

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Big Data and AI Revolution in Precision Agriculture: Survey and Challenges

SHOWKAT AHMAD BHAT¹, NEN-FU HUANG¹

¹Institute of Communication Engineering, Department of Electrical Engineering and Computer Science, National Tsing Hua University, 30013, Taiwan;

¹Department of Computer Science, National Tsing Hua University, Hsinchu, 30013, Taiwan;

Corresponding author: Showkat Ahmad Bhat (e-mail: showkatbhat1994@gmail.com).

This study is supported by the Ministry of Science and Technology (MOST) of Taiwan under the grant “MOST 109-2622-8-007-019-TE2”.

ABSTRACT Sustainable agricultural development is a significant solution with fast population development through the use of information and communication (ICT) in precision agriculture, which produced new methods for making cultivation further productive, proficient, well-regulated while preserving the climate. Big data (machine learning, deep learning, etc.) is amongst the vital technologies of ICT employed in precision agriculture for their huge data analytical capabilities to abstract significant information and to assist agricultural practitioners to comprehend well farming practices and take precise decisions. The main goal of this article is to acquire an awareness of the Big Data latest applications in smart agriculture and be acquainted with related social and financial challenges to be concentrated on. This article features data creation methods, accessibility of technology, accessibility of devices, software tools, and data analytic methods, and appropriate applications of big data in precision agriculture. Besides, there are still a few challenges that come across the widespread implementation of big data technology in agriculture.

INDEX TERMS precision agriculture, big data analytics, machine learning, sustainable agriculture, smart farming, digital agriculture

I. INTRODUCTION

The total populace, as revealed in November 2020 is 7.8 billion according to United Nations estimates. It is assessed that this number will project to 8.5 billion by 2030 and 9.9 billion by 2050. With the rapid growth in the total populace, food consumption is also growing rapidly worldwide. Agriculture is already producing about 17% more yield than it used to produce just three decades ago. However, about 821 million people in the world suffer from a lack of food security. Increasing agriculture or food production rapidly for meeting the growing food supply demands is not an easy task. Several factors contribute to this problem, such as decade-old agriculture practices, poor storage, marketplaces, and political upheaval. As the global population is growing higher, the food and agriculture organizations calculate that agriculture production will need to increase by 70% by 2050 to feed the growing population of the world. It is not just about feeding people; we need to provide them with highly nutritious food without harming the environment. Since the volume of arable land is not increasing, the groundwater levels are going down, and the

soil quality is not increasing, so we need to increase agriculture production responsibly.

To deliver sustainable agriculture production, the agriculture sector needs to employ cutting-edge technologies like blockchain [1] [2], IoT [3], and AI [4]. Data-driven agriculture with these technologies is the most promising approach to solve these problems. If we could generate a huge quantity of data from the farm and use that data to drive some of the agricultural decisions. It can help to solve some of these food problems globally. For instance, if we could enable farms to build data sets or maps for soil moisture, temperature and humidity in the area, availability of water, and other environmental factors around the farm, it would enable techniques like smart farming, precision agriculture [4], vertical farming, etc. Data-driven agriculture has been shown to improve crop yield, reduce cost, and ensure sustainability [5]. These are not limited to agriculture but have potential solutions for several challenges faced by livestock farming also. Digitanimal is a company to enhance livestock farm productivity, sustainability, and animal welfare along with providing thorough monitoring solutions based on IoT

wearable's powered by firmware, AI, satellite images, and blockchain technology provides farmers with relevant information on health, location, feeding, and reproduction conditions of their animals. Thus by incorporating technological-based agriculture techniques increases their yield, reduces costs, and improves the income of farmers, thereby improves their quality of life. Smart agriculture deems it essential to address these problems that have attracted a lot of technological attention, from sowing to watering of crops to health, harvesting, and traceability in supply chain management. Figure 1 shows an overview of these technologies concerning the smart agriculture ecosystem.

Big data empowers agricultural practitioners and related industries to gain information about different factors that influence agricultural production and take efficient decisions in daily farming. It keeps them up to date about the market price, demand of a particular crop, and the new technologies in the agriculture sector. Recently, big factory farms have embraced different technologies like IoT and blockchain with an intent to produce greater production in the farming practice. Blockchain technology is being implemented in the management of the agri-food supply chain to make available features such as transparency, security, immutability, and reliability of all operations. Blockchain also assists in addressing several IoT security and reliability challenges.

IoT assists in data collection at each stage of agriculture production and supply chain [6]. Therefore, it would likewise be valuable to perform big data analytics on the data collected during farming, processing, logistics, and marketing. Information driven agrarian sector would significantly revolutionize farming and customer behavior. For example, mobile agriculture expert systems and agriculture predictive analytics all depend on big data to provide intelligent recommendations to growers towards precision agriculture. Precise risk valuation could assist agriculture practitioners to better handle agriculture risks concerning production, market, institutional risk, accompanying individual and monetary risks. Furthermore, big data can be utilized to address several challenges like food safety, supply management, food security, along with food loss and wastage.

Similar to other sectors, the agricultural industry has pursued innovations by employing convergence technologies. Big data and AI have demonstrated their potential and usage throughout the industry. However, agricultural big data faces various limitations as discussed in Section V of this paper. These limitations must be addressed to offer the right methodology of agricultural solutions for the next generation. Therefore, a strategy and a scheme for future advancement should be recognized by reviewing the latest developments, trends, and potential of technological innovations. In this article, we provide a comprehensive analysis to provide intuitions into important research works in smart agriculture employing big data and AI with a emphasis on precision farming. The prospective for

exploiting big data and AI in processing actual field data for soil mapping, crop monitoring and estimation, sustainable resource utilization, disease and weed detection, weather monitoring, etc. in agriculture is extremely promising. Further, this paper provides a discussion on various available software tools for agriculture big data analysis and provides a discussion and comparison of different machine learning techniques in agriculture. Furthermore, we provide the complete discussion on data operational cycle right from data acquisition to decision making and implementation (actuation). We have also presented a in depth discussion on several potential applications and limitations of big data in precision agriculture.

The remainder of the review article is systematized as follows. In section II, we present the overview of big data and AI technologies in precision agriculture. In section III, we present comparison and discussion on machine learning techniques in agriculture. In section VI, we described the big data operating cycle in agriculture environment. In section V, we report potential applications. Then, we deliberated some open issues and research areas in section VI. Lastly, section VII presents the conclusion of the paper.

TABLE I. DIGITAL LIBRARIES AND WEB SOURCES

NAME	DIGITAL LIBRARY	WEB SOURCE
IEEE Xplore Digital Library	✓	
ACM Digital Library	✓	
ScienceDirect (Elsevier)	✓	
Springer	✓	
Google Scholar		✓
Web of Science		✓
Research Gate		✓

A. RESEARCH METHODOLOGY

The primary purpose of this systematic review article is to identify pertinent research in the field of study. The research methodology process used in this systematic literature review consists of planning, implementing, and result investigation. The initial stage involves formulation the review, recognizing its requirements, and outlining its rules including a) research questions, b) paper extraction, c) and selection of relevant papers for review. The second stage comprises extracting the relevant information from the selected papers. Lastly, present discussion, conclusion, and future work. We have identified some well-known digital libraries and web sources that we used to extract the relevant works mostly for 2000-2020. Table I presents the digital libraries and web sources. Table II contains some keywords and some concepts related to our field of study and used certain connectors to build search strings. On the subject of article selection, we start with

papers that include keywords related to agriculture (such as agriculture, farming, smart agriculture, agri-chain, food-chain, etc.) and others related to big data and AI. Then we rejected papers that are not relevant to the agriculture sector directly. Finally, we filter the papers for duplicated from different sources, other than the English language and student thesis. Thus, we obtained a total of 77 studies that are relevant to the research goal of this review article.

TABLE II. KEYWORDS UTILIZED TO EXTRACT RELEVANT PAPERS DURING THE REVIEW

AREA	KEYWORD	CORRELATED CONCEPTS
Agriculture	Agriculture, Farming, Farms	Smart-agriculture, e-agriculture, Smart-farming, Green revolution
Big data	Big data Massive data, Data mining	Big data, Data mining,
Artificial Intelligence	AI, Machine learning, Deep Learning	AI, Machine learning, Deep Learning, ANN

II. LITERATURE REVIEW

A. ARTIFICIAL INTELLIGENCE

The application of AI in the food sector is getting progressively significant owing to its capability to assist in minimizing food wastage, improves production hygiene, enhances the cleaning process of machines, disease, and pest control; therefore, there are numerous instances of employing AI and ML in the agri-food industry [7]. Automated frameworks can collect a huge amount of data in a matter of a few seconds on a single food item and analyze it rapidly. Even though agriculture practice is broad, some major areas of the agriculture sector where AI finds its application such as supply chain management, soil, crop, diseases, and pest management. [8] summary all the proposed models using AI techniques with their limitations (a) for soil management: Fuzzy-logic based SRC-DSS (Soil Risk Characterization Decision Support System) [9] for soil classification, MOM (Management-oriented modeling) [10] for minimization of nitrate leaching, ANN (artificial neural network) [11] to estimate soil enzyme activity and soil structure classification, etc., (b) for crop management: CALEX [12] to formulate scheduling guidelines, PROLOG [13] to remove redundant tools from the farm, ANN [14] to detect nutrition disorders in crops, ANN [15] to predict rice yield accurately, etc., (c) for disease management: computer vision system (CVS) [16] to detect multiple diseases at high speed, Fuzzy logic based database [16] are accurate in test environments, ANN-GIS [17] has got an accuracy of 90%, the expert system using rule-base in disease detection [18] for faster detection and treatment of disease, etc., (d) for weed control: invasive weed

optimization (IWO) [19], big data abased ANN-GA [20], support vector machine [21], etc. All these methods did not consider all the parameters; they are all application-specific towards a particular crop or environmental parameter. There is a need to design AI frameworks using multiple parameters and that can be used for multiple crops.

B. BIG DATA ANALYTICS

Big data analysis is outlined as a system in which cutting-edge analytic methods operate on huge data sets. Therefore, it is a combination of two technical entities massive amount of data sets, and a collection of analytics tool categories including data mining, statistics, AI, predictive analytics, natural language processing (NLP), etc. forming an important component of business intelligence. Lately, big data turn out to be a subject of broad and current interest equally in academic research and industry. It characterizes enormous and unstructured data generated by a large number of sources. Several out of the most prevalent data processing techniques employ big data techniques. Big data is depicted by the subsequent attributes which are shown in figure 2. Big data is being used in numerous fields such as big services business industries like Amazon to learn customer behavior and needs more precisely to tailor product prices accordingly, enhance operational productivity, and cut down personal costs. Even social networking sites Facebook, Twitter, and other networking sites utilize big data analytics to study your social behavior, interests, and social connections and then endorse the specific products. In an intelligent transportation system, big data techniques can handle the enormous quantity of diverse and complex data generated over the period to provide safe and superior facilities aimed at drivers and passengers in the transportation system. In the agriculture field, big data shows a huge potential for solving many challenges of farming and consequently boosting the agriculture production quality and quantity. Big data analytics can be used to determine the soil quality, diseases and pest interruption, water requirement, and can predict harvesting time for crops.

There has been a critical pattern to ruminate about the utilization of massive data procedures and strategies to agribusiness as a significant opportunity for utilization of the ICT pack, for financing, and for achieving added significance inside the agriculture sector [22] [23-26]. Applications of massive data in agriculture are not sternly regarding primary cultivation, but also assume a significant part in enhancing the effectiveness of the whole supply chain, thus reduces food security worries [23] [24]. Right now, discussion on the applications of big data present in literature is occurring fundamentally in America [27], Canada [28], Europe [29], and China [30] [26]. Considering the developing consideration and interest that appeared in the different works, nonetheless, the number of use cases are likely to grow promptly in different nations like Australia [31], Morocco [24], etc. Massive data is the center of in-depth,

progressed, game-evolving business intelligence, at a scale and speed that the old methodology of duplicating and removing every last bit of it into a warehouse is not, at this point appropriate. Prospects for Big Data use in agribusiness incorporate benchmarking, IoT-based sensor network implementation and analytics, prediction models, and utilizing enhanced models to oversee crop failure risks and to lift feed efficacy in livestock farming. Thus, big data technology is to offer prescient insights to upcoming farming outcomes, enables real-time effective decision making, and modernize business measures for rapid, state-of-the-art actions, and game-changing business models [32]. Big data is

predicted to modify both the degree and the organization of agriculture [33]. While there are questions about whether agriculture practitioners' information is going to be supplanted by algorithms, applications of Big Data are probably going to change how agriculture farms are managed and operated [34] [35]. Key domains of progress for precision agriculture such as real-time forecasting, tracing of agri-food products, and remodeling of business practices [36]. More extensive big data application is probably going to transform both farm organizations and the more extensive supply chain in unfamiliar ways.

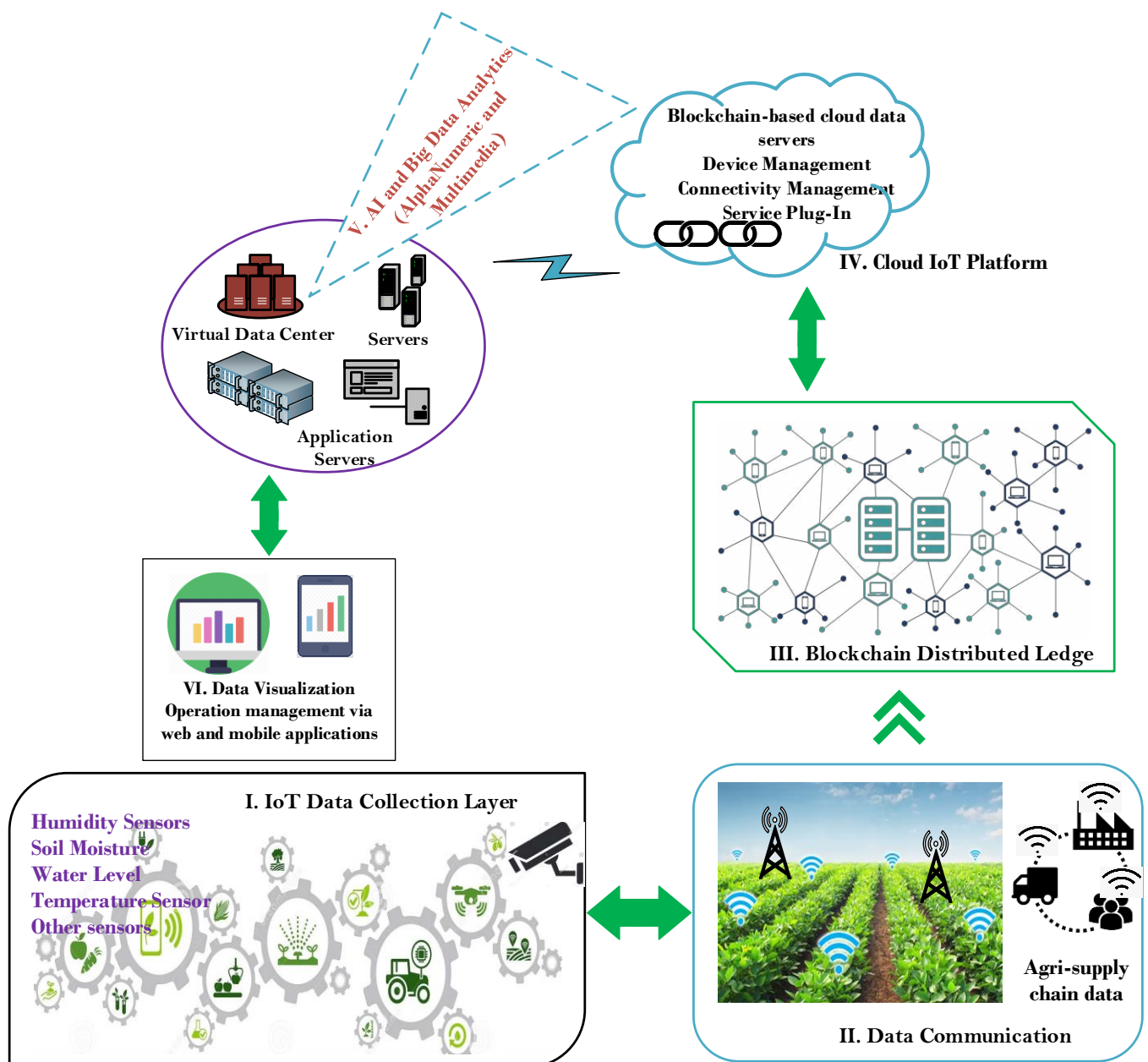


FIGURE 1. Illustration of smart agriculture ecosystem



FIGURE 2. Big data characteristics (10 V's)

III. COMPARISON AND DISCUSSION ON MACHINE LEARNING TECHNIQUES IN AGRICULTURE

There is a broad literature on different machine learning algorithms that have been employed in diverse application areas in agriculture. Identifying the ideal method for guaranteeing accuracy and constancy for a specific application in agriculture is significant. SVRs demonstrated robustness with outliers and noise presence with better estimation accuracy upon comparison with ANN [37]. ANN and SVRs when used for mapping of soil organic stocks(SOC) produced comparable performance [38]. Several regression models were evaluated to find appropriate techniques that realize great accuracy and better generality for yield prediction abilities. Neural networks, despite their site-dependency, ascertained robustly, however, the SVR model employed was highly accurate though being fast computationally [39]. ANNs, RFs, and SVMs have mostly been testified as classifiers, yielding great accuracies [40] [41]. Deep learning techniques are the utmost promising models for segmentation applications of agriculture image data sets. Finally, in Table III we discuss the reliability issues, computational characteristics, and threads of analysis of the models explained.

TABLE III. COMPARISON AND DISCUSSION ON MACHINE LEARNING TECHNIQUES

TECHNIQUE	COMPLEXITY	CHARACTERISTICS & LIMITATIONS	CURRENT APPLICATIONS
ANN	High	High power consumption; Susceptible to overfitting; Require large datasets; Parameter tuning procedure is time-consuming	Pattern classification and attribute mapping; Crop estimation; Soil properties analysis and estimation
DT's	Low	Sensitive to trivial variations in training data; Unstable occasionally; Susceptible to overfitting	Classification; Crop yield from soil variables
SVM'S	Low	Fast and accurate; Easy to implement; Require small training samples; Robust to noise in training data; Mostly used for classification; Less susceptible to overfitting	Classification of diseases; Soil mapping; Retrieval of vegetation attributes; Estimations
RFs	High	Overfitting problem; High efficiency; High prediction performance; Small training time; Easy parametrization	Attribute mapping; Classification and regression
DLs	High	Highly promising for agriculture applications; High accuracy; Require large datasets; High computational cost	Classification and regression applications; Measure features; Yield estimation, etc.

ANN = Artificial Neural Networks, DT = Decision trees
SVM = Support vector machines, DL = deep learning, RF= random forest

Graphical models are not apprehensive towards input-output pattern modeling; however, they model autocorrelation between the input parameters (variables) [42] [42]. They exist several variants of GMs that model data based on input-output interdependencies, for instance, conditional random fields (CRFs) [44] [45]. CRFs take up probabilistic systems unlike NNs and SVMs; i.e., they form input-out variable relations by a probability distribution $P(Y | X)$, instead of a definite capacity of the configuration $Y = f(X)$. Though, having a similar fundamental principle; the models learn the $P(Y | X)$, upon provided with training data (X_i, Y_i) , $i=1, 2, \dots, n$ by optimizing a proper loss function. Upon learning the best probability distribution from the training data, inference can be performed to estimate the target output. In agriculture, GMs show a substantial potential for modeling big data applications such as spatial disease distribution [46], operational decisions, modeling traits, etc.

The most specific challenges that occur with ML in precision agriculture are variable spatial-temporal resolutions and missing due to several reasons like IoT device malfunctioning, communication failure, bad weather prevented remote sensing image acquisition, etc. It is in this way important to have AI models that can adapt to missing information. All the recent ML and DL models designed for plant disease and pest detection are not suitable for the early detection of diseases and pests, thus unable to prevent the crops from early disease and pest attacks. Thus, deep learning models for the early classification of plant diseases and pests are important.

DL and CNNs have been progressively more employed in agriculture remote sensing applications [47] [48] [49]. CNN needs a hefty volume of data towards generating hierarchical features to make available semantic statistics at the output [50]. With the expanding access to enormous quantities of aerial images from unmanned aerial vehicles (UAVs) [51] and satellites, CNNs can assume a significant part in the analyses of all this information to extract significant information. Though, the UAV-based technological adaption by farmers for specialty crops is very low [52]. There are two facades for low adaption such as preprocessing and analysis of data, as it is very complex and time-consuming to produce precise and suitable information and the inability of available commercial tools to create enough useful information from the data for specialty crops.

As UAVs can accumulate a large and unstructured quantity of data, big data-based tools (analytics tools) and cloud computing has the potential to enhance the data processing efficacy, offer high data security, and scalability, and minimize cost. Applications based on cloud computing act as a potential solution having low upfront cost, proficient utilization of computational resources, and service costs [53]. UAVs with big data analytics methods such as CNNs can be used to detect tree characteristics (height, tree health, species, canopy area, etc.), leaf disease, crop estimation, etc.

Soil quality perform a significant part in influencing how healthy plants develop. Indeed, different types of plants grow in different soil conditions or types. Understanding the diverse characteristics such as texture, structure, and chemistry of soil assists agriculture practitioners to choose the best quality crops to cultivate in their farms. To study these characteristics of soil, IoT, and other sensor networks along with ML-based big data techniques like clustering and classification methods to label soil data. Spark Mlib comprising of several ML algorithms and utilities including logistic regression, and naive Bayes in classification, K-means, GMMs in logistics. Likewise, distributed parallel association rule mining techniques can be used to determine the growth of plants.

IV. BIG DATA OPERATING CYCLE IN THE AGRICULTURE ENVIRONMENT

The above discussion on the existing work on smart agriculture and the potential of integrating evolving technologies namely AI and big data to bring revolutionary changes, benefits and solve many problems of sustainable agriculture. In the technologically advanced big industrial farms, field management looks different from the traditional farms following the operating cycle represented in figure 3. In these smart farms, the management system employing smart decision-making by processing actual field data by deriving benefits of its inner variability (including both time and spatial-wise). The objective field data acquisition is done by deploying IoT devices, remote sensing, and other sensor networks. The data collected about soil, crops, weather, or ambient from the IoT sensor networks is stored on local or cloud storage. Where ML-based big data algorithms are used to abstract vital information and helping incorrect decision-making by the farmer. Finally, the required action recommended by the decision system is executed physically by the advanced machinery used based on the decision received through an intelligent control system. This cycle of processes remains repeating systematically till the harvesting stage.

The ML-based big data analytic tools [54] such as ANN, PROLOG, TOMRA, etc., at the processing stage, are used to abstract the significant information about the quality of soil like nutrient level and pH, analyze seed characteristics, sorting of food, the weather patterns, and the existence of food hazards by relating biotic or abiotic data with development and probabilistic existence of pathogens, pests, and toxicants. The different stages of this precision agriculture system are discussed in successive paragraphs below:

There is always assumed some level of spatial variability naturally for crop fields to existing irrespective of what way the crops will be managed. The weather of a current production year and the previous successive years influence the natural spatial variability, data from the previous numerous years can be fused to determine interested parameter trends, and therefore data turn out to be important for farm management. Thus, the need for crop monitoring

arises from the presence of variability, however, these variabilities require to be managed by a grower efficiently. Management zones having homogeneous characteristics are developed in a field to custom-made field practices for each subfield zones, ensuing in a realistic and economical method to precision agriculture. Thus, the implementation of subfield zones would cut the fertilizing cost, yield improvement, lessens pesticide usage, helps in building superior farm records, and make available vital data for decision making. The size of the field decides the natural variability function and subsequently determines the size of these management zones and management factors. The selection of several specific parameters to be pursued should be done at the early stage of the process. However, some use cases with very low spatial variability might be there in which a distinct mapping event can be satisfactory.

Different IoT sensors apart from traditional sensors are used for crop monitoring and collecting required important data from them. These sensor devices can be directly deployed in the fields, agriculture robots, autonomous platforms, machines, or weather stations. Different parameters can be calculated in real-time by using IoT sensor networks [55] enabled with a high-speed data network. Remote sensing from artificial satellites has performed a vital

part in the development of precision farming by making field data remotely accessible. American Landsat satellites, the European Sentinel 2 system, RapidEye constellation satellite system, GeoEye-1 system, and WorldView-3 are the important satellites supplying agriculture [56] data in the form of multispectral data, multispectral RGB imagery, RGB and NIR data, etc. The application of unmanned aerial vehicles (Drones and Remotely-piloted Aircrafts) in agriculture production is gradually increasing as a measure of an efficient method to sustainable agricultural management allowing growers, agri-engineers, and agronomists to assist simplify their procedures, utilizing robust information analytics to achieve valuable insights into their crops. Drones have made careful crop monitoring easier over large areas of agricultural lands, in identifying suitable crop recommendations, the emergence of plant and population, as more precise data can assist in decisions regarding replanting, pruning, and thinning activities, and yield estimation. UAV's [51] are very useful but still face certain challenges such as carry a limited payload, limit the use of sensors onboard, challenging data and image post-processing, vegetation shadowing during gathering imagery data, etc.

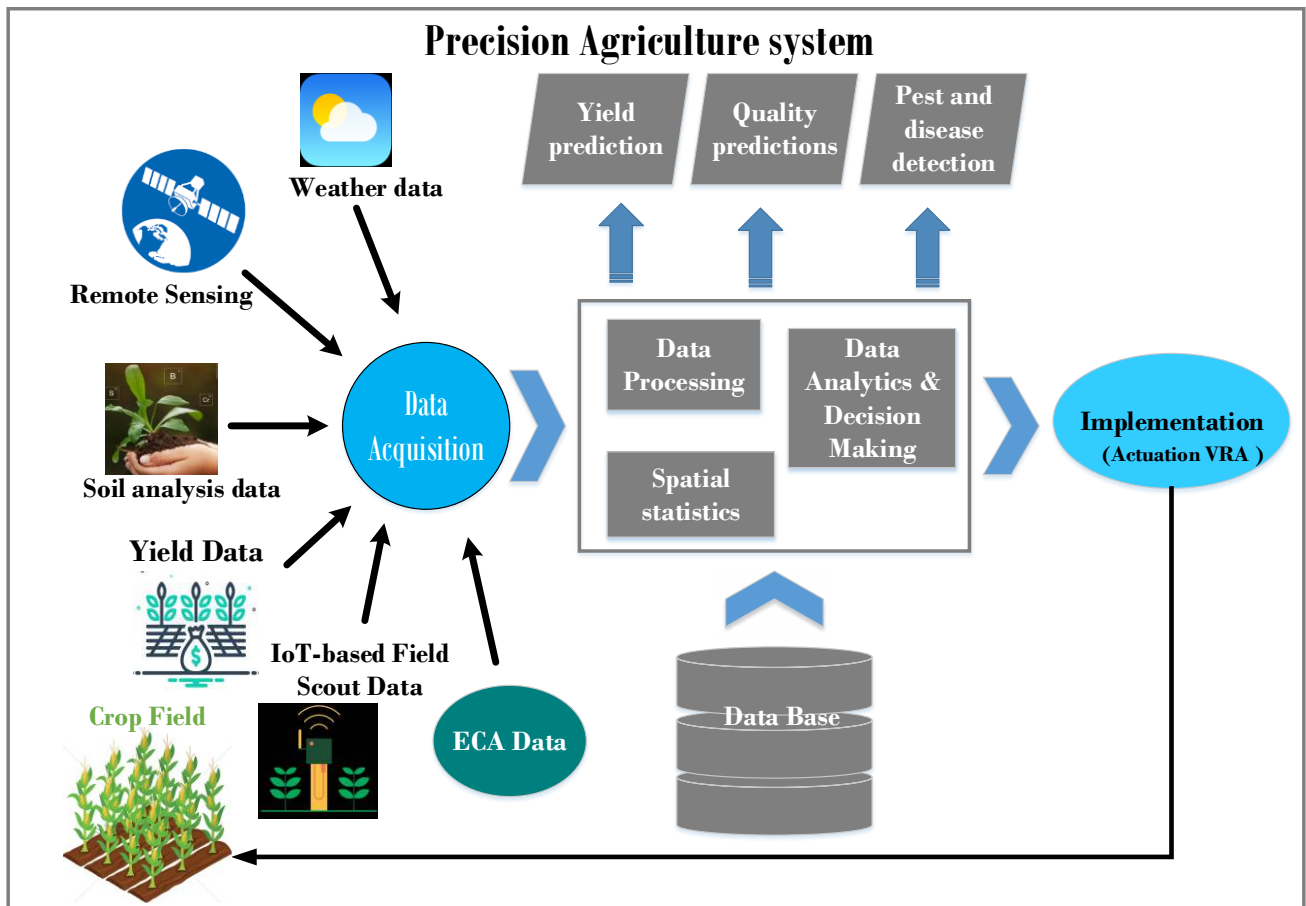


FIGURE 3. Big data-based precision agriculture system representation

In proximal sensing, the ground platforms such as Unmanned ground vehicles (UGV) and robots that operate close to crops increase the accuracy of acquired data and one or two high resolutions of samples per unit area are reasonable [57]. With UGVs applications requesting real-time data like weed detection and removal, selective pesticide spraying, soil analysis, pest control, and crop scouting are possible. Scouting robots are used for performing specific tasks such as robot Oz (mechanical weeding), GUSS autonomous sprayer, RowBot system (fertilizing, mapping, seeding, etc.), VineRobot (vineyard management), etc. Researchers and industries are working on different projects to converge UAV and UGV for better sustainable development.

The application of different wireless data collection technologies has created massive data in agriculture. But the huge quantity of data poses a significant challenge to manage, as important information may be imperceptible by noise. Presenting information in a coherent shape is vital for end-users to comprehend the different processes in the field [58]. Mapping is the most useful technique to express spatial developments and homogeneous subfields from the agriculture data. Maps assist in creating management zones with interesting parameters for the efficient application of custom-made field practices for each subfield zones. Kriging is a commonly used interpolation method to get manageable size subfields. Considering the enormous quantity of data generated by smart agriculture, there are several software applications employed to handle interpolation. Local Tangent Plane (LTP) [59] featuring Euclidean geometry, enables a user to establish origins and utilizes intuitive coordinate set up east-north. The systematic quantization of the LTP coordinates is allowed by grids in maps in the efficient management of agriculture production data, enables the data sharing among succeeding seasons, and different field parameters of a management zone.

Software-based farm management solutions [60] for instance Geographic Information Systems (GIS) encourage the automation of data collection and analysis, supervising, planning, record keeping, decision making, and farm operation management. These tools also help in basic tasks for record-keeping such as produce harvests and yields, scheduling farm tasks, profits-losses, tracking of soil nutrients, weather prediction, and mapping of the field, and other complicated functions for automating field management. A specific GIS data management system [61] known as farm management information system (FMIS) was developed for different applications of precision agriculture. FMIS helps growers with several tasks like operational planning, record keeping, implementation, and evaluation of executed fieldwork. Its objective is to decrease costs of production, comply with farming

standards, great product quality, and safety maintenance, and supervising the farmers to make the best decisions. There are several commercial agriculture information management systems for instance ADAPT, WinGIS, SpiderWeb GIS, AGERmetrix, FieldView, SST software, AgVerdict Inc., Trimble, etc. addressed not only for farmers or producers, nonetheless to other participants' in the agriculture sector i.e., in the agriculture supply chain from farm to fork. However, the efficiency of recommendations from these software tools depends upon the parameters encompassed in the design of algorithms of that particular software platform. For example, DSSAT produces outputs by taking experimental data for crop model evaluations, permitting users or growers to compare the simulated results and observed results, which is significant if real-world decisions are established on modeled outcomes. There are other wide varieties of big data analysis software tools available in agriculture [62] (see Figure 4).

It is practically difficult for people to manage complex agriculture data to make better decisions due to several field parameters involved in farm management. In such scenarios, AI with DL, genetic algorithms, ML, or expert systems can assist with its reasoning, and modeling abilities can perform a vital role in precision farming, facilitating to understand of all the available data. Therefore, precision agriculture presents a huge application space for all types of core technologies in AI because agents functioning in uncontrolled situations. A fuzzy logic-based decision support system is designed in [63] for kiwi, potato, and corn with input variable parameters as rain forecast and soil moisture. Likewise, [64] developed a DSS to estimate the weekly irrigation requirement for citrus orchards by taking soil moisture and climate data; it uses a real-time soil parameter measurement control system to avoid errors. DSS is the most robust and dependable by considering several parameters, however, some processes stay controversial as different objectives can result in diverse solutions at different times depending on the need set by users or others engaged with the procedure. Numerous DSS systems have been proposed in the literature considering different use cases having different objectives. Thus, the use of DSS tools is influenced by their usability, performance, cost-viability, significance to growers, and suitability with compliance requests. The application of software tools for decision-making in precision agriculture is considered valuable as these tools enhance management efficiency than other tools. Though, there is yet far to make innovation-based tools adequately attractive, simple, intuitive, and nice for farmers to adopt. On the other hand, producers need to be trained appropriately until these technological tools can be easily managed.

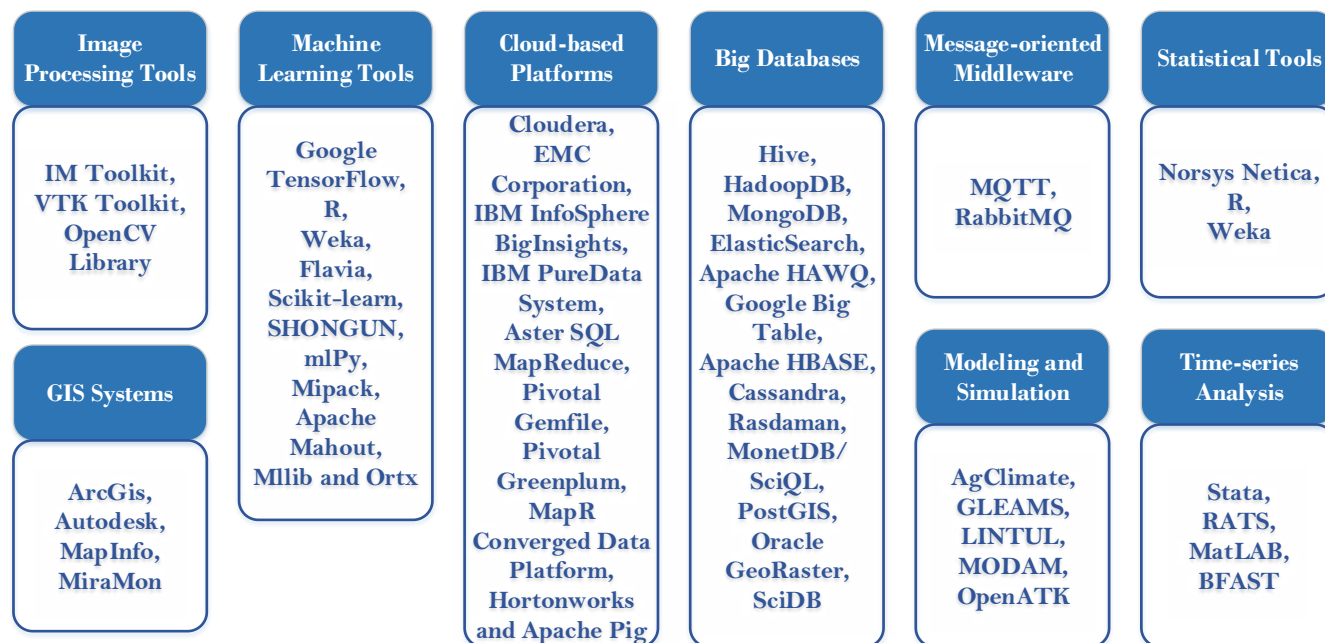


FIGURE 4. Software tools employed in agriculture big data analysis

Actuation on the crop is the last step to complete the crop management cycle as shown in figure 3. It is done by considering the recommendations of the DSS through advanced equipment's/machines able to receive data signals from the control unit. Variable-rate machines are capable of executing several agriculture errands operated by automatic systems. The application of variable rate technology (VRT) for site-specific crop management (SSCM) can improve profit and diminish the environmental impact by executing tasks precisely. Delineation technique usage in management zones can increase the efficiency of farms for instance applied delineation methods for variable rate nutrient use cases, which enhances the farm efficacy than traditional uniform-rate use cases, and impact on the environment was minimized. Different machinery manufacturing companies such as CLAAS, CEBIS MOBILE ISOBUS, etc. are developing various VRT based commercial solutions to perform different applications of precision agriculture. Variable-rate harvesting (VRH) or automatic differential harvesting is the other type of variable actuation, that tries to harvest according to initially specified management zones. Other than efficiency and usefulness, the cost is one of the vital parameters to reflect for the acceptance of these technologies. Thus, the pervasive availability of economical electronic components will favor the adoption of these digital applications throughout the world including small farm holders.

V. PROSPECTUS OF BIG DATA AND AI IN PRECISION AGRICULTURE

A. BIG DATA-BASED DECISION SUPPORT SYSTEM FOR CROP SELECTION

The proposed system architecture maintains the data collected at every stage of agriculture production and supply chains such as soil moisture, weather, and environment data, crop yield and harvest, demand and supply data from the supply chain, food processing data from food processing industries, and pesticides used by the farmer. Figure 4 represents the overview of the decision support system for crop selection in the proposed system. This data is at cloud storage and other local databases, are used to abstract the relevant information about the quality of soil like nutrient level and pH, analyze seed characteristics, sorting of food, the weather patterns, marketing and trade management, and the existence of food hazards by relating biotic or abiotic data with development and probabilistic existence of pathogens, pests, and toxicants. Big data analytics recommend the best appropriate crop for agriculture practitioners to select for which there will be demand. The system keeps track and maps the crops with the corresponding demand and prevents farmers from overabundance harvesting of the crops with the predicted demand.

B. CROP MANAGEMENT, GROWTH MONITORING, AND PRODUCE QUALITY

The raw estimation of vital parameters from the farm data necessities to be processed efficiently to convert numbers and pictures into beneficial info. The growth and quality of the crops can be monitored from the image data collected from the farm and then applying different image processing techniques such as OpenCV, Matplotlib, Scikit-image, etc., in Python. We can measure the height and width of plants along with the quality of the crops by extracting the different features of crops based on the color of the leaves and crops from the images. In the meantime, other IoT

devices are used to collect all other environmental parameters and are stored in the cloud. IoT networks deployed in the food processing industries collect information about each step in food production from the quality of raw material used to other ingredients used to produce the final food product. All these values can be stored in the network in any file format such as CSV file format, etc. After initial analysis and determination of how indispensable the extracted features are. We can employ some of the abstracted data to train our machine learning or deep learning models. Then these trained deep learning algorithms/models are used to extract the required features from large data sets collected from the field.

The crop yield estimation targets to analyze factors that affect and influence the production, like irrigation, natural soil composition, and its physical structure and topography, climate and weather, crop stress, crop diseases, and pests, etc. It facilitates efficient management of resources; a timely and precise estimation of products can offer a reliable base to decision-makers to ascertain if there will be a scarcity or excess, therefore, to respond appropriately according to the conditions.

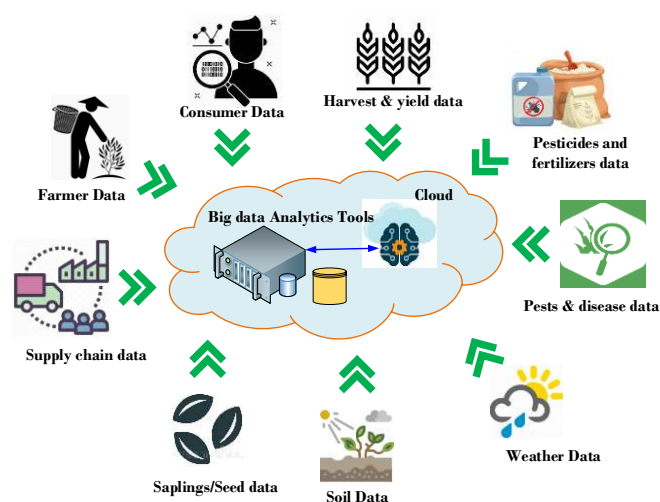


FIGURE 4. Big data-based decision support system

C. SUSTAINABLE USE OF RESOURCES

Driving advancements in the technologies, such as AI, IoT, and drones using big data in its processes to increasingly enhance sustainability in agriculture. Since the volume of arable land is not increasing, the groundwater levels are going down, and the soil quality is not increasing these technologies could make sure optimal utilization of arable land, water, and other resources to meet global requirements at the same time conserving resources for the future generations. Thus, big data have the potential to provide solutions (practical and scalable) that can assist in natural resource conservation, thus could sustain agriculture.

D. REDUCE PESTICIDE USAGE

With the implementation application of computer vision, machine learning, and robotics, agriculture practitioners can employ AI to manage weeds. From the data collected from the farms with the help of IoT devices, AI can assist in data abstraction to locate the weeds in fields and spray only to the specified locations where the weeds are. Thus, reduces the consumption of pesticide spraying an entire field and leaves less chemical on the agriculture produce relatively the amount of chemicals typically sprayed.

E. PLANT DISEASE DETECTION

Agriculture production and quality face a major threat from plant diseases. Recently, several deep learning-based neural networks have been developed to identify plant diseases but all the models work when the diseases are fully developed in the crop and thus have minimum effect on increasing the quality of the produce. In the fields, plant disease detection and monitoring at early stages are difficult, time extensive, and costly. But to increase the crop quality and yield, deep learning algorithms for timely classification and recognition of disease are required. This can be achieved by preparing plant pathology datasets with diseases in the early stages.

F. SYSTEM FOR RISK MANAGEMENT

Managing risks because of the farm location, type of soil, and mostly to heat stress or freeze involves essential significance in precision agriculture. A particular condition for cultivation is the impact of the climate and particularly its volatility. The amalgamation of different datasets is a critical process for data interpretation for this use case. Provincial climate patterns are utilized to join data from global patterns with local and provincial meteorological histories to give climate information to more modest spatial units and support real-time adoption of the environment and climate changes. For example, [65] discussed a situation of big data employed in the forecasting of rainfall by taking benefits of meteorological big datasets. The outcomes show a considerable prospective of data fusion in precision agriculture.

G. AGRICULTURE MANAGEMENT SYSTEM

ICT empowers farmers to share data, set up collaboration, and work together. As agriculture practitioners become connected, software-defined management frameworks arise. Rural administration frameworks emerge to give accounting administrations, linking growers with farm owners and administrators, and offer benchmarking capacities to agriculturalists by associating them. Their point is to support farm administrators and agribusinesses throughout the world, incorporate, and examine a tremendous quantity of information from real-time sources to help their decisions in business. Such frameworks give smart cultivating solutions. Smart cultivating is a term that broadens precision agriculture by putting together administration errands concerning handle explicit information as well as information upgraded by setting and circumstance awareness, set off by real-time events [62]. Taking examples of studies conducted on small farms in the developing world specify that farmers are unable to trade

their produce because of oversupply or inadequate information. Tools for higher productivity and require estimations can facilitate crops to be incorporated into the global supply chain [63].

V. BIG DATA CHALLENGES IN PRECISION AGRICULTURE

Gathering and examining huge data produced via IoT networks and wireless sensor networks, comprising digital images and more data from UAVs, satellites, and data fusion with existing data present difficulties to the effective execution of smart farming. Arising technologies about data mining techniques and artificial intelligence techniques are potential methods to achieve intuitions from said information [64] - [69]. These techniques can assist with dissecting greater and more unpredictable information, uncover covered-up models, and uncover trends quickly and precisely. The capability of these strategies in massive data investigation has not been sufficiently valued in farming for various reasons inspected underneath.

The greater part of the accessible open frameworks referenced already result from the latest projects, their issue is still extensively wide embraced, to conclude the final accomplishment. Large numbers of them may in any case be underdeveloped and have not attained their maximum capacity yet. Most of these applications of big data are suited for large industrial farms (such as Monsanto) that now employ big data in process of decision making and have the infrastructure to access data, resources, and most importantly access to finance [70]. There is very little work done on small farms around the developing world. Big data has the potential to support non-industrial farms, however moral and ethical queries regarding availability, cost, and funding need to be addressed to attain these advantages. If this trend continues, the benefits from data-driven precision agriculture will remain available to only big industrial farms.

A. DATA COLLECTION CHALLENGES

In precision agriculture use cases, enormous data is generated from different sources. Merging data from an assortment of sources raises worries about the issue of information quality and information merging, and the access to gathered huge information raises apprehensions about safety and protection.

Data-driven techniques request uncontaminated and applicable information be used. Incomplete datasets obliterated information, and the presence of exceptions or inclinations in the training set influences model precisions. The analysis of information quality requests critical human contribution and expert knowledge. However, even semi-computerized approaches are not useful concerning huge volumes of information. The act of large information assortment likewise increases apprehensions over availability and security. The capability of academicians and researchers to perform big-data-oriented examinations firmly relies upon the accessibility of farm data. Suggestions for administering security, ownership of data,

protection of data, and information use ought to be set by farmers' coalitions and farming technology suppliers.

B. CHALLENGES IN BIG DATA ANALYSIS TECHNIQUES

Big data requires extraordinary methods to proficiently process a huge quantity of data with infinite running time. Hypothesis analysis and ML are the largely employed methods for data exploration [71]. Agricultural data analysis is mostly statistical. AI techniques do not consider any pre-established relations among variables from the hypothesis however start from the data to look for potential connections between variables [72] [73].

The gathered datasets are huge and complex making it hard to manage normal AI procedures. Such methods regularly perform inadequately when applied to agricultural data. Scalable and versatile methods are expected to adapt to voluminous information. Besides, enormous information gathered in agriculture disrupt normal suppositions basic a few AI and analytic techniques, for example, the independence and identical distribution of data. Big data generated from farms shows spatial-temporal autocorrelation, has heterogeneity and high dimensionality, is nonstationary, and as a rule, must be handled in a constant manner [67] [74].

To control the datasets that go along with precision or smart farming, analytics techniques need to an extent in aligned and distributed means, high computational complication. Technological developments in cloud computing capabilities and distributed storage models can assist in this course. Distributed computing may be employed to incorporate data sources in various areas, and afterward, the data can be apportioned into an appropriated and parallel model. The integration of AI and distributed computing execution procedures gives potential approaches to deal with huge data. Established models, ought to remain compatible with distributed computing, but not all AI models are suitable for execution in distributed form. As an effective model, parallel SVM (PSVM) [75] decreases memory and time utilization and in [76] a versatile AI administration is presented for stream handling and real-time processing.

C. AVAILABILITY OF COMPUTING INFRASTRUCTURE

Apart from novel analytical standards for information abstraction from big data-compatible distributed frameworks and advanced wireless communication solutions are also required to implement big data in precision agriculture. Management of farms faces several challenges in executing real-time analysis and delivery of heterogeneous and multi-dimensional data channels from various sensor networks [77] Platforms for real-time data analysis are required to manage data collected from remote sensing online and fuse offline data with it from other distributed data sources. Since precision farming greatly depends on event monitoring which necessitates data analysis and therefore needs lesser latency and greater bandwidth. Hadoop (open source) is right for parallel processing and applications for execution of cutting-edge

analytics on enormous data stored have mostly developed on Hadoop (Apache Software Foundation, 2019b). However, Hadoop is inappropriate for real-time data processing applications. Apache Storm, Spark, and Flink are appropriate for real-time data stream processing. Several modules such as Mlib and GraphLab for providing ML operations, while Tensorflow like tools are intended to develop sophisticated ML models such as CNNs, and DNNs.

D. MANAGING GROWING DATA AND REAL-TIME SCALABILITY

An immense quantity of images and video is produced progressively through several devices during plant growth monitoring, which prompts several challenges in storing and processing all that data. Moreover, most of the data generated in agriculture are amorphous or semi-structured, not stable for storing in customary databases like MySQL, SQL Server. Management of such enormous unstructured data is considered a big challenge.

It is imperative to give consumers visual data in a real-time mode to empower them with efficient fast decision-making capabilities. This demands advanced big data platforms with real-time data handling capabilities across the network stages such as collecting, processing or analyzing, and visualizing. But real-time analyzing such an enormous amount of data is challenging near the source due to inefficient computing infrastructure.

E. DATA MANAGEMENT LANDSCAPE UNCERTAINTY

One disruptive aspect of massive data is the utilization of variability of advanced data management techniques whose intentions are to strengthen operational and analytical processing significantly. These methodologies are generally grouped into the NoSQL framework category that is distinguished from a traditional relational storage management system. There are various NoSQL methods. Some support a hierarchical object interpretation employing standing encoding techniques like XML, BSON, or JSON related to respectively managed data entities. While others are using the key-value concept of data storage, fundamentally supporting a schema-less model. Databases based on graphs maintain the interrelated relationships between various objects. And several other standards are evolving continuously.

In fact, inside every one of these NoSQL classifications, many models are being created by several organizations both for-profit and nonprofit. Each approach is fit diversely to key performance measurements—a few models give extraordinary adaptability; others are excellently scalable as far as execution while others support a more extensive scope of functionality.

To be specific, the wide assortment of NoSQL tools and designers, and the market status impart a prominent level of uncertainty to the data management landscape. Selecting a NoSQL technique is difficult, but choosing the wrong data management framework can result in enormous errors and loss if the selected NoSQL tool from a particular organization doesn't satisfy the expectations or if a different

data management system is adopted by a third party for application development. Therefore, to select big data management techniques users need to consider their respective applications and performance requirements along with the mitigation of risks of the underlining technology.

VII. CONCLUSION AND RECOMMENDATIONS

The ever-growing accessibility of information through developments in ICT appears promising for improving innovations on indispensable decision-making through enhancing precision and generalization capability of models. Besides, learning from the enormous quantity of data generated from precision agriculture practices is anticipated to create substantial opportunities and transformational perspectives for precision farming. With the advancement in big data, traditional learning methods are not naturally proficient or scalable adequately to process huge quantities of heterogeneous, multi-dimensional, and spatiotemporal data. Innovative ML techniques such as CNN, big data analysis methods present higher precision, flexibility, vigor, and performance. We have provided a comparison and discussion on the different ML techniques in precision agriculture.

Agriculture production challenges are growing, creating the necessity to comprehend the complicated agriculture environments more crucial than ever before. Several ML techniques due to their data mining capabilities from agriculture data are extensively being employed in smart farming. Several challenges facing by big data and AI in precision agriculture are classified appropriately.

Automation and application of AI, drones, IoT, robots, and big data are anticipated to perform a significant function in various agriculture areas in addition to precision farming. Employment of high-performance data-driven scalable learning methods provides better real-time decision-making capabilities and automates various agriculture processes, and thus can transform conventional farm management into artificial intelligence systems. Emerging domains of cutting-edge ML and data mining converged with accessible datasets and strategy structures are required to act instrumental in addressing the challenges of agrarian production regarding sustainability, efficiency, climate change, and food security.

A data-driven system benefits every single stakeholder engaged with the agriculture business right from agriculture practitioners (farmers) to consumers, financial institutions, food processing industries, and several others. Even though the best of its capabilities are still unexplored that it has to offer for value generation, it has already begun to get enormous revolutions in the agriculture industry. Some of the various benefits AI and big data offers include the development of healthier and superior products because of the availability of the new plant genome sequencing techniques, precision agriculture methods help in inferring conversant decision making, and the utilization of IoT sensor devices and analytics techniques help in thwarting the food wastage and food-borne diseases.

REFERENCES

- [1] B.-Almeida, M. C.-Rodriguez, T. S.-Cobo, E. F.-Gómez, R. C.-Cabezas, and W. B.-Vera, "Blockchain in agriculture: A systematic literature review," In Proc. Int. Conf. on Techn. and Innovation, pp. 44-56, Springer, Cham, Nov. 2018.
- [2] A. Kamilaris, A. Fonts, and F. P.-Boldv, "The rise of blockchain technology in agriculture and food supply chains," Trends in Food Science & Techn., Vol. 91, pp.640-652, 2019.
- [3] A. Vangala, A. K. Das, N. Kumar, and M. Alazab, "Smart secure sensing for IoT-based agriculture: Blockchain perspective," IEEE Sensors Journal, pp. 1558-1788, Jul. 2020.
- [4] R. Mark, "Ethics of using AI and big data in agriculture: The case of a large agriculture multinational," The ORBIT Journal, 2(2), pp.1-27, 2019.
- [5] T. Mohamed and A. E. Hassanein, "Integrating blockchain and the inter of things in precision agriculture: Analysis, opportunities, and challenges," Computers and Electronics in Agriculture, Vol. 178, Nov. 2020.
- [6] H. M. Kim, and M. Laskowski, "Agriculture on the blockchain: Sustainable solutions for food, farmers, and financing" Supply Chain Revolution, Barrow Books, 2018.
- [7] K. Demestichas, N. Peppes, T. Alexakis, and E. Adamopoulou, "Blockchain in Agriculture Traceability Systems: A Review," Applied Sciences, vol. 10, No. 12, pp. 1-22, 2020.
- [8] A. S. Patil, B. A. Tama, Y. Park, and K. H. Rhee, "A framework for blockchain-based secure smart greenhouse farming," Adv. in Computer Sci. and Ubiquitous Computing, Springer, Singapore, pp. 1162-1167, 2017.
- [9] A. Di Vaio, F. Boccia, L. Landriani, and R. Palladino, "Artificial intelligence in the agri-food system: Rethinking sustainable business models in the COVID-19 scenario," Sustainability, Vol. 12, No. 12, p.48-51, 2020.
- [10] N. C. Eli-Chukwu, "Applications of artificial intelligence in agriculture: A review," Eng., Techn. & Applied Sci. Research, Vol. 9, No. 4, pp.4377-4383, 2019.
- [11] E. M. López, M. García, M. Schuhmacher, and J. L. Domingo, "A fuzzy expert system for soil characterization," Environment International, Vol. 34, No. 7, pp.950-958, 2018.
- [12] M. Li and R. S. Yost, "Management-oriented modeling: optimizing nitrogen management with artificial intelligence," Agri. Syst., Vol. 65, No. 1, pp.1-27, 2000.
- [13] S. Tajik, S. Ayoubi, and F. Nourbakhsh, "Prediction of soil enzymes activity by digital terrain analysis: comparing artificial neural network and multiple linear regression models," Envir. Eng. Sci., Vol. 29, No. 8, pp.798-806, 2012.
- [14] H. Lal, J. W. Jones, R. M. Peart, and W. D. Shoup, "FARMSYS—a whole-farm machinery management decision support system," Agricultural Systems, Vol. 38, No. 3, pp.257-273, 1992.
- [15] H. Song and Y. He, "Crop nutrition diagnosis expert system based on artificial neural networks," In Proc. Int. Conf. on Infor. Techn. and App., (ICITA'05), IEEE, Jul. 2005, Vol. 1, pp. 357-362.
- [16] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," The J. of Agri. Sci., Vol. 145, No. 3, pp. 249, 2007.
- [17] K. Ballela, D. Satyanvesh, N. V. Sampath, S. P. Varma, and P. K. Baruah, "Agpest: An efficient rule-based expert system to prevent pest diseases of rice & wheat crops," In Proc. Int. Conf. on Intelligent Sys. and Control (ISCO), IEEE, Jan. 2014, pp. 262-268.
- [18] G. Liu, X. Yang, Y. Ge, and Y. Miao, "An Artificial Neural Network-based Expert System for Fruit Tree Disease and Insect Pest Diagnosis," In 2006 IEEE Int. Conf. on Networking, Sensing and Control, Apr. 2006.
- [19] M. Y. Munirah and M. Rozlini, "An expert system development: its application on diagnosing oyster mushroom diseases," In Proc. Int. Conf. on Control, Automation and Sys. (ICCAS 2013), IEEE, Oct. 2013, pp. 329-332.
- [20] M. Branch, "A multi-layer perceptron neural network trained by invasive weed optimization for potato color image segmentation," Trends in Applied Sci. Research, Vol. 7, No. 6, pp.445-455, 2012.
- [21] A. Tobal and S. A. Mokhtar, "Weeds identification using Evolutionary Artificial Intelligence Algorithm," J. Comput. Sci., Vol. 10, No. 8, pp.1355-1361, 2014.
- [22] M. Dakshayini, and B. B. Prabhu, "An Effective Big Data and Blockchain (BD-BC) Based Decision Support Model for Sustainable Agriculture System," In Int. Conf. Big Data Innovation for Sustainable Cognitive Computing, Springer, Cham, 2020, pp. 77-86.
- [23] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko, "IoT, big data and artificial intelligence in agriculture and food industry," IEEE Internet of Things J., pp. 1-19, May 2020.
- [24] H. Tarik, H. and O. J. Mohammed, "Big Data Analytics and Artificial Intelligence Serving Agriculture," In Int. Conf. on Adv. Intelligent Sys. for Sustainable Development, Springer, Cham, pp. 57-65, Jul. 2019.
- [25] N. Tantalaki, S. Souravlas, and M. Roumeliotis, "Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems," J. of Agri. & Food Inf., Vol. 20, No. 4, pp.344-380, 2019.
- [26] M. N. Sarker, M. Wu, B. Chanthamith, S. Yusufzada, D. Li, and J. Zhang, "Big Data-Driven Smart Agriculture: Pathway for Sustainable Development," In Proc. Int. Conf. on Artificial Intelligence and Big Data (ICAIBD), IEEE, May 2019, pp. 60-65.
- [27] H. Channe, S. Kothari, and D. Kadam, "Multidisciplinary model for smart agriculture using internet-of-things (IoT), sensors, cloud-computing, mobile-computing & big-data analysis," Int. J. Computer Techn. & App., Vol. 6, No. 3, pp.374-382, 2015.
- [28] M. Stubb, "Big data in US agriculture," Washington, DC, Congressional Research Service, 2016.
- [29] K. Bronson, and I. Knezevic, "Big Data in food and agriculture. Big Data & Society," Vol. 3, No. 1, pp. , 2016.
- [30] S. Wolfert, L. Ge, C. Verdouw, and M. J. Bogaardt, "Big data in smart farming—a review," Agri. Sys., Vol. 153, pp.69-80, 2017.
- [31] Z. Cui, X. U. E, Fei, S. Zhang, X. Cai, Y. Cao, W. Zhang, and J. Chen, "A hybrid BlockChain-based identity authentication scheme for multi-WSN," IEEE Transactions on Services Computing, Vol. 13, No. 2, pp.241-251, 2020.
- [32] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," J. of Big data, Vol. 4, No. 1, p.20, 2017.
- [33] H. Es. Van and J. Woodard, "Innovation in agriculture and food systems in the digital age," The global innovation index, pp.97-104, 2017.
- [34] M. Stubb, "Big data in US agriculture," Washington, DC, Congressional Research Service, 2016.
- [35] X. Pham, and M. Stack, "How data analytics is transforming agriculture," Business Horizons, Vol. 61, No. 1, pp.125-133. 2018.
- [36] B.Devlin, " The Big Data zoo—taming the beasts: the need for an integrated platform for enterprise information," Cape Town, 9sight Consulting, 2012.
- [37] L. Pasolli, C. Notarnicola, & L. Bruzzone, "Estimating soil moisture with the support vector regression technique," IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 6, pp. 1080-1084, Nov. 2011.
- [38] O. Penatti, K. Nogueira, & J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" In Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 44–51.
- [39] G. Rub, "Data mining of agricultural yield data: A comparison of regression models," In Proc. Perner (Ed.), pp. 24–37, 2009.
- [40] R. Chinchuluun, W. Lee, J. Bhorania, & P. Pardalos, "Clustering and classification algorithms in food and agricultural applications: A

- survey,” In .M. Pardalos and P.J. Papajorgji (Eds.), 2009, vol. 25, pp. 433–455.
- [41] I. Nitze, U. Schulthess, and H. Asche. “Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification,” In Proc. of the 4th GEOBIA, vol. 35, 2012.
- [42] M. I. Jordan, “Graphical models,” *Statistical Science*, vol. 19, pp. 140–155, 2004.
- [43] M. J. Wainwright & M. I. Jordan, “Graphical models, exponential families, and variational inference” *Foundations and Trends in Machine Learning*, vol. 1, pp. 301–305, 2008.
- [44] J. Lafferty, A. McCalum & F. Periera, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” In Proc. of int. conf. on machine learning (ICML). Vol. 1, pp. 282–289, 2001.
- [45] M. Wytock & J. Z. Kolter, “Sparse gaussian conditional random fields: algorithms, theory, and application to energy forecasting,” In Proc. ICML, vol. 3, pp. 1265–1273, 2013.
- [46] R. R. Vatsavia, A. Ganguly, V. Chandola, A. Stefanidis, S. Klasky, & Shekhar, “Spatiotemporal data mining in the era of big spatial data: algorithms and applications,” In Proc. ACM SIGSPATIAL international workshop on analytics for big geospatial data, pp. 1–10, 2012.
- [47] N. Kussul, L. Mykola, S. Skakun, A. Shelestov, “Deep learning classification of land cover and crop types using remote sensing data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, 2017.
- [48] A. C. Cruz, A. Luvisi, L. De Bellis, Y. Ampatzidis, “An effective application for detecting olive quick decline syndrome with novel deep learning methods,” *Frontiers, Plant Sci.*, vol. 10 Oct., 2017.
- [49] Y. Ampatzidis, & V. Partel, “UAV-based High Throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence,” *Remote Sens.*, vol. 11, no. 4, pp. 4–10, 2019.
- [50] A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Information. Proces. Syst.* pp. 1097–1105, 2012.
- [51] J. Kim, S. Kim, C. Ju and H. I. Son, “Unmanned Aerial Vehicles in Agriculture: A Review of Perspective of Platform, Control, and Applications,” in *IEEE Access*, vol. 7, pp. 105100–105115, July 2019.
- [52] S. Ghatrehsamani, T. Wade, Y. Ampatzidis, “The adoption of precision agriculture technologies by Florida growers: a comparison of 2005 and 2018 survey data. In *Int. Horticultural Congress, II International Symposium on Mechanization, Precision Horticulture, and Robotics*, Aug. 2018, Istanbul Turkey, pp. 12–16.
- [53] V. Jinesh, “Best practices in architecting cloud applications in the AWS cloud, *Cloud Computing: Principles and Paradigms*, vol. 18, 459–490, 2011.
- [54] K. L. -M. Ang and J. K. P. Seng, “Big Data and Machine Learning With Hyperspectral Information in Agriculture,” in *IEEE Access*, vol. 9, pp. 36699–36718, 2021.
- [55] M. S. Farooq, S. Riaz, A. Abid, K. Abid and M. A. Naeem, “A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming,” in *IEEE Access*, vol. 7, pp. 156237–156271, 2019.
- [56] M. E. Brown, “Satellite remote sensing in agriculture and food security assessment,” *Procedia Environmental Sciences*, vol. 29, pp. 307, Aug. 2015.
- [57] Abbas, Farhat, et al., “Crop yield prediction through proximal sensing and machine learning algorithms,” *Agronomy*, vol.10, no. 7, pp. 1046–1056, July 2020.
- [58] X. Cui and Z. Gao, “A Standard Architecture of Agricultural Big Data for Deep Learning,” *IEEE Int.Conf. Adv. in Electrical Eng. and Computer Apps. (AEECA)*, Dalian, China, Aug. 2020, pp. 908–911.
- [59] Wu, Fan, et al., “Research on Measurement and Control System of Common Parameters of Agricultural Equipment Based on Wireless Transmission,” *Int. J. Agricultural and Environmental In. Syst. (IJAEIS)*, vol.12, no. 2, pp. 73–86, 2021.
- [60] J. Tummers, A. Kassahun, & B. Tekinerdogan, “Reference architecture design for farm management information systems: a multi-case study approach,” *Precision Agriculture*, vol. 22 pp. 22–50, Jun. 2020.
- [61] Liu, Shengping, et al. “Key Technology and Application Platform of Agricultural Economic Spatial Information Service—Chinese Agricultural Economic Electronic Map,” *China’s e-Science Blue Book*, Springer, Singapore, pp. 269–289, 2020.
- [62] A. Kamilaris, A. Kartakoullis, & F. X. Prenafeta-Boldú, “A review on the practice of big data analysis in agriculture,” *Computers and Electronics in Agriculture*, vol. 143, pp. 23–37, 2017.
- [63] E. Giusti, S. Marsili-Libelli, “A Fuzzy Decision Support System for irrigation and water conservation in agriculture,” *Environ. Model. Softw.*, no. 63, 2015, pp. 73–86, 2015.
- [64] H. Navarro-Hellín, J. Martínez-del-Rincon, R. Domingo-Miguel, F. Soto-Valles, & R. Torres-Sánchez, “A decision support system for managing irrigation in agriculture,” *Comput. Electron. Agric.*, no. 124, pp. 121–131, 2016.
- [65] T. Kashiwao, K. Nakayama, S.Ando, K. Ikeda, M. Lee, A. Bahadori, “A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency,” *Applied Soft Computing*, Vol. 56, 2017, Pp. 317–330, 2017.
- [66] S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, “Big Data in Smart Farming – A review,” *Agriculture systems*, vol. 153, pp. 69–80, 2017.
- [67] A. Kamilaris, A. Kartakoullis, & Prenafeta-Bold, “A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agri.*, vol. 143 , pp. 23 – 37, 2017.
- [68] S. Shekhar, Z. Jiang, R. Ali, E. Eftelioglu, X. Tang, V. Gunturi, & X. Zhou, “Spatiotemporal data mining: A computational perspective,” *ISPRS Int. J. of Geo-Information*, vol. 4, pp. 2306–2338, 2015.
- [69] R. Ip, L. Ang, K. Seng, J. Broster, & J. Pratley, “Big data and machine learning for crop protection,” *Computers and Electronics in Agriculture*, vol. 151 , pp. 376–383, 2018.
- [70] A. Weersink, E. Fraser, D. Pannell, E. Duncan, & S. Rotz, “Opportunities and challenges for big data in agricultural and environmental analysis,” *Annual Review of Resource Economics*, vol. 10, pp. 19–37, 2018.
- [71] I. M. Carbonell, “The ethics of big data in big agriculture,” *Internet Policy Review*, vol. 5, 2016. URL: <https://ssrn.com/abstract=2772247>.
- [72] D. P. C. Peters, K. M. Havstad, J. Cushing, C. Tweedie, O. Fuentes, & N. Villanueva-Rosales, “Harnessing the power of big data: Infusing the scientific method with machine learning to transform ecology,” *Ecosphere*, vol. 5 , pp. 1–15, 2014.
- [73] H. Zheng, L. Chen, X. Han, X. Zhao, & Y. Ma, “Classification and regression tree (cart) for analysis of soybean yield variability among fields in northeast china: The importance of phosphorus application rates under drought conditions,” *Agriculture, Ecosystems Environment* , vol. 132 ,pp. 9– 105, 2009.
- [74] K. H. Coble, A. K. Mishra, S. Ferrell, & T. Griffin, “Big Data in agriculture: A challenge for the future,” *Applied Economic Perspectives and Policy*, vol. 40 , pp. 79–96, 2018.
- [75] E. Y. Chang, “Psvm: Parallelizing support vector machines on distributed computers,” In *Foundations of Large-Scale Multimedia Information Management and Retrieval: Mathematics of Perception*, Berlin, Heidelberg, 2011, pp. 213–230.
- [76] A. Baldominos, E. Albacete, Y. Saez, & P. Isasi, “A scalable machine learning online service for big data real-time analysis,” In *IEEE Symposium on Computational Intelligence in Big Data (CIBD)*, 2014, pp. 1–8.

- [77] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, & Y. Zhu, "Big data for remote sensing: Challenges and opportunities," Proceedings of the IEEE, vol. 104, pp. 2207–2219, 2016.



SHOWKAT AHMAD BHAT received his bachelor's degree in Electronics and Communication from Punjab Technical University, Punjab, India, and his master's degree from Lovely Professional University, Punjab, India. Currently pursuing a Ph.D. in High-Speed Network (HSNL) Laboratory, National Tsing Hua

University to pursue his Ph.D. degree under the supervision of Prof. Huang Neng-Fu. His primary research interests include Edge Computing, Blockchain Technology, Artificial Intelligence, Big Data, and Wireless Sensor Networks.



NEN-FU (Fred) HUANG received a Ph.D. degree in computer science from National Tsing Hua University (NTHU), Taiwan, in 1986. He is now serving as the Dean of EECS College of NTHU. From 1997-2000, he was the Chairman of the Department of Computer Science, NTHU and since 2008, he is a Distinguished Professor of NTHU. From

2015-2016, he was the director of the Computer and Communication Research Center (CCRC) of NTHU. His current research interests include Internet of Things (IoT) sensor networks, Smart Agriculture and agriculture blockchain, AI-based crops classification technologies, LPWAN (LoRaWAN/NB-IoT) transmission technologies, and applications, AI-based Big Data Analysis for MOOCs (Massive Open Online Courses), network security, SDN/NFV networks, and network applications/flow identification technologies.

He received the Outstanding Teaching Award from the NTHU in 1993, 1998, and 2008, the Outstanding University/Industrial Collaboration Award from the Ministry of Education, Taiwan in 1998, and the Outstanding IT people Award from ITmonth, ROC in 2002. He received the Technology Transfer Award from the National Science Council (NSC) of Taiwan in 2004. He received the Technology Creative Award from Computer and Communication Research Center (CCRC), NTHU in 2005, and the Outstanding University/Industrial Collaboration Award from the NTHU in 2010. He is the founder of Broad web Corp., (www.broadweb.com) and from 2002 to 2006, he served as the CEO/Chairman of

Broad web, which was acquired by TrendMicro, an international security company, in 2013. He is also the founder of ShareCourse, an AI-based MOOCs platform (www.sharecourse.net), which provides MOOC courses including IoT, AI, Big Data Analysis, FinTech, Robot related courses. Dr. Huang has published more than 200 journal and conference papers, including more than 50 papers in IEEE INFOCOM/ICC/GLOBECOM flag-ship conferences. Dr. Huang also has granted/pending more than 60 patents. He is a Fellow of the BCS (British Computer Society) and a senior member of the IEEE.