IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 8, 2021, accepted March 25, 2021, date of publication March 29, 2021, date of current version April 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069449

Towards Paddy Rice Smart Farming: A Review on Big Data, Machine Learning, and Rice Production Tasks

RAYNER ALFRED[1], JOE HENRY OBIT[1], CHRISTIE PEI-YEE CHIN[1], HAVILUDDIN HAVILUDDIN[2], (Member, IEEE), AND YUTO LIM[3], (Member, IEEE)

[1]Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu 88400, Malaysia
[2]Department of Informatics, Universitas Mulawarman, Samarinda 75123, Indonesia
[3]School of Information Science, Security and Networks Area, Japan Advanced Institute of Science and Technology, Nomi 923-1292, Japan

Corresponding author: Rayner Alfred (ralfred@ums.edu.my)

This work was supported by the Universiti Malaysia Sabah Internal Grant under Grant DFK2004-2020.

ABSTRACT Big Data (BD), Machine Learning (ML) and Internet of Things (IoT) are expected to have a large impact on Smart Farming and involve the whole supply chain, particularly for rice production. The increasing amount and variety of data captured and obtained by these emerging technologies in IoT offer the rice smart farming strategy new abilities to predict changes and identify opportunities. The quality of data collected from sensors greatly influences the performance of the modelling processes using ML algorithms. These three elements (e.g., BD, ML and IoT) have been used tremendously to improve all areas of rice production processes in agriculture, which transform traditional rice farming practices into a new era of rice smart farming or rice precision agriculture. In this paper, we perform a survey of the latest research on intelligent data processing technology applied in agriculture, particularly in rice production. We describe the data captured and elaborate role of machine learning algorithms in paddy rice smart agriculture, by analyzing the applications of machine learning in various scenarios, smart irrigation for paddy rice, predicting paddy rice yield estimation, monitoring paddy rice growth, monitoring paddy rice disease, assessing quality of paddy rice and paddy rice sample classification. This paper also presents a framework that maps the activities defined in rice smart farming, data used in data modelling and machine learning algorithms used for each activity defined in the production and post-production phases of paddy rice. Based on the proposed mapping framework, our conclusion is that an efficient and effective integration of all these three technologies is very crucial that transform traditional rice cultivation practices into a new perspective of intelligence in rice precision agriculture. Finally, this paper also summarizes all the challenges and technological trends towards the exploitation of multiple sources in the era of big data in agriculture.

INDEX TERMS Rice production, big data analytics, Internet of Things, machine learning, smart farming, precision agriculture, agriculture supply chain.

I. INTRODUCTION

The current global population of 7.8 billion (2020) persons is expected to reach 9.7 billion by 2050 [1]. It is expected that the world would require 70% more food than what available at the moment with less natural resources like land and water due to urbanization, soil erosion, climatic changes, water shortages and excessive use by livestock. It is estimated that there is about 33% wastage of agriculture production due to poor logistics and storage [2]–[4]. As result, the key for coping strategies in the contexts of climate change and food security is to implement precision agricultural or smart farming. Precision agriculture is a technology-enabled approach to farming management that observes, measures, and analyzes the needs of individual fields and crops [5]. Smart farming is defined as the application of information and data technologies for optimizing complex farming systems. It focuses on how the collected agriculture related information can be used in a smart way, rather than the storage of data, access to data and the application of these agriculture data. Big data and

The associate editor coordinating the review of this manuscript and approving it for publication was Mario Luca Bernardi.

machine learning algorithms are two main important components in paddy rice smart farming. Big Data can be defined as a data that can be described using three key concepts: volume, velocity, and variety. Volume refers to size of the data, variety refers to the various types of data (e.g., text, numbers, images, videos and audios) and Velocity refers to the increasing speed at which big data is created (e.g., live stream data). Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Applying smart farming technologies will definitely assist farmers in various tasks to increase crop productions. In order to narrow down the scope of this paper, we focus on paddy rice smart farming as rice is an increasingly important staple food in Asia Pacific region and other parts of the world.

The changes over time in land use and soil salinity levels have significant impact on the production of rice yields [6]. In addition to that, unpredictable weather conditions and inefficient techniques to predict weather conditions are amongst the few factors that reduce rice yields production [7], [8]. For instance, most farmers in Myanmar face heavy rains during the rice growing season and that crop damage and yield losses due to heavy rains cause extensive losses among farmers [9]. As a result, the ability to predict weather or climate trends and environmental factors (e.g., soil nutrient) is very important in enhancing paddy farmers' productivity [7].

The current practices, which heavily rely on fertilizers and pesticides to increase productivity are not supporting the sustainable rice yields production because these activities are not environmentally friendly farming systems [10]. In addition to that, the timing for rice yields harvesting also influences the production of rice yields as the best timing for paddy harvesting showed a linear relationship with grain loss [11]. As a result, monitoring the growth of paddy is very crucial to sustain rice yields production.

Rice production in coastal areas is frequently affected by typhoons. The lack of ability to manage impacts from natural events and disasters that include contamination of water bodies, loss of harvest, and destruction of irrigation systems and other agricultural infrastructure is another shortcoming that requires attention [12], [13]. With smart farming, the application of data mining and analytical techniques designed so far for prediction, detection and development of appropriate disaster management strategy based on the collected data from disasters can be used to manage these impacts and consequently support agriculture or farming activities with more effectively and efficiently.

Variation within farms and region based on resource endowments, location topography and farmers circumstances make it difficult to apply the same strategy in maximizing rice yields productivity. Towards the end of the twentieth century, precision agriculture began to be utilized that applies information technologies to capture and integrate data from multiple sources (e.g., farmers, sensors) in order to have a more robust strategy associated with crop management and thus can be used to maximize agriculture productivity [14].

Unsustainable rice yields production [6], inefficient techniques to predict weather conditions, lack of ability to manage calamities [7], [8], [12], variation within farms and region [14] and poor logistics and storage [2]–[4] are among the reasons why smart farming should be adopted to sustain and optimize rice yields productivity.

In this paper, we conduct a systematic literature review (SLR) of the latest research on intelligent data processing technology involved in rice smart farming focusing on the rice production and post-production phases of the agriculture supply chain. We describe the main datasets or features extracted for data modelling. We then elaborate role of machine learning algorithms in smart agriculture, by analyzing the applications of machine learning in various scenarios in the rice production and post-production phases of the agriculture supply chain. This paper also presents a framework that maps the activities defined in rice smart farming, datasets or features used in data modelling and machine learning algorithms used to analyze these features for each activity defined in the early stage of agriculture supply chain.

The remainder of this article is organized as follows. Section II provides the literature review of the most recent reviews conducted related to smart farming. Section 3 describes the existing frameworks related to agriculture supply chain. Section 4 provides an in-depth analysis of the type of big data used in rice smart farming agriculture focusing on the variety of sources used, the variety of machine learning algorithms used and finally the variety tasks involved in the rice production and post-production phases of smart farming. The research work presented in this section is classified based on the sources and types of data that are used, the types of tasks involve in smart farming and also the type of machine learning algorithms used to model these data. Section 5 presents a framework that maps the activities defined in smart farming, datasets or features used in data modelling and machine learning algorithms used to analyze these features for each activity defined in the early stage of agriculture supply chain. Finally, Section 6 concludes this paper and presents challenges and technological trends towards the exploitation of multiple sources in the era of big data in agriculture.

II. LITERATURE REVIEW

A survey has been conducted to look into the global coverage in terms of innovation related to smart farming and the usage of machine learning in smart farming. The survey was conducted by using two methods; looking at the trends of number of scholarly works over time related to *Smart Farming*, *Machine Learning in Smart Agriculture*, *Artificial Intelligence in Smart Agriculture* and *Internet of Things in Smart Agriculture*, and reviewing all review studies that were conducted on several elements of I4.0 and its applications in smart farming for improving the productivity.

Firstly, the trends of number of scholarly works over time related to *Smart Farming*, *Machine Learning in Smart*

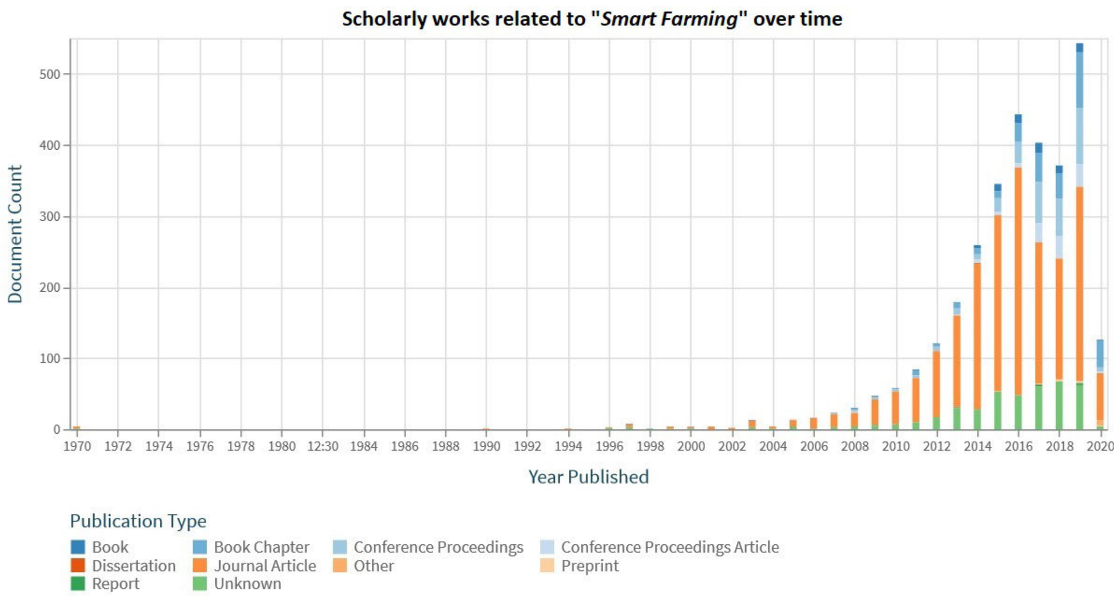

FIGURE 1. The trends of scholarly works over time related to *Smart Farming*.

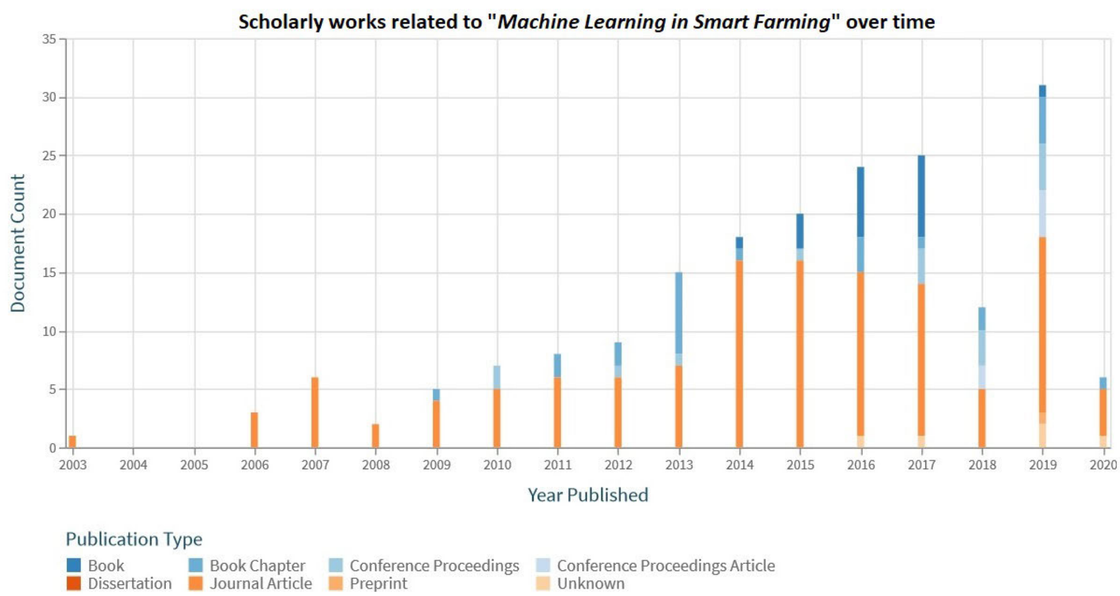

FIGURE 2. The trends of scholarly works over time related to *Machine Learning in Smart Farming*.

Agriculture, *Artificial Intelligence in Smart Agriculture* and *Internet of Things in Smart Agriculture*, can be obtained by using the Lens website (https://www.lens.org). Lens provides open datasets of patent documents, scholarly research works and any inventions related to machine learning, artificial intelligence, internet of things and smart farming disclosed in patents [15]. The Lens serves global patent and scholarly knowledge as a public resource to make science- and technology-enabled problem solving more effective, efficient and inclusive. This knowledge may help show ways forward such as new or repurposed ideas and inventions, better strategies and targeted partnerships for collective action. Based on these four keywords used in searching for trends in smart farming research, the usage of Machine Learning (ML) in smart farming, the usage of Artificial Intelligence (AI) in smart farming and the usage of Internet of Things (IoT) in smart farming, Fig. 1 through Fig. 4 display the increasing trends of number of scholarly works over time related to these keywords. For instance, based on these Fig. 1, Fig. 2, Fig. 3 and Fig. 4, several scholarly works have been filed

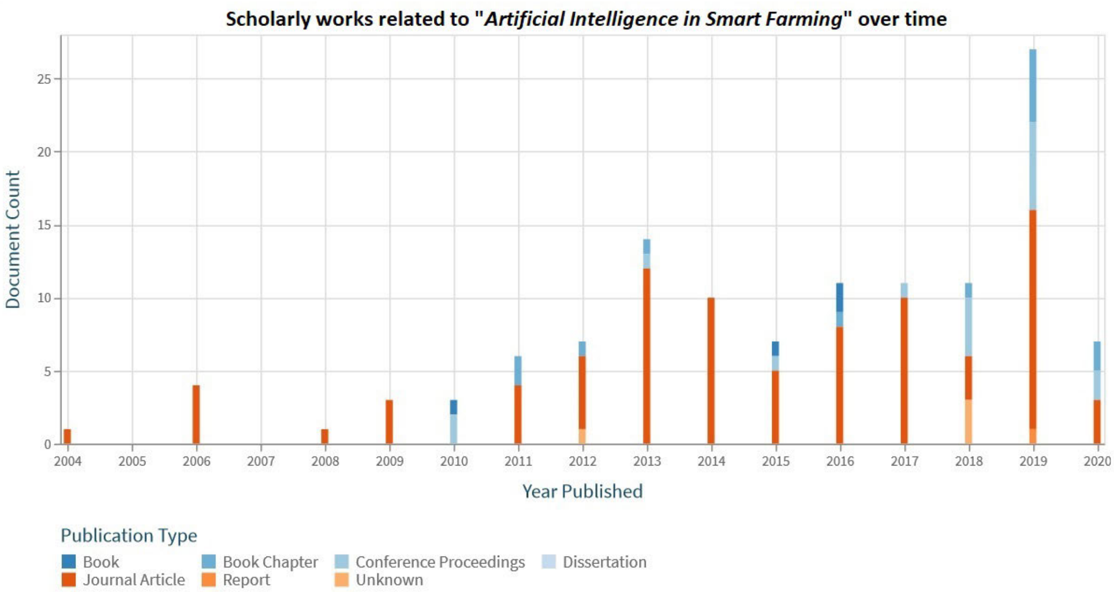

FIGURE 3. The trends of scholarly works over time related to ***Artificial Intelligence in Smart Farming***.

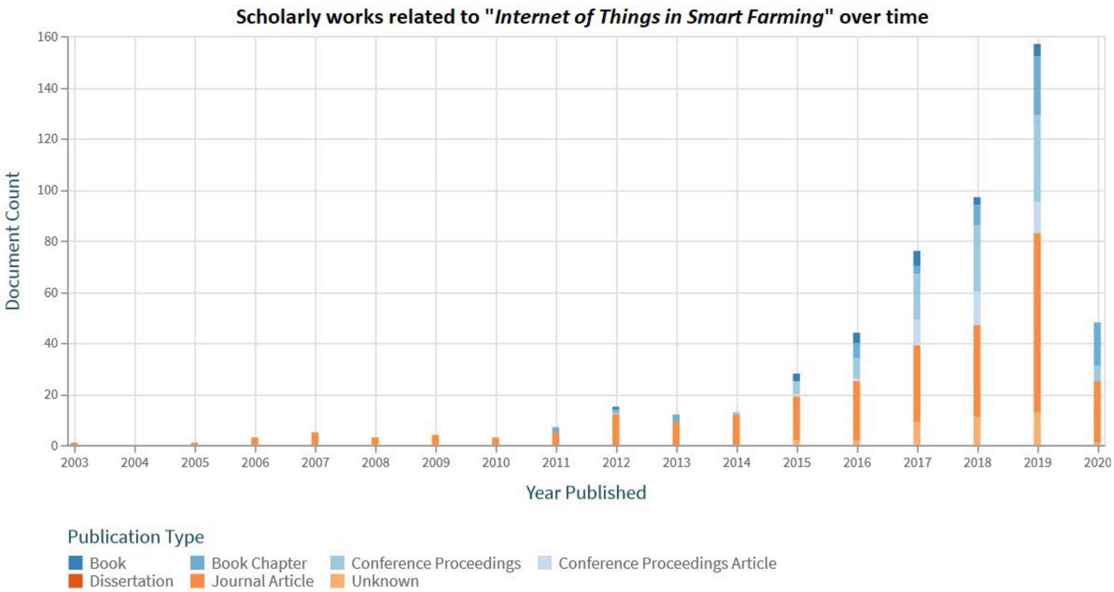

FIGURE 4. The trends of scholarly works over time related to ***Internet of Things in Smart Farming***.

and recorded and the number of scholarly works filed has increased between the year of 2018 and 2019.

Secondly, in the past, few review studies were conducted on several elements of I4.0 and its applications in smart farming for improving the productivity in agriculture sectors as mentioned in Table 1. These studies have focused on I4.0 applications in the smart farming covering specific aspects like Internet of Things, Cloud Computing and Big Data Analytics (Machine Learning). Table 1 shows several reviews that were conducted recently that are related to smart farming or precision agriculture. Several reviews conducted were focusing on the application of machine learning algorithms in smart farming [16], [17]. For instance, Sharma *et al.* investigated the current state of research on machine learning (ML) applications in Agriculture Supply Chain (ASC) that includes the application of ML in four different phases in ASC; pre-production, production, processing and distribution [16]. It was concluded that all three ML algorithms can be leveraged to develop a sustainable ASC. A Machine Learning-Agriculture Supply Chain performance framework

TABLE 1. Review papers: Elements of I4.0 and its applications in smart farming.

Study	Title	Period	Papers	Industry 4.0	Objective	Findings
[16]	A systematic literature review on machine learning applications for sustainable agriculture supply chain performance	2002–2019	93	Big Data Analytics (Machine Learning)	Review of the current state of research on machine learning (ML) applications in Agriculture Supply Chain	ML algorithms can be leveraged to develop sustainable ASCs.
[17]	Review–Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture	2016–2020	20	Machine Learning, Internet of Things	Review of the application of different machine learning algorithms in sensor data analytics within the agricultural ecosystem.	AI must be leveraged to increase the automation of tasks in agriculture and improve the yield while optimizing the use of natural resources.
[18]	Deep Learning Applications in Agriculture: A Short Review	2016–2019	29	Big Data Analytics – Deep Learning	A survey of different deep learning techniques applied to various agricultural problems	Deep learning provides high accuracy results.
[19]	Deep learning for smart agriculture: Concepts, tools, applications, and opportunities			Big Data Analytics – Deep Learning	Provides a concise summary of major deep learning algorithms applied in smart farming	
[20]	A comprehensive review on automation in agriculture using artificial intelligence	1998–2018	> 50	Big Data Analytics, Internet of Things	Survey the current implementation of automation in agriculture	Using deep learning is an added advantage over other machine learning methods.
[21]	Deep learning in agriculture – A survey	2014–2017	40	Big Data Analytics	Survey on deep learning techniques, applied to various agricultural challenges.	Deep learning provides higher accuracy.
[22]	Internet of Things in arable farming: Implementation, applications, challenges and potential	2008–2018	167	Internet of Things	Survey of the current and potential application of Internet of Things in arable farming	Interoperability is a main challenge throughout the whole IoT architecture
[23]	Smart poultry management: Smart sensors, big data, and the internet of things			Internet of Things, Big Data Analytics	Review new precision livestock farming sensor technologies	Incorporating precision livestock farming technologies, big data analytics, and the IoT increasing productivity.
[24]	Towards automated aquaponics: A review on monitoring, IoT, and smart systems	2004–2019	52	Internet of Things	Review of monitoring, smart and IoT systems in aquaponics.	
[25]	Survey, comparison and research challenges of IoT application protocols for smart farming			Internet of Things	Survey of IoT messaging (application) protocols applied in smart farming	Message Queue Telemetry Transport (MQTT) outperforms the other IoT protocols.
[26]	Big Data in Smart Farming – A review	2010–2015	110	Big Data	Review of Big Data applications in smart farming	Challenges: Data ownership and quality, Intelligent processing and analytics, Sustainable Integration of Big Data Sources, Business models and openness of platforms.
[27]	Towards smart farming: Systems, frameworks and exploitation of multiple sources			Big Data	Survey of the state-of-the-art agriculture systems and big data architectures	
[28]	Internet of Things (IoT) and Agricultural Unmanned Aerial Vehicles (UAVs) in Smart Farming: A Comprehensive Review	2013–2019	65	Internet of Things	Survey of the last research on IoT and UAV technology applied in agriculture.	The introduction of IoT technology in various farming practices has improved the overall metrics.

was introduced in which the machine learning algorithms are mapped into all four different phases in ASC based on the type of data used. However, these data are not explained and categorized comprehensively.

Mekonnen *et al.* conducted a review on the application of various machine learning methods in analyzing data captured from sensors within the agricultural ecosystem [17]. In this review, a limited number of machine learning algorithms is listed based on the data that are captured using different types of Wireless Sensor Networks (WSN) (e.g., ZigBee WSN, GSM and GPS WSN, LoRa WSN, Wifi and MQTT Sensor based with Raspberry pi and Arduino) and also remotely sensed data (multispectral or hyperspectral data) and vegetation indices. Based on the trend obtained from this review, there will be an increased use of more advanced techniques like distributed (or edge) deep learning.

Several reviews also conducted focusing on the application of deep learning algorithms only in smart farming [18]–[21]. One of the findings from these reviews is that the deep learning algorithms are proven to be better in providing high accuracy results compared with other machine learning algorithms in terms of accuracy when applied to various agricultural problems, such as disease detection and identification, fruit or plants classification and fruit counting among other domains.

The evolution of agriculture systems involves the adoption of incoming data from various sources [27] and also the application of big data applications in smart farming [26]. Together these big data technologies and the capability of machine learning algorithms in forecasting certain outcomes will cause major changes in the scope and organization of smart farming [16], [17], [26]. Lytos *et al.* conducted a survey paper that covers the state-of-the-art big data architectures and agriculture systems in order to bridge the knowledge gap between agriculture systems and exploitation of big data. However, in this review, the authors list out the name of the databases and features used in the agriculture systems only without outlining how these data are processed or analyzed.

The quality and type of dataset collected from sensors greatly influence the performance of the forecasting algorithm in predicting the crop yields. For instance, in optimizing the performance of forecasting crop yields, Fabrizio Balducci *et al.* have investigated the performance of several machine learning algorithms based on different subsets of features extracted from the environmental sensors [29].

There are quite a number of reviews conducted related to Internet of Things (IoT) technologies [38]. The emergence of Internet of Things (IoT) technologies has also improved the performance of a real-time monitoring of the data related to smart farming [20]. IoT are mainly used in monitoring crop, soil and weather, forecasting disease and crop yields, controlling irrigation machinery and autonomous vehicles and robots [22]. Based on several reviews, it can be concluded that incorporating precision livestock farming technologies (Sensors), big data analytics, and the IoT in smart farming practices puts forth a possible solution to assist us to improve agriculture productivity and meet projected global agricultural product demands [23]–[25], [28]. The increasing amount and variety of data captured and obtained by these emerging technologies in IoT offer the smart farming strategy new abilities to predict changes and identify opportunities.

However, to the best of our knowledge, no such studies are conducted to review comprehensively the present mapping of datasets or features and machine learning algorithms based on different types of tasks involved in the paddy rice production and post-production phases of the ASC. There are several related works conducted about the applications of machine learning on the paddy rice smart farming. Several researches have been conducted that apply machine learning algorithm (e.g., Support Vector Machine (SVM) [46], [98], [127], Convolutional Neural Network (CNN) [94], [95], and hybrid approaches [103], [124]) in paddy rice sample recognition and classification using high-resolution images. Remotely sensed, vegetation indices and climate data are commonly used to predict paddy rice yield estimation [34], [35], [48], [76], [77], [109] and to monitor paddy rice growth [63], [73], [84] using artificial neural networks and its variants and also linear regression approaches. In addition to that, hyperspectral and high-resolution images have been used to accurately and affectively monitor paddy rice disease [40], [41], [87], [88], [118] and assessing quality of paddy rice [93], [104], [105] by using deep learning algorithms.

In this paper, we will present a framework that maps three elements which include a) Paddy rice production and post-production activities defined in the ASC, b) datasets or features related to agriculture components captured from sensors, and c) machine learning algorithms used to analyze these features for each activities defined in the early stage of ASC. This is done by

i Identifying the phases and tasks involved in the paddy rice smart farming that require intelligent data processing technologies.
ii Describing the main datasets or features captured and used by intelligent data processing technologies in each task identified in the paddy rice smart farming.
iii Elaborating the roles of machine learning technology in paddy rice smart agriculture, by analyzing the applications of machine learning in various tasks and phases in the paddy rice smart farming.

III. PHASES AND TASKS IN PADDY RICE SMART FARMING

This paper focuses on the smart farming technologies used in paddy growth and production. This section elaborates the selected phases and tasks involved in the paddy rice smart farming [16]. The applications of machine learning algorithms and smart technologies in the agriculture supply chain can be divided into 4 phases and include pre-production, production, post-production and finally distribution phases [16], [30], [31]. However, in this work, we focus on several tasks that require intelligent data processing technologies that can be fully utilized to improve the production of paddy rice. Thus, this review focuses on the rice production and post-production phases.

In the rice production phase, several activities are conducted sequentially such as planting, managing water, monitoring soil fertility, managing weed and finally managing pests and diseases. Then, in the rice post-production phase, the activities can be divided into harvesting, drying, storage and milling and processing. Based on these phases and tasks, we will look into the features or datasets that are applied by machine learning algorithms in this rice production processes. The SLR framework used for presenting the review findings is presented in Fig. 5. Fig. 5 highlights two main categories which are the paddy growing activities and the smart farming activities associated with the paddy growing activities. The paddy growing activities can be divided into production and post-production phases. The first step in rice production phase is planting. Rice crops can either be direct seeded or transplanted. Next, ensuring the rice plant to get adequate water is very important since rice is extremely

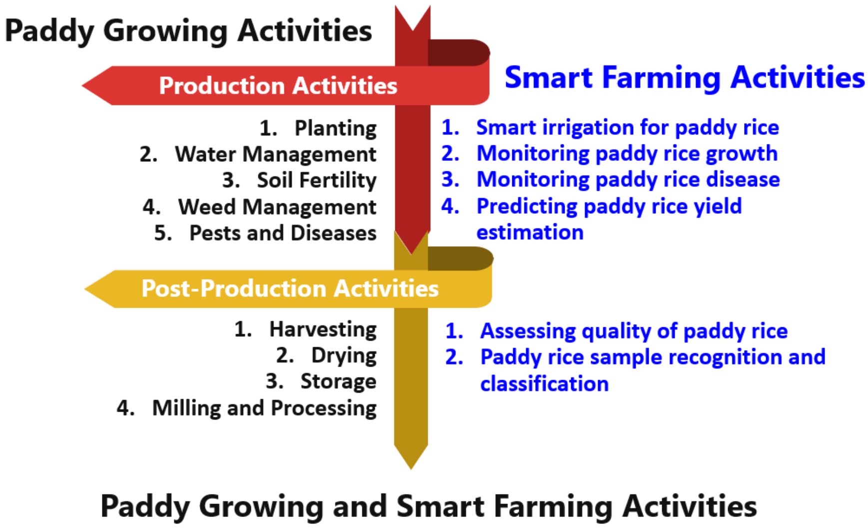

FIGURE 5. Rice production and post-production phases in paddy rice smart farming.

sensitive to water shortages. Managing good practices for smart paddy irrigation is very critical to maximize water efficiency and yield. Smart irrigation for paddy rice deals with maintaining a predetermined water height in paddy fields automatically based on the growth stages of the paddy rice [32], [33]. Managing weeds is crucial to reduce the amount of weed pressure in the field. Next, monitoring soil fertility is also very essential to optimize the growth of a rice plant. At the same time, timely and accurate diagnosis of paddy diseases and managing pests are highly required to reduce loses. Monitoring paddy rice disease involves activities such as detection and recognition of diseases from paddy plant leaf images [37], [39] or classifying, detecting, and predicting infestation patterns of the Brown Planthopper in rice paddies [40], [41]. Generally, monitoring the growth of paddy rice involves analyzing the growth of paddy rice based on climate data or remotely sensed data and vegetation indices. This also includes developing an approach for mapping rice-growing areas at field level using phenology-based rice crop classification or paddy growth stages classification [34]–[36]. Predicting paddy rice yield estimation involves tasks such as yield assessment of paddy fields using machine learning algorithms [42] or mapping rice planted area using the hyperspectral data or remotely sensed data and vegetation indices [43].

In the rice post-production phase, paddy harvesting activities include reaping, stacking, handling, threshing, cleaning, and hauling. Harvesting paddy should be performed efficiently as the speed of paddy harvesting showed a linear relationship with grain loss [11]. When rice is harvested, it will contain up to 25% moisture. High moisture level during storage can lead to grain discoloration, encourage development of molds, and increase the likelihood of attack from pests. It can also decrease the germination rate of the rice seed. Assessing the quality of paddy rice can be performed by using any machine learning algorithms. Assessing the quality of paddy rice usually involves activities such as assessing the quality of the rice [44] or investigating the impact of climate change on paddy rice production [45]. Next, drying process will involve the process of drying paddy by using traditional or mechanical systems. It is important to dry rice grain as soon as possible after harvesting (ideally within 24 hours). After that, these dried rice grain will be stored to prevent grain loss caused by adverse weather, moisture, rodents, birds, insects and micro-organisms like fungi. Finally, the last activity in the post-production phase is the milling process which remove the husk and the bran layers, and produce an edible, white rice. Paddy rice sample recognition and classification can be applied to perform the milling process. In paddy rice sample recognition and classification, the main task is to separate and classify objects of rice sample based on color and texture features with the help of image processing and machine learning techniques [46], [47].

IV. APPLICATION OF BIG DATA AND MACHINE LEARNING IN RICE PRODUCTION TASKS

A. BIG DATA USED IN RICE PRODUCTION TASKS

Data that are commonly used in paddy rice smart farming can be categorized into sensor data, remotely sensed data and vegetation indices, drone based data and finally paddy rice leaf analysis data. Table 2 tabulates the types of data and features used in paddy rice smart farming according to the smart farming activities described in Fig. 5.

1) SENSOR DATA

Firstly, the typical types of sensor data captured that can be used in monitoring paddy rice growth or yield estimation of paddy rice are data related to meteorological.

TABLE 2. Types of data and features used in paddy rice smart farming.

Tasks	Types of Features and Studies
Predicting paddy rice yield estimation	**Sensor Data**: Wind speed and direction [45], [48], [112], Temperature (air, water, soil) [45], [48], [110]–[112], Relative humidity [45], [48], [112], Precipitation [45], [48], [110], [112], Rainfall [45], [111], [112], [155], Soil moisture [48]. **Remotely Sensed Data and Vegetation Indices**: LSWI [35], [62], EVI [35], [43], [58], [59], [62], [67], NDVI [43], [57], [59], [67], [68], [109], MNDWI [57], Hyperspectral Images (Band 1 ∼ 4) [75]–[77], C-Band Synthetic Aperture Radar (SAR) [82] **Drones Based Data**: High-resolution Images [42], [84]–[86]
Monitoring paddy rice disease	**Sensor Data**: Wind speed and direction [41], Temperature (air, water, soil) [41], [36], Relative humidity [41], Rainfall [41]. **Remotely Sensed Data and Vegetation Indices**: Hyperspectral Images (Band 1 ∼ 4) [80], [81], C-Band Synthetic Aperture Radar (SAR) [40] **Drones Based Data**: High-resolution Images [37], [39], [87]–[92], [117], [118]
Monitoring paddy rice growth	**Remotely Sensed Data and Vegetation Indices**: LSWI [63],EVI [62], [63], NDVI [63], [66], [69], LAI [69], [71]–[73], C-Band Synthetic Aperture Radar (SAR) [83]
Assessing quality of paddy rice	**Sensor Data**: Soil moisture [105], Soil pH [105], [164], Soil Nutrients [105], Sonar [102]–[104], Nitrogen [105] **Remotely Sensed Data and Vegetation Indices**: LSWI [44], EVI [44], NDVI [44], LAI [105], Hyperspectral Images (Band 1 ∼ 4) [69], [71], [72], [78], [79] **Drones Based Data**: High-resolution Images [102]–[104]
Paddy rice sample recognition and classification	**Drones Based Data**: High-resolution Images [46], [47], [93]–[101]

Meteorological data (or climate data) can be used to monitor paddy rice growth [45], [48] and disease [41]. For instance, Guruprasad *et al.* conducted a yield estimation modeling paddy crop at different spatial resolution (SR) levels based on weather and soil data as input features. These features include day and night temperature (min, max, mean), diffused irradiance, precipitation (cumulative), relative humidity, wind speed, rainfall, pH, soil moisture and temperature (0-40cm) [48].

It was observed that the disease incidence on paddy rice growth is also directly affected by the level of temperature, wetness duration [50], [51]. Paddy rice production is also affected by the level of precipitation. For instance, the paddy rice production was found to be affected by the decreasing post-monsoon precipitation as this time coincides with the sensitivity of the paddy fruiting and ripening stages [54]. Besides that, winds may also affect the growth and production of paddy rice plants as strong winds are very detrimental to the growth and production of rice plants, especially when they occur during the flowering and ripening phases of rice [49]. Rainfall was found to be the main climate driver of the paddy rice yield [111]. The suitable soil pH for rice cultivation is at pH 6.0 [52] or 6.25 [53]. Analyzing the nitrogen level of paddy rice can also be used to assess the quality of paddy rice [105].

2) REMOTELY SENSED DATA AND VEGETATION INDICES

Secondly, remote sensing data or remotely sensed data and vegetation indices can be used in different ways in estimating paddy yield, monitoring paddy growth and diseases. Many studies are based on the mapping of rice-growing are [35], [43] [57], mapping cropping patterns [57]–[59], mapping paddy vulnerability to flooding [58].

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensors have a total of 36 spectral bands and seven of them are related to vegetation and land surfaces that include several ranges [60]. Seven of the most used spectral bands includes

i) Red (620–670 nm) - Band 1
ii) Near Infrared One (NIR1) (841–875 nm) - Band 2
iii) Blue (459–479 nm) - Band 3
iv) Green (545–565 nm) - Band 4
v) Near Infrared Two (NIR2) (1230–1250 nm) - Band 5
vi) Shortwave Infrared One (SWIR1) (1628–1652 nm) - Band 6
vii) Shortwave Infrared Two (SWIR2) (2105–2155 nm) - Band 7

Based on these spectral bands, several measurements can be derived and computed such as Land Surface Water Index (LSWI), Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI), Leaf Area Index (LAI) (see Table 2).

LSWI is sensitive to the total amount of liquid water in vegetation and its soil background. LSWI was developed by considering two bands of the shortwave infrared (SWIR) and the NIR regions of the electromagnetic spectrum to compute the estimation of water content of the land surface [61]. LSWI is computed based on Eq. 1;

$$LSWI = \frac{\rho NIR - \rho SWIR1}{\rho NIR + \rho SWIR1} \tag{1}$$

where ρNIR is the reflectance in the NIR, ρSWIR1 is the reflectance in the Shortwave Infrared One. LSWI can be used to detect and classify paddy rice phenology in paddy fields with complex cropping patterns [35], [62]. It was also used to assess the damage of regional rainfed paddy rice after severe floods [44] and monitoring rice growth [63]. Liou and Sha found that the value of LSWI increases and becomes higher that NDVI and EVI [44].

EVI can be used to quantify vegetation greenness [64]. Son *et al.* have constructed a time-series EVI and LSWI data in order to perform the phenology-based rice crop classification [35]. EVI can be measured as follows;

$$EVI = G \times \frac{\rho NIR - \rho Red}{\rho NIR + (C_1 \times \rho red - C_2 \times \rho Blue) + L} \quad (2)$$

where ρNIR is the reflectance in the NIR, ρRed is the reflectance in the red, ρBlue is the reflectance in the blue, C_1, C_2, and L are coefficients and G is the gain factor. The coefficients adopted in the MODIS-EVI algorithm are; $L = 1$, $C_1 = 6$, $C_2 = 7.5$, and $G = 2.5$. EVI is normally combined with other vegetation indices (e.g., NDVI, LSWI) to predict paddy rice yields' estimation [35], [58] [59], [62], assess damage of regional rainfed paddy rice [44] and monitor rice growth [62], [63]. The results of applying MODIS-based paddy rice phenological detection algorithm in classifying paddy growth stages are found to be encouraging and can be used to monitor paddy rice agriculture at a larger scale [62], [63].

Indices that correlate with vegetation cover are also used in estimating paddy yield and monitoring paddy growth such as the NDVI, which is mostly used to predict paddy rice yields' estimation [43], [57], [59], assessing damage of regional rainfed paddy rice [44] and monitoring rice growth [63]. NVDI is used to measure the level of greenness and biomass of vegetation. NDVI measurements are most often taken from satellites in orbit around the Earth. NDVI can be computed based on differences in the response patterns of vegetation in the red and NIR ranges as follows [65];

$$NDVI = \frac{\rho NIR - \rho Red}{\rho NIR + \rho Red} \quad (3)$$

where ρNIR is the reflectance in the NIR and ρRed is the reflectance in the red and this NDVI ranges between -1 (no vegetation) and +1 (green vegetation). Both the NDVI and EVI are most commonly used vegetation indices to monitor the health of vegetation on the fields [35], [43], [57], [59], [66]–[69]. However, some researchers have reported that EVI is often preferred than NDVI as EVI is more responsive to biophysical variables, such as LAI [35], [67]. For instance, EVI is more robust in capturing the difference in well-vegetated areas [67].

MNDWI is computed based on differences in the response patterns of vegetation in the green and SWIR1 ranges for the enhancement of open water features [70] and can be measured as follows;

$$MNDWI = \frac{\rho Green - \rho SWIR1}{\rho Green + \rho SWIR1} \quad (4)$$

where ρNIR is the reflectance in the NIR, ρSWIR1 is the reflectance in the Shortwave Infrared One. The integration of NDVI and MNDWI from Sentinel-2A image has shown increased accuracy of predicting the paddy rice yield estimation [57].

LAI is a dimensionless quantity that characterizes plant canopies that typically can be defined as the ratio of one sided leaf area per unit ground area (m^2/m^2) and can be considered as a measure of paddy crop growth and productivity since it characterizes plant canopy structure and gives an idea of the amount of biomass available in a field. LAI can be measured using a plant canopy analyzer [71]. Some works have been conducted to estimate paddy rice LAI with a fixed point continuous observation of near infrared reflectance using a calibrated digital camera [71], [72]. Estimating paddy rice LAI can also be done using machine learning methods [69] and also statistical methods [73] based on hyperspectral data. The leaf area index (LAI) and plant nitrogen concentration (PNC) were also used to estimate the nitrogen nutritional index (NNI) in paddy rice [74].

Hyperspectral images (Red, Blue, Green and Near Infrared One) can also be used to predict paddy rice yield estimation [75]–[77], assess the quality of paddy rice [69], [71], [72], [78], [79] and monitor paddy rice disease [80], [81].

Another remotely sensed data that is widely used in smart farming called *C-Band Synthetic Aperture Radar (SAR)* data. C-Band SAR data can be obtained from the Sentinel-1A satellite which provides a collection of data in all-weather, day or night. C-Band SAR data has been used in a wide range of applications that include sea and land monitoring. For instance, C-Band SAR has been used in predicting paddy rice yield estimation [82], monitoring paddy rice growth [83] and monitoring paddy rice disease [40].

3) DRONE BASED DATA

Next, drone based data include all imageries captured using the drone technology. The high resolution images captured using drone can be used to estimate the paddy rice yield [42], [84]–[86], monitor paddy rice disease [37], [39], [87]–[92], classify paddy rice samples [46], [47], [93]–[101] and also assess the quality of paddy rice [102]–[104]. For instance, a near real-time deep learning approach for detecting rice phenology has also been designed based on high resolutions images taken by using drones [86].

For instance, a SVM classifier can be used to perform segmentation and classification of paddy rice samples [46]. The prediction of nitrogen deficiency of rice crop can also be done to access the quality of the rice using deep learning methods [104].

B. APPLICATIONS OF MACHINE LEARNING ALGORITHMS IN PADDY RICE SMART FARMING

This section elaborates the roles of machine learning technology in paddy rice smart agriculture, by analyzing the applications of machine learning algorithms and smart technologies in various scenarios in the paddy rice production and post-production phases of the ASC. As mentioned earlier, intelligent data processing technologies can be applied in various scenarios in all the paddy rice production and post-production phases of the ASC and these tasks include smart irrigation for paddy rice, predicting paddy rice yield estimation, monitoring paddy rice growth, monitoring paddy

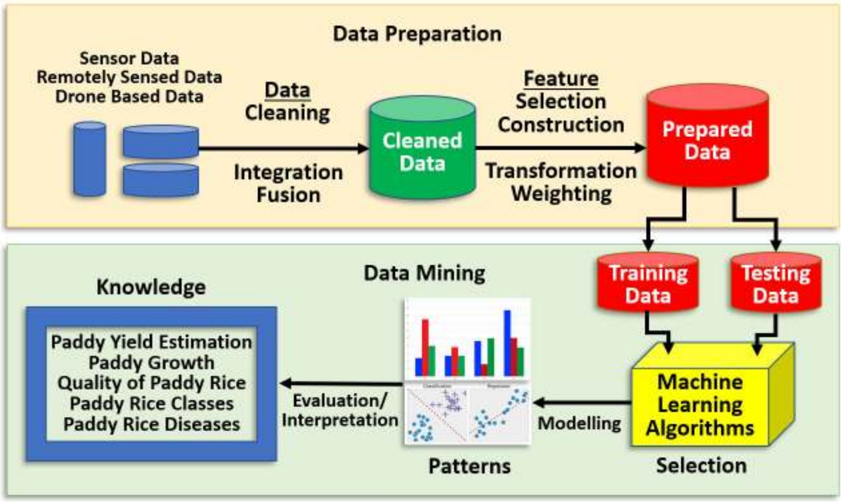

FIGURE 6. State-of-the-art for the tasks involved in the smart paddy rice farming.

TABLE 3. Applications of smart technologies (e.g., Internet of Things (IoT)) in various scenarios in the paddy rice pre-production and production phases of the ASC.

Tasks	Study	Devices	Findings
Estimating Paddy Rice Yield	[113]	Smart sensors for humidity, temperature, pH, soil moisture and light intensity	• Maximum quality of paddy rice harvested can be obtained.
	[114]	Smart sensors for climate, RFID, load Sensor, GPS	• The supply for paddy rice can be monitored in real time.
Monitoring paddy rice growth	[36]	Using smart sensors to monitor soil pH, lux and temperature	• It monitors soil pH and provides exact amount of water at different growth stages.
	[36]	MCU (ESP32) sensors for temperature, soil pH, sonar and light	
Smart Irrigation System	[32]	Microcontroller (LPC2148) and Eclipse SDK 3.6.2 compiler	• The proposed system comparatively requires a lesser amount of water and it outperforms than the existing systems.
	[33]	Water level sensor (FloodTech®)	• Automatic irrigation system also may cause a significant increase of rice production by making more arable land available for paddy rice plantation.
	[106]	Wireless Sensor and Actuator Network (WSAN).	• Water can be conserved up to 81%.
	[107]	Arduino/GSM, SIM900	• Saving water consumption by 19.91%

rice disease, assessing quality of paddy rice, paddy rice sample recognition and classification.

The state-of-the-art for the tasks involved in the smart paddy rice farming is illustrated in Fig. 6. First, all the acquired data (Sensor, Remotely sensed data and vegetation indices, and drone based data) will be cleaned, fusioned or integrated. Then, the dimensionality of the data can be reduced using feature selection, construction, transformation and weighting processes [162]–[164]. Next, once the data are prepared, then they will be divided into training and testing data depending on the types of task (e.g., classification, regression or clustering) or machine learning algorithms (e.g., estimation, linear and non-linear methods) used to model the data. Finally, model evaluation and interpretation will be performed to extract knowledge that supports the tasks in the smart paddy rice farming (e.g., Paddy Yield Estimation, Monitoring Paddy Growth, Assessing the Quality of Paddy Rice, Determining Paddy Rice Classes and Monitoring Paddy Rice Diseases).

1) SMART IRRIGATION SYSTEM FOR PADDY RICE

Automatic drip irrigation system requires a lesser amount of water to maintain a predetermined water height in paddy fields [32], [33] and this system can be controlled based on the captured climate data (e.g., temperature, humidity, light and rain) from sensors. Using a wireless sensor and actuator network (WSAN) to build a smart irrigation system for paddy fields can also conserve significant amount of water [106], [107]. Automatic irrigation system can cause a significant increase of rice production by making more arable land available for paddy rice plantation [33].

Besides that, smart sensors for climate and soil [36], [36], Radio-frequency identification (RFID), load Sensor and Global Positioning System (GPS) are also used in estimating paddy rice yield [113], [114]. Table 3 tabulates the applications of smart sensors (Internet of Things (IoT)) in various tasks involved in paddy rice smart farming.

TABLE 4. Applications of machine learning algorithms in predicting paddy rice yield estimation.

Study	Machine Learning (ML)	Data used in ML model	Findings	Learning Type
[34]	PLS, MLS	NIR, SWIR	• The estimation error of the crop yield was 12.78 kg/10a.	Regression
[35]	PLS	EVI, LSWI	• The R^2 explained 94% and 88% of variability in the data for the first and second crops, respectively.	Regression
[48]	ANN, RF, SVM	Climate and soil data	• ANN achieved better average (3.14%) and maximum error (9.66%).	Classification
[57]	CART	NDVI, MNDVI	• The integration of NDVI and MNDWI from Sentinel-2A image with temporal backscatter increased the accuracy by 0.08.	Classification
[58]	ISO	EVI	• Spatial distribution and phenology metrics derived from the EVI can be used to generate rice cropping patterns.	Clustering
[67]	SW	NDVI, EVI	• EVI is more robust in capturing the difference in well-vegetated areas.	Classification
[68]	EKF, UKF, MHE, MHE-PE	NDVI	• MHE-PE achieved better overall MAE of 15.41.	Regression
[75]	PCR, PLS	Hyperspectral bands 1 $\sim$ 4	• PLS produced higher R^2 of 0.984	Regression
[76]	CNN	Hyperspectral bands 1 $\sim$ 4	• CNN produced accuracy result of 81.68%	Classification
[77]	ANN	Hyperspectral Images	• ANN is able to estimate biomass with an average correlation of 0.76.	Regression
[82]	RF	C-band SAR	• Different spatial resolution or polarization combinations have significant impacts on mapped rice area.	Classification
[108]	OTSU	NDVI	• OTSU's method produce overall accuracy of 83.33%.	Classification
[109]	CNN, SVM	NDVI	• CNN achieved better overall accuracy of 97.06%.	Classification
[110]	ANN	Climate & Demography Data	• ANN achieved MAE of 0.0526.	Regression
[111]	MLR, ANN	Climate Data	• The study showed the robustness of ANN to forecast yields.	Regression
[112]	CNN	Climate Data	• Deep learning approach requires a large amount of time-series data to improve the prediction performance.	Regression
[155]	RNN (Sup), LSTM (Dem)	Climate Data	• RNN and LSTM achieved MSE values of 0.11 and 0.37 respectively.	Regression

2) PREDICTING PADDY RICE YIELD ESTIMATION

Most researches model the paddy rice estimation based on the hyperspectral and climate data in predicting paddy rice yield estimation (see Table 4). These studies conducted using various types of remotely sensed data and vegetation indices to predict paddy rice yield estimation [34], [35], [57], [58], [67]. Thus, one of the issues is determining the best combination of data obtained from remotely sensed data and vegetation indices to improve the accuracy of predicting paddy rice yield estimation. For instance, the integration of NDVI and MNDWI from Sentinel-2A image with temporal backscatter increased the accuracy by 0.08 [57]. Combining hyperspectral data (e.g., NDVI and MNDWI) will also increase the accuracy of estimating the paddy rice yield by using Classification And Regression Trees (CART) [57]. CART is one of the variants of Decision Tree (DT) classifiers that can be used for classification or regression predictive modeling problems [57], [66], [144]. DT is one of important types of algorithm for supervised learning, particularly in predictive modeling [78], [81], [126]. DT are constructed via an algorithmic approach that optimizes the splitting of a data set based on different conditions of the data features. In addition to that, using multi-features fusion method can also improve the accuracy of predicting paddy rice yields using a deep learning approach [112].

Partial Least Squares (PLS) algorithm can be found in many researches conducted to estimate the paddy rice yields [34], [35], [75]. For instance, short wave infrared region was found to be very essential for estimating the paddy yield using PLS algorithm [34]. PLS was developed based on the principal component regression that can be used to build models that can predict more than one dependent variable [63], [69], [136]. PLS was also found to produce higher R^2 of 0.984 compared to Principal Components Regression (PCR) in predicting paddy rice yield estimation [75]. PCR is based on Principal Component Analysis (PCA) that is used to analyze the multiple regression data that suffer from multicollinearity [132] (e.g., predicting paddy rice yield estimation [75]). Before any modelling can be performed, PCA can be used to extract features of the datasets [128]. PCA is a well-known technique used for reducing the dimensionality of the datasets [129]. This is done to increase the interpretability but at the same time minimizing information loss [130], [131].

A few variants of deep learning algorithms have also been used to predict paddy rice yield estimation based on NDVI [109], climate data [48], [110]–[112], [155] and hyperspectral data (Bands 1 $\sim$ 4) [75]–[77] with higher accuracy results. These deep learning algorithms include Artificial Neural Network (ANN) [48], [77], [110], [111], Convolutional Neural Network (CNN) [76], [109], [112], Recurrent Neural Network (RNN) [155]). For instance, neural network algorithms achieved better overall accuracy compared to Random Forest (RF) and Support Vector Machine (SVM) using either the hyperspectral or climate data [48], [109], [110]. Inspired by the way biological nervous systems, ANN is basically an information processing technique that works like the way human brain processes information [150]. An ordinary neural network may consist of hidden layers and weights while CNN has filters which collectively make up the convolution layers. CNN is most commonly applied to analyze images and it is a class of deep neural networks. CNN is suitable to be used for spatial data such as images. In contrast, RNN is suitable to be used for temporal data which is also called sequential data. Compared to ANN, RNN is able to learn time-series data since it has a recurrent connection on the hidden state and this looping constraint ensures that sequential information is captured in the input data [151], [155]. Although, deep learning algorithms are known to be very effective and robust to forecast yields paddy rice yield estimation, [76], [77], [109], [111], [155] they require a large amount of time-series data to improve the prediction performance [112].

RF requires two parameters namely the number of trees and the number of features to split the data set based on different conditions [143]. RF has been found to be effective in predicting paddy rice yield estimation and monitoring paddy rice growth [48], [82]. Several works related to applying SVM in paddy rice smart farming have been reviewed in this paper [48], [109]. However, they produced lower accuracies compared to deep learning algorithms. SVM is a supervised machine learning model that can be used for binary classification tasks [146]. The objective of the SVM is to find the optimum hyperplane in an N-dimensional space that can distinctly classifies the data points.

Unsupervised learning algorithm can also be used to predict the paddy rice estimation using the hyperspectral data [58]. For instance, Iterative Self-Organizing (ISO) has been used to generate paddy cropping pattern to predict paddy rice yield estimation [58]. ISO is an unsupervised learning algorithm that can be used to generate rice cropping patterns [58]. The ISO algorithm is a modification of the k-means clustering algorithm. The merging and splitting of clusters are based on a predefined threshold by the user. If the difference of distance in multispectral feature space is less than the predefined threshold, the merging or splitting of clusters will be performed [153].

There are several optimization approaches that produce estimates of unknown variables or parameters based on a series of measurements observed over time, such as the Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF) and Moving horizon estimation (MHE) [133], that can be used to predict paddy rice yield estimation. Moving Horizon Estimator with Pre-Estimation (MHE-PE) is an optimization-based estimator introduced and use an auxiliary estimator to describe the dynamics of the state over the horizon [134], [135]. MHE-PE is found to be more effective compared to MHE [68] for crop start date estimation in tropical area [68].

Some of the limitations found in these studies include the resolution limitations, topographic effects and limited and small size of time-series data that lead to estimation errors. For instance, the low fractional coverage of small-size rice paddies in the complex and hilly landscapes could also lower the probability of identification using the OTSU's algorithm [108]. OTSU's method is an image segmentation algorithm that segments a gray level image with only one modal distribution in gray level histogram [100], [148]. Stepwise classification (SW) is another classification approach that applies a strategy that combines two heterogeneous data sets in a novel way, and this can be used in estimating rice yields production [67]. Table 4 tabulates the applications of machine various learning algorithms found in some of works to predict paddy rice yield estimation.

3) MONITORING PADDY RICE GROWTH

Monitoring the growth of paddy rice can be performed by mapping paddy rice and assessing the growth stages of the paddy rice. One of the issues or challenges in monitoring paddy rice growth using machine learning algorithms is to determine the optimum features combination. With optimum features combination, the overall accuracy of the classification results can be improved [115]. For instance, the optimum features combination can be achieved by using the robust adaptive spatial temporal fusion model (RASTFM) [116]. NDVI [63], [66], [69], [115], EVI [63] and Hyperspectral bands $1 \sim 4$ [73], [78] are the most commonly used in monitoring the growth of paddy rice.

The Multilayer Perceptron (MLP) [63], a class of feedforward ANN, and RF [69], [115] algorithms show better accuracies [69], [115] compared to PLS, SVM [63], [78], [83] and Support Vector Regression (SVR) [69] in performing the paddy growth stages classification. SVR is characterized by the use of kernels, sparse solution, and the original control of the margin and the number of support vectors [141]. SVR trains using a symmetrical loss function, which equally penalizes high and low misestimates and it has been proven to be an effective algorithm in estimating real-value [69].

Least-squares support-vector machines (LS-SVM) is found to produce better results compared to Multiple Linear Regression (MLR) and PLS, in estimating LAI of paddy rice from optimal hyperspectral bands [73]. LS-SVMs are least-squares versions of SVM which can be used for classification and regression analysis problems [73], [123], [140]. MLR is a statistical technique that uses several independent variables to predict the outcome of the dependent variable [34], [73], [111], [137]. Multiple regression is an extension of linear (OLS) regression that applies only one independent variable.

Besides remotely sensed data, vegetation indices, climate and soil data obtained from smart sensors are also used in monitoring paddy rice growth [36], [36] (see Table 3). Table 5 tabulates the applications of machine various learning algorithms in monitoring paddy rice growth.

4) MONITORING PADDY RICE DISEASE

The color of the paddy rice leaves will change when they are infected by any disease and these colored spots are

TABLE 5. Applications of machine learning algorithms in monitoring paddy rice growth.

Study	Machine Learning (ML)	Data used in ML model	Findings	Learning Type
[63]	MLP, PLS, SVM, RF	EVI, NDVI, LSWI	• MLP using multiple regularization Dropout and Batch Normalization achieves the highest accuracy.	Classification
[66]	CART	NDVI	• Predicting nutrient value for the plants is feasible using hyperspectral data.	Classification
[69]	SVR, RF, ANN, PLS	NDVI, LAI	• RF models showed highest potential for estimating the LAI of paddy rice.	Classification
[73]	MLR, PLS, and LS-SVM	Hyperspectral bands $1 \sim 4$	• LS-SVM models gave best results with RMSE of 0.855.	Regression
[78]	DT, SVM	Hyperspectral bands $1 \sim 4$	• Rice canopy hyperspectral imagery can be effectively classified by using DT and SVM.	Classification
[83]	SVM	SAR	• RADARSAT-2 quad polarization exhibited good potential for monitoring rice growth.	Classification
[115]	RF	NDVI, LAI	• With optimum features combination, the overall accuracy and Kappa coefficient of the classification results are higher than 95% and 0.93.	Classification
[84]	CNN	High-resolution images	• Automatic navigation of machinery can be performed more efficiently.	Classification

created on leaves. For that reason, most of the researches used high-resolution images in monitoring the paddy rice disease [37], [39], [41], [87]–[92], [117]–[120] and hyperspectral images [80], [81] to detect and assess the paddy rice diseases. The ANN algorithm and its variants, CNN, are found to be very effective in classifying task for monitoring the paddy rice diseases [39]–[41], [81], [87], [88], [90], [118]. For instance, the ANN achieved better classification results compared to FC and SVM algorithms [39] and the calibrated CNN model still showed good classification ability in a small-scale sample set and it was selected as the best classification model compared to DT, k-NN and SVM [81]. However, CNN requires a large number of samples for training purposes [88], [112]. In fuzzy classification (FC) applications, once a set of classes has been defined, one can determine the degree of membership of every object x under consideration [149]. Fuzzy classification allows object x to belong to two or more classes.

k-Nearest Neigbour (k-NN) algorithm is also very effective in detecting diseases from paddy plant leaf images and identifying Brown Planthopper in paddy field and other classification problems [37], [81], [92], [96], [117]. Given an unknown sample, k-NN finds k samples that are nearest to this unknown sample based on certain distance functions (e.g., Euclidean or Cosine distance methods) and take the average of the response variables from these k samples as the label (class) of the unknown samples [145]. k-NN can be used for paddy rice sample classification [59], [99], [103]. Compared to SVM, k-NN produces better accuracy in detecting and recognizing diseases from paddy plant leaf images [37].

Some combined approaches show promising results that involve deep learning approaches [40], [41] and SVM algorithms [40], [91]. For instance, a combination approach of two machine learning algorithms (e.g., CNN + SVM) has been used to identify the cultivated paddy regions (e.g., Using CNN), and to detect areas damaged (e.g., Using SVM) by Brown Planthopper attacks [40]. Other works include building a semantic framework that models an ontology related to rice plant knowledge and applying this framework to help farmers to identify rice diseases, receive early warnings of possible spreadable diseases, and receive treatments based on multiple observations [121].

Minimum Distance Classifier (MDC) achieved better accuracy compared to k-NN in classifying high-resolution images for monitoring paddy rice disease [117]. MDC classifies unknown sample data to classes which minimize the distance between this sample data and the class in multi-feature space [147]. One of the works reviewed has applied MDC to classify images in the task of monitoring and controlling rice diseases using Image processing techniques [117].

There are also researches conducted on developing expert systems using optimized fuzzy inference system (OFIS) [122] and forward chaining [89] for monitoring paddy rice disease. Table 6 tabulates the applications of machine various learning algorithms in monitoring paddy rice disease.

5) ASSESSING QUALITY OF PADDY RICE

The quality of paddy rice can be assessed using the hyperspectral data [74], [79], climate and soil data [105] and also high-resolution images of the paddy rice [93], [104], [123]. SVM and CNN algorithms are the two most commonly used machine learning algorithms for assessing the quality of paddy rice [79], [93], [104], [105], [123]. CNN is found to be more effective compared to SVM algorithm in assessing the quality of the paddy rice [93]. However, a combination of classical artificial neural networks and SVM also has been used to predict nitrogen deficiency of rice crop [104].

Fuzzy c-means (FCM) has also been used to assess the quality of the paddy rice. FCM is a method of clustering which allows one piece of data to belong to two or more clusters [74], [154]. Table 7 tabulates the applications of machine various learning algorithms in assessing quality of paddy rice.

TABLE 6. Applications of machine learning algorithms in monitoring paddy rice disease.

Study	Machine Learning (ML)	Data used in ML model	Findings	Learning Type
[37]	k-NN, SVM	High-resolution images	• k-NN produced better accuracy compared to SVM with accuracies of 93.33% and 91.10% respectively.	Classification
[39]	FC, ANN, SVM	High-resolution images	• ANN achieved better classification results.	Classification
[40]	CNN + SVM	SAR	• The combined approach shows better results with an accuracy of 96.31%	Classification
[41]	CNN + LSTM	High-resolution images	• The CNN + LSTM model achieved accuracy rates of 89.33%.	Classification
[59]	k-Means	NDVI, EVI	• k-Means-Euclidean- achieved better accuracy than k-Means-Mahalanobis.	Classification
[80]	PCA	Hyperspectral Images	• Qualitative analysis by PCA showed the spectral difference and separability between healthy and infected rice kernels.	
[81]	DT, k-NN, SVM, CNN	Hyperspectral Images	• The calibrated CNN model still showed good classification ability in a small-scale sample set and it was selected as the best classification model.	Classification
[87]	CNN	High-resolution images	• CNN achieved maximum average stress classification accuracy of 95.08% by learning over 6000 images.	Classification
[88]	CNN	High-resolution images	• The Accuracy rate of training reached to 98%, 99% and 96% respectively for model Inception-V3, MobileNet-V1 and ResNet-50. The validation accuracy rate of ours is about 98%.	Classification
[89]	Expert System: Forward Chaining	High-resolution images	• The expert system produced accuracy of 73.81%	Classification
[90]	ANN	High-resolution images	• ANN achieved accuracy of 99.0%	Classification
[91]	AdaBoost + SVM	High-resolution images	• AdaBoost+SVM achieved an 85.2% detection rate and a 9.6% false detection rate.	Classification
[92]	k-NN	High-resolution images	• k-NN achieved precision 0.97, recall 0.96, accuracy 0.97 and F score 0.96.	Classification
[117]	MDC, k-NN	High-resolution images	• k-NN and MDC achieved accuracies of 87.02% and 89.23% respectively.	Classification
[118]	BPNN, SVM, CNN	High-resolution images	• CNNs model has a better training performance, faster convergence rate, as well as a better recognition ability than the other model.	Classification
[119]	SVM	High-resolution images	• SVM achieved prediction accuracy of 93%.	Classification
[120]	SVM	High-resolution images	• SVM achieved 83.0% accuracy	Classification
[122]	Optimized Fuzzy Inference System (OFIS)	High-resolution images	• OFIS provided accuracy of 95%.	Classification

TABLE 7. Applications of machine learning algorithms in assessing quality of paddy rice.

Study	Machine Learning (ML)	Data used in ML model	Findings	Learning Type
[74]	FCM	LAI + PNC	• The generated NNI maps based on LAI & PNC can be used to monitor crops in near-real.	Clustering
[79]	SVM	Hyperspectral bands $1 \sim 4$	• Hyperspectral images are vital for real-time monitoring soil heavy metal contamination.	Classification
[93]	CNN, SVM	High-resolution images	• CNN achieved better accuracy of 99.5%.	Classification
[104]	CNN + SVM	High-resolution images	• ResNet-50+SVM is the best classification method for prediction of Nitrogen deficiency of rice crop compared to other variants of deep learning methods.	Classification
[105]	CNN	Soil's pH and moisture, nitrogen, organic carbon	• CNN produced better results with larger dataset.	Regression
[123]	LS-SVM	High-resolution images	• LS-SVM produced accuracy of 98.20%.	Classification

6) PADDY RICE SAMPLE CLASSIFICATION

Machine learning algorithms are normally combined with computer vision techniques to perform paddy rice sample classification with more effectively. Applying computer vision and machine learning techniques to recognize and classify rice varieties is a method that can be used to increase the accuracy of classification process in real applications. Several studies have been conducted that apply and examine several morphological and textural features of rice seeds' images to evaluate their efficacy in identification of rice varieties [97] and classification of paddy rice adulteration levels [96]. In most studies related to the application of machine learning algorithm for paddy rice sample classification, deep learning algorithms are found to be very effective in classifying rice samples [94]–[97], [99], [101], [124].

The classification of the paddy rice samples can be improved with PCA-based reduced features [96], [103], [124]. PCA can be combined with other classifiers to

TABLE 8. Applications of machine learning algorithms in paddy rice sample classification.

Study	Machine Learning (ML)	Data used in ML model	Findings	Learning Type
[46]	SVM	High-resolution images	• The accuracy of segmentation and classification is 96.0% and 88.0%.	Classification
[94]	CNN (AlexNet, VGG-16, Inception-v3, ResNet-34)	High-resolution images	• ResNet-34 produced highest F1-score of 91.89%.	
[95]	CNN, AdaBoost and SVM	High-resolution images	• CNN has better performance than AdaBoost and SVM with different light intensity.	Classification
[96]	PCA+BPNN, PCA+k-NN, PCA+SVM	High-resolution images	• The maximum average adulteration level classification accuracy of 93.31% is obtained using the PCA+BPNN.	Classification
[97]	MLP	High-resolution images	• The maximum classification accuracy of 84.62% can be obtained using a three-layered feed forward type (MLP).	Classification
[98]	SVM, RF	Hyperspectral Images of Paddy Leaf	• SVM model achieved better accuracy of 100%, 100%, and 92% recognition rates for barnyard grass, weedy rice and rice, respectively.	Classification
[99]	SVM + CNN + k-Means	High-resolution images	• The detection accuracy of spike patches by CNN reached 86.7%.	Classification
[100]	OTSU + SVM	High-resolution images	• The overall accuracy of the algorithm reached 96.4% on average.	Classification
[101]	CNN	High-resolution images	• CNN achieved accuracies ranging from 93% to 99%.	Classification
[103]	PCA + k-Means	High-resolution images	• H channel data can provide clusters with considerable separability as measured using separability index measures.	Clustering
[124]	PCA + BPNN	High-resolution images	• PCA + BPNN achieved better accuracy than BPNN.	Classification
[127]	SVM	Rice organic elements	• SVM produces high accuracy of classifying organic rice samples with 98%.	Classification

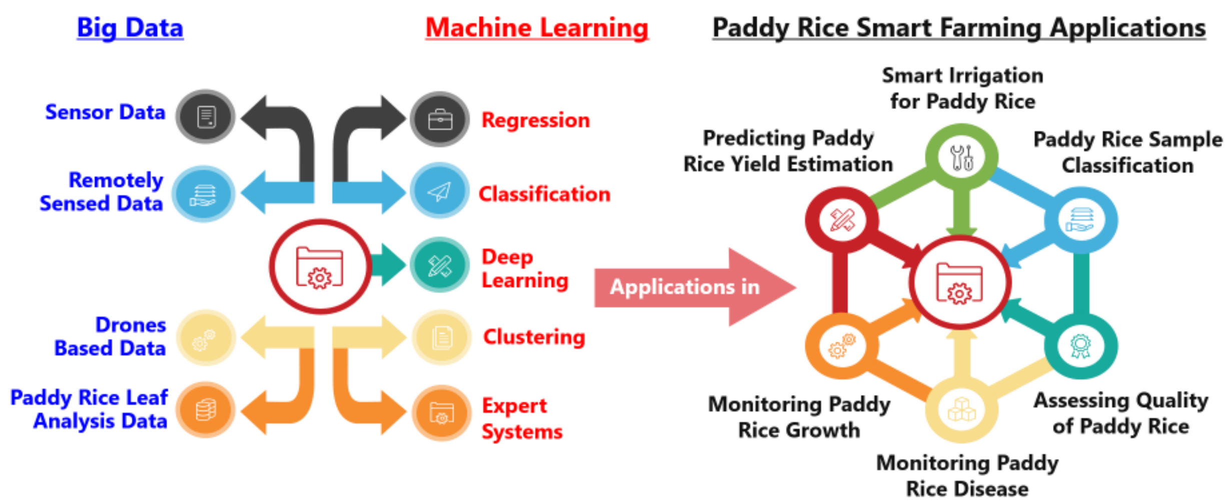

FIGURE 7. Mapping of big data, machine learning and paddy rice smart farming tasks.

improve the accuracy of paddy rice sample classification [96], [103], [124] and also to perform qualitative analysis in monitoring paddy rice disease [80].

Deep learning algorithms (e.g., BPNN, CNN) produced better accuracy compared to SVM algorithm [46], [101] in classifying paddy rice samples [95], [96]. When the label or number of varieties are not available, an unsupervised learning algorithm, such as clustering algorithm, can also be used to cluster paddy rice samples. For instance, k-Means clustering algorithm provides clusters with considerable separability as measured using separability index measures [103] based on the PCA-based reduced features. In k-means clustering, n observations are partitioned into k clusters in which each observation is assigned to the nearest cluster centroid.

The k-means clustering is also known as a method of vector quantization [152]. By using the k-means clustering method in paddy rice sample classification, the H channel data can provide clusters with considerable separability as measured using separability index measures [103]. k-means clustering also can be used as part of the approach to classify the annual cropping patterns of paddy crop based on k number of classes [59].

Adaptive Boosting (AdaBoost) has been used to classify paddy rice samples. AdaBoost algorithm combines multiple weak classifiers to form a single strong classifier [91], [95], [142]. AdaBoost is also known as ensemble method as it consists of multiple weak classifiers. However, deep learning algorithms are found to be more superior than AdaBoost algorithm in classifying paddy rice samples [91], [95].

A multi-classifier cascade based rice spike detection method has also been proposed that consists of SVM, CNN and k-Means algorithm [99]. Other works include training machine learning algorithms to predict weight and size of rice kernels [125], application of machine learning algorithm in detecting adulterated admixtures of white rice based on mass spectrometry data [126] and classifying organic rice samples using original rice elements [127]. Table 8 tabulates the applications of machine various learning algorithms in paddy rice sample classification.

V. RESULTS AND CONCLUSION

Based on the reviews of several works in this paper, a new framework is proposed that maps three entities that include big data, machine learning and paddy rice smart farming tasks. In this review, the types of machine learning algorithms used are highly dependent on the availability of data. At the same time, the type of data required depends directly on the type of tasks stated in each production and post-production phases of paddy rice smart farming. These machine learning algorithms are used to perform the intelligent data processing that will assist farmers in various tasks mentioned in the production and post-production phases. Based on the findings summarized in the previous sections, machine learning algorithms and smart technologies can be used to improve the overall efficiency of the paddy rice production system. The potential benefits lead to an improvement in the return of investment (ROI) for all paddy rice production systems by minimizing the losses or costs involved in the production of paddy rice. As a result, we use these findings found in the literature to map these three components (e.g., datasets, machine learning algorithms and tasks stated in the production of paddy rice) and develop a Big Data-ML-Task applications framework that can be used by the practitioners. The proposed framework shown in Fig. 7 has three main components, the types of datasets, types of ML algorithms, the types of tasks in paddy rice smart farming and paddy rice supply chain performance.

With smart irrigation system, the usage of water can be reduced and at the same time fully utilized to increase the paddy rice yield [32], [106], [107]. Automatic irrigation system also may cause a significant increase of rice production by making more arable land available for paddy rice plantation [33].

The task of estimating the yield of paddy rice precisely is very important for national food security and development evaluation. The development of an integrated aerial crop monitoring solution using an Unmanned Aerial Vehicle (UAV) has motivated researchers to apply vegetation indices retrieved from hyperspectral images to estimate paddy rice yield [77]. Several studies have estimated the paddy rice yield based on time-series climate data [48], [110]–[112], [155]. Rainfall was found the main climate driver of the rice yield [111]. Other studies considered hyperspectral data to estimate the yield of paddy rice [34], [35], [57], [58], [67], [68], [75], [76], [108], [109]. Deep learning algorithms were found to be more effective compared to other machine learning algorithms for modeling paddy rice yield [48], [76], [77], [109]–[112], [155]. Maximum quality of paddy rice harvested can be obtained by using sensors to monitor humidity, temperature, pH, soil moisture and light intensity in real [113], [114].

Monitoring the growth of paddy rice is critical for understanding the growing status and yield estimation of paddy rice. For instance, the self-sufficient level (SSL) for paddy rice in Malaysia is only 70%. As the world population is increasing, intensifying paddy rice farming is more preferable over the expansion of agriculture land due to limited arable land [156]. Monitoring the growth of paddy rice is difficult for traditional farmers due to climate change, soil conditions, age of the farmers and time consumed to monitor the whole area. With remotely sensed data, creating paddy rice crop growth map is possible using the hyperspectral images [66] and synthetic aperture radar (SAR) data [83], [115]. For example, the paddy rice growth based on rice growth parameters (e.g., rice height and biomass) can be monitored with the backscattering coefficient from RADARSAT-2 data [83]. The paddy rice leaf chlorophyll contents can also be retrieved from the rice canopy hyperspectral imagery to analyze the paddy rice plant growth [63]. Leaf area index (LAI) is commonly used as a surrogate for productivity in precision agriculture (PA) and is widely used in plant growth [69], [73]. In short, the applications of machine learning algorithms have enabled us to timely and accurately monitor paddy rice planting area for national food security and management [115]. Using smart sensors to monitor soil pH, lux and temperature also provides insight in understanding the stages of paddy rice growth [36], [36].

Due to the lack of knowledge and awareness of suitable management to rectify rice plant leaf diseases, the rice production is being reduced in recent years [157]. The manual detection of plant diseases based on naked eye observation of experts is very time consuming, expensive and sometimes it produces an error when identifying the disease type [158]. Machine learning (ML) algorithm can be used to provide early warnings to anticipate rice blast and detect its presence, thus supporting the applications of biocidal chemical compounds or biological organisms used to kill parasitic fungi or their spores. Based on several studies reviewed in this paper, the applications of ML, in detecting the presence of rice blast, has also provided suitable solutions for preventive remedial actions targeting the mitigation of yield losses and the reduction of fungicide use [159]. This review will be beneficial for modelers, farmers and stakeholders, to guide them in model development and selection for the most suitable models for the effective paddy rice disease detection and forecasting. The identification of paddy diseases may also assist farmer in providing them the remedies based on the types of disease [160].

The quality of paddy rice production depends highly on the quality of soil properties. These soil properties include soils' pH and moisture, nitrogen and organic carbon content of the soil. For instance, a CNN produced promising results in assessing the nitrogen deficiency of paddy rice crop [104]. These soil properties can be captured using sensors or retrieved from the hyperspectral images [74]. SVM and CNN are the two most common machine learning algorithms used in assessing the quality of paddy rice. Compared to SVM, CNN produced better assessment accuracy [93]. Besides soil properties [105], some studies have conducted the assessment of the paddy rice quality based on the high-resolution images of the paddy rice leaf [93], [104], [123] and the hyperspectral images of the paddy rice field [74], [79].

Improving the management and productivity of the paddy rice farming is important to strengthen the food security initiatives. Due to variation in economic value of different varieties of rice, rice quality identification is very important in the international and national rice market [97], [100], [101]. The quality of the rice is used to evaluate the milling process. Rice sample may consist of full rice, broken rice, damaged rice, paddy, stones and foreign objects. Image processing and machine learning techniques can be used to separate and classify objects of rice sample [46]. Other than hyperspectral images [98], most of the studies related to paddy rice sample classification use high-resolution images and apply machine learning techniques such as SVM [46], [96], [98]–[100], [127] and deep learning algorithms [94]–[98], [101], [124]. Combining efficient feature extraction method (e.g., PCA) [103] with neural network algorithm (e.g., Back-propagation Neural Network (BPNN)) shows better accuracy results in paddy rice sample classification [96], [124] and also better clustering results for paddy rice grade identification [103]. Other image pre-processing such as histogram of oriented gradients (HOG) also affects the performance of the classifiers [94], [99]. Combining features in paddy rice sample classification also improves the classification accuracies [96].

Multi-classifier cascade can also be used to improve the performance of the paddy rice sample classification [99]. In order to get a good model, low bias and variance are required in order to have high accuracies or lower errors. An optimal balance of bias and variance would never overfit and underfit the model. Reducing variance of the final classifier model can be achieved by fitting multiple final models or using hybrid approaches [99], [103], [124] or increase the training size. In addition to that, example of low-bias machine learning algorithms include DT, k-NN and SVM. Based on the findings of this review, the performance of all these three machine learning algorithms are very competitive in predicting paddy rice yield estimation [109], monitoring paddy rice growth [78], [83], monitoring paddy rice disease [37], [40], [81], [91], [117], [119], [120], assessing quality of paddy rice [79], [104], [123] and paddy rice sample classification [46], [98], [99], [127].

VI. CONCLUSION

This paper provides a structured overview of the recent applications of machine learning algorithms and smart devices for paddy rice smart farming. In addition to that, this paper has proposed a framework that maps big data, machine learning and paddy rice smart farming tasks. The review study reveals considerable benefits to the production of paddy rice that have applied the machine learning techniques and smart devices in the paddy rice smart farming. As with any research, here, we also summarize the following guidelines based on the findings obtained from this review for future works.

First, there is a need to explore further the capability of ensemble models or hybrid models based on deep learning methods using multi-source data, as these have been shown to improve the performance of the base model. However, deep learning methods require large number of samples to come up with efficient models. For instance, in predicting paddy rice yield and monitoring paddy rice disease using the deep learning approach, a large amount of time-series data is required to improve the prediction performance [88], [112]. Since most of the studies conducted for paddy rice sample classification are based on image processing, the optimization of the classification accuracy (e.g., using hybrid or ensemble approach) is another issue that requires more explorations. For instance, more works on the variety of wavelet transforms for texture analysis and different classification techniques (decision tree, random forest) for paddy rice sample classification can be explored [96].

Second, a limited number of investigations conducted in the area of the application of machine learning algorithm based on multi-sources data as the findings from existing studies have shown that a more comprehensive understanding can be obtained by integrating multi-sources data or determining the optimum features combination. We can produce better modelling results comprehensively by analysing these complex relationships among multi-sources data or by finding the optimum features combination. For instance, using only spectral reflectance, shape and texture of paddy rice will not provide better results and additional ground truth data is required in order to classify and differentiate paddy rice accurately [161]. Using multi-features fusion (e.g., combining Landsat and SAR Time Series Data) can also improve the accuracy of predicting paddy rice yield using a deep learning approach [112]. Limited works are found in exploring and combining multiple sources of data (e.g., Sensored data (climate and soil properties), Remotely sensed data, vegetation indices and drone-based data (e.g., high-resolution images)) to improve the modelling of data for smart irrigation for paddy rice, predicting paddy rice yield estimation, monitoring paddy rice growth, monitoring paddy rice disease, assessing quality of paddy rice, paddy rice sample classification.

Finally, a more comprehensive analysis needs to be conducted to investigate the efficiency of processing software to perform image preprocessing for modelling. For example, Monitoring the growth of paddy rice based on spectral reflectance has limitations of the processing software and the complicated steps to process the images [66]. More researches need to be conducted to acquire high resolutions remotely sensed time series imagery data in both time and space through effective and efficient image segmentation process using data blending approaches [108].

REFERENCES

[1] *Growing at a Slower Pace, World Population is Expected to Reach 9.7 Billion in 2050 and Could Peak at Nearly 11 Billion Around 21*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html

[2] Sharala Axryd and Chari TVT. *A Digital Solution Towards Data-Driven Agriculture in Malaysia, Digital News Asia*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.digitalnewsasia.com/insights/digital-solution-towards-data-driven-agriculture-malaysia

[3] C. Lieber. *A Scientist on the Myth of Ugly Produce and Food Waste*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.vox.com/the-goods/2019/2/26/18240399/food-waste-ugly-produce-myths-farms

[4] J. Reddy. *Post Harvesting Technology of Vegetables, AgriFarming*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.agrifarming.in/post-harvesting-technology-of-vegetables

[5] *What is the Difference Between Precision, Digital and Smart Farming?* Accessed: Apr. 15, 2020. [Online]. Available: https://www.agrocares.com/en/news/precision-digital-smart-farming/

[6] M. M. Morshed, M. S. Islam, H. D. Lohano, and P. Shyamsundar, "Production externalities of shrimp aquaculture on paddy farming in coastal Bangladesh," *Agricult. Water Manage.*, vol. 238, Aug. 2020, Art. no. 106213, doi: 10.1016/j.agwat.2020.106213.

[7] D. Boansi, J. A. Tambo, and M. Müller, "Analysis of farmers' adaptation to weather extremes in West African Sudan Savanna," *Weather Climate Extremes*, vol. 16, pp. 1–13, Jun. 2017, doi: 10.1016/j.wace.2017.03.001.

[8] T. W. Reynolds, S. R. Waddington, C. L. Anderson, A. Chew, Z. True, and A. Cullen, "Environmental impacts and constraints associated with the production of major food crops in sub-Saharan Africa and South Asia," *Food Secur.*, vol. 7, no. 4, pp. 795–822, Aug. 2015, doi: 10.1007/s12571-015-0478-1.

[9] S. Mar, H. Nomura, Y. Takahashi, K. Ogata, and M. Yabe, "Impact of erratic rainfall from climate change on pulse production efficiency in lower Myanmar," *Sustainability*, vol. 10, no. 2, p. 402, Feb. 2018, doi: 10.3390/su10020402.

[10] M. Yasar, C. Siwar, and R. B. R. Firdaus, "Assessing paddy farming sustainability in the Northern Terengganu integrated agricultural development area (IADA KETARA): A structural equation modelling approach," *Pacific Sci. Rev. B, Humanities Social Sci.*, vol. 1, no. 2, pp. 71–75, Jul. 2015, doi: 10.1016/j.psrb.2016.05.001.

[11] S. A. Mokhtor, D. El Pebrian, and N. A. A. Johari, "Actual field speed of rice combine harvester and its influence on grain loss in Malaysian paddy field," *J. Saudi Soc. Agricult. Sci.*, vol. 19, no. 6, pp. 422–425, Sep. 2020, doi: 10.1016/j.jssas.2020.07.002.

[12] W. Wu, W. Wang, M. E. Meadows, X. Yao, and W. Peng, "Cloud-based typhoon-derived paddy rice flooding and lodging detection using multi-temporal Sentinel-1&2," *Frontiers Earth Sci.*, vol. 13, pp. 682–694, Dec. 2019. [Online]. Available: https://doi-org.ezproxy.ums.edu.my/10.1007/s11707-019-0803-7

[13] T. Chapagain and M. N. Raizada, "Impacts of natural disasters on smallholder farmers: Gaps and recommendations," *Agricult. Food Secur.*, vol. 6, no. 1, p. 39, Dec. 2017, doi: 10.1186/s40066-017-0116-6.

[14] D. Jiménez, H. Dorado, J. Cock, S. D. Prager, S. Delerce, A. Grillon, M. A. Bejarano, H. Benavides, and A. Jarvis, "From observation to information: Data-driven understanding of on farm yield variation," *PLoS ONE*, vol. 11, no. 3, Mar. 2016, Art. no. e0150015, doi: 10.1371/journal.pone.0150015.

[15] *Solving the Problems of Problem Solving*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.lens.org/

[16] R. Sharma, S. S. Kamble, A. Gunasekaran, V. Kumar, and A. Kumar, "A systematic literature review on machine learning applications for sustainable agriculture supply chain performance," *Comput. Oper. Res.*, vol. 119, Jul. 2020, Art. no. 104926, doi: 10.1016/j.cor.2020.104926.

[17] Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali, "Review—Machine learning techniques in wireless sensor network based precision agriculture," *J. Electrochem. Soc.*, vol. 167, no. 3, 2020, Art. no. 037522. 10.1149/2.0222003JES.

[18] L. Santos, F. N. Santos, P. M. Oliveira, and P. P. Shinde, "Deep learning applications in agriculture: A short review," in *Robot 2019: Fourth Iberian Robotics Conference* (Advances in Intelligent Systems and Computing), vol. 1092, M. F. Silva, J. L. Lima, L. P. Reis, A. Sanfeliu, and D. Tardioli, Eds. Cham, Switzerland: Springer, 2020.

[19] N. Zhu, X. Liu, X. Liu, K. Hu, Y. Wang, J. Tan, M. Huang, Q. Zhu, X. Ji, Y. Jiang, and Y. Guo, "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 4, pp. 32–44, 2018, doi: 10.25165/j.ijabe.20181104.4475.

[20] K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artif. Intell. Agricult.*, vol. 2, pp. 1–12, Jun. 2019, doi: 10.1016/j.aiia.2019.05.004.

[21] A. Kamilaris and F. X. Prenafeta-Boldu, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.

[22] A. Villa-Henriksen, G. T. C. Edwards, L. A. Pesonen, O. Green, and C. A. G. Sørensen, "Internet of Things in arable farming: Implementation, applications, challenges and potential," *Biosystems Eng.*, vol. 191, pp. 60–84, Mar. 2020.

[23] J. Astill, R. A. Dara, E. D. G. Fraser, B. Roberts, and S. Sharif, "Smart poultry management: Smart sensors, big data, and the Internet of Things," *Comput. Electron. Agricult.*, vol. 170, Mar. 2020, Art. no. 105291, doi: 10.1016/j.compag.2020.105291.

[24] A. R. Yanes, P. Martinez, and R. Ahmad, "Towards automated aquaponics: A review on monitoring, IoT, and smart systems," *J. Cleaner Prod.*, vol. 263, Aug. 2020, Art. no. 121571, doi: 10.1016/j.jclepro.2020.121571.

[25] D. Glaroudis, A. Iossifides, and P. Chatzimisios, "Survey, comparison and research challenges of IoT application protocols for smart farming," *Comput. Netw.*, vol. 168, Feb. 2020, Art. no. 107037, doi: 10.1016/j.comnet.2019.107037.

[26] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming–a review," *Agricult. Syst.*, vol. 153, pp. 69–80, May 2017, doi: 10.1016/j.agsy.2017.01.023.

[27] A. Lytos, T. Lagkas, P. Sarigiannidis, M. Zervakis, and G. Livanos, "Towards smart farming: Systems, frameworks and exploitation of multiple sources," *Comput. Netw.*, vol. 172, May 2020, Art. no. 107147, doi: 10.1016/j.comnet.2020.107147.

[28] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, S. Wan, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, G. Karagiannidis, and S. K. Goudos, "Internet of Things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review," *Internet Things*, Mar. 2020, Art. no. 100187, doi: 10.1016/j.iot.2020.100187.

[29] F. Balducci, D. Impedovo, and G. Pirlo, "Machine learning applications on agricultural datasets for smart farm enhancement," *Machines*, vol. 6, no. 3, p. 38, Sep. 2018.

[30] O. Ahumada and J. R. Villalobos, "Application of planning models in the agri-food supply chain: A review," *Eur. J. Oper. Res.*, vol. 196, no. 1, pp. 1–20, Jul. 2009, doi: 10.1016/j.ejor.2008.02.014.

[31] V. Borodin, J. Bourtembourg, F. Hnaien, and N. Labadie, "Handling uncertainty in agricultural supply chain management: A state of the art," *Eur. J. Oper. Res.*, vol. 254, no. 2, pp. 348–359, Oct. 2016, doi: 10.1016/j.ejor.2016.03.057.

[32] S. R. Barkunan, V. Bhanumathi, and J. Sethuram, "Smart sensor for automatic drip irrigation system for paddy cultivation," *Comput. Electr. Eng.*, vol. 73, pp. 180–193, Jan. 2019, doi: 10.1016/j.compeleceng.2018.11.013.

[33] D. Masseroni, P. Moller, R. Tyrell, M. Romani, A. Lasagna, G. Sali, A. Facchi, and C. Gandolfi, "Evaluating performances of the first automatic system for paddy irrigation in Europe," *Agricult. Water Manage.*, vol. 201, pp. 58–69, Mar. 2018, doi: 10.1016/j.agwat.2017.12.019.

[34] I. Han-Ya, K. Ishii, and N. Noguchi, "Monitoring rice growth environment by low-altitude remote sensing using spectroradiometer," *IFAC Proc. Volumes*, vol. 43, no. 26, pp. 184–189, 2010.

[35] N.-T. Son, C.-F. Chen, C.-R. Chen, and H.-Y. Guo, "Classification of multitemporal Sentinel-2 data for field-level monitoring of rice cropping practices in Taiwan," *Adv. Space Res.*, vol. 65, no. 8, pp. 1910–1921, Apr. 2020, doi: 10.1016/j.asr.2020.01.028.

[36] A. R. Arko, S. H. Khan, M. H. Biswas, A. Imran, A. H. Kafi, and R. S. I. Antara, "IOT based smart water and environment management system of paddy rice at different growth stages," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IoTaIS)*, Bali, Indonesia, Nov. 2019, pp. 154–160, doi: 10.1109/IoTaIS47347.2019.8980424.

[37] K. Jagan, M. Balasubramanian, and S. Palanivel, "Detection and recognition of diseases from paddy plant leaf images," *Int. J. Comput. Appl.*, vol. 144, no. 12, pp. 34–41, Jun. 2016.

[38] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?" *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 41–49, Sep. 2018, doi: 10.1109/MSP.2018.2825478.

[39] R. P. Narmadha and G. Arulvadivu, "Detection and measurement of paddy leaf disease symptoms using image processing," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Coimbatore, India, Jan. 2017, pp. 1–4, doi: 10.1109/ICCCI.2017.8117730.

[40] D. Lakmal, K. Kugathasan, V. Nanayakkara, S. Jayasena, A. S. Perera, and L. Fernando, "Brown planthopper damage detection using remote sensing and machine learning," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Boca Raton, FL, USA, Dec. 2019, pp. 97–104, doi: 10.1109/ICMLA.2019.00024.

[41] C. Harris and Y. A. Trisyono, "Classifying, detecting, and predicting infestation patterns of the brown planthopper in rice paddies," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Boca Raton, FL, USA, Dec. 2019, pp. 246–251, doi: 10.1109/ICMLA.2019.00046.

[42] N. C. Tri, H. N. Duong, T. Van Hoai, T. Van Hoa, V. H. Nguyen, N. T. Toan, and V. Snasel, "A novel approach based on deep learning techniques and UAVs to yield assessment of paddy fields," in *Proc. 9th Int. Conf. Knowl. Syst. Eng. (KSE)*, Hue, Vietnam, Oct. 2017, pp. 257–262, doi: 10.1109/KSE.2017.8119468.

[43] P. Li, C. Xiao, and Z. Feng, "Mapping rice planted area using a new normalized EVI and SAVI (NVI) derived from Landsat-8 OLI," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1822–1826, Dec. 2018, doi: 10.1109/LGRS.2018.2865516.

[44] Y.-A. Liou and H.-C. Sha, "Using MODIS imagery to estimate the damage of rainfed rice in Northeastern Thailand," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Munich, Germany, Jul. 2012, pp. 6601–6604, doi: 10.1109/IGARSS.2012.6352086.

[45] W. Pantanahiran, "The impact of climate change on food security and agricultural production in the Pak Phanang river basin, Thailand," in *Proc. 6th Int. Conf. Agro-Geoinformatics*, Fairfax, VA, USA, Aug. 2017, pp. 1–6, doi: 10.1109/Agro-Geoinformatics.2017.8047028.

[46] N. Nagoda and L. Ranathunga, "Rice sample segmentation and classification using image processing and support vector machine," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Rupnagar, India, Dec. 2018, pp. 179–184, doi: 10.1109/ICIINFS.2018.8721312.

[47] P. Wijerathna and L. Ranathunga, "Rice category identification using heuristic feature guided machine vision approach," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Rupnagar, India, Dec. 2018, pp. 185–190, doi: 10.1109/ICIINFS.2018.8721396.

[48] R. B. Guruprasad, K. Saurav, and S. Randhawa, "Machine learning methodologies for paddy yield estimation in India: A case study," in *Proc. IGARSS-IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, Jul. 2019, pp. 7254–7257, doi: 10.1109/IGARSS.2019.8900339.

[49] A. F. Bouwman, "Agronomic aspects of wetland rice cultivation and associated methane emissions," *Biogeochemistry*, vol. 15, no. 2, pp. 65–88, 1991, doi: 10.1007/BF00003218.

[50] Q. Jia, B. Lv, M. Guo, C. Luo, L. Zheng, T. Hsiang, and J. Huang, "Effect of rice growth stage, temperature, relative humidity and wetness duration on infection of rice panicles by Villosiclava virens," *Eur. J. Plant Pathol.*, vol. 141, no. 1, pp. 15–25, Jan. 2015, dopi: 10.1007/s10658-014-0516-4.

[51] M. A. K. Shafaullah, N. A. Khan, and Y. Mahmood, "Effect of epidemiological factors on the incidence of paddy blast (Pyricularia oryzae) disease," *Pakistan J. Phytopathol.*, vol. 23, no. 2, pp. 108–111, 2011.

[52] E. A. Azman, S. Jusop, C. F. Ishak, and R. Ismail, "Increasing rice production using different lime sources on an acid sulphate soil in Merbok, Malaysia," *Pertanika J. Trop. Agric. Sci.*, vol. 37, no. 2, pp. 223–247, 2014.

[53] N. A. Halim, R. Abdullah, S. Karsani, N. Osman, Q. Panhwar, and C. Ishak, "Influence of soil amendments on the growth and yield of rice in acidic soil," *Agronomy*, vol. 8, no. 9, p. 165, Aug. 2018, doi: 10.3390agronomy8090165.

[54] S. Aryal, "Rainfall and water requirement of rice during growing period," *J. Agricult. Environ.*, vol. 13, pp. 1–4, Feb. 2013, doi: 10.3126/aej.v13i0.7576.

[55] Lufft. *WS100 Radar Precipitation Sensor / Smart Disdrometer*. Accessed: May 10, 2020. [Online]. Available: https://www.lufft.com/products/precipitation-sensors-287/ws100-radar-precipitation-sensor-smart-disdrometer-2361/

[56] Landscape Technologies. *THE TDR-315H: World's First Fully Integrated TRUE TDR Digital Soil Moisture Sensor*. Accessed: May 10, 2020. [Online]. Available: https://www.landscapetechnologies.com.au/tdr-315-sensor-new?gclid=EAIaIQobChMI3Nuas7m36QIV2nwrCh05agTHEAAYASAAEgIap_D_BwE

[57] T. Talema and B. T. Hailu, "Mapping rice crop using sentinels (1 SAR and 2 MSI) images in tropical area: A case study in Fogera Wereda, Ethiopia," *Remote Sens. Appl., Soc. Environ.*, vol. 18, Apr. 2020, Art. no. 100290, doi: 10.1016/j.rsase.2020.100290.

[58] R. Sianturi, V. G. Jetten, and J. Sartohadi, "Mapping cropping patterns in irrigated rice fields in west java: Towards mapping vulnerability to flooding using time-series MODIS imageries," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 66, pp. 1–13, Apr. 2018, doi: 10.1016/j.jag.2017.10.013.

[59] A. Fatikhunnada, Liyantono, M. Solahudin, A. Buono, T. Kato, and K. B. Seminar, "Assessment of pre-treatment and classification methods for java paddy field cropping pattern detection on MODIS images," *Remote Sens. Appl., Soc. Environ.*, vol. 17, Jan. 2020, Art. no. 100281, doi: 10.1016/j.rsase.2019.100281.

[60] *MODIS Moderate Resolution Imaging Spectroradiometer*. Accessed: May 10, 2020. [Online]. Available: https://modis.gsfc.nasa.gov/about/specifications.php

[61] X. Xiao, S. Boles, S. Frolking, W. Salas, B. Moore, C. Li, L. He, and R. Zhao, "Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data," *Int. J. Remote Sens.*, vol. 23, no. 15, pp. 3009–3022, Jan. 2002.

[62] D. K. Sari, I. S. W. Ismullah, W. N. Sulasdi, and A. B. Harto, "Detecting rice phenology in paddy fields with complex cropping pattern using time series MODIS data," *J. Math. Fundam. Sci.*, vol. 42, no. 2, pp. 91–106, doi: 10.5614/itbj.sci.2010.42.2.2.

[63] I. H. Ikasari, V. Ayumi, M. I. Fanany, and S. Mulyono, "Multiple regularizations deep learning for paddy growth stages classification from LANDSAT-8," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Malang, Indonesia, Oct. 2016, pp. 512–517, doi: 10.1109/ICACSIS.2016.7872790.

[64] J. Dong, X. Xiao, M. A. Menarguez, G. Zhang, Y. Qin, D. Thau, C. Biradar, and B. Moore, "Mapping paddy rice planting area in Northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth engine," *Remote Sens. Environ.*, vol. 185, pp. 142–154, Nov. 2016, doi: 10.1016/j.rse.2016.02.016.

[65] *Agriculture and Rural Development: Earth Observation for Sustainable Development*. Accessed: May 10, 2020. [Online]. Available: https://www.eo4idi.eu/eo4sd-knowledge-portal/4-potential-uses-remote-sensing-smallholder-context/42-monitoring-crop-0

[66] N. N. C. Ya, L. S. Lee, M. R. Ismail, S. M. Razali, N. A. Roslin, and M. H. Omar, "Development of rice growth map using the advanced remote sensing techniques," in *Proc. Int. Conf. Comput. Drone Appl. (IConDA)*, Kuching, Malaysia, Dec. 2019, pp. 23–28, doi: 10.1109/IConDA47345.2019.9034916.

[67] J. Wang, K. Yu, M. Tian, and Z. Wang, "Estimation of rice key phenology date using Chinese HJ-1 vegetation index time-series images," in *Proc. 8th Int. Conf. Agro-Geoinformatics (Agro-Geoinformatics)*, Istanbul, Turkey, Jul. 2019, pp. 1–4, doi: 10.1109/Agro-Geoinformatics.2019.8820262.

[68] R. Suwantong, P. Srestasathiern, S. Lawawirojwong, and P. Rakwatin, "Moving horizon estimator with pre-estimation for crop start date estimation in tropical area," in *Proc. Amer. Control Conf. (ACC)*, Boston, MA, USA, Jul. 2016, pp. 3626–3631, doi: 10.1109/ACC.2016.7525476.

[69] L. Wang, Q. Chang, J. Yang, X. Zhang, and F. Li, "Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments," *PLoS ONE*, vol. 13, no. 12, Dec. 2018, Art. no. e0207624, doi: 10.1371/journal.pone.0207624.

[70] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, Jul. 2006, doi: 10.1080/01431160600589179.

[71] M. Shibayama, T. Sakamoto, E. Takada, A. Inoue, K. Morita, W. Takahashi, and A. Kimura, "Estimating paddy rice leaf area index with fixed point continuous observation of near infrared reflectance using a calibrated digital camera," *Plant Prod. Sci.*, vol. 14, no. 1, pp. 30–46, Jan. 2011.

[72] Y. Ge, Z. Liu, J. Chen, and T. Sun, "Estimation of paddy rice leaf area index using digital photography," in *Proc. 7th Int. Congr. Image Signal Process.*, Dalian, China, Oct. 2014, pp. 681–686, doi: 10.1109/CISP.2014.7003865.

[73] F.-M. Wang, J.-F. Huang, and Z.-H. Lou, "A comparison of three methods for estimating leaf area index of paddy rice from optimal hyperspectral bands," *Precis. Agricult.*, vol. 12, no. 3, pp. 439–447, Jun. 2011. [Online]. Available: https://doi-org.ezproxy.ums.edu.my/10.1007/s11119-010-9185-2

[74] F. Nutini, R. Confalonieri, A. Crema, E. Movedi, L. Paleari, D. Stavrakoudis, and M. Boschetti, "An operational workflow to assess rice nutritional status based on satellite imagery and smartphone apps," *Comput. Electron. Agricult.*, vol. 154, pp. 80–92, Nov. 2018, doi: 10.1016/j.compag.2018.08.008.

[75] S. M. Isa, S. Chandra, D. E. Herwindiati, and S. Mulyono, "Combining ground-based data and MODIS data for rice crop estimation in Indonesia," in *Proc. Int. Conf. Inf. Technol. Syst. Innov. (ICITSI)*, Bandung, Indonesia, Nov. 2015, pp. 1–5, doi: 10.1109/ICITSI.2015.7437710.

[76] Y. Duan, J. Zhong, G. Shuai, S. Zhu, and X. Gu, "Time-scale transferring deep convolutional neural network for mapping early rice," in *Proc. IGARSS-IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 1136–1139.

[77] C. A. Devia, J. P. Rojas, E. Petro, C. Martinez, I. F. Mondragon, D. Patino, M. C. Rebolledo, and J. Colorado, "High-throughput biomass estimation in rice crops using UAV multispectral imagery," *J. Intell. Robot. Syst.*, vol. 96, nos. 3–4, pp. 573–589, Dec. 2019. [Online]. Available: https://doi-org.ezproxy.ums.edu.my/10.1007/s10846-019-01001-5

[78] B. Wang, X. Zhang, Y. Dong, J. Zhang, J. Zhang, and X. Zhou, "Retrieval of leaf chlorophyll content of paddy rice with extracted foliar hyperspectral imagery," in *Proc. 7th Int. Conf. Agro-geoinformatics (Agro-geoinformatics)*, Hangzhou, China, Aug. 2018, pp. 1–5, doi: 10.1109/Agro-Geoinformatics.2018.8476024.

[79] S.-Y. Zhang, T. Fei, and Y.-H. Ran, "Diagnosis of heavy metal cross contamination in leaf of rice based on hyperspectral image: A greenhouse experiment," in *Proc. IEEE Int. Conf. Adv. Manuf. (ICAM)*, Yunlin, Taiwan, Nov. 2018, pp. 159–162, doi: 10.1109/AMCON.2018.8614938.

[80] N. Wu, H. Jiang, Y. Bao, C. Zhang, J. Zhang, W. Song, Y. Zhao, C. Mi, Y. He, and F. Liu, "Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions," *Sens. Actuators B, Chem.*, vol. 308, Apr. 2020, Art. no. 127696. [Online]. Available: https://d10.1016/j.snb.2020.127696oi.org/

[81] L. Zhang, H. Sun, Z. Rao, and H. Ji, "Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 229, Mar. 2020, Art. no. 117973, doi: 10.1016/j.saa.2019.117973.

[82] K. Lasko, K. P. Vadrevu, V. T. Tran, and C. Justice, "Mapping double and single crop paddy rice with sentinel-1A at varying spatial scales and polarizations in Hanoi, Vietnam," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 2, pp. 498–512, Feb. 2018, doi: 10.1109/JSTARS.2017.2784784.

[83] F. Wu, C. Wang, H. Zhang, B. Zhang, and Y. Tang, "Rice crop monitoring in south China with RADARSAT-2 quad-polarization SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 196–200, Mar. 2011, doi: 10.1109/LGRS.2010.2055830.

[84] S. P. Adhikari, G. Kim, and H. Kim, "Deep neural network-based system for autonomous navigation in paddy field," *IEEE Access*, vol. 8, pp. 71272–71278, 2020, doi: 10.1109/ACCESS.2020.2987642.

[85] J. R. Balbin, J. B. G. Ibarra, A. A. A. Quiniquini, R. G. J. Locsin, G. G. Tindogan, and H. V. L. Tee, "Development of scientific farming in determining the harvest time of NSIC 2010 Rc226 rice grain using image processing algorithms," in *Proc. IEEE 9th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ. Manage. (HNICEM)*, Manila, Philippines, Dec. 2017, pp. 1–5, doi: 10.1109/HNICEM.2017.8269452.

[86] Q. Yang, L. Shi, J. Han, J. Yu, and K. Huang, "A near real-time deep learning approach for detecting rice phenology based on UAV images," *Agricult. Forest Meteorol.*, vol. 287, Jun. 2020, Art. no. 107938, doi: 10.1016/j.agrformet.2020.107938.

[87] B. S. Anami, N. N. Malvade, and S. Palaiah, "Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images," *Artif. Intell. Agricult.*, vol. 4, pp. 12–20, 2020, doi: 10.1016/j.aiia.2020.03.001.

[88] M. H. Kamrul, P. Paul, and M. Rahman, "Machine vision based rice disease recognition by deep learning," in *Proc. 22nd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dhaka, Bangladesh, Dec. 2019, pp. 1–6, doi: 10.1109/ICCIT48885.2019.9038350.

[89] E. Agustina, I. Pratomo, A. D. Wibawa, and S. Rahayu, "Expert system for diagnosis pests and diseases of the rice plant using forward chaining and certainty factor method," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Surabaya, Indonesia, Aug. 2017, pp. 266–270, doi: 10.1109/ISITIA.2017.8124092.

[90] O. Barrero, D. Rojas, C. Gonzalez, and S. Perdomo, "Weed detection in rice fields using aerial images and neural networks," in *Proc. 21th Symp. Signal Process., Images Artif. Vis. (STSIVA)*, Bucaramanga, Colombia, Aug. 2016, pp. 1–4, doi: 10.1109/STSIVA.2016.7743317.

[91] Q. Yao, D.-X. Xian, Q.-J. Liu, B.-J. Yang, G.-Q. Diao, and J. Tang, "Automated counting of rice planthoppers in paddy fields based on image processing," *J. Integrative Agricult.*, vol. 13, no. 8, pp. 1736–1745, 2014, doi: 10.1016/S2095-3119(14)60799-1.

[92] N. A. N. Shah, M. K. Osman, N. A. Othman, F. Ahmad, and A. R. Ahmad, "Identification and counting of brown planthopper in paddy field using image processing techniques," *Procedia Comput. Sci.*, vol. 163, pp. 580–590, 2019, doi: 10.1016/j.procs.2019.12.140.

[93] C. Liu, C. Xu, S. Liu, D. Xu, and X. Yu, "Study on identification of rice false smut based on CNN in natural environment," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Shanghai, China, Oct. 2017, pp. 1–5, doi: 10.1109/CISP-BMEI.2017.8302016.

[94] Y. Chen, Y. Wu, J. Cheng, and D. Tao, "A deep multi-view learning method for rice grading," in *Proc. IEEE Int. Conf. Real-time Comput. Robot. (RCAR)*, Irkutsk, Russia, Aug. 2019, pp. 726–730, doi: 10.1109/RCAR47638.2019.9044007.

[95] Y. Wu, Z. Yang, W. Wu, X. Li, and D. Tao, "Deep-rice: Deep multi-sensor image recognition for grading rice," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Wuyishan, China, Aug. 2018, pp. 116–120, doi: 10.1109/ICInfA.2018.8812590.

[96] B. S. Anami, N. N. Malvade, and S. Palaiah, "Automated recognition and classification of adulteration levels from bulk paddy grain samples," *Inf. Process. Agricult.*, vol. 6, no. 1, pp. 47–60, Mar. 2019, doi: 10.1016/j.inpa.2018.09.001.

[97] S. Fayyazi, M. H. Abbaspour-Fard, A. Rohani, S. A. Monadjemi, and H. Sadrnia, "Identification and classification of three Iranian rice varieties in mixed bulks using image processing and MLP neural network," *Int. J. Food Eng.*, vol. 13, no. 5, May 2017, doi: 10.1515/ijfe-2016-0121.

[98] Y. Zhang, J. Gao, H. Cen, Y. Lu, X. Yu, Y. He, and J. G. Pieters, "Automated spectral feature extraction from hyperspectral images to differentiate weedy rice and barnyard grass from a rice crop," *Comput. Electron. Agricult.*, vol. 159, pp. 42–49, Apr. 2019, doi: 10.1016/j.compag.2019.02.018.

[99] X. Bai, Z. Cao, L. Zhao, J. Zhang, C. Lv, C. Li, and J. Xie, "Rice heading stage automatic observation by multi-classifier cascade based rice spike detection method," *Agricult. Forest Meteorol.*, vol. 259, pp. 260–270, Sep. 2018, doi: 10.1016/j.agrformet.2018.05.001.

[100] S. Chen, J. Xiong, W. Guo, R. Bu, Z. Zheng, Y. Chen, Z. Yang, and R. Lin, "Colored rice quality inspection system using machine vision," *J. Cereal Sci.*, vol. 88, pp. 87–95, Jul. 2019, doi: 10.1016/j.jcs.2019.05.010.

[101] M. Izquierdo, M. Lastra-Mejías, E. González-Flores, S. Pradana-López, J. C. Cancilla, and J. S. Torrecilla, "Visible imaging to convolutionally discern and authenticate varieties of rice and their derived flours," *Food Control*, vol. 110, Apr. 2020, Art. no. 106971, doi: 10.1016/j.foodcont.2019.106971.

[102] J. D. Bambole and K. A. Ghodinde, "Evaluation of chalkiness in basmati rice by virtual instrumentation," in *Proc. Int. Conf. Smart Technol. Manage. Comput., Commun., Controls, Energy Mater. (ICSTM)*, Chennai, India, May 2015, pp. 604–608, doi: 10.1109/ICSTM.2015.7225485.

[103] D. Vishnu, G. Mukherjee, and A. Chatterjee, "A computer vision approach for grade identification of rice bran," in *Proc. 3rd Int. Conf. Res. Comput. Intell. Commun. Netw. (ICRCICN)*, Kolkata, India, Nov. 2017, pp. 10–14, doi: 10.1109/ICRCICN.2017.8234473.

[104] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Nitrogen deficiency prediction of rice crop based on convolutional neural network," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 11, pp. 5703–5711, Nov. 2020. [Online]. Available: https://doi-org.ezproxy.ums.edu.my/10.1007/s12652-020-01938-8
[105] J. Padarian, B. Minasny, and A. B. McBratney, "Using deep learning to predict soil properties from regional spectral data," *Geoderma Regional*, vol. 16, Mar. 2019, Art. no. e00198, doi: 10.1016/j.geodrs.2018.e00198.
[106] N. M. Tiglao, M. Alipio, J. V. Balanay, E. Saldivar, and J. L. Tiston, "Agrinex: A low-cost wireless mesh-based smart irrigation system," *Measurement*, vol. 161, Sep. 2020, Art. no. 107874, doi: 10.1016/j.measurement.2020.107874.
[107] R. H. Faisal, C. Saha, M. H. Hasan, and P. Kumar Kundu, "Power efficient distant controlled smart irrigation system for AMAN and BORO rice," in *Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dhaka, Bangladesh, Dec. 2018, pp. 1–5, doi: 10.1109/ICCITECHN.2018.8631927.
[108] B.-L. Li, C.-P. Ti, and X.-Y. Yan, "Estimating rice paddy areas in China using multi-temporal cloud-free NDVI imagery based on change detection," *Pedosphere*, vol. 30, no. 6, pp. 734–746, 2020, doi: 10.1016/S1002-0160(17)60405-3.
[109] M. Zhang, H. Lin, G. Wang, H. Sun, and J. Fu, "Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting lake area, China," *Remote Sens.*, vol. 10, no. 11, p. 1840, Nov. 2018, doi: 10.3390/rs10111840.
[110] N. Gandhi, O. Petkar, and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks," in *Proc. IEEE Technolog. Innov. ICT Agricult. Rural Develop. (TIAR)*, Chennai, India, Jul. 2016, pp. 105–110, doi: 10.1109/TIAR.2016.7801222.
[111] L. Zhang, S. Traore, J. Ge, Y. Li, S. Wang, G. Zhu, Y. Cui, and G. Fipps, "Using boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in sahel," *Comput. Electron. Agricult.*, vol. 166, Nov. 2019, Art. no. 105031, doi: 10.1016/j.compag.2019.105031.
[112] Z. Chu and J. Yu, "An end-to-end model for rice yield prediction using deep learning fusion," *Comput. Electron. Agricult.*, vol. 174, Jul. 2020, Art. no. 105471, doi: 10.1016/j.compag.2020.105471.
[113] R. F. Rahmat, T. Z. Lini, Pujiarti, and A. Hizriadi, "Implementation of real-time monitoring on agricultural land of rice plants using smart sensor," in *Proc. 3rd Int. Conf. Electr., Telecommun. Comput. Eng. (ELTICOM)*, Medan, Indonesia, Sep. 2019, pp. 40–43, doi: 10.1109/ELTICOM47379.2019.8943912.
[114] S. Alifah, G. Gunawan, and M. Taufik, "Smart monitoring of rice l gistic employing Internet of Things network," in *Proc. 2nd Borneo Int. Conf. Appl. Math. Eng. (BICAME)*, Balikpapan, Indonesia, Dec. 2018, pp. 199–202, doi: 10.1109/BICAME45512.2018.1570509318.
[115] Y. Cai, H. Lin, and M. Zhang, "Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data," *Adv. Space Res.*, vol. 64, no. 11, pp. 2233–2244, Dec. 2019, doi: 10.1016/j.asr.2019.08.042.
[116] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42–62, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425718300154, doi: 10.1016/j.rse.2018.02.009.
[117] A. A. Joshi and B. D. Jadhav, "Monitoring and controlling rice diseases using image processing techniques," in *Proc. Int. Conf. Comput., Analytics Secur. Trends (CAST)*, Dec. 2016, pp. 471–476.
[118] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, Dec. 2017, doi: 10.1016/j.neucom.2017.06.023.
[119] S. Mittal, M. K. Dutta, and A. Issac, "Non-destructive image processing based system for assessment of rice quality and defects for classification according to inferred commercial value," *Measurement*, vol. 148, Dec. 2019, Art. no. 106969, doi: 10.1016/j.measurement.2019.106969.
[120] S. Watcharabutsarakham, I. Methasate, N. Watcharapinchai, W. Sinthupinyo, and W. Sriratanasak, "An approach for density monitoring of brown planthopper population in simulated paddy fields," in *Proc. 13th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Khon Kaen, Thailand, Jul. 2016, pp. 1–4, doi: 10.1109/JCSSE.2016.7748922.
[121] W. Jearanaiwongkul, C. Anutariya, and F. Andres, "A semantic-based framework for rice plant disease management," *New Gener. Comput.*, vol. 37, no. 4, pp. 499–523, Dec. 2019. [Online]. Available: https://doi-org.ezproxy.ums.edu.my/10.1007/s00354-019-00072-0
[122] M. G. Jayanthi and D. R. Shashikumar, "A model for early detection of paddy leaf disease using optimized fuzzy inference system," in *Proc. Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Tirunelveli, India, Nov. 2019, pp. 206–211, doi: 10.1109/ICSSIT46314.2019.8987955.
[123] D. Ngampak and P. Piamsa-Nga, "Image analysis of broken rice grains of Khao Dawk Mali rice," in *Proc. 7th Int. Conf. Knowl. Smart Technol. (KST)*, Chonburi, Thailand, Jan. 2015, pp. 115–120, doi: 10.1109/KST.2015.7051471.
[124] W. Zhong, L. Ma, and Y. Yin, "Research of rice appearance quality recognition," in *Proc. 7th Int. Conf. Adv. Commun. Netw. (ACN)*, Kota Kinabalu, Malaysia, Jul. 2015, pp. 56–59, doi: 10.1109/ACN.2015.24.
[125] S. K. Singh, S. K. Vidyarthi, and R. Tiwari, "Machine learnt image processing to predict weight and size of rice kernels," *J. Food Eng.*, vol. 274, Jun. 2020, Art. no. 109828, doi: 10.1016/j.jfoodeng.2019.109828.
[126] D. K. Lim, N. P. Long, C. Mo, Z. Dong, L. Cui, G. Kim, and S. W. Kwon, "Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice," *Food Res. Int.*, vol. 100, pp. 814–821, Oct. 2017, doi: 10.1016/j.foodres.2017.08.006.
[127] R. M. Barbosa, E. S. D. Paula, A. C. Paulelli, A. F. Moore, J. M. O. Souza, B. L. Batista, A. D. Campiglia, and F. Barbosa, "Recognition of organic rice samples based on trace elements and support vector machines," *J. Food Composition Anal.*, vol. 45, pp. 95–100, Feb. 2016, doi: 10.1016/j.jfca.2015.09.010.
[128] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.
[129] G. Ji-Xi, D. Fei-Zhou, and X. Bao, "The application of principal component analysis to agriculture soil contamination assessment," *Geographical Res.*, vol. 25, no. 5, pp. 836–842, 2006.
[130] J. Ma, "Principal component analysis method-based research on agricultural science and technology website evaluation," in *Computer and Computing Technologies in Agriculture IX* (IFIP Advances in Information and Communication Technology), vol. 479, D. Li and Z. Li, Eds. Cham, Switzerland: Springer, 2016.
[131] F. Muema, P. Home, and J. Raude, "Application of benchmarking and principal component analysis in measuring performance of public irrigation schemes in Kenya," *Agriculture*, vol. 8, no. 10, p. 162, Oct. 2018.
[132] I. T. Jolliffe, "A note on the use of principal components in regression," *J. Roy. Stat. Soc. C*, vol. 31, no. 3, pp. 300–303, 1982, doi: 10.2307/2348005.
[133] J. B. Rawlings, "Moving horizon estimation," in *Encyclopedia of Systems and Control*, J. Baillieul and T. Samad, Eds. London, U.K.: Springer, 2013.
[134] R. Suwantong, S. Bertrand, D. Dumur, and D. Beauvois, "Stability of a nonlinear moving horizon estimator with pre-estimation," in *Proc. Amer. Control Conf.*, Jun. 2014, pp. 5688–5693.
[135] R. Suwantong, "Development of the moving horizon estimator with pre-estimation (MHE-PE). application to space debris tracking during the re-entries," in *Proc. Supélec*, 2014.
[136] S. C. Peter, J. K. Dhanjal, V. Malik, N. Radhakrishnan, M. Jayakanthan, and D. Sundar, "Quantitative structure-activity relationship (QSAR): Modeling approaches to biological applications," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. New York, NY, USA: Academic, 2019, pp. 661–676, doi: 10.1016/B978-0-12-809633-8.20197-0.
[137] A. G. Fragkaki, E. Farmaki, N. Thomaidis, A. Tsantili-Kakoulidou, Y. S. Angelis, M. Koupparis, and C. Georgakopoulos, "Comparison of multiple linear regression, partial least squares and artificial neural networks for prediction of gas chromatographic relative retention times of trimethylsilylated anabolic androgenic steroids," *J. Chromatography A*, vol. 1256, pp. 232–239, Sep. 2012, doi: 10.1016/j.chroma.2012.07.064.
[138] Y. Jiang, Y. Huang, W. Xue, and H. Fang, "On designing consistent extended Kalman filter," *J. Syst. Sci. Complex.*, vol. 30, no. 4, pp. 751–764, Aug. 2017, doi: 10.1007/s11424-017-5151-7.
[139] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to non linear systems," in *Proc. Aero Sense, 11th Int. Symp. Aerosp./Defence Sens. Simulation Controls*, 1997.
[140] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999, doi: 10.1023/A:1018628609742.
[141] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines*. Berkeley, CA, USA: Apress, 2015.

[142] R. Wang, "AdaBoost for feature selection, classification and its relation with SVM, a review," *Phys. Procedia*, vol. 25, pp. 800–807, 2012, doi: 10.1016/j.phpro.2012.03.160.

[143] L. Breiman, "Random forests," *Mach. Learn*. vol. 45, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[144] L. Breiman and J. C. Friedman, *Stone Classification and Regression Trees* (Wadsworth Statistics/Probability). New York, NY, USA: Chapman & Hall/CRC, 1984.

[145] Y. Akbulut, A. Sengur, Y. Guo, and F. Smarandache, "NS-K-NN: Neutrosophic set-based K-nearest neighbors classifier," *Symmetry*, vol. 9, no. 9, p. 179, Sep. 2017, doi: 10.3390/sym9090179.

[146] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[147] M. Rudrapatna and A. Sowmya, "Feature weighted minimum distance classifier with multi-class confidence estimation," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 4304, A. Sattar and B. Kang, Eds. Berlin, Germany: Springer, 2006.

[148] X. Yang, X. Shen, J. Long, and H. Chen, "An improved median-based Otsu image thresholding algorithm," *AASRI Procedia*, vol. 3, pp. 468–473, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212671612002338, doi: 10.1016/j.aasri.2012.11.074.

[149] A. Amo, J. Montero, G. Biging, and V. Cutello, "Fuzzy classification systems," *European J. Oper. Res.*, vol. 156, no. 2, pp. 495–507, 2004, doi: 10.1016/S0377-2217(03)00002-X.

[150] S. Lek and Y. S. Park, "Artificial neural networks," in *Encyclopedia of Ecology*, S. E. Jorgensen and B. D. Fath, Eds. New York, NY, USA: Academic, 2008, pp. 237–245, doi: 10.1016/B978-008045405-4.00173-7.

[151] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," 2017, *arXiv:1702.01923*. [Online]. Available: http://arxiv.org/abs/1702.01923

[152] B. HH, "Clustering methods: A history of K-means algorithms," in *Selected Contributions in Data Analysis and Classification* (Studies in Classification, Data Analysis, and Knowledge Organization), P. Brito, G. Cucumel, P. Bertrand, and F. D. Carvalho, Eds. Berlin, Germany: Springer, 2007.

[153] M. K. Dhodhi, J. A. Saghri, I. Ahmad, and R. Ul-Mustafa, "D-ISODATA: A distributed algorithm for unsupervised classification of remotely sensed data on network of workstations," *J. Parallel Distrib. Comput.*, vol. 59, no. 2, pp. 280–301, Nov. 1999, doi: 10.1006/jpdc.1999.1573.

[154] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.

[155] M. R. S. Muthusinghe, S. T. Palliyaguru, W. A. N. D. Weerakkody, A. M. H. Saranga, and W. H. Rankothge, "Towards smart farming: Accurate prediction of paddy harvest and rice demand," in *Proc. IEEE Region 10th Humanitarian Technol. Conf. (R10-HTC)*, Malambe, Sri Lanka, Dec. 2018, pp. 1–6, doi: 10.1109/R10-HTC.2018.8629843.

[156] K. Neumann, P. H. Verburg, E. Stehfest, and C. Müller, "The yield gap of global grain production: A spatial analysis," *Agricult. Syst.*, vol. 103, no. 5, pp. 316–326, Jun. 2010.

[157] F. T. Pinki, N. Khatun, and S. M. M. Islam, "Content based paddy leaf disease recognition and remedy prediction using support vector machine," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–5.

[158] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, pp. 1–10, Sep. 2016.

[159] D. F. Nettleton, D. Katsantonis, A. Kalaitzidis, N. Sarafijanovic-Djukic, P. Puigdollers, and R. Confalonieri, "Predicting rice blast disease: Machine learning versus process-based models," *BMC Bioinf.*, vol. 20, no. 1, p. 514, Dec. 2019, doi: 10.1186/s12859-019-3065-1.

[160] D. Katsantonis, K. Kadoglidou, C. Dramalis, and P. Puigdollers, "Rice blast forecasting models and their practical value: A review," *Phytopathologia Mediterranea*, vol. 56, no. 2, pp. 187–216, 2017, doi: 10.14601/Phytopathol_Mediterr-18706.

[161] A. Tugi, A. W. Rasib, M. A. Suri, O. Zainon, A. R. M. Yusoff, M. Z. A. Rahman, N. A. Sari, and N. Darwin, "Oil palm tree growth monitoring for smallholders by using unmanned aerial vehicle," *Jurnal Teknologi*, vol. 77, no. 26, pp. 1–11, Dec. 2015, doi: 10.11113/jt.v77.6855.

[162] M. S. Sainin and R. Alfred, "A genetic based wrapper feature selection approach using nearest neighbour distance matrix," in *Proc. 3rd Conf. Data Mining Optim. (DMO)*, Putrajaya, Malaysia, Jun. 2011, pp. 237–242, doi: 10.1109/DMO.2011.5976534.

[163] R. Alfred, "Feature transformation: A genetic-based feature construction method for data summarization," *Comput. Intell.*, vol. 26, no. 3, pp. 337–357, Jul. 2010, doi: 10.1111/j.1467-8640.2010.00362.x.

[164] R. Alfred, "Optimizing feature construction process for dynamic aggregation of relational attributes," *J. Comput. Sci.*, vol. 5, no. 11, pp. 864–977, 2009.

RAYNER ALFRED received the Computer Science degree from the Polytechnic University of Brooklyn, New York, USA, the master's degree in computer science from Western Michigan University, Kalamazoo, USA, and the Ph.D. degree in computer science from York University, U.K., in 2008. His Ph.D. degree focuses on intelligent techniques using machine learning to model and optimize the dynamic and distributed processes of knowledge discovery for structured and unstructured data.

He is currently an Associate Professor in computer science with the Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia, that focuses on data science and software engineering programmes. He leads and defines projects around knowledge discovery, information retrieval, and machine learning that focuses on building smarter mechanism that enables knowledge discovery in structured and unstructured data. His work addresses the challenges related to big data problem: How can we create and apply smarter collaborative knowledge discovery and machine learning technologies that bridge the structured and unstructured data mining and cope with the big data problem? He has authored or coauthored more than 150 journals/book chapters and conference papers, editorials, and served on the program and organizing committees of numerous national and international conferences and workshops.

Dr. Rayner is currently a Certified Software Tester (CTFL) from the International Software Testing Qualifications Board (ISTQB) and also a certified IBM DB2 Academic Associate (IBM DB2 AA). He leads the Advanced Machine Intelligence (AMI) research group, UMS, and he has lead several projects related to knowledge discovery and machine learning on big data. He was a recipient of the Research Fellow of the Japan Advanced Institute of Science and Technology (JAIST), Japan. He was also a recipient of multiple Gold and Silver awards at national and international research exhibitions in data mining and machine learning-based solutions (face recognition and knowledge discovery), that include the International Trade Fair Ideas (iENA 2018), Nuremberg, Germany; the International Invention Innovation Competition (iCAN 2018), Toronto, Canada; and Seoul International Invention Exhibition (SIIF 2010), Seoul, South Korea. He has secured RM6,931.433.00 worth of project grants. He was also a recipient of the Myron M. Rosenthal Academic Achievement Award for the outstanding academic achievement in computer science, in 1994.

JOE HENRY OBIT received the Ph.D. degree in computer science from the School of Computer Science, University of Nottingham. His Ph.D. thesis was developing a novel meta-heuristic, hyper-heuristic, and cooperative search. He is currently an Associate Professor of computer science with the Department of Data Science, Universiti Malaysia Sabah. His main research interests include interface of operational research and computer science. In particular, the exploration and development of innovative operational research, artificial intelligence, and distributed artificial intelligence models and methodologies for automatically producing high quality solutions to a wide range of real world combinatorial optimization and scheduling problems.

CHRISTIE PEI-YEE CHIN received the Ph.D. degree in information systems from the University of South Australia. She is currently a Senior Lecturer with the Faculty of Computing and Informatics, Universiti Malaysia Sabah. Her research interests include enterprise social networks, public social networks, social analytics, knowledge management, and technology adoption and usage. Her work was published in various international journals, such as *Information Systems Frontiers* and *Journal of Organizational Computing and Electronic Commerce*, as well as proceedings of conferences, such as AMCIS and PACIS.

HAVILUDDIN HAVILUDDIN (Member, IEEE) was born in Loa Tebu, East Kalimantan, Indonesia. He graduated in the field of management information from STMIK WCD Samarinda, the master's degree in the field of computer science from Universitas Gadjah Mada, Yogyakarta, and the Ph.D. degree in computer science from the Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia. He is currently a member of the International Association of Computer Science and Information Technology (IACSIT), Institute of Advanced Engineering and Science (IAES), and Association of Computing and Informatics Institutions Indonesia (APTIKOM) societies.

YUTO LIM (Member, IEEE) received the B.Eng. (Hons.) and M.Inf.Tech. degrees from Universiti Malaysia Sarawak (UNIMAS), Malaysia, in 1998 and 2000, respectively, and the Ph.D. degree in communications and computer engineering from Kyoto University, in 2005. From November 2005 to September 2009, he was an Expert Researcher at the National Institute of Information and Communications Technology (NICT), Japan. In 2005, he was actively joining the standardization activities of IEEE 802.11s Mesh Networking. He and his team members have introduced two proposals, which currently adopted in the draft of IEEE 802.11s D1.04. He also led the RA-OLSR group in resolving a part of the comments. In 2006, he is also actively joining the standardization activities of next-generation home networks at Telecommunication Technology Committee (TTC), Japan. Since October 2009, he has been working at the Japan Advanced Institute of Science and Technology (JAIST), as an Associate Professor. From 2017 to 2019, he was actively joining the standardization activities of smart cities from the Bridging the Standardization Gap (BSG) Working Group of the TTC, Japan. He is honorably appointed as a TTC representative to report the ''Smart City'' activities in Asia-Pacific Region in both APT Standardization Program Forum (ASTAP) and APT Telecommunication/ICT Development Forum (ADF). His research interests include smart homes, smart cities, cyber-physical systems, the Internet of Things, future wireless communication and networks, and smart energy distribution. He is a member of IEICE and IPSJ.