topsoil depth could be investigated for the palm oil yield forecasting. The study of soil properties in the palm field from a remote sensing point of view could be an attractive area of further research. The soil characteristics could be monitored from a distance since the reflected light receives different responses. For example, active remote sensing depends on the dielectric characteristics of the spreading wave may take soil moisture content [217]. In order to build analytical relationships, the correlations between recorded signals and many soil characteristics, including soil texture, soil form and soil structure, can be investigated, which may contribute to minimizing the workload of the soil samples.

However, the estimated parameters such as soil moisture, LAI, greenness of the palm canopy and height that substantially add values in the yield prediction should be considered. Once the connection is defined, a precise yield forecasting model could be created. Yield forecasting performance could be improved by including historical yield data, various climatic data, vegetation data, fertilizer application information, and other types of satellite data.

Ground information can be acquired from a wide area in a short time via remote sensing. It collects spatial data without any direct contact and has been put to use in fields such as activities as biomass estimation, oil palm disease detection, urban areas, agricultural land, object detection, hazard prediction, and biodiversity monitoring [218]. Numerous studies on oil palm plantation mapping have been carried out in response to the increased demand for oil palm products using remote sensing [136].

8) SELECTION OF OPTIMUM ALGORITHM FOR OIL PALM YIELD PREDICTION

An accurate assessment of palm plantation mapping can have substantial economic and environmental benefits. Historically, traditional machine learning techniques, classical image processing methods and deep learning methods can be applied to tree crown detection. In traditional machine learning approaches, many algorithms have been employed to tree crown detection where the most applied classifiers are RF [87], [136] and SVM [133], [150]. Traditional machine learning techniques make great progress compared with classical image processing techniques, however sophisticated techniques or very-high-resolution UAV images are required in most of the cases in this technique. In classical image processing techniques, some factors are usually obtained including local maximum filter, image binarization, and image segmentation [219]. However, complex scenarios such as overlapping tree crowns may cause deterioration of detection outcomes since there is no requirement of labels in this technique. DL-based classifiers, which use multiscale computational methods, have gained widespread adoption in recent studies using remote sensing images [134]. Most advances studies utilize deep learning-based classifier combined with a sliding window-based method to detect tree crowns from satellite images [134]. It is renowned for its notable capacity of feature extraction, which can be accomplished using DL. Progress has been made in the research of object detection based on pretrained models for the detection of palm trees in the plantation area, especially with Faster RCNN [89]. Since the classifiers's predictive accuracy is greatly depends on the input hyper-parameters, the accuracy can be significantly enhanced by the optimization of classifiers' hyper-parameters. Hyper-parameter optimization improves the accuracy thus necessitates hyper-parameter optimization. Compared to single classifier including SVM, NN, CART, NB, MD, the IGSO-RF classifier, which uses the IGSO algorithm to optimize the parameters selected for the traditional RF model, delivered higher accuracies [87].

With the advances in the agricultural sector, disease image recognition has a key role to play in innovative agriculture. Researchers have studied a wide range of diagnostic methods for oil palm diseases. These categories include: SVM [90], CNN [88], ANN [58], NB [140], and MLR [141]. Thus, when applying traditional ML methods to monitor palm oil diseases, there are some disadvantages. For example, the existing techniques are highly dependent on the quality of the original disease images. In addition, applying these methods, as well as image preprocessing, image segmentation, feature extraction, and classification, usually incurs significant operations that add to the complexity and slow down and delay the implementation [220]. Moreover, the training process is very difficult when using traditional machine learning methods, especially if the training dataset is large. More recent advanced machine learning techniques, such as deep learning and transfer learning, have the potential to aid in the development of recognition of oil palm diseases.

Since oil palm plantations are time-consuming and labor-intensive to monitor, this is an extremely arduous process. Numerous machine learning approaches such as RF [81], [126], [128], ANN [125], and hybrid CNN-SVM [124] have been applied to monitor palm growth using remote sensing. Yet it is still critical to extract valuable information from remote sensing data for plantation owners. This may provide solutions through deep learning, transfer learning, and object recognition.

Although a large number of ML-based studies have been carried out in palm oil industry, very few researches are directly related to the palm oil yield prediction. Only five studies [59], [67], [143]–[145] have been recorded that focus palm oil yield prediction. Based on these studies, it is difficult to regard the best algorithm for palm oil yield prediction. However, ML-based some regression algorithms, including RF, ANN and SVR could be very effective to predict palm oil yield. Moreover, rather than a single algorithm, the ensemble of multiple algorithms should be investigated to increase the robustness of the prediction model.

9) STRENGTHS AND WEAKNESSES OF SOME POPULAR ALGORITHMS IN OIL PALM INDUSTRY

There are some strong reasons in favor of SVM that make its widespread utilization in the field of palm oil industry. Regularization may be the first factor in this success. In agricultural data, features are often noisy as well as contain outliers. Regularization may increase the classifier's generalization capabilities which may overcome this problem [82]. As a result, a regularized classifier such as linear SVM have performed better than non-regularized one such as LDA. A few hyperparameters are existed in SVM that need to be defined by hand such as the regularization parameter C and the RBF width if using kernel 2. These advantages are occurred due to the expense of a low speed of execution. Moreover, The SVM decision rule is a simple linear function in the kernel space, which also means it becomes stable and has a low variance [82]. Since features in agricultural data are greatly unstable over time, obtaining low variability is crucial for low classification error. The robustness of SVM with respect to the curse-of-dimensionality may be the last probable cause. For small training sets and high-dimensional feature vectors, this has enabled SVM to obtain excellent results [82]. The key benefits of RF are generalization performance, high-speed operation, and is an ensemble of tree-structured classifiers used to identify a non-linear pattern in the data [81]. It has low data preprocessing requirements in training because it is robust to unit differences and can make correct predictions on sparsely annotated data. For feature selection, RF algorithm exposes better outcome in yield. However, input hyperparameters play a significant role in RF. Thus, optimization of RF's hyper-parameters could help to solve this issue. The CNN presents its important and excellent function in the field of computer vision for the target recognition tasks and the remote sensing area [78]. For example, Lee and Kwon [221] proposed the target recognition tasks with hyperspectral images, whereas Jiang *et al.* [222] extended the CNN application to recognize the target in SAR images. Both works performed significantly in the traditional recognition model. Moreover, CNN can deal with different retrieval tasks, including SM retrieval [223]. In contrast to the traditional

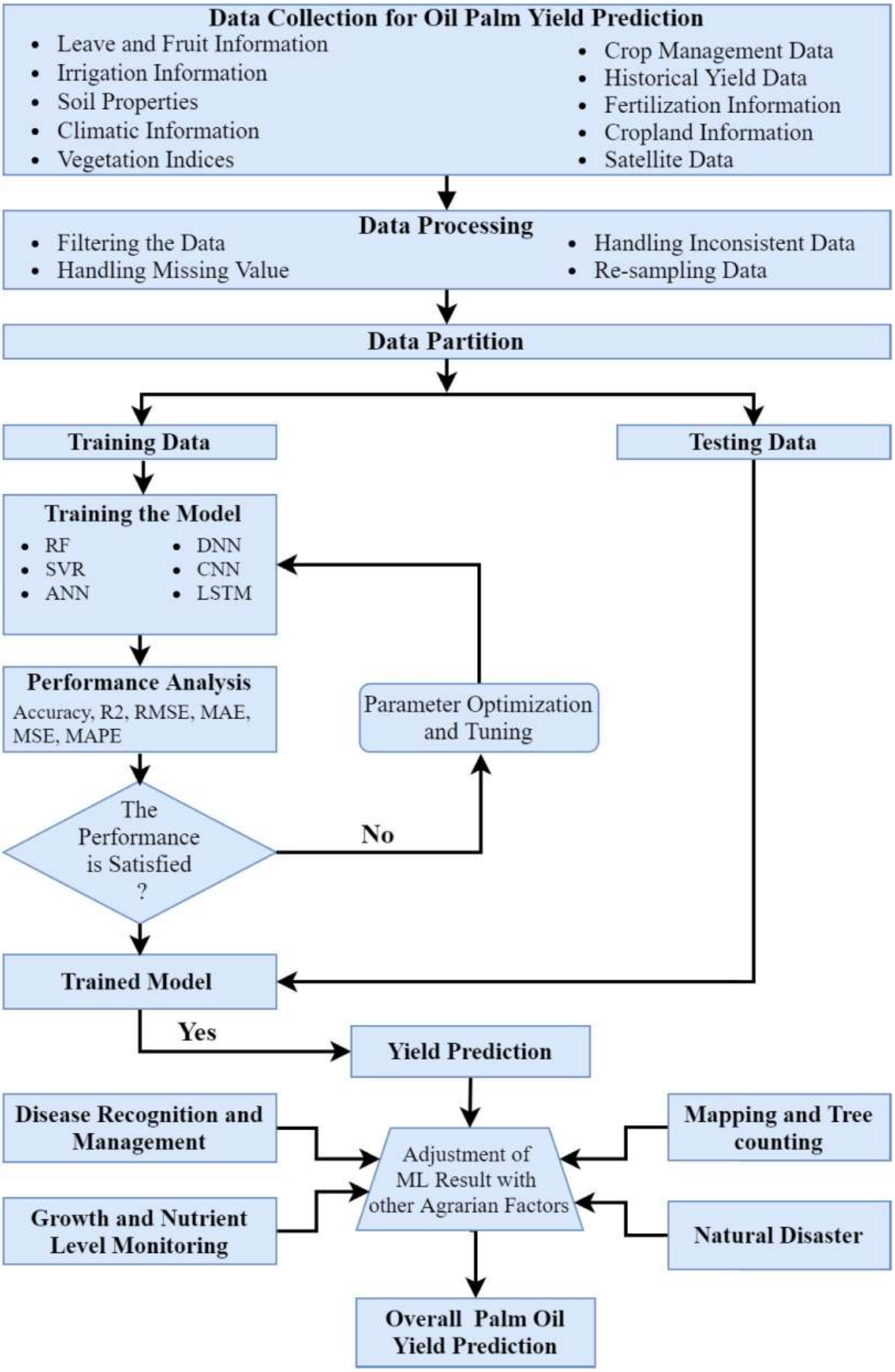

FIGURE 13. Prospective palm oil yield prediction architecture.

NN, CNN has a fully connected layer with multiple local connections. This approach eliminates the computation and automatically discovers relevant features from raw data. The shared convolution kernel is good for data processing with high-dimension. However, CNN demands large data and high processing cost.

10) ALGORITHM's ABILITY TO COPE WITH SPECIFIC PROBLEMS

SVM is one of the best classification algorithms for high-dimensional feature vectors. If a large number of time segments is a factor of high dimensionality, dynamic classifiers can also tackle the problem by considering sequence of feature vectors instead of a single vector of very high dimensionality. Moreover, RF, PCA and correlation coefficient can also handle high-dimensional datasets. This problem may be solved by a combination of classifiers since it lessens the variance, particularly when there is no stationarity in the dataset. The LDA classifier or SVM method is still effective but is likely to be outperformed by a classifier, combinations of LDA or SVM method. Simplistic techniques, such as LDA, may be employed due to the small number of training data set. In the palm oil industry, k-NN algorithms are not used much for the curse of dimension problem. However, if only low-dimensional feature vectors are considered in agriculture as well as the palm oil industry, or no feature vectors are needed, k-NN may be superior. The ANN is widely utilized to predict vegetation parameters and crop yield with remote sensing as it has the ability to retrieve the complex, dynamic and non-linear patterns from the data [10]. They are the earliest form of ML which are well-studied and readily available since many libraries, and software tools are available there. However, there are some drawbacks in practical applications such as learning rate, the selection of the numbers of neurons in hidden layers, overfitting problem and large training dataset. When dealing with large datasets, the process slows down. In training, back-propagation networks become slower compared to other types of networks in which a large number of epochs are required [80]. In the case of image data with a large training sample, CNN could be the best choice. For example, disease recognition using leaf image, mapping of oil palm plantation using remote sensing-based spectral images, nutrient level using plan images and so on. A range of advance CNN architecture including Faster R-CNN, R-FCN, and SSD, are also very effective for object recognition-based tasks. Regression-based algorithms including RF, ANN and SVR could be very effective to predict palm oil yield and oil palm price prediction. Moreover, rather than a single algorithm, the ensemble of multiple algorithms should be investigated to increase the robustness of the prediction model.

11) PROSPECTIVE ARCHITECTURE OF PALM OIL YIELD PREDICTION

In this study, we have reviewed huge number of articles related to the crop yield prediction. Based on critical assessment of related studies, we have proposed forthcoming trend of palm oil yield prediction framework. Fig. 13 illustrates the prospective framework of palm oil yield prediction. In order to predict the crop yield, a wide range of data including leave and fruit information, irrigation information, soil properties, climatic information, vegetation indices, cropland information, crop management data, historical yield data, fertilization information and satellite data is collected in the first step. After data collection, the data need to pre-process for further analysis. Once the data is pre-processed, the entire dataset is split into a training and a testing set. The training dataset is used to train the prediction model. Different ML based regression and classification algorithms are then employed in the model's training phase. If the performance of the trained model is not satisfied, the parameter of the prediction model is optimized. After getting the threshold performance, the trained model is tested through the testing dataset. There are some crucial agrarian factors that have significant impact on crop yield prediction and those agrarian factors include disease recognition and management, mapping and plant counting, plant growth and nutrition level monitoring and natural disaster. Finally, the output from ML is adjust with the agrarian factors' output to get the precise palm oil yield prediction.

VII. CONCLUSION

In order to feed a rising world population, new technology in the agricultural industry needs to be implemented. Apart from this, agriculturists need a proper guideline in time that will allow them to forecast crop yields so that they can formulate effective strategies to maximize crop yields. ML frameworks offer a clear insight into the process by assessing the vast sets of data and interpreting the obtained information. The models describing the correlations between constituents and actions are built through these technologies. In addition, the future reactions in a given situation can also be predicted through the ML models. The present review illustrates that a wide range of attributes is utilized by the selected articles, focusing on the data availability and research scope. Most of the referred articles explore yield forecasting through the ML algorithms. However, the core difference is the implementation of wide ranges of features. Additionally, the difference in crop, location and intensity has also been observed in the studies. The selection of the features relies on the dataset's accessibility and the research objective. The existing kinds of literature also depict that the utilization of extensive features in a model may not always offer the optimum output for the yield estimate. Although it is difficult to acknowledge the optimum method based on the existing findings, the widespread utilization of some ML algorithms and their promising performance are very important to get an overview. The most promising conventional ML architectures are LR, RF and NN. Besides these algorithms, some DL models, including DNN, CNN and LSTM, are also employed in the crop yield estimation. To come to a specific conclusion about the best performing model, some feature selection algorithms, together with existing outperforming models, should be investigated. The present review also shows that very few studies have been conducted to forecast palm oil yield prediction. Moreover, the existing palm oil yield prediction studies have utilized very few feature sets that cause a huge difference between the predicted and actual palm oil yield. It is very early to comment on the best feature set and prediction model to predict the palm oil yield. Hence more studies with a large number of features and a wide range of prediction algorithms should be investigated. This paper will presumably pave the groundwork for extensive research on how crop yield prediction and palm oil yield prediction are linked to each other.

REFERENCES

[1] D. Elavarasan and P. M. D. Vincent, "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications," *IEEE Access*, vol. 8, pp. 86886–86901, 2020.

[2] J. Huang, J. L. Gómez-Dans, H. Huang, H. Ma, Q. Wu, P. E. Lewis, S. Liang, Z. Chen, J.-H. Xue, Y. Wu, F. Zhao, J. Wang, and X. Xie, "Assimilation of remote sensing into crop growth models: Current status and perspectives," *Agricult. Forest Meteorol.*, vols. 276–277, Oct. 2019, Art. no. 107609.

[3] S. Li, S. Peng, W. Chen, and X. Lu, "INCOME: Practical land monitoring in precision agriculture with sensor networks," *Comput. Commun.*, vol. 36, no. 4, pp. 459–467, Feb. 2013.

[4] X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, and A. M. Mouazen, "Wheat yield prediction using machine learning and advanced sensing techniques," *Comput. Electron. Agricult.*, vol. 121, pp. 57–65, Feb. 2016.

[5] M. E. Holzman, F. Carmona, R. Rivas, and R. Niclòs, "Early assessment of crop yield from remotely sensed water stress and solar radiation data," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 297–308, Nov. 2018.

[6] A. Singh, B. Ganapathysubramanian, A. K. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," *Trends Plant Sci.*, vol. 21, no. 2, pp. 110–124, Feb. 2016.

[7] R. Whetton, Y. Zhao, S. Shaddad, and A. M. Mouazen, "Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI," *Comput. Electron. Agricult.*, vol. 138, pp. 127–136, Jun. 2017.

[8] Y. Dash, S. K. Mishra, and B. K. Panigrahi, "Rainfall prediction for the Kerala state of India using artificial intelligence approaches," *Comput. Electr. Eng.*, vol. 70, pp. 66–73, Aug. 2018.

[9] W. Wieder, S. Shoop, L. Barna, T. Franz, and C. Finkenbiner, "Comparison of soil strength measurements of agricultural soils in Nebraska," *J. Terramechanics*, vol. 77, pp. 31–48, Jun. 2018.

[10] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Comput. Electron. Agricult.*, vol. 151, pp. 61–69, Aug. 2018.

[11] B. Basso and L. Liu, "Seasonal crop yield forecast: Methods, applications, and accuracies," *Adv. Agronomy*, vol. 154, pp. 201–255, Jan. 2019.

[12] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, "Maize yield and nitrate loss prediction with machine learning algorithms," *Environ. Res. Lett.*, vol. 14, no. 12, Dec. 2019, Art. no. 124026.

[13] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting corn yield with machine learning ensembles," *Frontiers Plant Sci.*, vol. 11, p. 1120, Jul. 2020.

[14] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Comput. Electron. Agricult.*, vol. 156, pp. 585–605, Jan. 2019.

[15] D. Elavarasan, D. R. Vincent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," *Comput. Electron. Agricult.*, vol. 155, pp. 257–282, Dec. 2018.

[16] M. D. Johnson, W. W. Hsieh, A. J. Cannon, A. Davidson, and F. Bédard, "Crop yield forecasting on the canadian prairies by remotely sensed vegetation indices and machine learning methods," *Agricult. Forest Meteorol.*, vols. 218–219, pp. 74–84, Mar. 2016.

[17] K. L. Chong, K. D. Kanniah, C. Pohl, and K. P. Tan, "A review of remote sensing applications for oil palm studies," *Geo-Spatial Inf. Sci.*, vol. 20, no. 2, pp. 184–200, Apr. 2017.

[18] L. J. Young, "Agricultural crop forecasting for large geographical areas," *Annu. Rev. Statist. Appl.*, vol. 6, no. 1, pp. 173–196, Mar. 2019.

[19] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105709.

[20] L. S. Woittiez, M. T. van Wijk, M. Slingerland, M. van Noordwijk, and K. E. Giller, "Yield gaps in oil palm: A quantitative review of contributing factors," *Eur. J. Agronomy*, vol. 83, pp. 57–77, Feb. 2017.

[21] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018.

[22] B. Li, J. Lecourt, and G. Bishop, "Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield Prediction—A review," *Plants*, vol. 7, no. 1, p. 3, Jan. 2018.

[23] W. W. Guo and H. Xue, "Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models," *Math. Problems Eng.*, vol. 2014, pp. 1–7, Jan. 2014.

[24] N. Prasad, N. Patel, and A. Danodia, "Crop yield prediction in cotton for regional level using random forest approach," *Spatial Inf. Res.*, vol. 29, pp. 1–12, Jul. 2020.

[25] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction," *Frontiers Plant Sci.*, vol. 10, p. 1750, Jan. 2020.

[26] P. S. M. Gopal and R. Bhargavi, "A novel approach for efficient crop yield prediction," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104968.

[27] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," *Sensors*, vol. 19, no. 20, p. 4363, Oct. 2019.

[28] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, D. Altman, G. Antes, D. Atkins, V. Barbour, N. Barrowman, J. A Berlin, and J. Clark, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (Chinese edition)," *J. Chin. Integrative Med.*, vol. 7, no. 9, pp. 889–896, Sep. 2009.

[29] M. Shahbandeh. *Consumption of Vegetable Oils Worldwide From 2013/14 to 2019/2020, by Oil Type, Statista*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.statista.com/statistics/263937/vegetable-oils-global-consumption/

[30] M. Shahbandeh. *Production of Major Vegetable Oils Worldwide From 2012/13 to 2019/2020*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.statista.com/statistics/263933/production-of-vegetable-oils-worldwide-since-2000/

[31] MPOB Official Palm Oil Information Source. *Welcome to Malaysian Palm Oil Board // Malaysian Palm Oil Ind. // Washington, DC // 1-202-572-9768*. Accessed: Nov. 20, 2020. [Online]. Available: http://www.palmoilworld.org/about_malaysian-industry.html

[32] A. A. Karia and I. Bujang, "Progress accuracy of CPO price prediction: Evidence from ARMA family and artificial neural network approach," *Int. Res. J. Finance Econ.*, vol. 64, no. 64, pp. 66–79, 2011.

[33] N. F. Rahim, M. Othman, and R. Sokkalingam, "A comparative review on various method of forecasting crude palm oil prices," in *Proc. J. Phys., Conf.*, vol. 1123. Bristol, U.K.: IOP, 2018, Art. no. 012043.

[34] S. B. Hansen, R. Padfield, K. Syayuti, S. Evers, Z. Zakariah, and S. Mastura, "Trends in global palm oil sustainability research," *J. Cleaner Prod.*, vol. 100, pp. 140–149, Aug. 2015.

[35] *Malaysia Agriculture, Information about Agriculture in Malaysia*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.nationsencyclopedia.com/economies/Asia-and-the-Pacific/Malaysia-AGRICULTURE.html#ixzz6dfwX8w7Z

[36] *Production Volume of Palm Oil Worldwide From 2012/13 to 2019/20*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.statista.com/statistics/613471/palm-oil-production-volume-worldwide/

[37] *Palm Oil Consumption Worldwide From 2015/2016 to 2019/2020*. Accessed: Nov. 20, 2020. [Online]. Available: https://www.statista.com/statistics/274127/world-palm-oil-usage-distribution/

[38] X. Wang, J. Huang, Q. Feng, and D. Yin, "Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches," *Remote Sens.*, vol. 12, no. 11, p. 1744, May 2020.

[39] Y.-S. Shiu and Y.-C. Chuang, "Yield estimation of paddy rice based on satellite imagery: Comparison of global and local regression models," *Remote Sens.*, vol. 11, no. 2, p. 111, Jan. 2019.

[40] D. Lin, R. Wei, and L. Xu, "An integrated yield prediction model for greenhouse tomato," *Agronomy*, vol. 9, no. 12, p. 873, Dec. 2019.

[41] P. S. M. Gopal, "Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms," *Appl. Artif. Intell.*, vol. 33, no. 7, pp. 621–642, Jun. 2019.

[42] N. Kim, K.-J. Ha, N.-W. Park, J. Cho, S. Hong, and Y.-W. Lee, "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.*, vol. 8, no. 5, p. 240, May 2019.

[43] Y. Guo, Y. Fu, F. Hao, X. Zhang, W. Wu, X. Jin, C. R. Bryant, and J. Senthilnath, ''Integrated phenology and climate in rice yields prediction using machine learning methods,'' *Ecol. Indicators*, vol. 120, Jan. 2021, Art. no. 106935.

[44] M. B. and M. Louella, ''Bitter melon crop yield prediction using machine learning algorithm,'' *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 1–6, 2018.

[45] H. Yalcin, ''An approximation for a relative crop yield estimate from field images using deep learning,'' in *Proc. 8th Int. Conf. Agro-Geoinf. (Agro-Geoinf.)*, Jul. 2019, pp. 1–6.

[46] L. Zhang, Z. Zhang, Y. Luo, J. Cao, and F. Tao, ''Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches,'' *Remote Sens.*, vol. 12, no. 1, p. 21, Dec. 2019.

[47] A. G. Sánchez, J. F. Solís, and W. O. Bustamante, ''Predictive ability of machine learning methods for massive crop yield prediction,'' *Spanish J. Agricult. Res.*, vol. 12, no. 2, pp. 313–328, 2014.

[48] S. S. Panda, D. P. Ames, and S. Panigrahi, ''Application of vegetation indices for agricultural crop yield prediction using neural network techniques,'' *Remote Sens.*, vol. 2, no. 3, pp. 673–696, Mar. 2010.

[49] A. Crane-Droesch, ''Machine learning methods for crop yield prediction and climate change impact assessment in agriculture,'' *Environ. Res. Lett.*, vol. 13, no. 11, Oct. 2018, Art. no. 114003.

[50] J. Cao, Z. Zhang, F. Tao, L. Zhang, Y. Luo, J. Han, and Z. Li, ''Identifying the contributions of multi-source data for winter wheat yield prediction in China,'' *Remote Sens.*, vol. 12, no. 5, p. 750, Feb. 2020.

[51] S. Khaki and L. Wang, ''Crop yield prediction using deep neural networks,'' *Frontiers Plant Sci.*, vol. 10, p. 621, May 2019.

[52] J. Han, Z. Zhang, J. Cao, Y. Luo, L. Zhang, Z. Li, and J. Zhang, ''Prediction of winter wheat yield based on multi-source data and machine learning in China,'' *Remote Sens.*, vol. 12, no. 2, p. 236, Jan. 2020.

[53] D. Elavarasan, D. R. Vincent P M, K. Srinivasan, and C.-Y. Chang, ''A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling,'' *Agriculture*, vol. 10, no. 9, p. 400, Sep. 2020.

[54] F. Abbas, H. Afzaal, A. A. Farooque, and S. Tang, ''Crop yield prediction through proximal sensing and machine learning algorithms,'' *Agronomy*, vol. 10, no. 7, p. 1046, Jul. 2020.

[55] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, P. V. V. Prasad, and I. A. Ciampitti, ''Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil,'' *Agricult. Forest Meteorol.*, vol. 284, Apr. 2020, Art. no. 107886.

[56] B. Peng, K. Guan, W. Zhou, C. Jiang, C. Frankenberg, Y. Sun, L. He, and P. Köhler, ''Assessing the benefit of satellite-based solar-induced chlorophyll fluorescence in crop yield prediction,'' *Int. J. Appl. Earth Observ. Geoinf.*, vol. 90, Aug. 2020, Art. no. 102126.

[57] Y. Cai, K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, and B. Peng, ''Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches,'' *Agricult. Forest Meteorol.*, vol. 274, pp. 144–159, Aug. 2019.

[58] S. Ruslan, F. Muharam, D. Omar, Z. Zulkafli, and M. Zambri, ''Development of geospatial model for predicting Metisa plana's prevalence in malaysian oil palm plantation,'' in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 230. Bristol, U.K.: IOP, 2019, Art. no. 012110.

[59] Y. Y. Hilal, W. Ishak, A. Yahya, and Z. H. Asha'ari, ''Development of genetic algorithm for optimization of yield models in oil palm production,'' *Chilean J. Agricult. Res.*, vol. 78, no. 2, pp. 228–237, Jun. 2018.

[60] S. H. Bhojani and N. Bhatt, ''Wheat crop yield prediction using new activation functions in neural network,'' *Neural Comput. Appl.*, vol. 32, pp. 1–11, Mar. 2020.

[61] E. Khosla, R. Dharavath, and R. Priya, ''Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression,'' *Environ., Develop. Sustainability*, vol. 22, pp. 1–22, Aug. 2019.

[62] A. Mateo-Sanchis, M. Piles, J. Muñoz-Marí, J. E. Adsuara, A. Pérez-Suay, and G. Camps-Valls, ''Synergistic integration of optical and microwave satellite data for crop yield estimation,'' *Remote Sens. Environ.*, vol. 234, Dec. 2019, Art. no. 111460.

[63] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, ''Soybean yield prediction from UAV using multimodal data fusion and deep learning,'' *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111599.

[64] A. Kross, E. Znoj, D. Callegari, G. Kaur, M. Sunohara, D. R. Lapen, and H. McNairn, ''Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields,'' *Remote Sens.*, vol. 12, no. 14, p. 2230, Jul. 2020.

[65] P. Lykhovyd, ''Sweet corn yield simulation using normalized difference vegetation index and leaf area index,'' *J. Ecol. Eng.*, vol. 21, no. 3, pp. 228–236, Apr. 2020.

[66] P. Nevavuori, N. Narra, and T. Lipping, ''Crop yield prediction with deep convolutional neural networks,'' *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104859.

[67] A. W. Sugiyarto and A. M. Abadi, ''Prediction of indonesian palm oil production using long short-term memory recurrent neural network (LSTM-RNN),'' in *Proc. 1st Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2019, pp. 53–57.

[68] N. M. Nawi, A. R. M. Shariff, and S. A. Mehdizadeh, ''Application of artificial neural network classification to determine nutrient content in oil palm leaves,'' *Appl. Eng. Agricult.*, vol. 34, no. 3, pp. 497–504, 2018.

[69] S. Sousa, F. Martins, M. Alvimferraz, and M. Pereira, ''Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations,'' *Environ. Model. Softw.*, vol. 22, no. 1, pp. 97–103, Jan. 2007.

[70] W. Zhang and A. T. C. Goh, ''Multivariate adaptive regression splines and neural network models for prediction of pile drivability,'' *Geosci. Frontiers*, vol. 7, no. 1, pp. 45–52, Jan. 2016.

[71] R. R. Curtin, J. R. Cline, N. P. Slagle, W. B. March, P. Ram, N. A. Mehta, and A. G. Gray, ''MLPACK: A scalable C++ machine learning library,'' *J. Mach. Learn. Res.*, vol. 14, pp. 801–805, Mar. 2013.

[72] G. Tsoumakas, I. Katakis, and I. Vlahavas, ''Data mining and knowledge discovery handbook,'' *Mining Multi-Label Data*. Boston, MA, USA: Springer, 2010, pp. 667–685.

[73] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[74] K. Gr bczewski, *Meta-Learning in Decision Tree Induction*, vol. 1. Cham, Switzerland: Springer, 2014.

[75] L. Breiman, ''Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[76] P. Geurts, D. Ernst, and L. Wehenkel, ''Extremely randomized trees,'' *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.

[77] B. K. Zaied, M. Rashid, M. Nasrullah, B. S. Bari, A. W. Zularisam, L. Singh, D. Kumar, and S. Krishnan, ''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*, vol. 10, pp. 1–16, Oct. 2020.

[78] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, and L. Zhang, ''Deep learning in environmental remote sensing: Achievements and challenges,'' *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716.

[79] M. Rashid, M. Islam, N. Sulaiman, B. S. Bari, R. K. Saha, and M. J. Hasan, ''Electrocorticography based motor imagery movements classification using long short-term memory (LSTM) based on deep learning approach,'' *Social Netw. Appl. Sci.*, vol. 2, no. 2, pp. 1–7, Feb. 2020.

[80] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, ''Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data,'' *Remote Sens.*, vol. 7, no. 12, pp. 16398–16421, Dec. 2015.

[81] M. Yousefi D. B., A. S. M. Rafie, S. A. Aziz, S. Azrad, M. M. M. Masri, A. Shahi, and O. F. Marzuki, ''Classification of oil palm female inflorescences anthesis stages using machine learning approaches,'' *Inf. Process. Agricult.*, Dec. 2020.

[82] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, ''A review of classification algorithms for EEG-based brain–computer interfaces,'' *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007.

[83] M. Ali, R. C. Deo, N. J. Downs, and T. Maraseni, ''Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting,'' *Comput. Electron. Agricult.*, vol. 152, pp. 149–165, Sep. 2018.

[84] N. Deepa and K. Ganesan, ''Hybrid rough fuzzy soft classifier based multi-class classification model for agriculture crop selection,'' *Soft Comput.*, vol. 23, no. 21, pp. 10793–10809, Nov. 2019.

[85] A. F. Torres, W. R. Walker, and M. McKee, ''Forecasting daily potential evapotranspiration using machine learning and limited climatic data,'' *Agricult. Water Manage.*, vol. 98, no. 4, pp. 553–562, Feb. 2011.

[86] V. Rousson and N. F. Go oniu, "An R-square coefficient based on final prediction error," *Stat. Methodol.*, vol. 4, no. 3, pp. 331–340, Jul. 2007.
[87] K. Xu, J. Qian, Z. Hu, Z. Duan, C. Chen, J. Liu, J. Sun, S. Wei, and X. Xing, "A new machine learning approach in detecting the oil palm plantations using remote sensing data," *Remote Sens.*, vol. 13, no. 2, p. 236, Jan. 2021.
[88] H. Alaa, K. Waleed, M. Samir, M. Tarek, H. Sobeah, and A. A. Salam, "An intelligent approach for detecting palm trees diseases using image processing and machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 7, pp. 434–441, 2020.
[89] W. Wu, J. Zheng, H. Fu, W. Li, and L. Yu, "Cross-regional oil palm tree detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 56–57.
[90] D. Montero, W. Arenas, S. Salinas, and C. Rueda, "Development of a system based on aerial images for the morphological patterns classification using support vector machine," in *Proc. J. Phys. Conf.*, vol. 1702. Bristol, U.K.: IOP, 2020, Art. no. 012010.
[91] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, and S. Yuan, "Large-scale oil palm tree detection from high-resolution remote sensing images using faster-RCNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1422–1425.
[92] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," *IEEE Access*, vol. 9, pp. 4843–4873, 2021.
[93] D. R. Vincent, N. Deepa, D. Elavarasan, K. Srinivasan, S. H. Chauhdary, and C. Iwendi, "Sensors driven AI-based agriculture recommendation model for assessing land suitability," *Sensors*, vol. 19, no. 17, p. 3667, Aug. 2019.
[94] Y. Rao, Z.-H. Jiang, and N. Lazarovitch, "Investigating signal propagation and strength distribution characteristics of wireless sensor networks in date palm orchards," *Comput. Electron. Agricult.*, vol. 124, pp. 107–120, Jun. 2016.
[95] Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali, "Machine learning techniques in wireless sensor network based precision agriculture," *J. Electrochemical Soc.*, vol. 167, no. 3, Jan. 2020, Art. no. 037522.
[96] G.-Z. Hong and C.-L. Hsieh, "Application of integrated control strategy and Bluetooth for irrigating romaine lettuce in greenhouse," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 381–386, 2016.
[97] A. G. Mohapatra and S. K. Lenka, "Neural network pattern classification and weather dependent fuzzy logic model for irrigation control in WSN based precision agriculture," *Procedia Comput. Sci.*, vol. 78, pp. 499–506, Dec. 2015.
[98] J. Gutiérrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. A. Porta-Gándara, "Automated irrigation system using a wireless sensor network and GPRS module," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 166–176, Jan. 2014.
[99] D. Ilie-Ablachim, G. C. Pătru, I.-M. Florea, and D. Rosner, "Monitoring device for culture substrate growth parameters for precision agriculture: Acronym: MoniSen," in *Proc. 15th RoEduNet Conf., Netw. Educ. Res.*, Sep. 2016, pp. 1–7.
[100] H. Jawad, R. Nordin, S. Gharghan, A. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: A review," *Sensors*, vol. 17, no. 8, p. 1781, Aug. 2017.
[101] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour, and E.-H.-M. Aggoune, "Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk," *IEEE Access*, vol. 7, pp. 129551–129583, 2019.
[102] T.-G. Vågen, L. A. Winowiecki, J. E. Tondoh, L. T. Desta, and T. Gumbricht, "Mapping of soil properties and land degradation risk in africa using MODIS reflectance," *Geoderma*, vol. 263, pp. 216–225, Feb. 2016.
[103] L. Zhang, I. K. Dabipi, and W. L. B. Brown Jr., "Internet of Things applications for agriculture," *Internet Things A Z, Technol. Appl.*, vol. 18, pp. 507–528, May 2018.
[104] H. Liu, X. Wang, and J. Bing-Kun, "Study on NDVI optimization of corn variable fertilizer applicator," *INMATEH-Agricult. Eng.*, vol. 56, no. 3, pp. 193–202, 2018.
[105] S. Kim, M. Lee, and C. Shin, "IoT-based strawberry disease prediction system for smart farming," *Sensors*, vol. 18, no. 11, p. 4051, Nov. 2018.
[106] *FarmTRX | Precision Yield Monitors and Automated Yield Maps*. Accessed: Mar. 13, 2021. [Online]. Available: https://www.farmtrx.com/
[107] L. Manfrini, E. Pierpaoli, M. Zibordi, B. Morandi, E. Muzzi, P. Losciale, and A. C. Grappadelli, "Monitoring strategies for precise production of high quality fruit and yield in apple in Emilia-Romagna," *Chem. Eng. Trans.*, vol. 44, pp. 301–306, Sep. 2015.
[108] C. J. Ransom, N. R. Kitchen, J. J. Camberato, P. R. Carter, R. B. Ferguson, F. G. Fernández, D. W. Franzen, C. A. M. Laboski, D. B. Myers, E. D. Nafziger, J. E. Sawyer, and J. F. Shanahan, "Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations," *Comput. Electron. Agricult.*, vol. 164, Sep. 2019, Art. no. 104872.
[109] A. S. Terliksiz and D. T. Altỳlar, "Use of deep neural networks for crop yield prediction: A case study of soybean yield in Lauderdale county, Alabama, USA," in *Proc. 8th Int. Conf. Agro-Geoinf. (Agro-Geoinf.)*, Jul. 2019, pp. 1–4.
[110] D. Jiang, X. Yang, N. Clinton, and N. Wang, "An artificial neural network model for estimating crop yields using remotely sensed information," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1723–1732, May 2004.
[111] J. R. Romero, P. F. Roncallo, P. C. Akkiraju, I. Ponzoni, V. C. Echenique, and J. A. Carballido, "Using classification algorithms for predicting durum wheat yield in the province of buenos aires," *Comput. Electron. Agricult.*, vol. 96, pp. 173–179, Aug. 2013.
[112] A. K. Prasad, L. Chai, R. P. Singh, and M. Kafatos, "Crop yield estimation model for iowa using remote sensing and surface parameters," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 8, no. 1, pp. 26–33, Jan. 2006.
[113] M. Awad, "Toward precision in crop yield estimation using remote sensing and optimization techniques," *Agriculture*, vol. 9, no. 3, p. 54, Mar. 2019.
[114] S. Fukuda, W. Spreer, E. Yasunaga, K. Yuge, V. Sardsud, and J. Müller, "Random forests modelling for the estimation of mango (*Mangifera indica* l. Cv. Chok Anan) fruit yields under different irrigation regimes," *Agricult. Water Manage.*, vol. 116, pp. 142–150, Jan. 2013.
[115] J. H. Jeong, J. P. Resop, N. D. Mueller, D. H. Fleisher, K. Yun, E. E. Butler, D. J. Timlin, K.-M. Shim, J. S. Gerber, V. R. Reddy, and S.-H. Kim, "Random forests for global and regional crop yield predictions," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0156571.
[116] N. F. Rahim, M. Othman, R. Sokkalingam, and E. Abdul Kadir, "Forecasting crude palm oil prices using fuzzy rule-based time series method," *IEEE Access*, vol. 6, pp. 32216–32224, 2018.
[117] L. Chai, H. Xu, Z. Luo, and S. Li, "A multi-source heterogeneous data analytic method for future price fluctuation prediction," *Neurocomputing*, vol. 418, pp. 11–20, Dec. 2020.
[118] K. Kasturi, R Krishnan, U. R Hashim, M. B Masaina, S Indradevi, and M Jarrah, "A comparative study on univariate time series based crude palm oil price prediction model using machine learning algorithms," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5802–5806, Aug. 2020.
[119] A. K. Myat and M. T. Z. Tun, "Predicting palm oil price direction using random forest," in *Proc. 17th Int. Conf. ICT Knowl. Eng. (ICT KE)*, Nov. 2019, pp. 1–6.
[120] S. Ahmad, "Forecasting on the crude palm oil and kernel palm production: Seasonal ARIMA approach," in *Proc. IEEE Colloq. Humanities, Sci. Eng.*, Dec. 2011, pp. 939–944.
[121] K. Suppalakpanya, R. Nikhom, T. Booranawong, and A. Booranawong, "Forecasting oil palm and crude palm oil data in thailand using exponential time-series methods," *Eng. Appl. Sci. Res.*, vol. 46, no. 1, pp. 44–55, 2019.
[122] A. L. M. Hanapi, M. Othman, R. Sokkalingam, and A. Sakidin, "Developed a hybrid sliding window and garch model for forecasting of crude palm oil prices in Malaysia," in *Proc. J. Phys., Conf.*, vol. 1123. Bristol, U.K.: IOP, 2018, Art. no. 012029.
[123] K. C. Teng, J. Y. Koay, S. H. Tey, K. S. Lim, H. T. Ewe, and H. T. Chuah, "A dense medium microwave backscattering model for the remote sensing of oil palm," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3250–3259, Jun. 2015.
[124] C. M. Toh, S. H. Tey, H. T. Ewe, and G. Vetharatnam, "Classification of oil palm growth status with l band microwave satellite imagery," in *Proc. Photon. Electromagn. Res. Symp.-Fall (PIERS-Fall)*, Dec. 2019, pp. 1824–1831.
[125] M. S. M. Alfatni, A. R. M. Shariff, O. M. B. Saaed, A. M. Albhbah, and A. Mustapha, "Colour feature extraction techniques for real time system of oil palm fresh fruit bunch maturity grading," in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 540. Bristol, U.K.: IOP, 2020, Art. no. 012092.

[126] F. Ellsäßer, A. Röll, J. Ahongshangbam, P.-A. Waite, B. Schuldt, and D. Hölscher, "Predicting tree sap flux and stomatal conductance from drone-recorded surface temperatures in a mixed agroforestry system—A machine learning approach," *Remote Sens.*, vol. 12, no. 24, p. 4070, Dec. 2020.

[127] H. M. Asraf, N. Tahir, K. N. Dalila, and A. Hussain, "An agricultural telemonitoring method in detecting nutrient deficiencies of oil palm leaf," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 227–230, 2018.

[128] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, M. F. Ismail, N. P. Tan, and D. S. Karam, "Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105221.

[129] A. D. Amirruddin, F. M. Muharam, M. H. Ismail, N. P. Tan, and M. F. Ismail, "Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105768.

[130] Y. Cheng, L. Yu, A. P. Cracknell, and P. Gong, "Oil palm mapping using landsat and PALSAR: A case study in malaysia," *Int. J. Remote Sens.*, vol. 37, no. 22, pp. 5431–5442, Nov. 2016.

[131] W. Li, H. Fu, and L. Yu, "Deep convolutional neural network based large-scale oil palm tree detection for high-resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 846–849.

[132] K. Djerriri, M. Ghabi, M. S. Karoui, and R. Adjoudj, "Palm trees counting in remote sensing imagery using regression convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 2627–2630.

[133] Y. Wang, X. Zhu, and B. Wu, "Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 7356–7370, Oct. 2019.

[134] J. Zheng, H. Fu, W. Li, W. Wu, Y. Zhao, R. Dong, and L. Yu, "Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 154–177, Sep. 2020.

[135] S. Puttinaovarat and P. Horkaew, "Deep and machine learnings of remotely sensed imagery and its multi-band visual features for detecting oil palm plantation," *Earth Sci. Informat.*, vol. 12, no. 4, pp. 429–446, Dec. 2019.

[136] N. S. N. Shaharum, H. Z. M. Shafri, W. A. W. A. K. Ghani, S. Samsatli, H. M. Prince, B. Yusuf, and A. M. Hamud, "Mapping the spatial distribution and changes of oil palm land cover using an open source cloud-based mapping platform," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 7459–7476, Oct. 2019.

[137] N. S. N. Shaharum, H. Z. M. Shafri, W. A. W. A. K. Ghani, S. Samsatli, M. M. A. Al-Habshi, and B. Yusuf, "Oil palm mapping over peninsular malaysia using Google earth engine and machine learning algorithms," *Remote Sens. Appl., Soc. Environ.*, vol. 17, Jan. 2020, Art. no. 100287.

[138] R. Dong, W. Li, H. Fu, L. Gan, L. Yu, J. Zheng, and M. Xia, "Oil palm plantation mapping from high-resolution remote sensing images using deep learning," *Int. J. Remote Sens.*, vol. 41, no. 5, pp. 2022–2046, Mar. 2020.

[139] C. Fahy, F. Caraffini, and M. Gongora, "A multi-agent system for modelling the spread of lethal wilt in oil-palm plantations," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2020, pp. 1–7.

[140] N. A. Husin, S. Khairunniza-Bejo, A. F. Abdullah, M. S. M. Kassim, D. Ahmad, and M. H. A. Aziz, "Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning," *Agronomy*, vol. 10, no. 11, p. 1624, Oct. 2020.

[141] N. A. Husin, S. Khairunniza–Bejo, A. F. Abdullah, M. S. M. Kassim, and D. Ahmad, "Study of the oil palm crown characteristics associated with basal stem rot (BSR) disease using stratification method of point cloud data," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105810.

[142] S. A. Ruslan, F. M. Muharam, Z. Zulkafli, D. Omar, and M. P. Zambri, "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*, vol. 14, no. 10, Oct. 2019, Art. no. e0223968.

[143] B. Siregar, I. Butar-Butar, R. Rahmat, U. Andayani, and F. Fahmi, "Comparison of exponential smoothing methods in forecasting palm oil real production," in *Proc. J. Phys., Conf.*, vol. 801. Bristol, U.K.: IOP, 2017, Art. no. 012004.

[144] N. D. Kartika, I. W. Astika, and E. Santosa, "Oil palm yield forecasting based on weather variables using artificial neural network," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 3, no. 3, pp. 626–633, 2016.

[145] Ismail, "Neural network in modeling malaysian oil palm yield," *Amer. J. Appl. Sci.*, vol. 8, no. 8, pp. 796–803, Aug. 2011.

[146] S. Shanmuganathan, A. Narayanan, M. Mohamed, R. Ibrahim, and A. Khalid, "A hybrid approach to modelling the climate change effects on malaysia's oil palm yield at the regional scale," in *Recent Advances on Soft Computing and Data Mining*. Cham, Switzerland: Springer, 2014, pp. 335–345.

[147] C. M. Toh, H. T. Ewe, S. H. Tey, and Y. H. Tay, "A study on leaf area index and SAR image of oil palm with entropy decomposition and deep learning classification," in *Proc. Prog. Electromagn. Res. Symp.-Fall (PIERS-FALL)*, Nov. 2017, pp. 271–278.

[148] U. Adizue, S. Nwanya, and P. Ozor, "Artificial neural network application to a process time planning problem for palm oil production," *Eng. Appl. Sci. Res.*, vol. 47, no. 2, pp. 161–169, 2020.

[149] I. Bonet, F. Caraffini, A. Pena, A. Puerta, and M. Gongora, "Oil palm detection via deep transfer learning," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2020, pp. 1–8.

[150] Z. Y. Chen and I. Y. Liao, "Evaluation of feature extraction methods for classification of palm trees in UAV images," in *Proc. Int. Conf. Comput. Drone Appl. (IConDA)*, Dec. 2019, pp. 13–18.

[151] A. D. Amirruddin and F. M. Muharam, "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.*, vol. 34, no. 7, pp. 735–749, Jun. 2019.

[152] S. A. Saeed Alzaeemi and S. Sathasivam, "Examining the forecasting movement of palm oil price using RBFNN-2SATRA Metaheuristic algorithms for logic mining," *IEEE Access*, vol. 9, pp. 22542–22557, 2021.

[153] M. Culman, S. Delalieux, and K. Van Tricht, "Individual palm tree detection using deep learning on RGB imagery to support tree inventory," *Remote Sens.*, vol. 12, no. 21, p. 3476, Oct. 2020.

[154] A. Tuerxun, A. R. M. Shariff, R. Janius, Z. Abbas, and G. A. Mahdiraji, "Oil palm fresh fruit bunches maturity prediction by using optical spectrometer," in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 540. Bristol, U.K.: IOP, 2020, Art. no. 012085.

[155] N. Rashid, K. Hoong, A. Shamsuddin, N. Rosely, M. Noor, K. Jin, M. Lee, and A. Hamid, "Quality prediction and diagnosis of refined palm oil using partial correlation analysis," in *Proc. IOP Conf., Mater. Sci. Eng.*, vol. 884. Bristol, U.K.: IOP, 2020, Art. no. 012018.

[156] M. S. M. Alfatni, A. R. M. Shariff, S. K. Bejo, O. M. B. Saaed, and A. Mustapha, "Real-time oil palm FFB ripeness grading system based on ANN, KNN and SVM classifiers," in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 169. Bristol, U.K.: IOP, 2018, Art. no. 012067.

[157] A. Condro, Y. Setiawan, L. Prasetyo, R. Pramulya, and L. Siahaan, "Retrieving the national main commodity maps in indonesia based on high-resolution remotely sensed data using cloud computing platform," *Land*, vol. 9, no. 10, p. 377, Oct. 2020.

[158] R. Chapman, S. Cook, C. Donough, Y. L. Lim, P. Vun Vui Ho, K. W. Lo, and T. Oberthür, "Using Bayesian networks to predict future yield functions with data from commercial oil palm plantations: A proof of concept analysis," *Comput. Electron. Agricult.*, vol. 151, pp. 338–348, Aug. 2018.

[159] S. Poudel and R. Shaw, "The relationships between climate variability and crop yield in a mountainous environment: A case study in lamjung district, nepal," *Climate*, vol. 4, no. 1, p. 13, Mar. 2016.

[160] J. Maqsood, A. A. Farooque, X. Wang, F. Abbas, B. Acharya, and H. Afzaal, "Contribution of climate extremes to variation in potato tuber yield in prince edward island," *Sustainability*, vol. 12, no. 12, p. 4937, Jun. 2020.

[161] X. Liu, L. Guanter, L. Liu, A. Damm, Z. Malenovský, U. Rascher, D. Peng, S. Du, and J.-P. Gastellu-Etchegorry, "Downscaling of solar-induced chlorophyll fluorescence from canopy level to photosystem level using a random forest model," *Remote Sens. Environ.*, vol. 231, Sep. 2019, Art. no. 110772.

[162] Y. Zeng, G. Badgley, B. Dechant, Y. Ryu, M. Chen, and J. A. Berry, "A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence," *Remote Sens. Environ.*, vol. 232, Oct. 2019, Art. no. 111209.

[163] G. Duveiller, F. Filipponi, S. Walther, P. Köhler, C. Frankenberg, L. Guanter, and A. Cescatti, "A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity," *Earth Syst. Sci. Data*, vol. 12, no. 2, pp. 1101–1116, May 2020.

[164] B. Liu, X. Chen, Q. Meng, H. Yang, and J. van Wart, "Estimating maize yield potential and yield gap with agro-climatic zones in China—Distinguish irrigated and rainfed conditions," *Agricult. Forest Meteorol.*, vol. 239, pp. 108–117, May 2017.
[165] K. Yang, Y. Ryu, B. Dechant, J. A. Berry, Y. Hwang, C. Jiang, M. Kang, J. Kim, H. Kimm, A. Kornfeld, and X. Yang, "Sun-induced chlorophyll fluorescence is more strongly related to absorbed light than to photosynthesis at half-hourly resolution in a rice paddy," *Remote Sens. Environ.*, vol. 216, pp. 658–673, Oct. 2018.
[166] L. Liu, X. Yang, H. Zhou, S. Liu, L. Zhou, X. Li, J. Yang, X. Han, and J. Wu, "Evaluating the utility of solar-induced chlorophyll fluorescence for drought monitoring by comparison with NDVI derived from wheat canopy," *Sci. Total Environ.*, vol. 625, pp. 1208–1217, Jun. 2018.
[167] X. Hu, H. Ren, K. Tansey, Y. Zheng, D. Ghent, X. Liu, and L. Yan, "Agricultural drought monitoring using European space agency sentinel 3A Land Surface Temperature and normalized difference vegetation index imageries," *Agricult. Forest Meteorol.*, vol. 279, Dec. 2019, Art. no. 107707.
[168] D. Eleftheriou, K. Kiachidis, G. Kalmintzis, A. Kalea, C. Bantasis, P. Koumadoraki, M. E. Spathara, A. Tsolaki, M. I. Tzampazidou, and A. Gemitzi, "Determination of annual and seasonal daytime and nighttime trends of MODIS LST over Greece–climate change implications," *Sci. Total Environ.*, vols. 616–617, pp. 937–947, Mar. 2018.
[169] K. Katsura and Y. Nakaide, "Factors that determine grain weight in rice under high-yielding aerobic culture: The importance of husk size," *Field Crops Res.*, vol. 123, no. 3, pp. 266–272, Sep. 2011.
[170] P. Benincasa, L. Reale, E. Tedeschini, V. Ferri, M. Cerri, S. Ghitarrini, B. Falcinelli, G. Frenguelli, F. Ferranti, B. E. Ayano, O. Porfiri, and A. Rosati, "The relationship between grain and ovary size in wheat: An analysis of contrasting grain weight cultivars under different growing conditions," *Field Crops Res.*, vol. 210, pp. 175–182, Aug. 2017.
[171] M. N. Kiboi, K. F. Ngetich, A. Fliessbach, A. Muriuki, and D. N. Mugendi, "Soil fertility inputs and tillage influence on maize crop performance and soil water content in the Central Highlands of Kenya," *Agricult. Water Manage.*, vol. 217, pp. 316–331, May 2019.
[172] Y. Zhang, R. Wang, H. Wang, S. Wang, X. Wang, and J. Li, "Soil water use and crop yield increase under different long-term fertilization practices incorporated with two-year tillage rotations," *Agricult. Water Manage.*, vol. 221, pp. 362–370, Jul. 2019.
[173] Y. Chen, Z. Zhang, F. Tao, P. Wang, and X. Wei, "Spatio-temporal patterns of winter wheat yield potential and yield gap during the past three decades in north China," *Field Crops Res.*, vol. 206, pp. 11–20, May 2017.
[174] J. Zhao, X. Yang, and S. Sun, "Constraints on maize yield and yield stability in the main cropping regions in China," *Eur. J. Agronomy*, vol. 99, pp. 106–115, Sep. 2018.
[175] Y. Kang and M. Özdoğan, "Field-level crop yield mapping with landsat using a hierarchical data assimilation approach," *Remote Sens. Environ.*, vol. 228, pp. 144–163, Jul. 2019.
[176] X. Zhang and Q. Zhang, "Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 191–205, Apr. 2016.
[177] M.-J. Lambert, P. C. S. Traoré, X. Blaes, P. Baret, and P. Defourny, "Estimating smallholder crops production at village level from sentinel-2 time series in Mali's cotton belt," *Remote Sens. Environ.*, vol. 216, pp. 647–657, Oct. 2018.
[178] Z. Jin, G. Azzari, C. You, S. Di Tommaso, S. Aston, M. Burke, and D. B. Lobell, "Smallholder maize area and yield mapping at national scales with Google earth engine," *Remote Sens. Environ.*, vol. 228, pp. 115–128, Jul. 2019.
[179] J. J. Lawler, D. White, R. P. Neilson, and A. R. Blaustein, "Predicting climate-induced range shifts: Model differences and model reliability," *Global Change Biol.*, vol. 12, no. 8, pp. 1568–1584, Aug. 2006.
[180] D. R. Cutler, T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, Nov. 2007.
[181] R. Srinet, S. Nandy, and N. R. Patel, "Estimating leaf area index and light extinction coefficient using Random Forest regression algorithm in a tropical moist deciduous forest, India," *Ecological Informat.*, vol. 52, pp. 94–102, Jul. 2019.
[182] Y. Pan, J. Jiang, R. Wang, and H. Cao, "Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds," *Chemometric Intell. Lab. Syst.*, vol. 92, no. 2, pp. 169–178, Jul. 2008.
[183] P. Feng, B. Wang, D. L. Liu, C. Waters, and Q. Yu, "Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia," *Agricult. Forest Meteorol.*, vol. 275, pp. 100–113, Sep. 2019.
[184] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.
[185] S. Rasp, M. S. Pritchard, and P. Gentine, "Deep learning to represent subgrid processes in climate models," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 39, pp. 9684–9689, Sep. 2018.
[186] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org.
[187] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[188] N. Joshi, M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. Jepsen, T. Kuemmerle, P. Meyfroidt, E. Mitchard, J. Reiche, C. Ryan, and B. Waske, "A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring," *Remote Sens.*, vol. 8, no. 1, p. 70, Jan. 2016.
[189] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sens.*, vol. 9, no. 1, p. 22, Dec. 2016.
[190] W. Li, R. Dong, H. Fu, and A. L. Yu, "Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks," *Remote Sens.*, vol. 11, no. 1, p. 11, Dec. 2018.
[191] L. Li, J. Dong, S. Njeudeng Tenku, and X. Xiao, "Mapping oil palm plantations in cameroon using PALSAR 50-m orthorectified mosaic images," *Remote Sens.*, vol. 7, no. 2, pp. 1206–1224, Jan. 2015.
[192] L. Waranon, P. Karnngandee, and R. Ritronnasak, "Development of automatic G/T measurement program for THAICHOTE ground station," in *Proc. Int. Symp. Antennas Propag. (ISAP)*, 2016, pp. 970–971.
[193] S. Agustin, P. A. R. Devi, D. Sutaji, and N. Fahriani, "Oil palm age classification on satellite imagery using fractal-based combination," *J. Theor. Appl. Inf. Technol.*, vol. 89, no. 1, p. 18, 2016.
[194] S. U. Okoro, U. Schickhoff, J. Böhner, and U. A. Schneider, "A novel approach in monitoring land-cover change in the tropics: Oil palm cultivation in the niger delta, nigeria," *DIE ERDE–J. Geographical Soc. Berlin*, vol. 147, no. 1, pp. 40–52, 2016.
[195] T. P. M. Lazecky, S. Lhota, and D. Klushina, "Application of sentinel-1 satellite to identify oil palm plantations in balikpapan bay," in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 169. Bristol, U.K.: IOP, 2018, Art. no. 012064.
[196] A. Oon, H. Z. M. Shafri, A. M. Lechner, and B. Azhar, "Discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands using vegetation indices and supervised classification of LANDSAT-8," *Int. J. Remote Sens.*, vol. 40, no. 19, pp. 7312–7328, Oct. 2019.
[197] A. Poortinga, K. Tenneson, A. Shapiro, Q. Nquyen, K. S. Aung, F. Chishtie, and D. Saah, "Mapping plantations in myanmar by fusing landsat-8, sentinel-2 and sentinel-1 data along with systematic error quantification," *Remote Sens.*, vol. 11, no. 7, p. 831, Apr. 2019.
[198] I. Carolita, S. Darmawan, R. Permana, D. Dirgahayu, D. Wiratmoko, T. Kartika, and A. Arifin, "Comparison of optic landsat-8 and sar sentinel-1 in oil palm monitoring, case study: Asahan, north sumatera, indonesia," in *Proc. IOP Conf., Earth Environ. Sci.*, vol. 280. Bristol, U.K.: IOP, 2019, Art. no. 012015.
[199] A. C. Morel, J. B. Fisher, and Y. Malhi, "Evaluating the potential to monitor aboveground biomass in forest and oil palm in Sabah, Malaysia, for 2000–2008 with Landsat ETM+ and ALOS-PALSAR," *Int. J. Remote Sens.*, vol. 33, no. 11, pp. 3614–3639, Jun. 2012.
[200] D. J. Murphy, "Oil palm: Future prospects for yield and quality improvements," *Lipid Technol.*, vol. 21, nos. 11–12, pp. 257–260, Dec. 2009.
[201] K. V. Sankaran, P. D. Bridge, and C. Gokulapalan, "Ganoderma diseases of perennial crops in India? An overview," *Mycopathologia*, vol. 159, no. 1, pp. 143–152, Jan. 2005.
[202] H. D. Franqueville, "Oil palm bud rot in Latin America," *Experim. Agricult.*, vol. 39, no. 3, p. 225, 2003.
[203] G. Torres, G. Sarria, F. Varon, M. Coffey, M. Elliott, and G. Martinez, "First report of bud rot caused by phytophthora *palmivora* on African oil palm in Colombia," *Plant Disease*, vol. 94, no. 9, p. 1163, 2010.

[204] B. J. Wood and C. G. Fee, "A critical review of the development of rat control in malaysian agriculture since the 1960s," *Crop Protection*, vol. 22, no. 3, pp. 445–461, Apr. 2003.
[205] G. O. Bedford, "Advances in the control of rhinoceros beetle, *Oryctes rhinoceros* in oil palm," *J. Oil Palm Res.*, vol. 26, no. 3, pp. 183–194, 2014.
[206] C. L. Puan, A. W. Goldizen, M. Zakaria, M. N. Hafidzi, and G. S. Baxter, "Relationships among rat numbers, abundance of oil palm fruit and damage levels to fruit in an oil palm plantation," *Integrative Zool.*, vol. 6, no. 2, pp. 130–139, Jun. 2011.
[207] C. C. Lelong, J.-M. Roger, S. Brégand, F. Dubertret, M. Lanore, N. Sitorus, D. Raharjo, and J.-P. Caliman, "Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data," *Sensors*, vol. 10, no. 1, pp. 734–747, Jan. 2010.
[208] Y. Lin and M. Herold, "Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data," *Agricult. Forest Meteorol.*, vol. 216, pp. 105–114, Jan. 2016.
[209] J. Trochta, M. Krů ek, T. Vrška, and K. Král, "3D forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0176871.
[210] H. Croft and J. Chen, "Leaf pigment content," in *Reference Module in Earth Systems and Environmental Sciences*. Oxford, U.K.: Elsevier, 2017, pp. 1–22.
[211] R. Paterson, L. Kumar, F. Shabani, and N. Lima, "World climate suitability projections to 2050 and 2100 for growing oil palm," *J. Agricult. Sci.*, vol. 155, no. 5, p. 689, 2017.
[212] R. R. M. Paterson, L. Kumar, S. Taylor, and N. Lima, "Future climate effects on suitability for growth of oil palms in Malaysia and Indonesia," *Sci. Rep.*, vol. 5, no. 1, pp. 1–11, Nov. 2015.
[213] A. Rival, "Breeding the oil palm (Elaeis guineensis Jacq.) for climate change," *Ocl*, vol. 24, no. 1, p. D107, 2017.
[214] K. S. Chong, K. Abdullah, M. Z. M. Jafri, and S. L. Hwee, "Combination of radar and optical remote sensing data for land cover/use mapping," in *Proc. IEEE Int. Conf. Space Sci. Commun. (IconSpace)*, Jul. 2013, pp. 224–227.
[215] P. J. Zarco-Tejada, M. L. Guillén-Climent, R. Hernández-Clemente, A. Catalina, M. R. González, and P. Martín, "Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV)," *Agricult. Forest Meteorol.*, vols. 171–172, pp. 281–294, Apr. 2013.
[216] K. Uto, H. Seki, G. Saito, and Y. Kosugi, "Characterization of rice paddies by a UAV-mounted miniature hyperspectral sensor system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 851–860, Apr. 2013.
[217] C. S. Kang, K. D. Kanniah, Y. H. Kerr, and A. P. Cracknell, "Analysis of *in-situ* soil moisture data and validation of SMOS soil moisture products at selected agricultural sites over a tropical region," *Int. J. Remote Sens.*, vol. 37, no. 16, pp. 3636–3654, Aug. 2016.
[218] J. Gambo, H. M. Shafri, N. S. N. Shaharum, F. A. ZainalAbidin, and M. T. A. Rahman, "Monitoring and predicting land use-land cover (LULC) changes within and around krau wildlife reserve (KWR) protected area in Malaysia using multi-temporal landsat data," *Geoplanning, J. Geomatics planning*, vol. 5, no. 1, pp. 17–34, 2018.
[219] A. Chemura, I. van Duren, and L. M. van Leeuwen, "Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, ghana," *ISPRS J. Photogramm. Remote Sens.*, vol. 100, pp. 118–127, Feb. 2015.
[220] Y. Yuan, L. Chen, H. Wu, and L. Li, "Advanced agricultural disease image recognition technologies: A review," *Inf. Process. Agricult.*, vol. 8, no. 2, pp. 1–12, Jan. 2021.
[221] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
[222] T. Jiang, Z. Cui, Z. Zhou, and Z. Cao, "Data augmentation with Gabor filter in deep convolutional neural networks for SAR target recognition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 689–692.
[223] T. Wang, J. Liang, and X. Liu, "Soil moisture retrieval algorithm based on TFA and CNN," *IEEE Access*, vol. 7, pp. 597–604, 2019.

MAMUNUR RASHID received the B.Sc. degree in electrical and electronic engineering from the Pabna University of Science and Technology, Bangladesh, in 2016, and the M.Sc. degree from the Department of Electrical and Electronics Engineering, University Malaysia Pahang, Malaysia. His research interests include brain–computer interface, machine learning, deep learning, biomedical signal processing, and wireless communications. He received the Graduate Research Scheme Scholarship from the University Malaysia Pahang for his M.Sc. degree.

BIFTA SAMA BARI received the B.Sc. degree in electrical and electronic engineering from the Pabna University of Science and Technology, Bangladesh, in 2016, and the M.Sc. degree from the Department of Electrical and Electronics Engineering Technology, University Malaysia Pahang, Malaysia, in 2020. Her research interests include brain–computer interface, biomedical signal processing, machine learning, microwave imaging, and wireless communications. She was a recipient of the Graduate Research Scheme Scholarship from the University Malaysia Pahang.

YUSRI YUSUP received the B.Eng. degree (Hons.) in chemical engineering and the M.Sc. degree in chemical process technology, in which he developed a graphical user interface application using C++ for a refined air pollution model (USEPA's ISCST3), from Universiti Sains Malaysia (USM), in 2002 and 2004, respectively, and the Ph.D. degree in chemical and process engineering from Universiti Kebangsaan Malaysia, in 2008. In his Ph.D. thesis, he reported the structure of turbulence in the tropical atmospheric surface layer. He is currently an Associate Professor with the Environmental Technology Program, School of Industrial Technology, USM.

MOHAMAD ANUAR KAMARUDDIN received the B.Sc. degree in civil engineering and the M.Sc. and Ph.D. degrees in environmental engineering from Universiti Sains Malaysia. His research interests include on waste management, landfill leachate treatment, and industrial wastewater treatment. He has participated in several international conferences organized by reputable bodies worldwide. He also actively engaged in community empowerment through multidisciplinary research that allow him to communicate better with the bottom billion. He has published peer-reviewed articles in various reputable publishers. He actively involved in numerous technical visits and environmental protection works related to his expertise. He is a member of several prestigious bodies worldwide. He also sits in various organizing committees and technical committee.

NUZHAT KHAN received the M.S. degree in information technology from the Balochistan University of Information Technology, Engineering and Management Science (BUITEMS), Quetta, Pakistan, in 2019, through the Higher Education Commission (HEC) Scholarship. She is currently pursuing the Ph.D. degree with the Department of Environmental Technology, Universiti Sains Malaysia. Her research interests include natural language processing, applied linguistics, statistical linguistics, network science, graph theory, artificial intelligence, machine learning, and deep learning. She is currently involved in research on application of artificial intelligence in oil palm industry.